

5th International Conference on Corpus Linguistics (CILC2013)

Preparation of a Multilingual Terminology (Spanish, Catalan, English, Arabic and Chinese) for Industrial and Craft Sectors in the *Comunidad Valenciana*: the Case of the Turrón Industry

María Isabel Santamaría-Pérez, José Joaquín Martínez-Egido *

Universidad de Alicante, Campus San Vicente, E-03080 Alicante, Spain

Abstract

The ultimate purpose of this work is to develop a terminological base (Spanish, Catalan, English, Chinese and Arabic) for the industrial and craft sector of turrón, chocolate and cocoa derivatives with the aim of facilitating knowledge communication and transfer between specialists, linguistic mediators and entrepreneurs in the sector and contributing to mitigate the communication problems involved in international commercial transactions. From the methodological point of view, our intention is to carry out a systematic multilingual work (Spanish-Catalan-English-Arabic-Chinese) which can cover the terms used in the selected thematic area applying the methodology required for a task like this (Cabré, 1993 and 1999), which includes several stages, although the present paper focuses on the stage related to the documentation and creation of a corpus with texts that are representative of the field under study.

© 2013 The Authors. Published by Elsevier Ltd.
Selection and peer-review under responsibility of CILC2013.

Keywords: terminology; terminography; corpus; specialized languages

1. Introduction

This article is going to focus on the presentation of a recently initiated terminological project: the preparation of a multilingual terminology database (Spanish, Catalan, English, Arabic and Chinese) for the industrial and craft areas

* Corresponding author. Tel.: +34965909400; fax: +34965909452.
E-mail address: jj.martinez@ua.es

in the Valencian Autonomous Region –including sectors such as turrón and toys,¹ where tradition combines with new technologies. The fact that the terminology corresponding to the turrón industry constitutes a more specific and well-defined field guided our decision to deal with this sector first.

Prior to starting with an overview of the project, it is necessary not only to establish the theoretical premises from which our work starts but also to contextualize and justify the pertinence of such a research initiative. Therefore, the initial explanation of the theoretical foundations on which our work is supported will be followed by a detailed description of our project and, more precisely, of the text collection stage which is being carried out at this very moment.

2. Disciplinary framework: Theoretical premises

Without going into too many theoretical considerations, it seems advisable for us to establish the theoretical premises on which our terminographic application is based. Terminology is understood here as the field of knowledge to which the study, description and collection of the terminological units (TUs) used in the different specialized areas (Cabré 1993) is entrusted, and not only as a mere practical activity intended for the preparation of glossaries or oriented toward the resolution of equivalence problems in translation.

Terminology is a clear example of an interdisciplinary matter which takes its theoretical foundations from the disciplines with which it is interrelated: linguistics, the theory of knowledge and the theory of communication. Its study objects –terms– are also interdisciplinary units and must be analyzed from different perspectives: from a theory of knowledge which can explain how reality is conceptualized; from a theory of communication which can describe and clarify the characteristics of specialized communication; and from a theory of language which, based on discourse, can detect the specialized lexical units and describe them grammatically, semantically and pragmatically.

The starting point for our group is an eclectic and linguistics-based conception of terminology, along the lines of Cabré's Communicative Theory of Terminology (1999) for the development of terminographic applications from the data provided by the specialized discourse in various pragmatic situations. Taking this linguistics-based conception of terminology with a communicative purpose as a reference, concepts cannot be starting units; instead, work is carried out from the terms appearing in texts to the concepts which allow us to recognize one object in the real world (referent). The study objects are 'in vivo' terms; that is to say, terms in use. Hence our decision to choose the texts elaborated by specialists as our starting point.

3. The project: Design, implementation and preparation of a terminological database for the context of turrón in Spanish, Catalan, English, Arabic and Chinese

Present-day society's communicative needs have generated an increase in the demand for standardized terminological vocabularies in different scientific and terminological areas, as a result of which terminology is now a field that raises growing interest. The technological breakthroughs achieved during the last decades have favored a

¹ Contacts with the *Consejo Regulador de las Indicaciones Geográficas Protegidas Jijona y Turrón de Alicante (CRIGPJTA)* [Regulating Board for the Geographical Designations of Origin Jijona and Turrón de Alicante], which has its headquarters in Jijona (Alicante province), started more or less one year ago. Contacts were similarly initiated with the *Asociación española de fabricantes de juguetes (AEFJ)* [Spanish Association of Toy Manufacturers], established in Ibi (also in the Alicante province). Both of them currently have a collaboration *framework agreement* with the IULMA [Spanish initials for Inter-University Institute for Modern Applied Languages] and the University of Alicante.

huge development in traditional specialized areas such as the preparation of turrón or the toy industry, which in turn has multiplied the need to embrace new terms in each of their respective specialized languages.²

From this standpoint, and considering the importance that these sectors have for the economy and development of the Alicante province, for the expansion of its domestic and foreign trade, and for the dissemination of its culture, it seemed to us that the implementation of a terminological project could help facilitate knowledge communication and transfer in the context of international trade, firstly for turrón and later for toys. These terminographic applications will provide the user with the terms utilized in these fields in Spanish along with their equivalents in Catalan, English, Chinese and Arabic.

3.1 Project justification and relevance for the sector

According to the latest socioeconomic data collected in the FEAD 2011 report (FEAD being the Spanish initials for Spanish Federation of Sweets Associations) provided by the *Asociación Española de Fabricantes de Turrónes y Mazapanes* [Spanish Association of Turrón and Marzipan manufacturers], there has been a 7% decrease in the consumption of turrón and marzipan since the beginning of the crisis in 2007. However, last year's turrón sales (2010/11) increased by 3% thanks to a 11% rise in exports with respect to the preceding season according to the data supplied by the CRIGPJTA [Spanish initials for: Regulating Board for the Geographical Designations of Origin Jijona and Turrón de Alicante] which are offered in the charts below.

Table 1. Market size in 2010-11 (volume and value)

	Tm 2010	%	Tm 2011	%
	Tons 2010	10/09	Tons 2011	11/10
Domestic Market	27,782	1.4	28,851	3.9
Exports	3,476	12.0	3,563	2.5
Total	31,258	2.5	32,414	3.7

Table 2. Market size in 2010-11 (economic profit)

	000€	%	000€	%
	2010	10/09	2011	11/10
Domestic Market	239,924	2.8	240,596	3.9
Exports	31,461	10	35,393	12.5
Total	271,385	1.5	275,989	1.7

The data provided by the Spanish Association of Turrón and Marzipan Manufacturers (TUMA) indicate that 85% of turrón and marzipan sales are concentrated during the Christmas holidays. An attempt is being made to make the product available to the consumer at sales outlets all year round in order to combat seasonality, and numerous products and formats are additionally being adapted to the consumer's current demands. This means, for example, making a modern line of small-sized, appetizer- or *snack*-like products.

² In fact, work in the last few years has been oriented toward vocabularies of this kind such as, for example, the group Olivaterm at the University of Jaén led by Professor Mercedes Roldán [www.olivater.com]; the research group Semainein led by Professor Miguel Casas [http://www2.uca.es/grup-invest/semainein/], director of the Doctoral Thesis written by María García Antuña and entitled *Consideraciones metodológicas sobre la elaboración de un Léxico específico de la piel* [Methodological considerations about the preparation of a specific Lexicon for leather] (2011) or the research project "*TXTCeram: Extracción semiautomática y análisis conceptual formal de términos de la cerámica a partir de un corpus electrónico. Su eficacia y utilidad en la mediación lingüística* [TXTCeram: Semi-automatic extraction and formal conceptual analysis of ceramics terms from an electronic corpus. Its effectiveness and usefulness in linguistic mediation]" (GV05/260), Jaume I University of Castellón, Valencian Regional Government led by Professor Amparo Alcina, etc.

A second strategy, before the stagnation experienced by turrón and marzipán sales in recent years, consists in orienting this market toward foreign countries. It is thought that one of the main challenges for this industry lies in promoting its products, boosting the image of tradition and quality outside Spain but without relinquishing innovation, since novelties do not usually affect the preparation of turrón and marzipán but other processes, such as packaging.

Turrón and marzipán are usually welcomed in foreign markets –in fact they are the best-known Spanish sweet product in some countries. Thus, exports approximately account for 9.3% of the total sales in this sector. The main destination of exports is the European Union (especially Germany, United Kingdom and France), with the United States and Latin America as the most important export destinations beyond the European borders –although the sector also exports to such diverse countries as Russia, Israel, Jordan, and even China and Japan. The Chinese market was opened for the first time in 2010, which means an important new market opportunity for a sector that, despite being consolidated, is also suffering the consequences of the current economic crisis.

This sector also enjoys a clearly positive balance in terms of foreign trade because its products are not made elsewhere in the world. A number of similar products –such as Italian torrone, French nougat or Danish marzipán– do exist in other countries, but none of them can boast the quality or quantity of raw materials (especially almonds) used by Spanish manufacturers.

The behavior of exports has been negative in recent years due to the strength of the euro regarding the dollar and to the crisis which has affected some countries –particularly Latin American ones– which had traditionally been great consumers of these products. Nevertheless, the sector continues to make a great effort to export, as exportation undoubtedly represents a possible path for growth in this sector.

This is the reason why the main target for turrón manufacturers is to expand abroad, for which they need to have instruments that can facilitate the commercialization and internationalization of their sector, such as our multilingual terminological application, which is currently being developed.

3.2. Methodology and working plan

The terminological project is divided into various stages which are necessary for the preparation of any terminological work.

1. Work definition and delimitation
2. Previous preparation and documentation
3. Conceptual system structuring
4. Text corpus preparation
5. Terminology preparation
6. Revision
7. Equivalence establishment
8. Treatment and resolution of problematic cases
9. Edition

A very concise summary of what has been carried out so far comes next. With regard to the first stage, definition and delimitation of tasks: aims, languages and addressees, once the characteristics as well as the current situation of the professional field in question have been established, it becomes necessary to define the characteristics of the terminographic application about to be developed.

Our overall *aim* is to compile, systematize and present the terminology corresponding to the thematic turrón area with a view to facilitate knowledge communication and transfer between specialists, linguistic mediators and entrepreneurs in the sector and help mitigate the communication-related problems which arise in international commercial transactions. It will consequently be a descriptive application (it collects the units appearing in texts) of a semi-specialized nature –given the characteristics of the field analysed as well as the diversity of users– and a useful application for the consultation, correction and drawing-up of texts, for translation and interpreting, and even for the terminological standardization of this sector.

Our work is intended for a wide and varied range of *users*:

- Entrepreneurs in the sector
- The average Spanish or foreign consumer
- Translators and interpreters
- Agri-food marketing professionals, etc.

Nevertheless, a selection of texts covering all the specialized sub-areas will only be possible after carrying out a conceptual delimitation of the turrón sector which links and organizes all the ramifications which are inherent to this area, so that the reality of the sector under study can be schematically reflected (see Fig. 1 below).

As might be expected, the degree of knowledge about the defined thematic area will differ considerably across these groups, which is why one of our challenges is to establish a level of specialized concept transmission that can successfully meet the interests of the different user types.

The selection of languages for the project (Spanish, Catalan, English, Arabic and Chinese) responds to the specific communicative needs identified among the addressees of this turrón glossary. In this sense, our work has as its first aim to serve the Spanish business and scientific community, providing it with appropriate and precise terminology in Spanish. This initiative additionally targets the international community; hence the presence of English as the lingua franca of knowledge and communication. However, the novelty in this project lies in the inclusion of two languages which have become essential in the present-day market, Arabic and Chinese –because of the extraordinary expansion prospects that these new markets represent nowadays. Catalan is equally included for social, political and economic reasons. Any democratic administration and, by extension, any professional organization, must be able to address its citizens in their own language and ensure that it becomes a vehicle for communication in every social and professional context.

With regard to the type of information that the aforesaid application is going to contain, users will be able to access it through a telematic consultation on the Internet or from a tablet or smartphone that will provide the terms along with their respective equivalents in other languages, complemented by a short definition, the grammatical category, the usage context, the reference sources, etc.

As far as the second stage of the project is concerned, the documentation available about this subject matter was gathered in order to prepare the scenario so that the actual applied work could be implemented more smoothly. On the one hand, this documentation is lexical; in other words, our search for information focused on precedents or existing dictionaries that could prove helpful to us as basic consultation or verification materials (reference or lexicographic corpus). Regarding the terminographic resources already available in the selected languages, only one glossary currently exists which collects the basic terms for the professional field of turrón. It is hosted on the web page of the Consejo Regulador de las Indicaciones Geográficas Protegidas Jijona y Turrón de Alicante [<http://www.jijona.com/app/seccion/consejo/acercade.aspx>].

There are also some lists of terms specific to the sector collected in papers dedicated to this topic. An example can be found in the study carried out by Manuel López Espí, *Turrónes diversos, mazapanes y derivados* [A variety of turrón, marzipans and derivatives], Asociación de Fabricantes de Turrón, chocolates y derivados de la Comunidad valenciana [Valencian Region Association of Manufacturers of Turrón, chocolate and cocoa derivatives] and the Cámara de Comercio de Alicante [Alicante Chamber of Commerce].

These glossaries include the entry in Spanish and a brief encyclopedic-type definition –in Spanish too. Hence the relevance of our paper, since no systematic work has been carried out so far from the current theoretical foundations of terminology that can be categorized as representative of the specialized discourse in this specific sector; nor has this field ever been approached from a multilingual perspective.

It is also worth highlighting that the documentation needed to develop our research has a textual nature; in other words, it is necessary for us to carry out some preliminary corpus collection work in the texts, which not only will allow us to extract the terms and study them *in vivo* but will also help us when it comes to illustrating their usage, their definitions and other pieces of information of conceptual, pragmatic or linguistic interest –the stage in which our work finds itself right now.

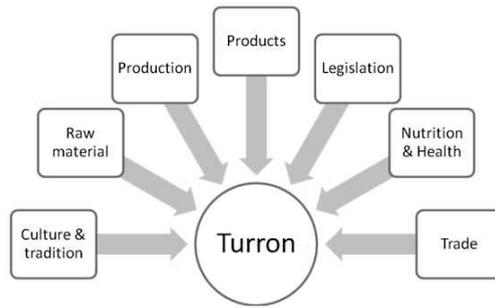


Fig. 1. Conceptual structure or field tree of the turrón sector

4. Preparation of a text corpus for turrón

4.1 Corpus creation criteria

Corpus linguistics is understood here as a set of tools and methods for empirical linguistic analysis based on the utilization of real linguistic usage samples coming from speakers of a language or one of its varieties. Terminology and corpora constitute the documentary basis for the preparation of terminological vocabularies in this particular field, where corpora have provided especially useful tools as they not only constitute an ensemble of texts containing potential candidates to become terms but also provide information about the context together with linguistic information about sources, word combinations, etc. as a result of their design.

Our starting point is the classical definition of corpus proposed by Sinclair (1991: 171): a collection of natural language texts selected to characterize the state of a language or one of its varieties. Since each corpus is created for a specific research purpose (Leech 1992: 116), texts must be collected using a strict selection criterion, for which a number of parameters need to be considered.

Firstly, the present project requires a specialized corpus because the research is focused on a specific field of study: the scientific and socioeconomic aspects related to the turrón industry. Thus, the documentation on which our work is based revolves around the turrón industrial sector, for which several types of texts (journalistic, informative or oriented to general interest, research-based, etc.) are used. This means that the corpus format is predominantly written, since most of the original documents come from the written language: newspapers, monographs, guides, brochures, etc. However, the fact that this is a traditional, autochthonous sector implies that there will be certain terms, particularly those referring to production and raw materials, which have to be checked by means of *in situ* research and compilation –with the additional peculiarity that many such terms are in Valencian, a dialectal variety of Catalan in this geographical area.

Secondly, our purpose is not to examine the evolution of this field but its current state, which is why the documentary sources need to be based on contemporary usage –in this sense, the corpus will be synchronous.

Thirdly, it has already been mentioned that each terminographic entry should offer a definition of the term, its real context of usage; the equivalent in English, the equivalents in Arabic, Chinese and Catalan, the references for the definition and context, as well as the synonyms and phraseologies where appropriate. Therefore, the corpus is going to be multilingual, including examples in Spanish, English, Arabic, Chinese and Catalan.

Finally, with regard to the degree of specialization in the texts, it must be remembered that this is a professional register used within a restricted context by specialized users (entrepreneurs, producers, manufacturers); in other words, it is a technical or semi-technical language. However, many of the texts are addressed to the general public, in which case we would be talking about informative, general-interest texts.

The next step after determining the corpus parameters is the collection of terms from the texts, which must be carried out in electronic format. Although some of the texts included in the corpus are in this format, those which are available only in paper format will have to be scanned and later transferred to electronic format using OCR software. From a total of 350 documents, 315 were in digital format and 35 in paper.

4.2 Stages in the creation of the turrón corpus

The first stage in the preparation of such a project implies a selection of texts based on the following variables: source; area; text specialization level; and type of publication. The main goal is to ensure that the study corpus is balanced with respect to these criteria.

This process –which becomes essential for the corpus to be reliable in the search for terms and contexts– is gradual and takes shape as new documents are added. By way of example, Table 1 shows some corpus documents along with the characteristics that were considered in order to incorporate them into the study.

Table 3. Corpus compilation variables

Text	Language	Level	Area	Source	Words
1	Spanish	high	RS	Jijona.com	4,318
2	Spanish	semi	PROD	Alicante Chamber of Commerce	9,286
3	Spanish	low	PRTO	http://jijona.com/producto/el-turrón/turrones-con-denominacion/	631
4	English	low	PRTO	http://www.exploretaca.com/eng/article.html?id=2089	1,157
5	English	low	PRTO	http://www.ehow.com/about_4577660_what-is-turrón.html	305

As can be seen, Table 2 collects different properties, some of which coincide across the texts, with the exception of sources. At present, most of the texts are in Spanish and some in English. Many texts belong to the informative type appearing in magazines, press releases, etc.; hence their reduced length, as opposed to those which have a more technical or professional nature and contain a much higher number of pages. What matters during this compilation stage is to know the overall data before adding any more documents so that the corpus can remain balanced. The ‘area’ column shows the different sub-contexts into which the turrón industry has been divided (with initials or abbreviations in Spanish: health regulation (RS); culture and tradition (CUL); raw materials (RM); production (PROD); product (PRTO); legislation (LEG); nutrition (NUT); and commercialization (COM)).

The selection of texts is followed by their digitization, the second stage within the corpus preparation process. The textual variables explained above are present in the coding system used to identify texts. Codes are represented in Table 3 below.

Table 4. Document identification codes

Title	Language	Specialization Level	Area	Source	Numbers
<i>Monograph on Turrón and Marzipans</i>	Sp	a	COM	MS	14
<i>Technical Health Regulation on turrón and marzipans</i>	Sp	a	LEG	CR	1
<i>Turrón, a healthy sweet</i>	Sp	c	NUT	FIT	2
<i>What is turrón?</i>	Eng	c	PRTO	HOW	2

It can be easily seen that each text is described according to four factors represented through codes, the sum of which makes up the file number for that document within the corpus. Each code is described using a fixed number of characters in order to ensure that the names of all documents are concise, homogeneous, comparable and easily identifiable after this stage. The first item is the language and includes the variables Spanish (SP), Catalan (CAT), English (ENG), Arabic (AR), and Chinese (CH). The second descriptor is the level of specialization, for which three levels are contemplated: high (A), semi (B) and low (C) –corresponding to the three degrees of specialization distinguished in our texts: technical, semi-technical and informative. The third field refers to the identification of text areas, with the six possible alternatives established within the field of turrón. In fourth place, a registration of the document source is carried out where the code introduction follows the addition of the text. And finally, the area

serves as the basis to number the texts and distinguish one from another. For example, the first text is number 14 within the commercialization area.

Once all the properties have been written down, the last stage within the corpus compilation process consists in renaming the document from the fixed codes. The texts in Table 2 have the following names: ESP-a-COM-MS-14; ESP-a-LEG-CR-1; ESP-c-NUT-FIT-2 and ENG-c-PRTO-HOW-2.

The most outstanding advantage provided by these codes is that, once they have been registered in different Microsoft Excel columns, it is possible to arrange the corpus documents automatically according to the variable chosen. For instance, if one wishes to examine all the texts belonging to the nutrition and health area, it suffices to order the corresponding column on the spreadsheet and the software program will organize the codes, placing all those which have the NUT label together.

After completing the corpus preparation, the latter will go through the terminological work station Terminus http://melot.upf.edu/Terminus2009/index_es.html for the automatic terminological extraction and the edition of terminological cards. Once the corpus has been shaped, this program allows us to draw up a list with the candidates to become terms, which will help us during the process of selection for the units that are going to form part of our vocabulary.

5. Conclusions

It has been shown here that corpora are an essential tool in the preparation of any terminological work, as they become a database for fundamental textual data, regarding both terminological extraction and the recovery of usage contexts.

The present paper has as its aim to establish the most suitable design for the compilation of corpora created for a terminographic purpose and specific to certain professional and academic fields. In order to achieve that aim, our treatment focused on the aspects directly related to the design of corpora for specific purposes, which were exemplified using the corpus created within the framework of our research. Our goal was to serve as a model or guide for the design and construction of a corpus in a specialized area. With this aim in mind, our paper starts with a brief explanation of the theoretical framework on which our work is based, followed by a definition of the characteristics of our corpus, which is meant to become a conceptually and pragmatically representative and balanced sample of the written and oral Spanish, Catalan and English used in the turron industrial sector. In short, it is a true collection of texts from the field explored in electronic format which has terminological research as its ultimate aim. In our opinion, the compilation of a corpus deserves to have a reflection period during which its design can be shaped and clarified. Thus, the corpus must be considered by the terminographer in different stages, as it is our raw material and the results and quality of the terminological application largely depend on this resource. All the decisions eventually adopted must be directly linked to the specific purpose of the corpus and refer, amongst other things, to considerations such as the text production channel, the final size, the textual distribution, the content, the size of samples, etc.

Thus, taking into account the importance of corpus linguistics with respect to terminology, the corpus created for this project is intended to become a representative corpus for the selected area. In accordance with the established parameters or criteria, this is a written, specialized, synchronous, closed and native corpus. Nevertheless, apart from what has been said above, some corpus characteristics cannot be known until the corpus has been completely finished; for instance, its volume, because it is impossible to foresee how many units it will contain.

References

- Agbago, A., Barrière, C. (2005). Corpus construction for terminology. *Proceedings from the Corpus Linguistics Conference Series* 1:1. In: <http://iit-iti.nrc-cnrc.gc.ca/iit-publications-iti/docs/NRC-48516.pdf>
- Arntz, R., Pitch, H. (1995). *Introducción a la Terminología*. Madrid: Fundación Sánchez Ruipérez.
- Cabré, M. T. (1993). *La terminología. Teoría, metodología, aplicaciones*, Barcelona: Antàrtida-Empúries.
- Cabré, M. T. (1999). *La terminología. Representación y comunicación*. Barcelona: IULA.
- Chung, t. (2003). "A corpus comparison approach for terminology extraction". *Terminology* 9: 2
- Díaz Negrillo, A., Fernández Domínguez, J. (2010). Corpus terminológicos. In Roldán Vendrell, M. (Ed.): *Bases para la terminología multilingüe del aceite de oliva* (pp. 37-50). Granada, Comares.

- Gómez, A., Vargas, Ch. (2004). Aspectos metodológicos para la elaboración de diccionarios especializados bilingües destinados al traductor. In González, L., Hernández, P. (Coords.). *Las palabras del traductor. Actas del II Congreso El español, lengua de traducción*. 20-21 de mayo de 2004 (pp. 365-398). Toledo, Brussels: Esletra.
- L'Homme, M. C. (2004). *La terminologie: principes et techniques*. Montreal: Les Presses de L'Université de Montréal.
- Leech, G. (1992). "Corpora and theories of linguistic performance". In Svartvik, J. (ed.). *Directions in corpus linguistics* (pp. 105-122) Berlin: Mouton de Gruyter.
- Leech, G. (1991). The state of the art in corpus linguistics. Aijmer K., Altenberg B. (Eds.). *English Corpus Linguistics: Studies in Honour of Jan Svartvik* (pp. 8-29). London: Longman.
- Maia, B. (2002). Corpora for terminology extraction- the differing perspectives and objectives of researchers, teachers and language services providers. In *Language resources for translation work and research, LREC 2002 Workshop Proceedings*. (pp. 25-28).
- Meyer, I., Mackintosh, K. (1996). The Corpus from a Terminographer's Viewpoint. *International Journal of Corpus Linguistics*, vol. 1(2), 257-285.
- Pearson, J. (1998). *Terms in Context*. Amsterdam/Philadelphia: John Benjamins.
- Peñas, A., Verdejo, F., Gonzalo, J. (2001). Corpus-based terminology extraction applied to information acces. In Rayson, P., Wilson, A., McEnery, T., Hardie, A., Khoja, S. (Eds.). *Proceedings of the Corpus Linguistics 2001 conference* (pp. 458-465).
- Santamaría Pérez, M.I. (2006). *La terminología: definición, funciones y aplicaciones*. Available at: <http://goo.gl/HsOop>.
- Santamaría Pérez, M.I. (2010). *La terminología*. In: <http://rua.ua.es>
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Vargas Sierra, Ch. (2006). Diseño de un corpus especializado con fines terminográficos: el corpus de la piedra natural. *Debate Terminológico*, 2 (7/2006). Paris: RITERM (Red Iberoamericana de Terminología). Available at: <http://goo.gl/bYYfE>.
- Vargas, Ch. (2005). A pragmatic model of text classification for the compilation of special-purpose corpora. In Mateo, J., Yus, F. (eds.). *Thistles. A homage to Brian Hughes Essays in Memoriam*, vol. 2 (pp. 295-315). Available at: <http://goo.gl/yFjx3>.