

Programa PIBIC/CNPq/UNESP

Relatório Final

(De 1º de março de 2011 a 31 de julho de 2012)

Rafael Luis Beraldo

Bolsista

Bento Carlos Dias-da-Silva

Orientador

1 Introdução

Este **Relatório Final** refere-se às atividades desenvolvidas pelo bolsista Rafael Luis Beraldo no período de 1º de março de 2011 a 31 de julho de 2012, após substituição da bolsista Débora Domiciano Garcia. O Plano de Atividades em que este relatório se pauta sustenta-se no projeto do orientador, “*O Desenvolvimento da Base de Substantivos da WordNet.Br e a sua Co-indexação com a WordNet de Princeton*”, que consiste na montagem de parte da rede WordNet.Br (WN.Br) — uma base relacional de unidades e expressões lexicais do português do Brasil, construída a partir da investigação de questões linguísticas (DIAS-DA-SILVA; OLIVEIRA; MORAES, 2002; DIAS-DA-SILVA, 2004), motivada pelo estudo da rede WordNet de Princeton (WN.Pr), na versão *WordNet 2.0*, e da metodologia empregada em sua construção por pesquisadores do Laboratório de Ciências Cognitivas da Universidade de Princeton, Estados Unidos (MILLER; FELLBAUM, 1991; FELLBAUM, 1998).¹

A WordNet.Br, como as demais wordnets, deverá constituir um recurso lexical que, além de servir de instrumento complementar para o estudo de línguas, poderá também ser utilizado em aplicações no âmbito do Processamento Automático de Línguas Naturais (PLN) e da Engenharia da Linguagem como, por exemplo, sistemas de tradução automática, motores de busca na Internet (Google, AltaVista, Lycos, entre outros), processadores e sumarizadores de texto, aplicativos de recuperação de textos e de informação em bases de documentos e, com a implementação do ILI (*Inter-Lingual-Index*), dicionários digitais inglês-português/português-inglês para acesso gratuito online na Internet.

Conforme os cronogramas de atividades (cf. os Planos do Bolsista em Anexo), de março de 2011 a julho de 2012, coube ao bolsista a realização de tarefas que permitiram o aprimoramento dos seus conhecimentos sobre o funcionamento das redes wordnet norte-americana e europeias, e a sua contribuição para o desenvolvimento da rede wordnet brasileira.

As atividades de pesquisa deste projeto realizaram-se no CELiC (Centro de Estudos Linguísticos e Computacionais da Linguagem), localizado na Sala 29, Prédio dos GPs, Faculdade de Ciências e Letras, UNESP, Campus de Araraquara, sede do Grupo de

¹O termo *wordnet* é internacionalmente aceito para designar um tipo particular de base de conhecimento lexical em que as relações que tecem a rede são relações léxico-conceituais (MARRAFA, 2001).

Pesquisa “Grupo de Estudos Linguístico-Computacionais da Linguagem Humana”, liderado pelo orientador, cadastrado no CNPq e certificado pela Universidade.

2 Objetivos

Os principais objetivos das atividades do bolsista, previstos no Plano de Atividades para o período de março de 2011 a julho de 2012, incluíram:

- a. Estudar os tópicos de semântica lexical referentes à classe dos substantivos, selecionados da Bibliografia Básica;
- b. Proceder à análise léxico-semântica dos synsets do português extraídos da base da WordNet.Br. (em construção), a partir da projeção da análise de mesma natureza de 700 synsets (números 300–499 e números 1000–1499) de substantivos do arquivo de base `06artifact(11256).doc` do inglês, extraídos da base da WordNet de Princeton, dando continuidade ao trabalho de análise anterior;
- c. Para cada synset do português selecionado em (b), analisar a sua boa-formação gráfica e a sua consistência léxico-conceitual, ou seja, corrigir os itens lexicais mal-grafados e verificar se todos eles lexicalizam um mesmo conceito; atividade de análise que se realiza pautada nos estudos de semântica lexical, na análise das concordâncias e na consulta aos dicionários Webster’s eletrônico, RHUD, Michaelis, Aurélio e Houaiss referidos na Bibliografia Básica; caso seja necessário, novos synsets do português são criados nesse processo;
- d. Participar de encontros semanais de orientação;
- e. Elaborar relatório científico e produzir um trabalho para a apresentação em congresso de IC.

3 Metodologia

Nessa seção listaremos os fundamentos teórico-aplicados utilizados na análise dos synsets, bem como elucidaremos as estratégias e convenções dessa análise. Esses dois aspectos da metodologia do projeto permitem o estabelecimento de um estilo padrão para os synsets da WordNet.Br. Também discutiremos a implementação do Git (GIT, 2012), software que foi selecionado para otimizar o armazenamento e o gerenciamento dos diferentes tipos de arquivo que são gerados na execução das atividades de construção da WordNet.Br.

3.1 Metodologia de base

Antes que qualquer análise de synsets seja feita, é necessário compreender os aspectos teóricos por trás da WordNet de Princeton (doravante WN.Pr), produzida para o inglês norte-americano, os pressupostos para o alinhamento de wordnets e as particularidades da WordNet.Br (doravante WN.Br). Essa necessidade foi suprida com o estudo dos textos da Bibliografia Básica.

Destacamos alguns textos fundamentais, como Fellbaum (1998), que descreve minuciosamente a WN.Pr, e Vossen (1998), uma edição especial do periódico *Computers and the Humanities*, que discute as dificuldades e estratégias na construção da EuroWordNet, uma mega-WordNet que associa diferentes wordnets em construção para as línguas da União Européia. Incluímos nessa lista Riemer (2010), que fornece o embasamento teórico que precedeu àquele dos textos acima, discutindo teorias e conceitos da semântica (grasso e lexical) que se provaram bastante úteis na compreensão das wordnets, em especial o capítulo 8.

3.2 Estratégias de execução

De acordo com os fundamentos teóricos fornecidos pela Bibliografia Básica, para cada synset selecionado da base da WN.Br, após análise léxico-semântica, propõe-se uma glosa.

A metodologia aplicada neste primeiro mês de pesquisa apresentou grande eficiência, o que diminui a probabilidades de que algum procedimento seja alterado até julho.

Procedimentos:

1. Iniciam-se os seguintes aplicativos:
 - a. 06artifact(11256).doc;
 - b. Arquivo modelo para posterior indexação (.txt);
 - c. WordNet 2.0;
 - d. Editor da WordNet.Br;
 - e. Dicionários de inglês/inglês e inglês/português;
 - f. Navegador de internet com acesso ao corpus (NILC) e à ferramentas de busca (Google, AltaVista, Yahoo).
2. Seleciona-se o synset do inglês que dará origem à análise ou construção do synset do português na lista de synsets do tipo 06artifact(11256).doc e localiza-se esse synset na base da WN.Pr; todo seu conteúdo (o synset e todas as informações associadas a ele, incluindo todos os seus hiperônimos) é copiado para o arquivo modelo em .txt.
3. Copia-se o synset do português semanticamente equivalente ao inglês, se ele já existir na base da WN.Br, para o arquivo referido em 2, analisando-o quanto à sua boa-formação ortográfica e semântico-conceitual (esta análise consiste em: ampliar ou reduzir o synset, segmentá-lo em dois ou mais synsets ou excluí-lo da base do português); caso não haja um synset equivalente na base do português, constrói-se um novo synset que se alinhe conceitualmente ao synset do inglês referido em 2, ou seja, procura-se construir o synset português cujas unidades lexicais sejam traduções das unidades lexicais do synset do inglês.
4. Durante os procedimentos em 3, observa-se qual destas análises que, previstas em Vossen et al. (1998), será a mais apropriada para especificar o alinhamento conceitual que resulta da análise: alinhamento por EQ_SYNONYM, por EQ_NEAR_SYNONYM, por EQ_HAS_HYPONYM ou por EQ_HAS_HYPERONYM.

5. Cria-se a glosa para o synset do português em construção, com base na glosa já existente para o synset do inglês, quando esta já estiver especificada no synset do inglês.
6. Para cada unidade lexical que constitui o synset do português, selecionam-se frases-exemplo em corpus. As frases localizadas no corpus do Núcleo Interinstitucional de Linguística Computacional (CORPUS..., 2012) têm prioridade sobre as frases localizadas pelas ferramentas de busca da web; usa-se o critério de maior frequência de ocorrência da unidade lexical em corpora para decidir a seleção das frases-exemplo; quando a frase é selecionada na web, sinaliza-se o fato, acrescentando-se “I]” (“proveniente da Internet”) a unidade lexical da frase que corresponde à unidade lexical do synset.
7. Define-se a chave do synset, que é a unidade lexical pertencente a ele e que é a mais frequente na busca em corpus ou a intuitivamente mais usada.
8. Conclui-se a análise e nomeia-se o arquivo segundo este padrão:

```
<nº_na_Base_WNBr>.<ILI>.<tipo_semântico>.  
<chave>.void|md|cr.HYPER|HYPO|NEAR|void>
```

3.3 Estratégia de gerenciamento dos arquivos

A dimensão do desafio que é gerenciar a base de arquivos da WN.Br pode ser apreciada considerando-se que, entre 2010 e 2012, produzidos e revistos cerca de 4100 arquivos, correspondendo ao mesmo número de synsets produzidos e alinhados. Cinco bolsistas de IC supervisionados pelo coordenador estiveram envolvidos nas tarefas, que significaram a exploração de diferentes análises até a escolha final de cada synset considerado ideal para configurar na base da WN.Br; análises cujos históricos foram documentados para que toda e qualquer decisão analítica pudesse ser recuperada, revista e atualizada.

A solução em teste para o desafio é a adoção do Git, um sistema de controle de versões de arquivos (CHACON, 2009), que deve viabilizar:

1. O controle do histórico da construção e do alinhamento de synsets (Que análise foi feita? Quem a fez? Quando?);
2. O registro de análises distintas para o mesmo synset, resultando em alinhamentos diferentes;
3. O acesso, via Internet, a todos os arquivos gerados pelos desenvolvedores do projeto, pois cada desenvolvedor pode copiar todos os arquivos da base da WN.Br (no servidor) para a sua máquina local durante o seu trabalho, utilizar aqueles que são relevantes para as suas análises sem interferir na integridade dos mesmos, pois trabalha com uma cópia e seu acesso é restrito no servidor;
4. A marcação e o resgate de uma versão do arquivo que registra o synset e o seu alinhamento considerados ideais;
5. A consolidação da base da WN.Br.

Um Sistema de Controle de Versões (SCV) resolve todos esses problemas, permitindo ao usuário reverter seus arquivos a um estado anterior, reverter o projeto inteiro a um estado anterior, comparar as mudanças ao longo do tempo, ver quem foi o último a modificar os arquivos, e recuperar arquivos que foram comprometidos ou deletados por acidente².

A escolha do sistema de controle de versões Git foi motivada por ser rápido, estável, possuir todas as características apontadas acima e por sua licença de software livre³. Prova de que o Git é capaz de gerenciar a base da WN.Br é sua história. O programa foi criado, originalmente, para gerenciar as mudanças no Linux, projeto para o qual mais de 7800 pessoas e quase 800 empresas já contribuíram (CORBET; KROAH-HARTMAN; MCPHERSON, 2012, p. 1).

Durante reuniões no final de 2010 e início de 2011, foram discutidos os detalhes iniciais para a implementação do Git (BERALDO; DIAS-DA-SILVA, 2011b, p. 4). O Git foi implementado e testado, usando um servidor cedido pelo NILC, no ICMC (Instituto de Ciências Matemáticas e Computação) da USP de São Carlos. Em reunião no início do mês de agosto, foi decidido que o repositório da WordNet.Br será movido do atual servidor de acesso privado no NILC para o servidor de acesso público no GitHub⁴. Esse movimento busca promover a divulgação do projeto e dos resultados obtidos até o momento, bem como promover a comunicação entre os pesquisadores e demais interessados. A figura 1 apresenta a interface de um repositório configurado no GitHub.

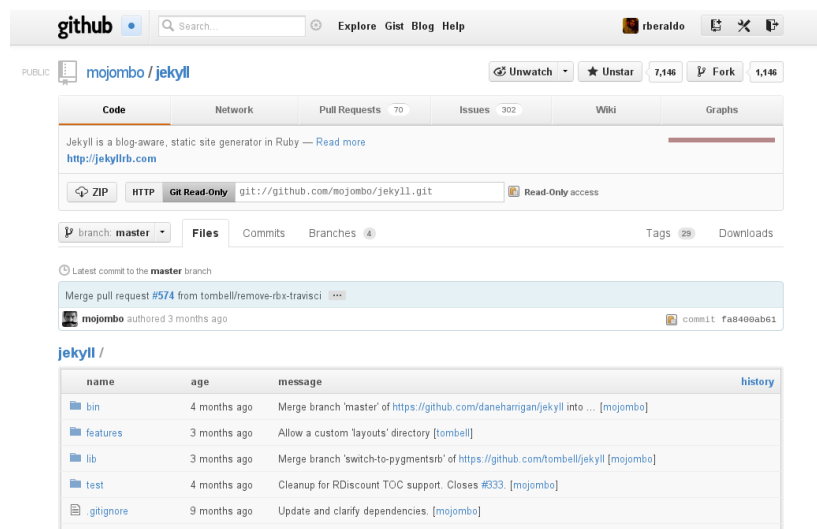


Figura 1: A interface de um repositório no GitHub

É possível adicionar pessoas a um projeto, para que possam contribuir ativamente; o único prerequisite é que possuam uma conta no GitHub. Também é possível aceitar alterações e melhorias em geral (*"pull requests"*) de outros usuários do GitHub. Desse modo, tem-se o controle de quem fez uma modificação ou adição,

²Chacon (2009): <http://git-scm.com/book/en/Getting-Started-About-Version-Control>

³<https://github.com/git/git/blob/master/COPYING>

⁴<https://github.com/>

ou mesmo deleção, quando foi feita e por qual motivo. O sistema do GitHub também permite que usuários copiem para suas contas o estado atual do projeto sem que nenhuma modificação seja feita nos arquivos originais. Usuários não cadastrados também podem fazer uma cópia local dos arquivos. Em nenhum desses casos há a possibilidade de alteração da base por pessoas não autorizadas.

Além disso, o GitHub possui servido de Wiki (figura 2) e um sistema de comunicação de *bugs*, ou seja, erros ou problemas (figura 3).

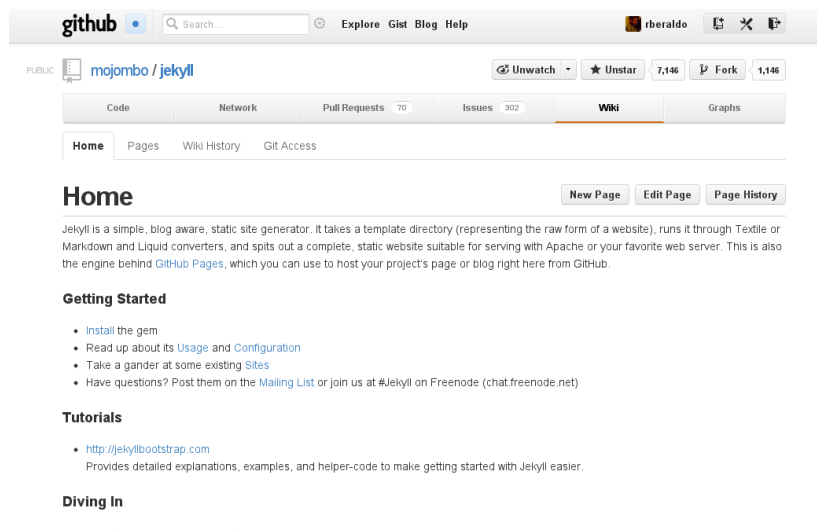


Figura 2: A Wiki fornecida pelo GitHub pode conter a documentação do projeto

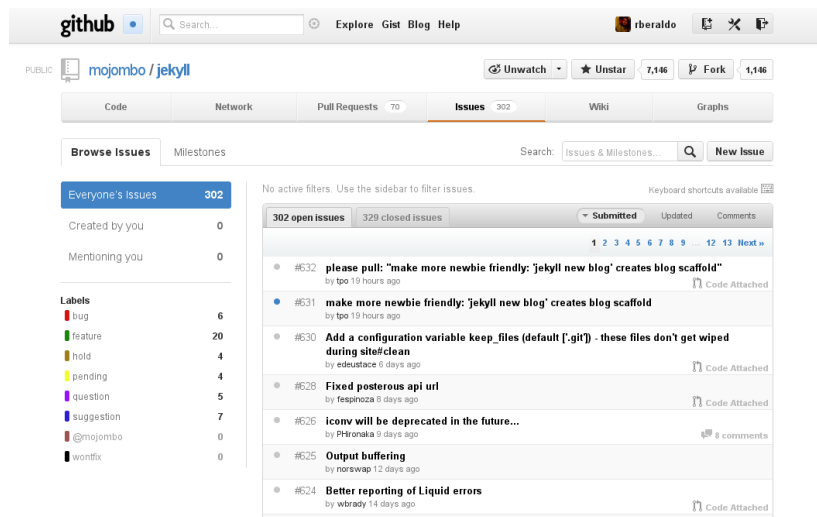


Figura 3: O *bug tracker* fornecido pelo GitHub pode ajudar a resolver problemas na WordNet.Br

A área de ajuda do GitHub⁵ possui uma página explicando a configuração necessária para acessar os repositórios disponíveis em todo o site⁶, que difere em cada sistema operacional. A área também possui uma parte dedicada ao uso do Git⁷, onde o usuário pode encontrar referências, entre livros e vídeos, que explicam o fluxo do trabalho usando os vários clientes disponíveis para o Git, dos mais tradicionais, como a linha de comando, aos mais amigáveis para o usuário, como os clientes fornecidos pelo GitHub⁸.

Com a adoção do sistema de controle de versões Git, espera-se a agilização da construção, do armazenamento, da revisão e da atualização da base da WN.Br e, sobretudo, a sua disponibilização e as suas atualizações.

4 Resultados

Ao fim de um ano e cinco meses de trabalho, verifica-se o cumprimento de todos os objetivos programados para esse período, incluindo o aprofundamento dos conhecimentos teóricos e metodológicos, a realização do trabalho aplicado de alinhamento entre substantivos das duas bases, proporcionando ao bolsista a oportunidade de adquirir experiências em desenvolver pesquisa teórico-empírica, e também o conhecimento relacionado às wordnets.

A tabela 1 indica os objetivos para os primeiros cinco meses de trabalho, ao passo que a tabela 2 apresenta os objetivos para os doze meses finais. Os Apêndices A e B apresentam as tabelas completas de todos os alinhamentos realizados para o período da bolsa.

Ao final de	Realizar as atividades	Com a previsão dos seguintes resultados
5 meses	Além dos estudos teórico-metodológicos, as atividades incluem, sobretudo, a especificação das frases-exemplos para cada um dos itens lexicais constitutivos de cada synset do português que puder ser indexado a cada um dos 200 synsets do inglês (ARTIFACT), a análise, a revisão e a glosagem dos synsets de substantivos do português e o seu alinhamento ao respectivo synset do inglês (a indexação).	Domínio das técnicas de análise léxico-semântica, inserção dos synsets indexados na base da WordNet.Br, estabelecendo da correspondência entre estes e os seus equivalentes semânticos na WordNet de Princeton e a especificação das relações de antonímia, hiponímia e meronímia entre os synsets do português, contribuindo assim para a construção da base dos substantivos do domínio ARTIFACT da WordNet.Br. (00300 a 00499)

Tabela 1: Panorama dos primeiros cinco meses

⁵<https://help.github.com/>

⁶<https://help.github.com/articles/set-up-git>

⁷<https://help.github.com/categories/19/articles>

⁸Cliente para Windows disponível em <http://windows.github.com/> e para Mac em <http://mac.github.com/>

Ao final de	Realizar as atividades	Com a previsão dos seguintes resultados
12 meses	Além dos estudos teórico-metodológicos, as atividades incluem, sobretudo, a especificação das frases-exemplos para cada um dos itens lexicais constitutivos de cada synset do português que puder ser indexado a cada um dos 200 synsets do inglês (ARTIFACT), a análise, a revisão e a glosagem dos synsets de substantivos do português e o seu alinhamento ao respectivo synset do inglês (a indexação).	Domínio das técnicas de análise léxico-semântica, inserção dos synsets indexados na base da WordNet.Br, estabelecendo da correspondência entre estes e os seus equivalentes semânticos na WordNet de Princeton e a especificação das relação de antonímia, hiponímia e meronímia entre os synsets do português, contribuindo assim para a construção da base dos substantivos do domínio ARTIFACT da WordNet.Br. (01000 a 01499)

Tabela 2: Panorama dos doze meses finais

5 Discussão

Na busca de um synset do português que seja conceitualmente equivalente ao do inglês, nem sempre ocorre um alinhamento direto, devido às especificidades de cada uma das línguas e, sobretudo, às lacunas lexicais que são constatadas durante o processo de análise.

Segundo Peters et al. (1998), há seis tipos de relações interlinguais. O alinhamento direto é chamado EQ_SYNONYM. As outras relações mais importantes são (PETERS et al., 1998, p. 225, tradução livre):

- EQ_NEAR_SYNONYM quando um sentido corresponde a vários ILIS simultaneamente;
- EQ_HAS_HYPERONYM quando um sentido é mais específico que qualquer ILI disponível;
- EQ_HAS_HYPONYM quando um sentido pode ser ligado apenas a ILIS mais específicos.

Exemplo de EQ_SYNONYM é o synset da WN.Pr { *bread knife* }, que se alinha diretamente com o synset da WN.Br { *faca de pão* }. Nesse caso, verificamos que as duas línguas — o inglês norte-americano e o português brasileiro — revestem lexicalmente o mesmo conceito de uma faca especial para cortar pães. Mesmo que esse tipo de alinhamento pareça bastante simples, existem algumas questões a serem levantadas. Algumas são de cunho linguístico e se relacionam às especificidades de cada língua, como no caso do synset { *brushz* } da WN.Pr. Nesse caso, o synset preenche lexicalmente um “utensílio que tem pelos ou cerdas firmemente presas a um cabo”. Em português brasileiro, existem ao menos dois possíveis synsets que se encaixam nessa descrição: { *escova* } e { *pincel* }. A dificuldade deriva do fato que, em português, o conceito é bipartido em função do uso das escovas e pincéis; podemos pensar que escovas *limpam* ao passo que pincéis *espalham* (tinta, claras de ovos etc.). No inglês, essa diferenciação ocorre nos termos *cleaning brush* e *paint brush*, ambos tipos de *brushes*.

Outro tipo de problema decorre de possíveis problemas na formação dos synsets da WN.Pr. Esse é o caso de { *bitmap*, *electronic image* }. Aqui é necessária a compreensão técnica do termo “*bitmap*”. Segundo (BITMAP, 2012):

[...] *bitmap* or *pixmap* is a type of memory organization or image file format used to store digital images. The term *bitmap* comes from the computer programming terminology, meaning just a map of bits, a spatially mapped array of bits.

O problema é, portanto, que *bitmap* é um *tipo* de imagem digital. Acreditamos que, por esse motivo, { *bitmap* } deveria ser um hipônimo de { *electronic image* }.

Finalmente, alguns synsets são, por natureza, impossíveis de serem alinhados por EQ_SYNONYM. É o caso dos synsets { *britches* } (termo informal para *breeches*, “calção”), { *bootz* } (termo britânico para “porta-malas”) e { *burthen* } (termo informal para *burden*, “fardo”). Esses synsets apresentam variações no registro (formal, informal, norte-americano, britânico) do inglês que não dizem respeito ao conceito que preenchem. Para efeito de comparação, embora o termo “Sampa” seja informal para “São Paulo”, ambas as palavras preenchem lexicalmente o mesmo conceito — o da capital do Estado de São Paulo — e devem, por esse motivo, formar um mesmo synset.

O alinhamento por EQ_NEAR_SYNONYM se dá quando um sentido pode se alinhar a múltiplos sentidos na WN.Pr. Já o alinhamento por EQ_HAS_HYPERONYM acontece quando um conceito falta no inglês ou na WN.Pr, como ilustrado na figura 4.



Figura 4: Relação do tipo EQ_HAS_HYPERONYM. Adaptado de Garcia e Dias-da-Silva (2010).

Quando a wordnet de Princeton apresenta um synset mais específico do que o presente na wordnet a ser alinhada, acontece o alinhamento por EQ_HAS_HYPONYM. O exemplo dado por Garcia e Dias-da-Silva (2010) é representado na figura 5.

Embora o alinhamento por EQ_SYNONYM seja o mais abundante em termos absolutos, o alinhamento por EQ_HAS_HYPONYM ocorre em grande número. Dos 500 synsets analisados nos últimos doze meses de trabalho, aproximadamente 150 deles ($\approx 30\%$) apresentam esse alinhamento. Exemplo desse tipo de alinhamento é o

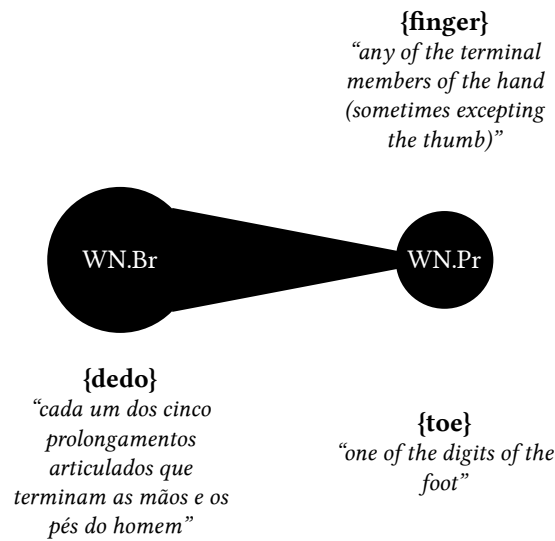


Figura 5: Relação do tipo EQ_HAS_HYPONYM. Adaptado de Garcia e Dias-da-Silva (2010).

synset da WN.Br { *bomba guiada por laser, bomba inteligente* }, que se alinha por EQ_HAS_HYPONYM ao synset da WN.Pr { *Bunker Buster, Guided Bomb Unit-28, GBU-28* }, que representa um tipo específico de bomba guiada por laser. Também { *arado* } foi alinhado dessa maneira a { *bull tongue* }, um tipo de arado pesado usado em plantações de algodão.

6 Atividades de divulgação

Durante o período da bolsa, realizou-se apresentação de painel (ARAUJO et al., 2011) no 59º seminário do Grupo de Estudos Linguísticos do Estado de São Paulo (GEL, em Bauru, SP, sob o nome de *As wordnets enquanto ontoléticos*, motivado por discussões em Huang et al. (2010).

Também em 2011, foi enviado resumo para o Congresso de Iniciação Científica (CIC) (BERALDO; DIAS-DA-SILVA, 2011a) da UNESP, sob o título de *Gerenciamento da rede semântica WordNet.Br*, motivado pela implementação do software Git para o gerenciamento dos resultados linguísticos do projeto WordNet.Br. Na mesma linha, dessa vez em 2012, foi enviado o resumo da apresentação *Gerenciamento lexical da WordNet.Br* (BERALDO; DIAS-DA-SILVA, 2012) para o CIC que ocorrerá nos primeiros dias de outubro na Faculdade de Ciências e Letras de Araraquara da UNESP.

7 Considerações finais

As tarefas previstas para o período de março de 2011 a julho de 2012 envolveram a análise, glosagem e indexação de 700 synsets do tipo ARTIFACT. Além da construção dessa parte da WordNet.Br, o bolsista realizou a divulgação do trabalho por meio da

apresentação de cartazes e eventos científicos. Também foram feitos esforços para permitir um melhor gerenciamento das várias partes do projeto, especialmente da base de synsets.

As atividades até então desenvolvidas se deram em contato direto com a equipe de pesquisa coordenada pelo orientador, e iniciação do bolsista em pesquisas teórico-empíricas sobre Semântica Lexical, com ênfase na aplicação desta no âmbito da Linguística Computacional que colaboraram para a qualificação das atividades metodológicas realizadas.

Cumpre, por fim, informar que, com o término das atividades do bolsista Rafael Luis Beraldo e na ausência de um bolsista substituto, a bolsa teve de ser cancelada. O projeto de construção da wordnet brasileira entrará, agora, em nova fase, com a utilização do GitHub para o gerenciamento do projeto e relacionamento com outros pesquisadores e interessados no projeto.

Bibliografia

- ARAUJO, F. F.; BERALDO, R.; GARCIA, D. D.; DIAS-DA-SILVA, B. C. As wordnets enquanto ontológicos. In: SEMINÁRIO DO GEL, 59. *Programação do 59º Seminário do GEL*. Bauru: GEL, 2011. Disponível em: <<http://gel.org.br/detalheResumo.php?trabalho=7628/>>. Acesso em: 21 jul. 2011.
- BERALDO, R.; DIAS-DA-SILVA, B. C. *Gerenciamento da rede semântica WordNet.Br*. São Paulo: UNESP, 2011. Trabalho apresentado ao 23º Congresso de Iniciação Científica da UNESP.
- BERALDO, R.; DIAS-DA-SILVA, B. C. *Relatório parcial*. Araraquara, 2011. 10 p.
- BERALDO, R.; DIAS-DA-SILVA, B. C. *Gerenciamento da base lexical da WordNet.Br*. São Paulo: UNESP, 2012. Trabalho apresentado ao 24º Congresso de Iniciação Científica da UNESP.
- BITMAP. Wikimedia, 2012. Disponível em: <<http://en.wikipedia.org/wiki/Bitmap>>. Acesso em: 26 ago. 2012.
- CHACON, S. *Pro Git*. New York: aPress, 2009. Disponível em: <<http://git-scm.com/book>>. Acesso em: 27 ago. 2012.
- CORBET, J.; KROAH-HARTMAN, G.; MCPHERSON, A. *Linux Kernel Development: How fast it is going, who is doing it, what they are doing, and who is sponsoring it*. San Francisco: The Linux Foundation, 2012. Disponível em: <<http://go.linuxfoundation.org/who-writes-linux-2012>>. Acesso em: 27 ago. 2012.
- CORPUS do NILC. 2012. Disponível em: <<http://www.linguatca.pt/acesso/corpus.php?corpus=SAOCARLOS>>. Acesso em: 23 ago. 2012.
- CRUSE, D. A. *Lexical semantics*. Cambridge, Mass.: Cambridge University Press, 1986.
- DIAS-DA-SILVA, B. C. *Montagem da Base da Wordnet para o Português do Brasil*. Araraquara: CELiC, FCL, UNESP, 2004. 50 p.

- DIAS-DA-SILVA, B. C.; OLIVEIRA, M. F.; MORAES, H. R. Groundwork for the development of the brazilian portuguese wordnet. In: RANCHHOD, E. M.; MAMEDE, N. J. (eds.). *Advances in natural language processing*. Berlin: Springer-Verlag, 2002. p. 189–196.
- EUROWORDNET. *Rede EuroWordNet*. Disponível em: <<http://www.illc.uva.nl/EuroWordNet/data/sampleData.html>>. Acesso em: 14 fev. 2009.
- FELLBAUM, C. (Ed.). *WordNet: an electronic lexical database*. Cambridge, Mass./London: The MIT Press, 1998.
- FERREIRA, A. B. H. *Dicionário Aurélio eletrônico século XXI (v. 3.0)*. São Paulo: Lexikon Informática Ltda., 1999.
- FLEXNER, S. B. (Ed.). *Random House Webster's unabridge electronic dictionary (v. 2.0)*. New York: Random House Inc., 1997.
- GARCIA, D. D.; DIAS-DA-SILVA, B. C. *O alinhamento de wordnets*. Araraquara: CELiC, FCL, UNESP, 2010.
- GIT. 2012. Disponível em: <<http://git-scm.com/>>. Acesso em: 23 ago. 2012.
- HANDKE, J. *The structure of the lexicon: human versus machine*. Berlim: Mouton de Gruyter, 1995.
- HOUAISS, A.; CARDIM, I. *Webster's Dicionário Inglês–Português*. Rio de Janeiro: Record, 2005.
- HOUAISS, A.; VILLAR, M. S. *Dicionário Houaiss da língua portuguesa*. Rio de Janeiro: Objetiva, 2001.
- HUANG, C.-R. et al. *Ontology and the lexicon: A natural language processing perspective*. Cambridge: Cambridge University Press, 2010. (Studies in Natural Language Processing).
- LEVIN, B. *English verb classes and alternations: a preliminary investigation*. Chicago: The University of Chicago Press, 1993.
- MARRAFA, P. *WordNet do Português: Uma base de dados de conhecimento linguístico*. Lisboa: Instituto Camões, 2001.
- MILLER, G.; FELLBAUM, C. Semantic networks of english. *Cognition*, v. 41, n. 1–3, p. 197–229, 1991.
- PETERS, W.; VOSSEN, P.; DÍEZ-ORZAS, P.; ADRIAENS, G. Cross-linguistic alignment of wordnets with an inter-lingual-index. *Computers and the humanities*, Kluwer Academic Publishers, Dordrecht/Boston/London, v. 32, p. 221–251, 1998.
- REDE WordNet de Princeton on-line. Disponível em: <<http://wordnetweb.princeton.edu/perl/webwn>>. Acesso em: 14 fev. 2009.
- RIEMER, N. *Introducing semantics*. New York: Cambridge University Press, 2010.
- SAINT-DIZIER, P.; VIEGAS, E. (eds.). *Computational lexical semantics*. Cambridge: Cambridge University Press, 1995.

- SCOTT, M. *WordSmith Tools version 3.0*. Oxford: Oxford University Press, 2001.
- TAYLOR, J. L. *Webster's Portuguese-English Dictionary*. Rio de Janeiro: Record, 2003.
- VOSSEN, P. Special issue on EuroWordNet. *Computers and the humanities*, Kluwer Academic Publishers, Dordrecht/Boston/London, v. 32, n. 2-3, 1998.
- VOSSEN, P. et al. Compatibility in interpretation of relations in EuroWordNet. *Computers and the humanities*, Kluwer Academic Publishers, Dordrecht/Boston/London, v. 32, p. 153-184, 1998.
- WEISZFLOG, W. (Ed.). *Michaelis português: moderno dicionário da língua portuguesa* (v. 1.1). São Paulo: DTS Software Brasil Ltda., 1998.