

VerbNet.Br: construção semiautomática de um léxico computacional de verbos para o português do Brasil

Carolina Evaristo Scarton

Núcleo Interinstitucional de Linguística Computacional (NILC)
Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo
Caixa Postal 668-13.560-970 – São Carlos – SP – Brasil

carol.scarton@gmail.com

***Abstract.** This article describes a semiautomatic method to build a computational verb lexicon for Brazilian Portuguese language (called VerbNet.Br). This lexical resource is based on the computational verb lexicon VerbNet for the English language, built according to Levin verb class theory. The method presented here is based on four steps: one manual and three automatic steps. Automatic steps use existing lexical resources and corpus-based knowledge. Also, this article presents the details of the implementation of one of the steps that uses existing lexical resources. The results of this step are the potential candidates of VerbNet.Br classes, yet without the validation.*

***Resumo.** Neste artigo é apresentado o método de criação semiautomática de um recurso léxico computacional de verbos para o português do Brasil, chamado VerbNet.Br, baseado na VerbNet, recurso léxico verbal para a língua inglesa, que segue a teoria das classes de Levin. O método de criação possui quatro etapas: uma manual e três automáticas. Nas etapas automáticas, recursos léxicos já existentes e conhecimento extraído de corpus são usados como informações de entrada. Também são apresentados neste artigo os detalhes de implementação da etapa automática que utiliza recursos léxicos já existentes. Os resultados obtidos, nesta etapa, são os candidatos a verbos membros das classes da VerbNet.Br, ainda sem validação.*

1. Introdução

A criação e disponibilização de recursos léxicos computacionais (RLC) é uma das atividades compreendidas pela área de Processamento de Língua Natural (PLN), pois eles são fundamentais para processar ou analisar a língua natural. Em especial, RLC's com informações sintáticas e/ou semânticas são importantes para tarefas como: animações de instruções em língua natural [Allbeck et al., 2002], construção de *parsers* semânticos [Shi and Mihalcea, 2005], desambiguação do sentido de palavras [Girju et al., 2005], dentre outras. Dentre os RLC's, um grupo especial são os que tratam de verbos, pois verbos possuem uma grande carga de informação sintática e semântica, sendo possível saber, a partir do comportamento do verbo, como a sentença se comporta.

Como a criação manual de RLC's é inviável dada a grande carga de trabalho e quantidade de tempo demandada, existem iniciativas que visam criar recursos léxicos automaticamente ou semiautomaticamente. As duas principais técnicas computacionais

utilizadas nesta tarefa são o uso de aprendizado de máquina em *córpus*¹ e o uso de recursos léxicos já existentes para outras línguas, numa abordagem cross-linguística.

A língua inglesa possui tradição na criação de RLC's. Os principais são: FrameNet [Baker et al., 1998], PropBank [Palmer et al., 2005], WordNet de Princeton (aqui chamada de WordNet.Pr) [Fellbaum, 1998] e VerbNet [Kipper, 2005], que foram criados manualmente ou semiautomaticamente. Em especial, a VerbNet é um RLC que trata especificamente de verbos e possui informações sintático-semânticas dos mesmos, seguindo a taxonomia de classe proposta por Levin (1993). Levin definiu que os verbos que compartilham o mesmo comportamento sintático (as mesmas alternâncias sintáticas) devem, também, compartilhar comportamento semântico. Como exemplos seguem as expressões (1) e (2), retiradas do trabalho de Levin (1993, p. 2), que apresentam a alternância locativa dos verbos em inglês *spray* e *load*.

(1)(a) *Sharon sprayed water on the plants.* (Sharon borrifou água sobre as plantas)

(b) *Sharon sprayed the plants with water.* (Sharon borrifou as plantas com água)

(2)(a) *The farmer loaded apples into the cart.* (O fazendeiro carregou maçãs na carroça)

(b) *The farmer loaded the cart with apples.* (O fazendeiro carregou a carroça com maçãs)

É possível observar nos exemplos (1) e (2) que os verbos *spray* e *load* compartilham o mesmo comportamento sintático e carregam o significado semântico de “cobrir/colocar”. Também é possível observar que a semântica fornecida pelas classes de Levin é uma semântica superficial (não se pode dizer, por exemplo, que os membros de uma mesma classe são sinônimos). A VerbNet supre essa lacuna semântica com mapeamentos para a WordNet.Pr, que é um RLC com informações sobre semântica verbal bem mais profundas (no nível de sinonímia, hiperonímia e outras relações).

Para o português do Brasil, existem algumas iniciativas na criação semiautomática de RLC's: a FrameNet Brasil [Salomão, 2009] e o FrameCorp [Bertoldi e Chishman, 2009] (baseados na FrameNet), a WordNet.Br [Dias-da Silva et al., 2008] (baseada na WordNet.Pr e diretamente alinhada com a mesma) e o PropBank.Br [Duran, 2009] (baseado no PropBank). Porém, nenhum destes recursos trata da interface sintático-semântica dos verbos, informação importante que tem sido um limitante para a área de PLN no Brasil. Por outro lado, alguns trabalhos de descrição do português apresentam as diferenças no português para algumas das classes de Levin: Cançado (1996) (verbos psicológicos); Chagas de Souza (2001) (uma construção própria da língua portuguesa, a construção adversativa); Moraes (2008) (verbos de movimento e suas alternâncias e classes) e Amaral (2010) (verbos de modo de movimento). Porém, esses e outros trabalhos tratam, geralmente, de um conjunto fechado de verbos e não disponibilizam os resultados da análise em formato eletrônico² para que possam ser aproveitados computacionalmente.

Há, também, iniciativas para a construção automática de RLC's verbais, utilizando aprendizado de máquina, como o trabalho de Joanis and Stevenson (2003) e Sun and Korhonen (2009) para o inglês, Merlo et al. (2002) para o italiano, Ferrer

¹ Neste trabalho escolheu-se o aportuguesamento da palavra *corpus/corpora* para *córpus/córpus*.

² Há uma iniciativa do Núcleo de Pesquisa em Semântica Lexical (NuPes), coordenado pela professora Márcia Cançado, da Universidade Federal de Minas Gerais, em disponibilizar de forma mais acessível os dados gerados nos trabalhos dos membros do grupo (Márcia Cançado, 2011, comunicação pessoal).

(2004) para o espanhol, Schulte in Walde (2006) para o alemão e Sun et al. (2010) para o francês. Todos estes trabalhos utilizam aprendizado de máquina não supervisionado.

Neste trabalho, são apresentadas as etapas de um método para a criação semiautomática de uma VerbNet para o português do Brasil (batizada de VerbNet.Br) a partir de recursos léxicos já existentes (VerbNet, WordNet.Pr e WordNet.Br) e diretamente alinhada com a VerbNet do inglês. Considera-se a hipótese de que há um potencial cross-linguístico nas classes de Levin – como evidenciado em Jackendoff (1990), comprovado em Merlo et al. (2002) (para o italiano) e Sun et al. (2010) (para o francês) e testado para ao português em Kipper (2005). Também se considera que uma construção semiautomática produza resultados mais precisos do que métodos baseados em aprendizado não supervisionado (*clustering*). Resultados de uma das etapas de criação também são apresentados.

Na Seção 2, são apresentados os recursos léxicos já existentes que serão utilizados para a construção da VerbNet.Br: WordNet.Pr, WordNet.Br e VerbNet. Na Seção 3, é apresentado o método de criação da VerbNet.Br composto de cinco etapas. Na Seção 4 são apresentados os primeiros resultados, mais especificamente, os resultados da etapa 3 do método, que considera os recursos léxicos mencionados anteriormente. Por fim, a Seção 5 fica reservada para as considerações finais.

2. Recursos Léxicos Computacionais e a VerbNet.Br

Dentre os recursos léxicos apresentados na Introdução, três deles merecem destaque, pois serão diretamente utilizados na construção da VerbNet.Br: a WordNet.Pr, a WordNet.Br e a VerbNet.

Uma *wordnet* é uma base de dados que sistematiza o conjunto dos verbos, substantivos, adjetivos e advérbios de um dado idioma em uma rede de quatro relações: sinonímia, antonímia, hiponímia/hiperonímia e meronímia/holonímia [Cruse, 1986]. A WordNet.Pr é a pioneira em recursos deste tipo (com 11.306 verbos e 13.508 *synsets*³ de verbos). A partir desta *wordnet*, *wordnets* para outras línguas começaram a ser construídas utilizando um índice inter-lingual (ILI) [Vossen, 2004] para manter o alinhamento com os *synsets* da WordNet.Pr, garantindo que algumas relações pudessem ser herdadas automaticamente da WordNet.Pr para a nova *wordnet* (como foi o caso da herança automática da relação de hiperonímia da WordNet.Pr para a WordNet.Br, realizada no trabalho de Scarton e Aluísio (2010)).

A Wordnet.Br é a *wordnet* para o português do Brasil. A construção de sua base de relações é feita (e já foi concluída para verbos) por meio de um alinhamento com a WordNet.Pr (usando o ILI). Seguindo o modelo proposto pelos criadores da EuroWordNet [Vossen, 2004], são possíveis quatro tipos de relações: EQ_SYNONYM (sinônimo perfeito), EQ_NEAR_SYNONYM (sinônimo imperfeito: quando mais de um ILI da WordNet.Pr está relacionado com um *synset* da WordNet.Br ou vice-versa), EQ_HAS_HYPONYM (um *synset* da Wordnet.Pr é hipônimo de um *synset* da Wordnet.Br) e EQ_HAS_HYPERONYM (um *synset* da Wordnet.Pr é hiperônimo de um *synset* da Wordnet.Br). Atualmente, a WordNet.Br possui a base de verbos criada, faltando apenas à validação dos *synsets* e dos alinhamentos.

³ Um *synset* é definido como um conjunto de palavras sinônimas

A VerbNet (<http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>) é um RLC verbal, para o inglês, com informações sintáticas e semânticas, que se utiliza das classes de Levin (1993). Este RLC possui 3769 lemas verbais e 274 classes (191 originais de Levin e 83 extensões). Cada classe verbal da VerbNet é descrita pelos componentes apresentados na Tabela 1 (membros, papéis temáticos e restrições seletivas, frames sintáticos e predicados semânticos).

Tabela 1. Classe Equip-13.4.2 da VerbNet

Equip-13.4.2		
Papéis temáticos e Restrições seletivas: Agent [+animate +organization], Theme e Recipient [+animate +organization]		
Membros: charge, invest, ply, arm, equip, rearm, redress, regale, reward, saddle, treat, armor, burden, compensate, encumber, overburden, weight		
Frames:		
NP V NP PP	Brown equipped Jones with a camera.	Agent V Recipient {with} Theme
Predicados semânticos:	(1) has_possession(start(E), Agent, Theme); (2) has_possession(end(E), Recipient, Theme); (3) transfer(during(E), Theme); (4) cause(Agent, E)	

Na primeira linha da Tabela 1 é apresentado o nome da classe (“equip-13.4.2”). A segunda linha contém os papéis temáticos e as restrições seletivas: *Agent* (Agente) (um executor de uma ação ou evento), *Theme* (Tema) (refere-se a participantes em um lugar ou submetidos a uma mudança de lugar) e *Recipient* (Receptor) (participante que é o alvo da transferência de uma entidade (concreta ou abstrata)). *Agent* e *Recipient* possuem a restrição seletiva [+animate | +organization] que define que este papel temático deve ser algo animado ou uma organização. Na terceira linha são apresentados os verbos membros da classe. A partir da quarta linha são apresentados os frames sintáticos, que descrevem as alternâncias sintáticas que a classe admite. Na quinta linha (coluna 1) há a apresentação da estrutura do único frame para esta classe: “NP V NP PP” (um sintagma nominal, seguido de um verbo, seguido de um sintagma nominal, seguido de um sintagma preposicional). Na linha 5 (coluna 2) há um exemplo para esta alternância (*Brown equipped Jones with a camera*). Na terceira coluna da linha 5 há a explicitação da alternância sintática (*Agent V Recipient {with} Theme*), ou seja, este frame pede um Agente seguido de um verbo que é seguido por um Receptor, seguido da preposição *with* que deve ser seguida de um Tema. Por fim, na linha 6, são apresentados os predicados semânticos (usados para denotar relações entre participantes e eventos). Como exemplo o predicado (1) (Tabela 1, linha 6, coluna 2) descreve que no início do evento (*start(E)*) o Agente tem a posse do Tema.

Cada membro das classes da VerbNet pode possuir alinhamentos com outros RLC’s como a WordNet.Pr, Xtag (XTAG Research Group, 2001), PropBank e FrameNet. No caso da WordNet.Pr, cada membro da VerbNet é alinhado a zero ou mais *synsets* da WordNet.Br. Na Figura 1 é apresentado um membro da classe “Equip-13.4.2” com os alinhamentos com a WordNet.Pr (etiqueta wn em XML).

```
<MEMBER name="charge" wn="charge&2:41:00 charge&2:32:01"/>
```

Figura 1. Membro da classe “equip-13.4.2” e os alinhamentos com a WordNet.Pr

3. Método de Criação da VerbNet.Br

O método de criação proposto é dividido em quatro etapas: uma manual e três automáticas. A etapa manual consiste da tradução das alternâncias sintáticas da VerbNet que podem ser diretamente traduzidas do inglês para o português. Então, utilizando o exemplo da Tabela 1, a estrutura NP V NP PP, em que PP corresponde à preposição *with* seguida de um NP, seria diretamente traduzida para NP V NP PP, com PP correspondendo a preposição “com” seguida de um NP. Nesta etapa, nos arquivos XML da VerbNet são inseridas etiquetas referente às alternâncias sintáticas do português para cada classe.

A segunda etapa (automática) consiste da busca das alternâncias sintáticas dos verbos em cópulas. Assim, para cada verbo encontrado no cópulas, haveria uma lista de alternâncias sintáticas. Por exemplo, para o verbo “trazer”, as seguintes alternâncias deveriam pertencer à lista: NP V NP (Maria trouxe o livro), NP V NP para NP (Maria trouxe o livro para a reunião), NP V NP de NP (Maria trouxe o livro de casa), e assim por diante. Resumindo, esta etapa consiste da busca por padrões “NP” “V” “PP” (com as preposições e os verbos explicitados) em um cópulas previamente anotado sintaticamente, como é o caso do PLN-BR-FULL (29.014.089 palavras) [Bruckschen et al., 2008] etiquetado com o *parser* PALAVRAS [Bick, 2000].

Já a terceira etapa (automática) consiste da busca dos candidatos a membros das classes da VerbNet.Br através dos mapeamentos entre a VerbNet e a WordNet.Pr e os alinhamentos entre a WordNet.Pr e a WordNet.Br (etapa apresentada na próxima seção).

Por fim, na figura 2 é ilustrada a quarta etapa (automática), na qual serão escolhidos os membros das classes da VerbNet.Br. Para esta escolha será realizada a comparação das alternâncias sintáticas admitidas por uma classe (etapa 1) com as alternâncias sintáticas das quais participam os verbos definidos como candidatos da classe (etapa 3) de acordo com a busca em cópulas (etapa 2). Os verbos que não participarem das alternâncias requeridas pela classe são descartados, sobrando somente os verbos considerados membros.

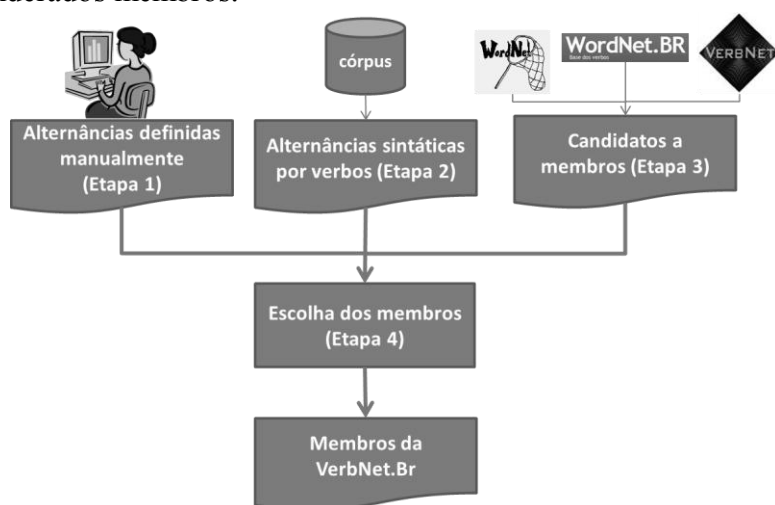


Figura 2. Ilustração da quarta etapa de construção da VerbNet.Br

Técnicas automáticas estão sendo exploradas para facilitar a realização da etapa 1. Para a etapa 2, estão sendo estudados trabalhos com frames de subcategorização

[Zanette, 2010; Altamirano and Alonso i Alemany, 2010] e, caso os resultados destes trabalhos sejam satisfatórios, eles serão utilizados nesta etapa. A etapa 3 já foi implementada e é apresentada na próxima seção.

4. Definição Automática dos Candidatos a Membros da VerbNet.Br

A terceira etapa de criação da VerbNet.Br consiste da definição de candidatos a membros das classes a partir dos mapeamentos entre a VerbNet e a WordNet.Pr e dos alinhamentos entre a WordNet.Pr e a WordNet.Br. Dentre os alinhamentos possíveis entre a WordNet.Br e a WordNet.Pr, escolheu-se abordar, inicialmente, o alinhamento ideal, EQ_SYNONYM, e o alinhamento que está próximo do ideal, EQ_NEAR_SYNONYM.

A VerbNet está disponível em formato XML, a WordNet.Br está armazenada em um banco de dados MySQL e a WordNet.Pr pode ser acessada através de arquivos texto estruturados. Esta etapa foi, então, implementada utilizando a linguagem C#.NET na plataforma Visual Studio 2008, o que facilitou o acesso aos arquivos XML's da VerbNet e ao banco de dados da WordNet.Br.

O método para definição dos candidatos a membros foi realizado seguindo o algoritmo (a Figura 3 ilustra este processo):

1. Para cada classe da VerbNet faça (arquivos em XML):
 - a. Para cada membro da classe faça:
 - i. Buscar os ILIs (índices inter-lingual) dos *synsets* da WordNet.Pr relacionados com o membro (arquivos texto estruturados)
 - ii. Para cada ILI encontrado faça:
 1. Buscar os *synsets* correspondentes na WordNet.Br (banco de dados MySQL)
 2. Para cada *synset* da WordNet.Br faça: Definir todos os verbos do *synset* como candidatos a membro

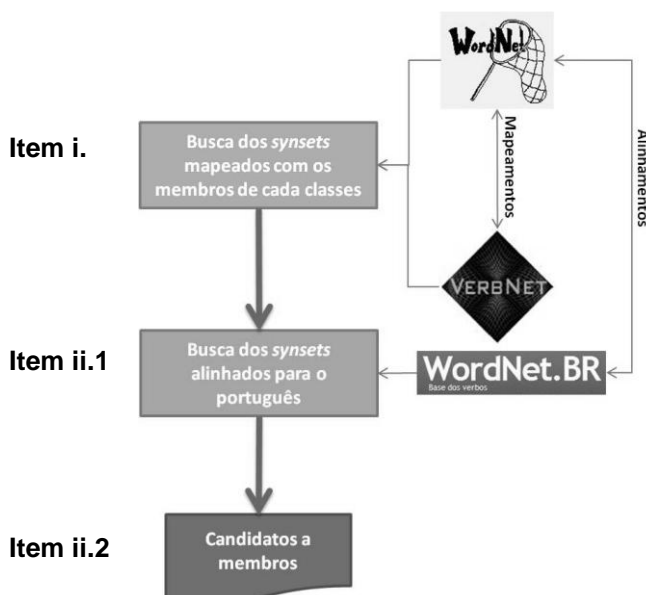


Figura 3. Ilustração da terceira etapa de construção da VerbNet.Br

Na Tabela 2 são apresentados os resultados do processo para alguns membros da classe “Equip-13.4.2”.

Tabela 2. Resultado do processo para alguns membros da classe “Equi-13.4.2”

Membros VerbNet	Mapeamentos WordNet.Pr	ILIs e <i>synsets</i> da WordNet.Pr	<i>Synset</i> WordNet.Br
Charge	charge%2:41:00	02401814 – {appoint, charge}	Sem relação
	charge%2:32:01	00725770 – {charge1, saddle, burden}	{incumbir, ocupar}
Invest	invest%2:41:03	02400129 – {endow, indue, gift, empower1, invest3, endue }	{beneficiar, bonificar, prove, dotar, favorecer, munir}
	invest%2:41:02	02315204 – { invest2, clothe, adorn}	Sem relação
Arm	arm%2:33:00	01052697 – {arm, build up, fortify, gird}	Sem relação
	arm%2:40:00	02267393 {arm}	{armar, municiar, munir}

Na coluna 1, são apresentados os membros da VerbNet cujos alinhamentos (coluna 2) possuíam ILI’s (coluna 3) associado aos *synsets* da WordNet.Br (coluna 4). Portanto, na última coluna da Tabela 2, são apresentados alguns dos candidatos a membros da classe “Equi-13.4.2” na VerbNet.Br. Como resultado deste processo obteve-se os candidatos a membros da VerbNet.Br somando 3908 verbos e 248 classes. Na Tabela 3, são apresentados 5 classes da VerbNet (“Accompany-51.7”, “Bring-11.3”, “Leave-51.2” e “Smell_emission-43.3”), seus membros (coluna 2) e os candidatos em português da VerbNet.Br (coluna 3).

Tabela 3. Exemplos de candidatos a membros da VerbNet.Br para 5 classes

Classe	Membros em inglês	Candidatos a membros em português
Accompany-51.7	accompany, conduct, escort, guide, lead, shepherd, steer	aconselhar, acompanhar, dirigir, encaminhar, seguir, escoltar, conduzir, ladear, guiar, levar orientar
Bring-11.3	take, bring	pegar, buscar, trazer, apanhar
Leave-51.2	abandon, split, desert, leave	fugir, sair, abandonar, abalar, desertar, desamparar, desaparecer, deixar, desabrigar, desproteger, largar
Smell_emission-43.3	reek, smell, stink	feder, catingar, tresandar

Outros resultados desta etapa são listas preliminares de verbos em português e os papéis temáticos que eles admitem. Na Tabela 4 são apresentados alguns exemplos de verbos no português que, a partir da VerbNet, devem admitir os papéis semânticos de Agente (coluna 1) e Experienciador (coluna 2) para seus complementos.

Tabela 4. Exemplos de verbos que admitem complementos com os papéis: agente ou experienciador

Agente	Experienciador
abrir, abusar, acabar, ampliar, assaltar, atirar, beber, bordar, bonificar, calcular, cancelar, cantar, capturar, cobiçar, danificar, dar, degradar, deixar, doar, educar, elogiar, enviar,...	achar, acalmar, acidentar, agonizar, amar, beijar, chatear, chorar, cobiçar, desdenhar, detestar, distrair, enamorar, enfiurecer, enjoar, ferir, gostar, ...

Todos estes resultados e listas estão disponíveis no portal PortLex (<http://www2.nilc.icmc.usp.br/portlex/>), destinado à divulgação de trabalhos com léxicos computacionais no português do Brasil.

5. Conclusão e Trabalhos Futuros

Com a conclusão da implementação de uma das etapas do método de construção da VerbNet.Br, pode-se começar a avaliação de seus resultados. Uma rápida análise dos

dados obtidos mostra que esta primeira etapa da criação da VerbNet.Br possui vários erros que são herdados da WordNet.Br. Apesar de alguns erros poderem ser eliminados na execução do método de criação (por exemplo, um verbo com grafia incorreta é eliminado por não ocorrer significativamente no corpus da etapa 2), uma validação da WordNet.Br é necessária e o apoio computacional a esta validação será realizado dentro do escopo do projeto VerbNet.Br. Como trabalhos futuros, tem-se a realização das outras etapas do método, a avaliação da VerbNet.Br comparada com métodos totalmente automáticos (aprendizado de máquina não supervisionado) e a avaliação extrínseca da VerbNet.Br em uma ferramenta de análise textual (Scarton e Aluísio, 2010).

A criação deste recurso léxico verbal para o português brasileiro será um importante avanço na área de semântica lexical no Brasil, pois com ele será possível realizar diversas tarefas de PLN que antes eram inviáveis. Dentre essas tarefas, podemos citar a desambiguação de palavras, e a extração e recuperação de informações. Além disso, se o método apresentar bons resultados, ele poderá ser reproduzido para qualquer língua que tenha uma *wordnet* alinhada através de ILI's com a WordNet.Pr, gerando um avanço na área de semântica lexical como um todo.

Agradecimentos

À FAPESP pela concessão de bolsas e de auxílios à pesquisa. Ao NILC-ICMC-USP pelo apoio institucional.

Referências

- Allbeck, J., Kipper, K., Adams, C., Schuler, W., Zoubanova, E., Badler, N., Palmer, M. and Joshi, A. (2002) "ACUMEN: Amplifying Control and Understanding of Multiple Entities", In *Proceedings of First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002)*, Bologna, Itália, p. 191-198.
- Altamirano, I. R. and Alonso i Alemany, L. (2010) "IRASubcat, a highly customizable, language independent tool for the acquisition of verbal subcategorization information from corpus", In *Proceedings of the Young Investigators Workshop on Computational Approaches to Languages of the Americas (YIWICALA '10)*, in conjunction with *NAACL HLT 2010*. Los Angeles, CA, USA, p. 84-91.
- Amaral, L. L. (2010) "O Verbos de Modo de Movimento no Português Brasileiro". Trabalho de Conclusão de Curso (Bacharel em Letras) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 53f.
- Baker, C. F., Fillmore, C. J. and Lowe, J. F. (1998) "The Berkeley FrameNet Project", In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, University of Montréal, Canadá, p. 86-90.
- Bertoldi, A. e Chishman, R. L. de O. (2009) "Desafios para a Criação de um Léxico baseado em Frames para o Português: um estudo dos frames Judgment e Assessing", In *Proceedings of the 7th Brazilian Symposium in Information and Human Language Technology (STIL 2009)*, São Carlos, SP, Brazil, 1 CD-ROM ISSN 2175-6201.

- Bick, E. (2000) "The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework". Ph.D. Thesis (Philosophy) – University of Aarhus, Dinamarca, 505f.
- Bruckschen, M., Muniz, F., Souza, J. G. C., Fuchs, J. T., Infante, K., Muniz, M., Gonçalves, P. N., Vieira, R. e Aluísio, S. M. (2008) "Anotação Lingüística em XML do Corpus PLN-BR". Série de Relatórios do NILC, NILC-TR-09-08, 39 p.
- Cançado, M. (1996) "Verbos Psicológicos: Análise Descritiva dos Dados do Português Brasileiro". *Revista de Estudos da Linguagem*, v. 4, n. 1, p. 89-114.
- Chagas de Souza, P. (2001) "Notas Sobre a Construção Adversativa", *Anais do 4º Encontro do Círculo de Estudos Linguísticos do Sul (CELSUL)*, Curitiba, PR, Brasil.
- Cruse, D. A. (1986) "Lexical Semantics". Cambridge, UK: Cambridge Textbooks in Linguistics, Cambridge University Press, 313p.
- Dias-da-Silva, B. C., Di Felippo, A. and Nunes, M. G. V. (2008) "The automatic mapping of Princeton WordNet lexicalconceptual relations onto the Brazilian Portuguese WordNet database", In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco, p. 1535-1541.
- Duran, M. S. (2009) "PropBank.BR: Regras Sintático-Semânticas para Mapeamento de Perguntas-Respostas de Verbos do Português e Anotação de Papéis Semânticos em um Corpus do Português do Brasil". Projeto de pós-doutorado aprovado pela FAPESP (processo: 2009/07394-9). ICMC-USP. Aprovado em maio de 2009.
- Fellbaum, C. (1998) "WordNet: An electronic lexical database". Cambridge, MA: MIT Press, 425p.
- Ferrer, E. E. (2004) "Towards a semantic classification of Spanish verbs based on subcategorisation information", In *Proceedings of the Workshop on Student research (ACLstudent 2004)*, in conjunction with *ACL 2004*, Barcelona, Espanha.
- Girju, R., Roth, D. and Sammons, M. (2005) "Token-level disambiguation of VerbNet classes", In *Proceedings of Interdisciplinary Workshop on the Identification and Representation of Verb Features and Verb Classes*, Saarbruecken, Germany.
- Jackendoff, R. (1990) "Semantic Structures". Cambridge, MA: MIT Press, 323p.
- Joanis, E. and Stevenson, S. (2003) "A general feature space for automatic verb classification", In *Proceedings of the 10th conference on European chapter of the Association for Computational Linguistics (EACL 2003)*, Budapest, Hungria, p. 163-170.
- Kipper, K. (2005) "Verbnet: A broad coverage, comprehensive verb lexicon". Ph.D. Thesis (Philosophy) - University of Pennsylvania, USA, 146f.
- Levin, B. (1993) "English Verb Classes and Alternations, A Preliminary Investigation", Chicago, IL: The University of Chicago Press, 348p.
- Merlo, P., Stevenson, S., Tsang, V. and Allaria, G. (2002) "A multilingual paradigm for automatic verb classification", In *Proceedings of the 40th Annual Meeting of the*

- Association for Computational Linguistics (ACL 2002)*, Philadelphia, PA, USA, p. 207-214.
- Moraes, H. R. (2008) “Aspectos sintaticamente relevantes do significado lexical: estudo dos verbos de movimento” Tese (Doutorado em Linguística e Língua Portuguesa) – Faculdade de Ciências e Letras, Universidade Estadual Paulista, Araraquara, 171f.
- Palmer, M., Gildea D. and Kingsbury, P. (2005) “The Proposition Bank: A Corpus Annotated with Semantic Roles”, *Computational Linguistics Journal*, v. 31, n. 1, p. 71-106.
- Salomão, Maria M. M. (2009) “FrameNet Brasil: Um trabalho em progresso”. *Revista Calidoscópio*, v. 7, n. 3, p. 171-182.
- Scarton, C. E. e Aluísio, S. M. (2010) “Análise da Inteligibilidade de textos via ferramentas de Processamento de Língua Natural: adaptando as métricas do Coh-Metrix para o Português”. *Linguamática (Revista para o Processamento Automático das Línguas Ibéricas)*, v. 2, n. 1, p. 45-61.
- Scarton, C. E. e Aluísio, S. M. (2009) “Herança Automática das Relações de Hiperonímia para a Wordnet.Br”. *Série de Relatórios do NILC. NILC-TR-09-10*, 48p.
- Schulte im Walde, S. (2006) “Experiments on the Automatic Induction of German Semantic Verb Classes”. *Computational Linguistics*, v. 32, n. 2, p. 159-194.
- Shi, L. and Mihalcea, R. (2005) “Putting pieces together: Combining FrameNet, VerbNet and WordNet for robust semantic parsing”, In *Proceedings of 6th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2005)*, Cidade do México, México, p. 99-110.
- Sun, L. and Korhonen, A. (2009) “Improving verb clustering with automatically acquired selectional preferences”, In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP 2009)*, Singapura, p. 638-647.
- Sun, L., Korhonen, A., Poibeau, T. and Messiant, C. (2010) “Investigating the cross-linguistic potential of VerbNet: style classification”, In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*, Beijing, China, p. 1056-1064.
- Vossen, P. (2004) “Eurowordnet: a multilingual database of autonomous and language specific wordnets connected via an interlingual-index”, *International Journal of Linguistics*, v. 17, n.2, p. 161-173.
- Zanette, A. (2010) “Aquisição de Subcategorization Frames para Verbos da Língua Portuguesa”. Trabalho de Conclusão de Curso (Bacharel em Ciência da Computação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 53f.