Enriching a lexicographic tool with domain definitions: Problems and solutions

María A. Barrios
Universidad Complutense de Madrid
Paraninfo Ciudad Universitaria s/n
28008-Madrid. Spain
auxiba@filol.ucm.es

Guadalupe Aguado de Cea
Ontology Engineering Group – UPM
Avda. Montepríncipe s/n
28660-Boadilla del Monte. Spain
lupe@fi.upm.es

José Ángel Ramos
Ontology Engineering Group – UPM
Avda. Montepríncipe s/n
28660-Boadilla del Monte. Spain
jarg@fi.upm.es

Abstract

Enriching linguistic resources with domain information has been considered one important target in natural language applications. However, automatic definition extraction of this domain information from specialized resources has revealed certain methodological problems in definition construction. This paper presents some problems encountered in automatic definition extraction that are mainly related to inconsistencies in definitions, different granularity of definitions and embedded definitions. To face these problems some Meaning-Text Theory tools have been used: (a) semantic labels as a solution for inferring knowledge, (b) lexical functions as a way of providing coherence to definitions and (c) the actancial structure as a tool for developing consistent and complete definitions. Our goal is to describe the problems and to show the solutions proposed.

Keywords

Definition extraction, ontology building, linguistic resource enrichment, Meaning Text Theory.

1. Introduction

Reusing and enriching existing resources are nowadays two key issues both in academy and in the business world. In several scientific disciplines such as ontology development, computational linguistics, web semantic, ontologies and computational terminology the interest has been focused on many different aspects ranging from reusing lexicons, thesauri to create ontologies to extracting semantic relations from domain corpora or enriching definitions from specialized texts. One of the current drifts tends to build ontologies extracting definitions from different sources. However, building new resources with linguistic information extracted from different domain sources has revealed a difficult task as quite often the domain sources can be useful for a certain task but may show certain inconsistencies for others. In this paper, we present the problems encountered when trying to reuse three domain resources for two different purposes: (a) to build an ontology and (b) to populate a general linguistic resource, a database, with specific information from domain documents. With the aim of developing a consistent linguistic resource for further use in natural language applications, we focus on achieving consistent definitions of domain terms. Accordingly, we resort to the Meaning-Text Theory (MTT) principles [16] to propose some systematic solutions in order to avoid the inconsistency problems when building a terminological resource that can later be used in ontology development. Thus, we have mainly focused on three fundamental aspects: (a) semantic labels as a solution for inferring knowledge, (b) lexical functions as a way of providing coherence to definitions and (c) the actancial structure as a tool for developing consistent and complete definitions.

The rest of the paper is organized as follows: In section 2 we provide the scenario in which we have based our research and the tools used. Section 3 focuses on definition extraction and the pitfalls faced in the process. Section 4 presents a short review on definition typology. The MTT tools used and the database, BADELE 3000, are described in section 5. The problems encountered and the solutions proposed are presented in section 6. Finally, some conclusions are outlined in section 7.

2. Background

The domain resources used in this project summarized in this section (for more details, see Gómez-Pérez *et al* [7]) relate to geographic and geospatial information. All geographic information (GI) resources contain data about real entities and how to represent them in a map. So, each entity corresponds to an instance of a geographic phenomenon (*feature*). Indeed, the most important concept for GI is the *feature* since the Open GeoSpatial Consortium (OGC) [19] has declared that a geographic feature is the starting point for modelling geospatial information. In other words, a *feature*, which is the basic unit of GI, is an abstraction of a real world phenomenon associated with a location relative to the Earth, about which data are collected, maintained and disseminated [11]. Features can include representations of a wide range of phenomena that can be located in time and space such as buildings, towns and villages or a geometric network, a geo-referenced image, pixel or thematic layer.

For modelling this domain we have decided to use an ontology. To achieve this target, we have used three domain resources provided by the National Geographic Institute of Spain (IGN-E): the Concise Gazetteer (NC) -scale 1:1,000,000-, the Numerical Cartographic Database (BCN25) -scale 1:25,000-, and the National Topographic Database (BTN25) -scale 1:25,000-.

The Concise Gazetteer is a basic corpus of standardized toponyms created by the Spanish Geographical Names Commission. The first version has 3667 toponyms. This gazetteer complies with the United Nations Conference Recommendations on Geographic Names Normalization. The Concise Gazetteer has been created by the Spanish Geographical Names Commission. For further details, refer to Nomenclátor Geográfico Conciso de España [18].

The BCN25 presents an abstraction of reality, represented in one or more sets of geographic data, as a defined classification of phenomena. It defines the feature type, its operations, attributes, and associations represented in geographic data. For more information on this document see Rodriguez [21].

The BTN25 is the latest IGN-E catalogue and intends to be a sort of BCN25 reorganization, following a structure similar to frames. The instance information is the same as in BCN25, but the phenomena classification and its attributes are completely different.

These resources have one characteristic in common: each resource has a domain dictionary with phenomena. In the first case, NC phenomena, there is a txt file with 22 definitions. In the second case, BCN25 phenomena, an Excel file contains 366 definitions developed after the catalogue. Finally, there is a PDF document with “Capture rules for GI to be included in BTN25” (a first version), which describes its phenomena with 292 definitions (the document is not complete). In all cases, definitions were formulated by specialists on geography to facilitate the classification of the real entities in order to be included in the instance set of each resource.

All definitions are grouped by labels, as illustrated in Table 1 with four examples. These definitions have been used to build the ontology, as explained in section 3.

Table 1. INDUSTRIAL INSTALLATION (source document)

Nouns	Definitions
Corral (corral)	***Construcción*** *creada para cobijarse los pastores o para recoger el ganado* (**Construction** created for shepherds or cattle shelter)
Granja (farm)	***Hacienda*** *de campo que consta de establos, huerta y casa habitable* (**Ranch** with stables, an orchard and a house)
Piscifactoría (fish farm)	***Instalación*** *en la que se crían diversas especies de peces y mariscos con fines comerciales* (**Installation** where fish or seafood are bred for commercial purposes)
Palomar (pigeon loft)	***Edificio*** *donde se recogen y crían palomas* (**Building** where pigeons live and are bred)

3. Definition extraction

Definition extraction, as used in this paper, is the process of extracting the definition for a term from different resources. In our case these definitions have not been taken from corpora using machine learning techniques, as in many natural language processing applications [3], but from other domain resources with explicit definitions for these terms, their term variants or other semantically equivalent terms. However, some problems have appeared in this definition extraction process that showed certain inconsistencies and loss of information.

The definition extraction process followed to build and enrich a domain ontology is as follows: (1) the application we have developed retrieves the term from “Capture rules for GI to be included in BTN25”; (2) it extracts its definition from the same document; (3) it searches for the term in the auxiliary domain dictionaries; and (4) it extracts the corresponding definitions to add them to the corresponding classes. All these actions are executed automatically. Fig. 1 shows the overall workflow of information.

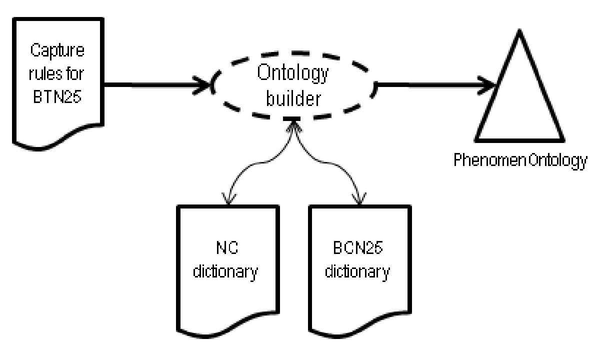

Fig. 1. Ontology building with definition extraction

As a result of this process, we obtained an ontology (called PhenomenOntology 3.5) which included 108 terms extracted from the documents mentioned and later transformed in 108 classes belonging to three groups: (a) classes without definitions; (b) classes with one definition; (c) classes with more than one definition. However, the retrieval ratio of definitions extracted from the auxiliary dictionaries was very low, although they belonged to the same domain. In fact, only 4 definitions were found in the NC dictionary (although it contains 22 definitions, which means that 18 definitions were lost in the process) and 33 definitions were found in the BCN25 dictionary (it contains 366 definitions, which means that 333 definitions were also lost in the definition retrieval process).

The origin of this low ratio mainly lies on the abundance of terminological variants and semantically equivalent terms. For example, when trying to retrieve definitions for '*río*' (river) in the ontology, the system cannot recognize definitions of term variants such as '*río 1ª categoría*' (river 1st category) and '*río 2ª categoría*' (river 2nd category), and consequently it does not retrieve any of these definitions. Moreover, semantically equivalent terms are not retrieved when incorporating definitions in the ontology, as the system cannot recognize the similarity of the definitions of '*río*' (river) and '*corriente fluvial*' (flowing current).

Therefore, the problem is not only the loss of certain definitions in the extraction process but also the overlapping of some of them with different granularity which led to inconsistencies. For example, '*río*' (river) was retrieved with two definitions: *recorrido de una corriente de agua natural y de caudal más o menos constante, que recoge los aportes de una cuenca fluvial* (taken from the original document BTN25: "stream of natural water, with more or less constant flow, which collects water from other water courses") and *curso natural de agua* (taken from the NC dictionary: "waterstream").

Although these terminographic resources have been originally compiled by different experts, they show many lexico-semantic divergences that hinder the automatic definition extraction process. Quite often specific domain lexicographic resources are generally built to share information within a project team and attention is not usually paid to terminological principles when defining new terminology.

In other words, when building ontologies, automatic extraction of classes implies the annotation of these classes with definitions which are also automatically extracted. The final result of the definition extraction process reveals some problems that we have tried to tackle as explained in the next sections. Nevertheless, ontology building problems are out of the scope of this paper, though they have served as test bed for our research on principles for definition writing.

4. Definition typology

According to the traditional aristotelic genus-species definition, a definition should describe the concept and its relations to other concepts in the concept system. This type of definition is traditionally called formal definition, or intensional definition [8, 9]. That is to say, it reflects the superordinate concept to which the designation belongs and its delimiting characteristics. However, there are also other ways of designating concepts, extensional, ostensive, lexical, precising, and stipulative definitions [8] as well as ontological definitions [4]. For a more exhaustive revision on definitions see [13, 12]. Although these definitions can be useful for certain purposes depending on the user's needs and the approach adopted, they do not conform to a certain defining formulation and hinder any possibilities of formalizing the knowledge expressed in definitions in order to be used for natural language applications, such as knowledge extraction, ontology enrichment, to mention just a few. For this reason, we claim that some recommendations regarding terminological definitions should be considered when preparing domain resources. As [9, 10] stipulates the selection of an appropriate superordinate is crucial for the intelligibilility of the defining statement. In Pearson's words [20] "the superordinate or closest generic concept should preferably be one step up in the hierarchy from the term being defined". Moreover, the same superordinate should be used for all terms that belong to the same class.

5. MTT lexicographic tools and BADELE.3000

In order to get more accurate systematic definitions, we decided to use the MTT tools. We considered two possible ways, (a) applying these tools directly to the ontology; (b) using them to enrich a general purpose lexicographic resource which could be later reused in other applications, for instance, for mapping the PhenomenOntology. At this point, we studied the advantages and disadvantages of the database BADELE.3000 [1, 2] that had been developed according to some MTT lexicographic tools.

BADELE.3000 is a database that contains the 3,000 most frequently used Spanish nouns. The information of each noun includes the definition and the combinatorial possibilities, among other linguistic information. A systematic process for the design of the database was followed; consequently the lexical data are well structured and separated from the applications that might use them. This way, the features of the data model and the subsequent database make them useful for different purposes, such as word sense disambiguation, machine translation and text generation.

As a result, the database contains a minimum of information useful for any type of ontology (because the general vocabulary includes some basic terms transversal to any specific domain) and more than 20,000 combinations. Besides, this resource allows us to infer knowledge potentially useful in real applications.

However, BADELE.3000 is a general-purpose resource with a low utility in commercial exploitations as it does not contain crucial information for real applications. The medium, long-term objective is to enrich this generic linguistic resource by formalizing definitions which can help infer conceptual knowledge

Thus, our aim is twofold: To solve the problems of definition extraction and to add domain knowledge to a general purpose linguistic resource. The process followed is presented in Fig. 2.

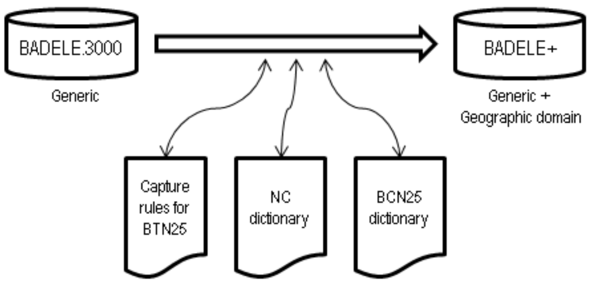

Fig. 2. Definition extraction and systematic lexicalization during BADELE upgrade

As for the lexicographic tools applied to BADELE.3000, we have resorted to three concepts proposed by the Meaning-Text Theory (MTT).

The first one is the *lexical function* (LF) [17: 39-40]: a LF associates a given lexical expression L (such as *sound*), which is the argument or keyword of F, with a set of lexical expressions –the value of F (such as loud, strong, heavy, deafening, etc). – expressing a specific meaning associated with F (for instance, 'intense' for the examples just mentioned which correspond to the LF known as **Magn**).

The second concept is the semantic label: a *semantic label* is the equivalent to the genus in traditional definitions by genus and differentia. For instance, *whale* could be defined as a 'sea mammal that breathes air through a hole at the top of its head and is hunted for meat and for other purposes, as a source of other materials'. The first part of this definition, 'sea mammal', the genus, is known in MTT approach as semantic label; the second part of this definition, the differentia, can be attached to some LFs.

The third concept is the actant [14, 15] and its derivate, the actantial structure. Actants correspond to beings or things that participate in the process expressed by a predicate: MTT approach considers that there is a sort of argument structure in all kinds of predicative words, which means that not only do the verbs have actants but also the adjectives, adverbs and the predicative nouns. The actantial structure reflects the syntactic expression of the actants, as shown in the example of *fleuve* (river) of Dicouèbe, in Table 2:

Table 2. *Fleuve* (river) Dicouèbe Actantial Structure

Nouns	Actantial Structure
Fleuve	[QUI COMMENCE AU lieu X, PASSE PAR LES lieux Z ET SE TERMINE DANS L'étendue d'eau Y]
River	[WHICH STARTS AT THE X place, FLOWS THROUGH THE Z places AND FINISHES AT THE Y area]

Among these three concepts, LFs have proved to be a specially helpful tool for lexicographic works such as the French dictionary *DECFC*[1], the French database *Dicouèbe*[2](developed in Montreal by Polguère and Mel'cuk) and the Spanish database *DiCE*[3] (developed in La Coruña by Alonso Ramos). Fontenelle [5] has also created (semiautomatically) a database but its originality derives from the fact that he takes as source bilingual dictionaries enriched with lexical-semantic information based on LFs. According to Frawley [6] the methodology followed by these resources is ideally suited to the compilation of specialized dictionaries.

6. Problems and solutions

In section 3 two problems have been pointed out when describing the definition extraction process. The low ratio of retrieved definitions can be solved by using linguistic resources (such as domain lexicons, WordNet, etc.) during the label search. So, term variants and semantically equivalent terms could be found and their definitions would be retrieved. The total number of definitions retrieved would increase. However, these definitions would show the same inconsistencies derived from the different granularity and specificity compared to existent ontology definitions. That is, the main problem in the whole process is the linguistic realization of definitions.

Thus, we have mainly focused on three subsidiary problems derived from the above mentioned problem and proposed some solutions according to MTT: (a) semantic labels as a solution for inferring knowledge, (b) lexical functions as a way of providing coherence to definitions and (c) the actancial structure as a tool for developing consistent and complete definitions.

6.1 Definitions and semantic labels

6.1.1 Problem: inconsistencies on the first part of definitions

The first problem that the technical definitions extracted from the knowledge resources used show is the inconsistencies between the name of the label of a group of terms (such as INDUSTRIAL INSTALLATION) and the first part of the definition, i.e. the superordinate of every single term under this label (such as construction, ranch, installation, building), because it differs from one to

[1] Information about the four volumes of this dictionary can be accessed at http://www.olst.umontreal.ca/decfr.html

[2] http://olst.ling.umontreal.ca/dicouebe/

[3] http://www.dicesp.com/

another, as Table 1 shows. The following question could be raised, why is a farm defined as a 'ranch', a corral as a 'construction', a fish farm as an 'installation' and a pigeon loft as a 'building'?

It is clear that the first part of every definition is used in an intuitive way as a quasi-synonym of the genus of the remaining definitions of the group. But in our view it is a false quasi-synonym. As a matter of fact, native Spanish speakers do not use ranch, building, container or installation as synonyms. This raises a second question, why all these words share the label but not the genus of the definition?

6.1.2 Solution: Semantic labels

The last question leads us to propose the use of semantic labels as envisaged in the MTT approach mentioned in section 5. A semantic label would correspond to the genus that matches the superordinate of the definition. Consequently, we propose the use of semantic labels as superordinates in the first part of the definition as a possible solution to avoid inconsistencies. In the examples in Table 2, we have used INDUSTRIAL INSTALLATION as a semantic label of the entire group, so all the definitions begin with the same superordinate. Table 3 shows our proposal.

Table 3. Our proposal for INDUSTRIAL INSTALLATION

Nouns	Definitions
Corral (corral)	***Instalación industrial*** *creada para cobijarse los pastores o para recoger el ganado* (**Industrial installation** created for shepherds or cattle shelter)
Granja (farm)	***Instalación industrial*** *que consta de establos, huerta y casa habitable* (**Industrial installation** with stables, an orchard and a house)
Piscifactoría (fish farm)	***Instalación industrial*** *en la que se crían diversas especies de peces y mariscos con fines comerciales* (**Industrial installation** where fish or seafood are bred for commercial purposes)
Palomar (pigeon loft)	***Instalación industrial*** *en la que se recogen y crían palomas* (**Industrial installation** where pigeons live and are bred)

6.2 Definitions and lexical functions

6.2.1 Problem: embedded definitions

Sometimes simple terms (nouns) or complex terms that share semantic features are defined differently. This inconsistency can be really subtle, as the example in Table 4 shows, based on the definitions of *bancal* (slope) and *ladera abancalada* (terrace slope).

Table 4. *Bancal* and *ladera abancalada* source definitions

Nouns	Definitions
Bancal (terrace)	*Rellano de tierra formado natural o artificialmente que frecuentemente se aprovecha para el cultivo* (Natural or artificial shelf that is frequently used for cultivation)
Ladera abancalada (terrace slope)	*Terreno pendiente con rellanos de tierra, naturales o artificiales, que se aprovecha para algún cultivo* (Natural or artificial terrace that is used for some kind of cultivation)

The two terms share all the semantic features, in other words, the basic characteristics. That would justify why the two definitions are almost equal. However, focusing on the object of the definitions, we find one definition is embedded in the other because a terrace slope is a set of slopes.

6.2.2 Solution: lexical functions

LFs are a powerful tool in order to give coherence to the definitions. Actually, the LF **Mult** could be quite useful in this and other similar cases. This LF expresses the sense 'set of X', where X is an argument that is usually filled by nouns, such as *grape*, or *flower*, as shown in (1):

(1) **Mult**(grape) = bunch of
Mult(flower) = bouquet of, bunch of

This LF can correspond to some lexical units that are not related syntagmatically (as examples above) but paradigmatically (in these cases, the value of the LF is preceded by the symbol //). Consequently, the final version of the entry of *bancal* in our database contains this LF, as shown in (2):

(2) **Mult**(bancal) = //ladera abancalada

The sense **Mult** is usually present at the beginning of definitions. For instance, the first sense of *bunch* is defined in the Oxford Dictionary as a number of things growing together, and the second one as a group of people. If we use the LF **Mult** in order to construct the definition, we should use *set of (bancales)* as the first part of *ladera abancalada*. Our proposal is shown in Table 5.

Table 5. *Bancal* and *ladera abancalada*: our proposal

Nouns	Definitions
Bancal (terrace)	*Rellano de tierra formado natural o artificialmente que frecuentemente se aprovecha para el cultivo* (Natural or artificial shelf that is frequently used for cultivation)

Ladera abancalada (terrace slope)	*Conjunto de bancales en terreno en pendiente* (Set of terraces on a slope)

6.3 Definitions and the actantial structure

6.3.1 Problem: different granularity in definitions

We have found definitions with different granularity in the domain resources used. This difference can derive from the fact that one definition is more explicit than another; or rather, it sometimes implies different entries in each document, such as *bus station* (present at BTN.25 document) and *depot station* (present at BCN.25 document), where *depot* is a hypernym of *bus*, as shown in Table 6.

Table 6. *Bus/depot station* definitions

Nouns	Definitions
Estación de autobuses (bus station) BTN.25	*Lugar donde hacen parada los autobuses para el trasiego de pasajeros y mercancías* Place where buses stop for picking up and dropping off passengers and goods or freight
Estación de transportes (depot station) BCN.25	*Edificio en el que están las oficinas y dependencias de las diferentes empresas encargadas de conducir personas y cosas de un lugar a otro. También alberga el sitio donde habitualmente hacen paradas los vehículos* Building or place where different transport companies that pick up and drop off passengers as well as goods or freight have their offices. It also refers to the place where buses usually have conventional stops

In the second case, we have to decide if the definition should include the sense of ‘offices of the enterprises’, as appears in the second one, or not.

6.3.2 Solution: the actantial structure

The actantial structure is a helpful tool when writing definitions. Actually, if we regard the actantial structure of “bus station”, in Table 7, we can see that each of the three actants is attached to some of the expressions, as shown in Table 8.

Table 7. *Bus station* actantial structure

Actantial structure	Bus Station X where the bus Y picks up the passengers Z

Table 8. Bus station actants and Spanish expressions

Actant	**Spanish expressions attached**
X (*place*)	*Estación de autobuses Méndez Álvaro* (Méndez Álvaro Bus station)
Y (*bus*)	*El autobús llega a la estación a las dos* (the bus arrives at the station at 2.00 o’clock)
Z (*passenger*)	*Juan coge el autobús de las dos* (John takes the bus at 2.00 o’clock)

As the complete sense of *bus station* is expressed by the three actants included in Table 8, we rule out the senses ‘offices and locals of the enterprises’; then we add the semantic label (‘place’) and propose a definition quite close to the first one in Table 6, in which the actantial structure is contained, as shown in Table 9.

Table 9. Our proposal: *Bus station* definition

Nouns	Definitions
Estación de autobuses (bus station) BTN.25	*Local en el que paran los autobuses para la subida y bajada de pasajeros y mercancías* (Place where the buses stop for picking up and dropping off passengers and goods …)

7. Conclusions and Future work

MTT has shown the potential advantages of using a systematic approach for defining terms as it builds on the relations established among the relevant information included in definitions and it allows for some sort of semantic network formed with all the elements present in the definitions. In the process of definition extraction from the domain resources used two problems appeared: semantic inconsistency between different definitions for a concept (term), and very low efficiency of automatic definition search in auxiliary dictionaries. These problems have been described and some solutions have been proposed. Thus, we can conclude that MTT tools are very powerful in order to define or redefine terms. Semantic labels have proved to be consistent as superordinates; LFs are useful when choosing the essential sense of some definitions; and, finally, the actantial structure helps to complete other incomplete definitions.

As future work, our proposal would aim at developing an extraction methodology that could be documented in order to set the steps for automatic extraction. Thus, the manual process above mentioned could be described in detail as the problematic cases are identified and solved so as to identify all the possible activities than can be automatized. To sum up, the final objective is to build a framework which supports definition extraction as automatically as possible. This framework will help experts in definition extraction and systematic lexicalization while adding domain knowledge to a generic lexicographic resource.

8. Acknowledgements

This work has been partially funded by the National Project “GeoBuddies: Anotación semántica colaborativa con dispositivos móviles en el Camino de Santiago” (TSI 2007-65677 C02) and the European Project “NeOn” (FP6-027595).

9. References

[1] Barrios Rodríguez, MA; Bernardos, MS. “*BaDELE.3000*: An implementation of the lexical inheritance principle”. In Gerdes *et al*, (eds.) Meaning-Text Theory 2007. Proceedings of the 3rd International Conference on Meaning-Text Theory. Wiener Slawistischer Almanach. Sonderband, 69. 2007. Pages: 97-106.

[2] Bernardos, MS; Barrios, MA. “Data model for a lexical resource based on lexical functions”. *Research in Computing Science*, vol. 27. 2008.

[3] Borigault, D. Jacquemin, C. & J’Homme, MC (eds.) Recent Advances in Computational Terminology, Amsterdam: John Benjamins, 2001.

[4] Cabré MT. *La terminología*. Barcelona: Empuréis. 1992

[5] Fontenelle T. “Using a Bilingual Dictionary to create Semantic Networks”. Practical lexicography: a reader. Oxford. Oxford University Press. 2008. Pages: 169-190.

[6] Frawley W. “Lexicography and the Philosophy of Science” Dictionaries, 3:18-27. 1980/1981.

[7] Gómez-Pérez A, Ramos JA, Rodríguez-Pascual AF, Vilches-Blázquez LM. ‘The IGN-E case: Integrating through a hidden ontology’, The 13th International Symposium on Spatial Data Handling (SDH 2008), June 23rd - 25th, 2008. Montpellier, France. 2008. Pages: 417-435.

[8] ISO/DIS 704. Terminology work — Principles and Methods. 2008.

[9] ISO 1087-1. Terminology work. Vocabulary: Theory & Application. 2000.

[10] ISO 1087-2. Terminology work. Computer Applications. 2000.

[11] ISO 19110. Geographic Information – Methodology for feature cataloguing. 2005.

[12] Malaisé, V. Zweigenbaum, P. & Bachimont, B. 2005. “Mining defining contexts to help structuring differential ontologies”. *Terminology,* Vol 11-1. 21-54.

[13] Martin, R. 1990. “La definition ‘naturelle’.” In Chaurand, J. & Mazières, F. (eds.) *La definition*. 86-95. Paris: Larousse.

[14] Mel'čuk I. “Actants in semantics and syntax I: Actants in semantics”. *Linguistics*, 42:1, 2004a. Pages: 1-66.

[15] Mel'čuk I. “Actants in semantics and syntax II: Actants in syntax”. *Linguistics*, 42:2, 2004b. Pages: 247-291.

[16] Mel’čuk, I and Polguère, A. 1987. “A formal lexicon in Meaning-Text Theory. Or how to do lexica with words”. Computational linguistics. Nº 13, vol.3, 4, July-December, 1987. Pages: 261-275.

[17] Mel’čuk I. and Wanner, L. “Lexical functions and lexical inheritance for emotion lexemes in German”. In Wanner, L. (ed.), *Lexical functions in lexicography and natural language processing*. Amsterdam/ Philadelphia. John Benjamin. 1996. Pages: 209-278.

[18] Nomenclátor Geográfico Conciso de España (versión 1.0). “Presentación y Especificaciones”. Instituto Geográfico Nacional. Octubre 2006. http://www.idee.es/ApliVisio/ Nomenclator/NGCE.pdf (ICC2005). A Coruña, Spain. 2006.

[19] OGC. OpenGIS Reference Model, Version 0.1.2, OGC Inc. Wayland, MA, USA. 2003.

[20] Pearson J. Terms in Contexts. Amsterdam/ Philadelphia: John Benjamins. 1998.

[21] Rodríguez Pascual AF, García Asensio L. “A fully integrated information system to manage cartographic and geographic data at a 1:25,000 scale”. XXII International Cartographic Conference (ICC2005). A Coruña, Spain. 2005.