

Anderson Bertoldi
andersonbertoldi@yahoo.com
Rove Luiza de Oliveira Chishman
rove@unisinos.br
Isa Mara da Rosa Alves
isamralves@gmail.com

A semântica adjetival e sua representação em uma ontologia de domínio jurídico^{1,2} Adjectival semantics and its representation in a legal ontology

RESUMO - Este trabalho faz parte de uma série de estudos que temos feito sobre a semântica dos adjetivos do domínio jurídico, visando a sua formalização e inclusão em uma ontologia que possa contribuir para o aperfeiçoamento de sistemas de busca e extração de informações jurídicas na web. Este artigo discute as dificuldades para a inclusão dos adjetivos em uma ontologia jurídica, bem como a validade do significado adjetival para a recuperação de informações. Para a análise lingüística dos adjetivos, optamos por uma abordagem baseada em relações semânticas. A ontologia proposta tem como objetivo futuro o agrupamento desses adjetivos sob conceitos de top-level e a interligação da ontologia a uma rede de ontologias jurídico-semânticas. Para a presente análise, utilizamos seis textos jurídicos extraídos, via internet, do Instituto das Tecnologias de Informação na Justiça de Portugal.

Palavras-chave: adjetivos, semântica lexical, ontologias, recuperação de informação.

ABSTRACT - This paper is part of some studies we have made about legal domain adjectives. We aim at formalizing and including these adjectives in a legal ontology, checking the possibility of this part-of-speech to contribute to the improvement of the legal information retrieval systems. In this paper we discuss the difficulties in including adjectives in a legal ontology, as well as the validity of the adjectival meaning to the information retrieval. Our approach to the adjectives was based on semantic relations. Our future objective is to group them under top-level concepts and interconnect this ontology to other legal ontologies. To this analyses we use six legal texts collected via web from the Instituto das Tecnologias de Informação na Justiça of Portugal.

Key words: adjectives, lexical semantics, ontologies, information retrieval.

Introdução

Ao contrário dos substantivos e dos verbos, os adjetivos têm despertado pouco interesse por parte dos estudiosos. Quando falamos em formalização do conhecimento em ontologias, os adjetivos se tornam muito menos atrativos. Nominais e eventos são as entidades que, por excelência, se destacam nesse tipo de aplicação. A inclusão da informação veiculada pelos adjetivos é tema controverso.

Enquanto os substantivos designam os seres e os verbos as suas ações, os adjetivos representam as propri-

edades dos seres. Assim sendo, como organizá-los dentro de uma ontologia? Se os adjetivos estão modificando um nome, formando um sintagma nominal, eles devem ter o mesmo *status* dos substantivos e dos verbos, entrando como uma terceira categoria dentro da ontologia? Ou então, por que não incluir na ontologia, juntamente com os eventos e os nomes, as propriedades dos nomes? Além dessas dificuldades de natureza organizacional, há outras de natureza semântica.

Nosso objetivo é avaliar a problemática da formalização da semântica dos adjetivos em uma ontologia jurídica, pois, tradicionalmente, uma ontologia se presta

¹ Este artigo é uma revisão e extensão do artigo "O significado adjetival para a construção de ontologias e redes jurídico-semânticas", apresentado oralmente no VII Encontro do CELSUL (2006).

² Este trabalho foi realizado com o apoio da CAPES.

bem à formalização dos nomes e dos eventos. Neste trabalho apresentamos um estudo preliminar dos adjetivos do domínio jurídico. Essa versão preliminar que apresentamos aqui está sendo desenvolvida de forma mais aprofundada em uma dissertação de mestrado.

A escolha desse tema está diretamente ligada à sua inserção nos projetos SEMANTEC³, DIRPI⁴ e LOIS. Adiantando o que exporemos mais à frente, o objetivo do LOIS é a criação de uma rede uma rede semântica multilíngüe para o Direito. É através dessa relação com o LOIS que objetivamos, para o futuro, uma ligação multilíngüe entre o nosso trabalho e outras ontologias e léxicos semânticos do domínio jurídico. Para desenvolvermos a parte aplicada, utilizamos seis acórdãos extraídos do *site* do Instituto das Tecnologias de Informação na Justiça⁵. Esses mesmos acórdãos foram estudados em Alves (2005), de forma que estamos dando continuidade a esse estudo. O nosso corpus apresenta 66 diferentes adjetivos, totalizando 132 ocorrências. A descrição semântica desses 66 *types* foi baseada em Borba (1996).

Para a realização deste experimento, escolhemos cinco adjetivos, segundo a frequência e a riqueza de informações a formalizar. Nosso propósito é descrever a semântica dos adjetivos que ocorrem nesses acórdãos e formalizá-los no editor de ontologias Protégé, criando uma ontologia de amostra do domínio jurídico. Para o futuro, o compromisso será a ampliação do corpus da pesquisa, uma nova extração dos adjetivos e a criação de uma ontologia com conceitos de *top-level* que abrigue sob si esses adjetivos, levando sempre em consideração a metodologia utilizada por outros léxicos jurídicos, a fim de se poder fazer uma ligação multilíngüe do Direito europeu.

Para atingir nossas metas, apresentamos nesta seção o tema da pesquisa e seus objetivos específicos. Em *Redes jurídico-semânticas*, abordamos o cerne motivador de nossa pesquisa. Nas seções *A semântica dos adjetivos nas wordnets* e *A semântica dos adjetivos nas ontologias*, fazemos uma revisão de alguns léxicos semânticos e ontologias que tratam do adjetivo. Na seção *A descrição semântica dos adjetivos*, tratamos das questões teóricas que embasam nossa análise. Apresentamos os resultados dos dados do corpus em *Os adjetivos do domínio jurídico*. Já em *Formalização dos adjetivos do domínio jurídico no Protégé* tratamos da formalização desses adjetivos no Protégé. Para finalizar, tratamos das considerações finais.

Redes jurídico-semânticas

Trataremos nesta seção de projetos voltados para a construção de léxicos semânticos do domínio do Direito. Para este trabalho escolhemos os seguintes projetos: LOIS e Jur-WN. Esta seção tem por objetivo situar no âmbito internacional o trabalho que estamos realizando. Poderemos, dessa forma, explicitar a relevância social desta pesquisa, além da relevância acadêmica representada pelo tema *adjetivos*.

LOIS: Lexical Ontologies for Legal Information Sharing

O LOIS (Lexical Ontologies for Legal Information Sharing)^{6,7} é um projeto de pesquisa europeu de recuperação de informação multilíngüe de base de dados legais, realizado entre os anos de 2004 e 2005 e coordenado pelo *Instituto de Teoria e Técnicas da Informação Jurídica (ITTIG-CNR)*⁸, de Florença, Itália. O projeto prevê a integração de bases de dados em seis línguas diferentes, a saber: o português, o italiano, o alemão, o neerlandês e o tcheco, cada qual representando o sistema jurídico dos países envolvidos no projeto.

O objetivo do LOIS é desenvolver uma ferramenta de acesso multilíngüe a bases de dados europeias de informação legal. Um instrumento de tal porte permitiria aos usuários, sejam cidadãos comuns, sejam profissionais do Direito, pesquisar leis europeias e documentos legais quebrando-se a barreira lingüística. Seria possível, então, digitar uma pergunta em uma língua e recuperar documentos em seis diferentes línguas.

Para se chegar a tal resultado, o projeto prevê a representação de conceitos legais através de representações formais, utilizando a experiência proporcionada pelo WordNet para a descrição do léxico e a do EuroWordNet para a integração das diferentes bases de dados⁹. Conceitos semelhantes em línguas diversas serão ligados, permitindo acesso aos diferentes sistemas jurídicos que compõem a União Européia, como o Sistema Românico, próprio dos países continentais, e a *Common Law*, tipicamente inglesa. Assim, também será possível a comparação dos sistemas judiciários e das leis nacionais, considerando-se que os termos jurídicos de um sistema nem sempre são compatíveis com os do outro.

As relações utilizadas pelo LOIS são de dois tipos: monolíngües, quando envolvem apenas as relações pró-

³ Projeto coordenado pela Profa. Dr^a. Rove Chishman.

⁴ DIRPI (Desenvolvimento e Integração de Recursos para Pesquisa de Informação), projeto de cooperação internacional entre Brasil e Portugal. Participam do projeto a UNISINOS (Professora Dr^a. Renata Vieira) e Universidade de Évora (Prof. Dr. Paulo Quaresma).

⁵ www.dgsi.pt

⁶ www.loisproject.org

⁷ Um dos resultados do LOIS, um léxico jurídico, já está sendo comercializado.

⁸ www.ittig.cnr.it

⁹ Adiante apresentamos, em breves linhas, ambos os projetos.

prias de uma língua, e multilíngües, quando as relações ligam conceitos de diferentes línguas. Dentre as relações lexicais monolíngües estão a sinonímia, a quase-sinonímia, a antonímia e a hiponímia/hiperonímia. Já entre as relações multilíngües estão a *eq_synonym*, que liga conceitos de diferentes línguas que podem ser livremente substituídos; a *eq_near_synonym*, que liga conceitos que apresentam muita sobreposição de sentidos, mas não podem ser livremente substituídos; e a *eq_has_hyperonym/eq_has_hyponym*, que liga um conceito mais específico em uma língua a um conceito mais geral em outra língua, e vive-versa.

O ponto de partida para o LOIS foi o Jur-WordNet. Na subseção seguinte, discutiremos esse projeto.

Jur-WordNet: uma wordnet para o Direito

O JurWordNet (Jur-WN) é outro projeto desenvolvido pelo Instituto de Teoria e Técnicas da Informação Jurídica. Trata-se de um léxico terminológico do domínio jurídico, organizado segundo relações semânticas, e ligado a um recurso genérico do italiano, o ItalWordNet¹⁰. Seu desenvolvimento é baseado no WordNet de Princeton e no EuroWordNet.

Segundo Sagri *et al.* (2003), o Jur-WN objetiva formar uma base de dados a ser utilizada como:

- uma fonte de metadados para o etiquetamento semântico de textos legislativos;
- um recurso de apoio para sistemas de recuperação de informação, facilitando o acesso a dados multilíngües e heterogêneos;
- uma interface entre a linguagem comum utilizada pelos cidadãos e a linguagem especializada utilizada pelos profissionais e pelos padrões legais;
- uma base de conhecimento conceitual, que possa ser empregada para uma grande variedade de aplicações, tais como extração de informação, sistemas de pergunta e resposta, etiquetamento automático, compartilhamento de conhecimento, comparação de normas etc.

A motivação para construir uma rede semântica de termos jurídicos é a facilitação da busca de informações jurídicas pelos profissionais. Já a sua conexão a um recurso de língua geral facilita também a recuperação de informações por parte de pesquisadores não especialistas, que desconhecem os termos do domínio jurídico. Dessa forma, o Jur-WN torna-se uma ferramenta útil tanto para os profissionais, quanto para o cidadão comum.

O Jur-WN é uma *wordnet* jurídica que, assim como as demais *wordnets* de língua geral, agrupa itens lexicais em um grupo de sinônimos representando um mesmo conceito. Esses agrupamentos de itens lexicais são chamados

de *synsets*. Porém, diferentemente das *wornets* de língua geral, a sinonímia no Jur-WN é limitada, pois, como se trata de uma rede semântica terminológica, os sinônimos de um termo técnico-científico são bastante reduzidos, ou inexistentes. Geralmente, a sinonímia na área terminológica está relacionada com a diferença sócio-dialetal.

Organizando os *synsets*, estão as relações semânticas. No Jur-WN elas podem ser de dois tipos: verticais taxonômicas (hiponímia) e horizontais associativas, tais como meronímia e papel (Gangemi *et al.*, 2005). Todos os *synsets* são organizados segundo uma hierarquia. Conceitos mais elevados e gerais abrigam sob si os termos próprios do Direito. Esses conceitos mais elevados compõem uma ontologia do domínio jurídico, chamada de *Core Legal Ontology*. A *Core Legal Ontology* do Jur-WN fornece um padrão de metadados para descrição de conteúdos (Sagri *et al.*, 2003; Gangemi *et al.*, 2003). Esses metadados são utilizados para o etiquetamento semântico automático de textos jurídicos.

A semântica dos adjetivos nas *wordnets*

Nesta seção, discutiremos as diferentes abordagens adotadas pelos léxicos semânticos para tratar do adjetivo. Para essa apreciação, consideramos três recursos diferentes, a saber: o WordNet, o EuroWordNet e o ItalWorNet. Ao contemplar esses léxicos, explicitamos as bases de construção do LOIS e do Jur-WN, além de levantarmos de antemão alguns rumos para as questões teóricas da semântica dos adjetivos, que servirão de parâmetro para o experimento a ser apresentado em *Formalização dos adjetivos do domínio jurídico no Protégé*. Abaixo, passaremos à descrição dos aspectos que interessam à nossa pesquisa.

O WordNet

O WordNet é uma rede semântica que agrupa os adjetivos em duas grandes classes, os descritivos e os relacionais (Miller, 1999). Os adjetivos descritivos são os adjetivos prototípicos. Eles atribuem ao nome um atributo. A antonímia e a sinonímia são as principais relações semânticas atribuídas a essa classe. Assim, um adjetivo como *rápido* tem como sinônimos *veloz*, *ligeiro* etc. O antônimo de *rápido* pode ser *lento*, que teria, por sua vez, *moroso*, *demorado* etc. como sinônimos. Essa classe também admite gradação.

Os adjetivos relacionais compõem uma ampla classe de adjetivos relacionados semântica e morfológicamente aos nomes. Esses adjetivos se diferem dos descritivos por não atribuírem uma propriedade ao nome com que se relacionam e também não serem graduáveis. Muitos adje-

¹⁰ O ItalWordNet será apresentado adiante, juntamente com o EuroWordNet.

tivos relacionais não se ajustam à separação proposta pelo WordNet em sinônimos e antônimos. Esse é o caso de adjetivos como *cardíaco*, *português* e *hospitalar*, que estão relacionados a nomes e, ao que tudo indica, não possuem antônimos. Esses adjetivos relacionais são mantidos em um arquivo separado, com indicadores para os nomes aos quais eles estão relacionados. Porém, os adjetivos relacionais que possuem sentidos bipolares, como, por exemplo, *físico* e *mental*, em “saúde física” e “saúde mental”, recebem o mesmo tratamento que os adjetivos descritivos, sendo, inclusive, agrupados junto desses.

O EuroWordNet e o ItalWordNet

O EuroWordNet (EWN) é uma base de dados lexicais multilíngüe estruturada segundo os mesmos princípios do WordNet (WN). A base de dados do EWN envolve oito diferentes línguas: em um primeiro momento foram desenvolvidas as *wordnets* para o neerlandês, italiano, espanhol e inglês; em uma segunda etapa, o projeto foi ampliado com a inclusão das *wordnets* para o francês, alemão, estoniano e tcheco (Vossen, 1998). Toda a sua arquitetura foi pensada para permitir a extração multilíngüe de informações, incluindo algumas das novas relações semânticas adicionadas ao EWN, que não existiam no WN.

O EWN classifica apenas verbos e substantivos (Alonge *et al.*, 1998; Alonge *et al.*, 2000; Vossen *et al.*, 1998). Embora não sejam tratados de forma sistemática, os adjetivos são representados em ocasiões especiais, quando servem como alvo de relações de nomes e verbos

(Alonge *et al.*, 2000). Esse é o caso, por exemplo, de alguns verbos derivados de adjetivos, como *avermelhar*, *branquear*, *clarear* etc. Nesses casos, a codificação dos adjetivos é feita a partir da codificação de algum verbo ou substantivo.

Como o EWN não trata dos adjetivos de forma sistemática, será discutida aqui a sua organização multilíngüe. Essa organização serve de base para a organização multilíngüe do LOIS. Além disso, se os adjetivos não são contemplados dentro do EWN, eles o são dentro do ItalWordNet, que é uma extensão italiana do projeto europeu.

O EWN é composto por módulos independentes de língua e módulos de línguas específicas, as *wordnets*. Cada módulo de língua específica é estruturado em *synsets* com relações semântica como hiponímia, meronímia, causa e papéis, como, *agente*, *paciente*, *instrumento*, *local* (Peters *et al.*, 1998). Os módulos independentes de língua são três: a Ontologia de Domínio (OD); a Ontologia de Top-level (OT); e o Índice Interlíngüe (ILI).

O ILI é usado no EWN como uma interlíngua, em inglês, que conecta as diferentes *wordnets* entre si e entre a Ontologia de Top-level e a Ontologia de Domínio (cf. Figura 1). O ILI é uma lista de conceitos não-estruturados retirados em grande parte do WordNet 1.5 (Peters *et al.*, 1998; Vossen *et al.*, 1998). Através do ILI é possível o alinhamento multilíngüe das diferentes *wordnets* e, conseqüentemente, a equivalência e comparação dos *synsets* nas diversas línguas que compõem o EWN (Peters *et al.*, 1998).

O ItalWordNet (IWN) é uma expansão do módulo de língua italiana do projeto EuroWordNet (EWN). A exten-

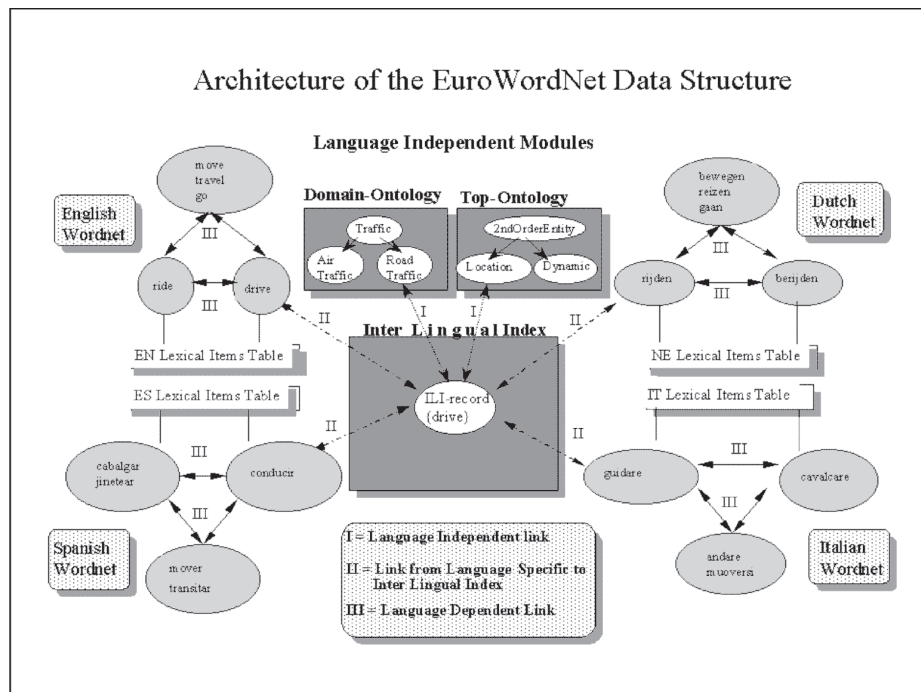


Figura 1. Arquitetura global da base de dados do EWN (Vossen *et al.*, 1998, p. 7).

são dessa base de dados foi feita acrescentando classes gramaticais não contempladas no EWN, como os adjetivos, os advérbios e os nomes próprios, aumentando a cobertura dos verbos e substantivos e criando um subgrupo terminológico econômico-financeiro (Alonge *et al.*, 2000; Roventini *et al.*, 2003) e outro jurídico (Sagri *et al.*, 2003).

Assim como o WN, o IWN organiza os adjetivos em *synset* (*sets of synonyms*), ou seja, grupos de sinônimos expressando um conceito. Porém, diferentemente do WN, que prega a relação de antonímia como a mais proeminente relação para a organização dos adjetivos, o IWN codifica também a hiponímia para essa classe gramatical. Alguns grupos de adjetivos são organizados em classes compartilhando o mesmo superordenado. Esse é o caso de adjetivos que indicam uma propriedade de “*contentor*”, como *aquoso*, *alcalino*, ou uma propriedade de “*apropriado para*”, como *defensivo*, *educacional*. Porém, as taxonomias construídas para os adjetivos são diferentes daquelas para os verbos ou substantivos, pois são mais simples, geralmente não ultrapassando dois níveis (Roventini *et al.*, 2003).

A seguir, apresentamos um quadro contendo as principais relações relacionadas aos adjetivos no IWN (Quadro 1).

A semântica dos adjetivos nas ontologias

Nesta seção, o objetivo é analisar as informações que as ontologias trazem sobre os adjetivos. Para tanto, analisaremos as ontologias SIMPLE e Mikrokosmos. O estudo dessas ontologias nos servirá de parâmetro para o experimento a ser apresentado em *Formalização dos adjetivos do domínio jurídico no Protégé*.

O SIMPLE

O SIMPLE é uma ontologia que divide os adjetivos em duas grandes classes: os extensionais e os intensionais (Peters e Peters, 2000). Os adjetivos extensionais são aqueles que atribuem uma propriedade ao nome que modificam, como, por exemplo, *americano*

em *Harry Truman foi um presidente americano*.

Os intensionais não atribuem propriedades aos nomes. Esses adjetivos estão geralmente associados a fatores temporais, modais, emotivos relacionados à predicação. É o caso de *atual*, em *Bush é o atual presidente dos Estados Unidos*, em que *atual* não atribui nenhuma propriedade ao nome, mas destaca o fator temporal envolvido na predicação. Os adjetivos podem ter um uso tanto extensional, como intensional. Um exemplo é *pobre*, que tem um uso extensional em *Mário é um homem pobre*, mas também aceita um uso intensional, em *O pobre Mário perdeu o emprego*. No último caso, *pobre* não atribui uma propriedade ao nome, mas ressalta um fator emocional envolvido na predicação.

Os adjetivos extensionais são, por sua vez, divididos em intersectivos e subsectivos. Os intersectivos são aqueles que atribuem ao hiperônimo de um nome a mesma propriedade que ao hipônimo. Um *carro vermelho* é um *automóvel vermelho*, porém, um *elefante pequeno* não é um *animal pequeno*. No caso de um adjetivo atribuir uma propriedade a um nome, que não é estendida ao seu hiperônimo, esse adjetivo é chamado de subsectivo. Assim, *pequeno* é um adjetivo subsectivo.

O Mikrokosmos

Segundo Raskin e Nirenburg (1998), os adjetivos no Mikrokosmos são divididos em cinco diferentes classes: graduáveis escalares, graduáveis não-escalares, não-escalares próprios, não-escalares deverbais e não-escalares denominais.

Raskin e Nirenburg (1998) consideram graduáveis aqueles adjetivos que podem ser usados comparativamente, como *bom*, em *Pedro é bom / Pedro é melhor que Paulo*. Os escalares são aqueles que podem ser mensurados através de uma escala de valores. Por exemplo, sabemos o que é *grande* através da comparação entre uma escala de valores em que *mínimo*, *pequeno*, *grande* e *gigante* se opõem. Os adjetivos escalares recebem um valor numérico no Mikrokosmos.

Quadro 1. Principais relações codificadas para os adjetivos no IWN (Alonge *et al.*, 2000, p. 46).

Relação	Classes Gramaticais ligadas	Exemplo
ANTONÍMIA	adj/adj	feliz/infeliz
ANTONÍMIA_GRADUÁVEL	adj/adj	bonito/feio
ANTONÍMIA_COMPLEMENTAR	adj/adj	vivo/morto
HIPONÍMIA	adj/adj	aquoso/contentor
PERTENCE_A	adj/subst	químico/química
É_UM_VALOR_DE	adj/subst	alto/estatura
ENVOLVIDO	adj/subst	dental/dente
CAUSA	adj/verbadj/subst	depurativo/depurar
SUJEITO_A	adj/verbadj/subst	processável/julgar

A partir de então, podemos compreender a divisão apresentada para os adjetivos acima. Os adjetivos graduáveis escalares são aqueles que podem ser usados comparativamente e que podem receber um valor numérico em uma escala de valores, como *grande*. Os adjetivos graduáveis não-escalares são adjetivos que não podem receber um valor numérico, mas podem ser usados comparativamente. Esses adjetivos, na verdade, são os adjetivos denominais que aceitam um uso comparativo, como *administrativo*, em *O estilo de Maria é mais administrativo que de Paulo*.

Fazem parte dos adjetivos não-escalares próprios os adjetivos pátrios como *brasileiro*, *português*, *francês* etc. Esses adjetivos não são nem escalares, nem graduáveis. São, na verdade, adjetivos denominais que indicam línguas e nacionalidades. Os adjetivos não-escalares deverbiais são aqueles que provêm de verbos, como *flamejante*, *comestível*, *amável*, *reciclável* etc. Os adjetivos não-escalares denominais, também chamados de verdadeiros relativos (Raskin e Nirenburg, 1998), são aqueles que derivam de nomes e que, diferentemente dos adjetivos graduáveis não-escalares, como *administrativo*, não podem ser usados comparativamente. É o caso de *civil*, *aeronáutico*, *ministerial* etc.

Em termos gerais, o WordNet divide os adjetivos em descritivos e relacionais, o SIMPLE em intensionais e extensionais e o MikroKosmos em escalares e não-escalares. Uma divisão em descritivos e relacionais tem a vantagem de dividir, de um lado, os adjetivos prototípicos que atribuem propriedades aos seres, e, de outro, adjetivos denominais, relacionando-os aos nomes que os originam. Porém, nem todos os adjetivos descritivos atribuem propriedades ao nome que acompanham. É o caso de *doce*. O adjetivo *doce*, em um *beijo doce* ou uma *bebida doce*, apresenta as mesmas informações semânticas que *água doce*. Não podemos dizer, no entanto, que *doce* atribua uma propriedade a *água*.

Uma divisão dos adjetivos em intensionais e extensionais, como no SIMPLE, resolve esse problema. Os adjetivos não são classificados segundo questões morfosintáticas, mas seguindo uma abordagem lógico-formal. Os adjetivos que atribuem propriedades ao nome que acompanham são extensionais; caso não haja essa atribuição de valores, o adjetivo é intensional. Essa abordagem é capaz de dar conta de fenômenos semânticos muito finos, como a intersecção, ou não, de atributos entre os componentes de uma classe hierárquica. É o caso, por exemplo, de *carro vermelho*, que é um *automóvel vermelho*, em que há a intersecção do atributo entre o hipônimo e o hiperônimo. O mesmo não acontece com *grande*. Um *rato grande* não é um *animal grande*.

A divisão escalares/não-escalares, adotada pelo MikroKosmos, está ligada à finalidade da ontologia, que é a tradução automática. Os escalares são aqueles que podem receber um valor, em uma escala de 1 a 10. Assim, *grande*,

em comparação com *pequeno*, *minúsculo* e *gigante*, receberia o valor 0.75 (Raskin e Nirenburg, 1998). Como os adjetivos escalares não são maioria em nosso corpus, uma abordagem como a empregada pelo MikroKosmos não contemplaria os nossos dados de maneira satisfatória. Assim, na seção 5 apresentamos uma proposta que acreditamos ser a mais apropriada para o tratamento de linguagens de domínios específicos, como o jurídico.

A descrição semântica dos adjetivos

Para este exercício, os adjetivos serão classificados em *qualificadores* e *classificadores*, seguindo Borba (1996). O motivo dessa escolha está ligado à natureza dos adjetivos do domínio jurídico, como veremos na seção 6, e à sua semelhança com a taxonomia adotada pelo WN. Através de uma pré-análise do nosso corpus, verificamos que a maioria dos adjetivos tinha a característica de criar classes para os substantivos que acompanhavam, ou seja, os adjetivos eminentemente classificadores. Os adjetivos *qualificadores* se incorporam à natureza do nome, representando uma avaliação subjetiva (1). Os *classificadores* apenas colocam o nome em uma classe, classificam as entidades do mundo (2).

- (1) interpretação *inadmissível*
- (2) código *civil*

Conforme Borba (1996), as principais características dos adjetivos *qualificadores* que os distinguem dos *classificadores* são: (i) a possibilidade de ocupar tanto a posição atributiva como a predicativa (3); (ii) a nominalização (4); (iii) a gradação (5); (iv) a combinação com verbos de avaliação (6); e (v) a construção exclamativa (7). Todas essas possibilidades são vetadas aos adjetivos classificadores.

- (3) a. interpretação *inadmissível* / a interpretação é *inadmissível*
b. código *civil* / *o código é *civil*
- (4) a. interpretação *inadmissível* / a *inadmissibilidade* da interpretação
b. código *civil* / *a *civilidade* do código
- (5) a. A interpretação é extremamente *inadmissível*. / A sua interpretação da lei é a mais *inadmissível* de todas.
b. *O ataque foi *muito* cardíaco. / *Esse ataque foi *mais* cardíaco que o outro.
- (6) a. Eu *acho* esta música muito bonita. / Eu *considero* este exercício difícil.
b. *Eu *acho* este ataque cardíaco. / *Eu *considero* este código *civil*.
- (7) a. Que interpretação *inadmissível*!
b. *Que código *civil*!

Assumimos aqui que adjetivos *descritivos* (Miller, 1999) e adjetivos *qualificadores* (Borba, 1996) correspon-

dem, grosso modo, à mesma classe de adjetivos, e que adjetivos *relacionais* (Miller, 1999), *classificadores* (Borba, 1996), *não-escalares denominais* (Raskin e Nirenburg, 1998) e *intensionais relacionados a objetos* (Peters e Peters, 2000) tratam-se, na verdade, de um único tipo semântico de adjetivos, os denominais, porém vistos sob óticas diferentes.

Em Raskin e Nirenburg (1998) e Peters e Peters (2000) vemos que os adjetivos denominais são codificados a partir da sua relação com o nome. Assim sendo, a principal informação atribuída a esses adjetivos é: “relativo a”, “pertencente a”. A relação “relativo a” torna-se um grande rótulo que agrupa sob si todos os casos de adjetivos denominais para os quais não se consegue codificar nenhuma outra relação. Aproximando essas duas abordagens a Borba (1996) poderemos enriquecer a descrição dos adjetivos denominais com a inclusão de informações adverbiais. Borba (1996, p. 186) cria uma subclasse entre os adjetivos classificadores chamada de *circunstanciais*. Os adjetivos circunstanciais são aqueles que provêm de uma função adverbial subjacente. Eles podem ser *locativos* (8), *temporais* (9), *instrumentais* (10), *causativos* (11) etc.

- (8) traumatismo *craniano* (que ocorreu no crânio)
- (9) publicação *mensal* (feita por mês)
- (10) conversa *telefônica* (por telefone)
- (11) doença *tropical* (causada pelo clima tropical)

Essas informações adverbiais podem ser formalizadas através de relações semânticas. A relação semântica que liga *craniano* a *traumatismo* é *ocorre em*. A relação *feita por* liga *mensal* a *publicação*. A relação *por meio de* liga *telefônica* a *conversa* e *causada por* liga *tropical* a *doença*. A formalização dessas informações amplia as relações “relativo a”, “pertencente a”, já adotadas pelos diferentes léxicos apresentados anteriormente.

Pensando em termos ontológicos, classificar os adjetivos em classificadores e qualificadores tem a vantagem de explicitar as relações taxonômicas que permeiam uma ontologia. Como os adjetivos classificadores colocam o nome que acompanham em uma classe, reconhecer os adjetivos classificadores nos ajuda a reconhecer os diversos níveis de classes possíveis dentro de uma ontologia.

O ponto forte para a formalização dos adjetivos qualificadores é que, como eles assumem a posição predicativa, são predadores. Assim sendo, eles podem ter o mesmo *status* que os verbos dentro da ontologia. Mas todos os adjetivos em posição predicativa são qualificadores? Segundo o que veremos em nossa análise, nem todos os adjetivos em posição predicativa são qualificadores.

Conforme vimos até agora, a distinção entre classificadores e qualificadores é muito simples e as possibilidades sintáticas de ambos os adjetivos estão muito bem definidas. Mas o que podemos dizer de um adjetivo como *ilegal*, que Borba (2002) considera como classificador, mas que ocorre na posição predicativa em nosso corpus? Ou seja, os adjetivos

classificadores podem ter também um uso qualificativo.

Outro dado que contraria a divisão sintática dos qualificadores como predicativos e classificadores como atributivos pode ser conferido na próxima seção. Dos seis adjetivos que ocorrem em posição predicativa em nosso corpus, apenas dois eram qualificadores. Isso significa que 66,67% dos adjetivos em posição predicativa eram classificadores, contra apenas 33,33% de qualificadores. É claro que uma amostra tão pequena não pode ser generalizada. A tendência de adjetivos classificadores ocorrerem em posição predicativa verificada neste trabalho é um indício que devem ser analisado futuramente. Para isso, necessitaremos de um corpus mais representativo e com um número maior de adjetivos em posição predicativa. Afinal, fica uma pergunta: o que diferencia adjetivos qualificadores de classificadores? Definitivamente, o critério sintático, conforme apresentado por Borba (1996), não é uma distinção relevante.

A delimitação entre classificadores e qualificadores é um ponto muito frágil. Conforme Borba (1996, p. 177), os adjetivos qualificadores “se incorporam (de forma acidental ou essencial) à natureza do nome, como se constituíssem um traço dele (...)”. Borba (2002) considera *suposto* como um adjetivo qualificador. Isso é o mesmo que dizer que *suposto*, em *um suposto assassino*, constituiria um traço intrínseco ao nome *assassino*. Na verdade, um *suposto assassino* não é nem *suposto* nem *assassino*, portanto considerá-lo como um qualificador não é uma escolha indubitável.

Classificar adjetivos em qualificadores e classificadores, conforme vimos no exemplo acima, não é uma informação suficientemente útil para sistemas de recuperação de informação. Na verdade, é necessária uma divisão geral, mais elevada, como *classificador* e *qualificador*, e, a partir dessa divisão, incluir informações trazidas pelo Mikrokosmos e pelo SIMPLE, como a escalaridade, os fatores modais, temporais e emotivos trazidos pelos adjetivos intensionais, e as propriedades físicas, sociais e psicológicas dos adjetivos extensionais. Informações temporais podem ser de muita utilidade para um sistema de pergunta e resposta. Supondo que um pesquisador queira descobrir alguma informação temporal, como “quem foi o presidente brasileiro de 1999 a 2002?”. Se o sistema de busca necessitar escolher entre duas sentenças “Lula é o presidente brasileiro” e “Fernando Henrique é o antigo presidente do Brasil”, o sistema deverá ter condições de inferir que o período de 1999 a 2002 é um período passado, e que o adjetivo “antigo” não caracteriza “Fernando Henrique” como presidente, mas como alguém, que em uma época passada, foi presidente do Brasil. Na próxima seção, demonstramos algumas das dificuldades expostas nesta seção.

Os adjetivos do domínio jurídico

Nesta seção apresentaremos a análise dos adjetivos de nosso corpus, em especial dos cinco adjetivos mais frequentes. Conforme já expusemos na introdução,

nosso corpus é composto por seis acórdãos coletados, via *web*, do site do Instituto das Tecnologias de Informação na Justiça, de Portugal. Como esse experimento compõe um primeiro esforço para a formação de uma ontologia *top-level*, a extração dos adjetivos foi feita manualmente. O critério sintático foi a primeira forma de classificarmos os adjetivos. Essa atitude se deve à importância que a informação sintática exerce na formalização dos adjetivos em ontologias, pois a posição dos adjetivos já nos indica a sua classificação semântica.

Em nosso corpus, encontramos um total de 66 adjetivos, considerando-se apenas os *types*. Esses adjetivos estão distribuídos da seguinte maneira: 61 em posição atributiva e seis em posição predicativa. Apenas um adjetivo se repete em ambas as posições. Para a análise, utilizamos a classificação semântica proposta por Borba (1996), que divide os adjetivos em qualificadores e classificadores. Para uma classificação mais rigorosa, não baseada na intuição, seguimos o *Dicionário de Usos do Português do Brasil*, de Borba (2002).

A Tabela 1 explicita uma das limitações da classificação que adotamos. Segundo Borba (1996), os adjetivos qualificadores são predicadores, enquanto os adjetivos classificadores são apenas atributivos. Na verdade, essa equação não corresponde aos dados encontrados no corpus. Podemos ver que os classificadores são a maioria no domínio jurídico: 59,09%, contra 40,91% de qualificadores. Mas o inesperado é que, na posição predicativa, os classificadores ocorrem em percentual muito maior, contradizendo a teoria de que classificadores ocorrem apenas na posição atributiva.

Para o experimento relatado na seção seguinte, optamos por analisar os cinco adjetivos mais frequentes. São eles: *civil*, *legal*, *supremo*, *patrimonial* e *sumário*. Dentre esses cinco, três são classificadores e dois são qualificadores, todos em posição atributiva.

Civil ocorreu 27 vezes em nosso corpus em quatro contextos diferentes: Código *civil* (17 ocorrências); Código processual *civil* (05 ocorrências); Responsabilidade *civil* (04 ocorrências); e Construção *civil* (01 ocorrência). Cada um desses contextos apresenta informações semânticas distintas, conforme Tabela 2.

O adjetivo *legal* aparece em segundo lugar (Tabela 3). Com oito ocorrências, sua distribuição dentro do corpus se dá em quatro diferentes contextos: Taxa *legal*

Tabela 2. Relações do adjetivo *civil*.

Civil: [Classificador] 1. que diz respeito às relações dos cidadãos entre si 2. relativo ao cidadão considerado em circunstâncias particulares dentro da sociedade

Contexto 1: Código *civil* Hipônimo de: *código*. Co-hipônimo de: *código penal*; *código comercial*; *código administrativo*; *código processual civil*. Civil se relaciona a: *cidadão*.

Contexto 2: Código processual *civil* Hipônimo de: *código*. Co-hipônimo de: *código penal*; *código comercial*; *código administrativo*; *código civil*. Civil se relaciona a: *cidadão*.

Contexto 3: Responsabilidade *civil* Hipônimo de: *responsabilidade*. Co-hipônimo de: *responsabilidade objetiva*; *responsabilidade subjetiva*. Civil se relaciona a: *cidadão*.

Contexto 4: Construção *civil* Hipônimo de: *construção*. Co-hipônimo de: *construção naval*. Civil se relaciona a: *cidadão*.

(três ocorrências); Juros *legais* (duas ocorrências); Presunção *legal* (duas ocorrências); e Regime *legal* (uma ocorrência).

O adjetivo *supremo* ocorreu sete vezes, em apenas um contexto, que também é um nome próprio: *Supremo Tribunal de Justiça* (Tabela 4).

O adjetivo *patrimonial* ocorreu seis vezes no corpus, em dois contextos diferentes: danos *não patrimoniais*, com quatro ocorrências; e danos *patrimoniais* com duas ocorrências (Tabela 5).

O adjetivo *sumário* teve quatro ocorrências no corpus (Tabela 6), aparecendo em três contextos diferentes, mas com o mesmo sentido para o Direito, são eles: Ação *sumária* (uma ocorrência); Forma *sumária* (uma ocorrência); e Processo *sumário* (uma ocorrência).

Seguindo o estudo apresentado aqui, passaremos para a formalização desses adjetivos no Protégé.

Formalização dos adjetivos do domínio jurídico no Protégé

O Protégé é uma ferramenta de código aberto, Java, com uma arquitetura que permite a edição de ontologias e sua conversão para o uso em aplicações baseadas em conhecimento. A grande vantagem do uso dessa ferra-

Tabela 1. Organização semântica dos adjetivos por *types*.

	Todos os adjetivos		Adj. Pos. Atributiva		Adj. Pos. Predicativa	
	nº	%	nº	%	nº	%
Classificadores	39	59,09	36	59,02	4	66,67
Qualificadores	27	40,91	25	40,98	2	33,33
Total	66	100	61	100	6	100

Tabela 3. Relações do adjetivo *legal*.

Legal : [Classificador] 1. prescrito pela lei.
Contexto 1: Taxa <i>legal</i> Hipônimo de: <i>taxa</i> . / Co-hipônimo de: taxa <i>ilegal</i> ; taxa <i>abusiva</i> . Legal se relaciona a: <i>lei</i> .
Contexto 2: Juros <i>legais</i> Hipônimo de: <i>juros</i> . / Co-hipônimo de: juros <i>ilegais</i> ; juros <i>abusivos</i> . Legal se relaciona a: <i>lei</i> .
Contexto 3: Presunção <i>legal</i> Hipônimo de: <i>conclusão</i> . / Co-hipônimo de: <i>conclusão de direito</i> . Legal se relaciona a: <i>lei</i> .
Contexto 4: Regime <i>legal</i> Hiperônimo de: regime <i>legal</i> do arrendamento urbano; regime <i>legal</i> da adoção. Legal se relaciona a: <i>lei</i> .

Tabela 4. Relações do adjetivo *supremo*.

Supremo [Qualificador] 1. extremo 2. último, derradeiro 3. máximo 4. do mais alto grau.
Contexto: <i>Supremo</i> Tribunal de Justiça Hipônimo de: <i>tribunal</i> . Co-hipônimo de: <i>tribunal da primeira instância</i> ; <i>tribunal da relação</i> . Supremo qualifica o nome.

Tabela 5. Relações do adjetivo *patrimonial*

Patrimonial [Classificador] 1. de ou relativo a patrimônio 2. que se baseia na propriedade.
Contexto 1: Danos não <i>patrimoniais</i> Hipônimo de: <i>danos</i> . / É sinônimo de: danos <i>físicos</i> ; danos <i>morais</i> . Co-hipônimo de: danos <i>patrimoniais</i> ; danos <i>indenizáveis</i> . Patrimonial se relaciona a: <i>patrimônio</i> .
Contexto 2: Danos <i>patrimoniais</i> Hipônimo de: <i>danos</i> . / É sinônimo de: danos <i>materiais</i> . Co-hipônimo de: danos <i>não patrimoniais</i> ; danos <i>indenizáveis</i> . Patrimonial se relaciona a: <i>patrimônio</i> .

Tabela 6. Relações do adjetivo *sumário*.

Sumário: [Qualificador] 1. formulado sem formalidades 2. simples; preliminar
Contexto 1: Ação <i>sumária</i> Hipônimo de: <i>processo declarativo comum</i> . É sinônimo de: forma <i>sumária</i> ; processo <i>sumário</i> . Co-hipônimo de: ação <i>ordinária</i> ; ação <i>sumaríssima</i> . Sumário qualifica o nome.
Contexto 2: Forma <i>sumária</i> Hipônimo de: <i>processo declarativo comum</i> . É sinônimo de: ação <i>sumária</i> ; processo <i>sumário</i> . Co-hipônimo de: forma <i>ordinária</i> ; forma <i>sumaríssima</i> . Sumário qualifica o nome.
Contexto 3: Processo <i>sumário</i> Hipônimo de: <i>processo declarativo comum</i> . É sinônimo de: ação <i>sumária</i> ; forma <i>sumário</i> . Co-hipônimo de: processo <i>ordinário</i> ; processo <i>sumaríssimo</i> . Sumário qualifica o nome.

menta é a possibilidade de conversão dos dados para linguagens como a *Ontology Web Language (OWL)*, o que torna a ontologia compatível com a *Web Semântica*. Com o auxílio desse editor, nossa proposta passa a ter um alcance maior na medida em que explicita esse conhecimento e possibilita o seu reuso.

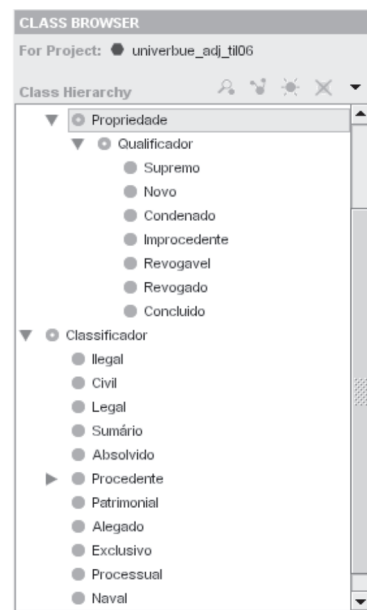
Para este exercício, partimos dos resultados apresentados por Alves (2005), que propõe recursos alternativos para contornar algumas limitações da ferramenta na sua proposta de inclusão da semântica verbal em uma ontologia para o mesmo domínio.

Na arquitetura do Protégé e na organização de ontologias, a relação de classe e subclasse é o principal critério estruturante. Criamos a superclasse *propriedades*, considerada pela ferramenta uma classe de papel abstrato, haja vista que não se aplica aos dados reais do corpus, servindo apenas para organizá-los na ontologia. O mesmo ocorreu com as categorias *classificadores* e *qualificadores*, consideradas *classes abstratas*.

Outra estratégia que se aplica a todos os adjetivos é a inclusão de definições, recurso denominado pelo Protégé de *documentação*. A Figura 2 ilustra esses aspectos.

A relação hiponímica é mais uma vez contemplada com a inclusão dos adjetivos classificadores. Isso porque os nominais formados por N + Adj Classificador, como *código civil*, *responsabilidade civil*, aparecem, no Protégé, como subclasses de nominais. Além dos nominais hipônimos, os adjetivos classificadores podem gerar outras relações. Incluímos, por exemplo, como um SLOT, a relação *relaciona-se a* para explicitar o elo entre o adjetivo *civil* e o substantivo *cidadão*.

No que tange à estruturação da semântica dos adjetivos qualificadores, vale enfatizar a possibilidade de

**Figura 2.** Hierarquia adjetivos.

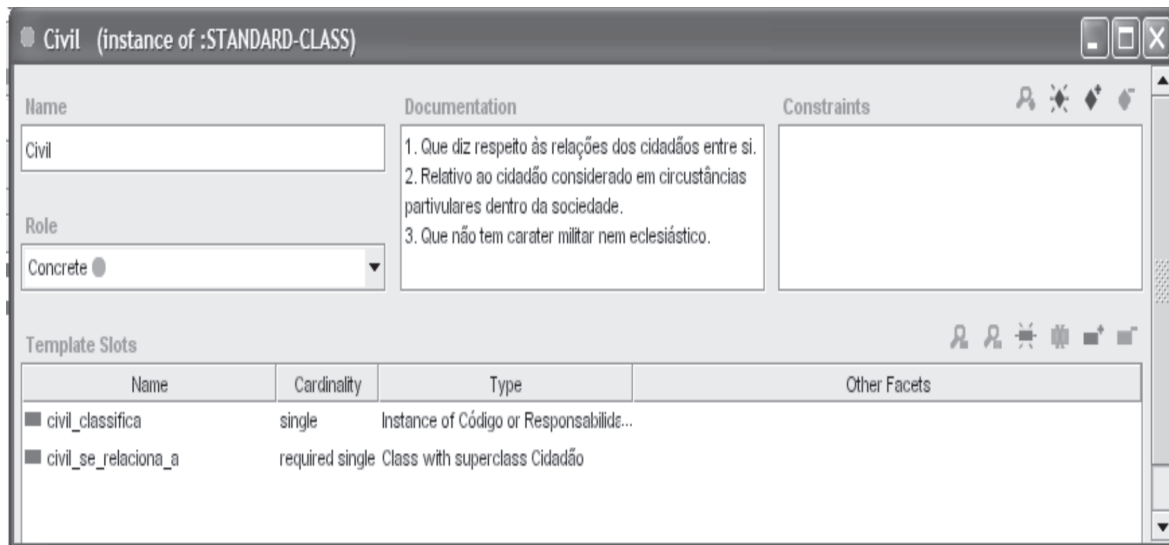


Figura 3. Adjetivo *civil*.

incluir as típicas relações de sentido, como a sinonímia e a antonímia. Os adjetivos *supremo* e *sumário* exemplificam essas possibilidades. A Figura 3 mostra a descrição das propriedades semânticas do adjetivo *civil*.

Considerações finais

O estudo teórico aqui apresentado, ainda que em caráter preliminar, trouxe-nos indicativos importantes no que tange à semântica dos adjetivos. Vimos que não é tarefa fácil escolher, dentre as tantas abordagens e taxonomias disponíveis, a que melhor se ajusta aos diferentes propósitos de pesquisa. Exploramos aqui uma dentre tantas perspectivas, que é a que parte de um critério sintático e subclassifica os adjetivos em classificadores e qualificadores. Há, contudo, outras distinções a serem investigadas: a distinção em graduáveis e não-graduáveis e a própria polissemia dos adjetivos são temas que teremos que enfrentar quando ampliarmos este trabalho.

Em termos aplicados, conseguimos chegar a uma primeira formalização no Protégé, ainda que modesta. Foi possível representar as particularidades de cada uma das subclasses: classificadores e qualificadores. Contudo, considerando que partimos de uma amostra pequena de adjetivos, nem todas as relações ilustradas na seção 6 tiveram correspondência na parte aplicada da pesquisa.

Por fim, como ponto positivo, enfatizamos o potencial da Semântica Lexical Computacional na construção de ferramentas computacionais.

Referências

ALONGE, A.; BERTAGNA, F.; CALZOLARI, N.; ROVENTINI, A. e ZAMPOLLI, A. 2000. Encoding information on adjectives in a lexical-semantic net for computational applications. In: FIRST CONFERENCE ON NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL

LINGUISTICS, Seattle, 2000. *Proceedings...* San Francisco, Morgan Kaufmann Publishers Inc. ACM International Conference Proceeding Series, 4:42-49.

ALONGE, A.; CALZOLARI, N.; VOSSEN, P.; BLOKSMA, L.; CASTELLON, I.; MARTI, M. A. e PETERS, W. 1998. The Linguistic Design of the EuroWordNet Database. *Computers and the Humanities*, 32(2-3): 91-115.

ALVES, I.M.R. 2005. *O uso da semântica verbal em sistemas de extração de informação: a construção de uma ontologia de domínio jurídico*. São Leopoldo, RS. Dissertação de Mestrado. Universidade do Vale do Rio dos Sinos – UNISINOS, 287 p.

BORBA, F.S. 1996. *Uma gramática de valências para o português*. São Paulo, Ática, 199 p.

BORBA, F.S. 2002. *Dicionário de usos do português do Brasil*. São Paulo, Ática, 1674 p.

GANGEMI, A.; SAGRI, M. T. e TISCORNIA, D. 2005. A constructive framework for legal ontologies. In: R. BENJAMINS (ed.), *Law and the Semantic Web: Legal ontologies, methodologies, legal information retrieval, and applications (Lecture Notes in Computer Science)*. Berlin/Heidelberg, Springer-Verlag, p. 97-124.

GANGEMI, A.; SAGRI, M. T. e TISCORNIA, D. 2003. Metadata for Content Description In Legal Information. In: Workshop on Legal Ontologies, 9th International Conference on Artificial Intelligence and Law (ICAIL-2003), Edinburgh, UK. Disponível em: www.loa-cnr.it/Papers/ICAIL.pdf. Acesso em: 13/09/2006.

PETERS, I. e PETERS, W. 2000. The treatment of adjectives in SIMPLE: theoretical observations. In: Second International Conference on Language Resources and Evaluation - LREC2000, Athens. Disponível em: http://www.ub.es/gilcub/SIMPLE/reports/papers/Adj_Peters.pdf. Acesso em: 13/09/2006.

PETERS, W.; VOSSEN, P.; DÍEZ-ORZAS, P. e ANDRIAENS, G. 1998. Cross-linguistic Alignment of Wordnets with an Inter-Lingual-Index. *Computers and the Humanities*, 32(2-3):221-251.

RASKIN, V. e NIRENBURG, S. 1998. An Applied Ontological Semantic Microtheory of Adjective Meaning for Natural Language Processing. *Machine Translation*, 13(2-3):135-227.

ROVENTINI, A.; ALONGE, A.; BERTAGNA, F.; CALZOLARI, N.; CANCELILA, J.; GIRARDI, C.; MAGNINI, B.; MARINELLI, R.; SPERANZA, M.; ZAMPOLLI, A. 2003. Italdwordnet: building a large semantic database for the automatic treatment

- of Italian. In: A. ZAMPOLLI; N. CALZOLARI e L. CIGNONI (eds.), *Linguistica Computazionale a Pisa*. Pisa/Roma, Istituti Editoriali e Poligrafici Internazionali, p. 745-791. [Edição especial de Linguistica Computazionale, **XVIII-XIX**].
- SAGRI, M.T.; TISCORNIA, D. e BERTAGNA, F. 2003. Jur-WordNet. In: P. SOJKA; K. PALA; P. SMRZ; C. FELLBAUM e P. VOSSSEN (eds.), *Second International Wordnet Conference - GWC 2004*. Brno, Masaryk University, p. 305-310.
- MILLER, K.J. 1999. Modifiers in WordNet. In: C. FELLBAUM (ed.), *WordNet: an electronic lexical database*. Cambridge, MIT Press, p. 47-67.
- VOSSSEN, P. 1998. Introduction to EuroWordNet. *Computers and the Humanities*, **32**(2-3):73-89.
- VOSSSEN, P.; BLOKSMA, L.; RODRIGUEZ, H.; CLIMENT, S.; CALZOLARI, N.; ROVENTINI, A.; BERTAGNA, F.; ALONGE, A. e PETERS, W. 1998. *The EuroWordNet Base Concepts and Top Ontology*. EuroWordNet (LE 4003) Deliverable D017, D034, D036. University of Amsterdam. Disponível em: <http://www.hum.uva.nl/~ewn>. Acesso em: 13/09/2006.

Submetido em:11/2006

Aceito em:11/2006

Anderson Bertoldi

UNISINOS, RS, Brasil

Rove Luiza de Oliveira Chishman

UNISINOS, RS, Brasil

Isa Mara da Rosa Alves

UNISINOS, RS, Brasil