Towards the Building of a Corpus of Definitional Contexts

Gerardo Sierra, Rodrigo Alarcón, César Aguilar, Alberto Barrón
gsierram,ralarconm,caguilar@ii.unam.mx; alberto@pumas.ii.unam.mx
Instituto de Ingeniería, UNAM
Torre de Ingeniería, Basamento
Ciudad Universitaria
México 04510, D.F.

Abstract
A repository of Definitional Contexts, even more than to be conceived of as traditional corpora, is very valuable tool for lexicography. This paper describes a corpus of Definitional Contexts, from the definition of the term to the description of the applications of the corpus in lexicography. The intended applications for the work, and how these affected the corpus design, are described. In addition, the methodology for corpus building and the corpus structure are outlined, and the preliminary results of the work in progress are presented.

1 The need for Definitional Contexts

Text corpora are widely used in lexicography because they provide information about words, such as their frequency, and allow the lexicographer to extract and analyse concordances, which can be useful for distinguishing the possible different senses of a word. However, for specialised lexical units, also called *terms*, viewing the context is not enough to clarify the meaning, and more information is needed to describe.

Previous studies have been carried out to provide the lexicographer with concordances of discursive structures that link a term with its definition in specialised texts. Among them, we should note the systematic search for definitions in specialised corpora (Pearson 1998; Meyer 2001), the work on metalinguistic information extraction (Rodríguez 2004), and the use of *énoncés définitoires* (Auger 1997; Rebeyrolle 2000).

For the purposes of this article, based on previous research by the Language Engineering Group (Alarcón & Sierra 2003), we consider Definitional Contexts (henceforth referred to as DCs): textual fragments in specialised texts in which information relevant for defining a term is given. A DC is formed by its minimal constitutive elements: a term, a definition, and usually a defining verb governing syntactic patterns. Also, some typographical and discursive markers may commonly be used to highlight the presence of either a term or a definition. Here we present an example of a DC:

La energía primaria, en términos generales, se define como aquel recurso energético que no ha sufrido transformación alguna, con excepción de su extracción.

In this case, we can see that the DC sequence is formed by the term *energía primaria*, the definition *aquel recurso...*, and the verbal predication *se define como* (Eng: is defined as), as well as other characteristic units such as the discursive marker *en términos generales* (Eng: in general terms) and the typographical marker (bold font) that in this case emphasises the presence of the term.

2 Applications of DCs

A corpus of DCs is an important tool for lexicographic work as well as work in other areas such as terminology, information extraction, and text mining. Here we outline the applications of the DC corpus for the Language Engineering Group at the Engineering Institute, whose central project is the creation of onomasiological dictionaries (Sierra and McNaught 2000), i.e. dictionaries that provide help to users who want to express a specific concept they have in mind but do not remember the term for it.

2.1 DCs for linguistic definition analysis

We can see a definition as a linguistic description of a concept represented by a term. This description establishes relationships with other terms, in order to delimit the meaning of the concept. We identify five types of definitions according to Figure 1:

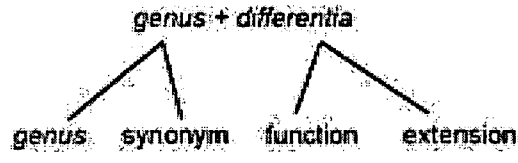

Figure 1. Typology of definitions

- Analytic definition: specifies genus + differentia.
- Exclusive genus definition: provides no description of the differentia.
- Synonymic definition: indicates a strong semantic relationship with the genus.
- Functional definition: includes differentia that indicate the function of the concept.
- Extensional definition: includes differentia that enumerate the parts composing the concept.

We identified regular specific patterns and structures associated with the representation of concepts in natural language. In Spanish, most definitions start with a noun phrase, although in functional definitions verbal phrases are also common.

- Noun phrases can start with a quantifier (*todos, algunos, cada uno, ninguno*), determiner (*un, una, unos, unas, el, la, los, las*), or demonstrative (*este, esta, estos, estas*).
- The genus may consist of a set of prepositional phrases after the initial noun phrase.
- The differentia may be introduced by subordinated clauses composed of noun, adjective and prepositional phrases.

Employing these patterns and structures is a productive method for recognising and classifying DCs in a corpus. Verbal predications are composed of a verb plus a grammatical particle and structures such as adverbs, adjectives, articles, prepositions or phrases. They establish a strong relationship with the type of definition introduced in a DC.

A DC corpus will be helpful in analysing – from a syntactic and semantic point of view – the relationships among these elements, as well as the arguments for each of the different types of definitions (e.g., agent/source, theme, goal).

2.2 DCs for automatic definition extraction

Terminography can be understood as the practice of creating dictionaries of specialised terms. Terms may be identified in corpora either manually or automatically. In the field of automatic term extraction, there are various possibilities for achieving this goal (Cabré, Estopà & Vivaldi 2001); one of these is a rule-based system that identifies recurrent syntactic patterns. Methodologies have also been developed for extracting definitions from text, e.g., based on searching for linguistic and metalinguistic patterns (Rodríguez 2004; Klavans & Muresan 2001).

We are working on the development of a tool for automatically extracting definitional contexts in Spanish from corpora. Previous work on a small corpus on engineering helped us to identify a collection of defining verbs commonly used in definitional contexts to link terms and definitions (Sierra and Alarcón 2002). This repository of syntactic patterns has been analysed, modified and extended by a methodology that will be explained in Section 3.

A corpus of DCs should provide a basis for analysing the constitutive elements of a definitional context. Furthermore, it could be a starting point for developing extraction algorithms for a rule-based system. This means that the DC corpus is helpful for analysing the structure of terms, definitions and characteristic elements, and will help to formulate rules for extracting candidate DCs from annotated corpora.

2.3 DCs for onomasiological dictionaries

An onomasiological dictionary is a user-oriented tool for accurately retrieving the terms associated with a concept. The dictionary consists of a system that analyses the information in lexical resources in order to create a lexical knowledge base (henceforth abbreviated as LKB), which is then matched with a query given by the user to provide the relevant information. The LKB provides all the knowledge necessary for onomasiological searching. In principle, it must represent what a person knows about both concepts and their corresponding terms. A person is able to determine the inherent properties of a concept and associate the conceptual properties with a set of common terms. Our LKB consists then of a set of terms, a set of definitions for each term, a set of keywords associated with the definitions and a set of lexical paradigms that group keywords with the same meaning. It includes not only the databases that constitute these sets of data, but also the interrelationships among all the sets.

Therefore, a corpus of DCs provides definitions (either lexicographical or terminological), which are the basic elements for the LKB and the primary source of linguistic knowledge to be considered. DC corpora are an important resource for attaining the high degree of completeness required in the LKB.

3 Methodology for DC identification

In this section, we present the different methodologies we used to obtain definitional contexts. These methodologies constitute the steps of DC corpus building, and are related to the need to identify and classify definitional contexts in different fields, in order to obtain a representative collection. For this purpose, we searched for contexts in which the verbs are linked to the Genus and Differentia. These defining verbs are: *comprender*, *concebir*, *conocer*, *considerar*, *definir*, *denominar*, *entender* and *identificar* (Eng: to comprehend, to conceive, to know, to consider, to define, to denominate, to understand and to identify).

3.1 Obtaining DCs through the Bwana

The Bwana is a search engine developed by the Instituto Universitario de Lingüística Aplicada (IULA). This engine allows users to search for linguistic occurrences in IULA's Technical Corpus (BwanaNet). We used this corpus to obtain part of our collection of DCs.

3.1.1 Search methodology

The first step was the search for the occurrences of the defining verb lemmas. We restricted to search to 500 occurrences of the lemma, using the random option to retrieve the contexts. The searches were carried out on all fields of the corpus.

Once we had retrieved contexts containing the defining verbs, we manually analysed them in order to find definitional contexts. We selected and classified each DC according to the inflexion of the verb. This taxonomy helped us to identify a set of definitional patterns.

3.1.2 Definitional patterns

Using this set of definitional patterns, we started a new search. We translated the patterns into regular expressions, according to the syntax of the EAGLES tags used by IULA's Technical Corpus.

Definitional patterns are composed of a defining verb and a grammatical element such as the Spanish adverb *como* and the pronoun *se*. The patterns present different structures according to the inflexion of the verb, and most of the defining verbs shared similar structures. Nevertheless, some definitional patterns followed particular structures depending on the verb.

Here, we show some examples of definitional patterns and the syntax we used to find them in the Bwana engine. We present three different cases related to the verbs *definir*, *entender* and *denominar*. In the next example, we can see a pattern common to these verbs:

```
[word="se"] [pos="R.*"] {0,1} [lemma="definir|entender|denominar" & pos="V[^IGC] ..."]
[word!="como"] {0,15} [word="como"]
```

The pattern is formed by the pronoun *se*, another optional pronoun expressed by the EAGLE tag "R", the lemma of the defining verb in inflected form, any optional word present in the next 15 words except the adverb *como*, and finally this adverb. The optional distance of 15 words takes into account the fact that the term or another element could appear between the defining verb and the adverb *como*.

With this pattern, we can retrieve contexts such as the following, in which we found the term, *hemólisis*, within the optional distance:

Estudio de la hemólisis: <search pattern>se define la hemólisis como</search pattern> la reducción de la duración de la vida de los hematíes, que normalmente es de 110 + 10 días.

3.1.3 Results

Some examples of the definitional patterns are shown below. Each one is specific to a defining verb:

Definir:
[pos="Z"] [lemma="definir" & pos="HMS"] [word!="como"] {0,15} [word="como"]
Entender:
[pos="V.*"] {1,2} [lemma="entender" & pos="HMS"] [word!="como"] {0,15} [word="como"]
Denominar:
[pos="N.*"] [pos="J.*"] {0,2} [lemma="denominar" & pos="HMS"]

In these examples, the defining verb occurs as a past participle. In the first one, the tag "Z" represents any punctuation mark. In the second example, we take into account the occurrence of at least one auxiliary verb. Finally, in the third example the tags "N" and "J" represent a noun and an adjective, which are common elements in specialised terms.

Using the methodology described above, we retrieved a total of 10,589 contexts. In the first stage, the search for the defining verb lemmas, we found 4,352 contexts, of which 363 were DCs and 3,989 were not (NO-DC). Searching with definitional patterns, we found 3,095 DCs and 3,142 NO-DCs.

There is a substantial difference between approaches involving obtaining DCs using defining verb lemmas and those using definitional patterns; with the latter we can obtain better results, both in precision and recall.

Further, we measured the precision of each definitional pattern. This measure, widely used in Information Retrieval, was obtained in our case by dividing the number of DCs found by the total number of contexts retrieved. The closer to 1 the precision value is, the better the precision results, i.e., the less noise found in retrieving definitional context candidates.

The precision values of the Genus & Differentia verbs are, in decreasing order: *denominar*, 0.720; *conocer*, 0.5183; *concebir*, 0.5116; *entender*, 0.4559; *definir*, 0.4483; *identificar*, 0.2133; *comprender*, 0.1580; and *considerar*, 0.1267.

We should note that searching for definitional patterns allows us to retrieve more DCs with less noise.

3.2 Obtaining DCs through the CLI

While we searched for and integrated candidate DCs from the BWANA Corpus, we also obtained other candidates from an extract of the Linguistic Corpus on Engineering (CLI, in Spanish) (Medina *et al.* 2004). This extract contains approximately 500,000 words. The CLI

is a corpus composed of technical documents (papers, reports, dissertations, etc.) in different thematic areas of Engineering.

3.2.1 Methodology of the search for patterns

The recognition of DCs in the CLI extract was performed using specific verbal patterns. For this task, we employed an electronic tool developed in Python by Rodríguez (2004).[1]

3.2.2 Search methodology

For the process of pattern extraction, we considered these verbal forms:

• Third person singular and plural inflexions (*define, caracteriza, es/son...*, etc.)

• Past and present participles, because they are associated in Spanish with the construction of verbal periphrasis in past perfect tense (*se ha definido como, ha significado...*), or in passive constructions (*siendo entendido como, es visualizado como...*)

• The auxilliary *poder* (Eng: can) in verbal periphrasis (*puede referir a, puede ser considerado como...*)

3.1.3 Results

We can see an example of the results of this process in Figure 2:

Total of lines	Extracted	Percentage
12532	89	0.710181934248

Statistics by pattern

conocido	19
conoce	18
conocida	13
conocer	36
conociendo	2

Figure 2. XML table with a DC Candidate

4 Corpus structure

After the extraction and classification of possible candidates from both corpora of texts, the BWANA and the CLI, we established canonical patterns of DCs. These kinds of struc-

[1] The input for this system is a set of previously delimited text fragments. The output is a XML table with a list of patterns obtained, the particular token associated with the pattern, and the frequency of use in the CLI extract as a whole.

tures are described as those patterns that explicitly show a defining verb with its syntactic arguments (Sierra *et al.* 2003; Aguilar *et al.* 2004), e.g., contrasting verbs such as *definir*, *considerar* (to consider) or *ser* (to be). In the first case, *definir* and *considerar* show three possible arguments: a NP or *pro* with the role of Agent or Source (Jackendoff 1991; Fillmore *et al.* 2002), a NP representing the Term in the Theme role, and an Adjunct Phrase or Clause introduced by a adverb *como* (as/like) with the definition in the role of Goal. Some examples of these DCs are shown in (a) and (b):

a) [<El Minvu> <NP=Agent/Source>] [<define> <Verb>] [<Gobernabilidad Urbana> <NP=Theme>] [<como> <Adverb>] : [<identificar tareas a realizar y quien las hará, las funciones que deben descentralizarse y las que deben centralizarse> <Adjunt =Goal>]
b) [<Para esta dirección> <NP=Agent/Source>] [<el paisaje> <NP=Theme>] [<se concibe como> <Verb>] [<una entidad espacial, un ensamble de ecosistemas en interacción centrando su interés en los diferentes fenómenos relacionados con el intercambio entre los sistemas y la heterogeneidad espacial> <Adjunct =Predication>].

The verb *ser* includes one argument of the NP type in the Theme role, which identifies the Term, and a nominal predication as definition, e.g., the case of (c):

c) CUCHILLA FUSIBLE [<La cuchilla fusible> <NP=Theme>] [<es> <Predicación Verbal Definitoria>] [<un elemento de conexión y desconexión de circuitos eléctricos>. <FN/Definición>]

This corpus shows the constituent units of DCs. For this reason, we have been establishing a set of tags in XML that identify both the DC as a whole and the particular elements integrated within it. The tags allow us to display an XML archive divided into a Head and a Body. The fields declared in the Head are:

Tags	Function
Name	Name of unit found in a document, e.g., verb *ser*
Source	Reference to the original corpus: BWANA or CLI
Date	Date of compilation and tagging
Pattern	Three construction patterns: 1. Canonical pattern (e.g., the copula *ser* associated with the Aristotelian Definition: Term + Verb + Definition) 2. Insertion of a NP or Impersonal Pronoun (*se*, in Spanish) referring to Author of Definition (e.g., X *define* Y *como* Z). 3. Patterns associated with insertion of an adverbial or prepositional particle (e.g., X *se considera como* Y, X *refiere a* Y, X *sirve para* Y)
Type	Type of definition: Genus & Differentia, Exclusive Genus, Synonymy, Functional or Extensional
Compiler	Name of Compiler

Table 1. Tags associated with the Head of the XML archive

This information not only has a great value for identifying the verb associated with a DC

in a text, but also for managing our corpora. In Figure 3, we illustrate the tags related to the Body of the XML archive, each of which is explained in Table 2.

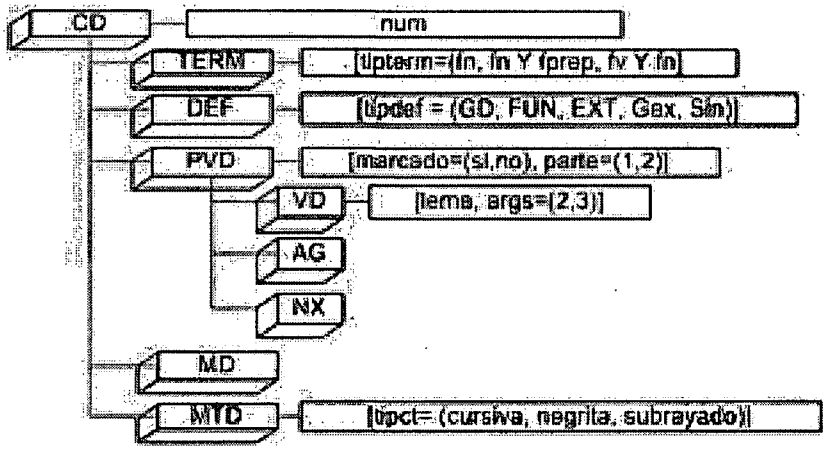

Figure 3. Three descriptions of XML tags in the Body

Tags	Signifi	Function
CD	Definitions Context	Indicates the start and the end of a DC.
TERM	Term	Refers to the term considered. The attribute tipterm alludes to three possible phrase structures: NP (e.g., article, noun and adjective), NP + PP, and VP (with a infinitive verb) + NP.
DEF	Definition	Indicates the definition associated with the term. For this case, we consider as attributes the type of definition: GD (Genus & Differentia), FUN (Functional), EXT (Extensional), Gex (Exclusive Genus) and Sin (Synonymy).
PVD	Definitional Verbal Predication	Alludes to the constituents of Verbal Predication, that is: a verb, an explicit Agent (Subject) that defines something, and the nexus that introduces the definition (an adverb or a preposition). The attribute *marcado/no marcado* (marked/unmarked), refers to canonical pattern considered: Agent + Term + Verb + Nexus + Definition is the unmarked case in contrast to other cases, such as passive constructions (Term + Aux + Verb + Nexus + Definition), constructions with an impersonal pronoun *se* (*se* + Term + Verb +Nexus + Definition), or inverted constructions (Definition + *se* + Verb + Nexus + Term). In addition, in some cases it is necessary to take into account the division of a verbal structure *se* + Verb + Nexus into two units, via the insertion of a Term, for example: <*se define*> [*se* + Verb] <*el coeficiente de bloqueo S*> [Term] <*consiste*> [Nexus] <*el cociente entre el área de la sección transversal del bagre y el área de la sección transversal del canal*> [Definition]. In this case, we assign the attributes Part 1 and Part 2 to situate both units as specific constituents of definitional verbal predication: *se* + Verb (Part 1) / Nexus (Part 2).
VD	Defining Verb	Identifies the defining verb that acts as the nucleus in the verbal predication. Depending of the semantic relationships mapped for the verb in a DC, syntactically it can establish a certain number of arguments (e.g. Agent/Source + Theme + Goal versus Theme + Predicate). If the verb contains three possible arguments, we introduce the number *3* associated with the attribute *arga*; we introduce the number *2* when the verb accepts only two arguments.
AG	Agent	Inserted in the case of verbs that have a semantic argument Agent, to recognise the author or authors of the definition or certain syntactic units that refer to this semantic role, such as

		personal pronouns (e.g., the first person plural pronoun *nosotros*) or the impersonal particle *se*.[2]
NX	Nexus	Indicates the adverb or preposition that functions as the nexus after the defining verb and introducing the definition.
MD	Discursive Marker	Defines discursive constructions that introduce a possible DC: *es decir, generalizar, de acuerdo con esta teoría, para los fines de la investigación*, etc.
MTD	Typographic Marker	Identifies all types of typographical features (italics, bold, underlining, capital letters, etc.) that highlight the term or the definition.

Table 2. Tags for identification of DC units

5 Work in progress

5.1 Typology of definitions and their relationships with verbal predications

Given our aim of obtaining terms and definitions from real texts in any natural language (in our case, Spanish), it is necessary to take into account the relationships established among such terms and definitions and the verbal predications that connect them.

This corpus allows us to perform a more detailed analysis of how a verbal predication determines the type of definition linked to a term in a DC. In this work, we have described the relationships that definitions of Genus & Differentia maintain with verbs such as *entender*, *concebir*, and *definir*, in the tri-argumental structure Agent/Source + Theme + Goal. Additionally, we have identified other verbs such as *referir, representar, ser* and *significar* (Eng: to refer, to represent, to be and to mean/to signify), related to definitions of Genus & Differentia. The important distinction between these groups of verbs is that the latter group establishes a bi-argumental structure of predication composed by a Theme (that is, the Term in the role of Subject of these predication) and the predicate introduced after the verb.

However, we believe that these two types of relationships are not the only ones. For example, verbs such as *contener, incluir* and *integrar* (Eng: to contain, to include and to integrate) establish the nexus in Extensional Definitions; verbs such as *emplear, servir* and *usar* (Eng: to employ, to serve and to use) are linked to functional definitions. This corpus offers an important set of data that helps to accomplish a better analysis of these relationships between verbal predications and types of definitions.

5.2 Concept extraction

DCs can be seen as the first stage in developing the conceptual framework for dictionary and glossary creation, as well as in determining semantic relations or finding new terms in specialised texts. Furthermore, DCs could help in many other lexicographic tasks, as we have seen in this paper. We have observed that searching for DCs using their definitional patterns is a good starting point for developing a rule-based DC-extractor system.

In order to identify these contexts automatically using corpora, we have recognised the need for a DC corpus that helps us with the job of identifying and classifying definitional

[2] For this purpose, we used the criteria described in *FrameNet* (http://framenet.icsi.berkeley.edu/) in relation to verbs such as to conceive, to define and to understand.

patterns. These patterns must at least help us to find DCs through the presence of defining verbs. However, a DC is a more complex structure that includes discursive markers, typographical marks, and pragmatic relations that highlight the presence of both term and definitions. Considering these characteristic elements would be helpful in developing rules for extracting and classifying the elements present in DCs.

Thus, this investigation contributes to a much-needed approach within terminographical research in the Spanish language, and adds new information to previous research based on the English language.

5.3 Defintions in the onomasiological dictionary

As the onomasiological dictionary is concept-oriented, in order to permit searching for terms by means of their meanings, the term and the concept must be extracted. Definitions refer to a concept and identify it with respect to all others, giving its essential properties and characteristics, along with some other relevant facts. Since definitions are easily identifiable in lexical resources, either dictionaries or specialised corpora, they constitute the basic information to be extracted. After recurrent linguistic patterns highlighting the presence of definitions in specialised texts were identified, an expert group was created to work on extracting definitions from different corpora in which terms are introduced and their concepts explained. Both DCs and definitions were captured for each term in the terminological data-bank (Sierra *et al.* 2003), on the subject of Linguistics, Physics (Lara *et al.* 2000), and sexuality (Medina and Sierra 2004).

For our purposes, definitions provide sufficient information for the onomasiological dictionary, as this type of dictionary contains the same information as a semasiological one, but with the difference that look-up is carried out using the definition rather than the term as a starting point. In this sense, lexical definitions provide a description of the necessary and sufficient characteristics for identifying a concept. If someone reads a lexical definition, they can often infer the word that denotes it. However, we take into account that the onomasiological dictionary must allow users to input the concept to be searched through the ideas they may have, using any words in any order; as such, there are several methods and a variety of words that may be used to formulate a concept. Therefore, the target for the terminological data-bank was to get the most definitions possible for each term, as a single lexical definition is not enough to retrieve the word; the individual definitions do not in themselves contain sufficient information. This paradoxical ambivalence is our reason for considering both dictionaries and corpora as the basic and essential resources of our LKB: instead of an isolated resource, we use as much data as necessary drawn from different resources to answer user queries.

Since the application of the onomasiological dictionary is restricted to specific domains, the best starting point is therefore a corpus describing the conceptual framework of that domain. As specialised corpora can extend the description of a concept beyond a useful kind of knowledge, the extraction of the properties requires the specification of a limit. This problem affects the compilation of terminological data, which relies on some initial principles such as the subject of the work, the users, the purpose of terminology processing, and the

scope and extension intended. Although the effort to identify concepts from a conceptual framework in a corpus is somewhat more difficult than using a dictionary as a database, it is also worthwhile to point out that the benefits of using a richer resource can far exceed the cost of extracting the information.

Acknowledgements

This paper is supported by the National Council for Science and Technology (CONACYT), Ref. 46832-H; and the Macro Project *Tecnologías para la Universidad de la Información y la Computación*, UNAM. The authors would like to thank Elizabeth Marshman for her kind help in checking and correcting the text.

References

A. Corpora

Cabré, T., Vivaldi, G. (Coords.)(1997), *Corpus Tècnic del IULA de la UPF* (CT-IULA). http://bwananet.iula.upf.edu/indexes.htm

Medina, A., Sierra, G., Garduño, G., Méndez, C., Saldaña, R. (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, México, Autonomous University of Puebla, pp. 203-208.

B. Other Literature

Aguilar, C. *et al* (2004), 'Reconocimiento y clasificación de patrones verbales definitorios en corpus especializados', *IX Simposio Iberoamericano de Terminología*, Nov 29-Dec 2, Barcelona.

Alarcón, R. Sierra, G. (2003), 'El rol de las predicaciones verbales en la extracción automática de conceptos', in *Estudios de Lingüística Aplicada* 38, pp. 129-144.

Auger, A. (1997), *Repérage des énonces d'intérêt définitoire dans les bases de données textuelles*, Thèse de Doctorat, Neuchâtel, Suisse, Université de Neuchâtel.

Cabré, T. Estopà, R. Vivaldi, J. (2001), 'Automatic term detection. A review of current systems', in Bourigault, D. Jaquemin, C. L'Homme, M. C. (eds.) *Recent Advances in Computational Terminology*, Amsterdam, Benjamins, pp. 53-87.

Fillmore, Ch. *et al* (2004), *FrameNet*, University of California at Berkeley: http://framenet.icsi.berkeley.edu/.

Jackendoff, R. (1991), *Semantic Structures*, Cambridge, Mass, MIT Press.

Klavans, J., Muresan, S. (2001), 'Evaluation of DEFINDER: a system to mine definitions from consumer- oriented medical text'. *Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries*. New York, ACM Press.

Lara, F. *et al* (2000), 'Sistema Inteligente Computarizado para el Aprendizaje Conceptual e Interactivo de la Física a Nivel Bachillerato'. *SOMI XV Congreso de Instrumentación*, Guadalajara, México.

Medina, A. Sierra, G. (2004), 'Criteria for the Construction of a Corpus for a Mexican Spanish Dictionary of Sexuality', in *Proceedings of 11th Euralex International Congress*, Université de Bretagne-Sud, July 6-10, Lorient, France.

Meyer, I. (2001), 'Extracting a knowledge-rich contexts for terminography: A conceptual and methodological framework', in Bourigault, D. Jaquemin, C. L'Homme, M.C. (eds.), *Recent Advances in Computational Terminology*, Amsterdam, Benjamins, pp. 279-302.

Pearson, J. (1998), *Terms in Context*, Amsterdam, Benjamins.

Rebeyrolle, J. (2000), 'Utilisation des contextes définitoires pour l'acquisition de connaissances à partir des textes', in *Actes des Journées Francophones d'Ingénierie des Connaissances, IC'2000*, Toulouse, France, pp. 105-114.

Rodríguez, C. (2004), *Metalinguistic Information Extraction from specialized texts to enrich computational lexicons*, Ph. D. Dissertation, Universitat Pompeu Fabra, Barcelona.

Sierra, G. Alarcón, R. (2002), 'Identification of recurrent patterns to extract definitory contexts', in *Lecture Notes in Computer Science* 2276, pp. 436-438.

Sierra, G. *et al* (2003), 'Definitional Contexts Extraction from Specialised Texts', in Lewandowska-Tomaszczyk, B. (ed.), *PALC 2003 Proceedings: Language, Corpora and E-Learning*. Frankfurt, Peter Lang, pp. 21-31.

Sierra, G. McNaught, J. (2000), 'Design of an onomasiological search system: A concept-oriented tool for terminology', in *Terminology* 1-6, pp. 1-34.