

# Relações semânticas em ontologias: estudo de caso do Blood Project

Maurício Barcellos Almeida\*

Lívia M. D. Teixeira\*\*

Kátia Cardoso Coelho\*\*\*

Renato R. de Souza\*\*\*\*

**Resumo** Nos últimos anos, a pesquisa em ontologias tem recebido destaque pelas possibilidades que oferece na organização da informação. No desenvolvimento de ontologias, a fase de conceitualização merece atenção especial por sua importância e complexidade. Esse artigo busca investigar possibilidades de melhorias na fase de conceitualização, adotando aportes da Linguística para verificar variações na semântica das relações entre termos. Apresenta-se uma proposta que abrange: i) um esquema linguístico para identificar relações semânticas em um texto; ii) o uso de uma ferramenta automática para extração de termos de textos médicos; iii) uma avaliação realizada por médicos sobre quais relações seriam mais adequadas. Investiga-se a existência de variações significativas na semântica das relações, o impacto dessa variação no desenvolvimento de ontologias e a validade das relações obtidas. Espera-se contribuir através de novas possibilidades na construção de instrumentos de organização da informação, bem como fornecendo alternativas para os profissionais envolvidos.

**Palavras-chave** ontologias, Ciência da Informação, processamento de linguagem natural, relações semânticas, organização da informação.

## Semantic relations on ontologies: the case study of the Blood Project

**Abstract** In the last years, research on ontologies has received much attention because of the possibilities it offers regarding information organization. During the ontology development process, the phase of conceptualization deserves special attention on account of its importance and complexity. This paper investigates possible improvements in the phase of conceptualization, relying on linguistics theories to verify variations of semantic relations among

---

\* Professor Adjunto, Depto de Teoria e Gestão da Informação, ECI/ UFMG. Escola de Ciência da Informação, Universidade Federal de Minas Gerais. Av. Antônio Carlos, 6627 - Campus Pampulha - CEP 31.270-901. Belo Horizonte - Minas Gerais - Brasil. Email: mba@eci.ufmg.br

\*\* Mestre em Ciência da Informação pelo Programa de Pós Graduação em Ciência da Informação, ECI-UFMG. Escola de Ciência da Informação, Universidade Federal de Minas Gerais. Av. Antônio Carlos, 6627 - Campus Pampulha - CEP 31.270-901. Belo Horizonte - Minas Gerais - Brasil. Email: [liviamarangon@gmail.com](mailto:liviamarangon@gmail.com)

\*\*\* Mestranda em Ciências da Informação pelo Programa de Pós-Graduação em Ciências da Informação, ECI-UFMG. Escola de Ciência da Informação, Universidade Federal de Minas Gerais. Av. Antônio Carlos, 6627 - Campus Pampulha - CEP 31.270-901. Belo Horizonte - Minas Gerais - Brasil. Email: [katiacocoelho@gmail.com](mailto:katiacocoelho@gmail.com)

\*\*\*\* Professor Adjunto, Depto de Tratamento e Organização da Informação, ECI/UFMG. Escola de Ciência da Informação, Universidade Federal de Minas Gerais. Av. Antônio Carlos, 6627 - Campus Pampulha - CEP 31.270-901. Belo Horizonte - Minas Gerais. Email: [rsouza@eci.ufmg.br](mailto:rsouza@eci.ufmg.br)

terms. We present a proposal including: i) a linguistic-based schema which identifies semantic relations in texts; ii) the use of an automatic tool which extracts terms from medical texts; iii) an assessment conducted with physicians requesting which relations are more appropriate. We investigate the possibility of significant variations in the semantic of those relations, the impact of such variation on the development of ontologies and the validity of the obtained relations. We hope to offer new possibilities regarding the construction of information organization instruments, as well as providing alternatives to involved professionals.

**Keywords** ontologies, Information Science, natural language processing, semantic relations, information organization

## Introdução

Nos últimos anos, a pesquisa em ontologias tem recebido destaque pelas novas possibilidades que oferece na organização da informação. Ontologias são estruturas baseadas em conceitos e relações, especificadas por um vocabulário representacional que pode ter níveis de formalização variáveis. Além do estudo da ontologia na Filosofia, as primeiras pesquisas sobre ontologia aplicada a sistemas de informação foram conduzidas na Inteligência Artificial (GENESERETH e NILSSON, 1987) na década de 1970 e na Engenharia de *Software* na década de 1980 (WAND e WEBER, 1990). A Ciência da Informação (CI), apesar da proximidade com o assunto, registra as primeiras iniciativas apenas na década de 1990 (VICKERY, 1997).

Ontologias têm sido construídas em diferentes domínios do conhecimento, como Direito, Engenharia, Biologia, Medicina, dentre outros. Segundo Smith (2003), as terminologias usadas na pesquisa biomédica e na prática clínica se desenvolveram a partir dos dicionários médicos surgidos ainda na década de 1930. Esses dicionários, criados para serem usados por pessoas, foram as primeiras iniciativas de padronização internacional para esse tipo de terminologia. Entretanto, o volume e a variedade de dados que os pesquisadores da Biomedicina manipulam têm suscitado o emprego de ferramentas automáticas. Dessa forma, as representações formais, como as ontologias, cresceram em importância para uso em computadores.

Entretanto, ontologias não têm sido usadas apenas em sistemas computacionais e as atividades ao longo de sua construção são oportunidades de pesquisa na CI (ALMEIDA, SOUZA e COELHO, 2009). De fato, no desenvolvimento de ontologias destaca-se uma fase em geral denominada “conceitualização”, aquela em que se estabelece uma estrutura conceitual. Na conceitualização, o responsável pela organização do vocabulário representacional efetua abstrações buscando representar a realidade de interesse para seus objetivos. A diversidade de conceitos e de relações existentes em um domínio do conhecimento não torna simples a tarefa de representar a realidade, ou a parte dela, que se deseja conceitualizar (WAND, STOREY e WEBER, 1999).

A conceitualização é uma atividade intelectual e ferramentas automatizadas não podem prover auxílio significativo. Para obtenção das entidades da ontologia, Almeida (2006) afirma que a seleção manual de termos é a melhor solução, mas em função do grande volume de dados não se pode prescindir de ferramentas automatizadas. Já para a criação de relações, Khoo e Na (2006)

destacam a necessidade de uma abordagem interdisciplinar onde se considerem aportes da Lingüística.

O presente trabalho insere-se nesse contexto e propõe-se a verificar variações na semântica de um grupo de relações extraídas automaticamente de textos da área médica e em selecionar aquelas mais adequadas à conceitualização de uma ontologia médica. A pesquisa foi conduzida no âmbito de projeto de pesquisa denominado *Blood Project*<sup>1</sup>. Trata-se de uma iniciativa para organização do conhecimento no domínio da Hematologia e Hemoterapia, no qual tem sido desenvolvida uma linguagem de representação formal sobre o sangue humano especificada por uma ontologia médica, bem como ferramentas e metodologias auxiliares.

Ao longo da pesquisa, foram estudadas diferentes abordagens da Linguística, de forma a aprimorar o entendimento das relações semânticas entre os termos de um texto e utilizada uma ferramenta automática para extração de termos de textos sobre hematologia. Finalmente, valeu-se da capacidade de avaliação humana para decidir quais termos seriam usados na conceitualização da ontologia médica. Buscou-se uma perspectiva de equilíbrio entre a qualidade na escolha dos termos, a qual é garantida pela atuação de pessoas, e a facilidade em manipular grandes volumes de dados, oferecida por sistemas. Pretendeu-se investigar, dentre outras questões: i) a existência real de variação significativa na semântica das relações analisadas; ii) o impacto da variação semântica das relações entre verbos (caso exista) no uso do verbo como relação entre conceitos na ontologia; iii) em que medida as relações *ad hoc* obtidas para construção da ontologia médica são adequadas.

A principal contribuição é a demonstração de que o estudo das variações semânticas entre verbos resulta em melhorias no processo de construção de ontologias, a medida que permite representar a realidade de forma fidedigna. Fornece-se assim ao responsável pela construção de ontologias, recursos adicionais para o refinamento do processo e para obtenção de uma representação detalhada de práticas médicas.

O restante do presente trabalho está organizado da seguinte forma: a seção 2 apresenta brevemente as principais abordagens à semântica; a seção 3 apresenta a pesquisa realizada, descrevendo a metodologia, a coleta de dados, os resultados, bem como uma breve discussão; finalmente, a seção 4 traz as considerações finais, conclusões e propostas para trabalhos futuros.

## **Semântica e relações semânticas**

O estudo da semântica é complexo e variado em função da existência de diversas linhas de pensamento desenvolvidas ao longo dos anos. Segundo Ilari e Geraldi (1985), a semântica é um domínio de investigação que explora fenômenos cujas relações nem sempre são claras. Em função disso, não se apresenta aqui um levantamento exaustivo do campo da semântica.

O restante desta seção está organizado conforme segue: a Seção 2.1 apresenta conceitos básicos da Linguística, menciona as principais linhas de pensamento da área, bem como a sua interseção

---

<sup>1</sup> Disponível na Internet em <http://mba.eci.ufmg.br/blood/>; projeto de pesquisa em desenvolvimento na Fundação Hemominas (Fundação Centro de Hematologia e Hemoterapia de Minas Gerais), uma instituição pública que tem por finalidade garantir à população a oferta de sangue e hemoderivados com qualidade.

com a CI; a Seção 2.2 discute as relações semânticas de um ponto de vista lexical, destacando a troponímia; a seção 2.3 apresenta um esquema lexical para extração de relações semânticas.

## Considerações básicas sobre semântica na Linguística e na CI

A semântica é um ramo da Linguística voltado ao estudo do significado de palavras e de expressões, levando em consideração as “funções e natureza das palavras no processo de comunicação [...]” (RECTOR, YUNES, 1980, p.11). Cançado (2005) divide o estudo do significado em: i) a *abordagem referencial*, que advoga a relação entre o referente no mundo e as palavras que o representam; ii) a *abordagem mentalista*, na qual o sentido está em um nível intermediário entre o mundo e as palavras, denominado nível de representação mental; iii) a *abordagem pragmática*, a qual estuda os usos situados da língua, ou seja, os usos que correspondem ao modo como as palavras são pronunciadas em certas circunstâncias. Um esquema das principais teorias semânticas, bem como respectivos autores significativos, é apresentado no restante da presente seção, bem como iniciativas da CI relacionadas a semântica.

A *abordagem referencial* fundamenta-se no processo de comunicação, nas formas através das quais as regras e os princípios gramaticais combinam palavras e expressões significativas, baseando-se em condições de verdade<sup>2</sup>. Cançado (2005, p.79) apresenta como exemplo dessa abordagem a sentença *Noam Chomsky é um famoso linguísta*, explicando que o nome *Noam Chomsky* refere-se a um único indivíduo no mundo, que tem como características *famoso* e *linguísta*. Exemplos dessa vertente são a semântica formal de Lyons (1977) e de Kempson (1977), bem como a semântica lexical de Cruse (1986) e de Evens (1988).

A *abordagem mentalista* fundamenta-se na construção do significado através de representações mentais criadas pelas pessoas a respeito do que dizem ou observam. Essas representações são esquemas mentais que podem ser imagéticos ou conceituais. Para Tuggy (2007) um esquema é um conceito super-ordenado que especifica um esboço básico comum a muitos conceitos mais específicos, os quais são denominados “instanciações”. Exemplos de abordagens dessa linha de pensamento são a semântica gerativista de Chomsky (1965), a semântica representacional de Jackendoff (1983) e a semântica cognitiva de Lakoff (1987).

Na *abordagem pragmática*, a linguagem “é mais do que a descrição dos estados das coisas, como na linguagem referencial, e mais do que estabelecer a relação entre o mundo e os conceitos mentais, como tratado na linguagem mentalista” (CANÇADO, 2005, p.127). O significado é, nesse caso, atribuído à prática e às convenções sociais. Como exemplo, a sentença *O cachorro está no jardim* tem um significado que, ao interagir com a capacidade humana de contextualização, permite sugerir atos distintos como: *Tire o cachorro do jardim!*, ou *Socorro!* Exemplos de abordagens relevantes dessa linha de pensamento são a teoria dos atos da fala de Austin (1962) e a semântica argumentativa de Ducrot (1987).

A evolução histórica dessas abordagens se deu com a sobreposição parcial de teorias, sendo que várias delas ainda coexistem e são utilizadas. Um quadro histórico bastante genérico poderia considerar primeiramente *Ferdinand de Saussure*<sup>3</sup>, considerado o pai da Linguística, que

<sup>2</sup> Abordagem à semântica da linguagem natural que admite que o significado de declarações é o mesmo, ou pelo menos reduzível à condições baseadas em uma definição indutiva de verdade.

<sup>3</sup> *Ferdinand de Saussure*, linguísta suíço (1857 - 1913).

afirmava que o signo lingüístico é a junção de significante e significado. Sua obra está vinculada ao *paradigma estruturalista* da linguagem, que a estuda como um sistema isolado, voltado para a estrutura e a dinâmica da língua. Nesse caso, os aspectos relacionados à representação do mundo pela linguagem são considerados extra-lingüísticos (SILVA, 1997).

Em meados no século XX, *Noam Chomsky*<sup>4</sup> introduziu o *paradigma gerativista*, no qual a linguagem é um componente autônomo e específico da mente, independente das outras faculdades mentais. Apesar de introduzir a noção da linguagem como um sistema cognitivo, o objeto de conhecimento é a linguagem como "sistema de conhecimento" (CHOMSKY, 1986, p.24), o que justifica o surgimento da gramática gerativa. Nessa abordagem, o conhecimento da linguagem é independente de outros tipos de conhecimento. Essa teoria abriu caminho para uma renovação da lingüística e para sua aplicação a diversas disciplinas do saber humano, como a Psicologia ou a Sociologia.

No final dos anos 70 e início dos 80, pesquisadores como *Lakoff*<sup>5</sup>, *Langacker*<sup>6</sup> e *Talmy*<sup>7</sup> repensaram a lingüística sob a ótica da significação, não mais considerando autônomas as estruturas formais da linguagem. O *paradigma da lingüística cognitiva* enfatiza o estudo da língua através de uma organização conceitual geral. A proposta considera princípios de categorização, mecanismos de processamento, influências da experiência e do próprio meio ambiente em que atua o indivíduo.

Segundo Croft e Cruse (2004), o *estruturalismo* e o *gerativismo* cederam lugar à *lingüística cognitiva*, uma vez que a última mantém os argumentos básicos das outras abordagens e ainda explora questões empíricas da sintaxe e da semântica. A forma como a lingüística cognitiva aborda a representação mental, através de *esquemas* e *frames*, facilita o entendimento da representação do conhecimento como subsídio para a criação de modelos.

No âmbito da CI, a questão da semântica, talvez influenciada pela variedade de visões e pela falta de um consenso teórico na Linguística, aparece de forma um tanto fragmentada (HJØRLAND, 2007a). Ainda assim, é possível encontrar uma variedade de estudos envolvendo semântica, alguns dos quais são apresentados no restante da presente seção.

De fato, Van Rijsbergen (1986) diagnostica certa negligência na definição de significado pelo campo da CI. Afirma que os estudos semânticos têm sido amplos e de difícil sistematização, pois as contribuições vêm de áreas diversas, como Filosofia, Psicologia e Cognição. Orrico (1999) entrelaça os paradigmas da CI relativos à natureza e cultura, ao subjetivo e objetivo, estabelecendo uma interface entre a CI e a Lingüística através da Terminologia. Mendonça (2000) estabelece relações entre a Lingüística e a CI através do levantamento de artigos temáticos em periódicos dessa última área, de acordo com: i) relação teórica entre a lingüística e a CI; ii) interseção entre a lingüística e a bibliometria; iii) organização do conhecimento e respectiva representação; iv) indexação automática e linguagem natural para recuperação de documentos; v) formação do profissional da informação; vi) usos de sistemas especialistas como provedores de informação; vii) lingüística no âmbito da CDU.

Green (2001) discute a possibilidade de um inventário de relacionamentos semânticos, incluindo distinções entre as tipologias. As relações semânticas usadas na organização do conhecimento são agrupadas em: i) relacionamentos entre as entidades de conhecimento registrado, baseado em

<sup>4</sup> *Noam Chomsky*, filósofo e linguísta norte-americano (1928 - ?).

<sup>5</sup> *George Lakoff*, linguísta americano (1941 - ?).

<sup>6</sup> *Ronald Langacker*, linguísta americano (1942 - ?).

<sup>7</sup> *Leonard Talmy*, linguísta americano (19xx - ?).

descrições bibliográficas; ii) relacionamentos em um mesmo texto e entre textos diferentes, através de citações e *hiperlinks*; iii) relacionamentos de assuntos em um tesauro ou outra estrutura classificatória; iv) pertinência da relação. Os relacionamentos são especificados simultaneamente pela identificação das relações semânticas e dos participantes envolvidos.

Lima (2003) agrega a *Ciência Cognitiva* aos estudos da CI, refletindo sobre a formação de um modelo de mundo pelo indivíduo, tanto para o processamento quanto para a recuperação da informação. A autora busca responder a questões sobre a maneira como “[...] as estruturas conceituais que formam o universo do usuário têm correspondência com a estrutura conceitual que forma o universo do sistema de recuperação” (LIMA, 2003, p.82). Souza e Alvarenga (2004) discutem a semântica sob a perspectiva da Web Semântica, na qual a lógica computacional estabelece possibilidades de relações de significação por meio de metadados descritivos em linguagens de marcação. Destacam-se as diferentes características de linguagens de marcação e padrões de metadados, os quais lhes conferem diferentes capacidades de expressão semântica.

Khoo e Na (2006) enfatizam a natureza, os tipos e a especificidade das relações semânticas para a construção de ontologias, representação, extração e recuperação de informação, enfatizando as perspectivas linguística e psicológica. Os autores advogam o uso do processamento da linguagem natural para a obtenção de relações semânticas para recuperação da informação, sugerindo estudos sobre os tipos de relacionamentos entre a necessidade de informação do usuário, a tarefa, situação e conteúdo do documento que determina a relevância e usabilidade do mesmo.

No âmbito da organização do conhecimento, Hjørland (2007a) suscita a necessidade do debate teórico, uma vez que as contribuições advindas da própria CI são distintas, não relacionadas e nem mesmo vinculadas a uma visão abrangente. Segundo o autor, o entendimento do mundo deve ser construído a partir das conexões criadas ao observar e listar os conceitos do ambiente ou objeto que se pretende representar. Atribuir relações significativas entre esses conceitos é a base para a construção do conhecimento em um domínio específico.

Uma vez apresentadas noções básicas sobre a semântica e algumas iniciativas relacionadas na CI, a seção seguinte apresenta a abordagem lexical, a qual é relevante para o estudo de relações semânticas e para os propósitos do presente trabalho.

## **Relações semânticas na abordagem lexical**

A abordagem lexical estuda uma língua através de seu vocabulário, suas palavras e expressões. Existem diversas propostas para a definição de relações semânticas na abordagem lexical, por exemplo: Hjørland (2007b) distingue *relações semânticas* entre conceitos e *relações lexicais* entre palavras; Khoo e Na (2006) consideram relações semânticas como relações significativas entre dois ou mais conceitos, entidades ou conjunto de entidades, podendo ainda fazer referência a relações entre conceitos mentais, entre elementos lexicais e entre parágrafos; Rector e Yunes (1975) consideram as *relações semânticas* como relações associativas que ocorrem dentro de um mesmo *campo semântico*, ou seja, dentro de um contexto onde os termos têm sentido e função.



Em uma distinção bastante conhecida, Cruse (1986) organiza as *relações semânticas* em dois tipos principais: as *relações sintagmáticas* e as *relações paradigmáticas*. A FIG.1 descreve como ocorrem essas relações, esclarecendo como os termos são usados para construir significados:



**Figura 1: Relações sintagmáticas e paradigmáticas**  
Fonte: adaptado do autor

As *relações sintagmáticas* ocorrem no sentido linear da frase, nas relações entre os léxicos que a constituem e propõem significado através de regras gramaticais e sintáticas. As *relações paradigmáticas* ocorrem entre as palavras de uma mesma classe gramatical (substantivo, advérbio, adjetivo, verbo, etc.), não necessariamente dentro de um contexto significativo, mas em um processo de categorização gramatical.

As relações entre os elementos do léxico podem ocorrer no mesmo contexto e constituir parte de da memória semântica (KHOO e NA, 2006), a qual corresponde a organização do conhecimento que uma pessoa possui sobre uma dada entidade. Exemplos desses tipos de relação são as relações de *hiponímia*, de *troponímia*, *meronímia*, *sinonímia* e *antonímia*. No restante da presente seção, descrevem-se essas *relações paradigmáticas*.

O par *hiponímia* – *hiperonímia* implica a relação hierárquica entre termos. Lyons (1977, p. 295) afirma que “a relação de hiponímia impõe uma estrutura hierárquica sobre o vocabulário e sobre campos particulares dentro do vocabulário [...]”<sup>8</sup>. Usualmente, a relação de hiponímia é denominada relação *is-a* (é um) e *a-kind-of* (um tipo de). Um exemplo de relação hiponímica é a frase *Um garanhão é um cavalo*.

O par *meronímia* – *holonímia*, também conhecido como relação *part-whole* (parte-todo), refere-se à relação entre uma entidade e suas partes constituintes. Winston, Chaffin e Herrmann (1987) enfatizam a relação expressa como *part-of* (parte-de) e citam seis tipos de relações meronímicas: i) componente-objeto, por exemplo, *A asa é parte da xícara*; ii) membro-coleção, por exemplo *Uma árvore é parte de uma floresta*; iii) porção- massa, *Essa fatia é parte da torta*; iv) matéria-objeto, *O álcool é componente do Martini*; v) tarefa-atividade, *Pagar é parte de comprar*; vi) lugar-área, *Os Everglades são parte da Flórida*.

A *relação de sinonímia* se manifesta quando existem duas expressões que, para todos os contextos, podem se substituir mutuamente sem alterar o significado da sentença. Lyons (1977) reforça que as sinonímias absolutas são raras, uma vez que os significados devem ser semanticamente equivalentes em todas as dimensões. As *relações de antonímia* são amplamente conhecidas e as pessoas têm facilidade em aprendê-las (JONES, 2002). Referem-se à oposição

<sup>8</sup> “The relation of hyponymy imposes a hierarchical structure upon the vocabulary and upon particular fields within the vocabulary [...]”.

no léxico verbal dentro de um mesmo campo semântico, na realização de uma mesma atividade. Segundo Fellbaum (1998), exemplos são os pares: *dar - tirar; comprar - vender*.

Existem ainda relações que ocorrem apenas entre verbos, as quais são estabelecidas através das *relações de modo*, ou seja, as maneiras pelas quais uma ação pode ser executada. Fellbaum (2002) denomina tais relações de *manner-of* (modo-de) ou relações de *way-of* (forma-de). Por exemplo, os verbos “*correr*” e “*voar*” podem ser considerados tipos de “*mover*”. Apesar de sua validação psicolinguística, as relações de troponímia são relações entre itens lexicais que apresentam variações provenientes de diferentes “modos-de” possíveis (FELLBAUM e CHAFFIN, 1990). Exemplos dessas variações são *modos de movimento, modos de função, modos de propósito*, dentre outros.

Segundo Pustejovsky (1995), a semântica de substantivos está em muitos casos associada a *papéis*. A relação entre um substantivo (*cão*) e um papel (*animal de estimação*) não é a mesma que entre um substantivo (*cão*) e outro substantivo (*animal*). No caso de verbos, existem situações em que, além de seu sentido estritamente denotacional, os verbos carregam aspectos de significado que dependem de contexto e que fazem variar sua função. Por exemplo, *correr, andar e nadar* são verbos com modos de movimento, enquanto *andar de bicicleta* é um verbo com modo de movimento usando um veículo. Entretanto, todos são modos de se exercitar, de forma que os dois modos originais (movimento e movimento com veículo) podem ser reunidos em outro modo de acordo com o contexto.

Segundo Levin (1993), as variações no modo dos verbos e a consequente variação semântica é refletida em um comportamento sintático compartilhado. Verbos são classificados de acordo com alterações sintáticas e os grupos resultantes apresentam coerência semântica. Uma observação atenta revela que os membros de uma classe incluem tanto verbos básicos, através dos quais a classe é nomeada, quanto verbos que aparecem em modos subordinados.

Fellbaum (1998) apresenta um abrangente trabalho sobre *troponímia* e *redes semânticas de verbos* na língua inglesa. A troponímia, do grego *tropos* e que significa *maneira*, indica a relação existente entre verbos que ocupam uma mesma posição hierárquica e são sobre-ordenados a um evento geral, com outros verbos de eventos mais específicos que os primeiros (sub-ordenados) em relação a uma mesma ação. Por exemplo, modos específicos de realizar a atividade de *andar* são: *a Maria caminha; a Maria manca; a Maria coxeia*.

A troponímia originou-se na observação da existência de relações de hiponímia também entre verbos, e não só entre substantivos. Constatou-se que não seria interessante identificar as duas relações (hiponímia entre verbos e hiponímia entre substantivos) com o mesmo nome, uma vez que possuem características diferentes (FELLBAUM, 1990). Em tal relação, aplica-se a fórmula: o *verbo1* está para o *verbo2* de uma maneira particular (FELLBAUM e MILLER, 1990). Por se tratar de relação em uma classe gramatical, a relação de troponímia é paradigmática e não se refere à atribuição de significado a uma sentença, mas a um processo de categorização. Para o estabelecimento de categorias entre os léxicos verbais, a troponímia, sendo uma *relação de modo*, explica como a ação foi executada e permite construir uma estrutura hierárquica.

Um dos instrumentos mais conhecidos que faz uso de relações de troponímia para verbos é a *Wordnet* (FELLBAUM e MILLER, 1990)<sup>9</sup>, um grande repositório de dados lexicais da língua inglesa, composto por adjetivos, verbos, advérbios e substantivos, agrupados em conjuntos de sinônimos e inter-relacionados por relações conceituais, léxicas e semânticas

---

<sup>9</sup> Disponível na Internet em <http://wordnet.princeton.edu/>. Acesso em 30 set. 2008.



Miller (1998) explica que a *WordNet* não é um dicionário convencional para computadores e Gomez-Perez et al (2004) consideram a *WordNet* como uma grande “ontologia linguística” que apresenta relações diversas entre as palavras que contém. Do ponto de vista prático, o resultado é uma teia de significados para palavras, inclusive verbos, disponível em um navegador. Por conter grande quantidade de verbos e relações entre eles, a *Wordnet* é uma alternativa para pesquisa sobre troponímia.

Fellbaum e Chaffin (1990) descrevem experimentos para a validação da troponímia na *WordNet*. Os resultados demonstram que a troponímia é uma relação perceptível por nativos em uma língua. Os verbos foram organizados em dois grandes tipos: verbos que denotam *ação e evento* e verbos que denotam *estado*. Esses dois tipos foram subdivididos entre quatorze domínios semânticos específicos, de acordo com classificação proposta por Miller e Johnson-Laird (1976): movimento, eventos climáticos, contato, comunicação, competição, mudança, cognição, consumo, criação, emoção, percepção, posse, funções corporais, percepção, posse, comportamento social, verbos de estado.

Mesmo considerando um único domínio semântico, os verbos nem sempre apresentam a mesma raiz morfológica. O mesmo verbo pode pertencer a dois domínios relacionados, por exemplo, *mover* pode significar *fazer um movimento* ou *viajar, deslocar-se*. Além da divisão em domínios semânticos, existem ainda na *WordNet* quatro eventos linguísticos que definem a organização da rede: i) a oposição semântica, por exemplo, *lembrar e esquecer*; ii) os conjuntos de sinônimos, por exemplo, *fechar e cerrar*; iii) o sentido metafórico, por exemplo *Ela vai a padaria* (ir como deslocar) e *O primeiro prêmio vai para ela* (ir como ser premiado); iv) os verbos polissêmicos.

As relações de modo, as formas pelas quais os verbos variam, são especialmente relevantes para os propósitos do presente artigo. Na seção seguinte, apresentam-se variações de verbos a partir dos advérbios que os modificam, construindo-se um esquema que será usado para extração automática de termos dos textos médicos.

## **Esquema para identificação e extração de verbos e modos**

Para que se possa extrair os verbos e os respectivos modos disponíveis em um texto, uma alternativa é considerar que advérbios modificam a semântica dos verbos através de modos (LEECH e SVARTVIK, 2002). Na presente seção, apresentam-se conceitos básicos e regras linguísticas para a construção de um esquema para obtenção dos verbos e suas variações. Na prática, a literatura adotada pelos médicos é essencialmente em inglês, de forma que as regras apresentadas remetem a gramática daquela língua. Alguns conceitos importantes para a criação do esquema foram obtidos em dicionários e gramáticas da língua inglesa como Yule (2009), Vince (2009), dentre outros.

Uma classe de sentenças importante nesse contexto são as *adverbial adjunct clauses*, as quais correspondem a *elementos da sentença* que estabelecem circunstâncias relativas a uma ação ou estado expresso por um verbo. Um elemento da sentença pode ser um *subject*, um *predicate*, um *object*, ou um *adverbial*. Por sua vez, um *adverbial* pode ser um: *adverb*, *prepositional phrase*, *noun phrase*, *clause with a finite verb*, *infinite clause*, *participle clause*, *verbless clause*. Um *adverbial adjunct* pode se apresentar através de uma palavra ou frase, como por exemplo, em

*She will leave tomorrow*, ou através de uma oração, como por exemplo em *She will leave after she has had breakfast*.

Além disso, as *adverbial clauses* normalmente se localizam no final da sentença após o verbo, caso não exista objeto ou complemento, por exemplo, em *I'd like to leave as soon as possible*; ou após o objeto ou complemento, por exemplo em *Please, don't call me before nine o'clock*. No esquema proposto, considerou-se as *adjunct adverbial clauses* localizadas no final da sentença e separadas da outra oração por uma conjunção (existem exemplos de conjunções no início da sentença, as quais não consideradas para fins de simplificação). A conjunção é assim o elemento sintático usado para identificar a *adjunct adverbial clause* por tipo. Dessa forma, a regra para a extração automática pode ser estabelecida como: A sentença é composta pela oração (1) que tem um VERBO e pode ter um complemento; após o complemento tem-se uma CONJUNÇÃO que conecta a oração (1) e a oração (2), sendo a oração (2) identificada por outro VERBO. A oração (2) modifica o verbo da oração (1) como um advérbio.

Em seguida, listam-se alguns tipos de *adjunct adverbial clause* identificadas e as conjunções correspondentes utilizadas para identificá-las:

i) *TIME* (relacionados a *when*)

Conjunções: *when, as, before, after, while, as soon as, once, now, since, until*

Exemplo 1: *It was almost totally dark when they arrived*;

Exemplo 2: *We'll let you know as soon as we've made up our minds*.

ii) *PLACE* (relacionados a *where*)

Conjunções: *where, wherever*

Exemplo: *I left it where I found it*

iii) *CONTRAST* (confrontam circunstâncias, *concessive clauses*)

Conjunções: *but, although/though, while, whereas, in spite of, despite*

Exemplo 1: *He hadn't eaten for days, but he looked strong and healthy*

Exemplo 2: *Mary was lively and talkative, whereas her sister was quite and reserved*.

iv) *CAUSE/REASON* (relacionadas a *why*), onde *CAUSE* diz respeito aos próprios eventos e *REASON* diz respeito a forma como uma pessoa interpreta os eventos.

Conjunções: *because, because of, on account of, since, as, given*

Exemplo 1: *The accident occurred because the machine had been poorly maintained*.

Exemplo 2: *The contest was abandoned on account of bad weather conditions*.

v) *PURPOSE*

Conjunções: *in order to, so that, in order that, in case*

Exemplo 1: *They advertised the concert so that everyone should know about it.*

Exemplo 2: *He left early in case he should miss the last train.*

vi) **RESULT**

Conjunções: *so, so that*

Exemplo: *The cleaner has gone on holiday vacation so that everthing is so dirty.*

vii) **CONDITIONAL**

Conjunções: *if, unless*

Exemplo 1: *I would lend Peter the money if he needed it.*

Exemplo 2: *You 'd be bored if you had no children.*

viii) **MANNER** (relacionado a *how*, expressando comparação)

Conjunções: *like this, like, this way, that way, the way*

Exemplo 1: *Pat cooks turkey in the way my mother did.*

Exemplo 2: *She sings like a professional.*

Na seção seguinte, seção 3, descreve-se a pesquisa realizada em que se aplicam os conceitos e as regras linguísticas apresentados acima na programação de ferramenta automática usada para extração de termos de textos médicos.

## **A pesquisa**

Conforme mencionado na seção 1, o objetivo da pesquisa diz respeito à obtenção de melhorias na fase de conceitualização da construção de ontologias. Segundo Almeida (2006), a *Methontology* (FERNANDEZ-LOPEZ et al 1997) é a metodologia mais madura para construção de ontologias. Dentre as diversas tarefas constantes na *Methontology* para a fase de conceitualização, considera-se aqui a possibilidade de melhorias na *tarefa de construir diagramas de relações binárias ad hoc*. Segundo Gomez-Perez et al (2004), essa tarefa corresponde a criação de relacionamentos entre conceitos, além da verificação da adequação do domínio e do escopo de cada um de seus argumentos. Os resultados que se espera obter serão aplicados à pesquisa desenvolvida no *Blood Project*.

O *Blood Project* (ALMEIDA, 2009a) é uma iniciativa para organização do conhecimento no domínio da Hematologia e Hemoterapia. No âmbito do projeto têm sido desenvolvidas as etapas necessárias para obtenção de uma linguagem de representação formal sobre o sangue humano. O projeto tem sido desenvolvido através de uma parceria entre a instituição pública de saúde de Minas Gerais responsável pela manipulação do sangue humano para fins terapêuticos (Fundação

Hemominas), uma universidade pública (Universidade Federal de Minas Gerais) e no âmbito de um projeto internacional (*Open Biomedical Ontologies*<sup>10</sup>).

Como resultado da aplicação dos aportes linguísticos a textos do domínio da Hematologia e Hemoterapia, espera-se obter futuramente princípios metodológicos para auxílio profissional responsável pela modelagem de relações na fase de conceitualização. As etapas da pesquisa são descritas em detalhe a seguir.

## **Metodologia**

A *primeira etapa* da pesquisa consistiu em selecionar a amostra de textos representativos do domínio do conhecimento sob estudo, ou seja, Hematologia e Hemoterapia. A amostra de textos foi selecionada a partir de entrevistas semi-estruturadas realizadas com médicos. Ao todo foram entrevistados quinze profissionais, sendo dez médicos que atuam diretamente ou indiretamente em atividades relacionadas às diferentes áreas envolvidas no processo de manipulação do sangue, três técnicos (profissionais não médicos) que atuam direta ou indiretamente no processo de manipulação de sangue e dois profissionais de sistemas de informação (sendo um da instituição e um colaborador externo, mas com conhecimento na área médica). A entrevista semi-estruturada foi simples e continha questões relativas a: i) projeto: breve descrição, levantamento preliminar e aquisição de conhecimento, escopo, resultado esperados; ii) fontes de dados e contexto: campo de atuação do pesquisador, fontes de dados, descrição das fontes de dados e forma de acesso a pessoas com *expertise*.

A *segunda etapa* consistiu na extração automática de termos dos textos selecionados, utilizando-se a ferramenta automática GATE<sup>11</sup>. A ferramenta foi configurada para extrair verbos candidatos a relações entre conceitos do domínio, bem como advérbios que os modificavam. Dessa forma, obtiveram-se as diversas possibilidades *manner-of* (modo-de) de interesse para a etapa de conceitualização. A extração dos termos foi conduzida pela extração de sentenças adverbiais, especificamente advérbios e verbos correspondentes, para posterior análise sobre as variações semânticas nas relações da ontologia médica. Os seguintes tipos de advérbios foram considerados, de acordo com as regras estabelecidas na seção 2.3: maneira, local, tempo-freqüência, tempo-duração, grau.

A *terceira etapa* consistiu na organização em planilhas dos diversos verbos e correspondentes advérbios extraídos dos textos pela ferramenta automática. Os itens das tabelas foram dispostos de acordo com os resultados gerados pelo GATE. A *quarta e última etapa* consistiu no agrupamento e organização das relações de modo propostas. Uma vez que as relações obtidas foram registradas em formulário, passou-se a análise sobre a utilidade dos modos obtidos para a conceitualização da ontologia. Tal avaliação se baseou em técnicas de avaliação do conteúdo de ontologias centradas no usuário (ALMEIDA, 2009b).

---

<sup>10</sup> Disponível na Internet em <http://www.obofoundry.org/>. Acesso em 22 Abril 2008.

<sup>11</sup> *General Architecture for Text Engineering* ou GATE é um *software* desenvolvido na *University of Sheffield*. Disponível na Internet em: <http://gate.ac.uk/>. Acesso em: 28 maio de 2009.

## Dados coletados

Na etapa 1 prevista foram selecionados textos da área médica de acordo com entrevistas conduzidas com especialistas. As seguintes fontes de dados (Quadro 1) da área de hematologia foram sugeridas:

Quadro 1: Fragmento do resumo das fontes de dados obtidas

#	Fonte de Dados	Tipo	Acesso
1	<i>The American Association of Blood Banks</i>	Instituição	Site
2	<i>National Center for Biotechnology Information</i>	Instituição	Site
3	<i>Blood Group and Red Cell Antigens</i>	Livro	Site
3	<i>Journal of American Society of Hematology</i>	Periódico	Site
4	<i>Glossary of Hematology and Blood Disorders, University of Virginia</i>	Glossário	Site
5	<i>The American Society of Hematology</i>	Instituição	Site
6	<i>BloodLine</i>	Periódico	Site
7	Associação Brasileira de Hematologia e Hemoterapia	Instituição	Site
8	Agência Nacional de Vigilância Sanitária	Instituição	Site
9	<i>Blood Essentials, The American Association of Blood Banks</i>	Livro	Digital
12	<i>Essentials of Clinical Immunology</i>	Livro	Digital
13	<i>Hematology: Basic Principles and Practice</i>	Livro	Digital
14	<i>Color Atlas of Hematology</i>	Livro	Digital
15	<i>Oxford Handbook of Clinical Haematology</i>	Livro	Digital
16	Hematologia: Fundamentos e Prática	Livro	Papel
17	<i>Atlas of Clinical Hematology</i>	Livro	Digital

18	Tratado de Fisiologia Médica	Livro	Papel
----	------------------------------	-------	-------

**Fonte: Elaborado pelos autores a partir da literatura especializada**

Cada texto foi fichado, de forma a obter-se um registro simples e prático de seu conteúdo, bem como de suas principais características. Essa atividade facilitou o entendimento do material, que mesmo considerado básico na área de hematologia, não é de fácil entendimento para não especialistas. O Quadro 2 apresenta um exemplo de descrição de um fonte de dados:

**Quadro 2: Exemplo de descrição de fonte de dados**

<b>Descrição de fontes de dados</b>		<b>Responsável:</b> K 1/09/2009	<b>Data:</b>
<b>Referência:</b> <i>American Society of Hematology</i>			
<b>Localização física:</b> Internet, <a href="http://www.hematology.org">http://www.hematology.org</a>		<b>Mídia:</b> digital	<b>Cod. Fonte:</b> 05
<b>Tema:</b> associação que define práticas de hematologia: diagnóstico, tratamento, e prevenção de doenças que afetam o sangue e sistemas correlatos. Promove pesquisa, educação, treinamento em hematologia.			
<b>Estrutura:</b>			
<b>Menu 1</b>	<b>Menu 2</b>	<b>Menu 3</b>	
<i>Home</i>	<i>Home</i>	<i>Home</i>	
<i>About</i>	<i>Research</i>	<i>Meetings</i>	
<i>Patients</i>	<i>Practise</i>	<i>Publications</i>	
<i>Media</i>	<i>Traning</i>	<i>Education &amp; Careers</i>	
...		...	
<b>Resumo descritivo:</b>			
Trata-se de um site com material básico que pode ser de utilidade no entendimento do domínio. Apresenta recursos diversificados como filmes, livros, pesquisa, buscas por profissionais, referências a sites, material básico sobre doenças no sangue, dentre outros.			
<b>Citações:</b>			
<b>Menu 1</b> , link <i>Patients</i> : tem diversos recursos didáticos (filmes, material básico, links)			
<i>Patients</i> → <i>Blood Basis</i> → <i>Advances in hematology</i> : apresenta progressos sobre a pesquisa em doenças do sangue, classificando-as em seis tipos de			



doenças;

*Patients* → *Blood Basis* → *other resources*: apresenta alguns artigos do livro *ASH Education Program Book* (<http://asheducationbook.hematologylibrary.org/>); traz lista de sites adicionais sobre cada tipo de doença, além de uma breve descrição de cada um deles: *Amyloidosis, Anemia, Autoimmune and Inflammatory Diseases, Bleeding Disorders, Blood Clotting disorders, Blood Donation, Blood Tests, Bone Marrow and Stem Cell Transplantation, Cancer, Clinical Trials, Genetic and Metabolic Disorders, Histiocytoses, Leukemia, Lymphoma, Multiple Myeloma and Waldenström Macroglobulinemia, Myelodysplastic Syndromes (MDS), Myeloproliferative Disorders (MPD), National Organizations, Pediatric Issues.*

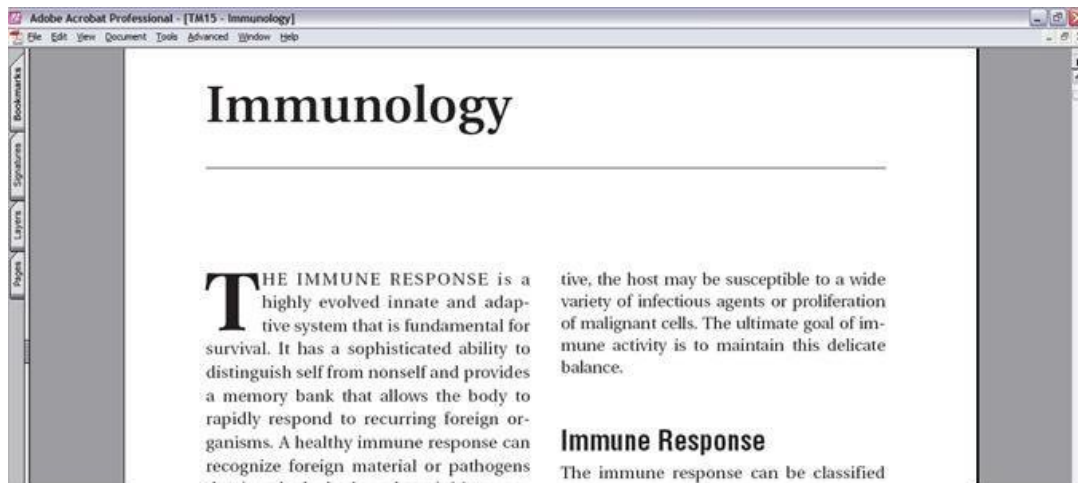
*Patients* → *For patients*: descreve alguns conceitos básicos sobre sangue, doação, transfusão e tipos de doenças do sangue.

...

**Fonte: Elaborado pelos autores**

A segunda etapa da metodologia previa a extração automática dos termos. Para efeito de apresentação da pesquisa descreve-se a fonte número 09 (listada no Quadro 1), que corresponde a um livro em formato digital com conceitos básicos sobre hematologia de cerca de vinte capítulos. Desses, foram separados doze capítulos, os quais eram fontes relevantes no âmbito do *Blood Project*. A extração automática de termos verbos e advérbios foi realizada através da ferramenta GATE e demandou cerca de trinta horas para ser concluída. A FIG. 2 apresenta um fragmento de capítulo da fonte de dados selecionada.

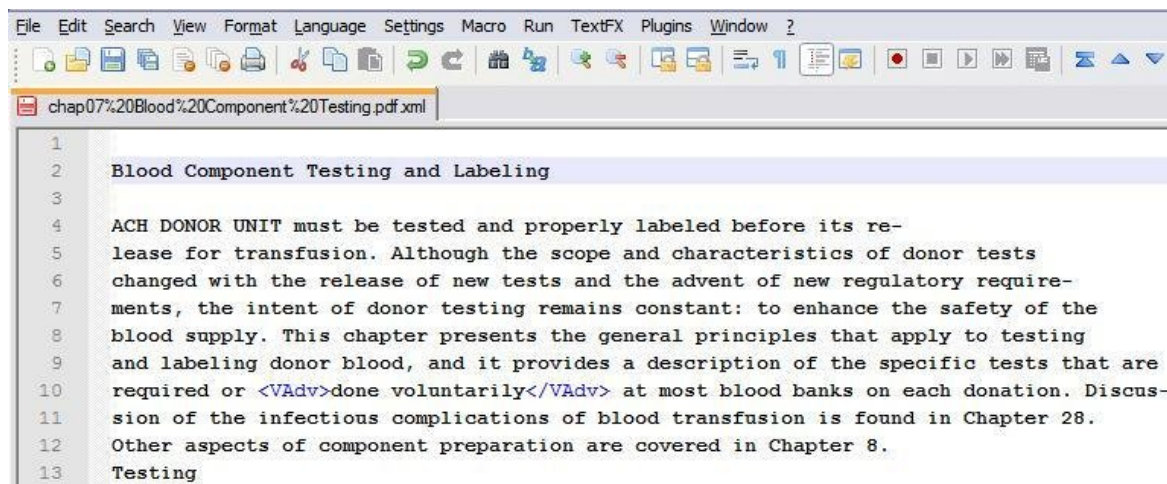
**Figura 2: Extrato de texto selecionado para extração de termos**



**Fonte: Adaptado do autor**

A FIG. 3 apresenta um extrato da marcação automática realizada pela ferramenta com um exemplo de texto marcado obtido após o processo de extração. As marcas que delimitam os termos de interesse aparecem no texto na forma <Vadv> </Vadv>.

**Figura 3: Texto da área médica marcado <Vadv> </Vadv> para extração**



**Fonte: Elaborado pelos autores**

A terceira etapa previa a organização dos termos extraídos em planilhas com verbos e advérbios correspondentes, demandando cerca de cinquenta horas de trabalho. Os Quadros 3 e 4 apresentam fragmentos das planilhas resultantes para dois textos médicos.

**Quadro 3: Tabela de termos extraídos (capítulo 4)**

#	Texto	Arquivo	Verbo	Advérbio	Verbo
1	Chapter 4: Allogeneic Donor ... ...	chap04.xml 1	<i>adjusted</i>	<i>Appropriately</i>	<i>Adjusted</i>
2			<i>agree</i>	<i>Not</i>	<i>Agree</i>
3			<i>Am</i>	<i>Potentially</i>	<i>AM</i>
4			<i>are</i>	<i>Always</i>	<i>Are</i>
5			<i>are</i>	<i>Not</i>	<i>Are</i>
6			<i>Are</i>	<i>Occasionally</i>	<i>Are</i>
7			<i>are</i>	<i>Seldom</i>	<i>Are</i>
8			<i>Be</i>	<i>As</i>	<i>be</i>

9			<i>be</i>	<i>No</i>	<i>Be</i>
10			<i>be</i>	<i>Particularly</i>	<i>Be</i>

Fonte: Elaborado pelos autores

Quadro 4: Termos extraídos (capítulo 3)

#	Texto	Arquivo	Verbo	Advérbio	Verbo
1	Chapter 3: Blood Utilization ...	chap03.xml 1	<i>Achieve</i>	<i>Evenly</i>	<i>Achieve</i>
2			<i>are</i>	<i>Often</i>	<i>Are</i>
3			<i>be</i>	<i>Clearly</i>	<i>Be</i>
4			<i>be</i>	<i>Periodically</i>	<i>Be</i>
5			<i>be</i>	<i>Safely</i>	<i>Be</i>
6			<i>be</i>	<i>Visually</i>	<i>Be</i>
7			<i>counted</i>	<i>Once</i>	<i>Counted</i>
8			<i>do</i>	<i>Not</i>	<i>Do</i>
9			<i>given</i>	<i>Instead</i>	<i>Given</i>
10			<i>made</i>	<i>Once</i>	<i>Made</i>

Fonte: Elaborado pelos autores

Finalmente, na quarta etapa, foi realizado o agrupamento e registro dos modos (Quadro 5), procedimento que despendeu cerca de cem horas.

**Quadro 5: Fragmento de tabela de agrupamento e registro dos modos**

<b>VERBO</b>	<b>ADVÉRBIO</b>	<b>REPETIÇÃO</b>
<i>HANDLE</i>	<i>Only</i>	1
<i>HAVE</i>	<i>Already</i>	3
	<i>Also</i>	9
	<i>Anecdotally</i>	1
	<i>Clinically</i>	2
	<i>Consistently</i>	2
	<i>Dramatically</i>	1
	<i>Essentially</i>	1
	<i>Excessively</i>	1
	<i>Less</i>	1
	<i>Little</i>	2
	<i>Little predictive</i>	1
	<i>Markely</i>	1
	<i>Never</i>	1
	<i>Not</i>	17
	<i>Not previously</i>	1
	<i>Often</i>	2
<i>Only</i>	3	

**Fonte: Elaborado pelos autores**

Posteriormente, procedeu-se à avaliação da importância das relações e suas variações obtidas pela equipe de especialistas.

**Figura 4: Fragmento de questionário utilizado para avaliação pelos médicos**

Leia as possibilidades e informe em que medida atendem à necessidades no seu contexto de trabalho, de acordo com a escala: 1 corresponde a “ <b>não atende</b> ” e 5 corresponde a “ <b>atende</b> ”.	
1) Os termos apresentados são importantes no contexto de seu trabalho?	- <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> + 1 2 3 4 5
2) As informações extras que qualificam esses termos são relevantes?	- <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> + 1 2 3 4 5
3) Quais questões são mais importantes no âmbito da caracterização desses termo? Tempo, duração freqüência, grau, ou outro?	- <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> + 1 2 3 4 5

**Fonte: Elaborado pelos autores**

O resultado dessa avaliação, realizada ao longo de vinte e cinco horas, correspondeu a conjunto de relações a ser considerado para a ontologia médica. Um fragmento do questionário aplicado aos médicos, baseado em Almeida (2009b), é apresentado na FIG. 4.

## Resultados e discussão

Os resultados obtidos não são quantitativos, como na maioria dos trabalhos que envolvem processamento de linguagem natural. De fato, a pesquisa quantitativa não foi considerada para a fase atual do projeto, mas será objeto de trabalhos futuros. Acredita-se, entretanto, que algumas combinações de resultados são úteis para determinar o encaminhamento do projeto. No momento, a principal contribuição é prover alternativas aos profissionais envolvidos na construção de ontologias. No restante da presente seção, apresentam-se considerações e resultados obtidos.

Na análise de resultados, em um primeiro momento, buscou-se uma combinação de caráter quantitativo, de forma a obter algum indicativo da importância desse ou daquele verbo bem como de suas variações. Frente ao grande volume de dados obtidos no resultado, tal análise se mostrou impraticável sem a consideração de técnicas estatísticas utilizadas processamento de linguagem natural. O Quadro 6 apresenta um esquema de organização por freqüência, que ilustra as tentativas de compilar resultados por via quantitativa.

**Quadro 6: Verbos com maior número de modos e freqüências correspondentes**

Verbo	Número de modos	Advérbio / Freqüência	Tipo
-------	-----------------	-----------------------	------

<i>TO BE</i>	194	<i>Usually</i>	21	Tempo-F
		<i>Also</i>	20	Grau
		<i>More</i>	18	Grau
		<i>Generally</i>	16	Tempo-F
		<i>Frequently</i>	10	Tempo-F
<i>TO HAVE</i>	34	<i>Not previously</i>	17	...
		<i>Recently</i>	04	...
		<i>Sometimes</i>	02	...
		<i>subsequently</i>	02	...
		<i>Sucessfully</i>	02	
<i>TO USE</i>	22	<i>Only</i>	08	
		<i>When</i>	06	
		<i>Instead</i>	03	
		<i>Primarily</i>	03	
		<i>Routinely</i>	03	
<i>TO OCCUR</i>	21	<i>When</i>	08	
		<i>spontaneously</i>	02	
		<i>Together</i>	03	
		<i>Almost</i>	01	
		<i>Independently</i>	01	
<i>TO GIVE</i>	16	<i>Immediately</i>	02	
		<i>Prophylatically</i>	02	
		<i>Subsequently</i>	01	
		<i>Concurrently</i>	01	
		<i>Cautiously</i>	01	

**Fonte: Elaborado pelos autores**



A alternativa adotada foi utilizar uma proposta para avaliação de ontologias centrada em usuários, solicitando-se que os especialistas fornecessem indicativos sobre a importância dos resultados obtidos. O Quadro 7 reúne um conjunto de verbos obtidos, os modos selecionados como os mais importantes pelos especialistas e uma classificação do tipo de sentença adverbial correspondente.

**Quadro 7: Amostra de verbos e modos diretamente relevantes para a prática médica**

<b>Verbo</b>	<b>Modo</b>	<b>Frequência</b>	<b>Tipo</b>
<i>ADMINISTER</i>	<i>intravenously</i>	01	Maneira
<i>APPEAR</i>	<i>serologically</i>	01	Maneira
<i>BE</i>	<i>aseptically</i>	01	Maneira
	<i>biochemically</i>	01	Maneira
	<i>biologically</i>	01	Maneira
	<i>clinically</i>	05	Maneira
	<i>genetically</i>	01	Maneira
	<i>genotypically</i>	01	Maneira
	<i>immunologically</i>	01	Maneira
	<i>phenotypically</i>	01	Maneira
	<i>serologically</i>	01	Maneira
<i>BECOME</i>	<i>bacterially</i>	01	Maneira
<i>BECOME</i>	<i>clinically</i>	02	Maneira
<i>BREATH</i>	<i>slowly</i>	01	Tempo-duração
<i>CAUSE</i>	<i>clinically</i>	01	Maneira
<i>COLLECT</i>	<i>perioperatively</i>	01	Maneira
<i>CONSIDER</i>	<i>clinically</i>	01	Maneira
<i>CONTAIN</i>	<i>negatively</i>	01	Maneira
<i>DETERMINE</i>	<i>serologically</i>	01	Maneira
<i>EXAMINE</i>	<i>macroscopically</i>	01	Maneira

<i>FIND</i>	<i>serologically</i>	01	Maneira
<i>GIVE</i>	<i>intravenously</i>	01	Maneira
	<i>orally</i>	01	Maneira
	<i>prophylactically</i>	01	Maneira
<i>HAVE</i>	<i>clinically</i>	02	Maneira
	<i>little predictive</i>	01	Maneira
	<i>physiologically</i>	01	Maneira
	<i>serologically</i>	01	Maneira
<i>INCLUDE</i>	<i>splenomegaly</i>	01	Maneira
<i>INDUCE</i>	<i>immunologically</i>	01	Maneira
<i>INJECT</i>	<i>intramuscularly</i>	01	Maneira
<i>LACK</i>	<i>serologically</i>	01	Maneira
<i>LINK</i>	<i>noncovalently</i>	01	Maneira
<i>LOCATE</i>	<i>hemostatically</i>	01	Maneira
<i>MAINTAIN</i>	<i>perioperatively</i>	01	Maneira
<i>MONITOR</i>	<i>clinically</i>	01	Maneira
<i>PROCESS</i>	<i>aseptically</i>	01	Maneira
<i>PRODUCE</i>	<i>genetically</i>	01	Maneira
<i>REACT</i>	<i>preferentially</i>	01	Maneira
<i>SELECT</i>	<i>serologically</i>	01	Maneira
<i>SHED</i>	<i>perioperatively</i>	01	Maneira
<i>SWITCH</i>	<i>intraoperatively</i>	01	Maneira
<i>TEST</i>	<i>reliably</i>	01	Grau
<i>TREAT</i>	<i>prophylactically</i>	01	Maneira
<i>USE</i>	<i>intravenously</i>	01	Maneira
	<i>prophylactically</i>	01	Maneira

<i>VIEW</i>	<i>microscopically</i>	01	Maneira
<i>WEAR</i>	<i>chemically</i>	01	Maneira

**Fonte: Elaborado pelos autores**

Uma avaliação visual dos dados do Quadro 7 indica a preponderância de modo verbal *maneira, meio ou instrumento* como relevante na visão dos especialistas. Apesar de intuitivamente esse resultado ser esperado, uma vez que os médicos vêem diferentes formas de executar uma mesma ação, observou-se que eles eram influenciados pela presença de quaisquer termos médicos e ficou clara a tendência a marcá-los como relevantes. É razoável afirmar que um modo relacionado a tempo seria importante, apesar de não ter aparecido nos resultados. Por exemplo, considerar que um processo médico deve ser executado depois de outro, ou simultaneamente, ou como pré-requisito, de forma a não causar danos à saúde de um paciente. Outras observações similares podem ser obtidas, por exemplo: alguns processos na prática médica têm locais para execução, tem durações específicas e diferentes graus de intensidade.

Os resultados e observações são indicativos de que a proposta é válida, viável, a pesquisa é significativa e relevante, mas ainda incipiente. Traçou-se um roteiro metodológico, ao qual devem ser agregados recursos adicionais para a obtenção de resultados generalizáveis. Tal proposta é apresentada nas perspectivas de trabalhos futuros, na seção seguinte.

## **Conclusão e perspectivas futuras**

O presente trabalho teve por objetivo verificar variações na semântica de relações e usar os resultados de tal verificação para a melhoria da conceitualização de ontologias. Para tanto, do ponto de vista teórico, abordou-se a semântica e as relações semânticas, enfatizando a abordagem lexical. Em seguida, apresentou-se pesquisa conduzida no âmbito de projeto da área médica de Hematologia e Hemoterapia, bem como a metodologia, os resultados e a discussão dos resultados. Finalmente, cabe oferecer considerações finais e conclusões, bem como perspectivas para trabalhos futuros. Faz-se necessário retornar às questões propostas na introdução, de forma a verificar em que medida foram atendidas.

A primeira questão dizia respeito à existência de uma variação semântica das relações analisadas. De acordo com os dados apresentados nas seções 3 e 4, conclui-se que existe variação real e significativa na semântica das relações analisadas. Essa variação foi observada ao longo das etapas da metodologia: as relações candidatas foram obtidas através de verbos disponíveis em textos médicos, tais verbos apresentavam diferentes modos, esses modos resultam em especificações distintas para relações na ontologia, ainda que representadas pelo mesmo verbo.

A segunda questão tinha por objetivo identificar se a variação semântica das relações entre verbos impactaria em seu uso como um termo candidato à relação na ontologia. A resposta à questão é positiva e decorre da resposta à primeira questão, além da confirmação obtida pela avaliação dos médicos. A variação semântica percebida impacta em seu uso na ontologia, uma vez que um verbo pode ter diferentes usos. Do ponto de vista prático, cabe refletir como essas diferenças semânticas serão passíveis de representação em editores de ontologias.

A terceira questão dizia respeito às relações *ad hoc* obtidas, indagando se seriam adequadas para uso em ontologias. Acredita-se que as relações obtidas de forma *ad hoc* são relevantes para o uso em ontologias. Entretanto, discernir em que medida não é tarefa trivial. O aparato computacional, representado aqui pelo GATE é importante para manipular o grande volume de dados, mas conforme mencionado, a etapa de conceitualização tem caráter eminentemente intelectual e o apoio automático é limitado. Determinar qual verbo é passível de uso como relação na ontologia remete à avaliação humana.

O tipo de distinção semântica obtida na pesquisa permite que o responsável pela construção de ontologias refine a representação do domínio do conhecimento observado. Caso verbos sejam extraídos de forma automática sem considerações sobre seus modos, diversas ações relevantes da prática médica não seriam consideradas na ontologia. A proposta aqui apresentada, dessa forma, contribui para refinar os processos que compõem a atividade de construção de ontologias médicas.

Finalmente, ainda que as questões propostas tenham sido abordadas, conclui-se que restam mais perguntas do que respostas. Ainda assim, acredita-se que será obrigatória a abordagem multidisciplinar entre CI e Lingüística, considerando o Processamento de Linguagem Natural, muitas vezes associado à Estatística. Nesse contexto, torna-se claro que a contribuição do trabalho é intermediária no contexto do projeto e depende da realização de trabalhos futuros complementares.

Ainda assim, cabe destacar o esquema linguístico desenvolvido para programação do GATE, que se apresenta como uma real contribuição e um bom modelo para uso em trabalhos futuros. Dentre outras possibilidades, algumas questões para investigação futura podem ser citadas de imediato: qual a granularidade ideal de um termo representativo de uma relação, para que ele seja adequado ao uso em ontologias? Qual verbo é passível de ser aproveitado como relação na ontologia e por quê? Como seriam os esquemas considerando-se a gramática portuguesa? Além disso, é preciso determinar amostras quantitativas e considerar o uso de outros esquemas linguísticos. Essas e outras questões serão objeto de etapas futuras do projeto.

Artigo recebido em 24/05/2010 e aprovado em 08/07/2010

## Referências

ALMEIDA, M.B. *Um modelo baseado em ontologias para a representação da memória organizacional*. 2006. 321p. Tese (Doutorado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

ALMEIDA, M.B.; SOUZA, R. R.; COELHO, K.C. Ontologias aplicadas a modelos de sistemas de informação. *Anais EDBICIC 2009*. Coimbra, Portugal. 2009.

ALMEIDA, M. B. Desenvolvimento de metodologia para avaliação de relações em ontologias no escopo do projeto *Open Biomedical Ontologies*. Projeto de Pós-Doutorado *State University of New York at Buffalo*. 2009a.

ALMEIDA, M. B. A proposal to evaluate ontology content. *Applied Ontology*. vol.4, n. 3 / 4, p. 245-265. 2009b.

AUSTIN, J.L. Locutinary, illocutinary, perlocutinary. In: HARNISH, R. *Basic topics in the philosophy of language*. New Jersey: Prentice Hall, p.30-39. 1962.

CANÇADO, M. *Manual de semântica: noções básicas e exercícios*. Belo Horizonte: Editora UFMG, 2005.

CHOMSKY, N. *Aspects of the theory of syntax*. Cambridge: MIT Press, 1965.

CHOMSKY, N. *Knowledge of language: its nature, origin, and use*. New York: Praeger, 1986.

CORRADI, F. M.; SOUZA, J. R. C. de; TRAVASSOS, L. C. P.; DIAS, R. de F. Nós, links, e redes... *Revista de Biologia e Ciências da Terra*, v.1, n.1, 2001. Disponível em:<[http://eduep.uepb.edu.br/rbct/sumarios/sumario\\_v1\\_n1.htm](http://eduep.uepb.edu.br/rbct/sumarios/sumario_v1_n1.htm)>. Acesso em: 28 out. 2008.

CROFT, W.; CRUSE, D.A. *Cognitive linguistics*. New York: Cambridge University Press, 2004.

CRUSE, D.A. *Lexical semantics*. New York: Cambridge University Press, 1986.

DUCROT, O. *O dizer e o dito*. Campinas: Pontes, 1987.

EVENS, M.W. *Relational models of the lexicon: representing knowledge in semantic networks*. Cambridge: Cambridge University Press, 1988.

FELLBAUM, C.; MILLER, G.A. Folk psychology or semantic entailment? Comment on Rips and Conrad (1989). *Psychological Review*, v.97, n.4, p.565-570. 1990.

FELLBAUM, C. English verbs as a semantic net. *International Journal of Lexicography*, v.3, n.4, p.278-301, 1990.

FELLBAUM, C.; CHAFFIN, R. Some principles of the organization of the verb lexicon. *Proceedings of the 12 th Annual Conference of The Cognitive Science Society*, 1990, p.420-428. Hillsdale: Erlbaum. 1990.

FELLBAUM, C. (Ed.). *Wordnet: an electronic lexical database*. Massachusetts: The MIT Press, 1998.

FELLBAUM, C. On the semantics of troponymy. In: GREEN, R.; BEAN, C. A.; MYAENG, S.H. *The semantic of relationships; an interdisciplinary approach*. Dordrecht: Kluwer, 2002. p.23-34.

FERNANDEZ-LOPEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, H. *Methontology: from ontological art towards ontological engineering*. 1997. Disponível em: <<http://citeseer.ist.psu.edu/context/544607/0/>>. Acesso em: 20 jul. 2005

GENESERETH, M. R.; NILSSON, N. J. *Logical foundation of artificial intelligence*. California: Morgan Kaufmann, 1987.

- GREEN, R. Relationships in the organization of knowledge: an overview. In: BEAN, A.; GREEN, R (Eds.) *Relationships in the organization of knowledge*. Dordrecht: Dordrecht: Kluwer Academic, 2001. p. 3-18
- GREEN, R.; BEAN, C. A.; MYAENG, S.H. Introduction. In: GREEN, R.; BEAN, C. A.; MYAENG, S.H. (Eds.) *The semantic of relationships; an interdisciplinary approach*. Dordrecht: Kluwer Academic, 2002. p.vii – xvi.
- GUARINO, N. *Formal ontology in information systems*.(1998). Disponível na Internet em: <<http://citeseer.ist.psu.edu/guarino98formal.html>>. Acesso: 03 janeiro de 2002.
- HJØRLAND, B. Semantics and knowledge organization. *ARIST*, v.41, cap.8, p.367-405, 2007a.
- HJØRLAND, B. *Semantic relations* (meaning relations). 2007. Disponível em: <[http://www.db.dk/bh/Lifeboat\\_KO/CONCEPTS/semantic\\_relations.htm](http://www.db.dk/bh/Lifeboat_KO/CONCEPTS/semantic_relations.htm)>. Acesso em: 16 jun. 2008. 2007b.
- ILARI, R., GERALDI, J.W. *Semântica*. São Paulo: Ática, 1985.
- JACKENDOFF, R. *Semantics and cognition*. Cambridge: The MIT Press, 1983.
- JONES, S. *Antonymy; a corpus-based perspective*. London: Routledge. 2002.
- KEMPSON, R. *Semantic theory*. Cambridge: Cambridge University Press, 1977.
- KHOO, C.S.G.; NA, J. Semantic relations in Information Science. *Annual Review of Information Science and Technology*, v.40, p.157-228, 2006.
- LAKOFF, G. *Women, fire and dangerous things: what categories reveal about the mind*. Chicago: University of Chicago Press, 1987
- LEECH, G., SVARTVIK, J. *A communicative grammar of English*. Essex: Longman, 1990.
- LEVIN, B. *English verb classes and alternations: a preliminary investigation*. Chicago: University of Chicago Press, 1993.
- LIMA, G. A. B. Interfaces entre a Ciência da Informação e a Ciência Cognitiva. *Ciência da Informação*, v.32, n.1, p.77-87, jan./abr.2003.
- LYONS, J. *Semantics*. 2 vols. London: Cambridge University Press. 1977.
- MENDONÇA, E. S. A lingüística e a ciência da informação. *Ciência da Informação*, v.29, n.3, p.50-70, set./dez.2000.
- MILLER, G.A, JOHNSON-LAIRD, P.N. *Language and perception*. Cambridge: Harvard University Press, 1976.
- MILLER, G. A. Nouns in WordNet. In: FELLBAUM, C. (Ed.). *Wordnet: an electronic lexical database*. Massachusetts: The MIT Press, 1998.p. 23-46.
- OLIVEIRA, R. P. de. *Semântica formal: uma breve introdução*. Campinas: Mercado das Letras, 2001.



- ORRICO, E.G.D. Interdisciplinaridade: ciência da informação e lingüística. In: PINHEIRO, L.V.R. (Org.). *Ciência da informação, ciências sociais e interdisciplinaridade*. Brasília : IBICT, 1999. p. 143-154.
- PUSTEJOVSKY, J. *The generative lexicon*. Cambridge:MIT Press. 1995.
- RECTOR, M.; YUNES, E. *Manual de semântica*. Rio de Janeiro: Ao Livro Técnico, 1980.
- SILVA, A. S. da. A lingüística cognitiva: uma breve introdução a um novo paradigma em lingüística. *Revista Portuguesa de Humanidades*, v.1, 1997. Disponível em: <http://www.facfil.ucp.pt/lingcognit.htm>. Acesso em: 10 set. 2008.
- SMITH, B. *Ontology and information systems*. 2003. Disponível em: <[http://www.ontology.buffalo.edu/ontology\(PIC\).pdf](http://www.ontology.buffalo.edu/ontology(PIC).pdf)>. Acesso em: 16 abr. 2009.
- SOUZA, R. R, ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. *Ciência da Informação*, v.33, n.1, p.132-141, jan./abr. 2004.
- TALMY, L. Lexicalization patterns: semantic structure in lexical forms (v.3). In: SHOPEN, T. *Language typology and syntactic description: grammatical categories and the lexicon* (vol.3). London: Cambridge University Press. 1990. p. 57-149.
- TUGGY, D. Schematicity. In: GEERAERTS, D., CUYCKENS, H. (Ed). *The Oxford handbook of cognitive linguistics*. New York: Oxford University Press, 2007. p.82-116.
- Van RIJSBERGEN, C. J. A new theoretical framework for information retrieval. *Proceedings of the annual international ACM SIGIR Conference on research and development in information retrieval*, 1986, p.194-200.
- VENTURA, P.; BRITO-MENDES, C.; MORAIS, J.; KOLINSKY, R. *A organização da memória semântica*. Lisboa: Fundação Calouste Gulbenkian, 2002.
- VICKERY, B. C. Ontologies. *Journal of Information Science*, v.23, n.4, p.227-286, jan. 1997.
- VINCE, M. *Advanced Language Practice*. New York: Macmillan, 2009.
- YULE, G. *Oxford Practice Grammar Advanced*. Oxford: Oxford University Press, 2009.
- WAND, Y.; WEBER, R. Mario Bunge's ontology as a formal foundation for information systems concepts. In: *Studies on Mario Bunge's treatise*. Amsterdam: Rodopi, 1990.
- WAND, Y.; STOREY, V.C.; WEBER, R. An ontological analysis of the relationship construct in conceptual modeling. *ACM Transactions on Database Systems*, v.24, n.4, 1999. p.494-528.
- WINSTON, M.E.; CHAFFIN, R.; HERRMANN, D. A taxonomy of part-whole relations. *Cognitive Science*, v.11, n.4, p.417-444, 1987. Disponível em: <<http://www.cogsci.rpi.edu/CSJarchive/1987v11/i04/p0417p0444/MAIN.PDF>>. Acesso em: 26 maio 2008.