Lingue e Linguaggi
Lingue Linguaggi 29 (2019), 581-597
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v29p581
http://siba-ese.unisalento.it, © 2019 Università del Salento

METHODOLOGICAL PROPOSAL TO BUILD A CORPUS-BASED ONTOLOGY IN TERMINOLOGY

ISABEL DURÁN-MUÑOZ
UNIVERSIDAD DE CÓRDOBA

Abstract – Corpora are an indispensable resource to improve quality both in the linguistic and conceptual dimension of terminological projects. However, while there is complete agreement that specialised corpora are vital in the linguistic dimension of any terminological project (e.g. to select real contextual examples), there are three different approaches with regard to the conceptual dimension and not all of them employ corpora in their projects. In an attempt to shed some light on the advantages that corpora bring to the representation of specialised knowledge in terminology, this research follows the ontoterminography methodology (Durán-Muñoz 2012) to propose the building of a corpus-based ontology within a terminological project, in particular a specialised resource about an adventure activity (canyoning) in English. More specifically, it describes the different steps that are required to create such an ontology, from the analysis of the specialised domain and the compilation of the corpus to the representation of the specialised knowledge in the form of a corpus-based ontology.

Keywords: specialised discourse; specialised corpus; corpus-based ontology; ontoterminography; terminology.

1. Introduction

Corpus Linguistics has been one of the linguistic disciplines that has been most influential in Terminology. So much so that nowadays the idea of employing specialised corpora in terminological projects is fully accepted in the community of terminologists, as corpus has become an indispensable resource for any work of this nature (Dash 2000, p. 27). The use of corpora has, therefore, become an essential tool for current terminological projects, including production of specialised dictionaries and databases, giving rise to what Leech (1992, p. 106) considered "a new way of thinking about language". Corpora can be used at any stage of a terminological project, from the selection of contexts to be included in the terminological resources to the study of concepts and the conceptual relations among them

to create knowledge representation and become familiar with the specific subject of their work, what Cabré Castellví (1999, p. 144) calls *cognitive competence*.

While the usage of specialised corpora in the linguistic dimension in terminology is now unquestionable, no clear consensus exists over their use in the conceptual dimension. There are some methodologies that employ specialised corpora in order to analyse concepts and conceptual relations in a specialised domain by means of the study of their designations (or terms) in context (see L'Homme 2008). In other words, these methodologies follow a bottom-up approach and consider the text (or corpus) as the starting point to reach the specialised knowledge. On the other hand, we also find methodologies that follow a top-down approach, in accordance with theoretical principles of the traditional terminology (see ISO 704 2000, ISO 1087-1 2000; Roche *et al*. 2009), and analyse specialised domains by means of the initial identification of concepts by domain experts. Hence, they follow an onomasiological approach and analyse specialised domains from concept to term. Finally, there is another type of methodology that follows a middle-out approach in order to examine the specialised domain, which combines both views: it starts from the corpus in order to extract frequent terms and gain knowledge about a particular domain, but, at the same time, it requires the assistance of domain experts to build the conceptual representation of the specialised domain (Temmerman, Kerreman 2003; Faber *et al*. 2005; Durán-Muñoz 2012).

This research revolves around this conceptual dimension and aims to contribute to this middle-out approach, which combines specialised corpora and expert knowledge, by illustrating the creation of a corpus-based ontology for a particular specialised domain: the adventure activity of canyoning. To do so, it will follow the ontoterminography methodology (Durán-Muñoz 2012), a middle-out approach inspired by Frame-based Terminology (Faber *et al*. 2005) and the theoretical principles of the Sociocognitive Theory of Terminology (Temmerman 2000).

The remaining part of the paper is structured as follows: Section 2 outlines the concept of ontology and its fruitful relationship with Terminology; Section 3 describes the methodology followed and the different steps: 1. Corpus compilation, 2. Extraction of candidate terms and conceptual relations; 3. Expert consultation, and 4. Creation of corpus-based ontology. Finally, Section 4 presents some concluding remarks and future work.

2. Ontologies and terminologies, a fruitful relation

Ontologies have aroused the interest of researchers in different fields, such as Terminology, Artificial Intelligence or Computational Linguistics, and ontology-oriented applications are commonly found as part of information systems, databases, natural language processing systems, knowledge based systems, etc. As Roche (2007, p. 47) states,

> Such a success is mainly due to what an ontology promises; it means a way of capturing and representing a shared understanding of a domain that can be understood and used by humans as well as by software.

The term *ontology* has its origins in the field of philosophy and it is considered a branch of metaphysics that deals with the nature and the organisation of reality. However, it has gained special relevance in the field of Artificial Intelligence, Knowledge Engineering, and Computational Linguistics in recent decades as a means of modelling knowledge, and it has acquired a more pragmatic and applied meaning derived from its original meaning.

In this computing-related field, the term *ontology* has numerous definitions, but the most accepted and cited by authors is the one proposed by Gruber (1993, p. 199): "An ontology is a formal, explicit specification of a shared conceptualisation". This definition, already considered as standard, includes the most relevant aspects of an ontology and its basic principles, namely:

- The term *conceptualisation* refers to an abstract model of the domain, or some phenomena of the world that it represents, which is intended to identify the most relevant concepts.
- The term *explicit* indicates that the concepts that constitute an ontology are defined in an unequivocal way.
- The adjective *formal* refers to the fact that the ontology must be expressed in some form of computer readable language by means of an identical formalism, so that it can be reused and understood by any machine regardless of the place, the platform or the language of the computer system that uses it.
- *Shared* reflects the notion that an ontology represents knowledge shared and accepted by the group or epistemological community to which it refers, and not only that of an individual.

As observed, *ontology*, also named *conceptualisation*, refers to a set of hierarchically organised concepts represented in a computer system whose purpose is to support various applications that require specific knowledge

on the subject that ontology represents (Moreno Ortiz 2004, p. 31). In the same vein, Roche (2007, p. 47) summarises: "an ontology is a shared description of concepts and relationships of a domain expressed in a computer readable language."

This latter meaning of the term *ontology* is gaining particular significance in the field of Terminology as it can be considered a bridge between knowledge representation in the mind and language. In Faber *et al*.'s (2005) words:

> Terminology is not only a matter of terms and term entries that endeavour, no matter how imperfectly, to represent slices of objective reality. In this sense, the representation of a specialized field should be more than a list or even a configuration of objects linguistically translated into either simple or compound nominal forms. It is necessary to situate concepts in a particular setting and within the context of dynamic processes that define and describe the principal event in the specialized field in question.

We are witnessing a great increase in the use of ontologies[1] to carry out research as well as to produce terminological resources, such as specialised dictionaries or databases, in an attempt to organise specialised knowledge (concepts and relations among them) in a clear and systematic way by means of computer applications. An example of this are the numerous conferences that focus on this line of research, to name a few: *Terminology and Artificial Intelligence* (TIA), *Terminology and Knowledge Engineering* (TKE), *ToTH: Terminology & Ontology: Theories and applications*, among others, which confirm the approach and cooperation between Terminology and Knowledge Engineering.

This relationship has led to the fruitful development of ontologies and computer applications for the management of specialised knowledge in the field of Terminology, which has facilitated a great advance in ontoterminography. In this context, specialised corpora are crucial to providing both conceptual and linguistic information and are currently the starting point of any terminological project aimed at producing specialised resources (dictionaries or databases), as we will see in the next section.

[1] In modern terminology we can see that, instead of the term *ontology*, other terms are employed that show similar features: knowledge representation, semantic categorisation or semantic frame.

3. Methodology: building a corpus-based ontology

As stated above, this research follows the ontoterminography methodology (Durán-Muñoz 2012) to propose the production of a corpus-based ontology in a terminological project. This methodology presents a middle-out approach and is inspired by the theoretical principles of the Sociocognitive Theory of Terminology (2000) and the Frame-based Terminology (*Faber et al*. 2005), a theory which employs semantic frames[2] as a basis for structuring expertise and creating separate representations of the working language. By way of an example, this paper depicts the different steps that are required to create such an ontology, from the compilation of the corpus to the representation of the specialised knowledge in a specialised resource about the adventure activity canyoning in English.

The ontoterminography methodology divides the terminological project in six main phases: 1) Design of the project and analysis of the specialised domain; 2) Compilation of the specialised corpus; 3) Ontoterminography management; 4) Elaboration of the ontoterminography database; 5) Validation, and 6) Edition of the terminological product (database, dictionary, etc.). The order established here does not imply that a strict sequence must be followed, as some tasks can overlap and occur simultaneously.

In view of the focus of the paper, only the first three steps will be considered, paying particular attention to corpus compilation. A specific adventure activity, canyoning, will serve to illustrate the building of such a corpus-based ontology.

3.1. Design of the project and analysis of the domain

This first step refers to the preparatory work of any terminological project and involves a series of decisions that serve as the basis for the entire process. In this initial phase, it is necessary to establish the pragma-linguistic variables that characterise the final resource and that need to be maintained throughout the entire process, always keeping in mind the

[2] Fillmore (1968) was one of the first to introduce the concept of *frame* in linguistics, considering it as a linguistic tool to represent extralinguistic events. Atkins and Rundell (2008: 145) explicitly define what a semantic frame is: "A semantic frame is a schematic representation of a situation type (e.g. speaking, eating, judging, moving, comparing, etc. - activities and situation which make up our everyday life) together with a list of the typical participations, props, and concepts that are to be found in such a situation; these are the semantic roles, or 'frame elements' (FEs)."

objectives of the research. These variables refer to the topic, the languages of the resource, the target users, the function to be covered, as well as the human and technical resources required. Besides the team working on the project, it is crucial to select domain experts in the working languages that assist terminologists during the whole process, but particularly during the conceptual representation of the domain (top-down approach).

Once the project has been designed, the analysis of the domain comes next. This consists of the study of the domain in a broad way and from different perspectives: socio-economic, pragma-linguistic and even a translation approach (if multilingual), so that terminologists can acquire broad knowledge of the domain and of the possible needs and difficulties of the project. To fulfil this stage, terminologists should consult domain experts[3], specialised journals, reports on the situation, relevant entities in the domain, existing legislation (if applicable), etc. It is also important to carry out an assessment of the terminological resources available at that time for that particular specialised domain. The analysis of available resources will allow terminologists to undertake an *in vitro* study of the terminology and learn from the definitions and other information included in those resources.

At this stage, terminologists have acquired a broad knowledge of the domain at stake and are capable of proposing a preliminary ontology, which can be enlarged and specified in subsequent steps.

In this particular research paper, adventure tourism was analysed from different approaches: the adventure activities provided and their features, the diverse textual genres, socio-economic factors and pragma-linguistic features[4]. Moreover, a set of domain experts were contacted and, thanks to their help, the following preliminary ontology was proposed (Figure 1):

[3] The author is grateful to the experts in physical and sport activities and active tourism that contributed to this research by providing their guidance and assistance in the development of this ontology.

[4] Due to space constraints, the results of this analysis are not presented here, but please see Durán-Muñoz (2014) for further information.

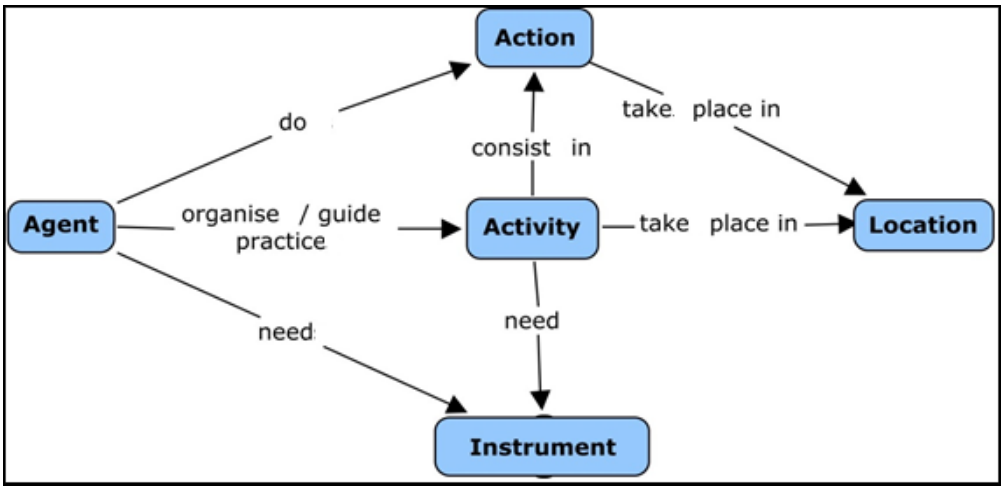

Figure 1
Preliminary ontology on adventure tourism.

This conceptual representation displays the prototypical situation in adventure tourism in a simplified way, indicating that all adventure activities (including canyoning) are organised in the same way. By means of this preliminary ontology, we can observe that, in order to understand and to define any concept within this frame, the activation of the other four categories that are part of this frame is also required. As such, an <agent>, i.e., a person who performs, guides or practises an activity is defined according to the activity and the actions which are performed along with the instruments employed for the activity; or for <activity>, it is necessary to explain the actions performed, the location in which it takes places and the instruments that are required. In short, the position of a concept in a particular context and with regard to other concepts is specified and, as a result, the ambiguity that may arise at language level disappears.

3.2. Corpus compilation

This step refers to the compilation of the specialised corpus, which provide conceptual and linguistic information to terminologists throughout the whole project. Consequently, the quality of this resource is of great importance to achieve successful results.

By way of an example, the semi-automatic compilation of a corpus about a particular adventure activity, canyoning, is provided in this subsection. This semi-automatic compilation was performed via *WebBootCat* tool (Baroni *et al*. 2006), a web-based automatic building tool

integrated in the Sketch Engine online system[5] (Kilgariff *et al*., 2004). Even though the corpus was automatically compiled by the system, the result was carefully checked and manually revised in order to avoid irrelevant or inappropriate data that could bias the final analysis.

Following Baroni *et al*.'s protocol (2006), the compilation was divided into the following steps:

- Step 1. Seed words, i.e. keywords for the domain of interest, were selected. In this case, the seeds concerned the adventure activity canyoning (see Table 1) and were chosen after the analysis carried out in the previous step.

adventure activity	gorge
adventure	rope
outdoor activity	ravine
canyoning	waterfall
instructor	river
rappel	canyon
adventure	descend

Table 1
Most frequent terms in the adventure activity "canyoning", employed as seeds.

- Step 2. The seed words were used by the system to create "tuples" for the queries, i.e. they are randomly combined into different multi-word sets automatically by the programme (e.g. "canyoning river adventure").
- Step 3. The system generated a list of potentially relevant webpages, which could additionally be checked, deleted or confirmed according to the project needs, before building the corpus.
- Step 4. The selected webpages were automatically downloaded and put together in a single file (the corpus). They were also deduplicated and cleaned, and spam text or non-text was eliminated to obtain high-quality text material. At this stage, the corpus was available to be downloaded or browsed through Sketch Engine.

The same process can be repeated again and again to extend the corpus by means of the *Keywords/terms* option, which uses the keywords extracted from the compiled corpus as seed words to launch the search again.

[5] URL: https://www.sketchengine.eu/

The greatest advantage that this automatic corpus builder brings to corpus compilation is that very large corpora can be compiled quickly and effortlessly, compared to traditional time-consuming manual compilation. However, a thorough and careful manual revision of several aspects of the process is required to refine the searches and to guarantee successful results, which increases the total time of corpus building. Consequently, besides the clean-up processing that the tool carried out automatically, the proposed URLs in Step 3 were carefully revised and refined manually before the corpus building took place in order to discard any irrelevant URLs and to select only those that were appropriate for the project. As a matter of example, URLs discarded in this study were those of *Wikipedia*, *Amazon*, social networks (e.g. *Facebook*, *Pinterest*), *Youtube*, *Scribd*, *eBay*, etc.; those that were not originally written in English; and those that were not promotional texts, such as articles, blogs, etc., since the chosen textual genre for the corpus was promotional texts published by public and private companies and administration. After this manual revision, 30% of the URLs proposed by the system were discarded. Once revised and compiled, the English CANYON corpus, containing 925,422 words, was ready to be exploited in the next stages.

3.3. Term extraction

Once the compilation of the specialised corpus was complete, a semi-automatic term extraction was performed in *Sketch Engine* with the aim of extracting the most frequent units from the corpus. Figure 2 displays the most frequent single and multi-word units proposed by Sketch Engine, ordered according to their keyness.[6]

[6] The keyness is automatically calculated by Sketch Engine using the British National Corpus (BNC) as reference corpus.

Single-word		Score	F	RefF
rappel	W	1,487.04	2,370	10,127
canyoning	W	748.83	900	2,067
Canyoning	W	508.92	581	840
abseil	W	436.26	619	6,560
canyon	W	353.54	2,219	106,668
Lip	W	305.35	596	17,565
belay	W	287.18	472	11,217
Waterfall	W	272.32	499	15,112
Rappelling	W	268.20	458	12,543
waterfall	W	259.62	2,264	157,051
canyoneering	W	237.03	270	839
Sheer	W	236.89	318	5,028
Descent	W	222.48	328	7,752
Canyoneering	W	211.88	235	232

Multi-word		Score	F	RefF
subjective opinion	W	154.66	177	5
outdoor activity	W	85.46	104	13
outdoor adventure	W	78.62	91	7
activity instructor	W	74.43	81	0
zip line	W	68.99	75	0
white water	W	59.87	129	109
rappel device	W	59.02	64	0
full day	W	57.79	129	117
gorge walking	W	49.95	54	0
canyoning adventure	W	44.52	48	0
outdoor education	W	41.71	49	10
team building	W	41.45	56	28
outdoor recreation	W	41.41	47	6
adventure activity	W	41.10	45	2

Figure 2
Single and multi-word candidate terms, ranked by keyness.

The list of candidate terms obtained by means of the automatic keyword extraction tool gives an account of the most frequent units in this domain. However, it also requires a manual revision in order to detect the terms, which belong to this adventure activity, and to delete the units that do not, such as terms related to other fields that were also extracted by the tool (e.g. insurance-, accommodation-, and travelling-related terms), flora and fauna, or countries and nationalities.

After this manual revision, the terms extracted were thoroughly analysed and organised according to the five broad conceptual categories (Table 2), which coincide with the categories that were included in the preliminary ontology (Figure 1). These categories refer to the person (AGENT), the place (LOCATION), the action performed (ACTION) and the instrument employed (INSTRUMENT) in this adventure activity (ACTIVITY).

Agent	Activity	Action	Location	Instrument	
				Clothing	Security
Instructor	canyoning	trek	gorge	neoprene suit	harness
experienced instructor	rappelling	climb	mountain gorge	wetsuit	safety harness
professional instructor	canyoneering	walk	river gorge	glove	rope
qualified instructor	gorge walking	slide	valley gorge	mountain boot	safety rope
skilled instructor		swim	waterfall	wetsuit boot	single rope
expert staff		rappel	(natural) pool	rock shoe	double rope
canyoner		abseil	natural park	waterproof trouser	eight
		descent	national park		helmet
			river		safety helmet
			puddle		karabiner / carabiner
			ravine		rappel device
					zip line

Table 2
Categorisation of selected terms after term extraction.

3.4. Building corpus-based domain ontologies

The preliminary frame-based ontology created during the first step of this ontoterminography methodology can now be completed with the terms extracted from the CANYON corpus. Hence, Figure 3 shows the combination of the top-down (step 1) and bottom-up (step 2) of this methodology.

By carrying out a further corpus-based bottom-up study, consisting of searching for terms in context with a KWIC tool (such as *AntConc*)[7] and studying their concordances and co-text, the recurrent linguistic patterns of the conceptual relations were identified (Table 3). This method was also enriched with top-down *in vitro* analysis, i.e. study of definitions included in terminological resources about canyoning[8] and domain expert consultation.

[7] *AntCon* is a freeware concordancer developed by Laurence Anthony. URL: http://www.antlab.sci.waseda.ac.jp/software.html

[8] There are very few terminological resources about canyoning, but some online glossaries including terminology related to this activity can be found. For example, *Dictionary of terms used in canyoning* (https://canyonmag.net/es/technical/essentials/dictionary/), *Glossary of canyoneering terms* (http://dyeclan.com/outdoors101/canyoneering101/?page=glossary-of-

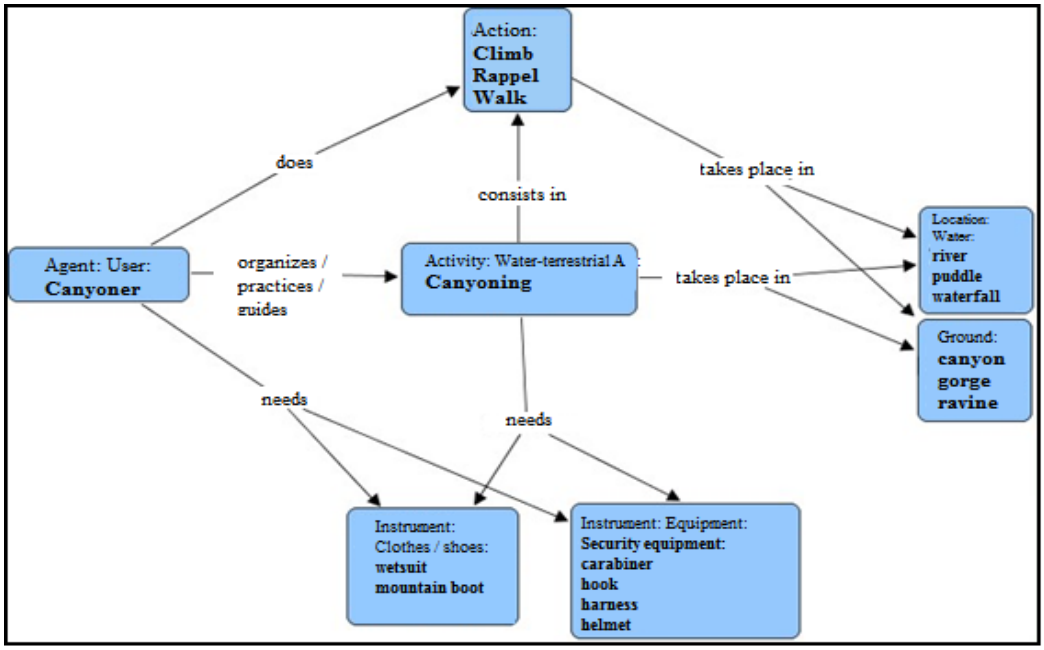

Figure 3
Preliminary frame-based ontology with canyoning-related terms (in bold).

With the acquisition of this further conceptual knowledge from the corpus, the preliminary conceptual representation was enlarged and became an extended corpus-based ontology. During this process, some subcategories were added (see Annex 1):

- In <agent>, three different agents were identified: 1. <organiser>, the person in charge of organising the adventure activity; 2. <technical staff>, dedicated to lead, coach or guide the activities, and 3. <user>, which corresponds to the person practicing the activity.
- In <instrument>, 1. <clothes/shoes> and 2. <security equipment> were identified, which were further subcategorised.
- <location> was also divided in different types of location (<water-air>, <air>, <water-ground>, <water>, <ground>).
- Category <activity> also was subdivided into <underground activity>, <water activity>, etc.

As observed, the extended corpus-based ontologies that result from the application of the ontoterminography methodology helps terminographers

canyoneering-terms) or *Canyoneering Glossary* (http://www.outdoorstogether.com/canyoneering_glossary.htm).

to organise the specialised knowledge in a clear and coherent way and becomes crucial in the subsequent steps of the methodology.

CONCEPTUAL RELATIONS	LINGUISTIC PATTERNS	CONCEPTUAL RELATIONS	LINGUISTIC PATTERNS
IS_A	- is a* - is a kind of - is a variant of - is a sort of - is a type of - is/are called - is an activity - such as - for example - known as - similar to - based on - named - a variety of	PART_OF	- contain* - composed of - compris* - is / are comprised of - a range of - consist* of - completed by - complemented by - classified - include* - including - is / are included
CONSIST_IN	- involv* - requir* - based on - with the aim of	TAKE_PLACE	- found at - found in - found throughout - in contact with - in the heart of - where - known in - located on - located in - situated in - situated on - tak* place - practi* on - practi* in - climb* up
NEED	- is / are required - required for - use* - using - the use of - mak* use of - utilis* - practis* with - is / are necessary - need* for - by means of - equipped with - be of help - with the help of - launch* by	ORGANISE GUIDE	- taught by - teach* - offer* - organis* - guid* - help* - lead* - made by - with the support of - the service* of - is / are responsible for - in the safe hands of - recommend* - advis*
		PRACTISE	- practi*ed by - done by

Table 3
Linguistic patterns of conceptual relations.

4. Final remarks

This paper applies the ontoterminography methodology to illustrate the different steps that are required to build a corpus-based ontology in a specialised domain, in this case the adventure activity of canyoning. It describes the different stages in which this middle-out protocol is divided and focuses on the first three, namely: 1. Design of the project and analysis of the domain, 2. Corpus compilation, 3. Term extraction and 4. Building of a corpus-based ontology.

As it has been proved, corpus-based ontologies in specialised domains provide users with a clear and organised representation of the specialised knowledge of a domain, including the main concepts and the conceptual relations among them, and are a convenient starting point to produce specialised resources. Corpora, for its part, are crucial in this methodology, both in the conceptual and linguistic dimension of any terminological project, since they are relevant information sources that contribute to the enlargement of conceptual representation as well as to the identification of terms, contexts, synonyms, equivalents, etc. This makes them indispensable for any terminological project nowadays.

With the aim of applying the methodology presented in this paper, the author is currently working on *DicoAdventure*, an online multilingual terminological database about adventure tourism which will contain terminology concerning the most common adventure activities, including canyoning, in English and Spanish. This project is the result of a collaboration with the OLST (*Observatoire de linguistique Sens-Texte*) at the University of Montreal.

Finally, it is worth mentioning that specialised knowledge is continuously evolving and changing over time, due to the emergence of new concepts, the change in meaning of previous concepts or the deletion of some others. Therefore, terminographers should not see any knowledge representation as stable and finite, but as a dynamic resource that may require adjustments in time.

Bionote: Lecturer (with tenure track) at the Department of English and German Philologies since 2016. BA in Translation and Interpreting from the University of Málaga and BA in English Studies from UNED. She obtained her PhD in June 2011, which was awarded with the best PhD dissertation award in the IV Biennial Iberian Association of Translation and Interpreting Studies [AIETI] and the Special doctorate award. Researcher in European, national, and regional R&D projects, and invited researcher at the Research Institute in Information and Language Processing (RIILP) of the University of Wolverhampton (UK), Erasmushogeschool Brussels (Belgium) and University of Montreal (Canada), among others. She has published in well-known national and international relevant conferences and journals. Some of her publications have received

awards for the quality of her research. Her main research lines are corpus linguistics, terminology, lexico-semantics, linguistic technologies and ICTs for foreign language teaching.

Author's address: iduran@uco.es

Acknowledgements: This research was partially carried out within the framework of the project *VIP: sistema integrado Voz-texto para IntérPretes* (ref. n. FFI2016-75831, 2017-2020) and the thematic network *TRAJUTEC: TRAducción JUrídica y TECnología* (University of Málaga).

References

Atkins S. and Rundell M. 2008, *The Oxford Guide to Practical Lexicography*, Oxford University Press, Oxford.

Baroni M., Kilgarriff A., Pomikálek J. and Rychlý P. 2006, *WebBootCaT: a web tool for instant corpora*, in *Proceedings of Euralex 2006*, Torino, Italy.

Cabré Castellví M.T. 1999, *Hacia una teoría comunicativa de la terminología: aspectos metodológicos*, in Cabré M.T. (ed.), *La Terminología: Representación y Comunicación. Elementos para una teoría de base comunicativa y otros artículos*, IULA, Universidad Pompeu Fabra, Barcelona, pp. 129-150.

Dash N.S. 2000, *Corpus Linguistics: An Introduction*, Pearson Longman, New Delhi.

Durán-Muñoz I. 2012, *La ontoterminografía aplicada a la traducción. Propuesta metodológica para la elaboración de recursos terminológicos dirigidos a traductores*, Peter Lang, Berlin.

Durán-Muñoz I. 2014, *Aspectos pragmático-lingüísticos del discurso del turismo de aventura: estudio de un caso*, in "Revista Normas" 4, pp. 49-69.

Faber P., Márquez Linares C., and Vega Expósito M. 2005, *Framing Terminology: A Process-Oriented Approach*, in "Meta" 50 [4], pp. 1-10.

Fillmore C. 1968, *The Case for Case*, in Bach E. and Harms R. (eds.), *Universals in Linguistic Theory*, Holt, Rinehart and Winston, New York, pp. 1-88.

Gruber T.R. 1993, *A Translation Approach to Portable Ontologies*, in "Knowledge Acquisition" 5 [2], pp. 199-220.

ISO (International Organization for Standarization) 2000, *ISO 704: 2000. Terminology Work. Principles and methods*, ISO, Genève.

ISO (International Organization for Standarization) 2000, *ISO 1087-1:2000. Terminology work – Vocabulary – Part 1: Theory and application*, ISO, Genève.

Leech G. 1992 *Corpora and theories of linguistic performance*, in Svartvik J. (ed.), *Directions in corpus linguistics. Proceedings of Nobel Symposium 82*, Mouton de Gruyter, Berlin/New York, pp. 105-122.

L'Homme M.C. 2008, *Le DiCoInfo. Méthodologie pour une nouvelle génération de dictionnaires spécialisés*, in "Traduire" 217, pp. 78-103.

Moreno Ortiz A. 2004, *Representación de la información terminológica en Ontoterm®: Un sistema gestor de bases de datos terminológicas basado en el conocimiento*, in Faber Benítez P. and Jiménez Hurtado C. (eds.), *Investigar en terminología*, Comares, Granada, pp. 25-70.

Roche C. 2007, *Saying is not modelling*, in *Proceedings of the 4th International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2007)*, Funchal, Portugal, pp 47-56.

Roche C., Calberg-Challot M., Damas L. and Rouard P. 2009, *Ontoterminology: A new paradigm for terminology*, in *Proceedings of International Conference on Knowledge Engineering and Ontology Development, KEOD 2009*, Madeira.

Temmerman R. 2000, *Towards New Ways of Terminology Description. The sociocognitive approach*, John Benjamins, Amsterdam.

Temmerman R. and Kerremans K. 2003, *Termontography: Ontology Building and the Sociocognitive Approach to Terminology Description*, in Hajicov E., Kotešovcov A. and Mírovský, J. (eds.), *Proceedings of the 17th International Congress of Linguists (CIL17)*, Matfyzpress, MFF UK (CD-ROM), Prague.

ANNEX 1. Extended semantic frame (or domain ontology)