

O projeto ONTOGOV-MT: contribuições lingüístico-tecnológicas para a implementação de *e-gov services*

Albano Dalla Pria e Gislaine Aparecida de Carvalho
UNEMAT-Universidade do Estado de Mato Grosso, Brasil

Abstract

The development of NLP systems has created a great demand for formally represented and linguistically motivated lexical knowledge bases. Proposing systematization of semantic and polyssemic relations of “government” domain and formalization of this knowledge in terms of a lexical-semantic data base which lends itself straightforwardly to NLP programming is the aim of the ONTOGOV-MT project presented in this paper to bridges part of the gap.

Keywords: Ontology, NLP, Semantic Relations, Polissemy, Government

Palavras-chave: Ontologia, PLN, Relações Semânticas, Polissemia, Governo

1. Introdução

Atualmente, com a massa de informação em meio digital disponível na *World Wide Web* (WWW), há grande interesse na geração de pesquisas que visam à implementação de sistemas (programas) de processamento da linguagem humana, tais como sistemas de tradução automática, *parsings*, *WordNets* e ontologias.

No intuito de contribuir com essas pesquisas, o projeto ONTOGOV-MT, que se insere no âmbito de pesquisas interdisciplinares de PLN, propõe-se chegar a uma representação lingüístico-computacional das relações léxico-semânticas e lógico-conceituais do domínio “governo”, ou seja, propõe-se a construção de uma base relacional de conhecimento lingüisticamente sistematizado e formalmente representado para implementação de sistemas computacionais de processamento do português brasileiro (PB).

No Brasil, o projeto ONTOGOV-MT¹, ainda em sua fase inicial, é pioneiro no desenvolvimento de pesquisas que visam à implementação de ontologias aos serviços governamentais disponíveis na WWW (ou *e-Gov Services*) e tem como objetivo

¹ Essa fase da pesquisa é financiada pela Fundação de Amparo à Pesquisa do Estado de Mato Grosso (FAPEMAT) através do projeto “Análise das relações semânticas e da polissemia sistemática dos nominais para a construção da ONTOGOV-MT”. O projeto conta com a colaboração dos pesquisadores Albano Dalla Pria, Gislaine Aparecida de Carvalho e Edileusa Gimenes Moralis, da Universidade do Estado de Mato Grosso, Mato Grosso, Brasil.

contribuir para o aperfeiçoamento de tais serviços para a superação das inconsistências, muitas vezes identificadas, entre as políticas governamentais e a administração pública, que comprometem o acesso democrático ao usuário através de e-portais.

Embora um projeto desse porte requeira trabalho alentado e seja ambicioso nos seus propósitos, projeto semelhante já foi desenvolvido para o domínio “universidade”² no intuito de contribuir para a implementação do sistema de busca da UNISINOS-Universidade do Vale do Rio dos Sinos, Brasil.

2. A inserção do projeto ONTOGOV-MT no universo da pesquisa lingüístico-tecnológica

2.1. O universo da pesquisa lingüístico-tecnológica

Na década de 50, com o surgimento dos primeiros computadores, com o reconhecimento do seu potencial para o processamento da linguagem natural, iniciaram-se os primeiros trabalhos de pesquisa, cujo objetivo sempre foi entender processos lingüísticos com vistas a dotar os sistemas computacionais com a habilidade de gerar e interpretar fragmentos de línguas naturais, de modo que pudessem realizar tarefas como traduzir e facilitar o acesso a imensas bases computacionais de armazenamento de dados (Grishman, 1986).

Ao perceber que não bastava aplicar ao processamento automático de línguas naturais teorias de modo isolado, lingüistas e projetistas logo integraram seus conhecimentos específicos em um campo interdisciplinar envolvendo Lingüística e Processamento Automático de Línguas Naturais (PLN). O PLN é uma subárea da Inteligência Artificial (IA) e visa à projeção e à implementação de sistemas computacionais para simulação do conhecimento lingüístico humano nesses sistemas (Menuzzi & Othero, 2005)³. Cabe à Lingüística, nesse contexto, a sistematização e a formalização do conhecimento lingüístico necessário para a implementação dos sistemas (Pria, 2005).

Em decorrência do trabalho colaborativo na área do PLN, surgiram, no âmbito lingüístico, teorias lingüísticas computacionalmente motivadas e, do mesmo modo, no âmbito computacional, percebeu-se que sistemas lingüisticamente motivados produzem melhores resultados do que programas desprovidos desse conhecimento.

Desde as primeiras tentativas de simulação do processamento das línguas naturais em sistemas computacionais, um dos desafios para o PLN é a especificação explícita e formal de informações semânticas. Esse fato se deve a dois princípios básicos das línguas naturais abordados tanto no campo lingüístico quanto no PLN: a ambigüidade e a polissemia (Leacock & Ravin, 1999). No campo lingüístico, a semântica lexical tenta

² O projeto APSCO-Análise da Polissemia Sistemática para a Construção de uma Ontologia, sediado na UNISINOS e coordenado pela professora Rove Luiza de O. Chishman, resultou na construção de representações semânticas para os itens lexicais constantes em ementas de disciplinas e no próprio site da universidade, resultando em uma ontologia para fins computacionais (Chishman & Alves, 2002).

³ Sistemas de PLN são programas computacionais “inteligentes” capazes de simular uma interação verbal através de uma língua natural (Winograd, 1992).

explicar a ambigüidade através da especificação das relações semânticas estabelecidas entre itens lexicais (Cruse, 1986). No campo do PLN, essas relações são formalmente representadas e implementadas em sistemas computacionais.

2.2. O papel das ontologias nos avanços lingüístico-tecnológicos

Dentre os fatores que, desde 1950, motivam a produção de conhecimento lingüisticamente sistematizado e formalmente representado para a implementação de sistemas de Processamento Automático de Línguas Naturais (PLN) está o aprimoramento da interação entre usuários e sistemas computacionais (disponibilizados, por exemplo, na WWW) por meio das línguas naturais (Pylyshyn *et al.*, 1980).

A WWW, concebida em 1991 como um serviço de gerenciamento de informação publicada na Internet (Berners-Lee, 2000; Berners-Lee, Hendler & Lassila, 2001) dispõe atualmente de um volume de informações que aumenta vertiginosamente. Estima-se que atualmente estejam postados na rede mais de um bilhão de documentos (Chaves, Vieira & Rigo, 2001). Se, por um lado, o elevado volume de informações publicadas na WWW representa o sucesso da proposta, por outro lado, dificulta o seu gerenciamento manual e compromete o acesso a tais informações. As dificuldades na busca de informações motivaram a reengenharia da concepção inicial da WWW, com o intuito de facilitar, otimizar e agilizar o acesso ao repositório de informações nela disponibilizadas.

O projeto *Semantic Web* (SW) (Berners-Lee, Hendler & Lassila, 2001), que propõe a reengenharia da WWW, prevê dentre os seus objetivos, a implementação de ontologias, i.e., bases de conhecimento léxico-semântico e lógico-conceitual (Evens, 1988) formalmente representado (Gruber, 1993) sobre situações variadas do mundo real, ao projeto inicial da WWW. Com efeito, espera-se aperfeiçoar a interação entre usuários e Sistemas de Perguntas e Respostas, serviços da Web Semântica e sistemas de busca e, ao mesmo tempo, facilitar o acesso à massa de informações em meio digital.

Segundo Zavaglia *et al* (2005: 2),

Na Web, o uso de ontologias pode fornecer uma base de informações comum, bem como padronizada, englobando conceitos-chave que possam ser utilizados por serviços requisitados para cada situação particular. Em comércio eletrônico, por exemplo, o conjunto de informações oferecido pela ontologia pode ser utilizado para unificar e integrar definições de produtos oferecidos pelos mais variados pontos de venda, com um formato padrão e único. Além disso, as ontologias podem ser utilizadas em sistemas de recuperação da informação para melhorar a precisão e revocação dos documentos recuperados numa busca. Precisão é dada pela razão entre o número de respostas (documentos) corretas obtidas e o número total de respostas recuperadas. Revocação é dada pela razão entre o número de respostas corretas obtidas e o número total de respostas corretas possíveis.

A construção de ontologias, entretanto, desafia lingüistas e cientistas da computação, porque não só requer pesquisa interdisciplinar e trabalho colaborativo, mas também demanda trabalho alentado (porém factível).

2.2.1. A ontologia no domínio lingüístico-tecnológico

No campo lingüístico, uma ontologia é uma base relacional de conhecimento que estabelece aproximações entre o conhecimento léxico-semântico e o lógico-conceitual do universo extra-lingüístico (Evens, 1988). No campo do PLN, “uma ontologia é uma especificação explícita e formal de uma conceitualização compartilhada” (Gruber, 1993). Uma ontologia é uma conceitualização porque se refere a um modelo abstrato do mundo que desejamos representar e é construída através da identificação de conceitos e de relações relevantes entre conceitos. Uma ontologia tem caráter explícito porque os tipos de conceitos e as restrições a eles impostas devem ser claramente explicitados. Uma ontologia deve ser formal porque deve ser processada em sistemas computacionais (Fensel, 2000).

O projeto ONTOGOV-MT, entretanto, propõe-se à construção de uma ontologia cuja dimensão não constitui um modelo abstrato do mundo como um todo, mas apenas por uma parte do mundo, i.e., o domínio “governo”.

O projeto ONTOGOV-MT tem como inspiração um projeto semelhante em desenvolvimento na Europa desde 2004: o projeto ONTOGOV⁴. No Brasil, entretanto, o projeto ONTOGOV-MT é pioneiro na investigação de uma base de conhecimento léxico-semântico para enriquecimento de *e-Gov Services*. Como tal, adota uma metodologia de pesquisa já praticada pela sua equipe e que se diferencia daquela adotada pelo projeto ONTOGOV.

3. A metodologia a ser empregada pelo projeto ONTOGOV-MT

Utilizar-se-á uma metodologia específica para o campo do PLN dada sua natureza lingüístico-tecnológica (Dias-da-Silva, 1996). Essa metodologia equaciona os problemas em três fases: **lingüístico**, **lingüístico-computacional** e **computacional**. Cada fase envolve um conjunto próprio de atividades e competências.

Embora a **fase lingüística** seja responsável pela explicitação ou sistematização do conhecimento lingüístico, reuindo investigações sobre um conjunto complexo de fatores de natureza fonético-fonológica, morfológica, sintática, semântica, pragmática ou discursiva, as atividades do projeto ONTOGOV-MT, na fase lingüística, restringem-se especificamente (i) à investigação do conhecimento lexical no processamento da linguagem e (ii) à investigação das relações semânticas e da polissemia sistemática dos nominais do domínio governo.

As investigações do conhecimento lexical no processamento da linguagem passaram por dois momentos. No primeiro, ainda nos primórdios da Lingüística Estrutural, o léxico era conceituado como um repositório de fatos idiossincráticos e imprevisíveis sobre os itens lexicais (Bloomfield, 1933) e não se atribuía nenhum tipo de organização ou estruturação a esse conjunto; em um segundo momento, quando se lançaram as bases da Hipótese Lexicalista, na década de 70, o léxico passa a ser o componente essencial de uma gramática (Briscoe, 1991). A passagem do primeiro para

⁴ Disponível em: <<http://www.ontogov.com/>>.

o segundo momento deve-se à influência exercida pelas Ciências Cognitivas (na década de 50) sobre a Lingüística. As Ciências Cognitivas consideravam o cérebro um órgão modular e essa idéia de modularidade estendeu-se aos estudos da linguagem⁵. Passou-se a acreditar que, no processamento mental da linguagem, existem módulos específicos para a manipulação das diferentes instâncias lingüísticas (Chomsky, 1965; 1981).

Embora reconhecesse o léxico ainda como um repositório de idiosincrasias, a Gramática Gerativa (Chomsky, 1965) identificou-o como uma dessas instâncias modulares, não-autônoma, integrada ao Componente de Base da gramática, demonstrando interesse em conferir algum tipo de organização ao léxico ao analisar categorias lexicais como complexos estruturados de traços categoriais (referentes à classe da palavra), traços de subcategorização (referentes às outras classes com a qual uma determinada classe co-ocorre) e traços seletivos (referentes aos tipos semânticos dessas outras classes). Como resultado dessas investigações, foram identificados, posteriormente (Chomsky, 1970), relacionamentos morfológicos, semânticos e sintáticos entre certos nomes (os deverbais) e os verbos dos quais eram derivados. Com isso, lançaram-se as bases da Hipótese Lexicalista (Chomsky, 1970) que institui um componente lexical dentro do qual são expressas generalizações lexicais (propriedades comuns a vários itens) através de regras de redundância, que servem para relacionar entradas lexicais e atribuir algum tipo de estruturação ao léxico (Briscoe, 1991) e a partir do qual os itens lexicais são gerados.

O estatuto de componente essencial da gramática atribuído ao léxico motivou o surgimento de arcabouços teóricos que codificam, no léxico, explicações para a maior parte dos fenômenos lingüísticos. Nesse contexto, surgiram, por exemplo a *Generalized Phrase Structure Grammar* (Gazdar *et al*, 1985), a *Lexical Functional Grammar* (Bresnan, 1982) e a *Head-driven Phrase Structure Grammar* (Pollard & Sag, 1994) e a *Generative Lexicon* (Pustejovsky, 1995). Nesses modelos, o léxico contém informações sobre o comportamento fonético-fonológico, morfológico, sintático e semântico das palavras.

Embora a **fase lingüístico-computacional** seja responsável pela representação formal (no sentido de precisa, explícita) e computacionalmente tratável do conhecimento lingüístico sistematizado no campo lingüístico, necessária a seleção de formalismos gramaticais capazes de representar de modo expressivo as informações quanto a um conjunto complexo de fatores, que articulem conhecimento lexical e conhecimento gramatical, as atividades do projeto ONTOGOV-MT, na fase lingüístico-computacional, restringem-se especificamente à investigação da Teoria do Léxico Gerativo como um formalismo que fornece subsídios à representação de um conjunto de propriedades léxico-semânticas e lógico-conceituais através da Estrutura Qualia.

A **fase computacional** é responsável pela implementação computacional do conhecimento lingüístico sistematizado e formalmente representado nas fases anteriores. Nesse campo, são definidos os módulos componentes dos sistemas, suas funções, sua disposição, sua organização, como a informação deve fluir entre os módulos e qual será a interface do sistema. As atividades do projeto ONTOGOV-MT não contemplarão essa fase da metodologia.

⁵ Fodor (1983) apresenta informações detalhadas a respeito da teoria da modularidade.

3.1. A investigação das relações semânticas e da polissemia sistemática dos nominais do domínio *governo*

3.1.1. As relações semânticas para o domínio *governo*

Um conjunto de relações semânticas pode ser identificado entre itens lexicais. Citem-se, por exemplo, as relações de hiponímia/hiperonímia, meronímia/holonímia, sinonímia, antonímia, homonímia, agentividade e polissemia (Cruse, 1986; 2000). As relações de hiponímia/hiperonímia e meronímia/holonímia têm caráter determinante para a construção de bases de conhecimento léxico-semântico e lógico conceitual como ontologias e *wordnets* (Fellbaum, 1998). A relação de hiponímia/hiperonímia, também conhecida por *é-um (tipo de)*, descreve uma relação de inclusão de significado de um termo (ou classe) mais específico no significado de um termo (ou classe) mais abrangente. Os hipônimos são termos mais específicos que dispõem de algumas, mas não todas as propriedades de seus hiperônimos. Sirvam-se como exemplos: (a) *A Casa Civil é um órgão estadual*, (b) *Blairo Maggi é um governador* e (c) *Secretário de estado de educação é um tipo de secretário*. *Casa Civil*, *Blairo Maggi* e *secretário de estado de educação* são hipônimos de *órgão estadual*, *governador* e *secretário*, respectivamente, e esses são hiperônimos daqueles.

A identificação dessas relações possibilita a especificação de inferências relevantes, por exemplo, para sistemas de busca de informações na WWW. Considerem-se os exemplos: (a) *A FAPEMAT é uma fundação*, (b) *As fundações são um tipo de órgão estadual*. Com base nas relações de hiponímia/hiperonímia estabelecidas em (a) e (b), pode-se derivar, por inferência, que *A FAPEMAT é um órgão estadual*.

A relação de meronímia/holonímia, conhecida como relação parte-todo, também tem se mostrado bastante produtiva nas línguas naturais. Os merônimos são termos cujo significado descreve partes constituintes do significado de outro termo, seu holônimo. Vejam-se os exemplos: (a) *Blairo Maggi é membro do governo*, (b) *O Secretário de estado de educação é membro da Secretaria de Educação* e (c) *Os órgãos estaduais são parte do governo*. *Blairo Maggi*, *Secretário de estado de educação* e *órgãos estaduais* são merônimos de *governo*, *Secretaria de Educação* e *órgãos estaduais*, respectivamente, e esses são holônimos daqueles.

As relações de meronímia/holonímia também possibilitam a especificação de inferências relevantes. Citam-se os exemplos: (a) *Secretário de estado de educação é membro da Secretaria de Educação*, (b) *Secretaria de Educação é parte dos órgãos estaduais* e (c) *Os órgãos estaduais são parte do governo*. Com base nas relações de meronímia/holonímia estabelecidas em (a), (b) e (c), pode-se derivar, por inferência, que *governo* tem uma *Secretaria de Educação* e um *Secretário de estado de educação*.

Uma das relações semânticas mais complexas, no estudo do léxico, é a polissemia. O desafio de explicar esse fenômeno nas línguas naturais tem motivado pesquisas que vão além da análise do comportamento de palavras ambíguas cujos significados distintos e não-relacionados são tradicionalmente denominados homônimos. Essa abordagem relacional capta, por exemplo, dois sentidos distintos e não-relacionados para o item *banco*, na frase *Hoje eu passei pelo banco*. Esses sentidos referem-se à *instituição financeira* ou *assento*. No entanto, verificou-se que *banco* pode apresentar sentidos relacionados recuperados no contexto em que ocorrem. Nas frases (a) *Este*

banco foi fundado em 1950, (b) Este *banco* está em greve e (c) O *banco* no qual deposito meu dinheiro fica ao lado do supermercado, *banco* refere-se a instituição, em (a), *peçoas*, em (b) e *instituição e edifício*, em (c). Os sentidos de *banco*, em (a), (b) e (c), estão todos relacionados a *instituição financeira*. A essa propriedade dos itens lexicais de apresentar significados relacionados, dependentes do contexto de ocorrência, Pustejovsky (1995) chamou polissemia complementar (ou lógica), em oposição à polissemia contrastiva (ou homonímia) que apenas distingue *banco*₁ (= *instituição financeira*) de *banco*₂ (= *assento*). Aplicada ao domínio governo, a polissemia complementar descreve significados distintos, porém relacionados, nas frases (a) *O governo republicano foi fundado em 15 de novembro de 1889*, (b) *O governo está em greve*, (c) *O governo teve início em 1º de janeiro*, (d) *O governo tem duração de 4 anos* e (e) *O governo está em obras*. Os exemplos de (a) a (e), descrevem aspectos do significado de *governo*: em (a) denota uma instituição, em (b) pessoas, em (c) o processo de governar, em (d) o evento de governar e em (e) uma edificação.

Pustejovsky (1995) explica que cada aspecto do significado de um item lexical refere-se a um “papel” relacionado ao significado desse item. Esses papéis descrevem os atributos essenciais de um objeto, eventos e relações que definem o item lexical (Abrahão, 1997). São quatro os papéis especificados pela teoria: Constituinte (CONST), Formal (FORM), Téliico (TELIC) e Agentivo (AGENT). CONST relaciona um objeto e suas partes constituintes (material, peso, partes e elementos componenciais); FORM distingue os objetos dentro de um domínio maior (orientação, magnitude, forma, dimensionalidade, cor, posição), TELIC especifica objetivo e função (objetivo de um agente na performance da ação, que especifica atividades) e AGENT especifica fatores envolvidos na origem ou construção de um objeto (criador, artefato, cadeia causal).

3.1.2. A formalização das relações semânticas para o domínio *governo*

A Teoria do Léxico Gerativo é uma teoria semântica que descreve não só propriedades do significado dos itens lexicais, mas também fornece subsídios à representação formal para esse conjunto de propriedades (i.e., precisa e explicitamente) através da Estrutura Qualia.

Considerem-se, por exemplo, os nominais *governador* e *Blairo Maggi*. A Estrutura Qualia do primeiro representaria que se trata de uma entidade que tem as características de uma pessoa (Papel Formal) e cuja função é administrar (Papel Téliico) alguma coisa (x, y) em um determinado evento (e). A Estrutura Qualia do segundo representaria que se trata de uma entidade que tem por característica ser governador (Papel Formal), cuja função é administrar (Papel Téliico) o Estado de Mato Grosso.

QUALIA=[governador(x) FORMAL: x:pessoa TÉLICO: administrar (e, x, y)]

Quadro 1: Estrutura Qualia para o nominal *governador*

QUALIA=[Blairo Maggi(x) FORMAL: x:governador TÉLICO: administrar (e, x:Blairo Maggi, y:Estado de Mato Grosso)]
--

Quadro 2: Estrutura Qualia para o nominal *Blairo Maggi*.

A Estrutura Qualia do nominal *Blairo Maggi* não só prevê que esse nominal é hipônimo do nominal *governador* (*Blairo Maggi é um governador*), mas também é capaz de esclarecer o uso criativo do nominal *Blairo Maggi* em novos contextos, como, por exemplo, em *Blairo Maggi administra o Estado de Mato Grosso*. Ou seja, a Estrutura Qualia especifica que a mesma pessoa é governador e administra o Estado de Mato Grosso.

3.2. O corpus, a extração dos termos e a edição da ontologia

Além de assumir pressupostos, o projeto ONTOGOV-MT empregará em suas estratégias de ação métodos semi-automático, a fim de otimizar o desenvolvimento das atividades tanto da fase lingüística quanto da fase lingüístico-computacional. Ou seja, parte das atividades necessárias para que se atinjam seus objetivos será realizada por humanos e parte será realizada automaticamente por ferramentas computacionais.

Após a montagem do *corpus* será necessária a definição do método mais eficiente para a extração de termos candidatos a compor a ontologia pretendida. Há várias metodologias empregadas nessa atividade, mas, em geral, emprega-se uma metodologia híbrida manual e automática, empregando-se alguma ferramenta computacional baseada em estatística. Embora a frequência não seja determinante para a inclusão de um termo na ontologia, pode ser um indício de que um termo é relevante no domínio em questão e que, portanto, pode ser um candidato a integrar essa base de conhecimento.

O aplicativo Protégé⁶ (Gennari *et al*, 2003) deverá auxiliar na representação formal das relações semânticas dos termos do domínio governo. Esse editor de ontologias, desenvolvido na Universidade de Stanford, representa o conhecimento semântico hierarquicamente em termos de classes, atributos e relações e dispõe de recursos que convertem sua base de conhecimento em formatos que permitem sua implementação em diversas aplicações tecnológicas, como os já mencionados Sistemas de Perguntas e Respostas, serviços da Web Semântica, Pesquisa Semântica Multimídia e sistemas de busca de informações.

4. Considerações finais

A execução do projeto ONTOGOV-MT deverá produzir resultados cujos impactos incidam nos âmbitos social, científico e tecnológico.

⁶ Esse aplicativo foi desenvolvido por pesquisadores do Centro de Pesquisa de Informática Biomédica da Universidade de Stanford. Protégé é um aplicativo gratuito e sua versão para *download* encontra-se disponível em: <<http://protege.stanford.edu/>>.

No âmbito social, o projeto contribuirá (i) para a democratização do acesso aos serviços públicos através de e-Government Services; (ii) para o fortalecimento do trabalho interdisciplinar envolvendo a Lingüística e o PLN, no Estado de Mato Grosso.

No âmbito científico, o projeto contribuirá (i) com a investigação de léxicos semânticos e ontologias, (ii) com a investigação de relações léxico-semânticas e lógico-conceituais, (iii) com o aprimoramento de metodologias de compilação, extração e construção de ontologias; (iv) com o aprimoramento do trabalho interdisciplinar envolvendo Lingüística e PLN.

No âmbito tecnológico, a ontologia poderá subsidiar (i) a elaboração de dicionários e *thesaurus* eletrônicos em língua portuguesa e (ii) o desenvolvimento de aplicativos (programas) computacionais. A redução, ainda que parcial, no custo da construção da base de conhecimento lexical contribui para a redução no custo total da implementação de conhecimento lingüístico em sistemas de PLN (Dias-da-Silva, Oliveira & Moraes, 2003).

Referências

- Abrahão, P. R. C. (1997) *Modelagem e implementação de um léxico semântico para o português*. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio Grande do Sul.
- Berners-Lee, T. (2000) *Weaving The Web: The original design and ultimate destiny of the World Wide Web by Its Inventor*. New York: HaperCollins.
- Berners-Lee, T., J. T. Hendler & O. Lassila (2001) *The semantic web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*. New York: Scientific American. Disponível em: <<http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>>. Acesso em 25 de set. de 2005.
- Bloomfield, L. (1933) *Language*. Nova York: Holt, Rinehart e Winston.
- Bresnan, J. (Ed.). (1982) *The mental representation of grammatical relations*. Cambridge: The MIT Press.
- Briscoe, T. (1991) Lexical issues in natural language processing. In E. Klein & F. VELTMAN (eds.) *Natural language and speech*. Berlin: Springer-Verlag, p.39-68.
- Chaves, M. S., R. Vieira & S. Rigo (2001) Uso de ontologias para gerenciamento e acesso a documentos na web. In *V Oficina de Inteligência Artificial, 2001*, Pelotas. *Anais da V Oficina de Inteligência Artificial*. Pelotas: EDUCAT.
- Chishman, R. L. O. & I. M. da R. Alves (2002) Ontologias e Relações Semânticas: Uma Aplicação. In XIII Simpósio Brasileiro de Informática na Educação - Workshop de Ontologias. 2002. São Leopoldo. *Anais do XIII Simpósio Brasileiro de Informática na Educação*. São Leopoldo: Editora UNISINOS, pp. 583-585.
- Chomsky, N. (1965) *Aspects of the theory of syntax*. Cambridge: The MIT Press.
- Chomsky, N. (1970) Remarks on nominalizations. In R. A. Javobs & P. S. ROSENBAUM (eds.) *Readings in English transformational grammar*. Waltham: Ginn and Company.
- Chomsky, N. (1981) *Lectures on government and binding*. Dordrecht: Foris.

- Cruse, A. (1986) *Lexical semantics*. Cambridge: Cambridge University Press.
- Cruse, D. A. (2000) *Meaning in language: an introduction to semantics and pragmatics*. New York: Oxford University Press.
- Dias-da-Silva, B. C. (1996) *A face tecnológica dos estudos da linguagem: o processamento automático das línguas naturais*. Tese (Doutorado em Lingüística e Língua Portuguesa) – Faculdade de Ciências e Letras de Araraquara, Unesp, Araraquara.
- Dias-da-Silva, B. C., M. F. Oliveira & H. R. Moraes (2003) Reusability of Dictionaries in the Compilation of NLP Lexicons. In N. J. Mamede, J. Baptista, I. Trancoso & M. G. V. Nunes (eds.) *Computational processing of the portuguese language*. Berlin: Springer-Verlag, pp. 78-85.
- Evens, M. (1988) *Relational models of the lexicon*. Cambridge: Cambridge University Press.
- Fellbaum, C. (1998) *WordNet: an electronic lexical database*. Cambridge: The MIT Press.
- Fensel, D. (2000) *Ontologies: silver bullet for knowledge management and electronic commerce*. Berlin: Springer-Verlag. Disponível em: <<http://www.cs.vu.nl/~dieter>>.
- Fodor, J. (1983) *The modularity of mind*. Cambridge: The MIT Press.
- Gazdar, G. et al. (1985) *Generalized phrase structure grammar*. Cambridge: Harvard University Press.
- Gennari, J. H. et al. (2003) The Evolution of Protégé: An Environment for Knowledge-Based Systems Development. *International Journal of Human-Computer Studies*, 58 (1), pp. 89-123.
- Grishman, R. & N. Calzolari (1995) Lexicons. In R. A. Cole (ed.). *Survey of the State of the Art in Human Language Technology*. Cambridge: Cambridge University Press.
- Gruber, T. R. (1993) A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition* 5 (2), pp. 199-220.
- Leacock, C. & Y. Ravin (eds.) (2000) *Polysemy: Theoretical and Computational Approaches*. Oxford: Oxford University Press.
- Menuzzi, S. de M. & G. de Av. Othero (2005) *Lingüística computacional: teoria e prática*. São Paulo: Parábola.
- Pollard, C. & I. Sag (1994) *Head-driven phrase structure grammar*. Stanford: CSLI/The University of Chicago Press.
- Pria, A. D. (2005) *Uma proposta de representação lingüístico-computacional do comportamento sintático e semântico de adjetivos no sintagma nominal do inglês e do português*. Dissertação (Mestrado em Lingüística e Língua Portuguesa) – Faculdade de Ciências e Letras da Universidade Estadual Paulista, Araraquara.
- Pustejovsky, J. (1995) *The generative lexicon*. Cambridge: The MIT Press.
- Pylyshyn, Z. W. et al. (1980) Understanding natural language. *IBM Journal of research and development* 35, pp. 1-20.
- Winograd, T. (1972) *Understanding natural language*. New York: Academic Press.
- Zavaglia, C. et al. (2005) *Avaliação de Métodos de Extração Automática de Termos para a Construção de Ontologias*. Relatório Técnico do NILC. São Carlos.