

Studies on Multilingual Lexicography

LEXICOGRAPHICA

Series Maior

Supplementary Volumes to the International Annual
for Lexicography
Suppléments à la Revue Internationale
de Lexicographie
Supplementbände zum Internationalen Jahrbuch
für Lexikographie

Edited by
Rufus Hjalmar Gouws, Ulrich Heid, Thomas Herbst,
Anja Lobenstein-Reichmann, Oskar Reichmann,
Stefan J. Schierholz and Wolfgang Schweickard

Volume 157

Studies on Multilingual Lexicography

Edited by
María José Domínguez Vázquez, Mónica Mirazo Balsa
and Carlos Valcárcel Riveiro

DE GRUYTER

ISBN 978-3-11-060467-2
e-ISBN (PDF) 978-3-11-060765-9
e-ISBN (EPUB) 978-3-11-060727-7
ISSN 0175-9264

Library of Congress Control Number: 2019945611

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>.

© 2020 Walter de Gruyter GmbH, Berlin/Boston
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Contents

María José Domínguez Vázquez, Mónica Mirazo Balsa and Carlos Valcárcel Riveiro
Studies on multilingual lexicography: an introduction — 1

Section 1: **Multilingual electronic lexicography in a new society**

Pedro A. Fuertes-Olivera and Henning Bergenholtz
Towards a new definition of multilingual lexicography in the era of internet — 9

Rufus H. Gouws
Metalexigraphic models for multilingual online dictionaries in emerging e-societies — 29

Sven Tarp
A dangerous cocktail: databases, information techniques and lack of vision — 47

Section 2: **Multilingual electronic dictionaries**

Olga Batiukova and Elena de Miguel
Multilingual Electronic Dictionary of Motion Verbs (DICEMTO): overall structure and the case of andar — 67

Maria Vittoria Calvi and Luis Javier Santos López
From the Linguaturismo glossary to the Dictionary of Food and Nutrition: proposal for a new electronic multilingual lexicography — 93

Gloria Corpas Pastor and Isabel Durán-Muñoz
INTELITERM: In search of efficient terminology lookup tools for translators — 113

María José Domínguez Vázquez and Carlos Valcárcel Riveiro
PORTLEX as a multilingual and cross-lingual online dictionary — 135

Isabel Durán-Muñoz and Gloria Corpas Pastor
Corpus-based multilingual lexicographic resources for translators: an overview — 159

Xavier Gómez Guinovart and Miguel Anxo Solla Portela
Construction of a WordNet-based multilingual lexical ontology for Galician — 179

Ignacio M. Palacios Martínez and Mario Cal Varela

Designing and compiling a terminological and multilingual dictionary for language teaching and learning: key issues and some reflections — 197

Janusz Taborek

Multilingual LSP dictionary — 213

Xavier Gómez Guinovart and Miguel Anxo Solla Portela

Construction of a WordNet-based multilingual lexical ontology for Galician

Abstract: This study describes the methodology used in the development of a WordNet lexicon for the Galician language, and its applications for language processing in the fields of terminology acquisition and ontology learning and management. First, we review the Princeton WordNet lexical model, its multilingual adaptation in the EuroWordNet framework, and its implementation in the Galician WordNet building. Second, we discuss the approach and the resources used in the design of Termonet, a tool for checking and verifying in technical corpora the specialty lexicons embedded in WordNet. This tool performs an identification of the synsets in WordNet belonging to a terminological domain from the semantic relations between the nodes of the lexical network, and validates the terms by means of a semantically disambiguated specialized corpus. Third, we analyze the process of construction of a new semantic categorization of WordNet based on epinonyms and generated automatically by exploring the relations from a terminological perspective. A WordNet epinonym is a noun synset in the semantic network representing the category of the semantic domain to which other synsets will be automatically assigned by algorithms that will evaluate their proximity from a terminological point of view through the cognitive processing of the lexical-semantic relations. Last, we present some applications of the RDF Galician WordNet in the Semantic Web by means of federated queries with lexical and ontological resources available as Linked Open Data (LOD) like DBpedia, BabelNet, Wiktionary and YAGO.

Keywords: computational lexicography, computational terminology, linked open data, multilingual language resources, ontologies, semantic web, WordNet

Xavier Gómez Guinovart: Grupo TALG, Universidade de Vigo, Departamento de Tradución e Lingüística, Campus Universitario, E-36310 Vigo, tel. +34 986812371, xgg@uvigo.es
Miguel Anxo Solla Portela: Grupo TALG, Universidade de Vigo, Departamento de Tradución e Lingüística, Campus Universitario, E-36310 Vigo, tel. +34 986812371, miguelsolla@uvigo.es

<https://doi.org/10.1515/9783110607659-010>

1 Introduction

This article¹ presents some aspects of the development of Galnet, a WordNet-based multilingual lexical ontology for Galician which is being built by the TALG Research Group (Galician Language Technologies and Applications) of the University of Vigo. Galnet is still under development but has already yielded interesting and useful results in the fields of Galician lexicology, semantics and automatic language processing. The following sections will describe the general lines of the project and its applications in the fields of terminology acquisition and ontology learning and management.

2 The WordNet model

WordNet is a lexical database of the English language, organized as a semantic network where the nodes are concepts represented as sets of synonyms and the links between nodes are semantic relations between lexical concepts (Fellbaum 1998, Miller et al. 1990). The nodes contain nouns, verbs, adjectives and adverbs grouped by synonymy. In WordNet terminology, a set of synonyms is called a synset, and each lemmatized synonym in a synset is called a variant and is considered a lexical variant of the same concept. Thus, each synset represents a distinct lexicalized concept and includes all the synonymous variants of this concept. Additionally, each synset may contain a brief definition or gloss, which is common to every variant in the synset, and, in some cases, one or more examples of the use of the variants in context.

In the WordNet model of lexical representation, the synsets are linked by means of lexical-semantic relations. Some of the most frequent relations represented in WordNet are hypernymy/hyponymy and holonymy/meronymy for nouns, antonymy and near synonymy for adjectives, antonymy and derivation for adverbs, and entailment, hypernymy/hyponymy, cause and opposition for verbs.

WordNet, which was originally developed for English, is now available in many other languages, although the English WordNet still stands as the most complete reference version. Created and maintained at Princeton University since 1985, version 3.0 contains 206,941 lemmas, i.e. synonymous variants (155,287 of which are unique, non-homographic forms) grouped into 117,659 sets of synonyms or synsets.

¹ This research has been carried out thanks to the project TUNER (TIN2015-65308-C5-1-R) supported by the Ministry of Economy and Competitiveness of the Spanish Government and the European Fund for Regional Development (MINECO/FEDER).

Today, WordNet is considered the most important resource in computational lexical semantics, especially in the field of natural language processing, where it is used in tasks of automatic semantic disambiguation (Agirre/Edmonds 2006), information retrieval (Zhao et al. 2012), answer extraction (Cai et al. 2016), machine translation (Vintar/Fišer/Vrščaj 2012), cross-language information retrieval (Agirre et al. 2007), cross-language question Answering (Ferrández et al. 2007), automatic text classification (Elberrichi et al. 2008), query expansion (Fang 2008), spell checking (Huang 2016), and automatic summarization (Plaza et al. 2010), among others. WordNet is also used with many benefits in computer applications related to language learning, for instance, in systems for the evaluation of the lexical competence of learners of English as a second language (Hu/Graesser 1998), for the automatic generation of gap-filling vocabulary exercises for smartphones (Knoop/Wilske 2013), or for WordNet-based vocabulary learning (Sun/Huang/Liu 2011).

Several WordNet versions are now available at different development stages in very diverse languages such as Hebrew (Ordan/Wintner 2007), Japanese (Isahara et al. 2008), Sanskrit (Bhingardive et al. 2014), Portuguese (Simões/Gómez Guinovart 2014), Spanish (Fernández/Vázquez 2010), Catalan (Oliver/Climent 2011) and Basque (Pociello/Agirre/Aldezabal 2011). Many of the WordNet versions in languages other than English follow the design model of EuroWordNet (Vossen 2002), where the synsets of a particular language are linked to the synsets of the other languages through an InterLingual Index (ILI) that is unique to each concept, and which is mainly based on the synsets of the English WordNet. Therefore, the set of WordNet lexicons in different languages allows the connection between the synsets of any pair of languages via the ILI, thus constituting a very useful resource in applications of linguistic technologies dealing with multilingual processing.

It is also worth noting that the concepts contained in the EuroWordNet database are categorized into domain hierarchies and ontologies, such as the WordNet Domains (Bentivogli et al. 2004), the Suggested Upper Merged Ontology (SUMO) (Pease/Niles/Li 2002) and the Top Concept Ontology (Álvarez et al. 2008), which allows the various applications benefiting from these semantic categorizations to make better use of the resource.

3 The Galnet project

The goal of the Galnet project (Gómez Clemente et al. 2013, Solla Portela/Gómez Guinovart 2015a, Álvarez de la Granja/Gómez Clemente/Gómez Guinovart 2016) is building a WordNet for Galician aligned with the ILI generated from the English WordNet 3.0, following the *expand model* (Vossen 2002) for the creation of new wordnets, where the variants associated with the Princeton WordNet synsets are obtained from different language resources (corpora, dictionaries, glossaries...)

using experimental methods of modern lexicography developed in the field of lexical knowledge acquisition. This project is part of a wider one aimed at the coordinated integration of the Spanish, Catalan, Galician, Basque and Portuguese versions of WordNet 3.0. The research groups participating in this project are IXA (Euskal Herriko Unibertsitatea/University of the Basque Country), TALP (Universitat Politècnica de Catalunya), GRIAL (Universitat Autònoma de Barcelona, Universitat de Barcelona, Universitat de Lleida and Universitat Oberta de Catalunya), IULA (Universitat Pompeu Fabra), and SLI/TALG (Universidade de Vigo), which is responsible for building Galnet.

Galnet is part of the Multilingual Central Repository (MCR) (González/Rigau 2013), a database that currently integrates wordnets from six different languages (English, Spanish, Catalan, Galician, Basque and Portuguese) using WordNet 3.0 as ILI and where each synset is classified under the WordNet Domains hierarchy, the SUMO ontology and the Top Concept Ontology.

Tab. 1: Synsets and variants by language

	English (WordNet 3.0)		Galician (Galnet 3.0.24)	
	<i>variants</i>	<i>synsets</i>	<i>variants</i>	<i>synsets</i>
<i>Nouns</i>	146,312	82,115	45,040	30,039
<i>Verbs</i>	25,047	13,767	6,541	2,785
<i>Adjectives</i>	30,002	18,156	10,039	6,135
<i>Adverbs</i>	5,580	3,621	1,038	706
<i>Total</i>	206,941	117,659	62,658	39,665
<i>Percent</i>	100%	100%	30%	34%
	Spanish		Portuguese	
	<i>variants</i>	<i>synsets</i>	<i>variants</i>	<i>synsets</i>
<i>Nouns</i>	101,027	55,227	17,149	10,047
<i>Verbs</i>	20,953	9,541	8,407	3,786
<i>Adjectives</i>	20,938	12,373	6,330	3,581
<i>Adverbs</i>	3,583	1,854	789	528
<i>Total</i>	146,501	78,995	32,675	17,942
<i>Percent</i>	71%	67%	16%	15%

	Catalan		Basque	
	variants	synsets	variants	synsets
Nouns	73,810	46,917	40,420	26,710
Verbs	14,619	6,349	9,469	3,442
Adjectives	11,212	6,818	148	111
Adverbs	1152	872	0	0
Total	100,793	60,956	50,037	30,263
Percent	49%	52%	24%	26%

The specific interface designed to query Galnet² extends the MCR functionalities by providing different types of navigation through domain hierarchies and ontologies, allowing an interactive tree-based visualization of synsets by their semantic relations, and including temporal values and sentiment scores for synsets from TempoWordNet³, SentiWordNet 3.0⁴ and ML-SentiCon⁵, a new presentation of information associated with synsets in Linked Open Data format with LodLive and Virtuoso Facets, a tool specifically designed for the extraction of lexical-semantic fields (Termonet), and a new terminology-oriented semantic categorization based on epinonyms. All these issues will be discussed in depth in the sections which follow.

Table 1 shows the current development status of WordNet in the six languages integrated in the project and available via the Galnet web interface. Table 1 includes the number of synsets and variants by language and part of speech, and shows the percentage progress obtained by each lexicon which is part of MCR with respect to the extension of English WordNet 3.0, and the updated data for Galician at Galnet version 3.0.24. It should be noted that the official distribution of Galnet done through the MCR platform, while being extremely important for the dissemination and use of the resource, is just a “frozen” version of the database, and the most up-to-date data can only be accessed directly from Galnet's web interface.

Beyond its internal network of semantic senses, Galnet is fully functional for users who are looking for purely lexicographic results. Galician variants in Galnet are linked to other multimedia resources highly relevant to Galician lexicography, such as the Diccionario da Real Academia Galega⁶, the Diccionario de Pronuncia da Lingua Galega⁷ (phonetic and audio transcription of monolexical lemmas), the

2 [<http://sli.uvigo.gal/galnet>]; last access: January 24, 2017].

3 [<https://tempowordnet.greyc.fr>]; last access: January 24, 2017].

4 [<http://sentiwordnet.isti.cnr.it>]; last access: January 24, 2017].

5 [<http://timm.ujaen.es/recursos/ml-senticon/>]; last access: January 24, 2017].

6 [<https://academia.gal/diccionario>]; last access: November 4, 2018].

7 [<http://ilg.usc.es/pronuncia/>]; last access: November 4, 2018].

encyclopedic contents of the Wikipedia structured through the Galician BDpedia⁸, or the photo gallery illustrating each synset obtained from Imagenet⁹.

The latest version of Galnet, together with other important lexical and textual resources, is also available for consultation on the RILG (Integrated Language Resources for Galician) website¹⁰.

4 Applications in terminology acquisition

WordNet was originally conceived in the context of psycholexicology, and structured by semantic relations between synsets belonging to different grammatical categories (Miller et al. 1990). It shows, therefore, many similarities with methodological aspects of terminology in terms of conceptual typology and structure. Similar semantic relations can be found in terminological repertoires such as the SNOMED Clinical Terms¹¹ or the Termoteca¹² terminological database (Gómez Guinovart 2012).

These similar features led us to the idea of reorienting the relations in WordNet towards strategies for exploring terminology in the lexical-semantic network. A revision of the Galnet lexical-semantic relations was then undertaken in order to examine the possibilities of building a hierarchical structure (Cabré 1992) from all the set of the WordNet synsets, acknowledging the difficulties of working with such generic synsets in terminology research, which tends to deal with more precise concepts.

Furthermore, WordNet synsets are distributed into four different grammatical categories (83,246 noun synsets, 18,156 adjective synsets, 13,885 verb synsets and 3,621 adverb synsets in WordNet 3.0) and, despite the high density of nouns, there is a significant number of synsets that cannot be considered a priori terminological concepts. A general approach would be to try to link adjectives and adverbs, whenever possible, with a terminologically relevant synset through transcategorical lexical relations; for instance, linking the concepts *surgical* and *surgically* with *surgery*. This approach shows numerous limitations, mainly because this lexical relation is sometimes missing in WordNet, and when these morphologically-based relations are codified in WordNet they occur between English-language variants rather than between concept nodes. Where this is not possible, other semantic relations such as near synonymy or antonymy would be used, despite their tendency to change the field of speciality with certain adjectives, because our strategy for exploring termi-

⁸ [<http://gl.dbpedia.org/wiki/>]; last access: November 4, 2018].

⁹ [<http://www.image-net.org/>]; last access: November 4, 2018].

¹⁰ [<http://sli.uvigo.gal/RILG/>]; last access: January 24, 2017].

¹¹ [<http://www.ihtsdo.org/snomed-ct/>]; last access: January 24, 2017].

¹² [<http://sli.uvigo.gal/termoteca/>]; last access: January 24, 2017].

nology in WordNet includes all the synsets of the lexical-semantic network regardless of their grammatical category.

To this end, it was assumed from the very beginning that it would be necessary to extend the cognitive approach to the study of the relations with an empirical verification of the results, by examining the presence of the variants in specialized text corpora from the perspective of communicative terminology.

Therefore, a methodology was designed to perform a parameterized browsing through a domain-specific lexicon given a synset representing that specific domain. Also, a method was sought to verify the empirical occurrence of the concepts in specialized Galician-language corpora. Terminology browsing from a given synset and verification in selected domain-specific corpora have been implemented in a freely accessible web application called Termonet¹³, which is integrated into Galnet's public interface (Solla Portela/Gómez Guinovart 2015b).

Termonet's features rely on two basic resources: the latest development version of Galnet and the Galician Technical Corpus¹⁴, a lemmatized, POS-tagged and semantically annotated terminology-oriented corpus of 18 million tokens containing contemporary specialized Galician texts in the fields of law, computing, economics, environmental science, sociology and medicine. The Galician Technical Corpus was tagged using Freeling¹⁵ and UKB (Agirre/Soroa 2009) and disambiguated with respect to the Galnet lexicon.

The main function of Termonet is to enable the extraction of domain-specific variants from WordNet. For this purpose, Termonet provides a query form that allows selecting a synset from the lexical-semantic network and extracting related terms according to the semantic relations defined in the configuration. Although Termonet allows extraction from any WordNet synset, due to its terminological nature, the application always tries to suggest the closest noun variants when initiating a search from a non-noun synset.

Termonet allows to indicate the source synset that will determine the domain-specific term extraction, and also to select the semantic relations that will be used to identify the domain-specific terms and the maximum distance or depth level to be reached for each relation type. Thus, Termonet will display the tree-like structure of relations from the source synset through the selected relations to the specified depth level in the tree. Figure 1 shows, for instance, the hyponymy relation found among linguistic terms, starting at the synset represented by the English variant *linguistics* according to the system default parameters. For reasons of space, the output of the extraction illustrated in Figure 1 includes only its first results, since the total number

¹³ [<http://sli.uvigo.gal/galnet/termonet.php>]; last access: January 24, 2017].

¹⁴ [<http://sli.uvigo.gal/CTG/>]; last access: January 24, 2017].

¹⁵ [<http://nlp.lsi.upc.edu/freeling/>]; last access: January 24, 2017].

of synsets identified by this query amounts to 422¹⁶. Each line of Figure 1 describes a synset in the domain of *linguistics* using the following information separated by spaces:

1. distance (or depth level) from the synset described to the source synset representing the selected area of exploration;
2. lexical-semantic relation linking the synset described with the previous synset in the extraction chain;
3. ILI that identifies the synset described;
4. variants of the synset in English;
5. variants of the synset in Galician, if any;
6. semantic class or epinonym to which the synset described is linked, with the distance from the synset to the epinonym indicated in brackets.

422 synsets. 152 deles (36.02 %) con 191 variantes en galego

```

[0] 06172789-n | eng: linguistics | glg: lingüística ((0) linguistics )
[+1] 1 has_hyponym 05999792-n | eng: science, scientific_discipline | glg: ciencia, disciplina_científica ((0) science )
[+1] 2 has_hyponym 06168552-n | eng: computational_linguistics ((1) linguistics )
[+2] 1 has_hyponym 06131503-n | eng: machine_translation, MT | glg: tradución_automática ((2) computer_science ((2) linguistics )
[+1] 3 has_hyponym 06168703-n | eng: dialect_geography, linguistic_geography ((1) linguistics )
[+1] 4 has_hyponym 06168855-n | eng: etymology | glg: etimoloxía ((1) linguistics )
[+2] 1 has_hyponym 06178574-n | eng: lexicostatistics ((2) linguistics )
[+3] 1 has_pertainym 02881246-s | eng: lexicostatistic ((3) linguistics )
[+2] 2 related_to 06514265-y | eng: etymologise, etymologize ((1) linguistics )
[+2] 3 related_to 01070458-y | eng: etymologise, etymologize ((2) animal )
[+2] 4 related_to 01020003-s | eng: etymological ((2) linguistics )
[+3] 1 pertains_to 06514621-n | eng: etymology | glg: etimoloxía ((4) intrans )
[+4] 5 related_to 10066314-n | eng: etymologist ((2) intrans )
[+1] 5 has_hyponym 06169050-n | eng: diachronic_linguistics, diachrony, historical_linguistics | glg: lingüística_diacrónica ((1) linguistics )
[+2] 1 related_to 02377418-a | eng: diachronic, historical | glg: diacrónico, histórico ((1) intrans )
[+2] 2 category_term 06744154-n | eng: derivation, deriving, etymologizing ((2) linguistics ((2) statement )
[+3] 1 related_to 00281791-y | eng: derive | glg: basearse_en_proceder ((2) alteration )
[+1] 6 has_hyponym 06178290-n | eng: neurolinguistics | glg: neurolingüística ((1) linguistics )
[+2] 1 related_to 10354053-n | eng: neurologist ((1) intrans )
[+1] 7 has_hyponym 06178492-n | eng: pragmatics | glg: pragmática ((1) linguistics )
[+1] 8 has_hyponym 06178792-n | eng: semantics | glg: semántica ((1) linguistics )
[+2] 1 has_hyponym 06178985-n | eng: dextis | glg: dexre ((2) linguistics )
[+3] 1 has_pertainym 02125402-a | eng: deictic ((2) word )
[+4] 3 related_to 06289693-n | eng: deictic, deictic_word ((4) word )
[+2] 2 has_hyponym 06180282-n | eng: formal_semantics ((2) linguistics )
[+2] 3 has_hyponym 06180412-n | eng: lexical_semantics ((2) linguistics )
[+2] 4 has_hyponym 06180549-n | eng: cognitive_semantics, conceptual_semantics, semasiology ((2) linguistics )
[+2] 5 has_pertainym 02842042-a | eng: semantic | glg: semántico ((2) linguistics )
[+3] 1 has_derived 02110646-r | eng: semantically ((3) linguistics )
[+2] 7 related_to 10877820-n | eng: semanticist, semiotician ((3) intrans )
[+1] 9 has_hyponym 06181123-n | eng: sociolinguistics ((1) linguistics )
[+2] 1 has_pertainym 02842881-a | eng: sociolinguistic ((2) linguistics )
[+3] 1 has_derived 02111018-r | eng: sociolinguistically ((3) linguistics )
[+2] 3 related_to 10620586-n | eng: sociolinguist ((1) intrans )
[+1] 10 has_hyponym 06181284-n | eng: structuralism, structural_linguistics ((1) linguistics )
[+1] 11 has_hyponym 06181448-n | eng: synchronic_linguistics ((1) linguistics )
[+1] 12 has_hyponym 06181584-n | eng: descriptive_linguistics ((1) linguistics )
[+2] 1 near_antonym 06181993-n | eng: prescriptive_linguistics ((1) linguistics )
[+2] 2 has_hyponym 06174404-n | eng: grammar | glg: gramática ((0) grammar )
[+3] 1 has_hyponym 06175829-n | eng: descriptive_grammar ((1) linguistics ((1) grammar )

```

Fig. 1: Term extraction

¹⁶ The full results of this query are available at [<http://sli.uvigo.gal/galnet/termonet.php?ili=ili-30-06172789-n>]; last access: January 24, 2017].

The application was designed to be interactive. The user can restrict the relations in WordNet in order to define his own conceptual field. After selecting the initial synset, which represents the area of the inquiry, the query form offers the different parameters of configuration for the groups of relations and for each relation in two distinct sections. First, the distance of the exploration is defined in order to limit the navigation towards conceptual nodes tied by means of ascending relations and to determine the extension of the descent and of the navigation in horizontal direction. The second section of the form allows to indicate the groups of relations or the individual relations that should be omitted from the query. It also offers the possibility of filtering the synsets with the grammatical categories selected for each relation.

This term extraction tool, albeit still in the testing phase, is already producing consistent and quantitatively significant results using rather simple search parameter settings. The extraction procedure is identical for both broad (e.g. biology) and narrow (e.g. microbiology) conceptual domains.

```

apendicite 14258512-n 5Z inflammation of the vermiform appendix
  • ili_p: 1
  • sense_p: 1
  • ili_f: 0.57
  • sense_f: 1
apnea 14389408-n 63 transient cessation of respiration
  • ili_p: 1
  • sense_p: 1
  • ili_f: 0.03
  • sense_f: 1
apnea do sono 14370267-n 22 apnea that occurs during sleep
  • ili_p: 1
  • sense_p: 1
  • ili_f: 0.22
  • sense_f: 1
aracnodactilia 14157967-n 2 an autosomal dominant disease characterized by elongated bones (especially of limbs and digits) and abnormalities of the eyes and circulatory system
  • ili_p: 1
  • sense_p: 1
  • ili_f: 0.02
  • sense_f: 1
arrefriado 14145501-n 84 a mild viral infection involving the nose and respiratory passages (but not the lungs)
  • ili_p: 1
  • sense_p: 1
  • ili_f: 0.84
  • sense_f: 1
arrefriado común 14145501-n 27 a mild viral infection involving the nose and respiratory passages (but not the lungs)
  • ili_p: 1
  • sense_p: 1
  • ili_f: 0.27
  • sense_f: 1
arrefriado común 14145759-n 27 a common cold affecting the nasal passages and resulting in congestion and sneezing and headache
  • ili_p: 1
  • sense_p: 0
  • ili_f: 0.27
  • sense_f: 0
arrefriado de cabeza 14145759-n 0 a common cold affecting the nasal passages and resulting in congestion and sneezing and headache
  • ili_p: 0
  • sense_p: 0
  • ili_f: 0
  • sense_f: 0
arterite 14258909-n 18 inflammation of an artery
  • ili_p: 1
  • sense_p: 1
  • ili_f: 0.18
  • sense_f: 1
artrite 14186541-n 22Q inflammation of a joint or joints
  • ili_p: 1
  • sense_p: 1
  • ili_f: 1
  • sense_f: 1

```

Fig. 2: Term evaluation

As previously mentioned, Termonet allows to verify the extraction results in a domain-specific textual corpus that has been lemmatized and disambiguated with WordNet senses. Termonet evaluates the occurrence of each monolexical or plurilexical term in the selected domain-specific corpus by applying four criteria scored from 0 to 1, and combines the results obtained across all the terms into a general score for each criterion. The four evaluation criteria are as follows:

1. *ili-p* index: the variant is present (1) or absent (0) as a lemma in the corpus and is semantically tagged with the corresponding synset.
2. *sense-p* index: the variant is present as a lemma in the corpus and is tagged with the most likely sense (1) or not (0) according to UKB.
3. *ili-f* index: absolute frequency of the variant in the corpus, awarding the maximum score (1) to the sense-tagged variants occurring at least 100 times, and the minimum score (0) to the variants not found in the corpus.
4. *sense-f* index: frequency of UKB assigning the maximum likelihood to the variant's synset tag, with the maximum score (1) for all the times and the minimum (0) for none.

Through a detailed analysis of the variants (Figure 2), Termonet offers the possibility of exploring their contexts of use in the selected domain-specific corpus of the Galician Technical Corpus (Figure 3), thus providing valuable terminological information about the real use of the terms.

A busca con lema <i>apnea_do_sono</i> ill 14370267-n (tipo: exacta) documentábase en 22 ocasións, en 22 frases, no corpus de medicina Medigal de 4.376.551 palabras.		
1	(1) (1)	i+i
Hai numerosas clasificacións, pero pódense agrupar en dous apartados: as disomnias, dificultades para concilia-lo sono reparador, nas que se inclúen o insomnio, a narcolepsia ou a apnea do sono, e as parasomnias, que se producen durante o sono, coma o somnambulismo, o bruxismo ou triscar dos dentes, o terror nocturno e a enuresis nocturna (miccionar durante o sono).		
Hai numerosas clasificacións, pero pódense agrupar en dous apartados: as disomnias, dificultades para concilia-lo sono reparador, nas que se inclúen o insomnio, a narcolepsia ou a apnea do sono <i>apnea_do_sono</i> NCF000 14370267-n-0.0095074, e as parasomnias, que se producen durante o sono, coma o somnambulismo, o bruxismo ou triscar dos dentes, o terror nocturno e a enuresis nocturna (miccionar durante o sono).		
2	(1) (1)	i+i
Neumoloxía: estudos e valoración previos a transplantes pulmonares, tratamento con rehabilitación pulmonar, estudos de apnea do sono e estudos de tumores.		
Neumoloxía: estudos e valoración previos a transplantes pulmonares, tratamento con rehabilitación pulmonar, estudos de apnea do sono <i>apnea_do_sono</i> NCF000 14370267-n-0.0147916 e estudos de tumores.		
3	(1) (1)	i+i
- Síndrome de apnea do sono.		
- Síndrome de apnea do sono <i>apnea_do_sono</i> NCF000 14370267-n-0.021331.		
4	(1) (1)	i+i
Cardiopatía isquémica Insuficiencia cardíaca congestiva Insuficiencia venosa de extremidades inferiores Enfermedade tromboembólica Insuficiencia respiratoria Apnea do sono Hernia de hiato Colelitiasis		
Cardiopatía isquémica Insuficiencia cardíaca congestiva Insuficiencia venosa de extremidades inferiores Enfermedade tromboembólica Insuficiencia respiratoria Apnea do sono <i>apnea_do_sono</i> NCF000 14370267-n-0.0125174 Hernia de hiato Colelitiasis		
5	(1) (1)	i+i
As técnicas ortodóncicas poden emendar problemas de saúde, como a patoloxía de mandíbula curta, que facilita a aparición do rouquén e da apnea do sono, amais de dificultar unha posible intubación orotraqueal.		
As técnicas ortodóncicas poden emendar problemas de saúde, como a patoloxía de mandíbula curta, que facilita a aparición do rouquén e da apnea do sono <i>apnea_do_sono</i> NCF000 14370267-n-0.00957028, amais de dificultar unha posible intubación orotraqueal.		
6	(1) (1)	i+i
Tamén se sabe que o neno que ronca pode padecer episodios de apnea nocturna (síndrome de apnea do sono), motivado en numerosas ocasións por unha obstrución das vías respiratorias altas, polo tamaño das amígdalas e as glándulas adenoideas.		
Tamén se sabe que o neno que ronca pode padecer episodios de apnea nocturna (síndrome de apnea do sono <i>apnea_do_sono</i> NCF000 14370267-n-0.0097992), motivado en numerosas ocasións por unha obstrución das vías respiratorias altas, polo tamaño das amígdalas e as glándulas adenoideas.		
7	(1) (1)	i+i
Novos riscos para a apnea do sono: Os ronquidos, algo máis ca ruidos molestos Para que se catalogue como enfermidade, unha persoa debe sufrir entre 5 e 10 interrupcións respiratorias (apneas/hipoapneas) por hora durante o sono, aínda que pode acadar, en casos extremos, ata 90 durante o mesmo período		
Novos riscos para a apnea do sono <i>apnea_do_sono</i> NCF000 14370267-n-0.00791801 Os ronquidos, algo máis ca ruidos molestos Para que se catalogue como enfermidade, unha persoa debe sufrir entre 5 e 10 interrupcións respiratorias (apneas / hipoapneas) por hora durante o sono, aínda que pode acadar, en casos extremos, ata 90 durante o mesmo período		

Fig. 3: Terms in context

5 Applications in ontology learning and management

5.1 Applications in WordNet semantic areas

A new semantic categorization of WordNet was devised to exploit the terminological implications of the relations between synsets. The adopted approach was based on tracing a path in the opposite direction to that used by Termonet to explore a domain from a synset, so each synset finds its way through the relations to an *epinonym* noun synset representing the semantic domain where it is to be included. Thus, we define an *epinonym* as a noun synset representing the category of the semantic domain to which other synsets will be automatically assigned by algorithms that will evaluate their proximity from a terminological point of view through the cognitive processing of the lexical-semantic relations.

There is a certain parallelism between the method used for the selection of the semantic classes designed for Basic Level Concepts (BLC) (Izquierdo/Suárez/Rigau 2015) and for epinonyms. In both cases, there is no preset semantic categorization, and synsets are selected by the fact of fulfilling certain criteria of selection. Furthermore, in both cases, one of the selection criteria is the computation of lexical-semantic relations. However, BLC and epinonyms have a substantially different objective. BLC are nominal and verbal synsets selected for being semantically representatives, and for being neither too concrete nor too abstract. For its part, epinonyms are nominal synsets selected by its aptitude to represent a semantic area, preferably a terminologically relevant area of specialty. The selection of epinonyms is limited to the grammatical category of nouns because only nominal synsets are considered appropriate to designate areas of specialty. In addition, its coverage tries to be representative of all semantic areas, examining the entire depth of the hyponymy tree of nouns.

This exploration of the terminological implications of the relations between synsets is carried out in two different phases:

1. Automatic selection of the epinonym noun synsets to become the categories representing the different semantic domains and grouping the synsets assigned to each of the domains.
2. Development of a set of algorithms to calculate the path from each WordNet synset through the lexical-semantic relations to the closest epinymic category or categories.

The algorithm used for selecting the set of epinonyms explores the noun synsets downward through the hyponymy relation; that is, it departs from the synset that corresponds to the concept *entity*, which is at the top (or level 0) of the hypernymy relations among WordNet noun synsets, and travels through the hyponymy rela-

tions assigning higher values of hyponymy as the distance to the top increases. The selection of the epinonyms is based on a score assigned to each synset according to the number of its semantic relations with other synsets, and to its hyponymy level. This approach yielded a hierarchical tree of 927 epinonyms, which represent 1.11% of the noun synsets in WordNet (83,246) and 0.78% of all WordNet synsets (118,868). By starting at the top level of hypernymy, the algorithm scheme preserves the inheritance in the WordNet hyponymy tree so that the relations between the selected epinonyms may be rebuilt into a hierarchical tree of subcategories, the structure and extension of which may be viewed online via Galnet's public interface¹⁷.

The selected set of epinonyms provides an overview of WordNet semantic areas. It is consistent with its own internal structure and is generated automatically by exploring the relations from a terminological perspective; therefore, there is no modeling of the lexical-semantic network to conform to preconceived categorial notions.

After a representative set of noun synsets was obtained, a methodology was developed to link each WordNet synset to the closest epinonym(s) according to terminological criteria based on the characteristics of the lexical-semantic relations. The algorithm used to assign epinonyms generated 128,986 pairings, which is slightly above the total number of WordNet synsets (108.51%), given that some synsets were linked with more than one epinonym. Furthermore, 99.80% of the WordNet synsets were able to automatically assign themselves to their corresponding semantic area(s) and only 239 synsets (225 adverbial, 8 verbal and 6 adjectival) were not paired with an epinonym, as shown in Table 2.

Tab. 2: Pairings between synsets and epinonyms

	Paired synsets	WordNet synsets	Percent
Nouns	83,246	83,246	100%
Verbs	18,150	18,156	99.97%
Adjectives	13,837	13,845	99.94%
Adverbs	3,396	3,621	93.79%
Total	118,629	118,868	99.80%

Owing to the distinctive ontological-relational nature of the results, alongside the ontologies already present in Galnet's interface, additional search methods were

¹⁷ [<http://sli.uvigo.gal/galnet/hierarchy.php?version=dev&ontology=epinonyms&category=entity##ili-30-00001740-n>]; last access: January 24, 2017].

included to enable exploring the design of WordNet semantic areas represented by the epinonyms. In addition, similarly to what is described in Section 4 for Termonet, a web application¹⁸ was built to enable the verification of Galician variants in specialized corpora by selecting all the variants of a category from any of the ontologies.

5.2 Applications in the Semantic Web

We also provide all contents of Galnet as RDF (Resource Description Framework) resources¹⁹ through a SPARQL endpoint²⁰ with free public access for users to explore our data using SPARQL queries. Turtle files with data corresponding to the latest public release of MCR and related ontologies can also be downloaded from our site²¹.

The RDF Galnet monolingual dictionaries conform to the Lemon (Lexicon Model for Ontologies) model²². The Galnet synsets are aligned with Princeton's WordNet synsets version 3.1²³, with Princeton's WordNet synsets version 3.0 in lemonUby²⁴ and with the Interlingual Index (ILI). In many cases, Princeton's WordNet also provides the alignment with a corresponding synset in lemonUby version 3.0. However, the alignment in RDF Galnet offers correspondences between all MCR synsets from version 3.0 and version 3.1 or lemonUby ones.

The RDF Galnet internal ontology is based on the ontology that uses the RDF Princeton Wordnet 3.1, revised and adapted to the EuroWordNet framework followed by the MCR project. Moreover, all the ontologies linked to the Galnet synsets were converted to the RDF data model: Adimen-SUMO, Top Ontology, WordNet Domains and Epinonyms.

Query 1: SPARQL federated query for Galician variant trabe

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
SELECT * WHERE {
  SERVICE <http://sli.uvigo.gal/sparql> {
```

¹⁸ [<http://sli.uvigo.gal/galnet/category.php>]; last access: January 24, 2017].

¹⁹ [<https://www.w3.org/RDF/>]; last access: January 24, 2017].

²⁰ [<http://sli.uvigo.gal/sparql/>]; last access: January 24, 2017].

²¹ [http://sli.uvigo.gal/lod/rdf_galnet_mcr30-2016.7z]; last access: January 24, 2017].

²² [<http://lemon-model.net>]; last access: January 24, 2017].

²³ [<https://wordnet.princeton.edu>]; last access: January 24, 2017].

²⁴ [<http://lemon-model.net/lexica/uby/>]; last access: January 24, 2017].

```

GRAPH <http://sli.uvigo.gal/rdf_galnet_glg> {
  ?offset rdfs:label "trabe"@glg .
  ?offset owl:sameAs ?pwn .
  FILTER (REGEX(str (?pwn), "^http://wordnet-rdf.princeton.edu/wn31/"))
}
}
SERVICE <http://babelnet.org/sparql/> {
  GRAPH <http://babelnet.org/rdf/> {
    ?babelnet <http://babelnet.org/model/babelnet#wiktionaryPageLink>
?wiktionary .
    ?babelnet skos:exactMatch ?pwn .
    ?babelnet skos:exactMatch ?eng_resource .
  }
}
SERVICE <http://gl.dbpedia.org/sparql> {
  GRAPH <http://gl.dbpedia.org> {
    ?glg_resource owl:sameAs ?eng_resource .
    FILTER (REGEX(str(?glg_resource), "^http://gl.dbpedia.org/resource/"))
) .
    FILTER (REGEX(str(?eng_resource), "^http://dbpedia.org/resource/")) )
.
}
}
}

```

The availability of the data from Galnet in the semantic web allows to link the resource with other Linked Open Data and to extract a wider knowledge through SPARQL queries. For instance, Query 1 demonstrates the power of SPARQL federated queries to link a Galician variant from Galnet with the corresponding resources in Princeton's WordNet, BabelNet, English Wiktionary and Dbpedia.

On the other hand, Query 2 starts from the English variant *penicillin* and extracts its ILI from the MCR, its equivalent variant in Galician, the corresponding resource in the Galician Dbpedia, the IUPAC (International Union of Pure and Applied Chemistry) designation in Galician of the chemical compound, the equivalent resource in the YAGO (Yet Another Great Ontology) knowledge base (Mahdisoltani/Biega/Suchanek 2015), and the 57 lexical forms in different languages listed in YAGO to designate the concept.

Query 2: SPARQL federated query for English variant penicillin

```

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>

```

```

PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT * WHERE {
  SERVICE <http://sli.uvigo.gal/sparql> {
    GRAPH <http://sli.uvigo.gal/rdf_galnet_eng> {
      ?eng_offset rdfs:label "penicillin"@eng .
    }
    GRAPH <http://sli.uvigo.gal/rdf_galnet> {
      ?ili owl:sameAs ?eng_offset .
      ?ilg owl:sameAs ?glg_offset .
      FILTER (REGEX(STR(?glg_offset),
"^\http://sli.uvigo.gal/rdf_galnet_glg")) .
    }
    GRAPH <http://sli.uvigo.gal/rdf_galnet_glg> {
      ?glg_offset rdfs:label ?glg_label .
    }
  }
  BIND (IRI(CONCAT("http://gl.dbpedia.org/resource/",
UCASE(SUBSTR(STR(?glg_label), 1, 1)), SUBSTR(STR(?glg_label), 2))) AS
?glg_resource) .
  SERVICE <http://gl.dbpedia.org/sparql> {
    GRAPH <http://gl.dbpedia.org> {
      ?glg_resource dbo:iupacName ?glg_iupac .
    }
  }
  BIND(CONCAT("1", SUBSTR(STRAFTER(STR(?eng_offset),
"http://sli.uvigo.gal/rdf_galnet_eng/"), 1, 8) ) AS ?yago_id) .
  SERVICE <http://linkeddata1.calcul.u-psud.fr/sparql> {
    GRAPH <http://www.yago-knowledge.org> {
      ?yago_resource <http://yago-knowledge.org/resource/hasSynsetId> ?ya-
go_id .
      ?yago_resource rdfs:label ?lang_label .
    }
  }
}

```

Finally, it is interesting to note that the joint exploitation of Linked Open Data can be an efficient, robust and reliable way to develop new wordnets or to expand the lexical coverage of existing ones (Solla Portela/Gómez Guinovart 2016).

6 Conclusion

Natural language processing plays an increasingly important role in the information society. The need for a computational representation of general and specialized lexical information capable of being exploited by language technologies (in automatic and computer-assisted translation applications, multilingual retrieval of information, semantic web, etc.) is a new challenge for the research in lexicology. The use of ontologies, understood as formal and shared specifications of the conceptualization of a domain that can be transmitted between people and/or systems, offers a suitable solution for this task. This is particularly relevant in a highly multilingual context such as the current information society, where the symbiosis between lexicon and ontologies allows us to work with powerful conceptual and methodological tools towards the best representation of general and specialized multilingual knowledge.

7 Bibliography

- Agirre, Eneko/Edmonds, Peter (2006): *Word Sense Disambiguation*. Berlin: Springer.
- Agirre, Eneko/Alegria, Iñaki/Rigau, German/Vossen, Piek (2007): MCR for CLIR. In: *Procesamiento del Lenguaje Natural* 38, 3–15.
- Agirre, Eneko/Soroa, Aitor (2009): Personalizing PageRank for Word Sense Disambiguation. In: Schlangen, David/Oflazer, Kemal (eds.): *Proceedings of the 12th Conference of the European Chapter of the ACL*. Athens: ACL, 33–41.
- Álvarez de la Granja, María/Gómez Clemente, Xosé María/Gómez Guinovart, Xavier (2016): Introducing Idioms in the Galician WordNet: Methods, Problems and Results. In: *Open Linguistics* 2 (1), 253–286.
- Álvez, Javier/Atserias, Jordi/Carrera, Jordi/Climent, Salvador/Oliver, Antoni/Rigau, German (2008): Consistent Annotation of EuroWordNet with the Top Concept Ontology. In: Tanács, Attila/Csendes, Dóra/Vincze, Veronika/Fellbaum, Christiane/Vossen, Piek (eds.): *Proceedings of the 4th Global WordNet Conference*. Szeged: Global WordNet Association, n.p.
- Bentivogli, Luisa/Forner, Pamela/Magnini, Bernardo/Pianta, Emanuele (2004): Revising WordNet Domains Hierarchy: Semantics, Coverage, and Balancing. In: Sérasset, Gilles/Armstrong, Susan/Boitet, Christian/Popescu-Belis, Andrei/Tufis, Dan (eds.): *Proceedings of COLING Workshop on Multilingual Linguistic Resources*. Geneva: ACL, 101–108.
- Bhingardive, Sudha/Ajotikar, Tanuja/Kulkarni, Irawati/Kulkarni, Malhar/Bhattacharyya, Pushpak (2014): Semi-Automatic Extension of Sanskrit Wordnet using Bilingual Dictionary. In: Orav, Heili/Fellbaum, Christiane/Vossen, Piek (eds.): *Proceedings of the Seventh Global WordNet Conference*. Tartu: Global WordNet Association, 324–329.
- Cabré, Maria Teresa (1992): *La terminologia. La teoria, els mètodes, les aplicacions*. Barcelona: Empúries.
- Cai, Qingqing/Gung, James/Guan, Maochen/Kurlandski, Gerald/Pease, Adam (2016): Word Substitution in Short Answer Extraction: A WordNet-based Approach. In: Barbu Mititelu, Verginica/Forăscu, Corina/Fellbaum, Christiane/Vossen, Piek (eds.): *Proceedings of the Eighth Global WordNet Conference*. Bucarest: University of Iași, 66–73.

- Elberrichi, Zakaria/Rahmoun, Abdelattif/Bentaalah, Mohamed Amine (2008): Using WordNet for Text Categorization. In: *The International Arab Journal of Information Technology* 5 (1), 16–24.
- Fang, Hui (2008): A Re-examination of Query Expansion Using Lexical Resources. In: Arisoy, Ebru/Maier, Wolfgang/Inoue, Keisuke (eds.): *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*. Columbus: Association for Computational Linguistics, 139–147.
- Fellbaum, Christiane (ed.) (1998): *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press.
- Fernández Montraveta, Ana/Vázquez, Gloria (2010): La construcción del WordNet 3.0 en español. In: Castillo, María Auxiliadora/García Platero, Juan Manuel (eds.): *La lexicografía en su dimensión teórica*. Málaga: Universidad de Málaga, 201–220.
- Ferrández, Sergio/Ferrández, Antonio/Roger, Sandra/López–Moreno, Pilar (2007): Búsqueda de respuestas bilingüe basada en ILI, el sistema BRILI. In: *Procesamiento del Lenguaje Natural* 38, 27–33.
- Gómez Clemente, Xosé María/Gómez Guinovart, Xavier/González Pereira, Andrea/Taboada Lorenzo, Verónica (2013): Sinonimia e rexistros na construción do WordNet do galego. In: *Estudos de Lingüística Galega* 5, 27–42.
- Gómez Guinovart, Xavier (2012): A Hybrid Corpus-Based Approach to Bilingual Terminology Extraction. In: Moskowich-Spiegel, Isabel/Crespo, Begoña (eds.): *Encoding the Past, Decoding The Future: Corpora in the 21st Century*. Newcastle upon Tyne: Cambridge Scholar Publishing, 147–175.
- González Agirre, Aitor/Rigau, German (2013): Construcción de una base de conocimiento léxico multilingüe de amplia cobertura: Multilingual Central Repository. In: *Linguamática* 5 (1), 13–28.
- Hu, Xiangen/Graesser, Arthur C. (1998): Using WordNet and latent semantic analysis to evaluate the conversational contributions of learners in the tutorial dialog. In: *Proceedings of the International Conference on Computers in Education*, vol. 2. Beijing: Springer, 337–341.
- Huang, Bill (2016): WNSpell: a WordNet-Based Spell Corrector. In: Barbu Mititelu, Verginica/Forăscu, Corina/Fellbaum, Christiane/Vossen, Piek (eds.): *Proceedings of the Eighth Global WordNet Conference*. Bucarest: University of Iași, 135–142.
- Ishahara, Hitoshi/Bond, Francis/Uchimoto, Kiyotaka/Utiyama, Masao/Kanzaki, Kyoko (2008): Development of the Japanese WordNet. In: Calzolari, Nicoletta/Choukri, Khalid/Maegaard, Bente/Mariani, Joseph/Odjik, Jan/Piperidis, Stelios/Tapias, Daniel (eds.): *Proceedings of the Sixth International Language Resources and Evaluation*. Marrakech: ELRA, n.p.
- Izquierdo, Rubén/Suárez, Armando/Rigau, German (2015): Word vs. Class-Based Word Sense Disambiguation. In: *Journal of Artificial Intelligence Research* 54, 83–122.
- Knoop, Susanne/Wilske, Sabrina (2013): WordGap - Automatic generation of gap-filling vocabulary exercises for mobile learning. In: Volodina, Elena/Borin, Lars/Loftsson, Hrafn (eds.): *Proceedings of the Second Workshop on NLP for Computer-Assisted Language Learning*. Linköping: Linköpings Universitet, 39–47.
- Mahdisoltani, Farzaneh/Biega, Joanna/Suchanek, Fabian M. (2015): YAGO3: A Knowledge Base from Multilingual Wikipedias. In: *Proceedings of the 7th Biennial Conference on Innovative Data Systems Research*. Asilomar: CIDR, n.p.
- Miller, George A. et al. (1990): Wordnet: An on-line lexical database. In: *International Journal of Lexicography* 3 (4), 235–244.
- Oliver, Antoni/Climent, Salvador (2011): Construcción de los WordNets 3.0 para castellano y catalán mediante traducción automática de corpus anotados semánticamente. In: *Procesamiento del Lenguaje Natural* 47, 293–300.
- Ordan, Noam/Shuly Wintner (2007): Hebrew WordNet: a Test Case of Aligning Lexical Databases Across Languages. In: *International Journal of Translation* 19 (1), 39–58.

- Pease, Adam/Niles, Ian/Li, John (2002): The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. In: Pease, Adam (ed.): *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*. Edmonton: AAAI, s.p.
- Plaza, Laura/Díaz, Alberto/Gervás, Pablo (2010): Automatic summarization of news using WordNet concept graphs. In: *IADIS International Journal on Computer Science and Information Systems* 5 (1), 45–57.
- Pociello, Elisabete/Agirre, Eneko/Aldezabal, Izaskun (2011): Methodology and Construction of the Basque WordNet. In: *Language Resources and Evaluation* 45 (2), 121–142.
- Simões, Alberto/Gómez Guinovart, Xavier (2014): Bootstrapping a Portuguese WordNet from Galician, Spanish and English wordnets. In: Navarro Mesa, Juan Luis *et al.* (eds.): *Advances in Speech and Language Technologies for Iberian Languages*. Berlin: Springer, 239–248.
- Solla Portela, Miguel Anxo/Gómez Guinovart, Xavier (2015a): Galnet: o WordNet do galego. Aplicacións lexicolóxicas e terminolóxicas. In: *Revista Galega de Filoloxía* 16, 169–201.
- Solla Portela, Miguel Anxo/Gómez Guinovart, Xavier (2015b): Termonet: Construcción de terminoloxías a partir de WordNet y corpus especializados. In: *Procesamiento del Lenguaje Natural* 55, 165–168.
- Solla Portela, Miguel Anxo/Gómez Guinovart, Xavier (2016): DBpedia del gallego: recursos y aplicaciones en procesamiento del lenguaje. In: *Procesamiento del Lenguaje Natural* 57, 139–142.
- Sun, Koun-Tem/Huang, Yueh-Min/Liu, Ming-Chi (2011): A WordNet-Based Near-Synonyms and Similar-Looking Word Learning System. In: *Educational Technology & Society* 14 (1), 121–134.
- Vintar, Špela/Fišer, Darja/Vrščaj, Aljoša (2012): Were the clocks striking or surprising?: using WSD to improve MT performance. In: Proceedings of the Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra) (EACL 2012). Stroudsburg: ACL, 87–92.
- Vossen, Piek (2002): WordNet, EuroWordNet and Global WordNet. In: *Revue française de linguistique appliquée* 7, 27–38.
- Zhao, Feng/Fang, Fei/Yan, Fengwei/Jin, Hai/Zhang, Qin (2012): Expanding approach to information retrieval using semantic similarity analysis based on WordNet and Wikipedia. In: *International Journal of Software Engineering and Knowledge Engineering* 22 (2), 305–322.