

Uma proposta de análise lógico-semântica dos verbos de “criação/transformação” com fins de aplicação computacional

Autora: Patrícia Tosqui – UNESP/Araraquara (PG) - patriciatosqui@uol.com.br
Endereço Postal: R. Cel. Orlando Pereira Barreto, 552 - Jd. Estrela CEP: 15070-190
São José do Rio Preto – SP

1. Introdução

O campo de estudos denominado Processamento Automático de Línguas Naturais (PLN) pode contribuir significativamente para o trabalho lingüístico, em especial para a lexicografia. Da mesma forma, esta contribui para o desenvolvimento de "léxicos computacionais" para fins de PLN. Kilgarriff (2000:02), afirma que “o PLN precisa de dicionários, e os lexicógrafos podem se aproveitar do PLN para fazer dicionários melhores, de modo que há grande potencial para sinergia entre as duas atividades”.

Em meu trabalho, proponho realizar uma análise léxico-semântica e lógico-conceptual dos verbos de "criação/transformação", com vistas a aplicações tecnológicas. A pesquisa vem sendo desenvolvida em dois domínios: lingüístico (a análise sintático-semântica propriamente dita) e lingüístico-computacional (proposição de uma forma de representação dos dados lingüísticos de modo que possam ser utilizados para a elaboração de aplicativos eletrônicos). A opção por essa classe semântica justifica-se pelo fato de que esses verbos são muito utilizados em textos técnicos e científicos, e creio que um estudo mais aprofundado sobre eles poderá ser aproveitado, no futuro, para a elaboração de aplicativos que auxiliem usuários em pesquisas de diversas áreas do conhecimento.

Não apresento aqui resultados ou conclusões sobre o tema acima explicitado. O objetivo deste trabalho é apresentar os questionamentos que me levaram a propor uma pesquisa de Doutorado, ainda em andamento, mostrar algumas análises iniciais e discutir possibilidades para se chegar a um resultado final. Apresento também a metodologia que será aplicada na parte lingüístico-computacional, a fim de possibilitar um debate e troca de experiências que possam vir a melhorar o trabalho.

2. Fundamentação Teórica

De acordo com Levin & Rappaport-Hovav (1991:124), o léxico tem assumido uma posição central em muitas teorias sintáticas recentes. Estudos apontam que muitos aspectos das configurações sintáticas de verbos e outros elementos parecem ser projeções de suas propriedades lexicais. Conseqüentemente, diversas pesquisas se dedicam a investigar a natureza de sua representação lexical. Para as autoras, uma entrada lexical ideal deveria minimizar o volume de informações necessárias para se depreender o sentido de um item. Isso pode ser feito fatorando-se qualquer informação previsível das entradas lexicais. O significado de uma palavra deve ser parte de sua entrada lexical, mas, será que só isso é suficiente para sua compreensão? A questão levantada pelas autoras é determinar se as propriedades sintáticas podem ou não ser deduzidas a partir do componente semântico de qualquer palavra. Com base na metodologia proposta por Levin & Rappaport-Hovav (1991) e Levin (1993), a estratégia usada nesta pesquisa será investigar a natureza do conhecimento lexical que o falante de português possui a respeito de certos verbos aparentemente relacionados

semanticamente. Para as autoras, apenas após a realização de diversos testes sintáticos de padrões de comportamento é possível determinar se dois verbos fazem parte do mesmo grupo ou de grupos semânticos diferentes. Os componentes de sentido relevantes para a caracterização de cada grupo são identificados isolando-se aqueles componentes de sentido compartilhados pelos membros de cada grupo.

Para melhor esclarecer, apresento exemplos de testes realizados por Levin (1989). Início com o verbo *construct*, do subgrupo dos “*build verbs*” aplicando testes ao verbo construir, em língua portuguesa:

- ❑ alternância material/produto:
 - (1) a. João construiu uma casa (de tijolos).
 - b.* João construiu tijolos em uma casa.
- ❑ alternância de beneficiário:
 - (2) a. João construiu uma casa para mim.
 - b. João construiu-me uma casa.
- ❑ alternância causativa:
 - (3) a. João construiu a casa.
 - b.* A casa construiu.
- ❑ alternância sujeito/matéria-prima:
 - (4) a. João construiu uma casa com aqueles tijolos novos.
 - b.* Aqueles tijolos construíram uma casa.

No exemplo, podem-se observar algumas propriedades do verbo construir em português. Tais propriedades são as mesmas que se aplicam ao verbo *construct*, em inglês. Essas características podem ser atribuídas a verbos de outros subgrupos, como por exemplo preparar, como na oração-modelo *João preparou um sanduíche de queijo.*), mas não a todos os verbos de criação/trans formação.

Vejamos outro exemplo, desta vez com verbos como cultivar e crescer, do subgrupo “*grow verbs*”:

- (5) a. João cultiva roseiras.
- b. * As roseiras cultivam.

- (6) a. * João cresce roseiras.
- b. As roseiras crescem.

Percebemos aqui uma diferença entre os usos do verbo crescer em português e *grow* em inglês. Em inglês, seria totalmente aceitável, por exemplo, a oração:

- (7) John grows rosebushes.

Por meio desses breves testes, pode-se apurar que, apesar de os verbos crescer e cultivar estarem aparentemente mais relacionados semanticamente, pois pertencem ao mesmo subgrupo, o verbo cultivar, em português, comporta-se sintaticamente de maneira semelhante ao verbo construir, de outro subgrupo.

Com a realização desse estudo, as autoras concluíram que a análise cuidadosa das propriedades léxico-semânticas e lógico-conceptuais dos verbos revela suas diferenças de sentido, possibilitando a identificação de várias subclasses semânticas dentro das

“classes” já existentes de verbos, e também da inclusão de verbos em classes diferentes. Aproveitando a metodologia proposta pelas autoras, pretendo chegar à identificação das classes referentes aos verbos de criação/transformação do português do Brasil.

3. Metodologia / Forma de Análise dos Resultados

Para a verificação das características sintáticas e checagem de uso atual das unidades, fase do projeto ainda não atingida, pode-se usar o recurso da Linguística de *Corpus*. Neste trabalho, pretendo fazer testes sintáticos com os verbos selecionados usando os programas de concordância. A pesquisa com *corpus* apresenta também estatísticas, o que pode contribuir na verificação dos usos mais comuns para cada um dos verbos em português, e aprendizado automático, que permite a checagem da eficácia das regras que serão estabelecidas na aplicação a todos os verbos de cada subclasse. O aplicativo usado será *WordSmith Tools* e serão usados, inicialmente, o *corpus* desenvolvido no Laboratório de Lexicografia da UNESP de Araraquara e o *corpus* disponibilizado pelo *Núcleo Interinstitucional de Linguística Computacional - NILC* (<http://nilc.icmc.sc.usp.br>). Ao fim das análises, creio que será possível chegar a uma proposta de classificação léxico-semântica e lógico-conceptual para os verbos de criação/transformação do português do Brasil.

Na fase lingüístico-computacional, os resultados obtidos na fase anterior serão projetados em sistemas formais computacionalmente tratáveis. Uma vez que esta será a última fase de minha pesquisa, esboço aqui uma forma de análise dos resultados baseada na proposta de Briscoe (1991). Como ainda não há dados a serem analisados, o objetivo desta seção é apresentar uma possibilidade de tratamento dos dados. Comentários e sugestões são muito bem vindos.

Briscoe afirma que teorias gramaticais devem ser expressas em linguagem formal com sintaxe e semântica apropriadas. Segundo o autor, muitas teorias gramaticais de base lexical consideram as categorias sintáticas como estruturas de traços (FS - *feature structures*) e adotam a **unificação** como modo de combinação da informações nas FSs. A unificação é uma forma de combinação de padrões bi-direcional, muito usada na prova de teoremas e na programação lógica, que foi introduzida nas "léxico-gramáticas" graças a trabalhos em PLN. O autor refere-se à linguagem pela qual o léxico é expresso como linguagem de representação lexical (LRL). Para ele, a LRL deve permitir que os aspectos de FSs comuns a uma mesma categoria, como os verbos transitivos, por exemplo, sejam expressos uma só vez, e não repetidos em cada entrada lexical individual. O autor defende o uso de estruturas lexicais para nomear e definir sub-partes de FSs comuns a classes de unidades lexicais e para abreviar as próprias entradas, sendo que os modelos podem ser expandidos nas FS, se necessário. Essa abordagem possibilita uma compactação das entradas lexicais e permite a expressão de certas generalizações, em particular porque os modelos podem se inserir em outros modelos de definição. Entretanto, esses modelos são recursos de abreviação, por isso não apresentam nenhuma organização específica do léxico ou caracterização mais aprofundada do conteúdo das FSs. Por essa razão, Briscoe afirma que muitos pesquisadores propõem um aumento nas funções lógicas das FSs, acrescentando operações como negação, disjunção, implicação condicional, etc. A livre adição dessas operações, contudo, pode complicar consideravelmente o tratamento computacional dos dados ou a interpretação semântica realizada formalmente, ou mesmo ambos, por isso o autor questiona até que ponto tais extensões são desejáveis lingüisticamente e

formalmente. Em um contexto mais amplo, não está claro que uma restrição aos formalismos desenvolvidos na unificação seja sustentável ou desejável. Assim, o autor afirma que os processos inferenciais envolvidos na compreensão de um língua ultrapassam os mecanismos limitados das formalizações unificadas; entretanto, não está claro se as operações lexicais em si exigem tais processos. Essas afirmações de Briscoe serão levadas em conta na execução da parte formal da pesquisa. As conclusões obtidas na parte lingüística serão tratadas de modo a permitirem uma formalização unificada.

4. Considerações Finais

Uma vez que a pesquisa está ainda em andamento, apresentei neste momento uma análise inicial, referente à parte lingüística de meu projeto, e não os resultados ou conclusões sobre o tema acima explicitado. O objetivo deste trabalho é apresentar os questionamentos que me levaram a propor uma pesquisa de Doutorado, ainda em andamento, mostrar algumas análises iniciais e discutir possibilidades para se chegar a um resultado final, a fim de possibilitar um debate e troca de experiências que possam vir a melhorar o trabalho.

5. Bibliografia

- BASILIO, M. *Teoria Lexical*. São Paulo: Ática, 1987.
- BOUGURAEV, B; BRISCOE, T. (ed.) *Computational Lexicography for natural language Processing*. London, New York: Longman, 1988.
- BRISCOE, T. Lexical issues in natural language processing. In: E. Klein; F. Veltman (ed.) *Natural language and speech*. Berlin: Springer-Verlag., 1991. p.1-22.
- CRUSE, D. A. *Lexical Semantics*. Cambridge: CUP, 1995.
- DIAS-DA-SILVA, B. C. A construção de um *thesaurus* eletrônico para o português do Brasil. V *PROPOR-Encontro para o Processamento Computacional da língua portuguesa escrita e falada*. Anais... São Carlos: ICMC, p. 1-10, 2000.
- _____. *A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais*. Araraquara, 1996. 272 p. Tese /Doutorado em Letras) – Faculdade de Ciências e Letras, UNESP.
- JACKENDOFF, R. S. *Semantic Structures*. Cambridge: MIT Press, 1990.
- KILGARRIFF, A. Business Models for dictionaries and NLP. *Papers from ITRI – University of Brighton*, 2000.
- LEVIN, B.; RAPPAPORT-HOVAV, R. M. Wiping the slate clean: a lexical semantic exploration. In: LEVIN, B.; PINKER, S. *Cognition. Special Issue: Lexical and Conceptual Semantics*. V. 42, n. 1-3, p. 123-151, 1991.
- LEVIN, B. *English Verb Classes and Alternations: a preliminary investigation*. Chicago: University of Chicago Press, 1993.
- PUSTEJOVSKY, J. *The Generative Lexicon*. 2ª ed. Cambridge: the MIT Press, 1996.
- THE EUROWORDNET Base Concepts and Top Ontology. Version 2, final, 1998. Deliverable D017, D034, D036, WP5, EuroWordNet LE2-4003
- SHIEBER, S. M. *An introduction to Unification-Based Approaches to Grammar*. Stanford: CSLI Pub, 1996.
- TOSQUI, P. *Advérbios Modalizadores: Subsídios para dicionários bilíngües*. 2002.144 p. Dissertação (Ms em Letras). FCL, UNESP, Araraquara.