

# PAPEL

---

Palavras Associadas Porto Editora Linguatca

Trabalho relacionado e relações  
semânticas em recursos semelhantes

Hugo Gonçalo Oliveira, Paulo Gomes  
Linguatca, pólo de Coimbra, DEI - FCTUC, CISUC

Diana Santos  
Linguatca, pólo de Oslo, SINTEF ICT

---

Dezembro 2007

# Índice

1	Introdução . . . . .	2
2	Trabalho relacionado . . . . .	3
2.1	WordNet . . . . .	3
2.2	EuroWordNet . . . . .	5
2.3	WordNet.BR . . . . .	5
2.4	WordNet.PT . . . . .	6
2.5	MindNet . . . . .	8
2.6	FrameNet . . . . .	8
3	Relações semânticas . . . . .	11
3.1	Sinónimos . . . . .	11
3.2	Hiperónimos e Hipónimos . . . . .	11
3.3	Todo e parte . . . . .	11
3.4	Causa e efeito . . . . .	12
3.5	Atributo . . . . .	12
3.6	Lugar/local geográfico . . . . .	12
3.7	Por meio de/Objecto para realizar uma acção . . . . .	12
3.8	Descrição da forma de realizar uma acção . . . . .	12
4	Agradecimentos . . . . .	13

# 1 Introdução

Este é o primeiro relatório do PAPEL (Palavras Associadas Porto Editora Liguatca), um projecto onde se pretende construir de forma automática uma ontologia lexical para o português, através de informação extraída do Dicionário da Língua Portuguesa da Porto Editora [dic05]. É apresentado algum trabalho relacionado, mais propriamente um conjunto de ontologias lexicais existentes e uma secção dedicada a exemplos de relações semânticas que poderão vir a estar contempladas no nosso recurso. Utilizando dicionário pretendemos extrair algumas das relações apresentadas, com base na combinação de padrões, inseridos em gramáticas específicas para cada relação.

## 2 Trabalho relacionado

Nesta secção apresentam-se alguns recursos linguísticos onde está representado conhecimento lexical e que são, de certa forma, semelhantes ao que queremos construir, o PAPEL. É dada especial atenção às relações presentes em cada um dos recursos referidos.

### 2.1 WordNet

Não há dúvida que os dicionários são um recurso bastante útil quando é necessário escolher palavras ou compreender o seu significado, mas mesmo os dicionários electrónicos falham quando se pretende organizar o conhecimento lexical de uma forma síncrona. O WordNet [MBF<sup>+</sup>90] é uma proposta para uma melhor combinação de informação lexicográfica tradicional com a computação moderna e pode dizer-se que se trata de um dicionário baseado em princípios psicolinguísticos. No léxico do WordNet existe uma divisão clara entre cinco categorias: nomes, verbos, adjectivos, advérbios e palavras funcionais. Os autores acreditam que tentativas para organizar todas as categorias de acordo com o mesmo princípio não serão capazes de representar a complexidade psicológica do conhecimento lexical.

A base do WordNet é uma rede de conceitos, correspondendo cada conceito a um conjunto de palavras que são sinónimos entre si (*synsets*). Os conceitos estão depois organizados de acordo com relações semânticas, que são representadas através de ligações entre eles.

Há ainda a dizer que foi construído manualmente, é distribuído gratuitamente<sup>1</sup> e tem também um interface *online* onde é possível efectuar buscas pela rede.

As relações semânticas existentes no WordNet encontram-se na Tabela 1. Segue-se uma pequena descrição de algumas, de acordo com [MBF<sup>+</sup>90]:

- **Sinonímia:** É a relação mais importante do WordNet e significa que duas palavras têm o mesmo significado (ou pelo menos um significado muito próximo) realizando por isso o mesmo conceito. Duas expressões são sinónimas num contexto linguístico *C* se a substituição de uma pela outra em *C* não altera o seu valor de verdade (são permutáveis). A última definição é uma das razões que leva à partição do WordNet em nomes, verbos, adjectivos e advérbios. Se os conceitos são representados por *synsets* e se além disso têm de ser permutáveis, conceitos de diferentes categorias gramaticais não podem ser sinónimos.

---

<sup>1</sup><http://wordnet.princeton.edu/>

Relação semântica	Categoria sintáctica	Exemplos
Synonymy (similar)	N, V, Aj, Av	(pipe, tube) (rise, ascend) (sad, unhappy) (rapidly, speedily)
Antonymy (opposite)	Aj, Av, (N, V)	(wet, dry) (powerful, powerless) (friendly, unfriendly) (rapidly, slowly)
Hyponymy (subordinate)	N	(sugar maple, maple) (maple, tree) (tree, plant)
Meronymy (part)	N	(brim, hat) (gin, martini) (ship, fleet)
Troponomy (manner)	V	(march, walk) (whisper, speak)
Entailment	V	(drive, ride) (divorce, marry)

Tabela 1: Relações do WordNet [Mil95]

- **Antonímia:** Segundo os autores é uma relação difícil de definir: O antónimo de uma palavra X, por vezes é não-X, mas nem sempre. A antonímia não é considerada uma relação semântica, mas uma relação entre palavras, porque apesar de um *synset* poder conter uma palavra antónima de uma palavra de outro *synset*, pode haver outras palavras dentro dos dois *synset* que não verifiquem essa relação.
- **Hiponímia/hiperonímia:** Também conhecida por subordinação/superordinação, subconjunto/superconjunto ou relação é-um (*is-a*). O hipónimo de um conceito se for um tipo mais específico desse conceito. Um hipónimo herda todas as características do conceito mais genérico (hiperónimo) e tem pelo menos mais uma que o distingue dos demais hipónimos desse conceito. A hiponímia é uma relação transitiva e assimétrica, havendo regra geral um único hiperónimo o que dá origem à construção de estruturas hierárquicas.
- **Merónimia/Holonímia:** Também conhecida como parte-de (*part-of*). O merónimo de um conceito X é um conceito Y que faz parte do primeiro. Por outras palavras, Y é uma parte de X.

Existem ainda **relações morfológicas** representadas, para que seja possível procurar por palavras não lematizadas (plural, feminino, conjugação de um verbo), associando-as ao(s) conceito(s) relativo(s) ao seu lema. Estas relações são necessárias essencialmente para quando é efectuada uma busca utilizando uma palavra não lematizada o sistema identificar o seu lema e fazer a busca por este último.

## 2.2 EuroWordNet

O EuroWordNet [Vos97] foi um projecto que consistiu na criação de uma base de dados de *wordnets* para as línguas inglesa, espanhola, holandês e italiana. As *wordnets* espanhol, holandês e italiano seriam depois ligados ao WordNet (inglês) através de relações de equivalência, dando origem a um recurso que pode ser utilizado na extracção de informação multilingue. Apesar da estrutura das *wordnets* não ingleses ser semelhante à estrutura do WordNet, existem pequenas diferenças devido a:

- Base de dados multilingue;
- Utilização da base de dados para extracção de informação;
- Natureza da informação contida em dicionários electrónicos utilizados para a derivação e eventual extracção (semi-)automática;
- Atingir um maior nível de compatibilidade entre os vários recursos;
- Tornar possível a manutenção de relações específicas de cada língua nos várias *wordnets*;
- Possibilitar os utilizadores a personalizar a base de dados para a sua aplicação sem necessitarem conhecer todas as línguas.

Este projecto terminou em Junho de 1999 e ao contrário do WordNet, a maior parte das *wordnets* que dele fazem parte não são gratuitamente disponibilizados na sua totalidade, podendo no entanto ser descarregadas amostras dos mesmos<sup>2</sup>.

## 2.3 WordNet.BR

O WordNet.BR [dS04] é a versão brasileira de uma estrutura semelhante ao WordNet. A sua compilação foi feita segundo uma metodologia que abrange três domínios:

- O **domínio linguístico**: neste domínio inserem-se as noções de *thesaurus* e condições para a definição de semelhança e criação de *synsets*. A forma de construir um *synset* é semelhante ao WordNet. Insere-se aqui também a utilização de dicionários como corpora base para a construção de *synsets*.

---

<sup>2</sup><http://www.ilc.uva.nl/EuroWordNet/data/sampleData.html>

- O **domínio representacional**: a forma escolhida para representar a sua estrutura foi baseada no WordNet, ou seja, o WordNet.BR é também um conjunto de *synsets* ligados através de relações semânticas. Cada entrada na base de dados é constituída por uma palavra, respectiva classe gramatical e pelo conjunto de *synsets* que dizem respeito aos vários significados que a palavra pode ter.
- O **domínio computacional**: existe uma aplicação gráfica onde os linguistas podem inserir informação no WordNet.BR (*synsets*, frases exemplo, ...).

Uma das formas utilizadas para extrair as relações presentes no WordNet.BR foi a extracção manual (e não automática) a partir de dicionários impressos[dSF06]. Em 2005 a base de do WordNet.BR estava apenas estruturada em função de relações de **sinonímia** e **antonímia**, apesar das restantes relações estarem previstas no seu desenvolvimento. Estava também previsto relacionar *synsets* semanticamente equivalentes entre a base do WordNet.BR e a base do WordNet americano, com vista à implementação de uma base bilingue.

## 2.4 WordNet.PT

Apesar de existir um sítio online do WordNet.PT<sup>3</sup>, onde se encontra uma listagem das relações por ele contempladas (Tabela 2), ao se efectuar uma pesquisa o utilizador depara-se com um ecrã de erro no servidor. Isto vem acontecendo (pelo menos) desde Setembro de 2007. Apesar da lista de relações se encontrar na página principal, a ajuda acerca de cada uma aparenta estar em construção. Os autores do WordNet.PT afirmam que os recursos linguísticos disponíveis para a língua portuguesa não são adequados à construção automática deste tipo de recurso, estando portanto a construí-lo manualmente [MAC<sup>+</sup>06], o que de certa forma vai contra a nossa proposta.

## 2.5 MindNet

O MindNet [RDV98] foi desenvolvido por um grupo de investigação da Microsoft e começou por ser uma base de conhecimento lexical construída de forma completamente automática com base na informação contida em texto de um conjunto de dicionários electrónicos. No entanto, o MindNet acaba por não ser apenas um recurso estático. Ele representa uma metodologia constituída por um conjunto de ferramentas para adquirir, estruturar, aceder e

---

<sup>3</sup><http://www.instituto-camoes.pt:8080/wordnet/>

<b>relações geral/específico</b>	<b>relações inversas</b>
x é hiperónimo (é supertipo) de	x é hipónimo (é um tipo) de
x é instanciado por	x é instanciação de

  

<b>relações todo/parte (holónimo/merónimo)</b>	<b>relações inversas</b>
x tem como parte	x é parte de
x tem como parte distinta	x é parte distinta de
x tem como substância/material	x é substância/material de
x tem como membro	x é membro de
x tem como porção	x é porção de
x tem como localização	x é localização de

  

<b>relações de equivalência</b>
x é sinónimo de
x tem correspondência transcategorial com
x é quase sinónimo de

  

<b>relações de oposição</b>
x é antónimo de
x é quase antónimo de

  

<b>relações de categorização</b>	<b>relações inversas</b>
x é caracterizável por	x caracteriza quanto a
x está relacionado com	x está relacionado com (?)
x é característica de	x tem como característica

  

<b>relações entre participantes num evento</b>	<b>relações inversas</b>
x co_relaciona-se com	
x agente/causa_resultado	x resultado_agente/causa
x instrumento_resultado	x resultado_instrumento
x objecto_resultado	x resultado_objecto
x agente_instrumento	x instrumento_agente
x objecto_instrumento	x instrumento_objecto
x agente_objecto	x objecto_agente

  

<b>relações definidoras da estrutura do evento</b>	<b>relações inversas</b>
x está envolvido em	x envolve
x é instrumento para	x envolve como instrumento
x é lugar para	x tem lugar em
x é ponto de origem de	x tem como ponto de origem
x resulta de	x tem como resultado
x tem como subevento	x é subevento de
x está implicado como agente em	x implica como agente
x está implicado como objecto em	x implica como objecto
x é meio físico para	x tem como meio físico
x é ponto de destino/chegada de	x tem como ponto de destino/chegada
x causa	x tem como causa
x é o modo de ocorrência de	x tem como modo de ocorrência

Tabela 2: Relações existentes no WordNet.PT [CLG07]



explorar informação semântica contida em texto, tendo entretanto já sido aplicada a texto que não de dicionário (por exemplo, Microsoft Encarta [VKSM05]).

Para a construção do MindNet foi utilizado o *parser* PLN utilizado na verificação gramatical do *Microsoft Word 97* que se revelou essencial e que é também a base para o crescimento do MindNet. Este *parser* produz árvores sintácticas e formas lógicas sobre as quais são aplicadas regras para a extracção de relações semânticas. A forma de encontrar semelhanças entre palavras combina a medição da “semelhança substitucional” e também na procura de *clusters* de palavras geralmente relacionadas. Há também um procedimento que, a partir das relações encontradas procura inferir novas relações. A maior parte da desambiguação de sentidos é feita com o auxílio de informação morfológica e sintáctica obtida na altura do *parsing*. Outras formas de desambiguação passam pela informação acerca do domínio de cada palavra (presente nos dicionário) e também na análise de relações entre as várias palavras. No MindNet existe a distinção entre as quatro classes gramaticais abertas, mas as relações não estão restringidas a nenhuma em particular. Existe um interface web para o MindNet, o MNEX (MindNet Explorer)<sup>4</sup> criado com o objectivo de facilitar as buscas feitas à estrutura e às relações do MindNet [VKSM05]. O utilizador insere duas palavras e o MNEX mostra todos os caminhos de relações semânticas entre essas duas palavras.

O MindNet define um vasto leque de relações semânticas e é actualmente constituído pelas relações na Tabela 3.

## 2.6 FrameNet

O FrameNet [BFL98] é um projecto que procura produzir a descrição de enquadramentos semânticos (*semantic frames*). Cada enquadramento corresponde a um conceito que pode descrever um objecto, um estado ou um evento através das relações sintácticas e semânticas de uma unidade lexical que representa uma situação e envolve participantes, propriedades e outros papéis conceptuais. A informação contida no FrameNet é extraída de corpora (com anotações semânticas feitas manualmente) e também na análise sistemática de padrões semânticos exemplificados por linguistas e lexicógrafos. Actualmente cobre apenas domínios como: *health care, chance, perception, communication, transaction, time, space, body (parts and functions of the*

---

<sup>4</sup><http://atom.research.microsoft.com/mnex/MnexHelp.aspx#relations>

<b>Relação</b>	<b>Descrição</b>
← <b>Attrib</b> →	Attributive modification
← <b>Cause</b> →	Causation
← <b>Classifier</b> →	Classifier
← <b>Goal</b> →	Goal
← <b>Hyp</b> →	Hypernymy
← <b>Intnsifs</b> →	Intensifier
← <b>Locn</b> →	Location
← <b>Manner</b> →	Manner
← <b>Means</b> →	Means
← <b>Mod</b> →	Modifier
← <b>Modals</b> →	Modals
← <b>Part</b> →	Meronymy (Part relation)
← <b>Possr</b> →	Possessor
← <b>Props</b> →	Proposition (like Purp, but relation left undetermined)
← <b>Purp</b> →	Purpose
← <b>Result</b> →	Result
← <b>Source</b> →	Source
← <b>Syn</b> →	Synonymy
← <b>Time</b> →	Time/Duration
← <b>Tind</b> →	Typical Indirect Object
← <b>Tobj</b> →	Typical Object
← <b>Tsub</b> →	Typical Subject

Tabela 3: Relações do MindNet

*body*), *motion*, *life stages*, *social context*, *emotion and cognition*.

Algumas diferenças entre o FrameNet e o WordNet [REP<sup>+</sup>06]:

- O FrameNet contém vários exemplos anotados de cada unidade lexical ou significado. Os exemplos consistem em texto natural, retirado de corpora.
- A análise contida no FrameNet é realizada enquadramento por enquadramento e não por lema.
- A rede de relações do WordNet é feita entre os seus nós, enquanto a rede de relações do FrameNet é feita entre os enquadramentos semânticos.

O FrameNet contém uma rede de relações entre *frames*. De seguida apresentam-se as mais importantes:

- **Inheritance** (herança): A típica relação *is-a* (hiponímia). Exemplo: *Revenge* (vingança) herda de *Rewards\_and\_punishments* (recompensas e punições).
- **Using**: a *frame* “filha” pressupõe a *frame* pai. Exemplo: *Speed* (velocidade) utiliza *Motion* (movimento).
- **Subframe**: a *frame* “filha” é um subevento da *frame* pai, que será um evento mais complexo. Exemplo: *Arrest* (detenção), *Arraignment* (acusação), *Trial* (julgamento) e *Sentencing* (sentenciação) utilizam *Criminal\_process* (processo criminal).
- **Perspective on**: a *frame* “filha” representa uma perspectiva da *frame* pai. Exemplo: *Hiring* (contratar) e *Get\_a\_job* (arranjar emprego) perspectivam a *frame* *Employment\_start* (início de um emprego) respectivamente do ponto de vista do empregador e do empregado.

Relation	Sub	Super
Inheritance	Child	Parent
Perspective_on	Perspectivized	Neutral
Subframe	Component	Complex
Precedes	Later	Earlier
Inchoative_of	Inchoative	State
Causative_of	Causative	Inchoative/State
Using	Child	Parent
See_also	Referring Entry	Main Entry

Tabela 4: Relações entre Frames, no FrameNet [REP<sup>+</sup>06]

### 3 Relações semânticas

Nesta secção apresenta-se um conjunto de relações semânticas, juntamente com exemplos na sua maioria adaptados para português a partir das ajudas do MNEX. Os exemplos apresentados são alguns daqueles que queremos ter representados no PAPEL.

#### 3.1 Sinónimos

←Sin→	diz-se	exemplos
X Sin→ Y	X tem um <b>Sinónimo</b> Y	colina <b>Sin→</b> (encosta, outeiro, monte, ...)
X ←Sin Y	X é um <b>Sinónimo</b> de Y	monte ← <b>Sin</b> (montanha, colina, ...)

#### 3.2 Hiperónimos e Hipónimos

←Hiper→	diz-se	exemplos
X Hiper→ Y	X tem um <b>Hiperónimo</b> Y	pássaro <b>Hiper→</b> (animal, vertebrado, criatura, ...)
X ←Hiper Y	X é um <b>Hiperónimo</b> de Y	pássaro ← <b>Hiper</b> (tordo, águia, faisão, ...)

#### 3.3 Todo e parte

←Parte→	diz-se	exemplos
X Parte→ Y	X tem <b>Parte</b> Y	pássaro <b>Parte→</b> (pena, asa, bico, pata, ...)
X ←Parte Y	X é uma <b>Parte</b> de Y	perna ← <b>Parte</b> (animal, calças, mesa, cadeira, ...)

Também chamados de holónimos e merónimos.

### 3.4 Causa e efeito

←Causa→	diz-se	exemplos
X Causa→ Y	X é <b>Causado</b> por/é um <b>Efeito</b> de Y	doença <b>Causa</b> → (microorganismo, deficiência, ...)
X ←Causa Y	X é uma <b>Causa</b> de Y	vírus ← <b>Causa</b> (doença, SIDA, gripe, ...)

### 3.5 Atributo

←Atrib→	diz-se	exemplos
X Atrib→ Y	X tem o <b>Atributo</b> Y	gato <b>Atrib</b> → (doméstico, selvagem, feroz, ...)
X ←Atrib Y	X é um <b>Atributo</b> de Y	azul ← <b>Atrib</b> (tinta, céu, plumagem, flor, ...)

Modificação através da atribuição de uma qualidade/propriedade.

### 3.6 Lugar/local geográfico

←Loc→	diz-se	exemplos
X Loc→ Y	X é <b>Local</b> de Y	restaurante <b>Loc</b> → (servir, mesa, refeição, conta, ...)
X ←Loc Y	X <b>Encontra-se/Localiza-se</b> em Y	restaurante ← <b>Loc</b> (estância, cidade, Grécia, ...)

### 3.7 Por meio de/Objecto para realizar uma acção

←Meio→	diz-se	exemplos
X Meio→ Y	Realizar X por <b>Meio</b> de Y	abrir <b>Meio</b> → (chave, código, interruptor...)
X ←Meio Y	X é um/uma <b>Meio</b> para realizar Y	chave ← <b>Meio</b> (abrir, fechar, ajustar, ...)

### 3.8 Descrição da forma de realizar uma acção

←Forma→	diz-se	exemplos
X Forma→ Y	Realizar X da <b>Forma</b> Y	correr <b>Forma</b> → (depressa, lentamente, de repente ...)
X ←Forma Y	X é uma <b>Forma</b> de realizar Y	lentamente ← <b>Forma</b> (cozinhar, andar, jogar, ...)

## 4 Agradecimentos

Este relatório foi escrito no âmbito da Linguateca, financiada pela Fundação para a Ciência e Tecnologia e pela União Europeia através dos projectos POSI/PLP/43931/2001 e POSC 339/1.3/C/NAC.

Agradecemos ainda ao Nuno Seco que iniciou o projecto e participou em várias discussões de ideias neste “reinício” dando o seu apoio sempre que necessário e também ao Núcleo de Investigação e Desenvolvimento da Porto Editora.

# Referências

- [BFL98] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics*, pages 86–90, Morristown, NJ, USA, 1998. Association for Computational Linguistics.
- [CLG07] CLG. Wordnet.pt - rede léxico-conceitual do português, 2007. <http://www.instituto-camoes.pt:8080/wordnet/>.
- [dic05] *Dicionário PRO da Língua Portuguesa*. Porto Editora, Porto, 2005.
- [dS04] Bento C. Dias da Silva. Wordnet.br: An exercise of human language technology research. *Revista PaLavra*, no. 12, 2004. *Série Linguagem*, Processamento Automático do Português:15–24, 2004.
- [dSF06] Bento C. Dias da Silva and Ariani Di Felippo. Extração de informações lógico-conceituais de dicionários para a elaboração de léxicos computacionais. *Revista Intercâmbio*, volume XV. São Paulo LAEL/PUC, ISSN 1806-275X, 2006.
- [MAC<sup>+</sup>06] P. Marrafa, R. Amaro, R. P. Chaves, S. Lourosa, C. Martins, and S. Mendes. Wordnet.pt new directions. In *Petr Sojka, Key-Sun Choi, Christiane Fellbaum, Piek Vossen (eds.), Proceedings of GWC'06: 3rd International Wordnet Conference, Jeju Island, Korea*, pages 319–320, 2006.
- [MBF<sup>+</sup>90] George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. Introduction to wordnet: An on-line lexical database\*. *Int J Lexicography*, 3(4):235–244, January 1990.
- [Mil95] George A. Miller. Wordnet: a lexical database for english. *Commun. ACM*, 38(11):39–41, 1995.
- [RDV98] Stephen D. Richardson, William B. Dolan, and Lucy Vanderwende. Mindnet: acquiring and structuring semantic information from text. In *Proceedings of the 17th international conference on Computational linguistics*, pages 1098–1102, Morristown, NJ, USA, 1998. Association for Computational Linguistics.

- [REP<sup>+</sup>06] J. Ruppenhofer, M. Ellsworth, M.R.L. Petruck, C.R. Johnson, and J. Scheffczyk. Framenet ii: Theory and practice. online publication, 2006.
- [VKSM05] Lucy Vanderwende, Gary Kacmarcik, Hisami Suzuki, and Arul Menezes. Mindnet: An automatically-created lexical resource. In *HLT/EMNLP*. The Association for Computational Linguistics, 2005.
- [Vos97] Piek Vossen. Eurowordnet: a multilingual database for information retrieval. In *Proc. of the DELOS workshop on Cross-Language Information Retrieval*, Zurich, 1997.