Recognition and extraction of definitional contexts in Spanish for sketching a lexical network

César Aguilar
Department of Linguistics
Autonomous University of Queretaro
Cerro de las Campanas, s/n, Queretaro, Mexico
CAguilar@iingen.unam.mx

Olga Acosta
Postgraduate School of Computer Science
UNAM
Ciudad Universitaria, Mexico City
OAcostaL@iingen.unam.mx

Gerardo Sierra
Language Engineering Group
Engineering Institute, UNAM
Ciudad Universitaria, Mexico City
GSierraM@iingen.unam.mx

Abstract

In this paper we propose a method to exploit analytical definitions extracted from Spanish corpora, in order to build a lexical network based on the hyponymy/hyperonymy, part/whole and attribution relations. Our method considers the following steps: (a) the recognition and extraction of definitional contexts from specialized documents, (b) the identification of analytical definitions on these definitional contexts, using verbal predications, (c) the syntactic and probabilistic analysis of the association observed between verbal predication and analytical definitions, (d) the identification of the hyponymy/hyperonymy, part/whole and attribution relations based on the lexical information that lies between predications and definitions and other types of phrases, in particular prepositional phrases mapped by the preposition *de* (Eng. of/from).

1 Introduction

Nowadays, the possibility of searching and recognizing lexical relations in definitions occurring in specialized text corpora is an important task in computational lexicography and terminology.

In this sense, authors such as Vossen and Copestake (1993), as well as Wilks, Slator & Guthrie (1995) are pioneers in offering a relevant set of experiments and techniques about how to identify hyponymy/hypernymy relations from analytical definitions, taking into account the underlying association that exists between terms and genus terms.

Complementary to these first attempts for identifying such lexical relations, Riloff & Shepherd (2004) argue that while these efforts have been oriented to extract lexical relations from corpus of general language, it is necessary to focus on domain-specific corpora, in order to obtain a specialized knowledge that is required for in-depth understanding of the subject matter.

In line with the argument formuled by Riloff & Shepherd, Buiteelar, Cimiano & Magnini (2005) have proposed several methods for building ontologies from text corpora, priorizing the automatic recognition of syntactic patterns that codify hyponymy/hyperonymy relations.

Following all these authors, we sketch here a research project to design a lexical network, focused on classifying scientific and technical concepts extracted from Spanish text corpora. In particular, we obtain these concepts by extracting definitional contexts (DCs) with terms and definitions clearly formulated, according to the theoretical framework developed by Sierra, Alarcon & Aguilar (2006).

After extracted these DCs, we propose a method to identify lexical relations between terms inserted into the DCs. The method considers, on the one hand, a grammatical analysis for detecting syntactic patterns that represent term and genus term, bearing in mind their association through lexical relations such as hyponymy/hyperonymy, part/whole or attribution relations. On the other hand, we proposed a semi-automatic evaluation to determine the degree of accuracy respect to the results obtained by our method.

The issues that we will deal in this paper are organized as follows: (a) as a starting point, we expose briefly the theoretical framework to extract DCs from Spanish corpora. (b) According to this framework, we describe how analytical definitions

Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas, pages 109–116, Los Angeles, California, June 2010.

linked to terms can be identified, considering the identification of verbal predications that function as connectors between such definitions and terms. (c) Thus, we offer a probabilistic evaluation for determining the degree of association between predications and analytical definitions. (d) After this evaluation, we sketch a method for exploiting this association between predications and definitions, in order to identify lexical relations, specifically hyponymy/hyperonymy, part/whole and attribution relations.

2 Theoretical framework: DC extraction

We situate our analysis within the framework of definitional contexts (or DCs) extraction. According to Sierra *et al*. (2008), a DC is a discursive structure that contains relevant information to define a term. DCs have at least two constituents: a term and a definition, and usually linguistic or metalinguistic forms, such as verbal phrases, typographical markers and/or pragmatic patterns. An example is:

(1) *La* ***cuchilla fusible*** [Term] *se define como* [Verbal Phrase] *un elemento de conexión y desconexión de circuitos eléctricos* [Definition]. (Engl. The fuse-switch disconnector is defined as an element of connection and disconnection of electric circuits).

In (1), the term *cuchilla flexible* is emphasized by the use of bold font, and it appears linked with the verbal predication *se define como*, and the definition *un elemento de conexión y desconexión de circuitos eléctricos.* Following to Sierra *et al.* (2008), we consider the term, the verbal phrase and the definition as the three main units constituting the syntactic structure of a DC.

This kind of syntactic structure introduces an analytical definition (in the Aristotle's sense), where the genus term is represented by a noun phrase (NP) *un elemento* and the differentia is represented by a prepositional phrase (PP) *de conexión y desconexión de circuitos eléctricos*.

In a detailed analysis on these syntactic structures, Aguilar (2009) explains that these structures are predicate phrases (PrP), according to the description proposed by Bowers (1993, 2001). A PrP is a phrase mapped by a functional head, and its grammatical behavior is similar to other functional phrases such as Inflexional Phrase (IP) or Complement Phrase (CP). A graphical tree representation of a PrP is:

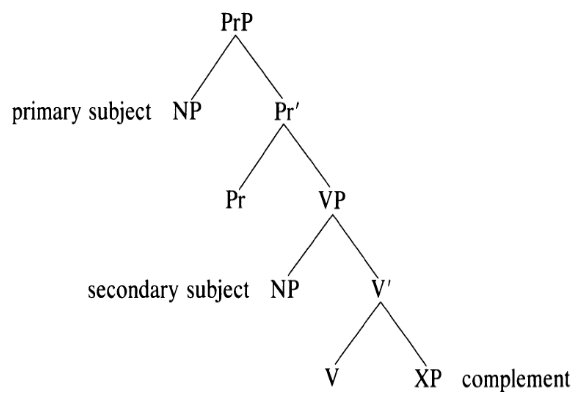

Figure 1: Tree representation for PrP, according to Bowers (1993: 596)

The Figure 1 describes the syntactic configuration of a PrP. We recognise a functional head with the feature $^{+/-}$ predicative (Pr). This head maps two subjects: a primary subject in the Specifier position of PrP (represented for a NP); and a secondary subject, in the Specifier position of verbal phrase or VP (often a NP). Finally, both subjects, the VP and the PrP are linked to one or several complements, which assume phrasal representations (e.g.: NP, IP, CP, and other types of phrases).

Based on this description about PrP, Sierra *et al.* (2008) and Aguilar (2009) observed that both primary and secondary predications have a close relation with analytical definitions expressed in specialized texts. Examples of this relation between PrP and analytical definitions are:

(2) [*Una computadora* [*es* [*un tipo de máquina electrónica que sirve para hacer operaciones* $_{PrP}$] $_{VP}$] $_{IP}$] (Eng. [A computer [is [a kind of electronic machine used to make operations $_{PrP}$] $_{VP}$] $_{IP}$]).

(3) [*Turing* [define una computadora [como un mecanismo electrónico que procesa conjuntos de datos $_{PrP}$] $_{VP}$] $_{IP}$] (Eng. [Turing [defines a computer [as a kind of electronic device that processes a set of data $_{PrP}$] $_{VP}$] $_{IP}$]).

We observe in (2) a canonical primary predication where the subject *una computadora* represents a term directly associated to predicate *es un tipo de máquina que*... This predicate introduces an analytical definition, conformed by a genus term *eletronic machine*, and the differentia *que sirve para hacer operaciones*. In (3), the predicate *como un mecanismo electrónico*... (Engl. *as a kind of electronic device*...) affects the secondary subject *una computadora* (Engl. *a computer*), in concordance with the explanation of Bowers (1993). Our analysis considers both types of predications as regular patterns that codify syntactically sequences of terms, verbal predications and definitions.

3 Searching analytical definitions in text corpora

We have adapted the predicative patterns deduced from our syntactic analysis, in order to search and find (semi-)automatically analytical definitions linked to these patterns. So, we conducted an experiment of identification of these definitions in two text corpora:

- Linguistic Corpus on Engineering (or CLI). The *CLI*, prepared by Medina and others (2004), is a collection of technical documents in different thematic areas of engineering, with an extension of 500,000 words, approximately.
- Corpus on Informatics for Spanish (or CIE). This corpus was built under the supervision of L'Homme and Drouin (2006). The *CIE* compiles several documents related to computer science and informatics. For our experiment we took a portion of *CIE*, which contains articles extracted from Wikipedia. This portion has an extension of 500,000 words.

Following to Aguilar *et al.* (2004) and *Sierra et al.* (2008), we selected a set of verbs that function as heads of predicative patterns in Spanish, taking into account the distinction between primary and secondary predications.

In the case of primary predication, the analytical definition is integrated in a sequence Term + Verbal Predication + Definition. This definition does not refer to possible author(s) of a definition. An example is:

(4) [El apartarrayos $_{\text{Term}}$] [es $_{\text{Verbal Predication}}$] [un dispositivo $_{\text{Genus Term}}$] [que protege las instalaciones contra sobretensiones de origen atmosférico $_{\text{Differentia}}$] (Engl. [The lightning conductor $_{\text{Term}}$] [is $_{\text{Verbal Predication}}$] [a device $_{\text{Genus Term}}$] [that protects electrical systems against surges of atmospheric origin $_{\text{Differentia}}$]).

Having in mind this sequence, we propose a grammatical description model for this relation:

Table 1: Construction patterns derived from the relation between primary predication and analytical definition

Definition	Genus Term	Differentia
Analytical (Primary Predication)	NP = Noun + {AdjP/PP}*	CP = Relative Pronoun + IP
		PP = Preposition + NP
		AdjP = Adjective + NP

The verbs that operate as head of these predications are: *referir* (to refer to), *representar* (to represent), *ser* (to be) and *significar* (to signify/to mean). In contrast, when a secondary predication introduces an analytical definition, this predication follows the sequence Author + Term + Verbal Predication + Definition, where the Author is equivalent to the primary subject, the Term assumes the position of secondary subject, and the definition is introduced after the Verbal Predication. In this case, the adverbial particle *como* (Eng. *as/like*), or the preposition *por* (Eng. *for/by*) indicates the place of the definition. An example is:

(5) [Carlos Godino $_{\text{Author}}$] [define $_{\text{Verbal Predication}}$] [la arquitectura naval $_{\text{Term}}$] [como la ciencia que se enfoca en la construcción de los buques $_{\text{Definition}}$] (Eng. [Carlos Godino $_{\text{Author}}$] [defines $_{\text{Verbal Predication}}$] [the naval architecture $_{\text{Term}}$] [as the science that focuses on the construction of ships $_{\text{Definition}}$])

Thus, the formal description of this sequence is:

Table 2: Construction patterns derived from the relation between secondary predication and analytical definition

Definition	Adverb/ Preposition	Genus Term	Differentia
Analytical (Secondary Predication)	*Como* *Por*	NP = Noun + {AdjP /PP}*	CP = Relative Pronoun + IP
			PP = Preposition + NP
			AdjP = Adjective + NP

The verbs linked to secondary predications are: *caracterizar + como/por* (Engl. to characterize + as/for), *comprender + como* (Engl. to comprehend

+ as), *concebir* + *como* (Engl. to conceive + as), *conocer* + *como* (Engl. to know + as), *considerar* + *como* (Engl. to consider + as), *definir* + *como* (Engl. to define + as), *describir* + *como* (Engl. to define + as), *entender* + *como* (Engl. to understand + as) , *identificar* + *como* (Engl. to identify + as) and *visualizar* + *como* (Engl. to visualize + as).

In order to recognize these sequences of predications and analytical definitions, we employed a system developed in Python by Rodríguez (2004). Broadly speaking, the input for this system is a set of previously delimited text fragments. The output is a XML table with a list of patterns, the verb used for searching these patterns, and the frequency of use in both corpora.

4 Results

Once we accomplished the process of searching and extracting of fragments with sequences of predication patterns of analytical definitions, we determined values of precision and recall for the CLI and CIE corpora based on the real number of analytical DCs in the corpus. This data was determined by a human expert through an exploration in the corpora mentioned above. In table 3 we showed DC candidates, as well as the real number of true DCs extracted from these candidates.

Thus, from CLI corpora we obtained a total of 1686 candidates. From these candidates, the human expert recognized a set of 111 true DCs to analytical definition linked to primary predication patterns. Our recall was 100% because we obtained all of the DCs with analytical definitions, but the precision achieved was very low (6.6%).

The main cause about this low precision is due to the verb *ser* (Eng. to be). The verb *ser* is highly productive in Spanish, however, much of the fragments found are not analytical definitions. In contrast, from secondary predication patterns, our recall was 100% and precision 9.4%. Thus, the CIE corpora showed measures of precision and recall higher than those of CLI corpora because most of documents were extracted from resources as Wikipedia. We suppose this factor is related with a definition scheme more canonical in scientific and technical documents.

Table 3: Sequence frequencies of predication patterns and analytical definitions

Analytical Definitions		**CLI**	**CIE**
Primary Predication	Candidates	1686	494
	DCs	111	127
	Recall	100%	100%
	Precision	6.6%	25.7%
Secondary Predication	Candidates	701	61
	DCs	66	11
	Recall	100%	100%
	Precision	9.4%	18.0%

We derived a frequency distribution of the verbs with type of predication for CLI and CIE corpora. The table 4 shows the relative frequency of use of each verb explored. Most of these verbs do not have been considered in automatic extraction tasks of hyponymy-hyperonymy relations, e. g.: Hearst (1992) or Wilks, Slator & Guthrie (1995).

Table 4: Frequency distribution of verbal predicate, and its use in analytical definitions

Predication	**Corpora**	
	CLI	**CIE**
Primary		
Referir(a)/To refer	0	0.02
Representar/To represent	0	0.04
Significar/To signify	0	0.03
Ser/To be	1	0.91
Secondary		
Caracterizar/To characterize	0.12	0.18
Concebir/To concibe	0.09	0
Conocer/To know	0.17	0
Considerar/To consider	0.21	0.27
Definir/To define	0.27	0.27
Describir/To describe	0.03	0.09
Entender/To understand	0.06	0.18
Identificar/To identify	0.03	0
Visualizar/To visualize	0.02	0

Once established this distribution, we have analyzed the degree of assurance to find a good candidate for analytical definitions. We have applied a method of conditional probabilities for primary and secondary predications. Our conditional probabilities are formulated by the hypothesis that the probability (*P*) of co-occurrence of predications (*Pred*) linked to analytical definition (*AD*) is high. Thus, we apply the following formula of conditional probability:

$$P(AD|Pred) = \frac{P(AD \cap Pred)}{P(Pred)}$$

Taking into account the formula mentioned above, we obtained the following results:

Table 5: Conditional probabilities of co-occurrence between predications and analytical definitions

Predication		CLI	CIE
Primary	Analytical definitions	93%	100%
	Not-analytical definitions	7%	0%
Secondary	Analytical definitions	95%	100%
	Not-analytical definitions	5%	0%

Therefore, we considered that the possibility to identify a good candidate of analytical definition is high, insofar as we took into account their relationship with primary and secondary predications.

In addition, Alarcón, Bach & Sierra (2007), propose a methodology for filtering true DCs from a set of candidates to DCs. An important advance provided for this work is the application of a filter phase that discards those syntactic patterns without true analytical definitions. For example, if we find a particle as *no* (Eng. not) or *tampoco* (Eng. either) in the first position before or after of a predication, there is a high probability these pattern do not introduce a good analytical definition. In Table 5 we showed some results in terms of precision and recall reported by authors only for analytical definition patterns.

Table 6: Precision & recall values

Verbal pattern	Precision	Recall
Concebir(como)/To conceive(as)	0.67	0.98
Definir(como)/To define(as)	0.84	0.99
Entender(como)/To understand(as)	0.34	0.94
Identificar(como)/To identify(as)	0.31	0.90
Significar/To signify	0.29	0.98

5 Sketching a method

In this section, we propose a method for recognizing lexical relations from the previous extraction of DCs. In particular, we assume that a good way to reach these relations is to improve the syntactic association observed between predications and analytical definitions inserted into these DCs.

This assumption is in line with the methodology proposed by Buitelaar, Cimiano & Magnini (2005) for building ontologies based on textual information obtained from corpora. These three authors conceive a chain of processes and sub-processes, represented with a layer cake scheme:

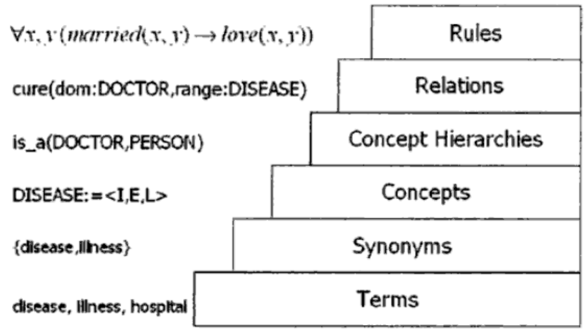

Figure 2: Ontology learning layer cake (according to Buitelaar, Cimiano & Magnani 2005)

Briefly, in this scheme Buitelaar, Cimiano & Magnini establish a sequence of 6 basic tasks for developing a possible ontology. Thus, the first task is the identification of a set of specific terms to a certain knowledge domain (in this case, a medical domain). After that, it is necessary to identify synonyms related to these terms (e.g., disease/illness). Given both sets of terms and synonyms, the following task is to determine concepts in a formal way. For delineating these concepts, in the next task are deduced lexical relations following lexical networks formulated by WordNet (Fellbaum 1998).

Once these lexical relations are established, the semantic relations are proposed, keeping this in mind, for example, first-order logic to represent predicate-arguments structures. The final process of this chain is to derive universal rules for building concepts, joining lexical and semantic relations deduced previously.

Thus, the recognition and extraction of concepts is a step towards the general goal proposed by Buitelaar, Cimiano & Magnini for building ontologies. For this particular phase, our proposal consists on identifying and extracting conceptual information through lexical-syntactic patterns as we mentioned above.

6 Towards the (semi-)automatic identification of lexical relations

In agreement with the methodology of Buitelaar, Cimiano & Magnini mentioned above, we propose to extract lexical relations from analytical definitions for covering the next step about hierarchical relations. Hiponymy/hypernymy and meronym/holonymy relations are considered as relations organizing a conceptual space in a hierarchical way (Winston, Chaffin & Herrmann 1987).

Additionally, our method provides a way to get more relations from a domain corpus through the application of a bootstrapping technique with the genus terms/wholes set as seed set.

- **Hyponymy/hyperonymy relations:** We consider works such as Hearst (1992), as well as Wilks, Slator & Guthrie (1995), because their methods allow combining linguistic and probabilistic criteria.

- **Part/whole relations:** In this case, we consider works such as Charniak & Berland (1999), as well as those results reported by Girju, Badulescu & Modolvan (2006). We propose a method exploiting the pattern with preposition *de*, due to its use frequency to link parts and wholes in Spanish. Table 6 shows examples about meronymy/holonymy relations using this pattern compared with other patterns worked in the literature.

Table 7: Number of hits returned by the search engine Google

Part	Whole	X is part of Y	Y has X	X of the Y
Mouse	Computer	27360	514	280400
Keyboard	Computer	60800	64730	1798000
Screen	Computer	58800	64100	556000

- **Attribution relations:** Attribution relations play an important role in disciplines involved with conceptual representation as artificial intelligence/knowledge representation, linguistics and psychology (Poesio & Almuhareb, 2005). So, we consider the work about the automatic extraction of attribution relations proposed by Poesio & Almuhareb (2004). They used an approach as that proposed by Charniak & Berland (1999) but to extract attribution relations using the pattern:

  “the * of the C [is|was]”

  Here, * represents a potential attribute for the concept C. In Spanish a common pattern to express attribution relations is the use of the preposition *de*, e.g.: *edad del paciente* (Eng. age of patient/patient's age), *altura del paciente* (Eng. height of the patient/ patient's height), and so on.

Summarizing, our methodology to extract lexical relations starts with the extraction of hyponymy-hypernymy relations from analytical DCs. For this phase we consider a lexical-syntactic approach due to the regularity of the definition schemes using predication patterns as those mentioned above.

Additionally, we propose a bootstrapping technique starting with the set of genus terms as a seedset to extract more lexical relations from a domain corpus. We use the preposition *de* to link genus term and other potential terms due to its importance to produce lexical relations of our interest.

For example, in a first phase exploring a domain corpus, a genus term as *dilatación* (Eng. dilation) links with a set of two elements {*vena*, *pupila*} (Eng. {vein, pupil}). In a next phase, the element *pupila* is linked to *ojo* (Eng. eye), and so on. Thus, on the one hand we have two relations IS-KIND-OF: *dilatación de la pupila* and *dilatación de la vena*. On the other hand, we have a meronymy-holonymy relation: *pupila*-*ojo*.

Integrating the three relations described above, we will implement a lexical network that allows organizing concepts related to terms. An example of this possible network is:

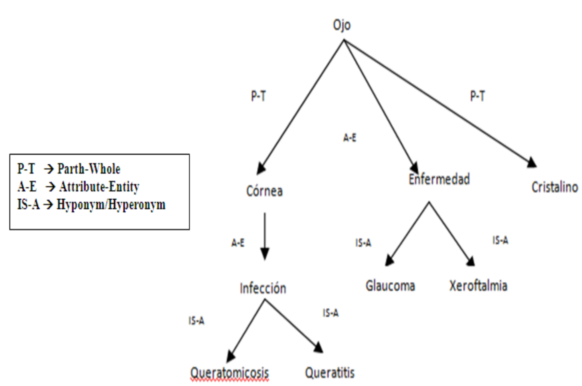

Figure 3: Example of a possible lexical network

In the figure, we can distinguish a set of sub-terms linked to the main term *Ojo* (Engl. Eye). These sub–terms operate as nodes, and the possible lexical relations are branches connected with the main term. Thus, based on a lexical Parth/whole relation, we can infer that *córnea* (Eng. cornea), is a constituent of eye. In contrast, the term *enfermedad* (Engl. *disease*) is an attribute of eye. Finally, the *glaucoma* is a type of disease that affects the eyes.

7 Work in progress and possible topics of collaborations

In this paper we proposed a method for recognizing lexical relations, taking into account the identification and extraction of analytical definitions situated into DCs in Spanish. This extraction considers verbal predications associated to these definitions. So, in order to explain this extraction, we have showed a formal syntactic analysis, based on the idea that these predications: (a) could be described in terms of predicative phrases, and (b), the association of predications and analytical definitions has a high frequency of use in specialized documents. For evaluating this frequency, we have exposed the results obtained for an experiment of extraction in two technical corpora.

Currently, we are situated in the phase to implement and evaluate a new experiment oriented to the detection of lexical relations between the term and the genus term formulated for analytical definitions. In particular, we are interested in discovering three types of relations: hyponym/hyperonymy, part-whole and attribution-entity.

We conclude suggesting some topics of collaborations for our potential colleges:

I. The construction of specialized texts corpora with good candidates of DCs, having in mind the basic features for identifying a DC.
II. The implementation of new linguistic and statistical methods for detecting and extracting lexical relations from text corpora.
III. The improvement of search systems, using these underlying lexical relations in electronic documents.
IV. Following to Wilks, Slator & Guthrie (1995), the design of lexical-semantic tags for recognizing and classifying concepts in taxonomies.

Similarly, according to Buitelaar, Cimiano & Magnini, we can use external lexical resources as Spanish WordNet and Spanish FrameNet (Subirats 2009) for determining and evaluating our lexical networks, in order to enrich the results that we could generate.

Acknowledgments

This paper was made possible by the financial support of the Consejo Nacional de Ciencia y Tecnología, CONACYT, and DGAPA–UNAM. Also, we wish to thank the anonymous reviewers for theirs comments and suggestions.

References

César Aguilar, Rodrigo Alarcón, Carlos Rodríguez and Gerardo Sierra. 2004. Reconocimiento y clasificación de patrones verbales definitorios en corpus especializados”. En Cabre T., Estopà R. & Tebé C. (Eds.). *La terminología en el siglo XXI*, IULA-UPF, Barcelona, Spain: 259-269.

César Aguilar. 2009. *Análisis lingüístico de definiciones en contextos definitorios*. Ph. D. Thesis, Department of Linguistics, UNAM, Mexico.

Rodrigo Alarcón, Gerardo Sierra and Carme Bach. 2007. Developing a Definitional Knowledge Extraction System. *Conference Proceedings of Third Language & Technology Conference LTC'07*, Poznań, Poland.

John Bowers. 1993. The Syntax of Predication, *Linguistic Inquiry*, 24(4): 591-636.

John Bowers. 2001. Predication. In Baltin, M. & Collins, C. (eds.), *The Handbook of Contemporary Syntactic Theory*, Blackwell, Oxford, UK: 299-333.

Paul Buitelaar, Philipp Cimiano and Bruno Magnini. 2005. *Ontology learning from text*. IOS Press, Amsterdam, The Netherlands.

Eugene Charniak and Matthew Berland. 1999. Finding parts in very large corpora. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*: 57-64.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*, MIT Press, Cambridge, Mass.

Roxana Girju, Adriana Badulescu and Dan I. Moldovan. 2006. Automatic Discovery of Part–Whole Relations. *Computational Linguistics*, 32(1): 83-135.

Marti Hearst. 1992. Automatic Acquisition of Hyponyms from Large Text Corpora. *Proceedings of the Fourteenth International Conference on Computational Linguistics*, Nantes, France.

Marie-Claude L'Homme and Patrick Drouin. 2006. C*orpus de Informática para el español*, Groupe Éklectick, OLST-Université de Montréal, Montréal, Canada: http://www.olst.umontreal.ca/.

Alfonso Medina, Gerardo Sierra, Gabriel Garduño, Carlos Méndez and Roberto Saldaña. 2004. CLI: An Open Linguistic Corpus for Engineering. In De Ita, G. Fuentes, O. & Galindo, M. (Eds.) *Proceedings of IX Ibero-American Workshop on Artificial Intelligence*, Puebla, BUAP: 203-208.

Massimo Poesio and Abdulrahman Almuhareb. 2004. Feature-Based vs. property-based KR: An empirical perspective. In *Proceedings of International Conference on Formal Ontology in Information Systems FOIS 2004*, Torino, Italy.

Ellen Riloff and Jessica Shepherd. 1999. A corpus-based bootstrapping algorithm for Semi-Automated semantic lexicon construction. *Journal of Natural Language Engineering* . 5(2): 147-156.

Carlos Rodríguez. 2004. *Metalinguistic Information Extraction from specialized texts to enrich computational lexicons*. Ph. D. Thesis, Universidad Pompeu Fabra, Barcelona, Spain.

Gerardo Sierra, Rodrigo Alarcón and César Aguilar. 2006. Extracción automática de contextos definitorios en textos especializados. In Inchaurralde, C. & Ibarretxe, I. (Eds.). *Memorias del XXII Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural*, University of Zaragoza, Zaragoza, Spain: 351-352.

Gerardo Sierra, Rodrigo Alarcón, César Aguilar and Carme Bach. 2008. Definitional Verbal Patterns for Semantic Relation Extraction. In Auger A. & Barrière C. (Eds.), *Pattern-based Approaches to Semantic Relation Extraction. Special issue of Terminology*, 14(1): 74–98.

Carlos Subirats (2009). Spanish Framenet: A frame semantic analysis of the Spanish lexicon. In Boas H. (Ed.), *Multilingual FrameNets in Computational Lexicography. Methods and Applications*, Mouton de Gruyter, Berlin/New York: 135-162.

Yorick Wilks, Brian Slator and Louise Guthrie. 1996. *Electric Words*. MIT Press, Cambridge, Mass.

Morton E. Winston, Roger Chaffin and Douglas Herrmann. 1987. A taxonomy of part-whole relations. *Cognitive Science*, 11(4): 417 – 444.