

Relações semânticas em português: comparando o TeP, o MWN.PT, o Port4NooJ e o PAPEL

*Diana Santos**, *Anabela Barreiro**, *Cláudia Freitas**, *Hugo Gonçalo Oliveira⁺*, *José Carlos Medeiros⁼*, *Luís Costa**, *Paulo Gomes⁺*
& *Rosário Silva**

*Linguateca

⁺ Centro de Informática e Sistemas da Universidade de Coimbra

⁼Porto Editora

Abstract

This paper compares individual resources that contain semantic relations in Portuguese as a first approach towards the representation of the Portuguese language through a general lexical ontology. We start with a short overview of five projects that make these semantic resources publicly available, namely WordNet.PT, TeP/WordNet.BR, MWN.PT, Port4NooJ and PAPEL. Although the accessibility and information represented in each one of these resources differs widely, we attempt to highlight the most evident similarities and differences. Based on individual relations, we present and discuss two initial comparative experiments. We also present some considerations about meronymy, based on how this semantic relation was handled in the resources studied and address briefly the semantic field of colour. Finally, we urge people to create collaboratively more complete, complex and dynamic resources for the description of Portuguese.

Keywords: lexical ontologies, Portuguese, evaluation.

Palavras-chave: ontologias lexicais, português, avaliação.

1. Ontologias lexicais e o seu interesse no processamento de linguagem natural

Desde muito cedo que o PLN reconheceu a necessidade de recursos de cobertura vasta para poder passar de protótipos para sistemas reais, e uma das formas de alimentar esta necessidade era recorrer a repositórios abrangentes já criados e que pudessem de certa forma colmatar a necessidade de grande esforço na criação de recursos.

Nessa demanda, surgiu o uso de dicionários ou outros recursos lexicográficos como base de estruturação de conhecimento (Gonçalo Oliveira *et al.*, 2009). Ao resultado desta transformação ou enriquecimento de materiais lexicográficos para uso no processamento da língua, ou seja, às estruturas representantes desse conhecimento damos o nome de ontologias lexicais (Veale, 2007)¹, embora ainda não haja uma terminologia completamente fixa, internacionalmente ou em português, sobre este assunto, e todos saibamos que tanto a noção de ontologia como de palavra (ou léxico) são complexas.² Para uma excelente discussão da diferença e relação entre ontologias e bases de dados lexicais, veja-se Hirst (2004).

Uma questão que se põe no entanto e sempre é a questão da cobertura ou abrangência de uma ontologia lexical, e também os tipos que a populam: ou seja, quais os membros da ontologia? Palavras, expressões, ou sentidos? Neste caso, desde a WordNet que os sentidos têm sido implicitamente definidos por um grupo de palavras sinónimas, o chamado *synset* em inglês, que aqui designamos por *nós* das ontologias. Outra questão é até que ponto recursos de cobertura vasta definidos com outros objectivos podem ser reformatados como ontologias lexicais. Para não excedermos o âmbito da nossa definição, não vamos considerar como ontologias lexicais nem o Port4Nooj nem as ontologias de cores e roupa definidas no âmbito da Linguateca para a anotação de corpos. Contudo, vamos aproveitar a existência desses recursos públicos para avaliar e comparar as ontologias lexicais (OL) propriamente ditas, ou seja, recursos que (i) pretendem cobrir o léxico todo e que (ii) o estruturam em termos de relações semânticas entre palavras. O que são palavras, itens lexicais ou nós de uma OL será aqui um dado empírico, baseando-nos nos nós das OL que vamos comparar. Veja-se de qualquer maneira Sampson (2000) e Gonçalo Oliveira *et al.* (2009) para uma discussão mais completa destas questões.

2. Breve apresentação dos recursos e sua descrição quantitativa

Passamos agora a descrever resumidamente os recursos que tratamos no presente artigo, quer as ontologias lexicais que focámos directamente, quer outros recursos relacionados, como é o caso do Port4Nooj, da WordNet.BR e da WordNet.PT.

O **PAPEL** (Gonçalo Oliveira *et al.*, 2008) é um recurso criado pela Linguateca a partir do Dicionário PRO de Língua Portuguesa da Porto Editora através de um protocolo de colaboração com o departamento de dicionários desta empresa, posto à disposição do público em <http://www.linguateca.pt/PAPEL/>, na sua versão 1.0, a 17 de Agosto de 2009.³

¹ Em 2007, Tony Veale ensinava, e assinava, a seguinte definição: *An ontology of lexical(-ized) concepts, used in NLP, serving as a lexical semantics.*

² Para os que preferem uma definição por exemplo, a WordNet (Fellbaum, 1998) e a MindNet (Richardson et al. 1998) são ontologias lexicais paradigmáticas.

³ Convém também talvez referir que o PAPEL foi iniciado pela Linguateca em 2005, ou seja, numa altura em que não havia nenhuma ontologia lexical para o português disponível. O protocolo com

Diferentemente de todas as outras ontologias lexicais para o português de que temos conhecimento, o PAPEL é público, grátis e utilizável por **todos** os actores de processamento da língua que o quiserem usar, e encontra-se aberto para subsequente melhoria pela comunidade. A versão 1.1 (ver tabela 1) já foi melhorada graças ao retorno dos utilizadores.

O conteúdo do PAPEL, correspondente a cerca de uma pessoa-ano de trabalho, foi obtido semi-automaticamente segundo inspiração da MindNet, o que significa que há muitas incorrecções e faltas neste recurso, a maior parte delas devidas à simplicidades dos processos empregues e não ao dicionário subjacente. Contudo, considerámos importante disponibilizá-lo como uma semente para trabalhos futuros, e também para desenvolver métodos automáticos ou semi-automáticos de avaliação ou revisão da extracção de relações entre palavras. Uma validação inicial do PAPEL com base quer no TeP quer em estudos com corpos foi apresentada em Gonçalo Oliveira *et al.* (2009). A validação humana de um conjunto de relações encontra-se em curso.

Tipo de relação	Tamanho
SINONIMO_DE	80429
HIPERONIMIA	63454
LOCAL_ORIGEM_DE	774
PARTE_DE	10645
PARTE_DE_ALGO_COM_PROPRIEDADE	3715
PROPRIEDADE_DE_ALGO_PARTE_DE	953
CAUSADOR_DE	1124
CAUSADOR_DE_ALGO_COM_PROPRIEDADE	16
PROPRIEDADE_DE_ALGO_QUE_CAUSA	505
CAUSADOR_DA_ACCAO	39
ACCAO_QUE_CAUSA	6421
PRODUTOR_DE	926
PRODUTOR_DE_ALGO_COM_PROPRIEDADE	31
PROPRIEDADE_DE_ALGO_PRODUTOR_DE	348
FINALIDADE_DE	2911
FINALIDADE_DE_ALGO_COM_PROPRIEDADE	23
FINALIDADE_DA_ACCAO	14
ACCAO_FINALIDADE_DE	5638
ACCAO_FINALIDADE_DE_ALGO_COM_PROPRIEDADE	266
PROPRIEDADE_DE_ALGO_REFERENTE_A	3694
PROPRIEDADE_DO_QUE	17028
MANEIRA_POR_MEIO_DE	1433

Tabela 1: Relações presentes no PAPEL v.1.1 e sua quantidade

A **MWN.PT** compra-se através da ELRA⁴, que comercializa este recurso através de duas licenças diferentes (uma para uso académico, outra para uso comercial), encontrando-se também acessível para consulta na rede a partir do endereço <http://mwnpt.di.fc.ul.pt>. Existe muito pouca informação ou documentação sobre a sua criação, mas o seu anúncio na lista `forum-lp` no princípio de Junho de 2009, assim como a página <http://mwnpt.di.fc.ul.pt/features.html>, consultada a 8 de Setembro de 2009, relata que a *MWN.PT - MultiWordnet of Portuguese (version 1)* contém 17.200 *synsets* (nós) validados manualmente, correspondentes a 16 mil lemas que dizem respeito ao português de Portugal e do Brasil. Além disso, na página de rede acima referida é indicado que este recurso está alinhado com a WordNet de Princeton, assim como, por transitividade, com as EuroWordNets do italiano, do espanhol, do romeno, do hebraico e do latim.

Quanto às relações incluídas, além de sinonímia, hiponímia e hiperonímia, é referido que a MWN.PT cobre as subontologias de *Person, Organization, Event, Location*, e *Art works*, assim como os 98 conceitos básicos (*Base Concepts*) sugeridos pela Global Wordnet Association, e os 164 conceitos nucleares básicos utilizados (*Core Base Concepts*) pela EuroWordNet. De acordo mais uma vez com a página acima referida, o acesso através da rede à MWN.PT existe desde Maio de 2008. A tabela 2 apresenta valores mais pormenorizados sobre o número de relações presentes na MWN.PT, calculados por nós após compra do recurso. A questão do cálculo do número de sinónimos foi feita da seguinte forma: para cada nó com mais do que um elemento, calcularam-se todos os pares possíveis. Depois, o formato original foi transformado em dois outros, correspondentes à expansão dos nós para triplos (ou seja, relações binárias entre duas “palavras”), no primeiro caso mantendo como elementos dos triplos os sentidos, no segundo caso juntando todos os sentidos da mesma palavra, que denominámos por “palavras ambiguadas” (P.A.).

Tipo de relação	Nós	Triplos	Palavras	Triplos	P.A.
IS-MEMBER-OF	5829	6330	6187	6325	6123
CONT	11220	6332	5941	6327	5883
HAS-HYPERNYM	17932	23869	18615	23389	15871
IS-PART-OF	400	513	592	504	542
IS-VALUE-OF	132	20	108	16	104
IS-SUBSTANCE-OF	12	9	19	9	19
HAS-HYPONYM	30010	23869	16322	23389	13605
HAS-PART	805	514	692	505	632
HAS-SUBSTANCE	35	9	47	9	46
Total	68747	68735	18615	66392	15871

Tabela 2: Relações presentes na MWN.PT: na contagem dos triplos apenas considerámos casos de palavras diferentes de GAP, PSEUDOGAP ou vazias

a Porto Editora foi assinado em Maio de 2006, mas os trabalhos definitivos foram apenas iniciados em Setembro de 2007.

⁴ http://catalog.elra.info/product_info.php?products_id=1101

Contudo, cedo nos demos conta de que grande parte das relações encontradas tinham apenas um dos elementos preenchido, ou seja, era possível encontrar casos de *fazenda*, *pano*, *tecido*, *tela* HAS-HYPERNYM: *artefacto* (PT), *artefato* HAS-PART:, em que o valor de HAS-PART: não continha nada.⁵ Além disso, em muitos casos os elementos relacionados contêm os lemas “GAP!” e “PSEUDOGAP!”. De acordo com a nossa interpretação⁶, os casos de “GAP!” indicam casos de palavras em inglês que não têm correspondente lexical em português. Por exemplo, na WordNet inglesa existe uma relação de hiperonímia entre *actor* e os dois nós *character_actor* (um actor especializado em papéis secundários) e *walk-on* (alguém que desempenha um pequeno papel numa produção dramática), correspondendo às seguintes relações: *actor* HAS-HYPONYM *character_actor* e *actor* HAS-HYPONYM *walk-on*. Mas, como os compiladores da MWN.PT consideraram que não existiam conceitos lexicalizados equivalentes em português nem para *character actor* nem para *walk/on*, a tradução das relações acima, sendo que *actor* é o correspondente em português (de Portugal) ao inglês *actor*, passou a *actor* HAS-HYPONYM GAP! e *actor* HAS-HYPONYM GAP!.

O TeP, Thesaurus Eletrónico de Português, foi desenvolvido em 2000-2001 pela equipa liderada por Bento Dias da Silva (Dias da Silva & Moraes, 2003, Dias da Silva *et al.*, 2000, 2002). O objectivo do trabalho era a construção de um recurso que pudesse ser usado como thesaurus de um processador de texto, para ajudar os utilizadores a encontrarem palavras diferentes para exprimirem as suas ideias.

A versão 2.0 do TeP surgiu em 2008. Consiste na mesma base de dados, mas encontra-se disponível na Web, com uma interface que permite o acesso interactivo aos mesmos dados (Maziero *et al.*, 2008). Todos os dados também se encontram disponíveis para download, para fins de pesquisa (os autores devem ser contactados para fins comerciais)..

Dada a aplicação que se pretendia dar aos dados, os autores optaram desde o início por catalogar apenas relações de sinonímia e antonímia. Contudo, embora o objectivo inicial fosse relativamente limitado, tomaram a decisão de representar os dados de forma semelhante à WordNet, na perspectiva de o recurso vir a evoluir nesse sentido, o que veio a acontecer com o projecto WordNet.BR (Dias-da-Silva *et al.*, 2006, 2008), iniciado em 2002 e actualmente ainda em curso.

Apesar das ferramentas usadas, a construção do TeP foi um trabalho essencialmente manual (por contraste com a metodologia usada no MindNet ou no PAPEL), sobre um corpo de referência constituído por vários dicionários electronicamente disponíveis, mas orientados apenas para o uso humano. Os dados encontram-se disponíveis em ficheiro de texto, com uma estrutura simples consistindo num grupo de sinónimos por linha, como

⁵ De notar, que devido à falta completa de documentação, esta é a nossa interpretação sobre o que acontece e qual o significado destas faltas, mas não podemos garantir que esteja correcta.

⁶ Do conteúdo de <http://multiwordnet.fbk.eu/english/whatin.php>, visitado em Novembro de 2009.

a seguir se mostra: 11519. [Substantivo] {aberta, saída, solução} <16854>, em que o primeiro campo representa o identificador do nó e o segundo a sua categoria gramatical. O conjunto representado entre chavetas é o grupo propriamente dito e o identificador que pode surgir no final consiste numa referência para um nó com um sentido oposto (antónimo) ao corrente.

Existem 19.885 destes registos, dos quais 4.312 se encontram relacionados pela relação de antonímia. (Ou seja, existem 2.156 relações de antonímia entre nós.) O número médio de palavras por nó é 3,8, enquanto que a moda é 2. Consta-se que 85% dos nós têm menos de seis palavras, enquanto que o maior nó compreende 53 palavras. Na tabela 3 apresentamos a distribuição do tamanho dos nós no TeP 2.0.

1	2	3	4	5	6	7	8	9	10	>10
1088	7311	4355	2558	1472	874	592	413	265	209	693

Tabela 3: Tamanho dos nós – grupos de sinónimos, *synsets* – no TeP 2.0

Se fizermos o exercício de desmultiplicar estes grupos em triplos, seja de sinonímia, seja de antonímia, ignorando a informação relativa à categoria gramatical, constatamos que o TeP 2.0 representa 202.514 relações de sinonímia e 49.141 relações de antonímia, envolvendo um total de 44.325 palavras distintas (com 75.713 sentidos diferentes).

Na tabela 3 apresentamos a distribuição em termos de categoria gramatical do TeP, tanto dos nós, como do número de palavras incluídas (separando ou não (“P.A.”) as formas por sentido). Além disso ainda apresentamos na última linha o número de formas distintas presentes neste recurso (independentemente da categoria gramatical, ou seja *colher sf* e *colher v* contam como uma forma).

Categoria gramatical	Nós	Triplos	Palavras	Triplos P.A.	P.A.
Adjectivos	6647	66775	21821	64354	15001
Advérbios	566	2089	1495	2063	1138
Substantivos	8526	63953	29659	60417	17276
Verbos	4145	133303	22738	125362	10910
Todas as palavras com c.g.		266129	75713	252196	44325
Todas as palavras sem c.g.		266123	75713	251655	43118

Tabela 4: Tamanho do TeP 2.0, em termos de categoria gramatical e total

A **WordNet.PT** (Marrafa, 2001, 2002) é um recurso linguístico desenvolvido pelo Centro de Linguística da Universidade de Lisboa em colaboração com o Instituto Camões. Este recurso encontra-se teoricamente disponível para consulta em <http://cvc.instituto-camoes.pt/wordnet/>⁷, não estando disponível para uso noutros sistemas. De acordo com

⁷ Há mais de dois anos que os autores do presente artigo tentam contudo, sem sucesso, interrogar o recurso, que parece estar com problemas graves a nível informático.

a informação presente na página, o seu desenvolvimento começou em 1998 e estende-se até hoje, passando por diversas versões, sendo a actual a 1.5, que compreende cerca de 10.000 nomes (8.000 nomes comuns e 1000 nomes próprios), cerca de 600 verbos e 800 adjectivos.

As relações que a WordNet.PT integra são (i) geral/específico, (ii) todo/parte, (iii) internas à estrutura do evento (co-relações), (iv) função não eventiva, (v) internas à estrutura do evento (simples), (vi) de oposição.

Este recurso abrange vários subdomínios semânticos: actividades profissionais e artísticas, alimentos, áreas geográficas e político-administrativas, instituições, instrumentos, meios de transporte, obras de arte, saúde e actos médicos, seres vivos, vestuário e vias de comunicação. A estrutura da WordNet.PT baseia-se no modelo da EuroWordNet e utilizou a ferramenta Polaris da EuroWordNet para fazer a construção do recurso, basicamente manual, e inspirada na WordNet de Princeton.

O **Port4NooJ** (Barreiro, 2008) é um conjunto de recursos linguísticos construídos no ambiente de desenvolvimento linguístico do NooJ (Silberstein, 2008), tendo em vista o processamento automático do português, publicamente disponíveis no repositório da Linguatca desde Outubro de 2008⁸, e que são usados em várias ferramentas públicas para o português e outras línguas, tal como o Corpógrafo e o ReEscreve.

Estes recursos, cujo desenvolvimento começou em 2006, correspondem a léxicos e a gramáticas com finalidades diversas: análise morfológica, sintáctico-semântica, desambiguação, identificação de unidades lexicais multipalavra, parafraseamento e tradução. Apesar de estes recursos não se encontrarem ainda completamente depurados, pareceu tanto à autora dos mesmos como à Linguatca que a sua disponibilização seria útil para a comunidade do processamento da língua portuguesa, pelo seu potencial e pelas suas características únicas. No Port4NooJ, a cada palavra (entrada lexical) pode estar associada uma ou mais relações sintáctico-semânticas com outras palavras, estabelecendo com elas uma relação de sinonímia, meronímia, hiperonímia, etc. O léxico está classificado em mais de 1000 categorias distintas, baseadas na taxonomia do sistema Logos, SAL (*syntactic-semantic abstract language*)⁹. Além da classificação taxonómica, os recursos incluem também relações morfossintácticas regulares, tais como a relação entre um adjectivo e um advérbio dele derivado (*rápido* > *rapidamente*; *acessível* > *acessivelmente*), entre um substantivo e um adjectivo (*entusiasmo* > *entusiasmado*), ou entre um substantivo e um advérbio (*imaginação* > *imaginativamente*). As relações semânticas incluídas no Port4NooJ, contudo, ao contrário dos outros recursos discutidos no presente artigo, encontravam-se apenas implicitamente no Port4Nooj, ou melhor na sua aplicação ao parafraseamento. Em teoria, a exploração do léxico em conjunto com

⁸ <http://www.linguatca.pt/Repositorio/Port4NooJ/>

⁹ A descrição das categorias SAL pode ser consultada nos Arquivos do Sistema Logos em <http://logosystemarchives.homestead.com/> ou nos recursos do OpenLogos em <http://logos-os.dfki.de/>.

a ontologia SAL permitiria a obtenção de um sem-número de relações entre as palavras. Contudo, para o presente artigo limitámo-nos a explorar as que pareciam cobrir um maior número de itens e que pudessem ser extraídas de forma simples¹⁰.

Neste momento, o Port4NooJ conta com perto de 30 mil relações morfossintácticas entre elementos semanticamente relacionados. Note-se que os resultados apresentados na Tabela 5 não se referem à capacidade de parafraseamento, mas sim às relações entre itens lexicais. Os resultados totais para parafraseamento são significativamente superiores, dado que as gramáticas, aplicadas à informação descrita no dicionário, permitem reconhecer e analisar expressões como *de (um) modo rápido*, *de (uma) forma/maneira rápida* (que poderiam ser contabilizadas como relações entre adjectivo e advérbio, mas que o não foram aqui), e que contemplam também as formas flexionadas como em *dar uns passeios*, etc.

Relação	Quantidade
Hiponímia	14963
Sinonímia (entre verbos, nomes, adj e advérbios)	10395 (5367, 20, 34, 5014)
Acção de	3773
Resultado de	283

Tabela 5: Relações semânticas presentes no Port4NooJ v. 2.0 e sua quantidade

3 Comparação dos vários recursos

Em primeiro lugar, tentámos estabelecer algum relacionamento entre os diversos tipos de informação existentes nos recursos de que nos ocupamos aqui. Esta aproximação de mapeamento foi realizada com base tanto no nome das relações, como também na sua descrição e exemplos (quando disponíveis). A Tabela 6 mostra os recursos em termos de relações contempladas, os números indicando quantos subtipos numa família de relações:

Recurso	MWN.PT	PAPEL	Port4NooJ	TeP	WordNet.PT
Sinonímia	sim	sim	sim	sim	sim
Antonímia	não	não	não	sim	sim
Hiponímia	2	1	sim	não	4
Meronímia	6	3	não	não	2
Causa	não	5	não	não	4
Objectivo	não	5	sim	não	12
Lugar	não	1	não	não	2
Maneira	não	1	sim	não	2

Tabela 6: Tabela comparativa das ontologias lexicais em termos de relações contempladas

¹⁰ O resultado dessa extracção, que, salientamos, não foi ainda revisto, encontra-se publicamente acessível em http://www.linguateca.pt/Repositorio/Port4NooJ/relacoes_semanticas_explicitas/.

Outra variável que faz sentido comparar é a extensão dos vários recursos, tanto em termos de quantidade como de tipo de nós (palavras simples ou também expressões), como as variantes a que se refere, o que é feito na Tabela 7. Note-se que a MWN.PT não tem informação sobre categoria gramatical, e que os valores sobre a WordNet.PT não puderam ser confirmados. Não incluímos o Port4NooJ devido à sua organização diferente.

Recurso	MWN.PT	PAPEL	TeP	WordNet.
Palavras (P.A.)	21091	99783	44296	
Nós (“synsets”)	17281	-	19885	c. 11400
Triplos (P/P.A.)	129049/95231	-/195652	251656/252196	
Variantes (Port/Br)	PP/PB	PP	PB	PP
Expressões	8599-40,8%	18165-18,2%	344-0,8%	sim
Substantivos (P.A.)	-	55372	17276	c. 10000
Adjectivos (P.A.)	-	24089	15001	c. 600
Verbos (P.A.)	-	18933	10910	c. 800
Advérbios (P.A.)	-	1389	1138	-

Tabela 7: Tabela comparativa das ontologias lexicais em termos da sua extensão

Finalmente, na Tabela 8 tentamos produzir informação sobre outras características importantes destes recursos, embora nem sempre seja fácil obter informação fidedigna.

Recurso	MWN.PT	PAPEL	TeP	WordNet.PT
Disponibilidade	paga e c/restrições	total	só para inv	Não
Validação	completa	incipiente	completa ¹¹	completa?
Tempo usado	?	c. 1 ano	?	?

Tabela 8: Tabela comparativa das ontologias lexicais em termos de outras qualidades

Um dos problemas da comparação de recursos de cobertura vasta é a impossibilidade de verificar individualmente que cobrem exactamente as mesmas questões com as mesmas hipóteses e análises. Por isso, uma das primeiras experiências que fizemos foi comparar aleatoriamente a cobertura dos diferentes recursos, usando duas formas diferentes de compilar triplos para análise.

Na primeira, cada autor do artigo forneceu independentemente dez relações que lhe pareciam dever estar numa ontologia lexical do português, chegando pois a 80 casos¹², dos quais 3 eram apenas aceitáveis em PP. Estamos contudo plenamente conscientes de que, estando muitos dos autores envolvidos na criação ou teste de alguns dos recursos (e não doutros), não se pode considerar que a amostra seja independente. Os resultados

¹¹ No sítio do TeP, <http://www.nilc.icmc.usp.br/tep2>, menciona-se que este recurso está em contante aprimoramento, o que significa que é possível que novas versões sejam lançadas.

¹² No anexo 1, além de em <http://www.linguateca.pt/PAPEL/comparacaoOL/>.

encontram-se na tabela 9. Consideramos como existente quer uma relação directa quer uma inversa. Raciocínio simples (não aplicável ao TeP) significa seguir automaticamente uma sequência de relações, tal como: cão HIPÓNIMO-DE mamífero e mamífero HIPÓNIMO-DE animal para obter cão HIPÓNIMO-DE animal, ou país HIPÓNIMO-DE espaço SINÓNIMO-DE lugar, para aceitar país HIPÓNIMO-DE lugar.

Recurso	MWN.PT	PAPEL	Port4NooJ	TeP
Existência directa	4	22	15	27
Existência após raciocínio simples	4	28	16	27

Tabela 9: Tabela comparativa em relação aos 80 primeiros casos

Para conseguir uma amostra totalmente independente, criámos outro conjunto de 80 triplos – no anexo 2 – extraindo aleatoriamente 20 casos de cada um dos quatro recursos, cujos resultados se encontram na tabela 10.

Recurso	MWN.PT	PAPEL	Port4NooJ	TeP
MWN.PT	20	3	0	2
PAPEL	0	20	1	0
Port4NooJ	0	0	20	0
TeP	0	4	0	20
Total em 80 /possíveis	20 / 48	27 /52	21 / 31	22 / 32

Tabela 10: Tabela comparativa em relação aos segundos 80 triplos: quantas das 80 se encontram

Se à primeira vista estes resultados parecem muito maus, um pouco de reflexão (e consideração das listas e das diferenças entre os recursos) permite-nos compreender várias coisas: Em primeiro lugar, a segunda comparação demonstra bem que estes recursos serão muito mais complementares do que comparáveis, e em segundo, o maior sucesso da primeira experiência parece dever-se ao facto de que as relações extraídas por eliciação de falantes de português são bem mais relevantes do que relações extraídas aleatoriamente de vocabulários correspondentes ao conhecimento passivo (se tanto) da língua.

4. Estudo qualitativo de alguns campos semânticos

Não pretendemos obviamente esgotar a análise ou avaliação destes assuntos neste artigo mas pareceu-nos interessante fazer um estudo inicial de algumas áreas com as quais tínhamos alguma familiaridade, e que passamos a descrever brevemente aqui.

A relação de **meronímia** – ou parte-todo – costuma ser caracterizada pela sua abrangência e, conseqüentemente, dificuldade de formalização, veja-se Cruse (1986). Iris *et al.* (1988:261) chegam inclusive a descrevê-la como “particularmente misteriosa e controversa”. Ao considerá-la uma família de relações, Cruse (1986) distingue, em um primeiro momento, as relações de meronímia canónicas (*selim* é necessariamente parte de

bicicleta, e *bicicleta* inclui necessariamente *selim*) e as não canônicas (o clássico exemplo da *maçaneta* – toda *maçaneta* pertence a uma porta, mas ter *maçaneta* não é condição necessária para algo ser uma *porta*), embora esclareça que não são poucos os casos de relações “não canônicas”. Cruse aponta ainda uma série de dimensões que poderiam ser levadas em consideração na análise do amplo conjunto de relações do tipo parte-todo: a concretude (*carros, árvores* vs. *coragem, adolescência*); o nível de diferenciação (partes de um *carro* vs. partes de uma *equipa*); o nível de integração entre as partes (*membros* de uma *equipa* vs. *livros* em uma *biblioteca*); e a propriedade de os itens analisados serem contáveis ou não, o que reflectiria diferentes níveis de individuação – por exemplo, *areia* (incontável) é composta por *grãos* (contável). Já Iris et al, para ficarmos apenas nestes dois exemplos, com base na análise da expressão linguística da relação de parte-todo em um dicionário, consideram-na uma família de quatro relações, em que as seguintes dimensões são consideradas: “functional part” (*selim-bicicleta*), “segmented whole” (*gomo de laranja*), “collection element” (*navio – esquadra*) e “set-subset” (*maçã – fruta*), sendo os dois primeiros grupos os mais genéricos. Dada a variedade de recortes para o fenómeno, não é de estranhar que diferentes trabalhos e, conseqüentemente, recursos codifiquem a relação de diferentes maneiras. Além dos quatro tipos de Iris et al. (1988), Winston et al. (1987), por exemplo, sugerem seis diferentes tipos de meronímia, mas, no trabalho de 1988 (Chaffin et al., 1988) são oito os tipos propostos.

No caso dos recursos analisados aqui, a EuroWordNet (com a qual a WordNet.PT se alinha) distingue cinco tipos de meronímia, a MWN.PT distingue três e o PAPEL outros três. Embora essa variedade ofereça um rico material para a investigação linguística, limita o alcance de uma comparação efectiva entre os diferentes recursos.

A seguir detalhamos a expressão da meronímia nos recursos analisados. A MWN.PT exprime a noção de meronímia por meio de três relações distintas: CONT (e a inversa IS-MEMBER-OF), HAS-PART (e a inversa IS-PART-OF), e HAS-SUBSTANCE (e a inversa IS-SUBSTANCE-OF). A relação CONT / IS-MEMBER-OF ocorre entre grupos e seus membros, HAS-PART/IS-PART-OF relacionam um todo e as partes que o compõem, e HAS-SUBSTANCE/IS-SUBSTANCE-OF ocorre entre uma substância e os elementos que a compõem: Assim temos por exemplo *alcateia, bando, gang, gangue, quadrilha* CONT: *gangster, mafioso*; *alcateia, bando, gang, gangue, quadrilha* IS-MEMBER-OF: *submundo do crime e alegria, animação* IS-PART-OF: *feitio, temperamento, índole*. Na tabela 1 podíamos ver uma visão global da meronímia no MWN.PT, correspondendo a 18401 casos.

Como já apontado, o diferente tratamento que o PAPEL dá à meronímia torna difícil a comparação entre este recurso e a MWN.PT. As diferenças são a dois níveis: a primeira refere-se ao nome da relação: No PAPEL, os diversos tipos de relação parte-todo abrangidos pela meronímia estão disponíveis, actualmente, sob o rótulo único – e mais genérico – PARTE-DE. Deste modo, as seguintes relações do tipo PARTE-DE: a) *tigre* PARTE_DE *Felídeos*; *fadista* PARTE_DE *fadistagem*, b) *oxigénio* PARTE_DE *água*; *hidrocarboneto* PARTE_DE *vaselina*, e c) *cabo*

PARTE_DE vassoura; virologia PARTE_DE microbiologia, poderiam ser classificadas como membro_de, substancia_de, parte_de, respectivamente.

A segunda diferença diz respeito à categoria gramatical envolvida na relação. No PAPEL, a relação de meronímia ocorre não apenas entre substantivos (as relações PARTE_DE, 9.970 casos), mas também entre substantivos e adjetivos, correspondendo às relações PARTE_DE_ALGO_COM_PROPRIEDADE (3.806 casos) e PROPRIEDADE_DE_ALGO_PARTE_DE (900 casos), como exemplificado em leite PROPRIEDADE_DE_ALGO_PARTE_DE lácteo. É preciso também relembrar que os dados do PAPEL foram obtidos automaticamente e que em muitos casos os verbos empregues nas gramáticas de extracção (tal como *possuir* ou *ter*) são eles próprios vagos entre vários dos casos. Parece-nos pois que estes merecem uma revisão aturada.

Por fim, por não ter sido possível aceder aos dados da WordNet.PT, baseamos a nossa análise exclusivamente na documentação deste recurso. Como mencionado, a Wordnet.PT, por estar alinhada com a EuroWordNet, considera cinco tipos de meronímia: PARTE, MEMBRO, PORÇÃO, MATÉRIA, LOCAL. Quanto à categoria gramatical das palavras envolvidas na relação, na WordNet.PT, tal como na MWN.PT, consideram-se apenas substantivos.¹³

Dos nossos estudos anteriores sobre a **cor** (Inácio et al., 2009) já tínhamos algumas ideias sobre a forma como estas palavras (palavras denotando cor ou a sua explícita negação) se poderiam estruturar, e de facto os recursos semânticos sobre a cor (Silva & Santos, 2009) que estamos a desenvolver (contendo 1706 palavras agrupadas em 44 classes) podem ser considerados como uma ontologia especial sobre esse campo. Contudo, embora esta área tenha já recebido muita atenção quer em português quer internacionalmente, demo-nos conta de que não era especialmente tratada nos recursos e/ou ontologias lexicais aqui comparados, e por isso limitamo-nos aqui a umas breves considerações.

Além das relações padrão de sinonímia, hiponímia, antonímia e meronímia, exemplificadas respectivamente por vermelho-hemoglobina SINÓNIMO-DE vermelho-sangue, azul-mal HOPÓNIMO-DE azul, incolor ANTÓNIMO-DE colorido, e branco PARTE-DE azul-e-branco, podemos também definir as seguintes relações envolvendo palavras de cor, mas que não vimos contempladas em nenhum dos recursos semânticos aqui tratados: azul-pacífico COR-ASSOCIADA-A-SENTIMENTO PAZ; verde-oliveira COR-ILUSTRADA-PELO-VEGETAL oliveira; e verde-rã COR-ILUSTRADA-PELO-ANIMAL rã.

¹³ Embora a WordNet.PT também estabeleça relações semânticas entre palavras de diferentes categorias gramaticais, o que é aliás uma vantagem inegável deste recurso, tais relações são do tipo “relações de função”, e envolvem, por exemplo, a relação entre um verbo e um instrumento (*telefonar/telefone*) ou entre um verbo e um agente (*jogar/jogador*).

Além disso, nos variados campos ou áreas em que as palavras de cor têm outra acepção que não a simplesmente visual, palavras de cor entram em relações diversificadas como negro SINÓNIMO-DE mau SINÓNIMO-DE sombrio SINÓNIMO-DE escuro (em expressões como *futuro negro*, *presságios sombrios*, *negócios escuros*), cinzento - sombrio - incolor - deslavado - chato - maçador - desinteressante (em expressões avaliativas) e verde ANTÓNIMO-DE maduro (no sentido temporal/agrícola).

Bacelar do Nascimento & Carvalho (1995) chamam a atenção para o facto de que, pese embora o carácter serial da oposição entre *preto* e *branco* (ou seja, estas duas cores são co-hipónimos), estas palavras são frequentemente citadas como opostos binários (e não seriais), e sugerem uma abordagem corpórea para redefinir a antonímia, no seguimento de Justeson & Katz (1991).

5. Comentários finais

Embora aparentemente exista bastante material sobre as relações entre palavras em português, estamos ainda longe de ter, para efeitos de processamento automático, material bem documentado e consensual sobre as várias relações semânticas existentes no léxico da nossa língua. Este artigo não pretende pois ser mais do que um pequeno contributo para esse objectivo. De facto, não apresentámos mais do que uma primeira exploração de vários recursos, para a qual enfatizamos a necessidade de uma maior explicitação na própria semântica das diversas relações, e apelamos à construção colaborativa de recursos mais completos, complexos, e dinâmicos que permitam aliás também raciocínio sobre o significado entretecido na língua portuguesa.

Agradecimentos

Este trabalho foi desenvolvido no âmbito da Linguateca, co-financiada pelo governo português, pela União Europeia (FEDER e FSE), sob o contrato POSC/339/1.3/C/NAC, e também financiada pela UMIC e pela FCCN. O trabalho de Hugo Gonçalo Oliveira foi financiado pela FCT, através de uma bolsa SFRH/BD/44955/2008.

Referências

- Bacelar do Nascimento, Maria Fernanda & Anabela Carvalho (1995) Preto e branco ou branco e preto? (Como se combinam os nomes de cores). In *Actas do XI Encontro Nacional da Associação Portuguesa de Linguística* (Lisboa, 2-4 de Outubro de 1995), Lisboa: APL/Colibri, pp. 367-380.
- Barreiro, Anabela (2008) “Port4NooJ: Portuguese Linguistic Module and Bilingual Resources for Machine Translation”. In *Proceedings of the 2007 International NooJ*

- Conference* (Barcelona, Espanha, 7-9 Junho, 2007), Cambridge Scholars Publishing, pp.19-47.
- Chaffin, Roger, Douglas J. Herrmann & Morton Winston (1988) An empirical taxonomy of part-whole relations: Effects of part-whole relation type on relation identification. *Language and Cognitive Processes* 3, 1, January 1988, pp. 17 - 48.
- Cruse, Alan (2004) *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford. Oxford University Press, 2004.
- Dias-da-Silva, Bento Carlos (2003) O TeP: construção de um thesaurus eletrônico para o português do Brasil. *Boletim da Associação Brasileira de Lingüística (ABRALIN)*. Fortaleza: Imprensa Universitária, 26, número especial, pp. 86 - 89.
- Dias-da-Silva, Bento Carlos & Helio Roberto de Moraes (2003) A construção de um thesaurus eletrônico para o português do Brasil. *ALFA* 47, num. 2, pp. 101-115.
- Dias-da-Silva, Bento Carlos, Mirna Fernanda de Oliveira & Helio Roberto de Moraes (2002) Groundwork for the development of the Brazilian Portuguese WordNet. In Nuno Mamede & Elisabete Ranchhod (eds.), *Advances in Natural Language Processing: Third International Conference, Proceedings (PorTAL 2002)* (Faro, Portugal, 23-26 June 2002), Berlin/Heidelberg: Springer-Verlag, pp. 189-196.
- Dias-da-Silva, Bento Carlos, Helio Roberto de Moraes, Mirna Fernanda de Oliveira, Ricardo Hasegawa, Daniela Amorim, Christie Passchoalino & Ana Cláudia Nascimento (2000) Construção de um thesaurus eletrônico para o português do Brasil. In Maria das Graças Volpe Nunes (ed.) *V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR 2000)*, pp. 1-10.
- Dias-da-Silva, Bento C., Ariani Di Felippo & Ricardo Hasegawa (2006) Methods and Tools for Encoding the WordNet.Br Sentences, Concept Glosses, and Conceptual-Semantic Relations . In Renata Vieira, Paulo Quaresma, Maria da Graça Volpes Nunes, Nuno J. Mamede, Cláudia Oliveira & Maria Carmelita Dias (eds.) *7th Workshop on Computational Processing of Written and Spoken Language (PROPOR '2006)* (Itatiaia, RJ, 13-17 de Maio de 2006), Springer, pp. 120-130.
- Dias-da-Silva, Bento C., Ariani Di Felippo & Maria das Graças Volpe Nunes (2008) The Automatic Mapping of Princeton WordNet Lexical-Conceptual Relations onto the Brazilian Portuguese WordNet Database. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, (Marraquexe, Marrocos, 28-30 Maio 2008), European Language Resources Association (ELRA).
- Fellbaum, Christiane (1998) (ed.) *WordNet: An Electronic Lexical Database*, with a preface by George Miller. The MIT Press.
- Gonçalo Oliveira, Hugo, Diana Santos, Paulo Gomes & Nuno Seco (2008) PAPEL: a dictionary-based lexical ontology for Portuguese. In António Teixeira, Vera Lúcia Strube de Lima, Luís Caldas de Oliveira & Paulo Quaresma (eds.) *Computational Processing of the Portuguese Language, 8th International Conference, Proceedings (PROPOR 2008)* (Aveiro, Portugal, 8-10 de Setembro, 2008), Springer Verlag, pp. 31-40.
- Gonçalo Oliveira, Hugo, Diana Santos & Paulo Gomes (2010) Extracção de relações

- semânticas entre palavras a partir de um dicionário: o PAPEL e sua avaliação. *Linguamática* 2, 1, Abril de 2010, pp. 77-94.
- Hirst, Graeme (2004) Ontology and the lexicon. In Steffen Staab & Rudi Studer (eds.). *Handbook on ontologies*, Springer, pp. 209-229.
- Inácio, Susana, Diana Santos & Rosário Silva (2008) COMPARando cores em português e inglês. In Sónia Frota & Ana Lúcia Santos (eds.) *Artigos seleccionados do XXIII Encontro da Associação Portuguesa de Linguística (APL)*, pp. 271-286.
- Iris, Madelyn A., Bonnie Litowitz & Martha Evans (1998) Problems of the part-whole relation. In Martha Evans (ed.) *Relational models of the lexicon: representing knowledge in semantic networks*. Cambridge: Cambridge University Press, pp. 261-288.
- Justeson, John S. & Slava M. Katz (1992) Redefining Antonymy: The Textual Structure of a Semantic Relation. *Literary and Linguistic Computing* 7(3), pp. 176-184.
- Marrafa, Palmira (2001) *WordNet do Português: uma base de dados de conhecimento linguístico*. Instituto Camões.
- Marrafa, Palmira (2002) Portuguese WordNet: general architecture and internal semantic relations. *DELTA: Documentação de Estudos em Linguística Teórica e Aplicada*, 18, pp. 131-146.
- Maziero, Erick G., Thiago Pardo, Ariani Di Felippo & Bento Carlos Dias-da-Silva (2008) A Base de Dados Lexical e a Interface Web do TeP 2.0 - Thesaurus Eletrônico para o Português do Brasil. In *VI Workshop em Tecnologia da informação e da linguagem humana (TIL)* (Vila Velha, ES, Brasil, 28-29 Outubro 2008), pp. 390-392.
- Richardson, Stephen, William B. Dolan & Lucy Vanderwende (1998) MindNet: acquiring and structuring semantic information from text. *Proceedings of the 17th International Conference on Computational Linguistics, COLING-ACL'98*, pp. 1098-1102.
- Sampson, Geoffrey R. (2000) Review of Christiane Fellbaum (ed.) *Wordnet: An Electronic Lexical Database*, 1998, *International Journal of Lexicography*, 13, pp. 54-59.
- Silberztein, Max (2008) Complex Annotations with NooJ. In Xavier Blanco & Max Silberztein (eds.) *Proceedings of the 2007 International NooJ Conference*. Newcastle, Cambridge Scholars Publishing, pp. 214-227.
- Silva, Rosário & Diana Santos (2009) Arco-íris: notas sobre a anotação do campo semântico da cor em português. <http://www.linguateca.pt/acesso/ArcoIris.pdf>.
- Veale, Tony (2007) Enriched Lexical Ontologies: Adding new knowledge and new scope to old linguistic resources, Curso na *ESSLLI 2007*, Dublin, Agosto de 2007.
- Winston, Morton E., Roger Chaffin & Douglas Herrmann (1987) A taxonomy of part-whole relations. *Cognitive Science* 11, 4, October-December, pp. 417-444.

Anexo 1: 80 exemplos criados pelos autores (Aplicável, Existente, Inferido)

			TeP		Port4NooJ			MWN		PAPEL		
			A	E	A	E	I	A	E	A	E	I
alegre	ANTÓNIMO_DE	triste	✓	✓								
alegre	SINÓNIMO_DE	bêbado	✓		✓			✓		✓		
alegre	SINÓNIMO_DE	contente	✓	✓	✓	✓		✓		✓	✓	
ambulância	HIPÓNIMO_DE	veículo			✓			✓		✓		
amor-perfeito	HIPÓNIMO_DE	flor			✓	✓		✓		✓		
amor-perfeito	HIPÓNIMO_DE	planta			✓	✓		✓		✓	✓	
andar	SINÓNIMO_DE	caminhar	✓	✓	✓			✓		✓	✓	
animal	HIPERÓNIMO_DE	galinha			✓			✓		✓		
ansiedade	HIPÓNIMO_DE	sentimento			✓			✓		✓		
areia	PARTE_DE	praia						✓		✓		
asa	PARTE_DE	ave						✓		✓		
baleia	HIPÓNIMO_DE	mamífero			✓	✓		✓		✓		✓
barata	HIPÓNIMO_DE	insecto			✓	✓		✓		✓		
bêbedo	SINÓNIMO_DE	ébrio	✓	✓	✓			✓		✓	✓	
beleza	ANTÓNIMO_DE	fealdade	✓	✓								
belo	SINÓNIMO_DE	bonito	✓	✓	✓	✓		✓		✓	✓	
bicicleta	HIPÓNIMO_DE	veículo			✓	✓		✓		✓		✓
cadeira	SINÓNIMO_DE	disciplina	✓		✓	✓		✓		✓		
cansar	SINÓNIMO_DE	fatigar	✓	✓	✓			✓		✓	✓	
cão	HIPÓNIMO_DE	animal			✓	✓		✓		✓		
capítulo	PARTE_DE	livro						✓		✓		
carro	SINÓNIMO_DE	automóvel	✓	✓	✓	✓		✓	✓	✓	✓	
casa	PARTE_DE	cidade						✓		✓		
chefiar	SINÓNIMO_DE	liderar	✓	✓	✓			✓		✓		
comida	HIPERÓNIMO_DE	chanfana			✓			✓		✓		
dedos	PARTE_DE	mão						✓		✓		
diamante	HIPÓNIMO_DE	mineral			✓	✓		✓		✓	✓	
disco	PARTE_DE	computador						✓		✓		
			TeP		Port4NooJ			MWN		PAPEL		
			A	E	A	E	I	A	E	A	E	I
distanciar-se	SINÓNIMO_DE	afastar-se	✓	✓	✓			✓		✓		
escrever	SINÓNIMO_DE	redigir	✓	✓	✓			✓		✓	✓	
escritor	HIPÓNIMO_DE	peessoa			✓			✓		✓		
feijão	PARTE_DE	feijoada						✓		✓		
feliz	ANTÓNIMO_DE	infeliz	✓	✓								
fofo	SINÓNIMO_DE	macio	✓	✓	✓		✓	✓		✓	✓	
folha	PARTE_DE	árvore						✓		✓		
futebol	HIPÓNIMO_DE	jogo			✓	✓		✓		✓	✓	
ganhar	SINÓNIMO_DE	vencer	✓	✓	✓			✓		✓	✓	

RELAÇÕES SEMÂNTICAS EM PORTUGUÊS: COMPARANDO O TEP, O MWN.PT, O PORT4NOOJ

gato	HIPÓNIMO_DE	animal			✓	✓			✓		✓		
gordo	ANTÓNIMO_DE	magro	✓	✓									
gritar	SINÓNIMO_DE	berrar	✓	✓	✓				✓		✓	✓	
guiador	PARTE_DE	bicicleta							✓		✓		
hipismo	SINÓNIMO_DE	equitação	✓		✓				✓		✓	✓	
índice	PARTE_DE	livro							✓		✓		
livro	HIPERÓNIMO_DE	sebenta			✓				✓		✓		
livro	HIPÓNIMO_DE	obra			✓				✓		✓		
livro	MEMBRO_DE	biblioteca							✓		✓		
local	HIPERÓNIMO_DE	país			✓	✓			✓		✓		✓
luz	ANTÓNIMO_DE	treva	✓										
macio	ANTÓNIMO_DE	áspero	✓	✓									
mamífero	HIPÓNIMO_DE	vertebrado			✓				✓	✓	✓	✓	
medroso	ANTÓNIMO_DE	corajoso	✓	✓									
militar	PARTE_DE	exército							✓		✓		✓
oferecer	SINÓNIMO_DE	proporcionar	✓	✓	✓				✓		✓	✓	
osso	PARTE_DE	esqueleto							✓		✓		✓
ovelha	MEMBRO_DE	rebanho							✓		✓		
parafuso	HIPÓNIMO_DE	ferramenta			✓				✓		✓		
partida	PARTE_DE	campeonato							✓		✓		
perceber	SINÓNIMO_DE	ver	✓	✓	✓				✓		✓	✓	
			TeP		Port4NooJ		MWN		PAPEL				
			A	E	A	E			A	E	A		
perfeito	ANTÓNIMO_DE	imperfeito	✓	✓									
perna	PARTE_DE	cadeira							✓		✓		
poema	HIPÓNIMO_DE	obra			✓				✓		✓	✓	
pônei	HIPÓNIMO_DE	cavalo			✓				✓		✓		
portátil	HIPÓNIMO_DE	computador			✓				✓		✓		
quarto	PARTE_DE	casa							✓	✓	✓		✓
química	HIPÓNIMO_DE	ciência			✓				✓		✓	✓	
rádio	PARTE_DE	esqueleto							✓		✓		
relâmpago	PARTE_DE	trovoada							✓		✓		
rico	ANTÓNIMO_DE	pobre	✓	✓									
robusto	ANTÓNIMO_DE	frágil	✓	✓									
roda	PARTE_DE	carro							✓		✓		
rolha	PARTE_DE	garrafa							✓		✓		
saltar	SINÓNIMO_DE	pular	✓	✓	✓				✓		✓	✓	
sarampo	HIPÓNIMO_DE	doença			✓				✓		✓	✓	
simpático	SINÓNIMO_DE	agradável	✓	✓	✓	✓			✓		✓		✓
soldado	HIPÓNIMO_DE	militar			✓				✓	✓	✓		

sovar	SINÓNIMO_DE	bater	✓	✓	✓			✓		✓		
tampo	PARTE_DE	mesa						✓		✓		
tecla	PARTE_DE	teclado						✓		✓		
televisão	HIPÓNIMO_DE	electro-doméstico			✓			✓		✓		
xícara	SINÓNIMO_DE	chávena	✓	✓	✓			✓		✓	✓	
Total			41	27	47	15	1	7	4	7	22	6

Anexo 2: 80 exemplos extraídos aleatoriamente dos recursos (Aplicáve, Existente)

				TeP		Port4N		MWN		PAPEL	
				A	E	A	E	A	E	A	E
TeP	altear	SIN	erguer			✓		✓		✓	✓
TeP	bobo	SIN	inepto			✓		✓		✓	
TeP	afável	SIN	desafectado			✓		✓		✓	
TeP	coruto	SIN	extremidade			✓		✓		✓	
TeP	apercebimento	SIN	prevenção			✓		✓		✓	✓
TeP	debilitar-se	SIN	estiolar-se			✓		✓		✓	
TeP	acalantar	SIN	ninar			✓		✓		✓	✓
TeP	desaparecer	SIN	mergulhar			✓		✓		✓	
TeP	desarmonia	SIN	discordo					✓		✓	
TeP	edulcorar	SIN	malear			✓		✓		✓	
TeP	desacerto	SIN	mal-entendido			✓		✓		✓	
TeP	areado	SIN	estramontado			✓		✓		✓	
TeP	abluir-se	SIN	depurar-se			✓		✓		✓	
TeP	limpar	SIN	polir			✓		✓		✓	✓
TeP	bazofiar	SIN	ostentar			✓		✓		✓	
TeP	corporificar-se	SIN	incorporar-se			✓		✓		✓	
TeP	cobarde	ANT	impávido								
TeP	pontualidade	ANT	atraso								
TeP	deleite	ANT	pesadume								
TeP	desamarrar-se	ANT	enlaçar-se								
P4N	aquário	HIP	coisa de natureza instrumental					✓		✓	
P4N	Bielorrússia	HIP	local geográfico					✓		✓	
P4N	calçado	HIP	equipamento					✓		✓	
P4N	campeonato	HIP	condição					✓		✓	
P4N	Canadá	HIP	entidade geográfica					✓		✓	

RELAÇÕES SEMÂNTICAS EM PORTUGUÊS: COMPARANDO O TEP, O MWN.PT, O PORT4NOOJ

P4N	citrato	HIP	coisa massiva					✓		✓	
P4N	colagem	HIP	Informação/conhecimento/dados					✓		✓	
P4N	creosoto	HIP	coisa não contável					✓		✓	
				TeP		Port4N		MWN		PAPEL	
				A	E	A	E	A	E	A	E
P4N	desidratar	SIN	fazer desidratação	✓				✓		✓	
P4N	fazer aborto	SIN	abortar	✓				✓		✓	
P4N	grade	HIP	coisa de natureza instrumental					✓		✓	
P4N	Labrador	HIP	entidade geográfica					✓		✓	
P4N	lima	HIP	coisa não contável					✓		✓	
P4N	Madagáscar	HIP	entidade geográfica					✓		✓	
P4N	malignamente	SIN	de um modo maligníssimo	✓				✓		✓	
P4N	realizar abdução	SIN	abduzir	✓				✓		✓	
P4N	rena	HIP	mamífero					✓		✓	
P4N	saída	HIP	coisa de natureza instrumental					✓		✓	
P4N	sujamente	SIN	de forma suja	✓				✓		✓	
P4N	tediosamente	SIN	De uma forma tediosíssima	✓				✓		✓	
MW	região francesa	HIPÓ	área geográfica			✓				✓	
MW	família Compositae	MEM	ordem Campanulales								
MW	parasita	SIN	parasito	✓	✓	✓				✓	
MW	peçoal da segurança	HIPÓ	peçoal			✓				✓	
MW	conjuntura	HIPÓ	estado			✓				✓	✓
MW	protetor~(BR)	HIPER	guardador			✓				✓	
MW	pé-rapado	SIN	zé-ninguém	✓		✓				✓	
MW	habitante de Wyoming	HIPÓ	americano			✓				✓	
MW	investigador	HIPÓ	cientista			✓				✓	
MW	pega	HIPER	cocote			✓				✓	
MW	viajante	HIPÓ	ser humano			✓				✓	
MW	partidário	HIPER	federalista			✓				✓	
MW	exército	SIN	tropa	✓	✓	✓				✓	✓
MW	fingido	HIPÓ	trampolineiro			✓				✓	
MW	afiançador	HIPÓ	indivíduo			✓				✓	
MW	pastor	HIPER	cabreiro			✓				✓	✓

				TeP		Port4N		MWN		PAPEL	
				A	E	A	E	A	E	A	E
MW	família Solanaceae	CONT	género Capsicum								
MW	família Poaceae	CONT	género Oryzopsis								
MW	família Dinornithidae	CONT	género Dinornis								
MW	chefia	HIPÓ	status			✓				✓	
PAP	peessoa	HIPER	sobredotado			✓		✓			
PAP	dispositivo	HIPER	amortecedor			✓	✓	✓			
PAP	silvícola	PROP	dizer_respeito_às_matas								
PAP	rentável	PROP	produzir_lucro								
PAP	incorreção	SIN	deselegância	✓		✓		✓			
PAP	nefelibata	PROP	não_ter_o_sentido_das_realidades								
PAP	peessoa	HIPER	calista			✓		✓			
PAP	actividade	HIPER	gestão			✓		✓			
PAP	sepultante	PROP	sepultar								
PAP	queimada	HIPER	ucha			✓		✓			
PAP	haste	HIPER	dardo			✓		✓			
PAP	acção	HIPER	barrunto			✓		✓			
PAP	templo	HIPER	anfiprostilo			✓		✓			
PAP	operação	HIPER	cementação			✓					
PAP	tríparo	PROP	reproduzir_em_grupos_de_três								
PAP	característica	HIPER	singularidade			✓		✓			
PAP	ardilosamente	SIN	estrategicamente	✓		✓					
PAP	brigar	SIN	enguedelhar	✓		✓					
PAP	dito	HIPER	lorpice			✓		✓			
PAP	pássaro	HIPER	pisco			✓		✓			
				32	2	31	1	48	0	52	7