

A semântica dos adjetivos e os sistemas de extração de informação na *web*

Anderson Bertoldi*
UNISINOS

Rove Luiza de Oliveira Chishman**
UNISINOS



1 Introdução

Este artigo apresenta um estudo preliminar sobre a semântica dos adjetivos com vistas à sua inclusão em uma base de dados lexicais de um sistema de busca e extração de informação na *web*. Vinculado a um projeto mais amplo, que reúne uma equipe interdisciplinar integrando pesquisadores da Linguística e da Informática,¹ este trabalho objetiva ilustrar como uma aplicação computacional pode se beneficiar de uma descrição linguística teoricamente bem fundamentada. Dando seguimento à tarefa de construção de uma ontologia que melhore o funcionamento da ferramenta de busca do *site* da Procuradoria Geral da República de Portugal, ocupamo-nos aqui da inclusão dos adjetivos. É importante ressaltar que uma proposta de representação da semântica dos verbos² já foi feita e parte dos nominais também.

* Mestrando no PPG Linguística Aplicada da UNISINOS.

** Professora do PPG em Linguística Aplicada da UNISINOS.

¹ Este estudo está vinculado ao projeto DIRPI (Desenvolvimento e Integração de Recursos para Pesquisa de Informação), de cooperação internacional entre Brasil e Portugal. Participam do projeto a UNISINOS (Professoras Rove Chishman e Renata Vieira) e Universidade de Évora (Prof. Paulo Quaresma).

² Ver dissertação de mestrado de Isa Mara da Rosa Alves (2005).

Em termos gerais, nosso desafio é avaliar o papel que os adjetivos podem desempenhar nos léxicos computacionais. Em termos mais específicos, pensando nos sistemas de busca e extração de informação, muitos estudos têm sido realizados sobre a semântica dos verbos e dos nominais. Como uma área de menor interesse, os adjetivos aparecem em desvantagem perante seus “primos ricos”. A questão consiste em definir que distinções referentes à semântica adjetival devem ser representadas na ontologia. Questões como a função dos adjetivos, a correlação entre a sintaxe e a semântica dos adjetivos, a natureza dos adjetivos relativos e a graduação adjetival devem ser levadas em conta.

Neste artigo, concentramo-nos no que podemos denominar de “etapa lingüística” da pesquisa, que consiste em avaliar, dentre as abordagens teóricas, a ou as que mais se ajustam à nossa demanda.

Para ilustrar a importância das questões acima abordadas, serão feitos um estudo da semântica lexical dos adjetivos e um exercício tomando como base seis acórdãos do Instituto das Tecnologias de Informação na Justiça de Portugal, visando demonstrar a importância da representação da semântica lexical dos adjetivos para PLN. Com isso, o presente trabalho objetiva demonstrar como o conhecimento do conteúdo semântico dos adjetivos ajuda a organizar a informação de textos de um domínio lingüístico específico, neste caso, o jurídico.

O tratamento do adjetivo para o processamento da linguagem natural pode não ser tão simples quanto o do verbo ou do substantivo. Essa dificuldade se deve principalmente à forma como se vai tratar a polissemia dessa classe gramatical. Se considerarmos o adjetivo como um modificador do nome, o seu significado se relaciona diretamente com nome que ele está modificando, ou seja, o significado do adjetivo é composto juntamente com o do substantivo. Assim sendo, uma abordagem composicional pode ser uma boa opção para a descrição da semântica dos adjetivos, como veremos adiante.

Em termos sintáticos, os adjetivos são divididos em predicativos e atributivos ou não-predicativos. Assim como os verbos, eles podem ou não receber complementos. Em termos semânticos, os adjetivos podem ser divididos em intersectivos ou absolutos, não-intersectivos e relativos. Outra característica importante que diferencia os adjetivos das outras classes gramaticais é a graduação. Um adjetivo pode ser gradativo ou não. Faremos, na seção seguinte, uma revisão de como a literatura tem tratado tradicionalmente os adjetivos. Porém, para aplicações computacionais essas divisões não dão conta da polissemia, uma das maiores dificuldades na formalização das línguas naturais em léxicos computacionais. Então, como

podemos representar o significado das palavras de uma forma econômica para que o computador possa trabalhar de forma eficiente e que dê conta da polissemia dos adjetivos?

A forma tradicional de se representar o significado das palavras é a enumeração de sentidos. Pustejovsky (1995) chama essas representações de Léxicos de Enumeração de Sentidos. Para exemplificar, consideremos o adjetivo *bom*. Conforme mencionado acima, o significado do adjetivo em posição atributiva é composto juntamente com o significado do adjetivo que o acompanha. Dessa forma, *uma boa faca* é uma faca que funciona bem, *um bom aluno* é um aluno que tem bom comportamento ou que tem bom rendimento, *uma casa boa* é uma casa que é confortável ou que tem uma estrutura forte. Assim, em um Léxico de Enumeração de Sentidos, como é o caso do dicionário, o número de sentidos de uma palavra como *bom* será tão grande quanto forem as possibilidades de interpretação dos diferentes contextos em que a palavra aparecer. Tal léxico não é um modelo de representação semântica útil para o processamento da linguagem natural, pois não é econômico.

Assim sendo, um modelo de representação semântica para os adjetivos deve levar em conta as informações tanto do adjetivo, quanto do substantivo que ele modifica, de uma forma econômica e que possa ser lido pelo computador. Para atingir esse objetivo, abordarei os adjetivos do domínio jurídico seguindo uma perspectiva lexical, conforme Raskin e Nirenburg (1995) e Bouillon e Viegas (1999). A noção de *co-composicionalidade*, segundo Pustejovsky (1995), é o ponto central de ligação entre o substantivo e o adjetivo que pretendo desenvolver neste trabalho, mostrando como o significado dos itens lexicais são composicionais, não enumerativos.

2 A Semântica Lexical dos Adjetivos

A descrição da semântica lexical dos adjetivos para fins computacionais deve se prender à representação daquelas características dos adjetivos que trarão benefícios para a aplicação que se tem em mente. Isso significa dizer que uma perspectiva teórica pode não se adequar à aplicação em desenvolvimento ou que talvez se tenha que recorrer a uma combinação de abordagens. Nirenburg & Raskin denominam esta visão de *demand-side lexical semantics*, em oposição à *syntax-driven lexical semantics*. Mas o que é importante saber sobre a semântica dos adjetivos quando falamos em aplicações computacionais? A resposta para essa pergunta ainda não é muito clara. Apresentaremos abaixo uma relação de cinco aspectos sintático-semânticos que são uma tentativa de esclarecer um pouco mais essa questão.

2.1 A posição do adjetivo

Do ponto de vista sintático, o adjetivo pode ser atributivo ou predicativo, dependendo da posição que ocupar dentro da sentença (Cruse, 2000; Bouillon e Viegas, 1999; Pustejovsky, 1995). Um adjetivo está em posição atributiva quando se liga de forma direta a um nome. Quando o adjetivo é acompanhado de verbo copular, ocupa a posição predicativa. Na primeira, o adjetivo compõe um sintagma nominal, ou seja, é periférico com relação ao núcleo do sintagma; na segunda, ocupa o núcleo de um sintagma verbal.

- (1) um homem alto (**posição atributiva**)
- (2) o homem é alto (**posição predicativa**)

Há adjetivos que podem ocupar ambas as posições, outros podem ocupar ou uma ou outra posição. Bouillon e Viegas (1999) subdividem os adjetivos atributivos e predicativos em três categorias, a saber: *somente atributivos*, *somente predicativos* e *centrais*.

- (3) *afraid people, people are afraid (**somente predicativo**)
- (4) the atomic scientist, * the scientist is atomic (**somente atributivo**)
- (5) the tall man, the man is tall (**central**)

Conforme vemos, os exemplos de Bouillon e Viegas (1999) estão em inglês. Essa tripartição dos adjetivos atributivos e predicativos, ao que tudo indica, não é normal em língua portuguesa. Segundo Borba (1996), qualquer adjetivo se coloca em posição atributiva, mas nem todo adjetivo ocupa posição predicativa (ex.: câmara municipal/*a câmara é municipal), ou seja, em língua portuguesa não teríamos, como temos na língua inglesa, adjetivos somente predicativos. Também é importante observar a posição que o adjetivo ocupa dentro do sintagma nominal. Na língua portuguesa a posição que o adjetivo atributivo comumente ocupa é a pós-nominal. Porém, em alguns casos o adjetivo só pode ocorrer em posição pré-nominal, em outros apenas na posição pós-nominal:

- (6) a bela casa, a casa bela (posição **pré-nominal** e **pós-nominal**)
- (7) o suposto criminoso, *o criminoso suposto (apenas posição **pré-nominal**)
- (8) *o premeditado crime, o crime premeditado (apenas posição **pós-nominal**)

2.2 As classes semânticas do adjetivo

As formas de se classificar os adjetivos varia muito de autor para autor. Tanto a nomenclatura utilizada pelos diversos teóricos, quanto os fatos por eles classificados divergem. Chierchia e Mc-Connel-Ginet (1990, p. 458-66) dividem os adjetivos, segundo suas características semânticas, em *intersectivos*, *subsectivos* e *não-predicativos*. Os exemplos apresentados pelos autores já são clássicos:

- (9) a. Pavarotti is a pink tadpole.
b. Pavarotti is pink.
c. Pavarotti is a tadpole.
- (10) a. Pavarotti is a large tadpole.
b. Pavarotti is large.
c. Pavarotti is a tadpole.
- (11) a. Pavarotti is a former tadpole.
b. *Pavarotti is former.
c. Pavarotti is a tadpole.
d. Pavarotti was a tadpole.

Como vemos em (9), (9a) acarreta tanto (9b), quanto (9c). Esses adjetivos são chamados de *intersectivos*. Já em (10), (10a) não acarreta (10b), mas acarreta (10c). Esses adjetivos são chamados de *subsectivos*. Por último, (11a) não acarreta nem (11b), nem (11c), mas acarreta (11d). Esses são chamados de *não-predicativos*.

O que percebemos em (9b) e (10b) é que *pink* e *large* podem assumir a posição predicativa, porém, o fato de Pavarotti ser um girino grande não acarreta a verdade de ele ser um animal grande. O adjetivo *pink* acarreta o fato de Pavarotti ser um animal rosa. O adjetivo *former* (*antigo*), por sua vez, não pode ser empregado em posição predicativa e o fato de Pavarotti ser um antigo girino não acarreta a verdade de que ele é um girino, mas que foi um girino.

Cruse (2000) classifica os adjetivos em *absolutos* e *relativos*. Para diferenciar essas duas classes Cruse sugere: “um teste simples para esta distinção é a seguinte: se Adj. + X (sempre) acarretar Adj. + Y, sendo que X é um hipônimo de Y, então o adjetivo é absoluto; se há casos claros em que o acarretamento falha, então o adjetivo é relativo (p. 290).³ Assim, *um rato preto* é um *animal preto*, dessa forma *preto* é um adjetivo *absoluto*, já *um rato grande* não é um *animal grande*, então *grande* é um adjetivo *relativo*.

³ Tradução minha.

Para Borba (1996), os adjetivos se dividem em *qualificadores* (QL) e *classificadores* (CL). Segundo o autor, a posição predicativa, geralmente, é vinculada aos adjetivos QL e a posição adnominal (atributiva) aos adjetivos CL. As relações que ligam adjetivos QL e os nomes são internas porque apresentam características essenciais ou acidentais da natureza de um nome, com se constituíssem um traço dele (menino *maroto*). Porém, as relações estabelecidas entre os nomes e os adjetivos CL são externas porque apenas colocam o nome numa determinada classe (parque *municipal*).

Borba (1996) considera QL como *descritivos* e CL como *definitórios*. Os adjetivos QL atribuem ao nome uma qualificação que está ligada ao julgamento pessoal. Por outro lado, os adjetivos CL representam o nosso modo de classificar entidades no mundo. Se somente os QL expressam opinião pessoal, somente eles podem receber gradação. Borba ressalta que alguns autores consideram os QL como verdadeiros adjetivos e os CL como pseudo-adjetivos.

Os adjetivos CL não podem assumir posição predicativa:

(12) a crise *ministerial* > a crise é *ministerial*

Somente em algumas situações especiais é que eles podem assumir a posição predicativa: "(i) com quantificados – (...) Algumas escolas são rurais; (ii) em construções contrastivas - Estas viaturas são municipais; aquelas, não; (...); (iii) com repetição do núcleo do sintagma nominal – (...) Estas flores são flores campestres (p. 179)".

Borba ainda afirma que muitos adjetivos classificadores podem ser usados como qualificadores, "basta que se tome o valor semântico básico do adjetivo como um traço, uma característica, isto é um atributo ou propriedade" (p. 181). Para se identificar um adjetivo CL exercendo função de CL é só aplicar a gradação e/ ou a posição predicativa. Conforme exemplo do próprio autor:

(13) política petista > Erundina é muito petista).

Raskin e Nirenburg (1995) classificam os adjetivos, segundo as suas características semânticas, em *relacionais* e *qualitativos*. Apesar de dividi-los em duas classes principais, os autores reconhecem a existência de uma terceira classe mista de adjetivos *relacionais/qualitativos*. Assim como Borba (1996) relaciona os adjetivos qualificativos à posição predicativa e os adjetivos classificadores à posição atributiva, Raskin e Nirenburg (1995) também relacionam os adjetivos relativos à posição atributiva e os adjetivos qualitativos à posição predicativa. Assim sendo, considerarei adjetivos *relativos* e *qualitativos* como sinônimo de *classificadores* e *qualificadores*.

Os adjetivos relacionais geralmente são identificados como adjetivos denominais (Borba, 1996; McNally e Boleda, 2004; Raskin e Nirenburg, 1995; Raskin e Nirenburg, 1996). Isso significa dizer que eles são derivados de nominais, talvez esse seja o motivo pelo qual muitos adjetivos relativos não podem ser transformados em substantivos:

- (14) a reforma *ministerial* > a reforma do *ministério*
- (15) a reforma *ministerial* > *a *ministerialidade* da reforma.

Judith Levi (*apud* Raskin e Nirenburg, 1995: 11-13) também considera os adjetivos relativos denominais. Ela aponta seis características que os adjetivos não-predicativos compartilham com os nominais, que são descritas abaixo.

Diferentemente dos adjetivos qualitativos, os adjetivos relativos não aceitam gradação:

- (16) a. *esta é uma reforma *muito ministerial*

Também não é possível unir na mesma oração, através de conjunção *e*, um adjetivo relativo e outro qualitativo:

- (16) b. *este é um engenheiro *civil e bom*.

Os adjetivos relativos só aceitam quantificação na forma de prefixos como *bi*, *tri*, *multi*, *pluri*, etc.:

- (16) c. esta é uma reunião *pluripartidária*.

Os adjetivos relativos, sendo denominais em sua maioria, herdam dos nomes as suas características semânticas:

- (16) d. o *pai* ama a filha (+ masculino) > o amor *paterno* (+masculino)

Esses adjetivos também herdam o papel temático (caso) do nome que lhe deu origem:

- (16) e. a reforma que o *ministério* fez (agentivo) > a reforma *ministerial* (agentivo)

Por último, como já foi exposto acima, os adjetivos relativos não admitem nominalização:

- (16) f. a política *partidária* > *o *partidarismo* da política

As relações entre os adjetivos relativos (denominais) e os nomes a que eles se referem podem ser de naturezas diversas. Segundo Raskin e Nirenburg (1996), existem relações do tipo *pertence-a*, *possuído-por*, *tem-como-parte*, *locação* etc.:

- (17) procedimentos *cirúrgicos* (**pertence-a cirurgia**)
- (18) viatura *policial* (**possuída-pela polícia**)
- (19) células *malignas* (câncer tem-como-parte células malignas)
- (20) viagens *espaciais* (localizadas no espaço sideral)

2.3 A Gradabilidade

Tradicionalmente, o fenômeno da gradabilidade é considerado como uma das principais características dos adjetivos, mas, na verdade, ele só atinge uma parte dos adjetivos, a saber: os qualitativos, ou qualificadores. Isso é possível pois esses adjetivos expressam, segundo Borba (1996), uma apreciação, um julgamento ou avaliação do mundo. Porém, nem todos os adjetivos qualitativos são graduáveis, nem intensificáveis:

- (21) *João é muito *solteiro*
- (22) *Maria está meio *viva*

Essa diferença entre adjetivos qualitativos graduáveis e não-graduáveis demonstra estar relacionada com as relações de *oposição*. Lyons (*apud* Raskin e Nirenburg, 1995) divide os termos antônimos em duas classes. Àqueles adjetivos que compõem uma escala gradual entre dois termos, tais como *bom/mau*, ele chamou de *contrários*. Já àqueles que não são escalares, ou seja, não existe meio termo, ou se está em um lado, ou se está em outro, Lyons chamou de *contraditórios*. São exemplos de termos contraditórios *vivo/morto*, *casado/solteiro* etc.

- (23) João é *muito ruim* / João não é *tão ruim* / João é *bom* (**contrários**)
- (24) João é *melhor* / *pior* que Paulo
- (25) *João está *muito vivo* / *João não está *tão vivo assim* / *João está *muito morto* (**contraditórios**)
- (26) *João está *mais vivo* / *mais morto* que Paulo

Assim, vemos que somente os adjetivos qualitativos contrários é que podem receber graduação e intensificação. Considerar esses dois fenômenos como característico da classe dos adjetivos é ser muito abrangente. Na realidade, os adjetivos contraditórios demonstram opiniões, são avaliativos.

2.4 A Valência Adjetival

A valência sintática define a quantidade de complementos que um adjetivo necessita. Pustejovsky (1995) divide os adjetivos em *unários* e *binários*. Os adjetivos unários são aqueles que não necessita de complemento, como *verde*:

(27) a folha *verde*

Já os adjetivos binários necessitam de um complemento. É o caso de *cheio* em:

(28) uma menina *cheia* de vida

Borba (1996) trata a valência sintática de forma mais ampla que Pustejovsky. Segundo Borba, os adjetivos podem ser monovalentes, divalentes, trivalente ou tetravalente, dependendo da quantidade de argumentos que necessitar. Assim sendo:

- (29) a. uma casa *grande* – **monovalente**
b. um carro *cheio* de gente – **divalente**
c. um capataz *doador* de sangue aos peões (cf. Borba, 1996:172) – **trivalente**
d. uma carga *transportável* do navio para os armazéns por guindaste (cf. Borba, 1996:172) – **tetravalente**

A valência semântica define qual será a natureza dos nominais que acompanharão o adjetivo:

(30) uma faca *cortante* – (faca – instrumento)

(31) um aluno *brincalhão* – (aluno – agentivo)

Mas a valência semântica não diz respeito apenas aos nominais que acompanham o adjetivo. Em caso de adjetivos relativos, como *governamental*, é possível descobrir qual é a relação subjacente entre o adjetivo denominal e o nome de que ele provém:

(32) a decisão *governamental* > o governo decidiu > **agentivo**

(33) a ajuda *governamental* > o governo ajudou > **agentivo**

(34) a ajuda *governamental* > a veio do governo > **fonte**

O conhecimento da valência semântica é extremamente importante para o processamento de linguagem natural, especialmente se esse processamento visa à extração de informações. No caso do trabalho em desenvolvimento, a incorporação dessas informações semânticas ao corpus pode ajudar a encontrarmos o agente das ações de forma mais rápida. Mas qual é a extensão dessas informações e a sua frequência dentro de um corpus é o que veremos adiante.

2.5 A Composicionalidade e Polissemia

A polissemia do léxico é uma das maiores dificuldades para o tratamento automático da língua. A abordagem mais comum até agora para tratar da polissemia, em especial dos adjetivos, conforme objetivos deste trabalho, tem sido a enumeração de sentidos. Desta forma, um adjetivo como *bom* teria uma quantidade infinita de interpretações, conforme fosse a sua capacidade de se juntar a novas palavras. Porém, essa quantidade imensa de sentidos diferentes só compromete o bom funcionamento dos programas computacionais.

Como uma alternativa para o tratamento da polissemia, Pustejovsky (1995) apresenta com detalhes a teoria do Léxico Gerativo, que não apresentarei aqui, mas me interessa uma de suas principais premissas, a questão da *composicionalidade*. Se tratarmos de um adjetivo como *bom* através de uma abordagem enumerativa, teremos como exemplo:

- (35) uma *boa* faca > 1º sentido: que corta bem
- (36) um *bom* relógio > 2º sentido: que funciona bem
- (37) uma *boa* pessoa > 3º sentido: que é amável, prestativa

Um léxico de enumeração de sentidos pode dar a falsa idéia de que certos adjetivos são polissêmicos, ou pior ainda, tem uma quantidade infinita de sentidos. Mas, se considerarmos a capacidade dos adjetivos de comporem sentidos a partir das informações disponibilizadas pelos nominais, teremos um número reduzido de entradas lexicais e, conseqüentemente, os programas computacionais podem trabalhar com mais agilidade.

Assim sendo, as informações que possuímos sobre *faca* nos dizem que ele serve para cortar. Portanto, interpretamos o sentido de *boa* conforme a função atribuída a uma *faca*. O nosso conhecimento sobre relógios nos diz que um *relógio* é um aparelho cuja função é marcar a hora. Da função de *relógio* vem o sentido de *bom*. Em *uma boa pessoa* remeterá a todas as qualidades positivas que costumamos atribuir a uma pessoa. Na verdade, *bom* não é um termo polissêmico em si mesmo, mas a sua composição com diferentes nominais o torna polissêmico.

A idéia de composicionalidade também é importante para a descrição da semântica de adjetivos que estão relacionados a eventos e não á características de um indivíduo. Esse é o caso do adjetivo *rápido*:

- (38) um datilógrafo *rápido* (cf. Pustejovsky, 1995, p. 44)

Nesse caso, *rápido* não se liga diretamente a *datilógrafo*, mas sim a sua função, que é datilografar. A sentença acima, pode ser interpretada da seguinte maneira: “um datilógrafo que datilografa *rapidamente*”. Isso significa dizer que esses adjetivos são representam propriedades dos indivíduos, mas sim das ações desses indivíduos.

Acabo de apresentar os pontos que considero, até o momento, mais importantes para o tratamento informatizado dos adjetivos. Na seção seguinte essas teorias serão empregadas na descrição dos adjetivos do domínio jurídico. Dessa forma, tentarei demonstrar quais os fatos relevantes a serem considerados na descrição desses adjetivos para fins computacionais.

3 Os Adjetivos no Domínio Jurídico: Um Estudo Aplicado

A revisão teórica que empreendemos na seção 2 converge para o estudo aplicado que será apresentado nesta seção. Para tanto, iniciaremos explicitando a forma como obtivemos o corpus analisado e, em seguida, a análise semântica desse corpus.

3.1 Pré-análise do corpus: questões metodológicas

O corpus de nosso experimento é composto por seis textos jurídicos disponibilizados eletronicamente pelo Instituto das Tecnologias de Informação na Justiça de Portugal.⁴ Esses textos pertencem ao gênero acórdão e não foram escolhidos ao acaso. Os mesmos seis acórdãos em questão já foram usados por Alves (2005), que tratou das informações verbais necessárias para a extração de informações via *web*. A nossa escolha pelos mesmos textos representa uma continuação do trabalho citado anteriormente, agora com o objetivo de estudar a importância dos nominais e dos adjetivos para a extração de informações.⁵

A extração dos adjetivos foi feita de forma manual. Procuramos retirar a maior quantidade possível de adjetivos dos textos analisados. Adjetivos que compunham sintagmas preposicionados, com função adverbial não foram considerados, é o caso da expressão *em especial*. Siglas, como CC, que significa *Código Civil*, também não foram contabilizadas. Devido à grande quantidade de numerais dos textos analisados, essa classe não foi considerada como adjetivo em nenhuma situação. Por último, procuramos não considerar os participípios como adjetivos naqueles casos em que eles podiam ser facilmente substituí-

⁴ www.dgsi.pt

⁵ O projeto SEMANTEC, coordenado por Rove Chishman engloba o estudo dos nominais e dos adjetivos. Esse projeto é um desdobramento de outro anterior, ONTO-VERB, que tratou da semântica dos verbos.

dos por uma oração desenvolvida de participio. Porém, em alguns casos, o participio foi mantido. É o caso de *amargurada*, em *a autora vive amargurada e triste*. Nesse caso, a coordenação de um participio com um adjetivo prototípico reforça a idéia de que esse adjetivo não constitui uma oração reduzida de participio no caso citado.

Os adjetivos extraídos foram divididos em duas categorias, de acordo com a sua posição, a saber: adjetivos em posição atributiva e adjetivos em posição predicativa. Para evitar que algum adjetivo fosse privilegiado, os cabeçalhos dos acórdãos e os títulos não foram utilizados na extração. Esse procedimento nos rendeu 354 ocorrências adjetivais, de um total de 8.368 palavras, assim divididas: 319 adjetivos em posição atributiva e 35 adjetivos em posição predicativa, sem excluir adjetivos repetidos.

Se excluirmos os adjetivos repetidos de cada uma das categorias temos: 165 adjetivos em posição atributiva e 23 adjetivos em posição predicativa. Isso demonstra a importância que os adjetivos em posição atributiva têm dentro dos textos analisados. Ao todo foram 184 ocorrências únicas em forma canônica, excluindo aqueles adjetivos que se repetiram ou na posição atributiva, ou na predicativa. Isso significa dizer que 4 adjetivos, na sua forma canônica, ocorreram tanto na posição atributiva, como na posição predicativa.

	repetidos	não repetidos
adjetivos em posição atributiva	319	165
adjetivos em posição predicativa	35	23
total de adjetivos	354	188
total de ocorrências canônicas		184
total de palavras	8.368	

A partir dos dados obtidos anteriormente, selecionamos os adjetivos a serem descritos. Devido à grande quantidade de adjetivos em posição atributiva em contraposição aos adjetivos em posição predicativa e às informações que esses adjetivos carregam sobre os nominais com os quais eles se relacionam, optamos por estudá-los neste trabalho. O critério utilizado para a seleção dos adjetivos que analisaremos na subseção seguinte foi o da maior frequência por ocorrência canônica. Assim sendo, surgiu o seguinte grupo de 5 adjetivos: *civil* (18 ocorrências, em 5 acórdãos), *legal* (13 ocorrências, em 5 acórdãos), *não patrimonial* (9 ocorrências, em 3 acórdãos), *presente* (7 ocorrências, em 4 acórdãos) e *próprio* (7 ocorrências, em 3 acórdãos).

Concluída a etapa de pré-análise do corpus, passamos agora para a aplicação da teoria aos adjetivos acima relacionados.

3.2 A análise semântica do corpus

Realizada a revisão bibliográfica na seção 2, tomaremos como base deste experimento a divisão apresentada por Borba (1996). Borba classifica os adjetivos em qualificadores (QL) e classificadores (CL). Os adjetivos qualificadores expressam julgamentos pessoais, já os adjetivos classificadores, como o nome diz, classificam os nominais que acompanham, colocando-os em uma categoria.

O adjetivo *civil* ocorreu 18 vezes, em 5 contextos diferentes: 11 vezes em *responsabilidade civil*, 3 vezes em *código civil*, 2 vezes em *código de processo civil*, 1 vez em *responsável civil* e 1 em *construção civil*. Esse adjetivo é classificador. O significado de *civil* se modifica nos diversos contextos em que é empregado, conforme o princípio da composicionalidade (Pustejovsky, 1995). Ao associar-se com *responsabilidade* o adjetivo *civil* significa “a responsabilidade que cabe ao cidadão”, ou seja, indica a fonte da responsabilidade. Em *código civil*, *civil* indica o objetivo, a meta a ser atingida, isto é, o cidadão, que deve respeitar as regras desse código em suas atividades sociais. Em *código de processo civil*, temos a mesma situação descrita em *código civil*, já que se trata de uma meta a ser atingida: a modulação do comportamento dos cidadãos. Em *responsável civil*, há uma relação de responsabilidade, alguém que pode exigir direitos ou ser responsabilizado, ou seja, pode ser agente ou paciente das ações tomadas em um processo. Por último, *construção civil* indica a finalidade da construção, que se destina ao benefício da população civil, em contraposição, por exemplo, à construção naval que seria a construção de barcos. (?)

O adjetivo *legal* ocorreu em 6 contextos diferentes, totalizando 13 ocorrências: 6 vezes em *taxa legal*, 2 vezes em *juros legais*, 2 em *presunção legal*, 1 em *regime legal*, 1 em *diplomas legais* e 1 em *vistos legais*. Tanto *taxa legal*, quanto *juro legal* significam “que é permitido por lei”. Em ambos os casos, *legal* indica a fonte do direito que se está requerendo, em outras palavras, o direito de se cobrar estas taxas é garantido na lei. Já em *presunção legal* o significado subjacente é “a presunção que foi tida pelos representantes da lei”, aqui indicando uma ação. Em *diplomas legais*, o sentido é “que é expedido legalmente/com o apoio da lei”. Nesse caso, devemos considerar quem, ou qual entidade, emite esses diplomas legais como o agente da ação, a finalidade desses diplomas e os seus beneficiários. *Vistos legais* significa “vistos que foram dados por representantes da lei”, nesse caso, o agente da ação. Todas essas informações que os adjetivos classificadores trazem, como temos visto até agora, podem ser utilizadas para melhorar a busca e extração de informações.

Em terceiro lugar, está o adjetivo *não patrimonial*, num total de 9 ocorrências em apenas 1 contexto: *danos não patrimoniais*. Aqui não separamos o advérbio de negação do adjetivo, pois o termo *não patrimonial* se opõe a outro adjetivo, *patrimonial*. Dessa forma, se retirássemos o *não* teríamos um significado não correspondente com o que se encontrava nos textos pesquisados. Nesse caso, a idéia expressa pelo nominal indica uma transformação negativa. O termo *não patrimonial* indica que o receptor dessa transformação negativa não é algo concreto.

Em quarto lugar, o adjetivo *presente* ocorreu em diversos contextos, porém o que mais chama a atenção é que, dentre 7 ocorrências, 6 apresentaram adjetivo em posição pré-nominal. *Presente* é mais um adjetivo classificador. A sua propriedade é demonstrar a temporalidade de alguma entidade, como no caso de *presente data*, em que há a idéia de um percurso no tempo, do passado até o presente. Ou então, *presente* aparece com termos anafóricos que remetem à lembrança visual de alguma entidade que é a representação material de algum ato imaterial. São os casos de *presente revista* e *presente decisão*, em que ação remete, na verdade, ao que está posto no papel.

Por último, o adjetivo *próprio* ocorreu 7 vezes. Assim como o adjetivo *presente*, também apresentou uma interessante variação de posição. Por exemplo, é estranha a ordem *o próprio risco da responsabilidade*, mas não *o risco próprio da responsabilidade*. Já *o legislador próprio* não soa bem, enquanto *o próprio legislador* é plenamente aceitável. Ao que tudo parece indicar, quando *próprio* está em posição pré-nominal ele indica uma certa hierarquia subjacente nas relações sociais. Quando em posição pós-nominal, indica uma característica intrínseca à entidade com que se relaciona.

Como vimos, todos os adjetivos de maior frequência, dentro de nosso experimento, foram adjetivos classificadores. Isso demonstrar uma necessidade de classificação das entidades desde o tempo até os participantes envolvidos no processo jurídico. Nesse sentido, se pudermos classificar esses adjetivos indicando quem são esses participantes, poderemos melhorar a extração de informações via *web*.

4 Considerações finais

Os objetivos que guiaram este trabalho foram: (i) estudar a semântica lexical dos adjetivos; (ii) aplicar essas teorias em um corpus do domínio jurídico e (iii) verificar a utilidade dos fatos descritos para o processamento da linguagem natural.

Apesar de a análise poder se aprofundar muito mais, as descrição exaustiva não se prestaria a um primeiro estudo do assunto. É importante ressaltar que este estudo trata-se apenas de um primeiro contato com o tema, que será desenvolvido em minha pesquisa de mestrado.

Embora muitos dos fatos lingüísticos apresentados no estudo teórico não terem sido abordados no estudo aplica, isso não significa que eles não sejam importantes. Um corpus mais abrangente pode apresentar uma diversidade maior de fatos a serem descritos e compreendidos, por isso é importante não excluí-los de um estudo teórico. Alguns fatos, como a complementação sintática e a gradação, por aparecer em pequena quantidade, não foram tratadas aqui.

Uma questão ainda fica em aberto para os próximos estudos: até onde vai o significado do adjetivo e onde começa o significado dos nominais? Isso é importante para se entender melhor a questão da composicionalidade. Assim sendo, um estudo mais aprofundado dos nominais é necessário.

Referências

ALVES, Isa Mara da Rosa. O Uso da Semântica Verbal em Sistemas de Extração de Informação: a Construção de uma Ontologia de Domínio Jurídico. Porto Alegre: PUCRS, 2005. Dissertação de Mestrado. Programa de Pós-Graduação em Lingüística Aplicada.

BORBA, Francisco da Silva. Uma gramática de valências para o português. São Paulo: Ática, 1996.

BOUILLON, Pierrette e VIEGAS, Evelyne. *The description of adjectives for natural language processing: theoretical and applied perspectives*. In the Proceedings of TALN 1999, Cargese, 1999. Disponível em: <http://citeseer.ist.psu.edu/342843.html>

CHIERCHIA, Genaro e McCONNEL-GINET, Sally. *Meaning and grammar: an introduction to semantics*. Cambridge, MA: MIT Press, 1990.

CRUSE, D. Alan. *Meaning in language: an introduction to semantics and pragmatics*. New York: Oxford University Press, 2000.

McNALLY, Louise e BOLEDA, Gemma. Relational adjectives as properties of kinds. In: *Empirical issues in formal syntax and semantics 5*. O. Bonami e P. Cabredo Hofherr (ed.), 2004. pp. 179-196. Disponível em: <http://www.cssp.cnrs.fr/eiss5>

PUSTEJOVSKY, James. *The generative lexicon*. Cambridge, MA: MIT Press, 1995.

RASKIN, Victor e NIRENBURG, Sergei. Lexical semantics of adjectives: a micro-theory of adjectival meaning. MCCS, 1995.

RASKIN, Victor e NIRENBURG, Sergei. Adjectival modification in text meaning representation. In: Proceedings of the 16th Conference on Computational Linguistics, Morristown, Association for Computational Linguistics, 1996, 2: 842-847.