



Bases lexicais verbais do português brasileiro

Brazilian Portuguese verbal databases

Roana RODRIGUES*

Marcella LEMOS-COUTO**

Francimeire Leme COELHO***

Isaac Souza de MIRANDA JUNIOR****

Oto VALE*****

RESUMO: Este artigo apresenta o levantamento e a análise de bases lexicais verbais do português brasileiro (PB), que podem ser utilizadas em tarefas de Processamento de Língua Natural (PLN). Para tanto, considerou-se para análise apenas bases de dados com extensão superior a 1.000 lexemas verbais, disponíveis de maneira on-line e gratuita e atualizadas nos últimos 10 anos. Sendo assim, o estudo recaiu sob a análise crítica e comparada de três bases lexicais: VerbNet.Br (SCARTON, 2013), Verbo-Brasil (DURAN; ALUÍSIO, 2015) e VerboWeb (CANÇADO *et al.*, 2018), destacando-se seus pontos comuns e divergentes. Acredita-se que esta pesquisa contribui com a atualização do estado da arte, no que se refere às bases lexicais verbais do PB da última década,

ABSTRACT: This paper presents a survey and an analysis of Brazilian Portuguese lexical verbal databases, which are possible to use in Natural Language Processing (NLP) tasks. For this purpose, we considered for analysis only databases with a dimension greater than 1.000 verbal lexemes, free online access, and updated in the last 10 years. Therefore, the study fell on the critical and comparative analysis of three lexical databases: VerbNet.Br (SCARTON, 2013), Verbo-Brasil (DURAN; ALUÍSIO, 2015) and VerboWeb (CANÇADO *et al.*, 2018), highlighting their commonalities and divergences. It is believed that this research contributes to updating the state-of-the-art, regarding the BP lexical verbal databases of the last decade, in addition to listing future investigations to create,

* Doutora em Linguística pela UFSCar. Professora do Departamento de Letras Estrangeiras da UFS. ORCID: <https://orcid.org/0000-0002-7748-8716>. roana@academico.ufs.br.

** Doutoranda em Linguística pela UFSCar. ORCID: <https://orcid.org/0000-0001-9669-9458>. marcella.couto@estudante.ufscar.br.

*** Mestranda em Linguística pela UFSCar. ORCID: <https://orcid.org/0000-0002-9333-9534>. fcoelho@estudante.ufscar.br.

**** Mestrando em Linguística pela UFSCar. ORCID: <https://orcid.org/0000-0002-4004-3182>. isaacmiranda@estudante.ufscar.br.

***** Doutor em Linguística e Língua Portuguesa pela UNESP. Professor do Departamento de Letras da UFSCar. ORCID: <https://orcid.org/0000-0002-0091-8079>. otovale@ufscar.br.

além de elencar ações investigativas futuras para criação, revisão e/ou ampliação de recursos descritivos linguísticos do PB.

revise and/or extend BP descriptive linguistic resources.

PALAVRAS-CHAVE: Processamento de Língua Natural. Lexicologia. Bases de dados verbais.

KEYWORDS: Natural Language Processing. Lexicology. Verbal databases.

1 Introdução

O interesse pela descrição dos verbos é uma constante nos estudos descritivos do português brasileiro, doravante PB. Pelo fato de o verbo ser o tipo de palavra com maior produtividade morfológica, ser, *grosso modo*, o elemento central nas frases, articulando os demais elementos, e ser o portador de informações como tempo, modo, aspecto, o verbo é sempre alvo de estudos nas mais variadas abordagens.

Nas gramáticas tradicionais, como as de Cunha e Cintra (2008 [1984]) e Azeredo (2008), o verbo ocupa um capítulo importante, sendo classificado por suas propriedades morfossintáticas, podendo ser regular, irregular, defectivo, abundante, auxiliar, principal/pleno, impessoal, pronominal, de ligação/copulativo, transitivo e intransitivo. Nas mais distintas abordagens teórico-metodológicas, há estudos linguísticos que visam à descrição e categorização dos verbos, segundo não só suas propriedades morfossintáticas, como também por suas características semânticas (CANÇADO; GODOY, 2012; RASSI; VALE, 2013; PERINI, 2016; CANÇADO; AMARAL, 2016).

O interesse descritivo sobre as particularidades dos verbos verifica-se também em trabalhos lexicográficos especializados. Sobre PB, pode-se mencionar, como uma obra de referência, o *Dicionário de verbos e regimes*, de Francisco Fernandes, publicado em 1940. De inquestionável valor, o dicionário abrange a definição de mais de 12 mil verbos, com abonações de textos literários. Em 1990, outra obra de igual repercussão

foi publicada, o *Dicionário Gramatical de Verbos do Português Contemporâneo do Brasil*, elaborado sob coordenação de Francisco da Silva Borba. Nessa obra, aplicam-se conceitos da Linguística de Corpus para a sua produção, culminando em uma seleção mais criteriosa dos verbos descritos. O dicionário apresenta cerca de 6 mil lexemas, com a proposta de uma classificação semântica para os verbos, categorizando-os como verbos de ação, processo, ação-processo e estado. Mais recentemente, em 2013, tendo como base os dados do dicionário de Borba (1990), foi publicado o *Catálogo de Verbos do Português Brasileiro*, coordenado por Márcia Cançado. O catálogo, que contém mais de 800 verbos de mudança do PB, organiza os lexemas verbais em 4 classes, de acordo com as suas propriedades sintático-semânticas. Essas obras lexicográficas compartilham tanto o empenho pela catalogação dos verbos do PB, quanto o público-alvo, que é o usuário comum.

Em consonância com as obras citadas, a presente pesquisa se preocupa com a descrição, sistematização e classificação dos verbos do PB, no entanto, direciona sua atenção a recursos descritivos que podem ser utilizados em tarefas de Processamento de Língua Natural (PLN). Segundo Gregghi (2002), as aplicações em PLN se realizam a partir de recursos linguísticos descritivos, como analisadores sintáticos (*parsers*), *corpora*, dicionários e bases lexicais. Para a autora,

“[as bases lexicais são] volumosas e abrangentes, compreendendo vários atributos linguísticos para cada item lexical, e não necessariamente servindo a uma aplicação específica, mas à centralização e organização das informações lexicais, a fim de apoiar a pesquisa e o desenvolvimento de aplicações de PLN para uma dada língua” (GREGHI, 2002, p. 2).

Portanto, as bases lexicais atuam como recurso basilar para a construção de ferramentas computacionais como revisores ortográficos e gramaticais, além de sumarizadores e tradutores automáticos, por exemplo. Salienta-se que essas bases

podem ser úteis também para os usuários comuns, diante da possibilidade de realização de consultas, daí a importância da construção de repositórios com uma interface de fácil interação.

Esta pesquisa visa, portanto, contribuir com a atualização do estado da arte, no que se refere às bases lexicais verbais do PB desta década. Para tanto, realizou-se o levantamento das bases, a partir de buscas na web, de acordo com os seguintes critérios: (i) extensão: a base deve conter pelo menos 1.000 lexemas verbais descritos; (ii) acesso: a base deve estar disponível para consulta on-line, de maneira gratuita; e (iii) atualização: a base deve ter sofrido atualizações nos últimos 10 anos. Ao todo, foram elencadas três bases lexicais, as quais serão analisadas nas próximas seções deste artigo, a saber: VerbNet.Br, Verbo-Brasil e VerboWeb.

De acordo com as particularidades de cada base, e considerando os seus arcabouços teórico-metodológicos, esta pesquisa apresenta também uma análise crítica sobre as bases de dados verbais do PB, com o intuito de discutir os seus pontos comuns e divergentes, assim como explorar as informações sintático-semânticas disponibilizadas por cada uma, com vistas a traçar planos para melhorias que possam ocorrer na área.

Sendo assim, este artigo organiza-se da seguinte maneira: na seção dois, são apresentadas as bases lexicais verbais selecionadas; em seguida, realiza-se uma análise qualitativa e comparativa sobre esses trabalhos, considerando os seus objetivos e delimitações. Por fim, são descritas as considerações finais desta pesquisa, assim como os encaminhamentos para investigações e ações futuras.

2 Bases de dados verbais do PB

A construção de bases lexicais é uma tarefa primorosa, que demanda tempo e mão de obra especializada. A participação ativa de linguistas se faz necessária, diante

da necessidade de uma descrição e/ou revisão pormenorizada de diferentes fenômenos da língua. Posto isso, elucida-se que este trabalho prestigia as atividades realizadas nos projetos descritos, que viabilizam reflexões sobre o comportamento morfossintático-semântico dos verbos do PB, além de atuarem como ricos repositórios descritivos que podem ser utilizados em diferentes empreendimentos na área de PLN.

As próximas seções apresentam as bases lexicais verbais selecionadas, com informações detalhadas sobre: (i) desenvolvedores; (ii) fundamentação teórico-metodológica; (iii) exemplo de descrição linguística; (iv) classificações e dados quantitativos; e (v) interface.

2.1 VerbNet.Br

A VerbNet (SCHULER, 2005) é um recurso léxico-computacional (RLC) que agrupa verbos do inglês em diferentes classes semânticas inspiradas nas classes propostas por Levin (1993). Em seu trabalho, Levin (1993) agrupou os verbos do inglês em 51 grandes classes (com diversas subclasses) de acordo com as alternâncias sintáticas e características semânticas que esses verbos compartilham. A hipótese é de que há uma relação entre os sentidos do verbo e seu comportamento sintático. A partir do RLC do inglês foram criados outros para várias línguas como espanhol, francês, alemão etc.

Para o PB, a pesquisadora Carolina Evaristo Scarton, durante seu Mestrado em Ciências da Computação e Matemática Computacional da Universidade de São Paulo, criou a VerbNet.Br¹, que também parte da VerbNet do inglês e que conta com 1.766 unidades lexicais verbais e 4.333 sentidos de verbos, distribuídos em 202 classes².

¹ VerbNet.Br. Disponível em: <http://143.107.183.175:21380/verbnnetbr/index.html>. Acesso em: 29 jun. 2021.

² Salienta-se que o RLC do inglês possui 3.769 unidades lexicais verbais e 274 classes.

A metodologia proposta por Scarton (2013) foi fundamentada em quatro etapas: uma etapa manual e três, realizadas de maneira automática. A primeira etapa, única realizada de maneira manual, consistiu na definição das alternâncias sintáticas do português para cada classe presente na VerbNet do inglês. A segunda, propôs a busca automática das alternâncias sintáticas em um *corpus* etiquetado. A terceira etapa serviu para definir automaticamente os candidatos a membros das classes por meio de alinhamentos entre VerbNet, WordNet e WordNet.Br³. Por fim, a quarta e última etapa consistiu em escolher, de fato, os membros das classes. Por conta dessa abordagem *cross-linguística* que foi utilizada na criação do recurso, as classes da VerbNet.Br apresentam-se em inglês. Para uma melhor exemplificação das informações presentes nas classes verbais, no Quadro 1, são apresentados os componentes presentes na classe *disassemble-23.3*.

Quadro 2 – Informações sintático-semânticas da classe *disassemble-23.3* da VerbNet.

<i>disassemble-23.3</i>	
<i>Members: 39, Frames: 3</i>	
Papéis temáticos e [restrições seletivas]	
<i>Agent [+animate +machine]</i>	
<i>Patient [+concrete]</i>	
<i>Co- Patient [+concrete]</i>	
Frames	
VerbNet.Br	VerbNet
V_NP	NP V NP - Basic Transitive
V_NP_PP[de]	NP V NP PP.patient2 - NP-PPSource-PP

Fonte: VerbNet.Br. Disponível em: <http://verbs.colorado.edu/verb-index/vn/disassemble-23.3.php>. Acesso em: 29 jun. 2021.

A classe reproduzida acima, com informações disponíveis na versão do PB, é composta por 39 membros e 3 frames. Os papéis temáticos e as restrições de seleção da

³ WordNets são bases de dados que visam sistematizar conjuntos de substantivos, adjetivos, advérbios e verbos em quatro relações: sinonímia, antonímia, hiponímia/hiperonímia e meronímia/holonímia. (SCARTON, 2013, p. 50).

classe são, respectivamente, Agente, com restrição seletiva “animado” e “máquina”; e Paciente e Co-Paciente, com a restrição que seleciona, para esses papéis temáticos, uma entidade “concreta”. Ainda no Quadro 1, são apresentados os frames sintáticos e alternâncias de transitividade tanto do português (VerbNet.Br), quanto do inglês (VerbNet).

A VerbNet do inglês contém uma lista de 23 papéis temáticos para dar conta dos argumentos selecionados pelos verbos; todos esses papéis foram incorporados na criação da VerbNet.Br. As restrições seletivas que, por sua vez, são impostas aos papéis temáticos, também não sofreram alterações na versão do PB. Os frames sintáticos, que descrevem, além da transitividade verbal, outros itens lexicais selecionados em alternâncias em particular (como a exigência de preposições), sofrem alterações em suas traduções. Por fim, é possível consultar os predicados semânticos, que fornecem informações sobre as relações entre os participantes e o evento da ação verbal. Os predicados semânticos são divididos em quatro classes:

Predicados gerais: inclui predicados como *motion* (movimento) e *cause* (causa) e são genéricos em uma grande quantidade de classes e, também, em várias línguas.

Predicados variáveis: predicados cujo significado admite uma relação um-por-um com um conjunto de palavras em uma língua.

Predicados específicos: carregam um sentido verbal específico.

Predicados para múltiplos eventos: são predicados usados para expressar relações entre eventos. (SCARTON, 2013, p. 65-66).

No Quadro 2 abaixo, estão reproduzidos os *frames* da classe *disassemble-23.3* e, em seguida, são discutidas as relações semânticas para melhor entendimento da metalinguagem semântica empregada.

Quadro 3 – Exemplo de *frames* da classe *disassemble-23.3* na VerbNet.

FRAMES	
NP V NP	
<i>example</i>	"I unscrewed the handle."
<i>syntax</i>	<u>Agent V Patient</u>
<i>semantics</i>	<i>cause</i> (Agent, E) <i>together</i> (start(E), physical, Patient, ?Co-Patient) <i>apart</i> (result(E), physical, Patient, ?Co-Patient)
NP V NP PP.co-patient	
<i>example</i>	"I unscrewed the handle from the box."
<i>syntax</i>	<u>Agent V Patient {from} Co-Patient</u>
<i>semantics</i>	<i>cause</i> (Agent, E) <i>together</i> (start(E), physical, Patient, Co-Patient) <i>apart</i> (result(E), physical, Patient, Co-Patient)
NP V ADV-Middle	
<i>example</i>	"That new handle unscrews easily."
<i>syntax</i>	<u>Patient <+plural> V ADV</u>
<i>semantics</i>	<i>property</i> (Patient, Prop) <i>Adv</i> (Prop)

Fonte: VerbNet. Disponível em: <http://verbs.colorado.edu/verb-index/vn/disassemble-23.3.php>.

Acesso em: 29 jun. 2021.

Como é possível observar, a VerbNet utiliza uma metalinguagem de representação semântica, a qual inclui informação temporal que desempenha a função de denotar se "o predicado é verdadeiro no início do evento (*start*(E)), na preparação do evento (*during*(E)), no final do evento (*end*(E)) ou no resultado do evento (*result*(E))" (SCARTON, 2013, p. 65). Sendo assim, é retomado, em (1a), seguido de sua tradução livre ao PB (1b), o primeiro *frame* (NP V NP):

- (1) a. I unscrewed the handle.
b. Eu desparafusei a maçaneta.

Para uma interpretação da metalinguagem *cause*(Agent,E) *together*(start(E), physical, Patient, ?Co-Patient) *apart*(result(E), physical, Patient, ?Co-Patient), em uma aplicação para (1b), leia-se: o Agente causa (*cause*) um evento que, em seu início (*start*),

é verdadeiro que Paciente e Co-paciente estavam unidos (*together*) fisicamente (*physical*); e, como resultado do evento verbal (*result*), é verdadeira a separação (*apart*) física (*physical*) de Paciente e Co-paciente, aqui, a “maçaneta desparafusada”. É possível notar que há uma interrogação antes de *Co-patient* na representação da metalinguagem, justamente porque o argumento do Co-paciente não é expresso nesse *frame* sintático. Ao passo que, no segundo *frame* da classe *disassemble-23.3* (**NP V NP PP.co-patient**), em sua equivalência do *frame* em português (**V_NP_PP[de]**), o Co-paciente é um argumento obrigatório e introduzido por preposição “de”, como em (1c):

(1) c. Eu desparafusei a maçaneta da porta.

Em Scarton (2013), foram avaliadas 16 classes de verbos com base em um *golden standard* produzido manualmente, tendo sido realizados 5 experimentos diferentes. Em seus resultados, todas as classes apresentaram uma boa cobertura (de até 96,17% em um dos experimentos), mas a maioria foi avaliada com uma precisão muito baixa (de, no máximo, 44,99% de precisão entre as 16 classes). A maior parte das falhas e erros de classificação da VerbNet.Br são decorrentes do caráter predominantemente automático de sua construção. É o que se verifica, por exemplo, na polissemia relacionada ao verbo *trabalhar*, que aparece como pertencente a quatro classes verbais distintas na VerbNet.Br, conforme é reproduzido no Quadro 3.

Quadro 4 – Verbo *trabalhar* na VerbNet.Br.

TRABALHAR				
Classes	build-26.1	carve-21.2	disassemble-23.3	other_cos-45.4
Papéis temáticos	Agent Asset Beneficiary Material Product	Agent Instrument Patient	Agent Patient Patient1 Patient2	Agent Instrument Patient
Restrições seletivas	Agent [+animate +machine] Material [+concrete] Product Beneficiary [+animate +organization] Asset [+currency]	Agent [+int_control] Patient [+concrete] Instrument [+concrete]	Agent [+animate +machine] Patient [+concrete] Co-Patient [+concrete]	Agent [+int_control] Patient Instrument Result
Alternâncias	V_SN, V, V_SN_SP[em], V_SN_SP[de], V_SN_SP[com], V_SN_SP[em], V_SN_SP[para], V_SN_SP[com]_SP[para], V_SN_SP[de]_SP[para], V_SN_SP[em]_SP[para], V_SN_SP[para], V_SN_SP[em], V_SN_SP[em]_SP[por], V_SN_SP[com]_SP[por], V_SN_SP[de]_SP[por].	all.	V_SN, V_SN_SP[de].	V_SN, V_SN_SP[com], V, V_SN, V_SN_SP[em], V_SN_SP[em]_SP[com].

Fonte: VerbNet.Br. Disponível em: <http://www.nilc.icmc.usp.br/verbnnetbr/glossario.php?letra=t>.

Acesso em: 29 jun. 2021.

O principal critério para a classificação dos verbos na VerbNet e na VerbNet.Br é o conjunto de alternâncias sintáticas que um determinado grupo de verbos admite. Na VerbNet.Br, esse conjunto de alternâncias é definido para cada classe. Existem, no entanto, 33 classes cujas alternâncias são definidas como "all", ou seja, admitem todas as alternâncias sintáticas possíveis, é o caso da classe *carve* 21.2. Por outro lado, há também classes muito genéricas, nas quais é difícil identificar a relação semântica a que está submetida. É o caso, por exemplo, da classe *other_cos-45.4*, que possui 368 membros diversos, tais como *adormecer*, *apertar*, *oxidar*, *estrangular*, *explodir*, *moer* etc. Tais classes demonstram que a base apresenta a necessidade de refinamento da classificação.

A interface dessa base de dados verbais é simples e intuitiva. As buscas são organizadas por ordem alfabética e podem ser realizadas tanto por verbos, quanto por classes. Informações como restrições seletivas, predicados semânticos e subclasses estão disponíveis somente na versão do inglês, no entanto, existe o direcionamento às páginas por meio de *hiperlinks* nos títulos das classes da VerbNet.Br. Algo importante a ressaltar é que, para os *frames* sintáticos presentes em cada uma dessas classes, não há exemplos reais do PB, diferente do RLC do inglês, que apresenta uma frase de exemplo para cada alternância sintática representada por novo *frame* sintático.

2.2 Verbo-Brasil

A Verbo-Brasil (DURAN; ALUÍSIO, 2015)⁴ é uma base de dados composta de 1.093 verbos com frequência acima de 1.000 no *corpus* PLN-Br, com a finalidade de apoiar a tarefa de anotação de papéis semânticos⁵ nos projetos PropBank-Br v1 e PropBank-Br v2 (DURAN; ALUÍSIO, 2012)⁶.

A Verbo-Brasil foi concebida pela pesquisadora Magali Sanches Duran em uma equipe supervisionada pela pesquisadora Sandra Maria Aluísio, ambas integrantes do Núcleo Interinstitucional de Linguística Computacional (NILC), da Universidade de São Paulo (USP). O repositório verbal fez parte do Pros@ (Processamento Semântico de Textos em Português Brasileiro)⁷.

⁴ Verbo-Brasil. Disponível em: <http://143.107.183.175:21380/verbobrasil/index.php?lang=pt-br>. Acesso em: 29 jun. 2021.

⁵ Os termos “papel temático” e “papel semântico” foram apresentados, ao longo deste trabalho, mantendo o emprego dado na fundamentação teórica de cada base verbal.

⁶ Projeto PropBank-Br. Disponível em: <http://143.107.183.175:21380/portlex/index.php/pt/projetos/propbankbr>. Acesso em: 29 jun. 2021.

⁷ Projeto Pros@. Disponível em: <http://www.nilc.icmc.usp.br/semanticnlp/index.php?id=principal&dir=includes&lang=pt-b>. Acesso em: 29 jun. 2021.

A base verbal surge como apoio ao projeto PropBank-Br (DURAN; ALUÍSIO, 2012), com o objetivo de construir uma camada de anotação de papéis semânticos em um *corpus* do PB, aproveitando a metodologia existente em língua inglesa para a tarefa: o repositório de informação semântica *Proposition Bank* (PALMER *et al.*, 2005)⁸. O Propbank-Br v1 contém anotado uma porção brasileira do *corpus* Bosque (AFONSO *et al.*, 2002); já o PropBank-Br v2 contém 8.350 instâncias anotadas do *corpus* PLN-Br (BRUCKSCHEN *et al.*, 2008) e uma amostra de 840 instâncias do *corpus* Buscapé (HARTMANN *et al.*, 2014). Ambos os projetos mencionados (Propbank-Br v1 e Propbank-Br v2) possuem anotação com rótulos de papéis semânticos realizada sobre as árvores sintáticas geradas pelo *parser* Palavras (BICK, 2000), seguindo os parâmetros do PropBank (PALMER *et al.*, 2005), a fim de identificar a estrutura argumental dos verbos de língua inglesa equivalentes aos verbos anotados no PB. No entanto, devido às particularidades do PB, foram realizadas decisões complementares (DURAN; ALUÍSIO, 2011).

Sendo assim, o desenvolvimento da Verbo-Brasil fundamentou-se nos projetos mencionados: PropBank (PALMER *et al.*, 2005) e PropBank-Br (DURAN; ALUÍSIO, 2012), com destaque para o primeiro. Além disso, para determinar a relação semântica, o projeto também contou com as classes da VerbNet de língua inglesa (SCHULER, 2005). Segundo Palmer *et al.* (2005, p. 73-74), os verbos do PropBank são classificados de acordo com os estudos de Levin (1993), em que são organizados em função de seus aspectos semânticos e propriedades sintáticas compartilhados.

Ainda de acordo com Palmer *et al.* (2005), os argumentos semânticos de um verbo são numerados, começando com zero. Para um verbo específico, *Arg0* é geralmente o argumento que possui características de um Agente *Prototípico*, enquanto

⁸ Base de dados da língua inglesa Propbank. Disponível em: <http://verbs.colorado.edu/~mpalmer/projects/ace.html>. Acesso em: 02 jun. 2021.

Arg1 é um Paciente ou *Tema Prototípico*. Um verbo polissêmico pode ter mais de um *frameset* ou *conjunto de quadros* (sintático-semânticos) quando possui significados diferentes, apresentando um *frameset* para cada conjunto de papéis semânticos. Para exemplificar, no Quadro 4, são apresentados os papéis semânticos do verbo *trabalhar* na base Verbo-Brasil.

Quadro 5 – Verbo *trabalhar* na Verbo-Brasil.

<p>Verbo: Trabalhar Exemplo: A República Movimento de Emaús trabalha há 20 anos com adolescentes carentes.</p> <hr/> <p>Frameset trabalhar.01. <i>Exercer atividade</i>. Mapeamento para o inglês: work.01 Arg0: <i>trabalhador</i> (vnrole: -agent) Arg1: <i>trabalho</i> (<i>trabalhar com, trabalhar em</i>) (vnrole: -theme) Arg2: <i>empregador</i> (<i>trabalhar para</i>) (vnrole: -beneficiary) Arg3: <i>co-trabalhadores</i> (<i>trabalhar com alguém</i>) Arg4: <i>instrumento</i> (<i>trabalhar com</i>) (vnrole: -instrument)</p>
--

Fonte: Verbo-Brasil. Disponível em: <http://143.107.183.175:21380/verbobrasil/textoFrames/trabalhar-v.html>. Acesso em: 29 jun. 2021.

Os ArgNs (Arg0, Arg1, Arg2, Arg3, Arg4) são previstos pela semântica dos verbos e cada sentido compreende os papéis numerados em palavras, segundo o Guia de Anotação do Propbank-Br (DURAN, 2014). O *frameset* do verbo *trabalhar* inclui os sentidos identificados a partir do Propbank e os papéis e as classes conforme a VerbNet (SCHULER, 2005). Segundo o Quadro 4, o conjunto de papéis semânticos do verbo *trabalhar*, ou *trabalhar.01*, prevê cinco ArgNs: o Arg0 é *trabalhador* ou Agente, o Arg1 é *trabalho - trabalhar com, trabalhar em* ou Theme⁹, o Arg2 é *empregador - trabalhar para* ou Beneficiário, o Arg3 é *co-trabalhadores - trabalhar com alguém* e o Arg4 é *trabalhar com* ou Instrumento.

⁹ O papel semântico Theme utilizado na base de dados Verbo-Brasil parece referir-se ao “assunto do trabalho”, distanciando-se da definição comumente utilizada na literatura, em que *Theme* se refere a uma entidade que se desloca (ou sofre deslocamento), como definido em Palmer *et al.* (2010, p. 4).

O repositório da Verbo-Brasil apresenta exemplos anotados para ilustrar a atribuição dos papéis. Embora previstos, os ArgNs não precisam ocorrer todos ao mesmo tempo, como se verifica no exemplo do Quadro 4, retomado em (2):

- (2) A República Movimento de Emaús trabalha há 20 anos com adolescentes carentes.

Em (2), tem-se: *Arg0* (Agente), que corresponde a “A República Movimento de Emaús” e o verbo em terceira pessoa do singular (*trabalha*), que expressa a relação entre constituintes sintáticos e os papéis semânticos na árvore sintática. O sintagma “há 20 anos” é etiquetado como *Argm-tmp* (modificadores de tempo) e o *Arg1* (*com adolescentes carentes*) possui a anotação de Paciente ou Tema. Cada *frameset* apresenta um conjunto de exemplos. Nesse caso, *trabalhar* possui 23 exemplos em um *frameset* na Verbo-Brasil. Os exemplos descritos no repositório foram extraídos de *corpora* utilizados no projeto Propbank-Br.

A interface da base de dados apresenta as entradas verbais em ordem alfabética, sendo a busca realizada de forma fácil e intuitiva. Em cada entrada verbal consta uma lista dos sentidos de cada verbo num arquivo (*framefile*) e, para cada sentido, um conjunto de papéis semânticos previstos (*roleset*) e um conjunto de exemplos para cada quadro sintático-semântico. As *tags* foram anotadas de acordo com o projeto Propbank-Br. Além disso, cada verbo possui um mapeamento para o Propbank de língua inglesa. Na interface também consta informações do projeto, dados da equipe e downloads dos *framefiles* do repositório.

2.3 VerboWeb

A VerboWeb é uma base de dados lexicais com a classificação sintático-semântica dos verbos do PB, criada pelas pesquisadoras Márcia Cançado, Luana Amaral e Letícia Meirelles, da Universidade Federal de Minas Gerais (UFMG). Trata-

se de uma base de dados em desenvolvimento, que se fundamenta nos dados do *Catálogo de verbos do português brasileiro*, publicado pelas criadoras em 2013 e nos trabalhos que são desenvolvidos pelo Núcleo de Pesquisa em Semântica Lexical da Faculdade de Letras da UFMG (NuPes)¹⁰.

Essa iniciativa pretende realizar a análise de cerca de 3.000 verbos e, até o momento, já divulga em seu site a descrição de 1.482 construções. As entradas lexicais verbais, disponíveis na base, foram retiradas do Dicionário do Borba (1990) e do Dicionário eletrônico Houaiss (2009) e as frases criadas, atestadas em *corpora* da Linguateca e na *web*.

Propondo uma descrição e classificação dos verbos do PB, a VerboWeb se pauta nos princípios da Semântica Lexical, mais especificamente da chamada *Interface Sintaxe-Semântica Lexical*, considerando, para a proposta de tipologia, nos moldes expostos em Cançado, Godoy e Amaral (2013): (i) *os papéis temáticos*, que consistem nas relações estabelecidas entre um predicado e seus argumentos; (ii) *o aspecto lexical*, que remete a como um evento verbal se desenrola no decorrer do tempo; e (iii) *a decomposição de predicados*, que se refere à metalinguagem formal e sistemática, que lida com os sentidos dos verbos, decompondo o sentido dos itens lexicais em um sistema de predicados primitivos.

As frases (3a) e (4a) exemplificam os comportamentos do verbo *partir*, seguidas pela representação lógica em (b) e sua paráfrase em (c). Os dados foram retirados da VerboWeb:

- (3) a. O peso do machado / O lenhador partiu a tora de madeira para fazer uma fogueira
 b. [[X ACT (volition)] CAUSE [BECOME [Y <STATE>]]]
 c. X (volicionalmente ou não) age causando Y ficar em determinado estado

¹⁰ VerboWeb. Disponível em: <http://www.letras.ufmg.br/verboweb/>. Acesso em: 29 jun. 2021.

- (4) a. As tropas partiram para o campo de batalha
 b. [BECOME<event> [X LOC Y]]
 c. X passa a ficar em Y através de um evento

Segundo Cançado *et al.* (2013, p. 107), “classificar verbos implica agrupá-los em classes que partilham certas propriedades não só semânticas, mas também sintáticas, ou, ainda, implica agrupá-los por propriedades semânticas que tenham impacto no seu comportamento gramatical”. Sendo assim, o lexema *partir*, conforme exemplificado acima, é duplicado por apresentar diferenças semânticas e sintáticas acentuadas em sua constituição. O Quadro 5 apresenta um exemplo detalhado da descrição sintático-semântica do verbo *trabalhar* na VerboWeb.

Quadro 6 – Verbo *trabalhar* na VerboWeb.

<p>Verbo: Trabalhar Exemplo: O rapaz trabalhava demais.</p> <hr/> <p>Classe: Verbos de atividade: internamente causados (inergativos) Propriedades da Classe:</p> <ul style="list-style-type: none"> - Conteúdo semântico recorrente na classe: x faz/produz um evento em si mesmo - Estrutura sintática básica: [SN V] (verbo intransitivo) - Estrutura de papéis temáticos: {Agente} - Estrutura de decomposição de predicados: [X DO <EVENT>] - Aspecto lexical básico: atividade - Licencia um objeto cognato: O rapaz trabalhava, trabalhos voluntários. - Licencia um adjunto equivalente ao objeto cognato: O rapaz trabalhava voluntariamente
--

Fonte: VerboWeb. Disponível em:

http://www.letras.ufmg.br/sistemas/verboweb_cliente/ver_verbo.php?id=1579. Acesso em: 29 jun. 21.

Conforme se verifica no Quadro 5, a partir da busca pelo lexema verbal *trabalhar*, obtém-se informações sobre a classe e as propriedades do verbo: conteúdo semântico, papéis temáticos, aspecto lexical, estrutura sintática, estrutura de decomposição de predicados e seus licenciamentos sintáticos, que, no caso, se referem à atuação com um *objeto cognato* e um *adjunto que equivale a esse objeto*.

Na VerboWeb, propõe-se uma classificação que consiste em 4 grandes categorias, 17 classes e 7 subclasses, a saber¹¹:

- (i) *Verbos de atividade (de ação)*: inergativos, como *correr* (subclasses: sujeito recíproco e expressão); inergativos com instrumento incorporado, como *esquiar*; contato com instrumento incorporado, como *martelar* (subclasse: instrumentais com dois instrumentos); contato mediado por instrumento, como *afiar* (subclasse: remoção); e contato mediado por corpo, como *abraçar*.
- (ii) *Verbos de causação*: de mudança de estado opcionalmente agentivo, como *quebrar* (subclasses: objeto recíproco e contato); mudança de estado não agentivo, como *cansar*; mudança de estado locativo, como *dependurar* (subclasse: criação de imagem); mudança de estado de posse, como *encher*; mudança de lugar, como *engarrafar*; mudança de posse, como *acorrentar*; transferência do tipo locatum, como *abençoar*; e estado psicológico, como *preocupar*.
- (iii) *Verbos de culminação (processo)*: de mudança de estado, como *amadurecer*; e mudança de lugar, como *entrar*.
- (iv) *Verbos de estado*: psicológicos, como *amar*; e existenciais, como *haver*.

Salienta-se que a página da VerboWeb é de fácil acesso e está atualizada. As buscas podem ser feitas por *verbo*, *classe*, *subclasse* ou a partir de alguma *propriedade não classificatória*. Além disso, é possível estabelecer relações de busca entre classe e subclasse ou cruzamento de propriedades, por exemplo. No site, encontram-se

¹¹ Dados baseados em Cançado, Amaral e Meirelles (2018) e na palestra *Um raio x dos verbos do PB: o projeto VerboWeb*, disponível em: <https://youtu.be/VnZrqNfW2QI>. Acesso em: 29 jun. 2021.

também informações sobre os pesquisadores envolvidos no projeto, referências bibliográficas e um tutorial sobre como fazer as buscas na base de dados. Pode-se afirmar, portanto, a existência de um recurso descritivo rico sobre os verbos da língua portuguesa, disponibilizado de maneira gratuita e acessível, o que a instaura como uma ferramenta interessante não só para linguistas e profissionais da computação, como também para usuários comuns.

3 As descrições sintático-semânticas verbais nas bases de dados do PB: reflexões e análise

Ao realizar a comparação entre as três bases de dados lexicais verbais do PB (VerbNet.Br, Verbo-Brasil e VerboWeb), é possível observar seus pontos comuns e divergentes, os quais serão descritos nesta seção.

Todas as bases analisadas consideram as classes semânticas propostas por Levin (1993) para a categorização dos verbos e apresentam, como objeto de descrição, o comportamento sintático e semântico de verbos *plenos* do PB. A VerbNet.Br e a VerboWeb se assemelham ainda na utilização de metalinguagens formais para a descrição dos fenômenos verbais.

Cada um dos recursos possui um número específico de entradas anotadas e categorizadas: VerbNet.Br, 1.766; Verbo-Brasil, 1.093; e VerboWeb, 1.486¹². No entanto, apresentam a descrição de apenas 298 comportamentos verbais em comum, incluindo verbos muito frequentes (*ir, mandar, pedir, voltar*) e verbos pouco frequentes (*afiar, congestionar, fascinar, rachar*) da língua.

¹² É importante mencionar que tanto a VerbNet.Br, quanto a VerboWeb, duplicam as entradas verbais quando o lema apresenta diferentes comportamentos sintático-semânticos (*partir 1; partir 2*). Já na Verbo-Brasil, não há casos de entradas duplicadas, pois é apenas no interior de cada lema verbal que se tem acesso às suas diferentes construções sintático-semânticas (*framesets*).

Para além das diferenças concernentes aos arcabouços teórico-metodológicos de cada recurso, verifica-se também critérios distintos para a criação de diferentes entradas com a mesma forma verbal. Enquanto a VerbNet.Br é uma base mais granular e relaciona um lema verbal a diferentes classes, a depender de suas alternâncias sintáticas e características semânticas, a Verbo-Brasil e a VerboWeb não apresentam muitos casos de duplicação, salvo quando os verbos possuem comportamentos muito destoantes (como o exemplo de *partir* na seção 2.3).

Conforme se verifica nos exemplos dados com *trabalhar*, na VerbNet.Br, o verbo está relacionado a quatro classes distintas (*build-26.1*, *carve-21.2*, *disassemble-23.3*, *other_cos-45.4*), ao passo que, na Verbo-Brasil e na VerboWeb, esse mesmo lexema possui apenas uma entrada: na Verbo-Brasil, embora esteja descrito em apenas um *frameset*, o verbo *trabalhar* apresenta ao todo 23 exemplos e propõe a informação sintática e semântica de quatro argumentos distintos: Agente, na posição de sujeito, e Tema, Beneficiário e/ou Instrumento, na posição de objeto; já na VerboWeb, *trabalhar* está descrito com o comportamento de uma construção intransitiva da classe de *verbos inergativos*, em que se descreve apenas o papel semântico do argumento que ocupa a posição de sujeito (Agente). A abrangência de informações sintático-semânticas da Verbo-Brasil também se destaca quando se observa a preocupação em anotar os sentidos não só de verbos plenos (e verbos pronominais), como também os usos de verbos auxiliares, de construções multipalavras e, ainda, de expressões idiomáticas verbais do PB.

Com base na análise dos dados, é possível contrapor dois projetos de investigação distintos, em que um se dedica à descrição linguística de fenômenos verbais (VerboWeb), e o outro, elabora propostas descritivas para fins computacionais, inserindo-se na área de PLN (VerbNet.Br e Verbo-Brasil).

Sendo assim, a VerboWeb, projeto ainda em desenvolvimento, ilustra o primeiro tipo de empreendimento descritivo, o qual tem como objetivo a realização de

descrições linguísticas sobre diferentes fenômenos verbais da língua portuguesa. São, portanto, dados produzidos de maneira manual, atividade que demanda tempo e mão de obra especializada, com a aplicação de resultados de pesquisas realizadas há mais de 20 anos na área da Linguística. O fato de o banco de dados não ser uma tradução ou adaptação de trabalhos anteriores e/ou realizados para outras línguas naturais garante uma maior uniformidade dos dados e o estabelecimento de um recurso coerente com a sua base teórico-metodológica.

Por outro lado, têm-se a VerbNet.Br e a Verbo-Brasil como exemplos de recursos elaborados para fins computacionais, os quais unem as competências profissionais de linguistas e cientistas da computação, privilegiando-se, desse modo, a otimização dos sistemas e visando aplicações na área de PLN. A Verbo-Brasil, que é um apêndice verbal do projeto Propbank-Br, partiu de um mapeamento verbal existente para a língua inglesa, assim como ocorreu com a elaboração da VerbNet.Br, criada a partir de dados do inglês e de forma semiautomática. Apesar de essa metodologia otimizar o trabalho desenvolvido, apresenta determinadas lacunas para o PB, como a ausência de verbos frequentes da língua (*atualizar, adiar, emprestar*¹³) e, no caso da VerbNet.Br, a proposta de classes muito genéricas, nas quais é difícil identificar a relação semântica e sintática de sua composição. Sobre isso, Scarton (2013, p. 98-99) enfatiza a importância da VerbNet.Br como um repositório de dados verbais do PB, mas reconhece a necessidade de validação linguística, principalmente por se tratar de um recurso desenvolvido com base na tradução. A anotação realizada por um linguista permitiria a revisão dos dados e a inserção de alternâncias específicas para a língua portuguesa.

¹³ Os exemplos mencionados se justificam pela frequência superior a 2.500 ocorrências no *corpus* Folha Kaggle, disponível em: <https://www.kaggle.com/marlesson/news-of-the-site-folhauol>. Acesso em: 29 jun. 2021.

A distinção entre esses “dois projetos de investigação” parece refletir também na disponibilização dos dados para consulta, já que a ferramenta de caráter mais linguístico (VerboWeb) possui uma interface mais acessível a especialistas e usuários comuns se comparada aos recursos elaborados com o propósito de servirem a aplicações de PLN (VerbNet.Br e Verbo-Brasil). Como já mencionado, a VerboWeb possui alternativas de buscas, para além da entrada verbal, e apresenta definições de informações sintático-semânticas anotadas nos itens lexicais, além de sugestões de referências acadêmicas sobre os temas. A Verbo-Brasil, por sua vez, embora não apresente atualizações nos últimos cinco anos, permite buscas simples com a entrada lexical verbal ou a listagem de verbos disponíveis, organizados em ordem alfabética, com a disposição das informações na página, que contribuem para o seu uso de maneira intuitiva. Já a VerbNet.Br parece ser o recurso com menor facilidade de acesso aos dados: apesar de ser possível realizar buscas por classes ou pelo lexema verbal e ainda consultar listas de verbos em ordem alfabética, as informações detalhadas sobre os predicados semânticos das classes estão em inglês. Portanto, para se ter acesso a essas informações, a página redireciona o usuário aos dados da VerbNet de língua inglesa.

No Quadro 6, são resumidas as características das bases comparadas neste trabalho.

Quadro 7 – Aspectos comuns e divergentes entre as bases lexicais verbais do PB.

Aspecto	Convergência	Divergência
---------	--------------	-------------

Quantidade de lexemas	Todas as bases possuem mais de 1.000 entradas verbais descritas.	Cada base apresenta uma quantidade distinta de entradas lexicais, assim como diferentes critérios para duplicação e classificação dos lemas: VerbNet.Br (1.766 entradas), Verbo-Brasil (1.093 entradas) e VerboWeb (1.486 entradas).
Base teórico-metodológica	Todas as bases se inspiram nas classes da semântica lexical de Levin (1993); VerbNet.Br e VerboWeb apresentam uma metalinguagem formal de representação sintático-semântica.	Embora se inspirem nas classes de Levin (1993), cada base é guiada ainda por outros aspectos sintático-semânticos.
Objetivo	Todas as bases objetivam propor uma descrição sintático-semântica dos comportamentos verbais do PB; VerbNet.Br e Verbo-Brasil são recursos criados para fins computacionais.	Apesar de também poder ser utilizada para fins computacionais, a VerboWeb não se construiu especificamente para essa função.
Organização dos dados	VerbNet.Br e VerboWeb, embora partam de critérios distintos, organizam os verbos em classes sintático-semânticas: VerbNet.Br, com 202 classes semânticas; VerboWeb, com 4 grandes categorias sintático-semânticas.	Verbo-Brasil apresenta a descrição sintática e semântica dos verbos, em <i>framesets</i> , mas não apresenta uma proposta de tipologia para a sua organização.
Origem dos dados	Verbo-Brasil e VerboWeb são recursos feitos sobre o PB e validados por linguistas brasileiras.	VerbNet.Br e Verbo-Brasil se apoiam em descrições linguísticas do inglês; VerbNet.Br e Verbo-Brasil possuem dados decorrentes de processos automatizados; VerbNet.Br apresenta dados que carecem de revisão realizada por linguistas; VerboWeb é um recurso feito inteiramente por linguistas.
Interface	Todas as bases de dados possuem páginas gratuitas de acesso aos dados, com possibilidade de buscas. VerbNet.Br e Verbo-Brasil não apresentam atualizações recentes. No entanto, as bases disponibilizam a opção de <i>download</i> dos dados.	VerbNet.Br redireciona o usuário para os dados da VerbNet do inglês; VerboWeb apresenta dados atualizados e possibilita diferentes opções de busca, sendo uma ferramenta interessante para usuários comuns. No entanto, a base não disponibiliza o <i>download</i> dos dados.

Fonte: autoria própria.

4 Considerações finais

Neste trabalho, foram descritas e comparadas três bases de dados verbais do PB, disponíveis de maneira gratuita atualmente e comandadas por pesquisadoras brasileiras nos últimos 10 anos: VerbNet.Br, VerboWeb e Verbo-Brasil. Os dados revelam a existência de trabalhos com descrições robustas com informações relevantes sobre a sintaxe e a semântica dos verbos do PB.

A análise comparada entre essas bases enaltece o papel imprescindível do linguista na elaboração e revisão de dados lexicais para fins computacionais: a base realizada de maneira inteiramente manual por linguistas – VerboWeb – demonstrou ser a que apresenta maior coerência com a sua fundamentação teórico-metodológica, enquanto que a VerbNet.Br, feita de forma automática e com base em traduções do inglês, é a base que necessita de maior revisão dos dados, conforme sugere sua própria idealizadora, Scarton (2013). Por sua vez, a Verbo-Brasil, embora tenha sido criada para cumprir funções de anotação fundamentalmente computacionais, parece ser a base que contém mais informações sintático-semânticas das construções verbais em seu acervo, já que inclui, em seu repositório, além do comportamento de verbos plenos, verbos auxiliares, construções multipalavras e expressões idiomáticas.

Nesse sentido, ressalta-se a necessidade de adaptação, ampliação e/ou criação de repositórios verbais que considerem a polissemia de determinados verbos e, ainda, o seu comportamento quando não atuam como verbos *plenos*. Sobre isso, pode-se citar trabalhos lexicogramaticais sobre as construções com nomes predicativos e verbos-suporte de Barros (2014), Santos (2015) e Rassi (2015), que descrevem os comportamentos dos verbos *fazer* (*fazer exercício*), *ter* (*ter apoio*) e *dar* (*dar trabalho*), respectivamente, com foco em sua atuação como verbo-suporte. Acredita-se que estas descrições precisam constar em bases de dados disponíveis para consulta por usuários comuns e também para aplicações linguístico-computacionais na área de PLN.

É importante salientar ainda que há, obviamente, para a língua portuguesa, outras bases de dados verbais disponíveis de igual - ou maior - impacto que as

selecionadas. Esses casos não constam em nosso artigo por não preencherem os critérios estipulados nesta pesquisa. A título de exemplo, menciona-se o projeto ViPER, uma base de dados de construções léxico-sintáticas dos verbos do português europeu, que culminou na publicação do *Dicionário gramatical dos verbos do português* (BAPTISTA, MAMEDE, 2020), o qual consta de 6.000 entradas verbais, com a apresentação de suas propriedades sintático-semânticas.

Outro trabalho de descrição dos verbos do PB importante de ser mencionado é o Projeto *Valências Verbais do Português* (VVP), sob a coordenação de Mário A. Perini. O objetivo do projeto é listar, em um dicionário, todos os conjuntos de construções (diáteses) que ocorrem no PB, assim como os verbos associados a cada um deles. A orientação do trabalho é fundamentalmente descritiva, procurando levantar e sistematizar dados, com base, tanto quanto possível, em pressupostos teóricos de ampla aceitação. Para isso, não há endosso de nenhuma corrente teórica, porque, segundo Perini (2008), as teorias devem basear-se no maior número de dados possível. Até o momento, o VVP contém cerca de 700 verbetes verbais descritos por linguistas associados ao projeto. Alguns resultados prévios podem ser conferidos em Perini (2015, 2016, 2019).

Como mencionado, o verbo é um elemento nuclear para as línguas naturais, daí a necessidade de elaboração de trabalhos acadêmicos e bases de dados lexicais dedicados especificamente ao comportamento sintático-semântico dessa classe de palavras. Por isso, como ações de pesquisas futuras, espera-se ampliar o reconhecimento e análise de bases de dados verbais disponíveis, principalmente de maneira gratuita, para o português, incluindo outras variantes da língua. Além disso, almeja-se verificar também como o verbo é tratado em diferentes recursos linguístico-computacionais, como a FrameNet.Br (SALOMÃO *et al.* 2013) e a anotação em Universal Dependencies (MCDONALD *et al.*, 2013) do *corpus* Bosque (AFONSO *et al.*, 2002), por exemplo.

Portanto, acredita-se que esta pesquisa contribui com a atualização do estado da arte sobre as bases de dados lexicais verbais do PB, dá visibilidade às obras analisadas e propõe interessantes desdobramentos de pesquisas e ações futuras para o incremento de recursos linguístico-computacionais do português, sobretudo da variante brasileira.

Agradecimentos

Os autores deste trabalho agradecem ao Centro de Inteligência Artificial (C4AI-USP) e o apoio da Fundação de Apoio à Pesquisa do Estado de São Paulo (processo FAPESP #2019/07665-4) e da IBM Corporation.

Referências

- AFONSO, S. BICK, E.; HABER, R.; SANTOS, D. Floresta sintá(c)tica: a treebank for Portuguese. *In*: RODRIGUES, M. G.; ARAUJO, C. P. S. (org.). **Proceedings of the Third International Conference on Language Resources and Evaluation (LREC 2002)**. Paris: ELRA, 2002.
- AZEREDO, J. C. **Gramática Houaiss da Língua Portuguesa**. São Paulo: Publifolha, 2008.
- BAPTISTA, J.; MAMEDE, N. **Dicionário gramatical de verbos do português**. Faro: Editora UAlg, 2020.
- BARROS, C. D. **Descrição e classificação dos predicados nominais com o verbo-suporte fazer em Português do Brasil**. Tese (Doutorado) - Universidade Federal de São Carlos (UFSCar), São Carlos, 2014.
- BICK, E. **The parsing system palavras**: Automatic grammatical analysis of Portuguese in a constraint grammar framework. Aarhus Universitetsforlag, 2000.
- BORBA, F. S. (coord.). **Dicionário gramatical de verbos do português contemporâneo do Brasil**. São Paulo: Editora UNESP, 1990.
- CANÇADO, M.; GODOY, L. Representação lexical de classes verbais do PB. **Alfa**, São Paulo, 56 (1), 2012. DOI <https://doi.org/10.1590/S1981-57942012000100006>

CANÇADO, M.; GODOY, L.; AMARAL, L. **Catálogo de verbos do português brasileiro**: classificação verbal segundo a decomposição de predicados. v. 1. Verbo de mudança. Belo Horizonte: Editora UFMG, 2013.

CANÇADO, M.; AMARAL, L. **Introdução à Semântica Lexical**. Petrópolis: Vozes, 2016.

CANÇADO, M.; AMARAL, L.; MEIRELLES, L. VerboWeb: uma proposta de classificação verbal. **Revista da Anpoll**, v. 1, 2018. DOI <https://doi.org/10.18309/anp.v1i46.1077>

CANÇADO, M.; AMARAL, L.; MEIRELLES, L. **Banco de Dados Lexicais VerboWeb**: classificação sintático-semântica dos verbos do português brasileiro, UFMG. Disponível em: <http://www.letras.ufmg.br/verboweb/>. Acesso em: 29 jun. 2021.

CUNHA, C.; CINTRA, L. **Nova gramática do português contemporâneo**. 5 ed. Rio de Janeiro: Lexikon, 2008 [1984].

DURAN, M. S.; ALUÍSIO, S. M. Propbank-Br: A Brazilian Portuguese corpus annotated with semantic role labels. *In: Proceedings of the 8th Symposium in Information and Human Language Technology*. Cuiabá, Brazil, 2011.

DURAN, M. S., ALUÍSIO, S. M. Propbank-Br: A Brazilian Treebank annotated with semantic role labels. *In: Proceedings of the 8th International Conference on Language Resources and Evaluation*. Istanbul, Turkey, 2012.

DURAN, M. S.; MARTINS, J. P.; ALUÍSIO, S. M. Um repositório de verbos para a anotação de papéis semânticos disponível na web. *In: Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology*. Fortaleza, Brazil, 2013. Disponível em: <https://www.aclweb.org/anthology/W13-4820.pdf>. Acesso em: 29 jun. 2021.

DURAN, M. S. **Guia de Anotação**: Propbank-Br. 2014. Disponível em: <https://docplayer.com.br/81173801-Guia-de-anotacao-propbank-br.html>. Acesso em: 29 jun. 2021.

DURAN, M. S.; ALUÍSIO, S. M. Automatic Generation of a Lexical Resource to support Semantic Role Labeling in Portuguese. *In: Proceedings of SEM 2015: The Fourth Joint*

Conference on Lexical and Computational Semantics. Colorado, US, 2015. Disponível em: <https://www.aclweb.org/anthology/S15-1026.pdf>. Acesso em: 29 jun. 2021.

FERNANDES, F. **Dicionário de verbos e regimes**. 45 ed. Porto Alegre: Globo, 2005 [1940]. DOI <https://doi.org/10.22456/2177-0018.7658>

GREGHI, J. G. **Projeto e desenvolvimento de uma base de dados lexicais do português**. Dissertação (Mestrado em Ciências da Computação e Matemática Computacional). – Universidade de São Paulo (USP), São Carlos, 2002.

LEVIN, B. **English Verb Classes and Alternations: A Preliminary Investigation**. Chicago: University of Chicago Press, 1993.

MCDONALD, R.; NIVRE, J.; QUIRMBACH-BRUNDAGE, Y.; GOLDBERG, Y.; DAS, D.; GANCHEV, K.; HALL, K.; PETROV, S.; ZHANG, H.; TÄCKSTRÖM, O.; BEDINI, C.; CASTELLÓ, N. B.; LEE, J. Universal dependency annotation for multilingual parsing. *In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Sofia, Bulgaria, 2013. Disponível em: <https://aclanthology.org/P13-2017.pdf>. Acesso em: 18 set. 2021

PALMER, M.; GILDEA, D.; KINGSBURY, P. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1), 71–106, 2005. DOI <https://doi.org/10.1162/0891201053630264>

PALMER, M.; GILDEA, D.; XUE, N. Semantic role labeling. *Synthesis Lectures on Human Language Technologies*, v. 3, n. 1, p. 1-103, 2010. DOI <https://doi.org/10.2200/S00239ED1V01Y200912HLT006>

PERINI, M. A. **Estudos de gramática descritiva**. São Paulo: Parábola Editorial, 2008.

PERINI, M. A. **Describing verb valency: practical and theoretical issues**. Switzerland: Springer, 2015. DOI <https://doi.org/10.1007/978-3-319-20985-2>

PERINI, M. A. Construindo o dicionário de valências: problemas e resultados. *Scripta (PUCMG)*, v. 20, 2016. DOI <https://doi.org/10.5752/P.2358-3428.2016v20n38p148>

PERINI, M. A. **Gramática descritiva do português brasileiro**. Petrópolis: Vozes, 2016.

PERINI, M. A. **Thematic Relations**. Switzerland: Springer, 2019. DOI <https://doi.org/10.1007/978-3-030-28538-8>

RASSI, A. P.; VALE, O. A. Tipologia das construções verbais em PB: uma proposta de classificação do verbo dar. **Caligrama**, Belo Horizonte, v. 18, n. 2, 2013. DOI <https://doi.org/10.17851/2238-3824.18.2.105-130>

RASSI, A. P. **Descrição, classificação e processamento automático das construções com o verbo dar em português brasileiro**. Tese (Doutorado) - Universidade Federal de São Carlos (UFSCar), São Carlos, 2015.

SALOMÃO, M. M. M.; TORRENT, T. T.; SAMPAIO, T. F. A Linguística Cognitiva encontra a Linguística Computacional: Notícias do Projeto FrameNet Brasil. **Cadernos de Estudos Linguísticos**, 55(1), 7-34, 2013. DOI <https://doi.org/10.20396/cel.v55i1.8636592>

SANTOS, M. C. A. **Descrição e classificação dos predicados nominais com o verbo-suporte ter em Português do Brasil**. Tese (Doutorado) - Universidade Federal de São Carlos (UFSCar), São Carlos, 2015.

SCARTON, C. E. **VerbNet.Br**: construção semiautomática de um léxico verbal online e independente de domínio para o português do Brasil. 2013. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) - Universidade de São Paulo (USP), São Carlos, 2013.

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

Artigo recebido em: 04.10.2021

Artigo aprovado em: 10.03.2022