

A Corpus-based Ontoterminological Tool for Tourist Translations

ISABEL DURÁN MUÑOZ

University of Málaga, Spain

ABSTRACT

In this paper, we will present our current project consisting of the creation of a corpus-based ontoterminological tool for tourist translations (German-Spanish) in the adventure tourism sector. Our project's main aim is to provide a tool of high quality which meets the requirements of specialised translators in the tourist sector.

This project is based on Termonotographical methodology (Temmerman and Kerreman, 2003; Temmerman and Geentjens 2006) and follows the theoretical and methodological framework described in the project HUM-892.¹ It involves the compilation of a bilingual comparable corpus in Spanish and German in order to elaborate a database with terminological information as well as other data useful to translators, such as context, examples, field, images, translation equivalents, collocations, phraseology and semantic information. This semantic information will be built in the form of conceptual relations by the creation of an ontology within this field.

Our project will try to repair the lack of qualified resources to translate tourist texts from German into Spanish, and vice versa, and to provide translators with valuable and useful information about the sector that they are working in. This work can only be managed by using representative corpora in both languages and analysing the results obtained by applying computational techniques and term extractors. In this sense, Corpus Linguistics, together with Computational Linguistics, becomes indispensable for our working field.

INTRODUCTION

Corpora as linguistic resources have been widely used within different fields, such as Linguistics or Lexicography, for the last decades. In contrast, they were not seriously considered as required resources within Translation Studies or Terminology until the 1990s (Meyer and Mackintosh 1996: 257).

Nowadays there can be no doubt as to the importance of using corpora in translation (Corpas Pastor 2008) and in terminological or terminographical work, and corpora are now recognised as an indispensable resource for these two disciplines.

The use of corpora in these disciplines has been accompanied by the development of linguistic technologies and resources, which contribute greatly to them. In this sense, electronic linguistic resources, suitable storage, and the selection of information that meets translators' needs have evolved over time (Cabré 2006: 93).

In the early days of electronic linguistic resources, terminological and lexical data banks were the most typical type of resources available, and translators used them frequently in order to find solutions to linguistic queries in translation contexts. Each record corresponded to a terminological unit, which appeared in a decontextualised form and was only sometimes accompanied by an example, usually in the form of a single context in which it had been found.

Text banks (or textual corpora) represent the second level in the development of resources which have become useful to translators. Text banks display data in authentic non-fragmented contexts. By using specialised tools, translators can access all occurrences of a unit in texts, search for collocations and phraseology, etc., which means that translators can view the use of units in context, in real texts and concordances.

Although textual corpora represent an advance in the evolution of linguistic resources in electronic form, they have also evolved to a new stage: knowledge databases. This current resource attempts to bring together grammatical, textual, terminological, documentary and semantic information, in order to provide valuable information to their users.

For translators (and other language professionals), knowledge databases can be very useful, since they provide rich and real information and help to solve linguistics as well as conceptual queries. According to Cabré (2004), knowledge bases “have become the automatic applications that give a more adequate response to the user’s needs.”

Apart from our project in the adventure tourism sector, several projects have been working on multilingual knowledge databases in different fields, either in Spain and abroad: GENOMA-KB, a specialized knowledge base in the human genome domain (<http://www.iula.upf.edu/>); Oncoterm, a bilingual specialized knowledge base in the medical subdomain of oncology (<http://www.ugr.es/local/oncoterm/>); COGNITERM, a specialized knowledge base in the domain of optical storage technologies (e.g. optical discs, drives, processes, etc.) (<http://aix1.uottawa.ca/~imeyer/research.htm>); ONTODIC, a research project whose aim is to elaborate a systematic methodology to create ontoterminological dictionaries (http://tecnoletra.uji.es/es/?page_id=35), among others.

Although these projects work in different domains, all of them share the same features and the same structure, that is: a text bank (or textual corpus), which contains relevant texts dealing with the particular specialised field; a documentary bank, which contains the bibliographic information; a terminological bank, which contains the term records for the terminological units, and an ontology, which represents the concept schema of the domain.

TERMONTOGRAPHY AND THE TRANSLATORS

During the last decade, Terminology and its applied branch, Terminography, have been subjected to several changes. First, the electronic revolution has affected terminology management and the terminographers’ working methods (Bouringault, Jacquemin, L’Homme 2001). Second, the principles of “traditional Terminology” have been seriously questioned (Cabré 1999; Temmerman 2000). The third change refers to the parallel working methods and mutual interests of ontology engineers and terminologists. These three changes produced a number of principles and methodologies which focus on flexibility and diversification of terminological resources depending on the user’s needs for each specific terminology project, and a new terminological approach emerged: Termontography.

Termontography is the terminological approach in which terminological information, retrieved from a corpus, is structured according to a framework of domain-specific knowledge (Kerremans 2007). This new approach is composed of the three following disciplines:

- Terminology: refers to the usage and study of terminological units of a specific domain.
- Ontology: the study of concepts in a particular domain.
- Terminography: the compilation of collections of the vocabulary of special languages.

This new theory is currently gaining ground in the Linguistics field. Its main purposes are the following two: on the one hand, describing ontologies with (multilingual) terminological information, and on the other hand, structuring

terminological resources with ontologies. In this case, “ontology” is understood as “a database describing concepts in the world or some domain, some of their properties and how the concepts relate to each other” (Weigand 1997).

The resources based on this new linguistic branch, such as specialised knowledge bases or ontoterminological databases, include the different types of information which translators, and other language professionals, require to work:²

- linguistic information of a terminological unit (definition, collocations, acronyms and abbreviations, form variants),
- pragmatic information (usage, domain, context), and
- cognitive (or ontological) information (semantic relations).

Translators, as creators and interpreters of texts (Rogers and Ahmad 1998), require context-sensitive information and subject knowledge in order to provide high quality translations. They need data that help them with the codification of the original text, but traditional multilingual terminological resources do not provide sufficient information. Looking at terms in context and within conceptual organisations helps translators to build up an image of the specialised field and to supplement their knowledge. This is why Termontography plays a crucial role in the elaboration of terminological resources for translators.

However, Termontography would not exist without Corpus Linguistics, and this is where this linguist discipline occupies a crucial place in this work.

ELABORATION OF THE ONTOTERMINOLOGICAL TOOL FOR TOURIST TRANSLATIONS

Nowadays, there are very few terminological resources that can be found in the tourist sector, especially for the German-Spanish language combination, and the existing ones fail to offer information according to translators’ needs, such as word use, equivalents, collocations, contexts, semantic relations, and so on. They only provide terminological information –sometimes doubtful – without giving more information to users. In this sense, the existing terminological resources for the domain of adventure tourism for German-Spanish do not satisfy translators’ needs,³ since translators require more information and more guidance to codify original texts.

Our tourist-domain ontoterminological tool is based on two main resources: a concept-based bilingual (German-Spanish) terminology database, and domain-specific corpora for both languages for information extraction. Furthermore, it contains all the required data elements that professional translators indicated in a study about terminological resources and their needs.⁴

The methodology followed to manage this tool is based on the theoretical and methodological framework described in the project HUM-892 (See Note 1) and comprises four major steps, which will be explained below:

1. Compilation of a specialised corpus on the tourist domain;
2. Extraction of terms;
3. Creation of a tourist-domain ontology, and
4. Completion of a terminological database with the extracted terms.

Compilation of a Specialised Corpus on the Tourist Domain

The first step in the process of building a knowledge base or an ontoterminological database is the compilation of a corpus, whose nature will depend on the interests and applications of the work. This step consists of the search, retrieval and storage of domain-specific texts from different sources, which can vary depending on the corpus goals.

Our aim with this corpus is the compilation of online promotional texts related to the adventure tourism sector in German and Spanish, in order to acquire the terminology used in this specialised field and in this language pair. In this sense, the nature of our corpus can be summarised as follows: a written electronic bilingual corpus, whose application is Termontography, following Austermühl's description of corpora (Austermühl 2001: 125-128).

This corpus is divided in two subcorpora, one for Spanish and one for German, and has been compiled by accessing documents on the Internet and converting them into plain text format (.txt). All the texts retrieved to this compilation present the same parameters: they are about the same theme, adventure tourism and its activities (such as trekking, horse riding, mountain bike, abseiling); they are all complete and original promotional texts; they are synchronic (collected from January to April this year); they are addressed to tourists and potential tourists, and their authors are tourist organisations, associations and tourist companies, located either in Spain or in Germany.

As said before, both subcorpora, the Spanish and the German one, are based on the same theme, the adventure tourism sector, and therefore both of them contain texts related to the same activities practised in this sector. This fact allows us to obtain the same terminology in both languages and to compile a balanced and homogeneous corpus. The topics included in the selected texts can be divided in the following activities:

Table 1. *Topics in the adventure tourism subcorpora.*

GERMAN	SPANISH
Abenteuertourismus	Turismo de aventura, turismo de riesgo
Aktivtourismus	Turismo activo
Sporttourismus	Deporte Turismo deportivo
Wassersporttourismus	Turismo náutico
Bergtourismus	Turismo de alta montaña
Wintertourismus	Turismo de nieve, Turismo blanco
Mountainbiking	Mountain Bike, bicicleta de montaña
Kanu	Canoa
Trekking, Wanderung	Trekking, Senderismo
Heißluftballonflüge	Globo, Globo aerostático
Rafting	Rafting
Bergsteig	escalada
Drachenfliegen	Vuelo en ala delta
Langlauf	Esquí de fondo, esquí nórdico
Ski	Esquí
Snowboard	Snowboard
Kajak	Kajak

The corpus also follows the criteria established by Sinclair in the EAGLES project (1996: 4): quantity, quality, and documentation.

The criterion about quantity presents some controversy, since some authors indicate it is not relevant for a terminological work (Meyer et. al. 1996) while others consider it essential (Biber 1993; Sinclair 1996). From our viewpoint, we believe that this criterion is important to the validity of a corpus and we have tried to meet this requirement.

In total, our Spanish corpus presents 781 files with 291,694 words and our German corpus consists of 612 files with 213.056 words. With this amount of words, our corpora reach the representativeness of this specialised field,⁴

understanding for “representativeness” Biber’s definition (1993: 243), that is to say, “the extent to which a sample includes the full range of variability in a population.”

Our corpus also fulfils the quality criterion, since all the texts selected satisfy the parameters established during the design phase, which are: a reliable authorship, same domain (adventure tourism), same typology of texts (promotional texts), same specialised level ((semi)-expert to beginner/tourist), same temporal period and limited geographical aspect (Germany and Spain).

The third criterion is documentation. It refers to the extra information included in the corpus to describe the texts and to facilitate the control and detection of those texts once storage has been carried out. This information includes: the text source, authorship, date, textual genre, specialisation level, etc.). All the texts in our corpus have been registered in a record sheet containing all these fields.

Table 2. *Example of the Spanish record sheet.*

Code	Title	Url	Textual Genre	Language	Tourist Segment	Translation	Domain	Authorship	Contact detail
30001 CSP	Multiaventura en las Zorreras	http://www.pangeacentral.com/multi_zorreras.html	Promotional	Spanish	Turismo de aventura	No	Regional (Andalucía)	Pangea Active Nature	Pasaje Cayetano nº 10 LOCAL D - Ronda (MALAGA). Teléfonos: 952 87 34 96 (Oficina) - 630 56 27 05 (Móvil). Email info@pangeacentral.com

Extraction of terms

Once the domain-specific corpus is compiled and codified according to a code established during the design stage, further analysis is performed to extract the terms which will be included in the knowledge base.

Before extracting the terms from the corpora, pre-processing is required. This pre-processing can be of different kinds; syntactic annotation, part-of-speech tagging or morphological annotation. In our case, we only used part-of-speech tagging (POS tagging), the most used corpus annotation scheme to date. This kind of annotation consists of assigning to each word in a text an unambiguous indication of the part of speech which this word belongs to in its context (Ooi 1998: 135). Although manual POS tagging is possible for this kind of annotation, there are several POS taggers that carry out tagging automatically and offer very good results, which reduces the tagging time and also the possible inconsistency of manual tagging.

In this research, we applied TreeTagger (<http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>), a free POS tagger available on the Internet developed at the Institute for Computational Linguistics of the University of Stuttgart. It provides good results and can be used in different languages (German, English, French, Italian, Dutch, Spanish, Bulgarian, Russian, Greek, Portuguese, Chinese and old French).

Table 3. *Sample of a tagged text with TreeTagger.*

POR	PREP	por
SIERRA	NC	sierra
DE	PREP	de
LAS	ART	las
NIEVES	NC	nieve
Senderismo	NC	senderismo
Descripción	NC	descripción
El	ART	el

Pinsapo	NP	<unknown>
es	VSfin	ser
un	ART	un
abeto	NP	abeto
endémico	ADJ	endémico
del	PDEL	del
sur	PE	sur
de	PREP	de
España	NP	España
Este	DM	esto
tipo	NC	tipo
de	PREP	de
árbol	NC	árbol
sólo	ADV	sólo

In this table, we can observe the result given by the POS tagger. It consists of three columns: the first one contains the words that appear in the text; the second one, the part-of-speech tag (PREP = preposition, ADJ = adjective, ART = article, etc.), and the third column indicates the lemma of the word in column one given by the tagger. For example, the tagger indicates that the neutral form of “es” is “ser”, its infinitive.

Once this pre-processing is carried out, we can extract the terms from the corpora. Nowadays, there are three different approaches to term extraction, which are usually classified as linguistic, statistical, or hybrid.

Terminology extraction tools following a linguistic approach try to identify terms by their linguistic (morphological and syntactic) structure. For this purpose, texts are annotated with linguistic information with the help of morphological analysers, part-of-speech taggers and parsers. Then, term candidates following certain syntactic structures are filtered from the annotated text by using pattern matching techniques. An example of this technique is represented by the program LEXTER (Bourigault 1992). Another commonly used technique is to filter term candidates by looking for commonly used text structures such as definitions and explanatory contexts like “X is defined as ...” or “X is composed of ...” (Pearson 1998).

The general assumption underlying the statistical approach to term extraction is that specialised documents are characterised by the repeated use of certain lexical units or morphosyntactic constructions. Terminological extraction tools based on statistics try to filter out words and phrases having a certain frequency-based statistic. Another common method is to compare the frequency of words and phrases in a specialised text to their frequency in general language texts assuming that terms tend to appear more often in specialised texts than in general language texts.

Terminological extraction tools following a purely linguistic approach tend to produce too many irrelevant term candidates (noise), whereas those following a purely statistical approach tend to miss term candidates that appear with a low frequency value (silence) (Ha, Fernandez, Mitkov, Corpas 2008).

More recently, approaches to automatic term extraction have moved towards using both statistical and linguistic information. Generally the main part of the algorithm is the statistical part, but shallow linguistic information is incorporated in the form of a syntactic filter which only permits phrases having certain syntactic structures to be considered as candidate terms.

This method is the one used in our research, that is, a hybrid method based on linguistic and statistic information. In this sense, we can extract different part of speech elements, such as nouns, adjectives, verbs, etc., either as single-word units or as multi-word units, and order them by frequency. After the term candidate extraction, we had to select the terms which would be included in our ontoterminological tool. The selection was made with the help of experts in the tourist domain, who have collaborated in our research.

Table 4. Sample of the first 50 terms extracted by linguistic and statistical information.

N	Word	Freq.
1	sierra	1418
2	ruta	1242
3	actividad	1218
4	agua	1027
5	hora	964
6	río	853
7	zona	797
8	camino	795
9	metro	721
10	material	711
11	lugar	673
12	viaje	522
13	barranco	459
14	programa	454
15	mar	444
16	cañón	412
17	carril	403
18	precio	392
19	pueblo	390
20	carretera	384
21	descenso	382
22	cuerda	381
23	servicio	359
24	ronda	353
25	forma	332
26	cueva	330
27	arroyo	317
28	punto	307
29	montaña	295
30	vista	286
31	pie	276
32	tiempo	269
33	vuelo	266
34	duración	265
35	aventura	261
36	cerro	252
37	reserva	252
38	cortijo	249
39	puerto	245
40	torre	242
41	consumidor	241
42	paisaje	241
43	caballo	237
44	kilómetro	234
45	sendero	234
46	equipo	233
47	cumbre	220
48	nivel	220
49	naturaleza	215
50	ropa	211

Creation of a Tourist-Domain Ontology and a Terminological Database

The two remaining steps were completed simultaneously. The terminological database was developed in parallel with the ontology, rather than being developed after the ontology was finished, as has been the case in some other projects. This methodology followed Cabré’s words (2006: 12) when the author states:

The only way to access concepts is through the terminological units that designate them. Therefore, for each term, there will be an associated concept in the ontology, which can be accessed via the corresponding term.

In this sense, a cyclic relation (Fig. 1) will be established between the term and the concept, with the concept the basic element and the term the access to it.

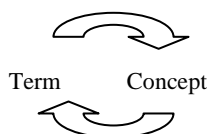


Figure 1. A cyclic relation between concept and term.

The system used to develop the terminological database and the ontology is Ontoterm (<http://www.ontoterm.com>), created by Professor Antonio Moreno of the University of Málaga. Ontoterm is composed of two primary modules: the Ontology Editor, where ontology building is carried out, and the Termbase Editor, where lexical mappings and term description take place. It also includes other tools which facilitate browsing, navigating, and reporting the ontology (Moreno and Pérez 2001).

As we previously defined, an ontology is a database including the concepts in a specific domain, their properties and the relations between those concepts, and therefore it is composed of the following components:

- Concepts (also called classes), which represent ideas about the physical or the abstract objects which configure a specific domain.
- Relations, which refer to the associations between the concepts of a specific domain.
- Instances, which represent individuals or specific elements in an ontology.
- Attributes, which refer to the properties of concepts and instances.

In our case, our tourist-domain ontology also presents these four categories. It is divided into five different concepts or classes: activity, such as trekking, ballooning, canoeing, kayaking; activity type, such as terrestrial, aquatic or aerial activity; place of interest, such as mountain, lake, ski center, natural park; suitable clothes, such as trainers, warm clothes; and transport, such as bus, boat, kayak or car.

It contains relations between the concepts, such as subclass-of, include, and IS-A, and attributes of each class and instance, such as duration, age, difficulty levels, habitat or necessary equipment.

Finally, it also includes instances, such as “Parque Natural de los Alcornocales” as an instance of “Parque natural”, in Spanish.

In the next figure (Fig. 2), the concept map and its elements can be observed very clearly. Here we have used another ontology management programme, Protégé, to show the semantic relations in graphics. Ontoterm does not allow the opportunity to create graphics, and therefore we have been forced to use another frequently used programme to visualise the relations between concepts.

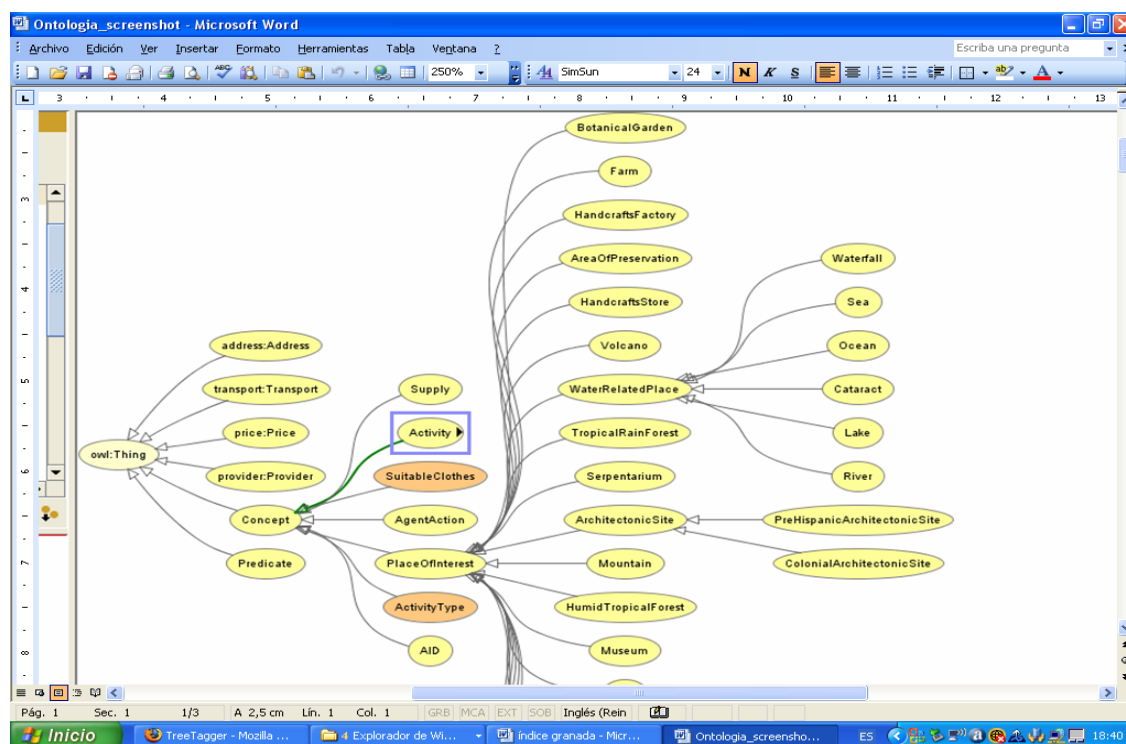


Figure 2. Sample taken from Protégé, an ontology management system.

At the moment, the ontology project is still a work in progress, but once it is finished, our tourist-domain ontology will help translators to structure tourist-domain knowledge, to clarify the relations established between different concepts; to simplify comparative analysis between concepts and designations in different languages, and to facilitate the codification of original texts.

As previously stated, the ontology module was developed parallel to the terminological module in Termbase Editor, where the task to describe the linguistic, usage and administrative categories is carried out. This module is subordinated to the ontology, and therefore the ontology module must be activated during the editing or browsing process of the terminological module.

For space reasons, we cannot go into the details here and will limit ourselves to a presentation of the different fields contained in the database. These fields were selected after analysing the results obtained in a survey about the preferences and needs of translators. To the question relating to the content of terminological resources “What do you think a good terminological resource for translators should offer?,” the 402 translators who participated in the survey answered the following:

Table 5. Percentage to the survey question “What do you think a good terminological resource for translators should offer?”

	Essential data	Desirable data	Irrelevant data
A great number of entries	50.94%	44.77%	4.29%
A great variety of units (nouns, verbs, adverbs, adjectives, proper nouns)	33.79%	55.22%	10.99%
Abbreviations and acronyms	55.26%	41.78%	2.96%
Definitions	74.86%	24.04%	1.09%
Clear and concrete definitions	69.15%	30.32%	0.53%
A definition in both languages (if bilingual)	45.11%	45.38%	9.51%

Etymological information	6.65%	38.23%	55.12%
Pronunciation	5.40%	33.52%	61.08%
Syllabification	4.29%	25.43%	70.29%
Examples	57.33%	40.53%	2.13%
A greater variety of examples	29.21%	60.11%	10.67%
Phraseological information (collocations and idiomatic expressions)	56.06%	40.70%	3.23%
Derivatives and compounds	51.51%	41.92%	6.58%
Domain specification	22.35%	41.92%	6.58%
More translation equivalents	54.97%	64.80%	12.85%
An explanation of each translation equivalent	34.60%	60.49%	4.90%
Grammatical information (part of speech, inflections, number, gender)	28.69%	58.22%	13.09%
Instructions for use	26.94%	50.83%	22.22%
Semantic information (synonyms, antonyms, hypernymy, hyponymy)	13.76%	53.65%	32.58%
Pictorial illustrations	21.17%	63.79%	15.04%
A great number of entries	13.56%	61.58%	24.86%

Taking into account these results, we select the fields to be included in our terminological module from the essential and desirable data and exclude the irrelevant data. After carrying out this step, we continue to fill in the selected fields with the information taken from the corpus.

To the best of our knowledge, this is the first terminological research study which takes into consideration the preferences and needs of translators and which creates a terminological tool from this information.

The idea behind this is to manage an ontoterminological tool which satisfies translators' needs and expectations in all their facets.

CONCLUSIONS

There is no doubt that information technologies have significantly changed the translation process. The consults to traditional resources, like hardcopy dictionaries, are equalled, or even exceeded, by online consults to electronic dictionaries or searches in search engines, like *Google* (<http://www.google.com>).

However, simultaneously with the development of powerful search engines and filters, linguistic engineering has been exploring more suitable applications to resolve language professionals' queries and problems. One of these new resources is the semantic-based terminological database, known as specialised knowledge databases or ontoterminological database. These new resources are based on a conceptually-structured terminology, and their aim is to facilitate the understanding of a specialised domain by providing terminological and semantic information mapped in a same database.

These conceptually organised terminology resources have already been applied to the management of knowledge in complex organisations, and have been proved to be useful for meeting the needs of translators.

In this sense, and as the basic resource to elaborate this kind of database, corpora are considered to be well-established and consistent if they meet the Corpus Linguistics criteria for their compilation (quantity, quality, and

documentation, cf. Sinclair, 1996: 4). They have been proved to be not only a source of linguistic evidence but also a source of domain knowledge, and therefore their design and compilation is very relevant to terminographical work.

To our knowledge, this ontoterminological tool on the adventure tourism domain will fill the existing gap of terminological resources in this field (especially for the language pair German-Spanish) and will satisfy tourist translators' needs within this domain.

FURTHER WORK

As further work, we intend to extend the specialised domain of this research, that is, the adventure tourism sector, and elaborate the ontoterminological tool for all tourism fields and sectors, such as rural tourism, cultural tourism, gastronomic tourism, sun and beach tourism, etc. We will also extend the number of languages involved to English, Italian and French.

NOTES

1. The research reported in this paper has been carried out in the framework of R&D Project for Excellence *La contratación turística electrónica multilingüe como mediación intercultural: aspectos legales, traductológicos y terminológicos* [Multi-lingual tourism e-contracts: legal, translational and terminological aspects]. Funding source: Andalusian Ministry of Education, Science and Technology. Ref. no. HUM-892 (2006-2009). Its methodology is also implemented in the current project ECOSISTEMA: *Espacio Único de Sistemas de Información Ontológica y Tesauro sobre el Medio Ambiente*, coordinated program by University of Málaga and University of Granada. Funding source: Spanish Ministry of Science and Technology (National R&D Program) (2008-2011). Ref. no. FFI2008-06080-C03-03.

2. A survey about terminological resources for professional translators and their needs has been carried out as part of an author's academic research project. The survey is available online under <http://clg.wlv.ac.uk/surveys>.

3. A non-published study about the existing terminological resources for German-Spanish in the domain of adventure tourism has been carried out as part of an author's academic research project.

4. See note 2.

5. The representativeness of this corpus is been measured using the ReCor tool, a computer application which estimates the minimum representativeness of a corpus by measuring its types and tokens. (See Corpas Pastor and Seghiri Domínguez 2006).

REFERENCES

- Biber, D. 1993. Representativeness in corpus design. *Literary and Linguistic Computing. Journal of the Association for Literary and Linguistic Computing*, 8/4, 243-257.
- Bouringault, D & C. Jacquemin & M-C. L'Homme 2001. *Recent Advances in Computational Terminology*. Amsterdam/Philadelphia: John Benjamins.
- Cabré Castellví, M. T. 1999. *Terminology. Theory, methods and application*. Amsterdam: John Benjamins.
- Cabré Castellví, M. T. 2004. A Specialized Knowledge Base: from Distributed Information to the Specialized Dictionary Construction". In *Proceedings of the 11th EURALEX International Conference*, pp. 867-872.
- Cabré Castellví, M. T. 2006. *From Terminological Data Banks to Knowledge Databases: The Text as the Starting Point*. In Bowker, L. (ed.), *Lexicography, Terminology and Translation. Text-based studies in honour of Ingrid Meyer*. (pp. 123-158). Ottawa: Ottawa University Press.

- Corpas Pastor, G. 2002. *Traducir con corpus: de la teoría a la práctica*. In M. García Palacios y T. Fuentes Morán (eds.), *Texto, terminología y traducción* (pp. 189- 226). Salamanca: Almar.
- Corpas Pastor, G. 2008. *Investigar con corpus en traducción: los retos de un nuevo paradigma*. Frankfurt: Peter Lang.
- Corpas Pastor, G. y Seghiri Domínguez, M. 2006. *El concepto de representatividad en la Lingüística del Corpus: aproximaciones teóricas y metodológicas*. 28 Sept. 2008. Available online: <<http://www.turicor.com/privada/documentos/TI-DT-2006-1.pdf>>
- EAGLES. 1996. Preliminary Recommendations on Corpus Typology. Eagles Document (Expert Advisory Group on Language Engineering) EAG-TCWG-CTYP/P, by J. M. Sinclair, University of Birmingham, Corpus Linguistics Group.
- Ha, L. A., Fernandez, G., Mitkov, R. and Corpas, G. 2008. Mutual bilingual terminology extraction. In *Proceedings of LREC'08*. Marrakech
- Meyer, I. and Mackintosh, K. 1996. The Corpus from a Terminographer's Viewpoint. *International Journal of Corpus Linguistics*, 1/2, 257-285.
- Meyer, I., Skuce, D., Bowker, L., and Eck, K., 1992. Towards a New Generation of Terminological Resources: An Experiment in Building a Terminological Knowledge Base. In *Coling'92: Proceedings of the 15th International Conference on Computational Linguistics*, 129-138.
- Moreno, A. and Pérez, Ch. 2001. From text to Ontology: Extraction and Representation of Conceptual Information". In *Actes des quatrièmes rencontres Terminologie et Intelligence Artificielle (TIA 2001)*, Nancy, 233-242.
- Ooi, V. B. Y. 1998. *Computer Corpus Lexicography*. Edinburgh: Edinburgh University Press.
- Pearson, J. 1999. *Terms in context*. Amsterdam: John Benjamins.
- Rogers, M. and Ahmad, K. 1998. The Translator and the Dictionary: Beyond Words? In *Using Dictionaries. Studies of Dictionary Use by Translators*. Lexicographica Series Mayor. Tübingen.
- Temmerman, R. 2000. *Towards New Ways of Terminology Description. The sociocognitive approach*. Amsterdam: John Benjamins.
- Temmerman, R. and Kerremans, K. 2003. Termontography: Ontology Building and the Sociocognitive Approach to Terminology Description. In: Hajicová, E., Kotešovcová, A., Mírovský, J. (eds.), *Proceedings of CIL17*, Matfyzpress, MFF UK (CD-ROM). Prague, Czech Republic. 28 Sept. 2008. Available online: <http://www.hf.uib.no/forskingskole/temmerman_art_prague03.pdf>
- Temmerman, R. and Geentjens, S. 2006. Termontography and domain-specific translation dictionaries. Tackling cross-linguistic and cross-cultural variation. *International Conference on Terminology: Terminology and Society*, Antwerp, Belgium, November 15-17.
- Varantola, K. 2002. Disposable Corpora as Intelligent Tools in Translation. *Cadernos de Tradução* 9 (1). 20 Sept. 2008. Available online: <<http://www.cadernos.ufsc.br/online/cadernos9/krista.pdf>>.
- Weigand, H. 1997. Multilingual Ontology-Based Lexicon for News Filtering –The TREVI Project. In K. Mahesh (1997), *Ontologies and Multilingual NLP. Proceedings from the 15th International Joint Conference on Artificial Intelligence (IJCAI)*, August 23-29, 1997. Nagoya, Japan. 138-159.

ISABEL DURÁN MUÑOZ
 DPTO. TRADUCCIÓN E INTERPRETACIÓN
 FACULTAD DE FILOSOFÍA Y LETRAS
 CAMPUS DE TEATINOS
 UNIVERSIDAD DE MÁLAGA
 29071 MÁLAGA
 SPAIN

