Busca e Extração de Informações através de Pergunta e Resposta: uma nova concepção de Web

Isa Mara da Rosa Alves[1], Rove Luiza de Oliveira Chishman[2], Paulo Miguel Torres Duarte Quaresma e José Saias[3]

[1] Programa de Pós-Graduação em Lingüística e Língua Portuguesa – Universidade Estadual do Estado de São Paulo (UNESP), Araraquara – SP – Brasil

[2]Programa de Pós-Graduação em Lingüística Aplicada – Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo – RS – Brasil

[3]Departamento de Informática -
Universidade de Évora, Évora – Portugal

isamralves@gmail.com, rove@icaro.unisinos.br, {pq, jsaias}@di.uevora.pt

Abstract. *This paper carries out a descriptive study of verbs of the juridical domain aiming at the construction of an ontology which may contribute towards the enhancement of an Information Retrieval System. The approaches which have been more useful for the description of verbal semantics include logical-semantic relationships, semantic roles, and frames. This ontology is going to be applied in the system of Procuradoria Geral da República de Portugal, that, from the basic search for key words will become able to interact with users in natural language through questions and answers in a more efficient way.*

Resumo. *Este trabalho propõe-se a realizar um estudo descritivo de verbos do domínio jurídico com vistas à construção de uma ontologia que possa contribuir para o aperfeiçoamento de sistemas de busca e extração de informações na Web. Das abordagens estudadas, as que mais se prestaram para a descrição da semântica verbal com os fins aqui desejados foram as relações lógico-semânticas, os papéis semânticos e os frames. A ontologia proposta aqui será aplicada no sistema da Procuradoria Geral da República de Portugal, o qual poderá interagir com o usuário em língua natural através de pergunta e resposta de maneira mais eficiente.*

1. Introdução

Com a constante difusão mundial da *World Wide Web*, a tentativa de criar sistemas de busca cada vez mais eficientes (ágeis e precisos) é fato marcante na comunidade científica. A fim de possibilitar que tais ferramentas sejam capazes de processar o conteúdo semântico de suas bases textuais, a W3C[1] criou uma nova concepção de Internet: a Web Semântica (do Inglês: *Semantic Web*).

[1] O World Wide Web Consortium (W3C).

É nesse contexto que se insere o trabalho a ser apresentado aqui. Discutiremos os resultados de um estudo que compreende a construção da ontologia de verbos de domínio jurídico – UNIVERBUE – que será aplicado na melhoria do desempenho de sistemas *on-line* de busca e extração de informações, particularmente, o sistema da Procuradoria Geral da República de Portugal (PGR) que se tornará apto a interagir com os usuários fornecendo a eles respostas mais exatas às questões em língua natural. Esta é uma pesquisa interdisciplinar, que visa a fazer um estudo dos verbos do domínio jurídico no âmbito da Semântica Lexical Computacional tendo em vista uma aplicação computacional específica. Propomos aqui um critério para a descrição das propriedades semânticas de verbos que serviu para a construção de uma ontologia, a qual foi convertida para Ontology Web Language (OWL) e será aproveitada para o aperfeiçoamento do sistema de busca e extração de informações da PGR de Portugal.

Este trabalho organiza-se da seguinte maneira: na segunda seção, apresentaremos alguns projetos de construção de ontologias e léxicos computacionais a fim de identificarmos, entre as metodologias empregadas, aquelas que descrevem informações lingüísticas úteis à construção da nossa ontologia; na terceira seção, trataremos de abordagens semânticas próprias para a representação da semântica verbal; na quarta seção, trataremos da etapa lingüística de construção da ontologia UNIVERBUE; na quinta seção, descreveremos como as informações semânticas representadas na estrutura ontológica proposta foram inseridas no editor Protégé; na sexta seção, por fim, apresentaremos uma situação que exemplifica a contribuição que a ontologia proposta dará para a melhoria do desempenho do sistema de busca de informação.

2. Ontologias

Há vários projetos que objetivam o aperfeiçoamento dos sistemas de busca tanto na comunidade científica quanto entre *sites* com fins comerciais. Esses esforços são motivados por uma nova concepção de Web – a Web Semântica -, a qual prevê a construção de ontologias contendo detalhadas informações semânticas estruturadas de maneira explícita para tornar acessível aos sistemas de Processamento Automático da Língua Natural (PLN) o conteúdo semântico de documentos até então legíveis apenas pelos usuários. Nesse contexto, a principal linguagem padrão para a marcação de hipertexto, HTML (HyperText Markup Language), será complementada pela OWL (Ontology Web Language). A OWL não é uma simples linguagem de marcação de hipertexto com informações estruturais apenas; ela possui semântica e sintaxe próprias para a modelação do conteúdo semântico de textos. Dessa maneira, o sistema estará apto a realizar inferências, estabelecendo relações de sentido entre as bases eletrônicas e as informações fornecidas pelo usuário na solicitação da busca.

É para auxiliar na interpretação do significado das informações textuais que os computadores usarão ontologias. Elas fornecem "pistas" ao sistema de PLN, guiando-o na busca pela interpretação mais adequada de um texto. No contexto do trabalho aqui proposto, uma ontologia representa um conjunto de asserções que definem as relações entre os conceitos e estabelecem regras lógicas de raciocínio sobre eles. Essas relações envolvem informações de cunho sintático-semântico e semântico-pragmático e habilitarão o sistema a realizar inferências sobre o conteúdo dos textos. Há diversas

informações que podem ser incluídas em uma ontologia, isso depende do domínio a ser representado e do objetivo a que o projeto se propõe.

Algumas vantagens da construção de uma ontologia são: (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.

Para definirmos que informações são úteis ao nosso objetivo, analisamos metodologias empregadas por projetos afins. Escolhemos para apresentar aqui ontologias[2] que têm sido referência para diversos estudos. Podemos classificá-las de duas formas: (i) ontologias independentes de língua (aquelas que não são específicas de nenhuma língua em particular, pois baseiam-se em uma interlíngua) ou não; e (ii) como ontologias de base formal ou de base lingüística.

As ontologias dos projetos Cycorp (Cyc)[3] e Mikrokosmos (MK) são ontologias independentes de língua de base formal. Caracterizam-se por serem ontologias comprometidas fundamentalmente com a Inteligência Artificial (IA), uma vez que têm suas raízes no âmbito de projetos de Ciências da Computação e o objetivo é a implementação de um sistema.

O projeto Cyc propõe-se a construir uma grande base de conhecimento enciclopédico que possa ser aplicada em diferentes sistemas de IA. A base fundamental do Cyc é a lógica, contudo, como o objeto manipulado é a língua, os recursos usados para a criação das regras lógicas são informações previstas em abordagens lingüísticas (fala-se em entidades *agente* e *relações*, por exemplo), mas não há, em nenhum momento, referência a uma análise dessas bases teóricas. Não se questiona o rigor descritivo, uma vez que o conhecimento lingüístico descrito por humanos toma como base a intuição a partir do conhecimento de mundo. O foco é a codificação do maior número possível de informações que possam ser diretamente aplicadas em diferentes sistemas de IA.

O MK, por sua vez, é um projeto de construção de um tradutor automático do Espanhol para o Inglês. No MK, nota-se um cuidado maior na manipulação dos dados lingüísticos. Os pesquisadores deixam evidentes as diferenças entre o estudo das teorias lingüísticas com fins descritivos (lingüística teórica) e o uso das teorias lingüísticas com vistas ao PLN. A equipe responsável pelo MK faz referência aos diferentes níveis de análise lingüística (morfológico, semântico, sintático e pragmático), representando-os no que eles chamam de *microteorias*, o que, na verdade, se trata de "um pedido de licença" às teorias lingüísticas para simplificar e unir diferentes abordagens com a justificativa da adequação à necessidade aplicada, no caso, a construção de um tradutor automático. A ontologia do MK une abordagens ligadas aos papéis temáticos, *frames* e relações semânticas.

---

[2] Ressaltamos que consideraremos projetos de construção de léxicos computacionais e bases de dados lexicais como projetos relacionados a ontologias.

[3] http://www.cyc.com

As *wordnets* (de Princeton[4], a EuroWordNet e a WordNet.Br) e o FrameNet[5] (FN) caracterizam-se tipicamente como ontologias lingüísticas dependentes de língua, desenvolvidas por pesquisadores comprometidos fundamentalmente com a Lingüística e preocupados com o rigor das descrições feitas. A aplicação final objetivada por esses projetos é a construção de uma base de dados formalizável. Vislumbram-se diversas possibilidades de uso desses recursos em ferramentas de PLN; porém, essa aplicação não é direta, uma vez que não é previsto nenhum modelo formal de codificação dos dados lingüísticos ali expressos.

As *wordnets* são bases de dados lexicais cujas arquiteturas foram construídas sob o viés de teorias lingüísticas e psicolingüísticas da memória lexical humana. Os termos encontram-se organizados hierarquicamente em conjuntos de sinônimos (*synsets*) que representam um conceito de uma língua, explicitando as relações de sentido (sinonímia, hiponímia, acarretamento, etc) e relações temáticas (agente, paciente, instrumento), além de informações categoriais e glosas para os synsets e para as relações.

O Berkeley FrameNet é um projeto de Charles J. Fillmore para o Inglês baseado em *frames semânticos* que objetiva documentar possibilidades combinatórias semânticas e sintáticas (valenciais) de cada palavra predicativa (nominais, verbos e adjetivos) em cada um dos seus sentidos. Um *frame semântico* constitui uma representação esquemática que formaliza o resultado das relações sintáticas e semânticas de uma unidade lexical que representa uma situação envolvendo vários participantes, propriedades e outros papéis conceituais que constituem cada elemento de um *frame.*

A partir da análise das informações lingüísticas empregadas pelas ontologias citadas, o critério de seleção das abordagens lingüísticas visa aos seguintes fins: (i) descrever o significado dos verbos em questão, portanto, percorrendo diferentes níveis de análise (semântico, sintático e pragmático) sem perder de vista o rigor teórico; e (ii) formalizar essas descrições em linguagem computacionalmente compatível, no caso, a OWL.

Passemos, então, à discussão dos aspectos de descrição verbal que constituirão nossa ontologia de verbos jurídicos.

3. Abordagens Lingüísticas para a Construção da UNIVERBUE

A UNIVERBUE foi construída com base em três perspectivas para tratar a semântica dos verbos jurídicos: (i) campos semânticos e relações lógico-semânticas, (ii) aspectos gramaticais do significado verbal e (iii) os *frames* semânticos.

3.1 Campos semânticos e relações lógico-semânticas

Lingüistas, psicolingüistas, antropologistas e cientistas da computação têm se utilizado de diferentes concepções de léxico, dependendo dos aspectos da língua que desejam focalizar. Ainda que defendam diferentes pontos de vista, a grande questão que

[4] http://www.cogsci.princeton.edu/cgi-bin/webwn
[5] http://www.icsi.berkeley.edu/~framenet

todos tentam responder é a seguinte: quais e quantas são as relações de sentido significativas? (EVENS, 1988)

Lehrer e Kittay (1981)[6] explicam que uma análise de *campos semânticos* – ou *áreas temáticas*, conforme Borba (1996) – é baseada na concepção de que o significado de uma palavra em um dado campo surge das *relações semânticas* – de similaridade e de contraste – que se estabelecem entre ela e as outras. Uma abordagem relacional aceita a existência de domínios e descreve como os elementos de um domínio estão relacionados a outros. Os nós que os concectam são chamados relações lexicais ou semânticas.

Tomando como base as relações propostas pelas *wordnets*, nosso propósito é identificar um conjunto de relações semânticas que possibilite organizar um léxico do domínio jurídico. A partir dos verbos do *córpus*, foram identificadas as seguintes relações: (a) acarretamento, (b) antonímia, (c) causa, (d) hiponímia e (e) sinonímia. As relações de hiponímia e sinonímia assumiram papel de destaque na descrição semântica dos verbos em questão.

Fellbaum (1998) também chama a atenção para a importância da relação de hiponímia para a construção de léxicos computacionais do tipo *wordnets*, uma vez que é a mais freqüente tanto entre verbos quanto entre nominais. Em linhas gerais, pode-se dizer que a hiponímia é uma relação lexical correspondente à inclusão de uma classe em outra. No caso de nossa ontologia, 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 Português Europeu) com *veículos automotores*.

A relevância da sinonímia em nossa ontologia 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.

3.2 Aspectos Gramaticais do Significado Verbal

Há informações fundamentais veiculadas pelas entidades verbais que só podem ser expressas a partir de um estudo sintagmático das relações de sentido. Informações de ordem sintático-semântica com base em tipos de situações – estados, eventos, ações – e participantes ou papéis semânticos auxiliarão o sistema a interpretar e a construir frases coerentemente. Inserir informações sobre a estrutura argumental de um verbo em um léxico computacional 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 por conceitos de determinado tipo (ex.: humano, animado, abstrato, etc.).

A literatura nos oferece diversas abordagens para a classificação das situações (Vendler, 1967; Dowty, 1979; Van Valin, 1990- 1997; Chafe, 1970; Frawley, 1992; Borba, 1996; Saeed, 1997)[7]. Ainda que não haja consenso quanto à terminologia

[6] Apud Miller e Fellbaum (1991).
[7] Apud Alves (2005).

adotada, todos compartilham de um mesmo pressuposto teórico: a centralidade do verbo.

Para descrever os verbos extraídos do córpus, optamos pela classificação proposta por Borba (1996) por descrever os verbos do Português e por partir de uma análise de córpus bastante completa, tomando como base a Gramática de Valência. O autor identifica quatro classes sintático-semânticas: verbos de ação, de processo, de ação-processo e de estado. Já os papéis semânticos são definidos por ele como noções relacionais que se apresentam como configurações estruturais, com estatuto comparável ao das noções de sujeito e objeto em muitas teorias gramaticais. Seguindo basicamente a proposta de Fillmore (1968), Borba enumera treze casos e procura caracterizá-los a partir de traços semânticos fundamentais.

A sistematização dos papéis semânticos que farão parte da nossa análise inspira-se também na organização proposta por Frawley (1992), que divide os papéis temáticos em papéis semânticos participantes, os quais se subdividem em atores lógicos, receptores lógicos e papéis espaciais, e papéis semânticos não-participantes, que dizem respeito aos papéis opcionais para a predicação. Observe-se o seguinte esquema que sistematiza as noções adotadas.

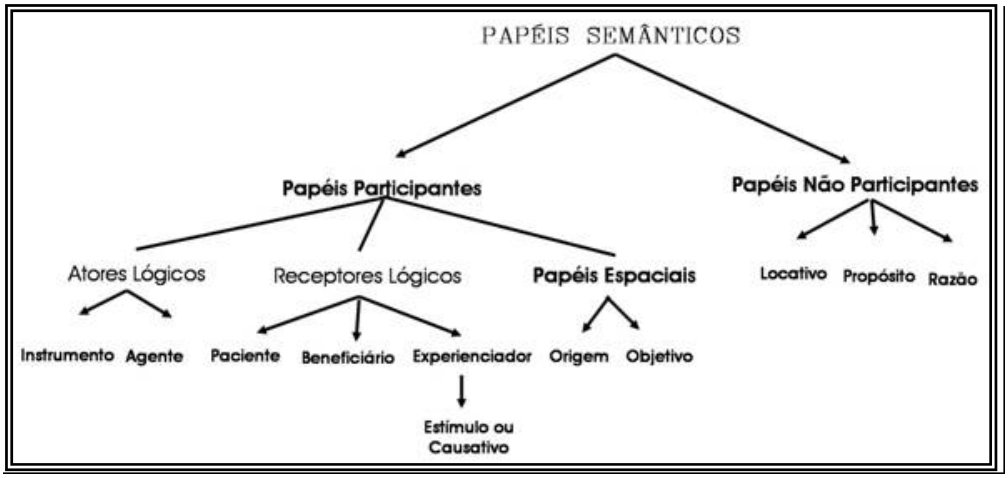

Figure 1. Papéis Semânticos

3.3 Aspectos Contextuais do Significado Verbal

Além das informações sobre as relações semânticas e temáticas, a possibilidade de incluir em uma ontologia informações de base enciclopédica de maneira estruturada deve ser considerada. O recurso que tem sido usado em PLN para inserir tais informações são as estruturas *frame*.

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*. Mas essa informação 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*. Tais informações podem ser fornecidas a partir de uma estruturação da hierarquia do léxico de maneira compatível com as informações

fornecidas nos *frames semânticos* do FN. Uma abordagem baseada em *frames* possibilitará ainda incluirmos 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.

4. A construção da UNIVERBUE: etapa lingüística

A ontologia proposta aqui - a UNIVERBUE - foi construída a partir da análise de um córpus e embasada nas abordagens citadas na seção 3. Com o objetivo de delimitar o córpus, realizamos as seguintes tarefas: (i) *seleção do córpus***,** seis Acórdãos Judiciais homologados no período 2002-2003 sobre o tema "acidentes rodoviários"; (ii) *contextualização do córpus*, estudo do gênero textual e do seu papel nos trâmites jurídicos; (iii) *extração automática dos verbos*, com o auxílio do EXTRACTOR[8]; e (iv) *seleção dos verbos a serem analisados*, de um total de 916 verbos distintos foram selecionados 10 que ocorrem em 99 concordâncias, com base na maior freqüência e relevância para o domínio. Concluída a preparação do *córpus*, passamos para a análise semântica que refere-se à construção da ontologia propriamente dita. Essa análise gerou uma estrutura ontológica que segue o seguinte modelo constituído de quatro níveis de representação: (i) definição, (ii) relações lógico-semânticas, (iii) papéis semânticos e (iv) elementos *frame*. A título de ilustração, apresentamos a estrutura ontológica do verbo *condenar*.

ENTIDADE: *condenar*
Definição: declarar culpado; pronunciar uma sentença a alguém em um tribunal, reconhecendo-o responsável pelo delito, crime ou falta e atribuindo-lhe uma pena (WN – tradução nossa; BORBA (2002) e Dicionário da Língua Portuguesa Contemporânea (2001)
Relações lógico-semânticas: sinônimos, antonímia, hiperonímia, termos coordenados
Papéis semânticos : agente, paciente, objetivo, instrumento, razão
***Frame* semântico**: avaliador, avaliado, meio, base legal, razão, local, condições, tempo, maneira

Figure 2. Estrutura Ontológica *Condenar*

5. A edição da UNIVERBUE no Protégé

Nesta seção apresentaremos uma proposta de como representar no editor Protégé as informações lingüísticas sistematizadas na etapa anterior de forma a possibilitar que os quatro níveis de análise propostos interajam cooperativamente sem que haja sobreposições.

O Protégé toma como base uma abordagem relacional para a inclusão das classes da ontologia e possibilita a especificação de uma *estrutura frame* (*template slot*) para cada verbo onde incluimos as informações ontológicas descritas na seção 3 de maneira integrada, conforme descrevemos nesta seção.

[8] EXTRACTOR é uma ferramenta de extração automática de termos criada em cooperação pela Universidade de Évora e UNISINOS.

5.1 A definição

A polissemia é uma riqueza da língua que acaba se tornando um dos desafios a serem superados em PLN. Embora a definição não seja um recurso previsto em uma ontologia, delimitar o significado focalizado é uma tarefa importante para o construtor da ontologia. Especialmente em se tratando de uma ontologia de domínio específico, pois isso interfere na representação ontológica das unidades lexicais. No Protégé, esse recurso é chamado de *documentação*.

5.2 As relações lógico-conceituais

A arquitetura geral da ferramenta Protégé utiliza como critério primeiro a relação de *hiponímia/hipernímia* entre as classes, suas subclasses e instâncias. Dessa forma, o início da construção da ontologia tratou da inclusão de entidades relacionadas por meio dessa relação. As demais relações identificadas no córpus foram inseridas como *slots* pertencentes da estrutura frame de cada classe já inserida no editor. Foram inseridas inicialmente 10 classes verbais as quais geraram um total de 120.

A inserção dos *slots* do tipo relações lógico-semânticas sofreu uma restrição devido a limitações da ferramenta. O Protégé não permite que sejam definidos *slots* iguais com valores diferentes, ou seja, não é possível inserirmos uma mesma propriedade (relação) que estabeleça ligação entre unidades lexicais diferentes. A solução encontrada foi inserirmos juntamente com o nome da relação o nome do verbo a que ela se refere. Assim temos: sinônimo_condenar (figura), sinônimo_absolver, sinônimo_julgar, etc.

5.3 Os papéis semânticos e os elementos *frame*

Tendo em vista que os papéis semânticos e os elementos *frame* referem-se a entidades semelhantes em uma situação de forma mais ou menos independente da estrutura sintática, o desafio, nessa etapa da construção da ontologia, foi encontrar uma forma de não inserir informações sobrepostas nem perder preciosas informações fornecidas por ambas as abordagens. Optamos por inserir os papéis semânticos e aqueles elementos *frame* que descrevem novas entidades na estrutura *frame* (*template slot*). Os demais elementos frames descritos na estrutura ontológica serviram para auxiliar na organização hierárquica dos nominais relacionados de alguma forma aos verbos que fazem parte da UNIVERBUE. Elementos como *avaliador* e *avaliado* serviram para subespecificar os papéis semânticos *agente* e *paciente*. Isso deu origem à *slots* do tipo: *agente_avaliador* (ex.: juiz) e *agente_avaliado* (ex.: réu). Outra limitação encontrada na versão utilizada do Protégé foi a impossibilidade de fazer o que se pode chamar de *redefinição* de um atributo, como o caso do atributo *sinônimo*. Isso faz com que tenhamos que inserir um novo atributo *sinônimo* para cada classe (ex.: sinônimo_condenar, sinônimo_absolver). Contudo, a grande vantagem do editor Protégé é a possibilidade de conversão automática do conteúdo lingüístico para a linguagem implementacional OWL, isso, obviamente, além de facilitar o próprio armazenamento e gerenciamento dos dados.

6. Uso da ontologia em sistemas de busca de informação

Conforme já anteriormente mencionado, o objetivo maior deste trabalho é construir uma ontologia de verbos jurídicos que possa ser aplicada ao aperfeiçoamento

de sistemas de extração de informação na Internet. A partir da inserção dos dados lingüísticos no editor de ontologias Protégé, as informações foram exportadas para a linguagem implementacional própria para a codificação de ontologias na era da Web Semântica, a OWL. Somente a partir desses dados formais é que a ferramenta de busca e extração de informações na Web fará uso do conhecimento descrito no Protégé. Tendo a ontologia jurídica representada em OWL, é possível aos sistemas de busca e extração de informação na Web usarem esta informação, de modo a melhorarem o seu desempenho e responderem de um modo mais exato às questões dos usuários. Suponhamos o seguinte exemplo de interação de um usuário com o sistema de busca de informação jurídica:

Usuário: Quem julgou o caso do algodão?

Na base de documentos existe uma frase relacionada com esta interrogação:

Documento: O Tribunal da Relação de Évora condenou o réu do caso do algodão ao pagamento de 500 euros.

Como é fácil de observar, um sistema de busca de informação que não possua conhecimento do domínio dificilmente conseguirá extrair a informação adequada. A utilização de ferramentas de PLN (análise sintática e semântica), é possível identificar na interrogação a intenção do usuário de ser informado sobre os agentes responsáveis por uma dada ação: "julgar o caso do algodão". A seguir, e recorrendo à ontologia jurídica desenvolvida no âmbito deste projeto, é identificado o verbo "condenou" como uma especialização (subclasse) de "julgou". Além disso, na ontologia, a ação "condenar" possui como atributos (slots) um agente responsável pela ação e um agente objeto da ação. Assim, utilizando a ontologia jurídica UNIVERBUE, o sistema de busca de informação passa a ter a capacidade de identificar adequadamente o agente responsável pela ação de julgar (ou de uma sua especialização), respondendo corretamente ao usuário: *Tribunal da Relação de Évora*.

Esta abordagem também tem sido adotada no âmbito de outros projetos do Departamento de Informática da Universidade de Évora, sendo, neste momento, possível efetuar automaticamente a análise sintática e semântica parcial de documentos escritos em Português, extraindo a informação relevante e criando instâncias de classes pertencentes a uma dada ontologia do domínio.

7. Conclusão

Este trabalho dedicou-se a descrever a semântica de verbos do domínio jurídico de forma a possibilitar a construção da ontologia UNIVERBUE. Tal ontologia contribui especialmente para o funcionamento mais eficiente de sistemas jurídicos *on-line* de busca e extração de informações que interaja com o usuário através de pergunta e resposta em língua natural. Como trabalho futuro, pretende-se estender a ontologia a outros conceitos (verbos e nominais) a fim de alargar o âmbito da sua utilização nas interações com o sistema de pergunta-resposta jurídico.

A análise do córpus evidenciou, fundamentalmente, que organizar uma ontologia com base em uma abordagem comprometida unicamente com a Semântica Lexical e a Semântica Lógica (relações lógico-conceituais) não é suficiente para a construção de uma ontologia que auxilie um sistema de extração de informações, uma vez que o sistema terá que interpretar e gerar língua natural estabelecendo interações

com o usuário através de pergunta e resposta. A abordagem adotada para a representação do conhecimento ontológico dos verbos permitiu ampliarmos o limite da nossa ontologia. O objetivo inicial era a descrição da semântica verbal; entretanto, de maneira indireta, acabamos por organizar informações semânticas dos nominais relacionados aos verbos do córpus. O que possibilitou essa expansão de resultado foi a inclusão de abordagens com as quais a Semântica faz interface: a Sintaxe e a Pragmática. A construção da UNIVERBUE partiu de seis Acórdãos Judiciais da PGR-PT; daí foram extraídas 359 ocorrências verbais diferentes e selecionados 10 verbos. Ao final da etapa apresentada aqui, estamos com 120 e 74 entidades não verbais.

Finalmente, gostaríamos de acrescentar que, além das referidas contribuições teóricas, este trabalho mostra os resultados de um trabalho bem sucedido integrado entre profissionais da Lingüística e da Computação. Com isso gostaríamos de enfatizar a importância de uma postura cooperativa entre profissionais dessas áreas para a construção de sistemas de PLN.

7. Referências

Alves, I. M. R. A. (2005). "O Uso da Semântica Verbal em Sistemas de Extração de Informação: A Construção de uma Ontolgia de Domínio Jurídico". Dissertação (Mestrado em Lingüística Aplicada) – Universidade do Vale do Rio dos Sinos (UNISINOS), São Leopoldo. 287 p.

Borba, F. S. (1996). "Uma Gramática de Valências para o Português". São Paulo: Editora Ática.

Borba, F. S. (2002). "Dicionários de Usos do Português do Brasil". São Paulo: Ed. Ática.

"Dicionário da Língua Portuguesa Contemporânea – Verbo". (2001). Vol. 1 e 2, Editorial Verbo e Academia das Ciências de Lisboa.

Evens, M. W. (1988). "Relational Models of the Lexicon". Cambridge: Cambridge University Press.

Fellbaum, C. (1998). A Semantic Network of English Verbs. In: Fellbaum, Christiane. "WordNet: An Electronic Lexical Database". Cambridge: MIT Press.

Fillmore, C.J. (1968). The Case for Case. In.: Bach and Harms (Ed.): "Universals in Linguistic Theory", Holt, Rinehart, and Winston, New York.

Frawley, W. (1992). "Linguistic Semantic". London: Lawrence Erlbaum Associates, Publishers.

Miller, G. A. *e* Fellbaum, C. (1991). Semantic Networks of English. In:. Levin e S. Pinker (eds), "Lexical and Conceptual Semantics". Cambridge, MA: Blackwell.