



Universidade do Minho
Instituto de Letras e Ciências Humanas

Catarina Ribeiro de Sousa | **Extração semiautomática de Hiperónimos e Merónimos**

Catarina Ribeiro de Sousa

Extração semiautomática de Hiperónimos e Merónimos

Uminho | 2013

Outubro de 2013



Universidade do Minho
Instituto de Letras e Ciências Humanas

Catarina Ribeiro de Sousa

Extração semiautomática de Hiperónimos e Merónimos

Tese de Mestrado
Mestrado em Linguística Portuguesa e Comparada

Trabalho efetuado sob a orientação de
Prof. Doutor Álvaro Iriarte Sanromán
e do
Prof. Doutor Alberto Manuel Brandão Simões

Aos meus pais

Agradecimentos

Ao Professor Doutor Álvaro Iriarte, pelo incentivo na escolha do tema e por todas as dicas e correções.

Ao Doutor Alberto Simões, pelo trabalho informático, pelas dicas, comentários, correções, paciência e tempo despendido.

Aos meus pais, pela importância que têm e sempre tiveram na minha vida e por sempre me terem incentivado e proporcionado condições para concluir este trabalho.

Aos meus irmãos, Joana e Daniel, e ao meu namorado, Filipe, por todo o amor, amizade e companheirismo.

A toda a minha família cuja presença na minha vida torna tudo mais agradável.

Resumo

A linguagem usada nos dicionários é uma linguagem tipicamente controlada, o que contribui para que o lexicógrafo escreva as entradas de uma forma rápida e coerente. Muitas destas construções definem relações entre diversos conceitos. Dado o uso consistente destas construções, torna-se possível a utilização de padrões para a extração de relações entre conceitos.

Este trabalho tem como objetivo a análise e a avaliação de um conjunto de padrões de *Hearst* para a extração semiautomática de hiperónimos/hipónimos e de merónimos/holónimos de um dicionário eletrónico *online*, o Dicionário Aberto. Em primeiro lugar, um conjunto de padrões foi definido e aplicado ao dicionário. As relações extraídas foram avaliadas manualmente, a fim de testar a sua produtividade e precisão. Os padrões utilizados para a extração de hiperónimos/hipónimos resultaram numa precisão de 96,30%, ao passo que os padrões utilizados para a extração de merónimos/holónimos resultaram numa precisão de 76,41%.

Abstract

Dictionaries use a controlled language for word definitions. This helps the lexicographer to write dictionary entries faster, and more coherently. Some of these constructions define relationships between different concepts. Given the consistent use of these constructions it becomes possible to use patterns for the extraction of relationships between concepts.

This work aims on the analysis and evaluation of a set of *Hearst* patterns for the semi-automatic extraction of hyperonyms/hyponyms and meronyms/holonyms from an on-line electronic dictionary, Dicionário Aberto. First, a set of patterns were defined and applied to the dictionary. The extracted relationships were evaluated manually, in order to test their productivity and precision. The patterns used for the hyperonyms/hyponyms relationship extraction resulted in 96,30% of precision, whereas the meronyms/holonyms relationship extraction resulted in 76,41% of precision.

Índice

1. Introdução	1
2. Relações lexicais	3
2.1. Hiponímia/hieronímia	3
2.1.1. Os co-hipónimos	5
2.1.2. Identificar hipónimos	6
2.2. Meronímia/holonímia	8
2.2.1. Lyons (1980)	8
2.2.2. Cruse (1986)	10
2.2.3. Winston, Chaffin e Herrmann (1987)	13
3. Extração de relações do DA	17
3.1. O Dicionário Aberto.....	17
3.2. Padrões de Hearst	18
3.3. Padrões utilizados	21
3.3.1. Para hiponímia/hiperonímia	21
3.3.2. Para meronímia/holonímia	23
3.4. Outros padrões	24
4. Avaliação dos resultados obtidos	27
4.1. Método	27
4.2. Resultados e avaliação	28
4.3. Últimas considerações	41
5. Conclusão	43
Referências bibliográficas	45

Índice de figuras

Figura 1: representação hierárquica de hipónimos de “animal”	5
---	---

Índice de tabelas

Tabela 1: tipos de meronímia central e tipos de meronímias periféricas.....	12
Tabela 2: caracterização de cada tipo de meronímia de Winston <i>et al.</i> quanto à funcionalidade, à homomeria e à separabilidade.....	14
Tabela 3: padrões utilizados não avaliados.....	25
Tabela 4: resultados do padrão 1.....	28
Tabela 5: exemplos de extrações do padrão 1	28
Tabela 6: resultados do padrão 2	29
Tabela 7: exemplos de extrações corretas do padrão 2	29
Tabela 8: exemplos de erros gerados pelo padrão 2	30
Tabela 9: resultados de “outros problemas” do padrão 2	31
Tabela 10: exemplos de “dúvidas do autor” com o padrão 2	32
Tabela 11: exemplos de “combinações fixas”	33
Tabela 12: exemplos de “incerteza”	34
Tabela 13: resultados do padrão 3.....	35
Tabela 14: erro gerado pelo padrão 3	35
Tabela 15: exemplos de extrações do padrão 3	36
Tabela 16: resultados do padrão 4	36
Tabela 17: exemplos de “dúvidas do autor” com o padrão 4	37
Tabela 18: exemplos de erros gerados pelo padrão 4	37
Tabela 19: exemplos de extrações corretas do padrão 4	38
Tabela 20: resultados do padrão 5	38
Tabela 21: exemplos de erros gerados pelo padrão 5	39
Tabela 22: exemplos de extrações corretas do padrão 5	39
Tabela 23: resultados do padrão 6	40
Tabela 24: erros gerados pelo padrão 6	40
Tabela 25: exemplos de relações bem extraídas pelo padrão 6	41
Tabela 26: comparação dos resultados por relação lexical.....	42

Lista de abreviaturas

DA	Dicionário Aberto
PLN	Processamento da linguagem natural
S	Substantivo
hip	Hipónimo
hiper	Hiperónimo
N	Nome
Adj	Adjetivo

1. Introdução

Este trabalho enquadra três áreas de estudos estritamente relacionadas: a Linguística Computacional, a Lexicografia e a Semântica Lexical.

A informação a que o mundo tem acesso é cada vez mais abundante e, conseqüentemente, torna-se mais difícil para as pessoas alcançá-la e assimilá-la.

A informática veio ajudar nessa tarefa, armazenando e selecionando alguma dessa informação de várias formas, nomeadamente, através de léxicos computacionais que, por sua vez, podem assumir a forma de dicionários, ontologias, *wordnets*, dicionários de tradução, etc. Este tipo de léxicos tem como objetivo auxiliar o ser humano a encontrar e processar mais rapidamente determinada informação. Para isso, é necessário fornecer à máquina a informação necessária que lhe permita realizar a tarefa pretendida. É, então, que surge o Processamento da Linguagem Natural – disciplina da Inteligência Artificial – que tem o duplo objetivo de permitir que a máquina possa compreender o Homem, mas também de ajudar o Homem a compreender a sua própria linguagem.

Diana Santos (2001) define PLN como:

“ (...) a utilização de conhecimentos sobre a língua e a comunicação humana, tanto para a comunicação com sistemas computacionais como para melhorar a comunicação entre os seres humanos.”

É nesta base que se torna possível fazer a extração automática, ou semiautomática, de relações lexicais – ou léxico-conceptuais¹ – a partir de *corpora* disponíveis em formatos eletrónicos.

Até ao presente, não são muitos, em Portugal, os trabalhos neste âmbito.

O recurso por excelência com informação sobre relações lexicais ou léxico-conceptuais, é a WordNet de Princeton (Miller, 1995) e (Fellbaum, 1998), que constitui uma base de dados lexical e funciona como que uma enciclopédia de relações léxico-

¹ Embora se empregue com mais frequência a expressão relações lexicais, estas são também denominadas relações léxico-conceptuais porque relacionam, além de palavras, conceitos.

conceptuais. Similar para a língua portuguesa, encontra-se a WordNet.PT (Marrafa 2001). Este recurso é bastante fiável, por ter sido construído manualmente, mas por esta mesma razão, relativamente limitado. Ao mesmo tempo, não se encontra disponível livremente, mas apenas para consulta na Internet.

Com o objetivo de colmatar estas lacunas surgiram alguns projetos similares, como o PAPEL (Gonçalo Oliveira *et al.*, 2008) que constitui um recurso lexical para o português, público e gratuito, composto por relações lexicais extraídas automaticamente do Dicionário PRO da Língua Portuguesa da Porto Editora. Este recurso dispõe de diversos tipos de relações lexicais, entre as quais, hiponímia/hiperonímia, meronímia/holonímia, causa, finalidade, local e sinonímia.

Posteriormente, surge o CARTÃO (Gonçalo Oliveira *et al.* 2011), aumentando em 70% o PAPEL. Na sua construção foram usadas as mesmas regras utilizadas na construção do PAPEL, mas teve como base três dicionários eletrónicos: para além do Dicionário Pro da Língua Portuguesa foram adicionados o Dicionário Aberto (Simões e Farinha, 2011) e o Wikcionário.

Ambos os recursos – o PAPEL e o CARTÃO – estão incorporados na Onto.PT² (Oliveira, 2013) – projeto que visa a construção automática de uma ontologia lexical para o português.

Neste trabalho, não se pretende elaborar uma ontologia lexical nem tão pouco uma enciclopédia de relações léxico-conceptuais, mas antes tentar perceber que tipo de regras ou padrões podem ser utilizados para a extração de relações lexicais de hiponímia/hiperonímia e de meronímia/holonímia em alternativa à extração manual, e, ao mesmo tempo, contribuir para melhorar a informação disponibilizada no Dicionário Aberto.

O próximo capítulo expõe os pressupostos teóricos deste trabalho, nomeadamente caracteriza as relações lexicais. Posteriormente será apresentado o conceito de padrão de *Hearst* e apresentados os padrões escolhidos para a extração de relações a partir do Dicionário Aberto. No capítulo 4, é feita a avaliação de produtividade e precisão dos padrões escolhidos. Finalmente, são tecidas algumas conclusões.

² Disponível em ontopt.dei.uc.pt

2. Relações lexicais

“Só pode determinar-se o valor de uma palavra definindo-a em relação com o valor das palavras vizinhas em contraste. Só tem sentido como parte do todo, porque só no campo é que tem significado.”
(Jost Trier *apud* Lyons, 1980,236)

O vocabulário de uma língua é composto por lexemas organizados que são a principal ferramenta utilizada pelos seres humanos para comunicar. Para esta tarefa ser realizada com sucesso, os lexemas têm de respeitar a morfologia, a sintaxe, a semântica e a pragmática de determinada língua. Caso contrário, não passam de palavras soltas carentes de sentido. Ao ocorrerem de forma diversa, em lugares diversos e em situações diversas, dependentemente das outras palavras com que coocorrem, as mesmas palavras provocam diferentes sentidos e efeitos, pois estão interrelacionadas – as palavras estabelecem relações com outras palavras.

Em finais do século XIX, surgiu a Semântica Lexical, enquadrada na Semântica Estrutural. Neste contexto, muitos estudiosos debruçaram-se sobre as relações entre o léxico e o significado e muitas teorias foram propostas, tais como a Teoria dos campos semânticos (Trier e Wittgenstein), a Teoria do Léxico generativo (Pustejovsky), e a Teoria das representações mentais (Jerry Fodor), (*apud* Brauner, 2005), que não serão aqui citadas por não se mostrarem relevantes para o trabalho. Neste trabalho, limitar-me-ei a estudar as relações lexicais hiponímia/hiperonímia e meronímia/holonímia.

2.1. Hiponímia/Hiperonímia

Os conceitos de hiperonímia/hiponímia referem-se à relação que se estabelece entre um conceito mais geral e um mais específico, tendo o segundo sempre as características do primeiro, mas o primeiro não tem todas as características do segundo. Por exemplo, “cão” é hipónimo de “animal” e “animal” é hiperónimo de “cão”. Tal significa que um cão tem as características de um animal e outras que o distinguem dos outros animais, mas um animal não tem as características particulares de um cão.

“A hiponímia é uma relação transitiva. Se x é hipónimo de y e y é hipónimo de z , então x é hipónimo de z ” (Lyons, 1980, 275). Imaginemos, então, que x corresponde a “pastor alemão”, y corresponde a “cão” e z corresponde a “animal”. Assim, “pastor alemão” é hipónimo de “cão” e de “animal”, “cão” é hipónimo de “animal” e hiperónimo de “pastor alemão” e “animal” é hiperónimo de “cão” e de “pastor alemão”.

A hiponímia não ocorre apenas em substantivos, pode também ocorrer em adjetivos, advérbios e verbos. Por exemplo, o adjetivo ‘obeso’ é hipónimo do adjetivo ‘gordo’ e o verbo ‘comprar’ é hipónimo do verbo ‘obter’. Uma pessoa obesa é, obrigatoriamente, gorda, mas uma pessoa gorda não é necessariamente obesa. O mesmo acontece com ‘comprar’ e ‘obter’. Se ‘gordo’ e ‘obter’ fossem hipónimos, respetivamente, de ‘obeso’ e ‘comprar’, a relação estabelecida não seria de hiponímia/hiperonímia, mas sinonímia ou quase sinonímia, que na Wordnet.PT³ estão definidas da seguinte forma:

Sinonímia: “*duas expressões são sinónimas num contexto linguístico C se a substituição de uma por outra em C não altera o valor de verdade de C.*”

Quase sinonímia: “*se A e B são quasi-sinónimos, então, (i) um A é uma espécie de B e (ii) um B é uma espécie de A.*”

Importa, aqui, expor estas relações porque, como se define em “sinonímia”, “(...) *num contexto linguístico C (...)*”, dependendo do contexto, relações de hiponímia/hiperonímia podem corresponder a sinonímia. Por exemplo, imagine-se que um indivíduo, o João, é obeso. As frases: *O João é gordo* e *O João é obeso* conservam os valores de verdade. Portanto, se a substituição de uma expressão permite que a frase se mantenha verdadeira, então, neste contexto, *gordo* e *obeso* não são hiperónimo e hipónimo, mas sinónimos.

Ainda na Wordnet.PT, o par hiperónimo/hipónimo está definido através das seguintes paráfrases: “A é hiperónimo de B, se A não é um (tipo de) B e B é um (tipo de) A”; “A é hipónimo de B, se A é um (tipo de) B e B não é um (tipo de) A.” Partindo destas definições, a principal característica da relação hiponímia/hiperonímia reside na sua falta de simetria, ao contrário do que acontece na sinonímia.

³ Disponível em <http://www.clul.ul.pt/clg/wordnetpt/>

2.1.1. Os co-hipónimos

Segundo a linguagem lógica, se A e B são hipónimos de C, então A e B são co-hipónimos.

Sendo a relação de hiponímia/hiperonímia uma relação hierárquica de inclusão de classes, os hipónimos e seus hiperónimos encontram-se em níveis. Assim, um hipónimo tem como co-hipónimos os elementos que se encontram no seu nível, e, também os elementos a outro nível exceto os níveis que se situem acima, na mesma linha. Assim, ‘caniche’ e ‘pastor alemão’ são hipónimos de ‘cão’, então, ‘caniche’ e ‘pastor alemão’ são co-hipónimos, pois são hipónimos do mesmo heterónimo. Mas, vejamos: ‘pastor alemão’ é hipónimo de ‘cão’ que, por sua vez, é hipónimo de ‘mamífero’, e ‘mamífero’ é hipónimo de ‘animal’. Embora todos estes termos sejam hipónimos de ‘animal’, não se pode afirmar que estamos perante co-hipónimos; ‘cão’ é um hipónimo de ‘mamífero’, então não pode ser seu co-hipónimo. A figura 1 ilustra o exemplo:

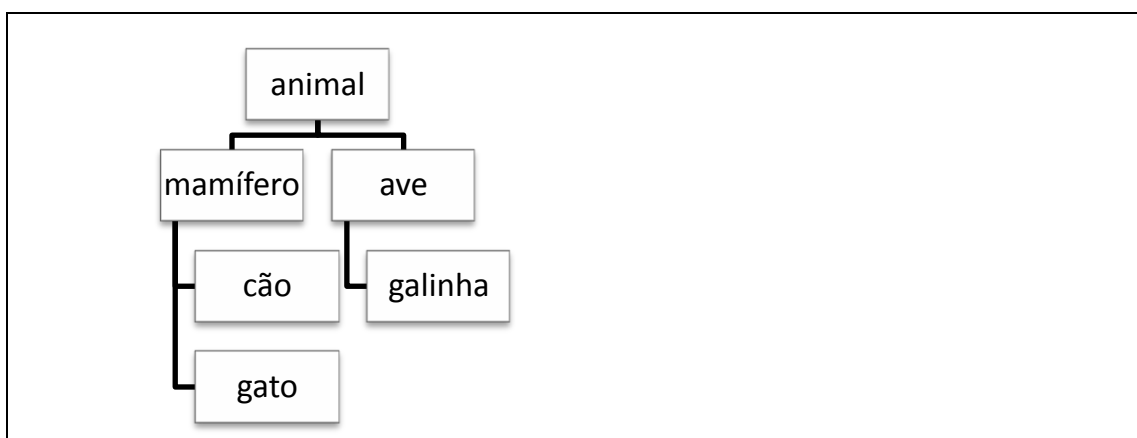


Figura 1: representação hierárquica de hipónimos de animal.

2.1.2. Identificar hipónimos

Por meio de paráfrases podemos construir testes que permitem perceber se existe a relação de hiponímia/hiperonímia entre conceitos. Os testes que apresento aqui são utilizados apenas em substantivos.

As paráfrases capazes de identificar substantivos hipónimos são compostas pelas expressões linguísticas *S é um/a S*, *S é um tipo de S*, e *S outros tipos de S*, ou apenas substituindo o hipónimo pelo seu hiperónimo. Por exemplo:

- 1) O rato^{hip} é um roedor^{hiper}.
- 2) A sapateira^{hip} é um tipo de armário^{hiper}.
- 3) Havia tigres^{hip} e outros felinos^{hiper}.
- 4) O João comprou esparguete^{hip} e outros tipos de massa^{hiper}.
- 5) Esqueci-me de dar água ao meu cão^{hip}. Quando cheguei a casa, o animal^{hiper} estava cheio de sede.

Se pegar em cada um deste par de palavras e tentar aplicar os outros testes, terei os seguintes resultados:

- a) O tigre^{hip} é um felino^{hiper}.
- b) Esparguete^{hip} é uma massa^{hiper}.
- c) A cama^{hip} é um armário^{hiper}.
- d) O cão^{hip} é um animal^{hiper}.
- e) O rato^{hip} é um tipo de roedor^{hiper}.
- f) O tigre^{hip} é um tipo de felino^{hiper}.
- g) Esparguete^{hip} é um tipo de massa^{hiper}.
- h) O cão^{hip} é um tipo de animal^{hiper}.
- i) Havia ratos^{hip} e outros roedores^{hiper}.

- j) Vendeu a sapateira^{hip} e outros armários^{hiper}.
- k) Comprei esparguete^{hip} e outras massas^{hiper}.
- l) Vi cães^{hip} e outros animais^{hiper}.
- m) Tenho medo a ratos^{hip} e outros tipos de roedores^{hiper}.
- n) Nessa loja, há sapateiras^{hip} e outros tipos de armários^{hiper}.
- o) Ele tem tigres^{hip} e outros tipos de felinos^{hiper}.
- p) Quero ter cães^{hip} e outros tipos de animais^{hiper}.
- q) Tenho um rato^{hip} na garagem. O roedor^{hiper} dá-me cabo de tudo.
- r) Já pus a sapateira^{hip} à venda há uma semana. Ainda ninguém quis comprar o armário^{hiper}.
- s) O Mário tem um tigre^{hip} em casa. O felino^{hiper} é muito feroz.
- t) Cozi a esparguete^{hip} em pouca água. A massa^{hiper} ficou toda pegajosa.

Só em modo de comparação, utilizo os mesmos testes, mas para um par de unidades lexicais que não estabelece a relação de hiponímia/hiperonímia – ‘chocolate’, ‘peixe’.

- i) #O chocolate é um peixe.
- ii) #O chocolate é um tipo de peixe.
- iii) #Adoro chocolate e outros peixes.
- iv) #Comprei chocolate e outros tipos de peixe.
- v) #O chocolate ficou ao sol. Quando cheguei ao carro, o peixe estava derretido.

Tendo em conta a aceitabilidade das frases (a – t) e a anomalia semântica das frases (i – v), podem ser utilizados estes testes de cada vez que for pretendido perceber se há, ou não, hiponímia/hiperonímia entre unidades lexicais. No entanto, devido à idiossincrasia de algumas, em alguns casos, é mais aceitável a presença do lexema ‘tipo/s de’ e noutros a sua ausência.

2.2. Meronímia/Holonímia

O termo meronímia foi proposto por Cruse, em 1986. Até então, a relação meronímia/holonímia era denominada parte-todo. Esta relação estruturadora do léxico relaciona as partes (merónimos) com o seu todo (holónimo). Para facilitar a compreensão, pense-se no corpo humano; as mãos, os pés, a cabeça *e as outras partes do corpo* são merónimos de corpo. Este exemplo apresenta apenas um tipo de meronímia, no entanto esta é uma relação lexical bastante complexa e nem sempre fácil de definir.

Neste capítulo, serão abordados, de forma não exaustiva, três trabalhos relacionados com o estudo da relação parte-todo, doravante, meronímia/holonímia. São eles o de Lyons (1980), o de Cruse (1986) e o de Winston, Chaffin e Herrmann (1987).

2.2.1. Lyons (1980)

No seu trabalho, Lyons preocupa-se não tanto em definir a relação parte-todo (colocar nota de rodapé: utiliza-se aqui parte-todo em vez de meronímia/holonímia, tendo em conta que era assim que Lyons definia a relação), mas em diferenciá-la de hiponímia/hiperonímia e em discutir a sua transitividade.

O autor considera que a relação entre objetos físicos discretos é claramente transitiva, mas afirma pode dar-se transitividade também em zonas do espaço físico e propõe os seguintes raciocínios:

*“Se uma coisa X forma parte de uma coisa Y
que, por sua vez, forma parte de uma coisa*

Z, então X pode sempre descrever-se como uma parte de Z.”

“Se X é um ponto ou uma zona que forma parte de uma zona Z, então X forma parte de Z.”

No entanto, para Lyons, as relações de transitividade que se estabelecem no vocabulário não são as mesmas que se estabelecem no mundo, afirmando que o facto de uma entidade poder ser descrita como parte de outra entidade não significa que se estabeleça necessariamente uma relação parte-todo entre os lexemas que referem essas entidades. Para ilustrar esta distinção utiliza como exemplo os pares ‘maçaneta, porta’; ‘porta, casa’; ‘maçaneta, casa’. Através dos raciocínios anteriormente propostos, se *a maçaneta é uma parte da porta e a porta é uma parte da casa, uma maçaneta é parte da casa*. Mas acha estranha a frase *a maçaneta da casa* e completamente inaceitável *a casa não tem maçaneta*, embora a maçaneta faça realmente parte das entidades existentes numa casa. Daí a distinção entre a transitividade para o mundo e a transitividade para o léxico. Por outro lado, com os pares ‘punho, manga’; ‘manga, casaco’; ‘punho, casaco’ já considera aceitáveis frases como *“os punhos do casaco”* ou *“o casaco não tem punhos”*, pois o vocabulário já permite a sua aceitabilidade. Como a transitividade é dependente de fatores linguísticos, prefere classificar a relação lexical parte-todo como uma relação lexical não transitiva.

Relativamente à distinção entre hiponímia e relação parte-todo, o autor contraria a ideia de que tal distinção é suficientemente clara no caso de nomes que denotam objetos físicos discretos, afirmando que esses são, por certo, os que geram mais discussões linguísticas. Para exemplificar, utiliza o conceito de ‘ouro’ que tanto pode ser uma parte de um todo como um hipónimo e formula orações como, *“esta substância contém ouro”* ou *“esta substância é composta por ouro (e outros metais)”* em que ‘ouro’, na primeira oração, corresponde a uma parte de um todo e, na segunda, ainda como parte de um todo, mas também como um tipo metal (hipónimo). Defende que, ao passo que estas orações são perfeitamente aceitáveis, constituem anomalia semântica as seguintes: *“#este animal é composto de vaca (e outros mamíferos)”* ou *“#este corpo é braço”*.

Estas dificuldades de distinção não ocorrem apenas nos casos de nomes concretos, mas também em nomes abstratos e em verbos. Como tal, a ‘honra’ tanto pode ser um tipo de virtude como uma parte da virtude. E o verbo ‘coser’ implica a inclusão de um conjunto de outros verbos como ‘alinhar’ ou ‘abainhar’, aqui como hipónimos de ‘coser’, que denotam também atividades que constituem partes da atividade ‘coser’.

Lyons (1980) reflete também sobre o lugar que os coletivos, como ‘gado’ ou ‘clero’, ocupam na estrutura do vocabulário. Começa, então, por afirmar que muitos coletivos “*atuam como hiperónimos para um conjunto de quase-hipónimos*”.

O coletivo ‘gado’ é hiperónimo de ‘vaca’, ‘touro’ e ‘novilho’, como o comprova a aceitabilidade de expressões como “*vacas, touros e outros tipos de gado*”. Do mesmo modo, ‘clero’ é hiperónimo de ‘bispo’ e ‘sacerdote’, mas o autor chama a atenção para o facto de ‘sacerdote’ e ‘bispo’ também estabelecerem uma relação parte-todo com ‘clero’ (cf. “*sacerdotes, bispos e outros membros do clero*”). Por outro lado, num coletivo como ‘rebanho’ face a ‘ovelha’, defende que não existe uma relação de hiponímia/hiperonímia, nem tão pouco uma relação parte-todo, dado que um coletivo como este desempenha uma função individualizadora, sendo que qualquer ovelha de um rebanho é um individual e o papel deste coletivo é *individualizar um conjunto de indivíduos indiferenciados*. Assim, associa o funcionamento deste tipo de coletivos ao funcionamento de conceitos como ‘grupo’, ‘conjunto’, ‘coleção’, entre outros, e justifica-o afirmando que dizer “um rebanho de ovelhas” é o mesmo que dizer “um grupo/um conjunto de ovelhas”, tal como acontece com muitos outros coletivos.

2.2.2. Cruse (1986)

Cruse considera que a relação meronímia/holonímia é uma relação complexa e afirma que:

“Meronym is subject to a greater number of complicating factors than taxonomic relations are; instead of there being a single clearly distinguished relation, there is in reality a numerous family of more-or-less similar relations.” (Cruse, D.A, 1986, “Lexical Semantics”, 160)

O autor estabelece uma distinção entre as relações *parte de* e *porção de*, partindo do exemplo duma máquina de escrever. Imagina, então, que se utiliza uma serra para dividir a máquina de escrever em porções, mas para obter as suas partes é necessário utilizar uma chave de fendas. Após analisar este exemplo, propõe alguns parâmetros:

- a. A relação *parte de* nem sempre é transitiva, ao passo que a de *porção de* é-o sempre;
- b. As *partes* são autónomas e as *porções* não, isto é, uma *parte* tanto pode ser utilizada noutra máquina de escrever, como pode nunca ter feito parte de nenhuma, o que não pode acontecer com uma *porção* da máquina de escrever;
- c. Os limites das partes não são arbitrários e os das porções são;
- d. As partes são funcionais e as porções não, ou seja, uma parte tem sempre uma função determinante.

A relação *porção de* não é considerada pelo autor como sendo meronímia, e separa as relações que considera verdadeiramente meronímia de outras relações próximas ou similares a ela; respetivamente, meronímia central e meronímias periféricas.

A meronímia central está relacionada com entidades físicas. Entidades abstratas, coletividades e massas referem-se a meronímias periféricas. O seguinte quadro apresenta alguns exemplos que, segundo este modelo, se classificam como meronímia central e meronímias periféricas

Merónimia Central	Merónimias periféricas	
gaveta, armário	Lugares	Portugal, Europa
folhas, livro	Tempo	puberdade, adolescência
teclado, computador	Medidas	metro, quilómetro
roda, carro	Coletividades	ovelha, rebanho
manga, casaco	Ingredientes	borracha, pneu
ponteiros, relógio	Partículas	grão, areia

Tabela 1: Tipos de merónimia central e tipos de merónimias periféricas.

De modo a verificar se determinado par de conceitos pode ser considerado merónimia, independentemente de central ou periférica, o autor propõe, inicialmente, que passem no seguinte teste:

“X é merónimo de Y se, e só se, orações com a forma um Y tem X(s) e um X é uma/Xs são parte(s) de um Y são normais quando os sintagmas nominais um X ou um Y são interpretados genericamente.”

Para exemplificar, ilustrou com dois pares de palavras, sendo que o primeiro – braço, corpo – passa no teste, constituindo merónimia, enquanto o segundo – mãe, filho – passa na primeira parte, mas não passa na segunda, não constituindo merónimia: “um corpo tem braços” e “um braço é parte de um corpo” contra “um filho tem uma mãe” e “#uma mãe é parte de um filho”.

No entanto, este teste não dava conta de alguns tipos de merónimos, como, por exemplo, o par sépala, flor, uma vez que nem todas as flores têm sépalas. De modo a

conseguir dar conta das meronímias que não cabiam no teste inicial, Cruse incluiu um novo teste:

“As partes de um Y incluem o(s) X(s) /o(s) Z(s), etc.”

2.2.3. Winston, Chaffin e Herrmann (1987)

Partindo do princípio de que há uma vasta gama de relações de meronímia/holonímia, estes autores dividiram o que consideram pertencer a esta relação, em seis grupos principais:

1. Componente integral – objeto (pedal – bicicleta)
2. Membro – coleção (navio – frota)
3. Porção – massa (fatia – piza)
4. Material – objeto (aço – carro)
5. Fase – atividade (pagar – comprar)
6. Lugar – área (Everglades – Flórida)

Cada um destes tipos de meronímia pode ser decomposto a partir de três traços: funcionalidade, homomeria e separabilidade. Assim, uma parte pode ou não ter um papel fundamental para o funcionamento do todo (+ ou – funcional), uma parte pode ou não ser igual às outras partes que compõem o todo (+ ou – homómera) e cada parte pode ou não ser separada das outras partes e do todo (+ ou – separável).

Através destes traços, cada tipo de meronímia é caracterizado do seguinte conforme mostra o seguinte quadro⁴:

⁴ Esta tabela representa um esquema simplificado por Salvador Climent (2000).

Tipo de meronímia	Funcionalidade	Homomeria	Separabilidade
Componente integral – objeto	+	-	+
Membro – coleção	-	-	+
Porção – massa	-	+	+
Material – objeto	-	-	-
Fase – atividade	+	-	-
Lugar – área	-	+	-

Tabela 2: Caracterização de cada tipo de meronímia de Winston *et al.* quanto à funcionalidade, à homomeria e à separabilidade.

De modo a mostrar a vastidão e heterogeneidade das relações de meronímia/holonímia existentes e facilitar a sua compreensão, os autores apresentam alguns exemplos para cada um dos seis tipos:

Componente integral – objeto:

“Uma pega faz parte de um copo.”

“As rodas são peça dos carros.”

“O frigorífico é parte da cozinha.”

“Os capítulos são partes de livros.”

“A Bélgica faz parte da NATO.”

“A Fonologia faz parte da Linguística.”

Membro – coleção:

“A árvore é parte de uma floresta.”

“Um jurado é parte do júri.”

“Este navio faz parte de uma frota.”

Porção – massa:

“Esta fatia é parte de uma torta.”

“A jarda é parte de uma milha.”

“Este pedaço é parte da minha argila.”

Material – objeto:

“Um Martini é parcialmente álcool.”

“A bicicleta é, em parte, aço.”

“A água é, em parte, hidrogénio.”

“A lente é feita de vidro.”

Fase – atividade:

“Pagar é parte de comprar.”

“A ovulação é parte do ciclo menstrual.”

“O namoro faz parte da adolescência.”

Lugar – área:

“Um oásis é parte de um deserto.”

“A linha de base é parte de um campo de ténis.”

Relativamente à análise da transitividade das relações de meronímia/holonímia, os autores criam o exemplo que se tornou célebre:

“O braço de Simpson faz parte de Simson.”

(componente – objeto)

“Simpson faz parte do Departamento de Filosofia.” (membro – coleção)

#“*O braço de Simpson faz parte do Departamento de Filosofia.*”

Este exemplo ilustra a não transitividade da relação, no entanto, há muitos casos em que se verifica transitividade, como, por exemplo, no tipo ‘porção – massa’. Isto causou um conflito na tentativa de definir a meronímia/holonímia como uma relação transitiva ou não transitiva, dado que tanto pode sê-lo como não pode.

Winston *et al.* (1987) justificam a dificuldade com o facto de haver diversos tipos de meronímia e concluem que a meronímia/holonímia é, de facto, uma relação transitiva, no entanto, a transitividade só ocorre dentro de um mesmo tipo de relação meronímica.

3. Extração de relações do DA

A extração de relações lexicais pode ser manual, automática ou semiautomática. A extração manual é, sem dúvida, a mais eficaz, pois consiste na análise minuciosa do *corpus* por um ser humano, sendo possível encontrar todas as relações pretendidas. No entanto, constitui um trabalho muito moroso e os custos com a mão-de-obra tendem a ser muito elevados. A extração automática consiste no uso de aplicações computacionais, através de padrões previamente selecionados, que extrairão as relações pretendidas. Este processo permite uma maior rapidez na recolha de informação, mas, por um lado, nem sempre é capaz de detetar as ocorrências de padrões existentes no corpus que definem relações (cobertura), e, muitas vezes, detetam padrões e relações não existentes (precisão).

O que diferencia uma extração automática de uma extração semiautomática é o facto de, na extração automática, o resultado não ser sujeito a revisão, quer porque não há interesse em obter relações 100% corretas, ou porque os padrões são tão específicos que não dão lugar a grandes erros da extração. A extração semiautomática requer uma seleção e avaliação das relações extraídas que terá de ser feita por um ser humano.

O tipo de trabalho aqui feito corresponde a uma extração semiautomática. Para tal, foram selecionados padrões de *Hearst* (1992) aparentemente capazes de extrair as relações pretendidas, e passou-se à sua validação e avaliação.

Neste capítulo, começo por apresentar sucintamente o Dicionário Aberto (Simões e Farinha, 2011), donde foram extraídas as relações. De seguida falo nos padrões de *Hearst*, e apresento os padrões utilizados.

3.1. O Dicionário Aberto

O DA é um dicionário eletrónico *online* que corresponde à modernização e digitalização de um dicionário antigo com direitos de autor já expirados – *O Novo Dicionário da Língua Portuguesa de Cândido Figueiredo* (1913). Tratando-se de um dicionário antigo, a grafia é bastante diferente da que utilizamos atualmente. Como tal, é também compreensível que a sua nomenclatura seja bastante divergente da de dicionários mais recentes, pois a língua está em constante mutação e o vocabulário de

há um século era muito diferente do atual. Sucede que, ao consultarmos o DA, somos capazes de encontrar imensos termos que já caíram há muito em desuso, ao passo que muitos termos posteriores, como é o caso de ‘informática’ não são lá contemplados.

O DA não constitui ainda um projeto acabado, sendo que está a ser desenvolvido e melhorado de modo a tornar-se mais interessante quer para o utilizador comum, quer para linguistas ou pesquisadores de PLN.

Atualmente, o DA contém 128 521 entradas para as quais dispõe de algumas relações lexicais, como *sinónimo de*, *antónimo de*, *relativo a*, *merónimo de*, *holónimo de*, *hipónimo de* e *hiperónimo de*.

As relações de hiponímia/hiperonímia, assim como meronímia/holonímia não foram ainda validadas, e só estão disponíveis nos casos em que o padrão encontrou apenas uma palavra.

3.2. Padrões de Hearst

A ideia de se extraírem relações de sentido entre palavras de forma automática a partir de *corpora* surgiu com Marti Hearst (1992) que propôs a utilização de padrões léxico-sintáticos previamente escolhidos.

A autora propõe que a escolha dos padrões seja feita tendo em conta a sua frequência de ocorrência e a não exigência de muito conhecimento pré-codificado. Isto é, propõe que, depois de se escolher a relação que se pretende extrair (por exemplo, hiperonímia), se faça um estudo no *corpus* de modo a tentar encontrar padrões que se repetem várias vezes para referir uma relação estabelecida entre conceitos. Por exemplo, sabendo que ‘rosa’ é hipónimo de ‘rosácea’ verifica-se qual o padrão que dá conta dessa relação. Se durante a observação do *corpus*, o mesmo padrão aparecer várias vezes, então será um bom candidato a ser utilizado.

De modo a codificar a relação de hiperonímia para o Inglês, Hearst selecionou os seguintes padrões⁵:

⁵ Nos padrões, NP₀ corresponde ao sintagma nominal hiperónimo e NP₁, NP₂ e NP₃ correspondem a sintagmas nominais hipónimos.

1) NP₀ such as {NP₁, NP₂... , (and|or)} NP_n

Este padrão tira partido da expressão *such as* para relacionar um hiperónimo com um ou mais hipónimos, como na frase:

“*The bow lute, such as the Bambara ndanga, is plucked and has na individual curved neck for each string.*” (O alaúde de arco, tal como o Bambara ndanga, é dedilhado e tem um pescoço curvo para cada uma das cordas).

Hiperónimo: alaúde de arco (*bow lute*) **Hipónimo:** *Bambara ndanga*

2) Such NP₀ as {NP , }* {(or|and)} NP

Este padrão tem a mesma função do anterior; relacionar um hiperónimo com um ou mais hipónimos. No entanto, diferencia-se por ter o hiperónimo entre as duas unidades lexicais que formam a expressão *such as*. A frase seguinte ilustra o exemplo:

“... works by such authors as Herrick, Goldsmith, and Shakespeare.” (...escrito por tais autores como Herrick, Goldsmith, e Shakespeare)⁶.

Hiperónimo: autores (*authors*). **Hipónimos:** *Herrick, Goldsmith, Shakespeare*

3) NP { , NP}* { , } or other NP

A expressão *or other* contida neste padrão permite inferir que os sintagmas nominais que o precedem sejam hipónimos do sintagma nominal que o sucede, como demonstra a frase:

“*Bruises, wounds, brokends bones or other injuries...*” (Hematomas, feridas, ossos partidos ou outros ferimentos...)

Hiperónimo: ferimentos (*injuries*). **Hipónimos:** hematomas, feridas, ossos partidos (*bruises, wounds, broken bones*)

⁶ Numa tradução da frase para português, excluir-se-ia *tais*. Daí os padrões nem sempre serem cambiáveis para outras línguas.

4) NP { , NP}* { , } and other NP

Este padrão só difere do anterior na medida em que a conjunção contida (*and*) é aditiva em vez de disjuntiva (*or*). De resto, têm exatamente o mesmo funcionamento, como se vê na frase:

“... *temples, treasuries, and other important civic buildings.*” (... templos, tesouros, e outras construções cívicas importantes.)

Hiperónimo: construções cívicas (*civic buildings*). **Hipónimos:** templos, tesouros (*temples, treasuries*)

5) NP { , } including {NP , }* {or|and NP

Este padrão pressupõe que os sintagmas nominais que sucedem a expressão *including* estejam incluídos na classe do sintagma nominal anterior a essa expressão, como no caso da frase:

“*All common-law countries, including Canada and England...*” (Todos os países de lei comum, inclusive o Canadá e a Inglaterra...)

Hiperónimo: países de lei comum (*common-law countries*). **Hipónimos:** *Canada, Inglaterra*

6) NP { , } especially {NP , }*{or|and} NP

A expressão *especially* entre sintagmas nominais permite que o(s) que a sucede(m) se relacionem com o que a precede como hipónimos, devido ao facto de o próprio termo referir conceitos mais específicos (hipónimos). O exemplo é ilustrado pela frase:

“...*most European countries, especially France, England, and Spain.*” (... a maioria dos países europeus, especialmente França, Inglaterra, e Espanha.)

Hiperónimo: países europeus (*European countries*). **Hipónimos:** *França, Inglaterra, Espanha*

Os contributos de Hearst foram importantes para trabalhos posteriores de quem trabalha em PLN, não só para se aproveitarem os padrões utilizados pela autora, mas também para se retirarem as suas ideias e, a partir daí, se criarem outros padrões eficazes. Por isso, o conceito “padrões de Hearst” é comumente utilizado para referir a técnica de uso de padrões para extrair relações.

3.3. Padrões utilizados

O documento do qual se pretende extrair as relações tem um papel importante na escolha dos padrões a utilizar. Sendo um dicionário o documento utilizado para este trabalho, é preciso ter em conta que o pretendido é encontrar dentro da definição um conceito ou conceitos com o qual/os quais a entrada do dicionário estabeleça a relação de sentido e não entre dois conceitos contidos na definição.

A escolha dos padrões para este trabalho foi feita com base na prévia observação do *corpus*, graças a uma função do DA: a *pesquisa reversa* que permite procurar qualquer expressão contida nas definições e, assim, ver se a quantidade de ocorrências justifica testar determinado padrão.

3.3.1. Para hiponímia/hiperonímia

1) que tem por tipo [ao]s P

Este padrão⁷ foi usado para a extração da relação *hiperónimo de*. Assim, o pretendido é que o termo correspondente à entrada no dicionário seja um hiperónimo do conceito que sucede o padrão. No seguinte exemplo, será possível inferir que as *Plumbagineas* são um hiperónimo de *plumbago* (hipónimo).

⁷ Em todos os padrões, *P* representa o termo/sequência de termos com o/s qual/ais se pretende que seja estabelecida a relação.

Plumbagineas

f. pl.

Família de plantas, que tem por tipo a plumbago.

(Fem. pl. de *plumbagíneo*)

2) espécie de P

Padrão testado para obter a relação *hipónimo de*. Foi escolhido pelo facto de um hipónimo ser, por natureza, uma espécie do seu hiperónimo. A frase seguinte permite, assim, inferir que um canapé uma *Marquesa* é um hipónimo de *canapé* (hiperónimo).

Marquesa

f.

Mulher de marquês.

Senhora, que tem marquesado.

Espécie de canapé largo, com assento de palhinha.

(...)

3) género de P

Este padrão foi testado para extrair a relação *hipónimo de*. A razão da sua escolha é idêntica à do padrão “espécie de”, pois um hipónimo é também um género do seu hiperónimo. Pode ser exemplificado através da seguinte frase, na qual *Cão-do-mar* é um hipónimo de *peixe* (hiperónimo).

Cão-do-mar

m.

Género de peixes, a que pertence o tubarão e a raia.

3.3.2. Para meronímia/holonímia

4) parte de P

Este padrão foi o primeiro a ser escolhido por ser a própria definição dos merónimos; um merónimo é uma **parte de** um todo. O pretendido é que a entrada corresponda a um merónimo do conceito que o sucede. No seguinte exemplo, *Canela* é um merónimo de *perna* (holónimo).

Canela

f.

Casca odorífera de uma planta de Ceilão.

Árvore, que produz canela; caneleira.

Designação de algumas árvores, semelhantes à caneleira.

Parte da perna, entre o pé e o joelho.

(...)

5) membro de P

Este padrão foi utilizado para selecionar a relação merónimo de, dado que um membro de algo é, naturalmente, uma das partes que o constitui. Assim, no seguinte exemplo, *Juiz* representa um merónimo de *júri* e de *poder judicial* (holónimos).

Juiz

m.

Aquele que tem competência legal para tomar conhecimento das causas dos litigantes e julgá-las por sentença.

Aquele que julga; julgador.

Árbitro.

Membro do júri.

Membro do poder judicial.

(...)

6) cada um[a] d[ao]s X d[eao]s P

Este padrão foi selecionado aquando da revisão das relações extraídas, por aparecer várias vezes nas definições. Como costumava ilustrar a relação de meronímia, resolvi testá-lo. É um padrão para *merónimo de*, em que *X* corresponde a uma unidade lexical não analisada e *P* ao holónimo da cabeça do verbete. A expressão cada um(a) dos(as) representa, normalmente, partes iguais (homómeras). No seguinte exemplo é possível inferir que uma *Apanha* é um merónimo de *tear* (holónimo).

Apanha

f.
Acto de apanhar.
Colheita: a apanha da azeitona.

Prov.
Cada um dos pedais do tear.

3.4. Outros padrões

À medida que ia sendo feita a avaliação das relações extraídas, foram surgindo no *corpus* outros padrões que se mostravam também capazes de extrair relações de hiponímia/hiperonímia e de meronímia/holonímia. Não seria possível, neste trabalho, avaliar todos os padrões. Ainda assim, foi feita a sua extração para uma avaliação futura.

O quadro que se segue apresenta os padrões extraídos, a relação que a entrada infere e alguns exemplos.

Padrão	Exemplos	Relação inferida
(da) família d[eao]s P	Acantorinas: <i>f. pl.</i> <u>Família de peixes</u> , que apresentam entre os olhos uma espécie de nariz, armado de agulhão.	As “acantorinas” são um <i>hipónimo de</i> “peixes”
(da) tribo d[eao]s P	Corvídeos: <i>m. pl.</i> <u>Tribo de pássaros conirostros</u> , que têm por tipo o corvo.	Os “corvídeos” são um <i>hipónimo de</i> “pássaros conirostros”
(da) ordem d[eao]s P	Abutre: <i>m.</i> Ave de rapina, <u>da ordem das diurnas</u> . [...]	O “abutre” é um <i>hipónimo de</i> “diurnas”
(da) classe d[eao]s P	Porco: <i>m.</i> Quadrúpede <u>da classe dos mamíferos</u> e da ordem dos paquidermes.	O “porco” é um <i>hipónimo de</i> “mamíferos”
variedade de P	Barqueira: <i>f.</i> <u>Variedade de maçã</u> .	A “barqueira” é um <i>hipónimo de</i> “maçã”
que serve(m) de tipo (ao à)s P	Búrsera: <i>f.</i> Planta das Antilhas, <u>que serve de tipo às burseráceas</u> .	A “búrsera” é um <i>hipónimo de</i> “burseráceas”
a que pertence(m) o(s) P	Acalefos: <i>m. pl. Zool.</i> Classe de zoóphytos, <u>a que pertence a alforreca</u> .	Os “zoóphytos” são um <i>hiperónimo de</i> “alforreca”
composto a de P	Acordeão: <i>m.</i> Instrumento, <u>composto de paletas metálicas</u> , [...]	O “acordeão” é um <i>holónimo de</i> “peças metálicas”

Tabela 3: padrões utilizados não avaliados.

4. Avaliação dos resultados obtidos

Neste capítulo, será analisada a produtividade dos padrões apresentados no capítulo anterior, bem como uma avaliação da sua precisão. Serão também expostos os diversos tipos de erro que impedem que determinado padrão consiga fazer a extração com sucesso.

4.1. Método

Depois dos padrões terem sido propostos, o Doutor Alberto Simões implementou uma aplicação que percorre todos os verbetes do DA, um de cada vez, acedendo diretamente à base de dados. Em cada um, a aplicação experimenta todos os padrões. Para cada padrão que funciona, é guardada a cabeça do verbe (entrada), a relação referente ao padrão, e a sequência de palavras P do padrão.

A avaliação das relações extraídas assenta nas seguintes convenções:

- a. Consideram-se *hiperónimo de* ou *hipónimo de* os termos que passarem em pelo menos um dos testes descritos em 2.1.2;
- b. A análise da relação *merónimo de* é feita com base nas leituras dos trabalhos mencionados no capítulo 2, sendo os tipos de meronímia propostos por Winston *et al.* os mais levados em conta;
- c. Se o conceito extraído estabelecer a relação selecionada pelo padrão usado, considera-se a extração correta;
- d. Considera-se extração errada (ou erro) quando o conceito extraído não corresponde, de todo, à relação selecionada pelo padrão;
- e. Outros problemas, como “dúvidas do autor”, “combinações fixas” e “incerteza”, serão anotados à parte, não sendo contabilizados nem nas extrações corretas, nem nos erros;
- f. A percentagem de precisão de cada padrão será feita assumindo apenas as extrações corretas em relação ao total de relações extraídas pelo padrão em causa.

4.2. Resultados e avaliação

Nesta secção, apresentarei os resultados e a avaliação de cada padrão, tendo em conta o número de ocorrências, as extrações corretas e o tipo de erros gerados.

Padrão 1: que tem por tipo [ao]s P

Ocorrências	Extrações corretas	Erros	Precisão (%)
293	293	0	100%

Tabela 4: resultados do padrão1.

Este padrão mostrou-se 100% preciso, uma vez que, em todos os casos, a cabeça do verbete corresponde a um *hiperónimo* dos conceitos por ele extraídos. O seu ponto fraco é pouca quantidade de ocorrências, o que o torna pouco produtivo.

Seguem-se alguns exemplos de extrações obtidas aplicando este padrão:

<p><i>“Bambusáceas: f. pl. Tribo de plantas gramíneas, que tem por tipo a <u>bambusa</u>.”</i></p> <p><i>“Columbino: (...) m. pl. Família de aves, que tem por tipo a <u>pomba</u>”</i></p> <p><i>“Felino: (...) m. pl. Família de animais mamíferos, que tem por tipo o <u>gato</u>.”</i></p> <p><i>“Isópode: (...) m. pl. Ordem de crustáceos, que tem por tipo o <u>bicho de conta</u>”</i></p> <p><i>“Malváceas: f. pl. Família de plantas, que tem por tipo a <u>malva</u>”</i></p> <p><i>“Pelicanídeo: (...) M. pl. Família de aves, que tem por tipo o <u>pelicano</u>.”</i></p>

Tabela 5: exemplos de extrações do padrão 1.

Padrão 2: espécie de P

Ocorrências	Extrações corretas	Erros	Outros problemas	Precisão (%)
3824	3518	138	168	92%

Tabela 6: resultados do padrão 2.

Este padrão, além de apresentar uma elevada percentagem de precisão, mostrou-se bastante produtivo, justificado pelo elevado número de ocorrências.

O padrão foi bem-sucedido em 3518 ocorrências, das quais são apresentados alguns exemplos, na tabela seguinte.

<p><i>“Abelha-flôr: f. Designação vulgar de uma espécie de <u>orquídeas</u>.”</i></p> <p><i>“Alce: m. Espécie de <u>veado das regiões do norte</u>.”</i></p> <p><i>“Barcelos: f. Espécie de <u>videira portuguesa</u>. (...)”</i></p> <p><i>“Cabaça: f. (...) Espécie de <u>abóbora</u>. (...)”</i></p> <p><i>“Diásporo: m. Mineral raro, espécie de <u>jaspe</u>.”</i></p> <p><i>“Gralha: (...) Pl. Espécie de <u>jogo popular</u>.”</i></p>

Tabela 7: exemplos de extrações corretas do padrão 2.

Os erros gerados por este padrão devem-se a problemas de natureza sintática, em que a expressão correspondente ao padrão utilizado não tem como referente o termo que constitui a cabeça do verbete, como se verifica nos seguintes exemplos:

“Arquiviola: f. Antigo instrumento músico, que se compunha de uma espécie de cravo, (...).”

“Molariforme: adj. (...) Bot. Diz-se de certos cogumelos, que têm a superfície coberta de uma espécie de dentes.”

“Falcata: f. Arma antiga, formada de uma haste, encimada por uma espécie de foice.”

“Epipterado: adj. Bot. Diz-se do fruto ou do grão, quando provido de uma espécie de asa no seu ápice.”

“Felose: f. Produção acidental de uma espécie de cortiça em alguns vegetais.”

“Hernuto: m. Membro de uma seita religiosa, também conhecida por irmãos morávios, a qual proclama uma espécie de comunismo e se distingue pela natureza dos seus costumes.”

“Laringóstomo: adj. Zool. Diz-se do animal articulado, cuja boca é uma espécie de tromba, formada pelo esófago.”

Tabela 8: exemplos de erros gerados pelo padrão 2.

Em “**arquiviola**”, “uma espécie de cravo” constitui um atributo (merónimo) de “arquiviola”, que não representa a relação para a qual o padrão é utilizado. A expressão “uma espécie de cravo” é igual ao padrão – uma espécie de P. O pretendido com este padrão seria P representar um *hiperónimo de* “arquiviola”, mas, como tal não acontece, a extração é considerada errada.

Como se verifica em “**molariforme**”, o padrão representa “uma espécie de dentes” não acontecendo de P, “dentes”, corresponder a um hiperónimo da entrada, mas antes de todo o padrão representar um atributo dela.

Em “**falcata**”, o padrão extraiu “foice” como um hiperónimo da entrada, no entanto, observa-se que “uma espécie de foice” é um elemento que entra na constituição de “falcata” (merónimo) e não há relação de hiponímia.

Em “**eipterado**”, “uma espécie de asa” não constitui, como se pode verificar, um hiperónimo da entrada, mas um atributo dela.

Embora nos três exemplos anteriormente descritos o conceito representado pelo padrão corresponda, coincidentemente, a um merónimo da cabeça do verbete, não pretendo propor nenhuma regra na qual este padrão seja utilizado para extrair merónimos. Como se pode ver nos exemplos de “**felose**”, “**hernuto**” e “**laringóstomo**”, não há qualquer relação entre a cabeça do verbete e a palavra ou sequência de palavras extraída pelo padrão.

Os “**outros problemas**” gerados por este padrão dizem respeito às extrações que, por alguma razão, não foram incluídas nem no grupo das corretas, nem no dos erros, inserem-se em três subgrupos:

Dúvidas do autor	Combinações fixas	Incerteza
20	13	135

Tabela 9: resultados de “outros problemas” do padrão 2.

Foram incluídos no grupo “**dúvidas do autor**”, os casos em que a definição continha um elemento (*talvez*, *?*, ou *parece que*) que mostrava que o autor do dicionário não tinha a certeza quanto àquela aceção. Tal verifica-se nos seguintes exemplos.

“**Carapo**: m. Ant. Espécie de tela?: «hũa taboa debuxada em carapo branco.» (De um testamento de séc. XVII)”

”**Forcarete**: m. Ant. Parece que era uma espécie de cobertor, pelo que se pode inferir das Provas da Hist. Geneal., vol. II, na descrição do enxoval de D. Beatriz.”

“**Paroleira**: f. Espécie de medida antiga? canastra? «levou quatro paroleiras de azeitonas.» (De um testamento do séc. XVII)

“**Taupla**: f. Espécie de antigo e luxuoso vestuário de mulher? cobertor precioso? Cf. Provas da Hist. Geneal., II, na descrição do enxoval de D. Beatriz.”

“**Malaia**: f. Ant. Espécie de saia?: «hũa malaya de pacotilho...» (De um testamento de 1692).

“**Macoco, (cô)**: m. Animal do Congo, talvez uma espécie de antílope.”

Tabela 10: exemplos de “dúvidas do autor” com o padrão 2.

O dicionarista incluiu em alguns dos verbetes certas abonações que o levaram a supor aquela aceção. No entanto, o próprio não assume que tal corresponda ao significado das entradas. Por esta razão, considero que não devo incluir estes casos nos grupos de extrações corretas nem dos erros por não ter certezas quanto à sua veracidade.

Outro dos problemas encontrados neste padrão diz respeito às “**combinações fixas**”. Nestes casos, a extração não pode ser considerada erro, uma vez que está presente uma relação de hiperonímia. No entanto, também não se pode considerar a extração correta porque o pretendido é ser a cabeça do verbete a estabelecer a relação de hiponímia com P. Acontece que P estabelece uma relação de hiperonímia com uma combinação fixa na qual entra a cabeça do verbete, mas não com ela. Estas considerações podem ser ilustradas pelos exemplos que se seguem⁸.

⁸As combinações fixas não se encontram a negrito no DA. Resolvi utilizá-lo de modo a que fosse possível perceber que a definição corresponde à combinação fixa, e não à cabeça do verbete.

“**Môsc**a: f. (...) **Asa de mosca**, espécie de prego, (...)”

“**Ponta**: f. (...) **Navalha de ponta e mola**, espécie de navalhas, cuja lâmina, de ponta aguda, é segurada quando aberta, por uma mola.”

“**Tejolo**, (jô): m. (...) **Tejolo burro**, espécie de tijolo grosseiro. (...)”

“**Terra**: f. (...) **Terra de Siena**, espécie de ocre amarelo. **Terra de Sevilha**, espécie de caparrosa, empregada para tingir de negro.”

“**Topázio**: m. (...) **Falso topázio**, espécie de quartzo amarelado.”

“**Vara**: f. (...) **Pano de varas**, antigo tecido de lã, de fabrico nacional, espécie de saragoça.”

Tabela 11: exemplos de “combinações fixas”.

O subgrupo “**incerteza**” é constituído pelos casos em que eu não fui capaz de decifrar se havia realmente uma relação de hiponímia ou se a relação existente entre os conceitos era apenas de semelhança. Tal deve-se, em grande parte, à própria definição de *espécie*, no Dicionário Aberto:

Espécie

f.

Qualidade.

Condição.

Carácter.

Aparência.

Semelhança externa.

Caso especial.

Gêneros alimentícios, que se dão em pagamento: *emprestei-lhe dinheiro, que pagou em espécie*.

Dinheiro.

Subdivisão de certas classificações.

Conjunto de seres que têm a mesma essência: *espécie humana*.

Casta.

Caso especial.

Especiaria.

Doce de amêndoa pisada.
Quantidade da mesma natureza, em aritmética.
Fam.
Surpresa, intriga: *isso faz-me espécie.*
(Lat. *species*)

Como se pode verificar, através desta definição, há várias aceções para definir *espécie*, assim como são variados os seus sentidos, entre os quais se encontram “aparência” ou “semelhança externa”. Isto significa que a expressão *espécie de* tanto pode ser utilizada para referir uma classe hierárquica na qual uma entidade se insere, como para expressar uma relação de semelhança entre essa entidade e outra qualquer. A minha dificuldade em perceber em qual destas duas relações se encaixam determinados termos deve-se ao facto de não ter, nem conseguir adquirir, mais conhecimento acerca dessas unidades lexicais, tendo em conta a época em que o Dicionário Aberto foi escrito.

A seguinte tabela apresenta alguns exemplos em que ocorreu o problema.

<p>“Beto: m. Prov. trasm. <i>Espécie de <u>pá de madeira</u>, com que se joga o toque-emboque.</i> (...)”</p> <p>“Jamaxi: m. Bras. <i>Espécie de <u>paneiro</u>, feito de timbó, e em que os seringueiros transportam suas mercadorias.</i>”</p> <p>“Lilinete, (nê): m. <i>Tecido, espécie de <u>lila</u>, mas menos encorpado.</i>”</p> <p>“Mascoto, (cô): m. <i>Grande martelo, com que, nas fábricas de moeda, se reduzem a pó os fragmentos de metal. T. de chapeleiro Espécie de <u>pisão</u>, em que se opera a fula.</i> (...)”</p> <p>“Nomo: m. <i>Divisão territorial do antigo Egipto, espécie de <u>distrito ou província</u>.</i>”</p> <p>“Oéta: f. Ant. <i>Vestuário de homem, espécie de <u>meia túnica</u>.</i>”</p>

Tabela 12: exemplos de “incerteza”

Mesmo depois de ter procurado, através de outros dicionários, perceber o que é, ou como é, cada um dos conceitos representados pelas cabeças dos verbetes, não fui capaz de discernir se a expressão “*espécie de*” se refere a um *tipo de* ou se pretende estabelecer uma relação de semelhança entre conceitos. De modo a não induzir em erro o utilizador do DA, optei por não incluir estes casos no grupo das extrações corretas. No entanto, não os incluí também nos erros porque podem não o ser. Resolvi, então, incorporar todos estes casos num subgrupo – “incerteza”.

Padrão 3: género de P

Ocorrências	Extrações corretas	Erros	Precisão (%)
4189	4188	1	99,98%

Tabela 13: resultados do padrão 3.

Este padrão mostrou-se quase perfeito, tendo em conta o elevado número de ocorrências e a elevada percentagem de precisão. Surgiu apenas um caso em que P não corresponde a um hipónimo da cabeça do verbe, como se verifica a seguir.

*“**Masculino:** adj. (...) Gram. Diz-se das palavras ou dos nomes e do género de palavras ou nomes, que, pela sua terminação ou pela sua concordância, designam seres que são masculinos ou se consideram tais, embora não tenham sexo.”*

Tabela 14: erro gerado pelo padrão 3.

Aqui, “masculino” não constitui um hipónimo de “palavras” nem de “nomes” – que seria o pretendido com este padrão – mas um atributo.

A tabela que se segue ilustra alguns exemplos de ocorrências deste padrão.

<p><i>“Bôa: f. Gibóia, género de <u>serpentes</u> da classe dos répteis. (...)”</i></p> <p><i>“Bovista: m. Género de <u>cogumelos</u>.”</i></p> <p><i>“Colúmbia: f. Género de <u>árvores tiliáceas</u>. (...)”</i></p> <p><i>“Fídia: f. Género de insectos <u>coleópteros tetrâmeros</u> da América.”</i></p> <p><i>“Hoteia: f. Género de <u>plantas saxifragáceas</u>.”</i></p> <p><i>“Mugem: f. Género de <u>peixes mugiloides</u>.”</i></p>
--

Tabela 15: exemplos de extrações do padrão 3.

Padrão 4: parte de P

Ocorrências	Extrações corretas	Erros	Dúvidas do autor	Precisão (%)
1007	746	259	2	74,08%

Tabela 16: resultados do padrão 4.

Embora com uma percentagem de precisão relativamente mais baixa do que a dos padrões utilizados na extração de hiperónimos e de hipónimos, este padrão fez, na minha opinião, um bom trabalho.

A respeito das “dúvidas do autor”, repete-se o que acontece com o padrão 2 (espécie de P); o autor demonstra incerteza na definição que faz. Os dois casos são apresentados na seguinte tabela.

“**Buquete:** *m. Agr.* Parte das máquinas de debulhar? Cf. *Gazeta dos Lavr.*, I, 17.”

“**Contrafio:** *m. Ant.* Armadilha, ou talvez parte de uma armadilha. Cf. Bernardes, *Luz e Calor*, 20.”

Tabela 17: exemplos de “dúvidas do autor” com o padrão 4.

Em ambos os casos, se não houvesse dúvidas pela parte do autor, considerar-se-ia que as extrações estavam corretas. Como tal não acontece, não posso incluí-las no grupo das extrações corretas porque corro o risco de um ‘buquete’ não ser um merónimo de uma ‘máquina de debulhar’, assim como ‘contrafio’ de ‘armadilha’.

Os erros gerados por este padrão surgem não apenas devido à estrutura das frases, como acontece com o padrão 2, mas também porque nem sempre uma *parte de* é considerada um *merónimo de*, o que justifica com a dificuldade na definição de meronímia. Alguns dos erros que o padrão gerou podem contemplar-se na tabela que se segue.

“**Alfitra:** *f. Ant.* Tributo, que pagavam os Moiros conquistados, e que constava da décima parte dos seus gados”

“**Azaqui:** *m.* Tributo, que entre nós pagaram os Moiros, e que consistia na décima parte dos frutos da terra.”

“**Dobra:** *f.* Parte de um objeto, que voltada, fica sobreposta a outra. (...)”

“**Fisconia:** *f. Med.* Tumefacção de uma parte do abdómen, sem timpanite, nem flutuação.”

“**Geodesia:** *f.* Ciência, que trata da forma e grandeza da Terra ou de uma parte da sua superfície.”

“**Ilutação:** *f.* Acto de cobrir de lodo uma parte do corpo para fins terapêuticos.”

Tabela 18: exemplos de erros gerados pelo padrão 4.

É possível que, devido à grande diversidade das relações de meronímia, haja quem considere que os conceitos representados nos três primeiros exemplos da tabela 18 constituam relações de meronímia. No entanto, a minha interpretação desta relação não me permite aceitar que exista efetivamente relação de meronímia.

Os restantes exemplos apresentam erros de natureza sintática.

Exemplos de extrações de sucesso deste padrão observam-se na tabela seguinte.

<p><i>“Higiene: f. Parte da <u>Medicina</u>, que trata dos meios de conservar a saúde. (...)”</i></p> <p><i>“Corôa: f. (...) Parte do <u>dente</u>, superior aos alvéolos. (...)”</i></p> <p><i>“Contra-almeida: f. Parte da <u>embarcação</u>, entre a barra de almeida e o parapeito das janelas da câmara,”</i></p> <p><i>“Doca: f. Parte de <u>um porto</u>, ladeado de muros ou caes, na qual se abrigam os navios, e onde tomam ou deixam a carga. (...)”</i></p> <p><i>“Frecha: f. (...) Parte do <u>raio</u> perpendicular à corda, entre esta e o arco, em Geometria. (...)”</i></p> <p><i>“Manga: f. Parte do <u>vestuário</u>, com que se cobre o braço. (...)”</i></p>
--

Tabela 19: exemplos de extrações corretas do padrão 4.

Padrão 5: membro de P

Ocorrências	Extrações corretas	Erros	Precisão
132	123	9	93,18%

Tabela 20: resultados do padrão 5.

Quanto à precisão, este padrão apresenta uma percentagem bastante elevada. Mas, por outro lado, é um padrão com poucas ocorrências, não tendo, portanto, grande cobertura.

Os erros encontrados refletem a situação que se verifica em quase todos os padrões; a expressão contida no padrão não se relaciona com a cabeça do verbete.

“Epanáfora: f. Rhet. Repetição da mesma palavra no princípio de cada membro de um período, ou no princípio de cada verso.”

“Amputar: v. t. Cortar (um membro do corpo). (...)”

“Militar: v. i. (...) Ser membro de um partido, ou partidário de uma doutrina.”

“Patente: adj. (...) f. (...) Diploma de um membro de confraria. (...)”

“Triplegia: f. Med. Hemiplegia, acompanhada da paralisia de um membro do lado oposto.”

“Epífora: (...) Rhet. Repetição de uma ou mais palavras, no fim de cada membro de um período.”

Tabela 21: exemplos de erros gerados pelo padrão 5.

Algumas relações extraídas com sucesso por este padrão apresentam-se na seguinte tabela.

“Confrade: m. Membro de confraria. (...)”

“Príncipe: m. Filho ou membro de uma família reinante. (...)”

“Lord: m. (...) Membro da câmara aristocrática do parlamento inglês. (...)”

“Juiz: m. (...) Membro do júri. Membro do poder judicial. (...)”

“Sectário: (...) m. Membro de uma seita. (...)”

“Vereador: m. Aquele que vereia; membro da Câmara Municipal; camarista.”

Tabela 22: exemplos de extrações corretas do padrão 5.

Conforme se pode verificar através dos exemplos, este padrão extrai, de um modo geral, a relação de meronímia correspondente ao tipo “membro-coleção” proposto por Winston *et al.*

Padrão 6: cada um/a d[ao]s X d[eao]s P

Ocorrência	Extrações corretas	Erros	Precisão (%)
188	145	43	77,13%

Tabela 23: resultados do padrão 6.

Os erros gerados por este padrão têm a ver, não apenas com a estrutura da frase, mas também pelo facto de, além de referir atribuição, este padrão referir posse – que não constitui meronímia. Exemplos de ambos os tipos de erro são apresentados na tabela que se segue.

<p>“Teor: m. (...) <i>Quím. Proporção, em que está cada um dos elementos de <u>um corpo composto.</u></i>”</p> <p>“Bordada: f. (...) Descanso de canhões, de cada um dos lados do <u>navio.</u> (...)”</p> <p>“Gênio: m. (...) <i>Cada um dos espíritos, que se supunha dominarem cada um dos elementos da <u>natureza.</u></i> (...)”</p> <p>“Lúnula: f. <i>Cada um dos satélites de <u>Júpiter ou Saturno</u></i> (...) ”.</p> <p>“Filípica: f. (...) <i>Cada uma das orações de <u>Cícero</u> contra Marco-António.</i> (...)”</p> <p>“Senadora: f. <i>Designação de cada uma das mulheres de <u>Heliogábalo.</u></i> Cf. Rui Barb., <i>Réplica, 158.</i>”</p>

Tabela 24: erros gerados pelo padrão 6.

Nos três primeiros exemplos, o padrão gerou erro devido à estrutura da frase. Nos outros três exemplos o erro deve-se ao facto de “Lúnula”, “Filípica” e “Senadoras” referirem apenas elementos de posse de “Júpiter e Saturno”, “Cícero” e “Heliogábalo”,

respetivamente, e não de atribuição (*Heliogábalos* tem *senadoras*, mas as *senadoras* não são uma parte de *Heliogábalos*⁹).

Embora com um número relativamente baixo de ocorrências, este padrão foi capaz de extrair uma elevada percentagem de relações com sucesso, algumas das quais são apresentadas a seguir.

“**Rízula:** f. Bot. Cada uma das radículas dos cogumelos.”

“**Pena:** f. (...) Cada uma das asas do rodízio do moinho, nas quais bate a água que o move.”

“**Vareta, (varê):** f. (...) Cada uma das pernas do compasso. (...)”

“**Falange:** f. (...) Anat. Cada um dos ossos dos dedos, especialmente o osso que se articula com o metacarpo. (...)”

“**Manipular:** m. Cada um dos soldados de um manípulo, entre os Romanos.”

“**Elo:** m. (...) Cada um dos anéis de uma cadeia. (...)”

Tabela 25: exemplos de relações bem extraídas pelo padrão 6.

4.3. Últimas considerações

Importa saber que a aplicação na qual foram inseridos os padrões para extração infere as relações inversas, isto é, *se um conceito x constitui um hipónimo ou um merónimo de y, então y constitui um hiperónimo ou um holónimo de x*.

Também o fator da transitividade tem influência na extração das relações; *se um conceito x constitui um hipónimo de y, e y constitui um hipónimo de z, então será também inferido que x é um hipónimo de z*. O mesmo não acontece na relação de meronímia devido à sua não transitividade (ou transitividade apenas dentro do mesmo tipo).

⁹ Este exemplo ilustra as considerações de Cruse (1986) “um filho tem uma mãe”, #”uma mãe é parte de um filho” (vd §2, p.12).

De modo a comparar os resultados obtidos nas diferentes relações – hiponímia/hiperonímia e meronímia/holonímia, apresento uma tabela, na qual constam o número de ocorrências, o número de erros e o número de extrações corretas obtidos pelos padrões em cada uma das relações:

Relação	Ocorrências	Extrações corretas	Erros	Percentagem de extrações corretas
Hiponímia/hiperonímia	8306	7999	3519	96,30%
Meronímia/holonímia	1327	1014	311	76,41%

Tabela 26: comparação dos resultados por relação lexical.

A tabela 26 demonstra uma diferença bastante significativa nos números gerados pelos padrões em cada uma das relações. A justificativa para tal diferença pode recair sobre a escolha dos padrões, ou seja, outros padrões gerariam, certamente, valores diferentes. Não obstante, na minha opinião, as diferenças não se devem apenas aos padrões, mas também à dificuldade em parafrasear a relação de meronímia, ou então pela dificuldade e vagueza de defini-la, contrariamente ao que acontece com a hiponímia.

5. Conclusão

Este trabalho teve como principal objetivo a avaliação de produtividade e precisão de um conjunto de padrões de *Hearst* para a extração semiautomática de relações léxico-conceptuais a partir de um dicionário eletrônico.

Foram experimentados seis padrões, três para a extração de hiponímia/hiperonímia e três para a extração de meronímia/holonímia.

Os padrões são capazes de extrair relações entre a cabeça do verbete do dicionário e uma palavra ou uma sequência de palavras. No entanto, muitas vezes, os padrões utilizados não são capazes de delimitar corretamente a sequência de palavras, limitando-as aos elementos constitutivos do termo pluriverbal. Por esta razão, optou-se por delimitar esta sequência de palavras pelo primeiro sinal de pontuação presente. Mesmo assim, várias relações continuam a incluir palavras que não fazem parte do termo que se relaciona com a cabeça do verbete. Embora se tenham avaliado todas as situações em que cada um dos padrões propostos ocorre, por enquanto só se encontram disponíveis *online* as relações em que o padrão extraiu uma relação com uma única palavra (uma vez que estas relações foram disponibilizadas antes da avaliação ter sido concluída).

Obviamente, o ideal seria que, em cada verbete em que ocorrem os padrões analisados, pudéssemos contemplar os conceitos com os quais a entrada estabelece as relações. Isso talvez fosse possível se conseguíssemos que o padrão extraísse a sequência de palavras correspondente à relação estabelecida, através de uma regra, como por exemplo, *espécie de* (Adj) + N + (Adj), segundo a qual o padrão “espécie de P” seria capaz de extrair como hiperónimo um nome, ou um nome precedido de um adjetivo, ou um nome sucedido por um adjetivo. No entanto, para isso seria necessário o uso de um analisador morfológico e, enquanto a modernização da grafia do Dicionário Aberto não for terminada, isso não é possível (não existe um analisador morfológico apto para o tratamento da língua portuguesa com a grafia pré 1940).

Embora alguns dos padrões utilizados neste trabalho tenham uma boa capacidade de cobertura e se tenham mostrado até bastante precisos, como o que se verificou com o padrão 3 – *género de P* –, há sempre lacunas que uma extração automática não é capaz de preencher.

No entanto, conseguiu-se uma precisão de 96,30% para as regras de hiponímia/hiperonímia e de 76,41% para as regras de meronímia/holonímia.

Na minha opinião, a extração semiautomática e a extração manual de relações lexicais não se substituem uma à outra. Se, por um lado, a extração semiautomática tem pouca cobertura, a extração manual é muito morosa, e uma extração semiautomática com mais cobertura e que não deixasse lacunas aumentaria, provavelmente, a quantidade de relações erradamente extraídas.

Por estas razões, a escolha do tipo de extração deve ser feita com base no objetivo que se pretende alcançar. Como o meu objetivo para este trabalho consistia, essencialmente, em perceber que tipo de padrões eram passíveis de extrair corretamente uma grande quantidade de relações, só fazia sentido efetuar uma extração semiautomática, mas, se por outro lado, o meu objetivo fosse cobrir todas as relações de hponímia/hiperonímia e de meronímia/holonímia presentes no DA, então, só uma extração manual seria capaz de me ajudar.

Referências bibliográficas

Brauner, G. (2005). Sobre a Semântica Lexical: Jerry Fodor e Ernest Lepore versus Pustejovsky. In: *V Semana de Letras*, PUCRS.

Climent, S. (2000). Individuación e Información Parte-Todo: Representación para el procesamiento computacional del lenguaje. In: *Estudios de Lingüística del Español*.

Cruse, D. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press.

Freitas, C., Santos, D., Oliveira, H., & Quental, V. (2010). *VARRA: um serviço para a Validação, Avaliação e Revisão de Relações semânticas no AC/DC*.

Hearst, M. (1992). Automatic acquisition of hyponyms from large text corpora. In: *Proceedings of the 14th International Conference on Computational Linguistics*, Nantes.

Lyons, J. (1980). *Semántica*. Barcelona: Teide.

Marrafa, P. (2001). *WordNet do Português: uma base de dados de conhecimento linguístico*. Lisboa: Instituto Camões.

Miller, G. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11), 39-41.

Oliveira, H., Costa, H., Pérez, L., & Gomes, P. (2011). Uma rede léxico-semântica de grandes dimensões para o português, extraída a partir de dicionários electrónicos. *Linguamática*.

Oliveira, H., Santos, D., & Gomes, P. (2010). *Extracção de relações semânticas entre palavras a partir de um dicionário: o PAPEL e sua avaliação*. *Linguamática*.

Oliveira, H. (2013). *Onto.PT: Towards the Automatic Construction of a Lexical Ontology for Portuguese* (Dissertação de Doutoramento). Coimbra: Universidade de Coimbra.

Santos, D., Barreiro, A., Freitas, C., Oliveira, H., Medeiros, J., Gomes, P., & Silva, R. (2010). Relações semânticas em português: comparando o TeP, o MWN.PT, o Port4NooJ e o PAPEL. *Linguamática*.

Santos, D. (2001). Introdução ao processamento da linguagem natural através das aplicações. In E. Ranchhod (ed.), *Tratamento das Línguas por Computador. Uma introdução à linguística computacional e suas aplicações*. Lisboa: Caminho, pp. 229-259.

Simões, A., Iriarte, A., & Almeida, J. (2012). *Dicionário-Aberto: A Source of Resources for the Portuguese Language Processing*. Acedido através do endereço <http://repositorium.sdum.uminho.pt/bitstream/1822/23570/1/da2012.pdf> em setembro de 2013.

Simões, A. & Farinha, R. (2011). Dicionário Aberto: um recurso para processamento de linguagem natural. *Vice-Versa*, 16,159-171

Winston, M., Chaffin, R., & Herrmann, D. (1987). A Taxonomy of Part-Whole Relations. *Cognitive Science*, 11, 417-444.