

INTEGRANDO LÉXICOS SEMÂNTICOS E ONTOLOGIAS: UMA APROXIMAÇÃO A FAVOR DA *WEB* SEMÂNTICA

INTEGRANDO LEXICOS SEMÁNTICOS Y ONTOLOGÍAS: UNA APROXIMACIÓN DE LA *WEB* SEMÁNTICA

Rove Luiza de Oliveira Chishman – rove@unisinors.br

Professora Adjunta e Pesquisadora do PPG LA e Curso de Letras, Universidade do Vale do Rio dos Sinos (UNISINOS – RS).
Doutora em Linguística Aplicada pela PUC RS.
Pós-doutoranda na University of Texas at Austin, com pesquisa na área da Lexicografia computacional.

Resumo

Este artigo tem como objetivo central chamar a atenção, por meio de relatos de três propostas de estruturação ontológica, para os benefícios em se buscar uma estreita relação entre ontologias e léxicos semânticos computacionais. Para tanto, apresenta uma visão panorâmica dos léxicos computacionais e ontologias mais representativos, com ênfase nos recursos computacionais ligados ao domínio jurídico, temática predominante nos experimentos que serão descritos. A segunda parte do trabalho apresenta alguns estudos empreendidos pelo grupo de pesquisa SEMANTEC, no sentido de ilustrar como estes dois componentes podem se beneficiar mutuamente. Como consequência desta linha de pensamento, está a escolha consciente de modelos linguísticos consistentes para a construção de ontologias. Percebe-se, também que para fazer a *Web* um recurso global, é inevitável, além da semântica, tratar as questões multilíngues.

Palavras-chave

Ontologia. Léxicos computacionais. *Web Semântica*.

1 INTRODUÇÃO

Nesses últimos dez anos, como reação ao excesso de recursos de informação, pudemos testemunhar o interesse crescente em torno da temática da *web* semântica (BERNERS-LEE; HENDLER; LASSILA, 2001). Segundo essa nova concepção de *web*, busca-se tratar as informações da *web* como uma rede de conceitos em contraposição a uma rede de documentos. A idéia básica é transformar o conteúdo da *World Wide Web*

em dados processáveis por máquinas através de uma rede de conceitos, permitindo que agentes de *softwares* acessem este conteúdo tendo em vista as mais variadas aplicações computacionais, tais como sistemas de busca de informações.

Nesse novo cenário, as ontologias surgem como componentes primordiais, haja vista que cabe a elas a tarefa de definir o vocabulário básico e modelar os conceitos hierarquicamente. Dada a complexidade, a criação de ontologias tem exigido o esforço de equipes de profissionais multidisciplinares, o que não significa que as dificuldades tenham sido minimizadas. Entre os principais problemas, destacamos aqueles relacionados à própria estruturação da ontologia, etapa que requer a escolha de modelos conceituais, assim como o grau de granularidade exigido pelo domínio em questão. As questões de implementação, como definição de código e linguagem para formalização, não deixam de ser igualmente complexas.

Neste artigo, voltamo-nos para as questões de ordem conceitual da ontologia. É nosso propósito chamar a atenção para a estreita relação que pode haver entre ontologias e léxicos computacionais e os pontos positivos em reforçar a interação entre estes dois componentes. Para tal, nos apoiamos em Lenci, Calzolari e Zampolli (2002), que ressaltam haver um *continuum* entre ontologias e léxicos semânticos, já que ambos representam o conteúdo lexical das palavras. Recorremos também a Guarino (1998), que também destaca os pontos de contato entre léxicos computacionais e ontologias. A principal idéia que estes autores compartilham é a de que a integração entre estes dois componentes é condição indispensável para a implementação da *web* semântica.

Neste trabalho, nossa meta é ilustrar, através da apresentação do próprio percurso trilhado pelo grupo de pesquisa SEMANTEC¹, como podemos contemplar, em termos aplicados, esta estreita relação. É nosso propósito também atentar para a importância da lingüística, em especial a área da semântica lexical, como suporte teórico para a tarefa de modelagem conceitual de uma ontologia. Para tanto, o artigo está organizado em cinco seções. A seção de Introdução chama a atenção para o cenário da *web* semântica e o enfoque adotado neste trabalho. A seção 2 apresenta uma visão panorâmica dos léxicos computacionais e ontologias, com ênfase nos recursos computacionais ligados ao domínio jurídico, temática predominante nos projetos que serão descritos a seguir. A seção 3, em forma de relato de pesquisa, apresenta os principais resultados de dois projetos de construção de ontologia jurídica empreendidos pelo grupo. A seção 4 é o

¹ O grupo de pesquisa SEMANTEC (CNPq) é liderado pela Profa. Dra. Rove Chishman. O objetivo do grupo é desenvolver estudos semântico-computacionais voltados para a construção de léxicos e ontologias.

ponto culminante deste trabalho, ao refletir sobre a aproximação entre FrameNet e ontologias e sua aplicação à representação do domínio jurídico. Por fim, apresentamos as considerações finais, reforçando a idéia de que a modelagem conceitual de uma ontologia requer a escolha consciente a partir de um leque complexo de opções teóricas.

2 ONTOLOGIAS E LÉXICOS COMPUTACIONAIS

A definição do termo *ontologia* não é consensual entre os estudiosos da área. Uma das definições mais citadas, e também mais criticadas, é a de Gruber (1993, p.199): “[...] uma ontologia é uma especificação explícita de uma conceitualização.” Essa discordância se deve a dois motivos: (i) o termo *ontologia* é utilizado por diferentes áreas do conhecimento e (ii) mesmo entre os estudiosos da mesma área, há diferentes formas de conceber o que é uma ontologia.

Aristóteles se tornou o primeiro filósofo a apresentar um tratamento sistemático da lógica e da ontologia (SOWA, 1995). Em sua obra *As categorias*, Aristóteles apresenta 10 categorias básicas para classificar qualquer coisa que possa ser dita ou predicada sobre qualquer coisa. São elas: *Substância, Quantidade, Qualidade, Relação, Lugar, Tempo, Posição, Estado, Atividade e Passividade*. A partir de então, muitos filósofos têm se debruçado sobre o estudo das categorias ontológicas que melhor descrevem o mundo.

Nas aplicações computacionais, as ontologias têm o objetivo prático de oferecer estruturas de “conhecimento” para os sistemas computacionais e possibilitar, assim, que eles resolvam problemas de conhecimento de mundo relacionados com o processamento de linguagem natural e o raciocínio lógico automatizado (BATEMAN, 1993). Segundo Zúñiga (2001), ontologia para sistemas de informação é uma linguagem formal, projetada para representar um domínio particular do conhecimento, e o seu propósito de construção é funcional, ou seja, uma ontologia para sistemas de informação é planejada para um ou mais propósitos específicos.

Os léxicos computacionais, do ponto de vista lingüístico, possuem uma relação estreita com as ontologias. As ontologias como estrutura conceitual, que apresentam relações de significados entre os diferentes conceitos que estruturam um determinado conhecimento de mundo, podem incluir ou não o conhecimento lingüístico. De uma forma geral, as ontologias que descrevem conceitos mais gerais são conhecidas como

ontologias de *nível superior*, ou *top-level*. As ontologias de domínio descrevem o vocabulário relacionado a uma área em especial (GUARINO, 1998).

Os léxicos computacionais são outra forma de organizar o conhecimento semântico das línguas naturais. Esses léxicos podem ter diferentes formatos, dependendo da fundamentação teórica que os embasa. As ontologias podem ser utilizadas como forma de organização da estrutura hierárquica dos léxicos computacionais. É importante acrescentar também que os léxicos computacionais representam componentes indispensáveis nas aplicações multilíngües desenvolvidas pela comunidade de Tecnologia de Linguagem Humana (TLH). Tradutores e sistemas de extração de informação que precisam processar mais de uma língua são exemplos de aplicações que requerem um léxico robusto para representar o conhecimento lexical compartilhado pelas diferentes línguas.

Dois dos léxicos computacionais mais conhecidos no momento são o WordNet (FELLBAUM, 1998) e o FrameNet (FILLMORE; JOHNSON; PETRUCK, 2003). A estrutura do WordNet está fundamentada na semântica relacional. Assim, a representação do conhecimento lingüístico é feita através de *synsets* – *synonym sets* – ou grupos de sinônimos. Os *synsets* representam conceitos através de grupos de sinônimos. O FrameNet não está estruturado em grupos de sinônimos, mas em *frames* semânticos. Os *frames* são estruturas esquemáticas de situações e dos participantes das situações (FILLMORE, 1982).

O WordNet, versão 1.6, é uma base de dados lexicais que contém aproximadamente 94.000 formas nominais distintas, 10.000 formas verbais, 20.000 formas adjetivais e 4.500 formas adverbiais (FELLBAUM, 1998). Os itens lexicais no WordNet estão agrupados respeitando as quatro classes sintáticas chamadas de abertas: *substantivos*, *verbos*, *adjetivos* e *advérbios* (MILLER, 1995). Esses itens são reunidos em *synsets*, respeitando as classes sintáticas. A ligação entre os *synsets* e os itens lexicais é feita através de relações semânticas.

A base de dados lexical FrameNet contém mais de 10.000 Unidades Lexicais da língua inglesa, sendo mais de 6.100 totalmente anotadas, e um conjunto de mais de 940 *frames*. No FrameNet, as unidades de análise lexical fundamentais são o *frame* e a unidade lexical. “Quando o sentido de uma palavra está baseado em um frame particular, nós dizemos que a palavra evoca o frame.” (FILLMORE; JOHNSON; PETRUCK, 2003, p.

236). A unidade lexical é o conjunto de uma palavra com um sentido, ou seja, no FrameNet, cada sentido diferente de uma palavra representa uma unidade lexical distinta.

A descrição do léxico da língua inglesa leva em consideração a valência sintática e semântica dos itens lexicais. A valência sintática especifica os tipos frasais (sintagma nominal preposicional, etc) e as funções gramaticais (sujeito, objeto, etc). A valência semântica é descrita em termos de entidades de um frame evocado por uma unidade lexical, tais entidades são chamados de “elementos frame” (FILLMORE; JOHNSON; PETRUCK, 2003). A unidade lexical *acusar* é a evocadora do frame Judgment (Julgamento). A descrição semântica do frame Judgment inclui os elementos frame: *Cognizer*, *Evaluee*, *Expressor* e *Reason*. A descrição da estrutura esquemática representada por um frame é feita através dos elementos frame.

A arquitetura do WordNet e do FrameNet tem inspirado largamente a criação de léxicos, tanto de língua geral quanto de linguagem especializada. Passamos abaixo à discussão de algumas ontologias e redes semânticas desenvolvidas especificamente para a representação do conhecimento jurídico.

2.1 A Representação do Conhecimento Jurídico em Ontologias

As ontologias jurídicas podem ter diversas funções. Um dos critérios para a criação de uma ontologia é justamente a função a que ela se propõe. Ela pode servir como fonte de metadados para indexação automática de textos jurídicos e para estruturação de redes semânticas (SAGRI; TISCORNIA; BERTAGNA, 2003). As ontologias também podem servir como estrutura de conhecimento para a integração de bases de dados e sistemas de raciocínio artificial (VALENTE; BREUKER, 1995). Ou ainda, podem ser usadas como fonte de conceitos para a criação de outras ontologias mais especializadas (BREUKER, 2004). A seguir, apresentamos duas ontologias jurídicas com arquiteturas diferentes.

A **Core Legal Ontology** (CLO) é uma ontologia jurídica desenvolvida pelo *Instituto de Teoria e Técnicas da Informação Jurídica* (GANGEMI; SAGRI; TISCORNIA, 2005). A CLO organiza os conceitos do *JurWordNet*, um léxico terminológico do direito italiano organizado em formato de rede semântica. Através de um trabalho de integração de ontologias, a CLO herda os conceitos de nível superior de outra ontologia, a *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE) (MASOLO et al., 2003). A partir dos conceitos de nível superior da DOLCE, são criados novos conceitos para

abrigar os termos jurídicos da *wordnet* terminológica JurWordNet. A ontologia, representando o nível conceitual, organiza o nível lingüístico, ou seja, o nível das palavras e dos termos de uma língua.

A **LRI-Core** é uma ontologia jurídica criada pelo *Leibniz Center for Law*. O objetivo dessa ontologia é servir de suporte para a construção de ontologias de domínio juridicamente bem estruturadas e permitir a herança de termos bem definidos para a indexação automática de textos jurídicos (BREUKER, 2004). Os idealizadores da LRI-Core, insatisfeitos com as ontologias de nível superior disponíveis até o momento da sua construção, optam em criar os conceitos de nível superior que reúnem os demais conceitos da ontologia. A camada “superior” desta ontologia se divide em: **ENTIDADE FÍSICA, ENTIDADE MENTAL, ENTIDADE ABSTRATA, PAPEL e OCORRÊNCIA**. Esta camada da LRI-Core cumpre a função de cobrir um largo espectro de conceitos relacionados às fundações do senso comum em que o Direito está fundamentado (BREUKER, 2004).

2.2 A Representação do Conhecimento Jurídico em Redes Semânticas

As *wordnets* jurídicas, assim como as demais *wordnets* de língua geral, são baseadas no pioneiro WordNet. Aqui serão apresentados dois projetos de redes semânticas para a linguagem jurídica que inspiraram nosso trabalho em ontologia jurídica: o JurWordNet e o LOIS. JurWordNet e LOIS se completam em seus objetivos de representar o conhecimento jurídico e ajudar na extração da informação legal. O JurWordNet é um léxico monolíngüe para o italiano. Um de seus maiores desafios é tratar das diferenças sócio-dialetais que atrapalham leigos na busca por informações jurídicas na web. O LOIS, como uma extensão multilíngüe do JurWordNet, agrega a essa preocupação a da extração de informações multilíngües.

O **JurWordNet** é um léxico terminológico do domínio jurídico, organizado segundo relações semânticas, e ligado a um recurso genérico do italiano, o ItalWordNet (SAGRI; TISCORNIA; BERTAGNA, 2003). Seu desenvolvimento é baseado no WordNet de Princeton e no EuroWordNet. O JurWordNet alinha-se a uma série de projetos visando à criação de uma *web* semântica. Alguns dos esforços necessários para a viabilização de uma *web* semântica são a construção de bases de conhecimento conceitual e a etiquetagem de textos com informações semânticas.

O JurWordNet 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 *wordnets* de língua geral, a sinonímia no JurWordNet 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 e os sinônimos são considerados variantes sócio-dialetais. Assim, ao se ligar um léxico terminológico jurídico a um léxico de língua geral, aumentam-se as possibilidades de sucesso na busca de informações.

O **LOIS** (Lexical Ontologies for Legal Information Sharing) é um projeto da União Européia cujo objetivo é facilitar a extração de informação em bases de dados jurídicos através da construção de uma *wordnet* jurídica multilíngüe (DINI et al., 2005). Um instrumento de tal porte permitiria aos usuários, sejam leigos ou profissionais do Direito, pesquisar leis européias e documentos jurídicos quebrando-se a barreira lingüística. Seria possível, então, digitar uma pergunta em uma língua e recuperar documentos de outra língua pré-selecionada (SCHWEIGHOFER; LIEBWALD, 2005).

Para se chegar a tal resultado, o projeto prevê a representação de conceitos jurídicos 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 dados. Conceitos semelhantes em línguas diversas serão ligados, permitindo acesso a leis e decisões judiciais dos vários países que integram o projeto. Através dessa ferramenta, também é 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.

O LOIS está ligado aos EuroWordNet através de relações de *plug-in* e os *synsets* de todas as *wordnets* nacionais estão ligados ao Índice de Interlíngua (ILI). O ILI é uma lista de conceitos sem estruturação, formando um “supra grupo” de todos os conceitos de todas as *wordnets*. Todos os *synsets* das *wordnets* nacionais estão ligados ao ILI através de relações de *equivalência completa*, *quase-equivalência* ou *equivalência* como *hiperônimo* ou *hipônimo*. São as diferentes relações de equivalência que garantem a interconexão de todas as *wordnets* nacionais (CURTONI et al., 2005).

As relações utilizadas pelo LOIS são de dois tipos: monolíngues, quando envolvem apenas as relações próprias de uma língua, e multilíngues, quando as relações ligam conceitos de diferentes línguas. As relações monolíngues são aquelas que são estabelecidas entre os conceitos de uma mesma *wordnet*. Já as relações multilíngues ligam conceitos de diferentes línguas. O LOIS é um projeto de pesquisa que tem como objetivo prático a criação de um produto comercial: um tesouro jurídico eletrônico. Esse produto já está sendo comercializado.

É importante encerrar esta seção reforçando a idéia de que há um continuum entre ontologias e léxicos computacionais, o que pode, dependendo do empreendimento de pesquisa, parecer algo ambíguo. Os projetos apresentados nesta seção sugerem exatamente isto, pois já foram concebidos com o propósito de caracterizar o léxico como um sistema dinâmico.

3 UMA ONTOLOGIA JURÍDICA PARA O PORTUGUÊS: OS VERBOS E OS ADJETIVOS

Nesta seção, apresentamos o relato de duas iniciativas do grupo no que tange ao desenvolvimento de ontologias jurídicas. As ontologias que apresentamos foram desenvolvidas no contexto dos projetos SEMANTEC, DIRPI e LOIS².

3.1 Verbos

A primeira tarefa para a construção da ontologia dos verbos foi o estudo do *corpus* composto por 6 Acórdãos Judiciais sobre o tema “acidentes rodoviários”, homologados entre 2002-2003, disponibilizados eletronicamente³ nas bases de dados jurídicas do Instituto das Tecnologias de Informação na Justiça de Portugal. Feito o reconhecimento do *corpus*, foram selecionados os 10 verbos mais relevantes para o domínio, e analisados em 101 concordâncias, levando sempre em consideração seu papel para a situação de

² Estes estudos foram realizados em parceria com o grupo de pesquisa do Prof. Dr. Paulo Quaresma, do Departamento de Informática da Universidade de Évora, Portugal, em projeto de cooperação internacional no período de 2005 a 2007.

³ <http://www.dgsi.pt>

comunicação como um todo. A escolha do tema deve-se a integração da pesquisa aos projetos ONTOVERB⁴ e LOIS⁵.

A ontologia de verbos contou com a sistematização dos aspectos da semântica verbal em quatro níveis: (i) definição; (ii) relações lógico-semânticas; papéis semânticos; e elementos *frame*. A definição é um nível de representação importante para uma melhor interação da ontologia com humanos, mas, para o sistema, esse nível não é compreensível.

Sistematizar informações sobre as relações lógico-semânticas possibilita que o sistema reconheça o significado dos conceitos definidos na ontologia através das relações expressas entre eles. A identificação de conceitos sinônimos, por exemplo, possibilitará que a consulta ao sistema de busca seja feita através de palavras equivalentes, facilitando o acesso do usuário à base de dados. Informações de ordem sintático-semântica com base em tipos de situação e papéis semânticos auxiliam o sistema a interpretar e a construir frases coerentemente. As relações lógico-conceituais representadas na ontologia foram as seguintes: (a) acarretamento; (b) antonímia; (c) causa; (d) hiponímia; (e) sinonímia. A codificação desse nível da semântica verbal possibilitou uma grande ampliação da ontologia, de 10, passou a 107 entidades verbais. A hiponímia e a sinonímia foram a relação mais produtiva do domínio. A hiponímia facilitará a realização de inferências por parte do sistema, possibilitando, por exemplo, a interpretação de perguntas como: *que veículos automotores mais se acidentam?* O sistema, para responder a esse tipo de pergunta, deve ser capaz de associar *carro*, *motocicleta*, *motociclo*, *autocarro* (“ônibus” em PE) com *veículos automotores*. A relevância da sinonímia na ontologia aqui proposta deve-se à possibilidade de auxiliar o usuário no momento da busca, na medida em que ele não precisará se limitar a empregar um termo jurídico específico, podendo realizar suas consultas através de termos equivalentes.

Inserir informações sobre a estrutura argumental de um verbo em uma ontologia possibilitará que o sistema de busca reconheça que determinado verbo exige a presença de um conjunto X de argumentos e que esses argumentos precisam ser representados

⁴ O projeto ONTOVERB propôs-se a estudar a semântica dos verbos a fim de incluí-lo em uma ontologia de domínio *universidade* com o intuito de aperfeiçoar sistema de busca de informação na web. Projeto financiado pela UNISINOS, FAPERGS e CNPq.

⁵ *Lexical Ontologies for Legal Information Sharing – LOIS* - é um projeto que se propõe a construir uma *WordNet* multilíngue especializada para o domínio Jurídico nas seguintes línguas européias: italiano, inglês, alemão, checo, português e holandês.

por conceitos de determinado tipo (exemplo: humano, animado, abstrato, etc). Para classificar os argumentos dos verbos, fez-se uma sistematização dos papéis propostos por diferentes autores (exemplo: BORBA, 1996; CHAFE, 1970; DOWTY, 1979; FILLMORE, 1968; FRAWLEY, 1992; JACKENDOFF, 1975; KEARNS, 2000; SAEED, 1997). Os papéis semânticos identificados no *corpus* foram: (a) agente; (d) instrumento; (b) beneficiário; (e) paciente; (f) objetivo; (c) origem; (d) locativo; (e) propósito; (f) razão.

Entretanto, há informações que auxiliam na representação dos aspectos semânticos das entidades verbais que não se encontram nos limites da Semântica Lexical nem na Sintaxe; são informações situacionais que uma abordagem baseada em *frames* permite codificar. A noção de *frame* recebe interpretações variadas tanto no escopo da Linguística quanto da Computação. Algumas abordagens mantêm-se mais ligadas à estrutura da sentença, como as *wordnets*), outras menos, como o FrameNet, abordagem aqui adotada. Os elementos *frame* identificados no *corpus* foram: (a) avaliador (juiz, magistrado, tribunal); (b) avaliado (réu, representante, autor); (c) argumentador (réu, representante, autor); (d) reconhecedor (réu, representante, autor // juiz, magistrado, tribunal) (e) meio; (f) base legal; (g) razão; (h) propósito; (i) evidência; (j) tópico; (k) conteúdo; (l) mensagem; (m) pedido. Além desses, incluímos elementos *frame* contidos no cabeçalho dos documentos: (n) tempo; (o) condição; (p) local; e (q) maneira.

Com o auxílio de uma abordagem baseada em papéis semânticos, descobrimos como se estrutura a predicação, percebendo que *condenar*, por exemplo, exige argumentos como agente (argumento externo) e paciente (um dos argumentos internos), sendo ambos do tipo *humano*. Essa informação, no entanto, não é suficiente para que o sistema diferencie se é uma entidade como *réu* que deve ser o agente ou uma entidade como *juiz*. É necessário fornecermos subsídios ao sistema para que ele reconheça que *juiz* se presta para ser o agente de verbos como *condenar* e *julgar*, mas não para ser agente de verbos como *recorrer* e *provar*. Já a entidade *réu* somente pode ser paciente de *condenar* e *julgar*, mas pode ser agente de verbos como *recorrer* e *provar*.

Os níveis de análise dedicados aos papéis semânticos e aos elementos *frame* organizam informações semelhantes na semântica verbal. Tanto a abordagem de papéis semânticos quanto a de *frames* explicitam papéis para os elementos envolvidos. A diferença fundamental é que uma abordagem de papéis semânticos focaliza a *predicação* e classifica os participantes pertencentes da estrutura argumental e outras entidades acessórias e uma abordagem baseada em *frames*, por sua vez, identifica participantes da

situação de comunicação expressos ou não lexicalmente, previstos ou não pela estrutura argumental. Por localizar-se na interface entre sintaxe e semântica, uma abordagem baseada em papéis semânticos mostra-se mais rígida do que uma abordagem baseada em *frames*, a qual amplia as possibilidades de identificação dos participantes da situação de comunicação em que o verbo se insere. Os elementos *frame* propostos pelo FrameNet codificam essas informações contextuais; portanto, possibilitam diferenciar o tipo de agente ou de paciente de determinado verbo, como nos casos descritos acima. Uma abordagem baseada em *frames* possibilitará ainda incluímos na ontologia papéis para elementos da situação que não estão presentes na estrutura das sentenças de maneira explícita, mas que se referem a informações que podem ser recuperadas em porções anteriores do texto, em elementos do co-texto (como o cabeçalho) ou até a partir do conhecimento de mundo do falante. Os papéis semânticos e os *frames* incluíram na ontologia 74 entidades entre nomes e adjetivos, ampliando assim a abrangência da ontologia.

De posse da descrição ontológica dos verbos, as informações semânticas foram armazenadas no editor de ontologias Protégé, que possibilita a conversão dos dados para a linguagem OWL.

Com o auxílio do Protégé, as descrições semânticas aqui propostas passam a ter um alcance muito maior do que apenas a proposição de uma metodologia para a descrição semântica de verbos. Será possível atingir, dessa forma, os principais objetivos da construção de uma ontologia, são eles: (i) compartilhar conhecimento estruturado de informações comuns entre pessoas e máquinas (sistemas computacionais); (ii) possibilitar o reuso do conhecimento de determinado domínio; (iii) tornar explícito o conhecimento sobre determinado domínio; (iv) separar o conhecimento de um domínio do conhecimento operacional de construção de um sistema; (v) analisar o conhecimento de um domínio.

As vantagens da editoração de uma ontologia em uma ferramenta desse tipo são indiscutíveis, mas não se pode negar que esse exercício de organização do conhecimento verbal revelou também algumas limitações do Protégé.

3.2 Adjetivos

O *corpus* da pesquisa realizada com os adjetivos foi constituído por sentenças, pareceres e acórdãos coletados, via internet, do Instituto das Tecnologias de Informação

na Justiça de Portugal. Ao todo, foram 40 textos que renderam 173.855 palavras. Os adjetivos desse *corpus* totalizaram 7.880 *tokens* e 998 *types*. Os 20 adjetivos mais frequentes foram escolhidos para a análise semântica. Esses adjetivos foram agrupados segundo as suas características.

O grupo dos adjetivos denominados de *classificadores* compreende os seguintes adjetivos: *público*, *administrativo*, *jurídico*, *legal*, *diplomático*, *judicial* e *cultural*. O que esses adjetivos possuem em comum para serem postos juntos é o fato de estabelecerem relações taxonômicas, pois criam classes de entidades. Por exemplo, um “funcionário *diplomático*” é um tipo de “funcionário *público*”, o “Ministério *Público*”, o “Supremo Tribunal *Administrativo*” ou o “Tribunal *Administrativo* Sul” são “entidades *públicas*”.

É importante ressaltar a grande utilidade dos adjetivos classificadores para a organização estrutural de uma ontologia de domínio. Esses adjetivos explicitam relações taxonômicas como a hiponímia/hiperonímia. Essas relações taxonômicas são fundamentais para que um sistema de busca e extração de informação possa recuperar informações mais amplas.

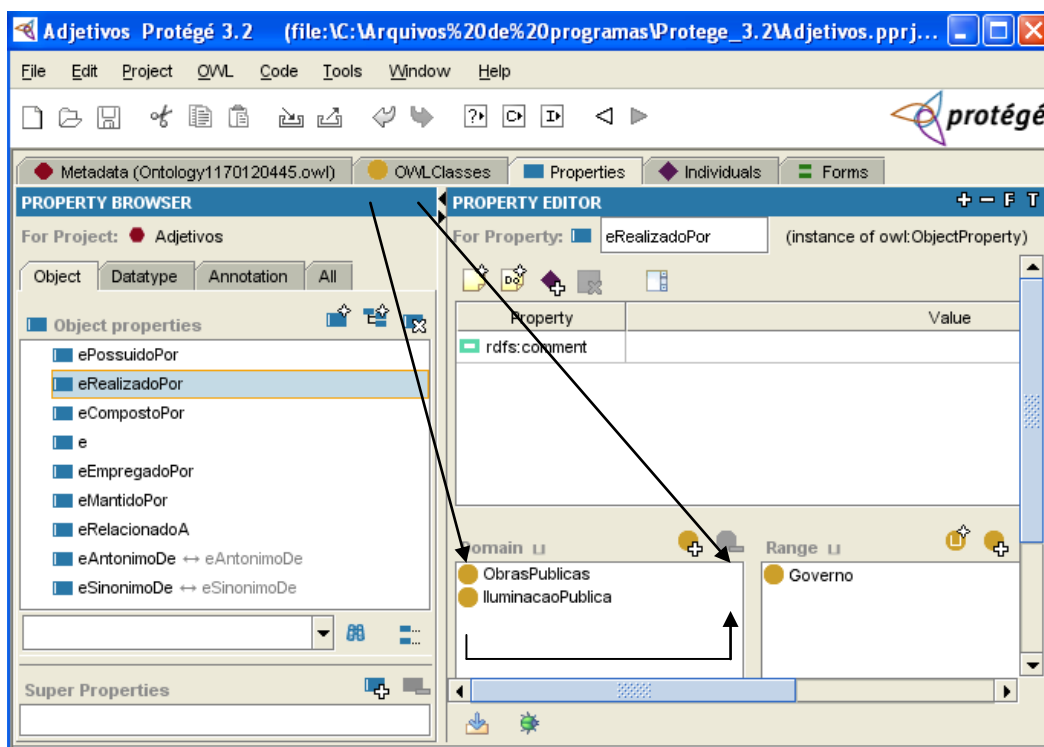


Figura 1. Formalização das relações para o adjetivo *público* no Protégé

Alguns adjetivos classificadores podem adquirir comportamento qualificador, como por exemplo, *público*, em “a rua é *pública*”, ou *cultural*, em “fiz uma viagem *cultural*”. O

contrário também é verdade. Adjetivos tipicamente qualificadores como *comum*, *geral* e *especial* ocorreram no *corpus* como classificadores apenas: “Justiça *Comum*”, “processo *comum*”, “regime *geral*”, “lei *geral*”, “regime *especial*”, “processo *especial*”. Através desses exemplos, percebe-se também que esses adjetivos classificadores, originalmente qualificadores, podem estabelecer relações de antonímia.

O que chamamos de adjetivos intensionais remissivos trata-se de uma ampla gama de classificadores que não se encaixam no grupo dos adjetivos classificadores prototípicos, vistos anteriormente. Esses adjetivos classificadores não possuem a característica prototípica atribuída a essa classe que é a classificação de entidades do mundo em subclasses mais especializadas.

Os adjetivos, quando em função remissiva, costumam apresentar um “esvaziamento” semântico. Alguns são mais vazios de significado, outros menos, porém, o que liga todos esses adjetivos é a remissão a uma entidade, do texto ou do mundo. Dentre os adjetivos do *corpus*, são intensionais remissivos os seguintes: *novo*, *anterior*, *presente*, *respectivo* e *próprio*. Esses adjetivos se ligam aos nomes para criar relações de remissão, não para lhes atribuir propriedades ou para classificá-los.

Sobre os adjetivos intensionais remissivos ficaram em aberto duas questões: (i) Como representar em uma ontologia essas relações de remissão? (ii) Como formalizar o conhecimento extralinguístico suscitado pelo adjetivo

Os adjetivos ditos *valenciais* são um grupo que compreende adjetivos relacionados semântica e morfologicamente a verbos e que herdam deles a necessidade de complementação, ou seja, a valência. Dentre os 20 adjetivos selecionados para esse estudo qualitativo, incluem-se nesse grupo *aplicável* e *relativo*.

Tomamos como exemplo o adjetivo *aplicável* e o seu verbo correlato *aplicar*. Pode-se dizer que *aplicar* é um verbo de *ação-processo* e, portanto, apresenta, pelo menos, dois argumentos: um sujeito *agentivo* ou *causativo* e um complemento *afetado* ou *efetuado* (BORBA, 1996). Em “x aplica z a y”, x é o *agentivo*, z é a *origem* e y é a *meta*, ou seja, “o juiz aplicou a lei ao caso”.

Em suma, o que fica de concreto sobre os adjetivos valenciais para a construção de ontologias é a possibilidade de definirmos a natureza da relação semântica que liga um adjetivo deverbal a um substantivo. A partir do conhecimento dessa relação, é

possível estabelecer o nível de relação entre esses itens lexicais, reconhecendo a natureza dos papéis temáticos: *meta*, *origem*, *paciente* etc.

Uma quarta categoria de adjetivos que consideramos na estruturação da ontologia são os *adjetivos modalizadores*. Neves (2000) trata os adjetivos modalizadores como uma subclasse dos qualificadores, dividindo-os em modalização epistêmica e modalização deôntica. Os adjetivos modais epistêmicos exprimem conhecimento ou opinião do falante, como, por exemplo, *óbvio*, *evidente*, *certo*, *provável*. Já os deônticos exprimem, por parte do falante, consideração de necessidade por obrigatoriedade, como *obrigatório*, *necessário*, *imprescindível*.

Dentre os 20 adjetivos selecionados como amostra do *corpus*, apenas um é modal: *necessário*. Trata-se de um adjetivo modal deôntico, ou seja, expressa obrigatoriedade. Se considerarmos que o Direito se caracteriza como uma ciência normativa baseada na infração de normas e punição, entenderemos a importância de formalizarmos a semântica dos adjetivos modalizadores, principalmente os deônticos.

Por fim, consideramos agrupamos como uma quinta categoria os *casos residuais*. Os casos residuais são, na verdade, adjetivos *qualificadores*. Esses adjetivos são os seguintes: *concreto* e *douto*. É próprio da natureza dos adjetivos qualificadores se agregarem aos substantivos compondo um traço intrínseco deles. Porém, o que se percebe no *corpus* é uma tendência ao uso intensional, ou seja, o atributo expresso pelo adjetivo não é inerente ao substantivo, mas é aliado ao substantivo em um determinado momento para restringir a sua significação.

Para tomarmos um exemplo do *corpus*, uma *douta sentença*, como em “[...] nenhum reparo há a fazer à *douta* sentença recorrida [...]”, não é algo que é *douto* e que é *sentença*, mas é algo que só é *douto* por representar a decisão de um ser humano. Os adjetivos que abordamos aqui não são qualificadores prototípicos e também não constituem uma subclasse bem delimitada. Assim, chamá-los de *qualificadores* seria desprezar a existência de um subgrupo que represente melhor os qualificadores. Ou seja, um subgrupo de adjetivos extensionais por excelência, o que é quase inexistente dentre os 20 adjetivos estudados.

4 FRAMENET E DOMÍNIO JURÍDICO: UMA INVESTIGAÇÃO PRELIMINAR

Nesta seção, apresentamos um estudo em andamento que parte da base de dados lexical FrameNet para a descrição do domínio jurídico. Ao contrário dos outros dois relatos, que levaram em conta modelagens teóricas variadas para descrever verbos e adjetivos, este trabalho se ocupa exclusivamente deste recurso lexical e não parte de uma categoria gramatical em particular.

Conforme já apresentamos na seção 2 deste trabalho, a base de dados lexical FrameNet, iniciada em 1997, está estruturada segundo a Semântica de *Frames* (FILLMORE, 1982). A Semântica de *Frames* é considerada um programa de pesquisa em semântica empírica que, além de enfatizar as continuidades entre a linguagem e a experiência, providencia uma estrutura para a apresentação dos resultados da pesquisa (FILLMORE, 1982; PETRUCK, 1996). Os *frames* são pequenas “cenas” ou “situações” abstratas, de forma que, para entender a estrutura semântica de um verbo, é necessário entender as propriedades de tais cenas esquematizadas. Assim, pode-se dizer que os *frames* são estruturas esquemáticas de situações e dos participantes das situações (FILLMORE, 1982).

A descrição do léxico da língua inglesa leva em consideração a valência sintática e semântica dos itens lexicais. A valência sintática especifica os tipos frasais (sintagma nominal preposicional, etc.) e as funções gramaticais (sujeito, objeto, etc.). A valência semântica é descrita em termos de entidades de um *frame* evocado por uma unidade lexical, tais entidades são chamados de “elementos *frame*” (FILLMORE; JOHNSON; PETRUCK, 2003). A unidade lexical *acusar* é a evocadora do *frame Judgment* (*Julgamento*). A descrição semântica do *frame Judgment* inclui os elementos *frame*: *Cognizer*, *Evaluee*, *Expressor* e *Reason*. A descrição da estrutura esquemática representada por um *frame* é feita através dos elementos *frame*.

O FrameNet aproxima muito o nível lexical do nível ontológico devido ao extenso número de relações codificadas entre os *frames*. Essas relações são de diferentes tipos, desde relações de herança, que organizam a hierarquia dos *frames* (por exemplo, o *frame Communication* é herdado pelo *frame Statement*), a relações de sucessão temporal (por exemplo, o *frame Criminal_process* é precedido pelo *frame Criminal_investigation*).

Outra vantagem do paradigma FrameNet em comparação a outros paradigmas de criação de léxicos computacionais, como o WordNet, é a descrição da estrutura

semântica dos verbos através dos participantes de uma determinada cena. Esses participantes são chamados de elementos *frame*.

Um *frame* pode apresentar as propriedades de diferentes *frames*. O *frame Judgment_communication*, por exemplo, como é a comunicação aberta de um julgamento, é uma combinação dos *frames Judgment* e *Communication*. Essa combinação é feita inclusive no nível dos elementos *frame* do *frame Judgment_communication*, que apresenta um *Communicator* no lugar do *Cognizer*, conforme quadro abaixo.

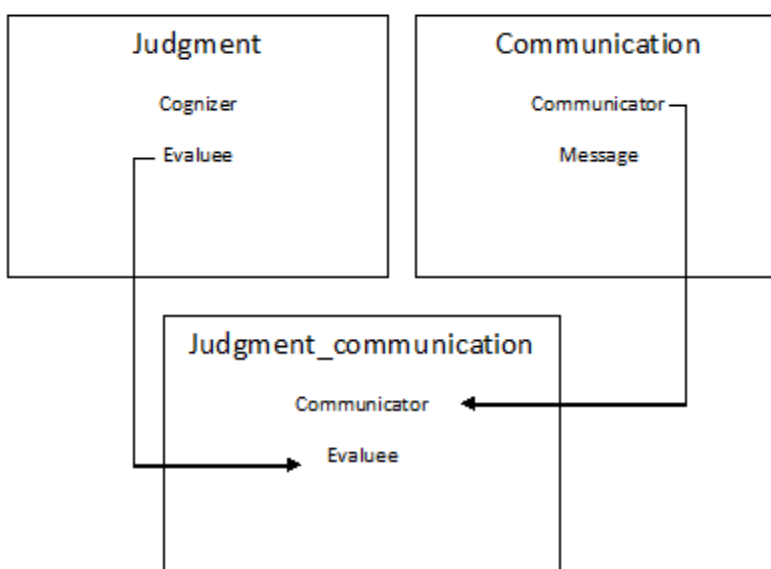


Figura 2: Herança de elementos *frame*

O trabalho de criação de bases de dados lexicais e ontologias com o FrameNet parte da comparação das unidades lexicais do Inglês e do Português, considerando-se o *frame* ao qual a unidade lexical pertence. Os *frames* semânticos são considerados não dependentes de língua (FILLMORE, 1976). Assim, parte-se de um *frame*, comparando as unidades lexicais da língua inglesa com a língua portuguesa. Os *frames* oferecem um modelo conceitual para estruturação do léxico. As relações entre *frames* podem ajudar a interpretar o significado das unidades lexicais, à medida que organizam a estrutura da ontologia.

Para exemplificar as possibilidades de uso do FrameNet para construção de ontologias e bases de dados lexicais, bem como os desafios de se partir de uma base de

dados da língua inglesa, tratamos de dois itens lexicais importantes para o domínio jurídico: *acusar* e *julgar*. Para este estudo que apresentamos aqui, utilizamos um *corpus* composto de 50 textos jurídicos, totalizando 1.055.535 *tokens*, estes apresentando 39.108 *types* (palavras distintas).

O estudo do item lexical *acusar*, que ocorreu nove vezes no *corpus*, aponta para 4 significados diferentes, o que caracterizaria 4 unidades lexicais distintas. Das nove ocorrências, duas remetem ao *frame* “Judgement” (exemplo 1) e três ao *frame* “Judgment_communication” (exemplo 2).

- (1) [...] **Acusa**-se a Polícia [evaluatee] como sendo a responsável pelo aumento da criminalidade [reason] [...].
- (2) [...] chegou ao exagero de dizer que **acusar** de corporativismo [reason] os advogados [evaluatee] [...].

Pode-se ver que o *frame Judgment_communication* é uma especificação do *frame Judgment*, pois requer uma comunicação do julgamento. Na sentença (1), o verbo *acusar* pode ser considerado como sinônimo de *avaliar*, *considerar*, enquanto que na sentença (2) o verbo *acusar* pode ser considerado como sinônimo de *criticar*, um verbo evocador do *frame Judgment_communication*.

As outras quatro ocorrências restantes estão assim divididas: uma ocorrência em que *acusar* é sinônima de *apontar*, *indicar* (exemplo 3) e três ocorrências em que *acusar* está ligada à linguagem especializada (exemplo 4), indicando o ato jurídico de indicar um suspeito como réu de um processo.

- (3) [...] uma tradição que já **acusava** desgaste [...].
- (4) [...] em virtude da separação das funções de **acusar** e julgar [...].

O FrameNet não apresenta toda essa polissemia formalizada em sua base de dados lexical, o que não significa que tais itens lexicais não sejam altamente polissêmicos também em Inglês. A polissemia do item lexical *acusar* no FrameNet está restrita a apenas dois sentidos: o *frame* “judgment” e o *frame* “judgment_communication”. Nesse sentido, um trabalho orientado à linguagem de especialidade pode criar bases de dados lexicais ainda mais robustas na língua alvo, à medida que se percebe essa polissemia e que se criam novos *frames* e novas relações para formalizar os novos sentidos ainda não formalizados no FrameNet.

O item lexical *julgar* também apresentou polissemia. Dois sentidos distintos foram identificados. O verbo *julgar* ocorreu 271 vezes no *corpus*: 244 vezes indicando o processo jurídico de avaliação de um fato ou uma pessoa, ou seja, indicando um uso especializado para o verbo *julgar* (exemplo 5) e 27 vezes indicando um processo de reflexão, ligado às crenças e costumes sociais, ou seja, um uso não especializado (exemplo 6). Ambos os usos remetem ao *frame* “Assessing”.

- (5) [...] a Justiça Militar Estadual [assessor] **julgaria** o PM [phenomenon] e a Justiça Comum [assessor], **julgaria** o Civil [phenomenon] [...].
- (6) [...] para fazer com que os homens [assessor] que ainda a cometem parem de se **julgar** superiores [feature] às mulheres [phenomenon] [...].

Neste ponto levantam-se algumas perguntas para as quais ainda não se possui resposta: se são sentidos diferentes não deveriam remeter a *frames* distintos? Deve-se considerar que *julgar* possui sentidos diferentes, ou apenas usos diferentes? O que devemos fazer com a linguagem especializada no contexto de formalização da língua em uma base de dados baseada em *frames*? Outro ponto de discussão chama a atenção ao se analisar o verbo *julgar*. Das 245 ocorrências em que o verbo remete a um conhecimento especializado, em 12 ocorrências o verbo *julgar* é utilizado para proferir o resultado da avaliação, como no exemplo (7), podendo ser também performativo, conforme exemplo (8). Nesses casos, o verbo *julgar* tem o valor de *declarar*.

- (7) [...] o MM. Juiz Federal Dr. José Godinho Filho [assessor] **julgou** improcedente [feature] o pedido [phenomenon] [...].
- (8) **Julgo** improcedente [feature] a ação direta [phenomenon].

Os dados lexicográficos apresentados aqui expressam os desafios e as vantagens de se empregar o paradigma FrameNet para a análise linguística visando à criação de bases de dados lexicais. A possibilidade de se criar um léxico estruturado através de relações semânticas, aproximando esse léxico da estrutura de uma ontologia, pode ser muito útil para a recuperação de informação. Dentre os desafios a serem gerenciados, há a polissemia, que, em se tratando de *corpus* de linguagem especializada, como o utilizado nesta pesquisa, está ligada não apenas aos diferentes sentidos de um item lexical na língua geral, mas também aos diferentes sentidos da língua geral em contraste com o sentido especializado. Há os diferentes padrões de lexicalização entre o Inglês e o Português e para um estudo mais profundo seria necessário um estudo de equivalências

de tradução. Por fim, a relação sintático-semântica entre os diferentes sentidos de um mesmo item lexical deve ser analisada para a inclusão desses itens em uma base de dados lexical.

5 CONSIDERAÇÕES FINAIS

Nosso propósito, neste artigo, foi chamar a atenção para a benéfica integração entre ontologias e léxicos semânticos computacionais. Podemos inclusive falar em convergência natural entre os dois componentes, o que é visível em muitos dos projetos apresentados na seção 2 deste trabalho. As iniciativas do grupo de pesquisa SEMANTEC, conforme o relato da seção 3, seguem à risca esta orientação. A representação semântica dos verbos na ontologia jurídica inclui informações tipicamente abarcadas por um léxico semântico, e não por uma ontologia. A inclusão da semântica dos adjetivos em uma ontologia, por sua vez, ocorre tendo em vista uma rica caracterização dos sentidos desta categoria. Aliás, nada típico em uma ontologia. Por fim, concluímos atentando para outra estreita integração que parece reforçar mais ainda o que nos propusemos a apresentar neste trabalho: a estreita relação entre web semântica e multilinguagem. Em outras palavras, para fazermos da web um recurso global, é inevitável que nos preocupemos, além da disponibilidade do conteúdo, das questões multilíngues.

REFERÊNCIAS

- BATEMAN, J. A. Ontology construction and natural language. In: WORKSHOP ON FORMAL ONTOLOGY IN CONCEPTUAL ANALYSIS AND KNOWLEDGE REPRESENTATION, 1993, Padova. *Proceedings...*, Padova, 1993. p.83-93.
- BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The semantic web. *Scientific American*, v.284, n.5, p.35-43, 2001.
- BORBA, F. S. *Uma gramática de valências para o português*. São Paulo: Ática, 1996. 199p.
- BREUKER, J. A. Constructing a legal core ontology: LRI-Core. In: FREITAS et al. (Ed.). *Proceedings Wonto-2004 on ontologies and their applications*. São Luís, 2004. p.115-126.
- CHAFE, W. L. *Meaning and the structure of language*. Chicago: University of Chicago Press, 1970.
- CURTONI, P. et al. Semantic access to multilingual legal information. In: SCHWEIGHOFER, E. (Ed.). *WORKSHOP "FREE EU INFORMATION ON THE WEB: THE*

FUTURE BEYOND THE NEW EUR-LEX” OF JURIX 2005: THE 18 TH ANNUAL CONFERENCE ON LEGAL KNOWLEDGE AND INFORMATION SYSTEMS. Proceedings... Brussels: Vrije Universiteit, 2005. p.1-11.

DINI, L. et al. Cross-lingual legal information retrieval using a WordNet architecture. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW, 10., 2005, Bologna. *Proceedings...* New York: ACM, 2005. p.163-167.

DOWTY, David. *Word meaning and montague grammar*. Dordrecht: Reidel, 1979.

FELLBAUM, C. A semantic network of English: the mother of all wordnets. *Computers and the Humanities*, v.32, n.2-3, p.209-220, 1998.

FILLMORE, C. J. *The case for case: universals in linguistic theory*. Edited by Emmon Bach and Robert Harms. New York: Holt, Rinehart and Winston, 1-88, 1968.

FILLMORE, C. J. Frame semantics and the nature of language. *Annals of the New York Academy of Science*, v. 280, p. 20-32, 1976.

FILLMORE, C. J. Frame Semantics. In: THE LINGUISTIC SOCIETY OF

KOREA (Ed.). *Linguistic in the morning calm*. Seoul: Hansinh Publishing, 1982. p. 111-137.

FILLMORE, C. J.; JOHNSON, C. R.; PETRUCK, M. R. L. Background to FrameNet. *International Journal of Lexicography*, v. 16, n. 3, p. 223-230, 2003.

FRAWLEY, W. *Linguistic semantic*. London: Lawrence Erlbaum Associates, Publishers, 1992.

GANGEMI, A.; SAGRI, M. T.; TISCORNIA, D. A constructive framework for legal ontologies. In: BENJAMINS, V. R. et al. (Ed.). *Law and the semantic web: legal ontologies, methodologies, legal information retrieval, and applications*, LNAI (3369). Berlin/Heidelberg: Springer-Verlag, 2005. p.97-124.

GRUBER, T. R. A translation approach to portable ontology specifications. *Knowledge Acquisition*, v.5, n.2, p.199-220, 1993.

GUARINO, N. Formal ontology and information systems. In: GUARINO, N. (Ed.). *Formal ontology in information systems: Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998*. Amsterdam: IOS Press, 1998, p.3-15.

JACKENDOFF, R. (1975) Morphological and semantic regularities in the lexicon. *Language*, n.5, p.639-71.

KEARNS, Kate. 2000. *Semantics*. New York: St. Martin's Press

LENCI, A.; CALZOLARI, N.; ZAMPOLLI, A. From text to content: computational lexicons and the Semantic Web. *AAAI Technical Report WS*, 2002.

MASOLO, C. et al. *Wonderweb deliverable d18: ontology library*. Trento: Technical report, ISTC-CNR, 2003.

MILLER, G. A. WordNet: a lexical database for English. *Communications of the ACM*, New York, v. 38, n. 11, p.39-41, 1995.

NEVES, M. H. M. *Gramática de usos do português*. São Paulo: UNESP, 2000. 1037p.

PETRUCK, M. Frame semantics. In: VERSCHUEREN, J. et al. (Ed.). *Handbook of pragmatics*. Philadelphia: John Benjamins, 1996.

SAEED, John I. *Semantics*. Malden: Blackwell Publishers, 1997.

SAGRI, M. T.; TISCORNIA, D.; BERTAGNA, F. Jur-WorNet. In: SOJKA, P. et al. (Ed.). *Second international wordnet conference - GWC 2004*. Brno: Masaryk University, 2003. p.305-310.

SCHWEIGHOFER, E.; LIEBWALD, D. Advanced lexical ontologies and hybrid knowledge based systems: first steps to a dynamic legal electronic commentary. In: LEHMANN, J. (Ed.). *Loait: legal ontologies and artificial intelligence techniques*. Nijmegen: Wolf Legal Publishers, 2005. p.71-81.

SOWA, J. F. Top-level ontological categories. *International Journal of Human-Computer Studies*, v.43, n.5-6, p.669-685, 1995.

VALENTE, A.; BREUKER, J. A. ON-LINE: An architecture for modeling legal information. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW. *Proceedings of the fifth international conference on artificial intelligence and law*. New York: ACM Press, 1995, p 307-315.

ZÚÑIGA, G. L. Ontology: its transformation from philosophy to information systems. In: INTERNATIONAL CONFERENCE ON FORMAL ONTOLOGY IN INFORMATION SYSTEMS, 2001, New York. *Proceedings...*New York: ACM Press, 2001. p.187-197.

Title

Integrating semantic lexicons and ontologies: an approximation in favour of the Semantic Web

Abstract

This article, through reports of three ontology proposals, aims to stress the benefits of a closest relationship between ontology and computational semantic lexicons. In order to reach this objective, firstly, an overview of the most representative computational lexicons and ontology is presented, with emphasis on the computational resources related to the juridical area, which is a predominant theme in the experiments to be described. In the second part of the work, some studies undertaken by the SEMANTEC research group are presented, to illustrate how these two components can be mutually benefic. As a main consequence of this line of thought is the choice of a consistent linguistic for ontology's building. It stresses out that to turn the Web a global resource, it is inevitable, besides semantic, to focus on multi-language's issues.

Keywords

Ontology. Computational lexicons. Semantic Web

Título

Integrando lexicos semânticos y ontologías: una aproximación a favor de la web semántica

Resumen

Este artículo tiene como objetivo principal llamar la atención, por medio de relatos de tres propuestas de estructuración de ontologías, los beneficios en buscar una relación angosta entre ontologías y léxicos semânticos computacionales. Para tanto, presenta una visión panorâmica de los léxicos computacionales y ontologías más representativos, con énfasis en los recursos computacionales conectados al dominio jurídico, tema predominante en los experimentos que serán descritos. La segunda parte del lo trabajo presenta algunos estudios emprendidos por el grupo de investigación SEMANTEC, en el sentido de ilustrar como estos dos componentes pueden beneficiarse mutuamente. Como consecuencia de esta línea de pensamiento, está la elección consciente de modelos lingüísticos consistentes para la construcción de ontologías. Se percibe también que para hacer la Web un recurso global, es ineludible, además de la semántica, tratar las cuestiones multilíngues

Palabras clave

Ontología; Léxicos Computacionales; Web Semántica

AGRADECIMENTOS

Aos doutorandos Anderson Bertoldi (CAPES – UNISINOS) e Isa Mara da Rosa Alves (UNESP – Araraquara), pela contribuição para a realização deste estudo.

A CAPES, pelo Auxílio de Estágio Pós-Doutoral na University of Texas at Austin, março a junho de 2009, período em que este artigo foi redigido.

Recebido em: 24.05.2009

Aceito em: 31.08.2009