

## A methodology for building a translator- and translation-oriented terminological resource

By Adonay Moreira (Polytechnic Institute of Leiria, Portugal)

### Abstract & Keywords

#### English:

The main goal of this paper is to provide insight into our parallel corpus-based approach to the creation of a term bank in the subject area of tourism. The term bank is based on a unidirectional parallel corpus of Portuguese-English tourist advertising material and it is conceived as a tool for all those interested in finding linguistic, conceptual and pragmatic information on the terminology of tourism. Thus, it can become particularly useful for translators, who need to master the specialized lexical items and find their appropriate foreign language equivalents, and tourism professionals, who work in an increasingly multilingual society and would gain from access to a 'ready-made' bilingual list of terms.

**Keywords:** Bilingual term banks, corpus linguistics, parallel corpus, tourism marketing, lexicography and terminology

---

### 1. Introduction

These last few years have witnessed an increase in research involving the compilation of large quantities of texts and their respective translations, as well as the development of techniques for processing those bilingual term banks (Bowker and Pearson 2002; Biber *et al.* 2004; McEnery and Wilson 2004). The present study is an example of such research as it uses a Portuguese-English parallel corpus as a starting point for the retrieval of terminology. Thus, *Turigal*, a parallel corpus of tourist advertising material, has been devised to support the creation of a bilingual term bank on tourism.

It is worthwhile to mention that, although tourism is one of the main sectors of the Portuguese economy, little attention has been paid to the promotional texts responsible for such an important cross-cultural and linguistic contact and the study of the terminology of tourism in Portugal is practically nonexistent. There are only two monolingual publications on the terminology of tourism in Portugal: *Dicionário Técnico do Turismo* (Domingues 1990) and *Prontuário Turístico* (Domingues 1997). Though extremely useful for professionals working in the tourist industry, these publications give no information about the origin of term definitions and they lack English equivalents, contexts of use, morphological information and frequency of terms. As far as term banks are concerned, there is only one linguistic resource that contains bilingual tourism terminology: *Lextec – Léxico Técnico do Português* (Instituto Camões 2010). *Lextec* contains term definitions and illustrative examples which display terms in context and thus facilitate the understanding of concepts. It also indicates a range of relationships between terms, such as generic-specific and part-whole relations, and English equivalents are provided, although there is no information about the origin of those equivalents. On the whole, *Lextec* evinces a normative, monolingual-oriented approach to terminology. Moreover, terms are not organized around a thematic structure, their frequency is not measured, and the illustrative examples are collected solely from the Internet and do not belong to a homogeneous technical corpus.

Our bilingual term bank of tourism terms is an attempt to mitigate these shortcomings, since it comprises pragmatic (context of use, relative frequency of terms), linguistic (gender, number, grammatical category, lemmas, synonyms) and conceptual information (thematic tree, semantic relations) in a specific context. Our comprehensive term bank could eventually assist translators working in this industry or tourism professionals who work in an increasingly multilingual society and would gain from access to a 'ready-made' bilingual list of terms.

### 2. Theoretical approach to the compilation of terminology

We take Michael Stubbs' statement as a starting point for our theoretical approach: "The meaning of words is created in texts: the instance has a history in a text. And ultimately, words change in meaning because their use in texts changes" (Stubbs 1996: 89). If words have no fixed meaning prior to their use in texts (Stubbs 1996: 92), the same happens to terminological units which, according to Teresa Cabré's Communicative Theory of Terminology, are simply lexical units that activate a specialized value in a certain pragmatic-discursive context (Cabré 1999: 90). In other words, it is the context of use of a term which provides it with a specialized value.

Cabré's approach recognizes the role of discourse and text (we perceive text as an embodiment of discourse) in the study of descriptive terminology. She acknowledges text as the natural habitat of terminology and points to the diversity of approaches to terms (Cabré 2005: 92). This diversity is consistent with the way Cabré (2003: 183) characterizes the nature of terms. Cabré describes terms or terminological units as many-sided: they are simultaneously units of knowledge, language and communication. The author uses the image of the polyhedron to show their multidimensional nature: "At the core of the knowledge field of terminology we, therefore, find the terminological unit seen as a polyhedron with three viewpoints: the cognitive (the concept), the linguistic (the term) and the communicative (the situation)" (Cabré 2003: 187). These linguistic, cognitive and communicative units should be studied in a given discursive context, because they only acquire a meaning and function in discourse.

Cabré's approach brings terminology closer to texts and this perspective is also supported by other authors, such as Temmerman (2000) and Pearson (1998). These authors move away from the idea that terms have an intrinsic specialized value and they share the view that it is texts that accord specialized status to terms. Since meaning is created in texts, from the methodological point of view it is vital to identify the context in which texts are produced and to characterize the texts that make up the corpus. Therefore,

our descriptive bilingual terminological products based on the use of a parallel corpus where terms can be analysed in their context.

While this textual context is important, terms are also selected within the context of a specific application. Cabré *et al.* consider that “[...] *term* designates the meaningful text unit in specialized discourse considered useful for an application” (Cabré *et al.* 2007: 2). In our view, this definition reveals a shift towards a textual terminology that focuses on the real needs of its users; this is the definition that best suits the purpose of this work. We extract text units, properly contextualized in the area of tourism, with the aim of creating a specific terminological application – a term bank on tourism – to bridge an economical and social need which has been previously identified.

Thus, one may conclude that what determines a lexical unit as a term is, on the one hand, the discursive context, represented in the selected corpus, and, on the other hand, both the users and objectives of the term bank. Ultimately, it is the terminologist and specialist who determine the choice of terminology in the context of their work and decide whether or not to include a term in a particular application, given the relevance of that term within the predefined objectives.

Since our terminological approach is corpus-based – specifically, based on a parallel corpus, where the meaning of terms arises from their context of use – and this corpus is determined by the purpose for which it will be used, we have named it “special purpose parallel corpus”. This expression is adapted from Pearson (1998: 48)’s term – “special purpose corpus” –, which designates a parallel corpus built for specific purposes. Thus, within the present research, a special purpose parallel corpus consists of a corpus of original texts and their translations, which is used for terminological purposes.[1]

### 3. Parallel corpora and bilingual terminology

The purpose of this section is to present the advantages of using parallel corpora for terminological purposes and to give some examples of research undertaken in this area. There is still some reluctance to use parallel corpora in the compilation of bilingual or multilingual lexicographical products. Although a parallel corpus is a relatively straightforward way to identify equivalents in another language, the influence of the source language on the target language is often feared. Thus, in lexicographical projects, parallel corpora are primarily used as secondary sources of words. Parallel corpora are used to confirm the use of certain equivalents in the target language or to identify the equivalents lexicographers did not detect with comparable corpora.

The fear of linguistic interference in translations has also justified their rejection as appropriate data to be used in the compilation of terminological products. Few studies show the usefulness of parallel corpora in bilingual terminology and this fact can also be ascribed to the scarcity of parallel corpora. Compiling parallel corpora is rather more complex and time-consuming than collecting comparable corpora. Still, in recent years, some research projects on the use of parallel corpora in bilingual terminology have been undertaken (Vintar 2001; González-Jover and Sierra 2002; Gómez Guinovart 2008).

Vintar (2001) developed a methodology for the automatic extraction of terms and simple compounds in Slovenian, as well as their translated English equivalents, applying statistical and syntactical methods to a parallel corpus. Her research shows that it is feasible to automatically extract bilingual terminology from a parallel corpus, in order to be used as a translation tool and as the basis of a term bank.

González-Jover and Sierra (2002) demonstrate the use of both parallel and comparable corpora in bilingual terminology extraction. Their main aim is to build a terminological term bank and a bilingual dictionary. On the one hand, parallel corpora are automatically aligned, allowing for the search of terminology. One writes down a word in the search button and all occurrences of that word in Spanish and its translations into English come out aligned. On the other hand, comparable corpora are the starting point for the extraction of bilingual terminology and phraseology. The extraction of terminology is carried out in both languages and it is followed by the grouping of Spanish and English terms which have been previously selected. The authors point out the complexity of this process, given the difficulty of finding equivalents for a particular term (González-Jover and Sierra 2002).

Research carried out by the Computational Linguistics Group (SLI) and the Neology Group of the University of Vigo has resulted in the creation of terminological and lexicographical resources from the Linguistic Corpus of the University of Vigo (CLUVI) (Gómez Guinovart 2003). CLUVI is an open set of parallel textual corpora of specialized registers of contemporary Galician language. With a current total length exceeding the 22 million words, CLUVI comprises six main parallel corpora belonging to four specialized registers (from fiction, computing, popular science and legal-administrative fields) and five different language combinations with Galician (Galician-Spanish, English-Galician, French-Galician, English-Galician-French-Spanish and Spanish-Galician-Catalan-Basque) (Gómez Guinovart and Sacau Fontenla 2004). The CLUVI corpus has been gradually enlarged to accept other linguistic combinations, such as English-Portuguese, English-Spanish and Portuguese-Spanish.

As far as terminology is concerned, the same research group at the University of Vigo created the *Terminological Databank of the University of Vigo* or *Termoteca* (Gómez Clemente *et al.* 2006). This terminological databank is based on both monolingual and parallel corpora – the *Technical Corpus of Galician (CTG)* and the CLUVI corpus respectively–, and it only provides terms and examples of use from corpora. Our term bank on tourism is included in *Termoteca*.

The three aforementioned studies – Vintar (2001), González-Jover and Sierra (2002) and Gómez Guinovart (2008) – demonstrate that the extraction of terminology from parallel corpora is not only feasible, but extremely useful, particularly in technical translation. Therefore, one can identify several advantages of using parallel corpora in bilingual terminology, as, for instance, the fact that there is a more direct semantic connection between texts and their translations, since most notions or meanings from the source text match those of the translated text. In addition, the process of finding equivalents is less complex than in a comparable corpus.

Vintar (2001) points out another advantage: extracting bilingual terminology from parallel corpora can help translators produce more suitable technical texts, since traditional terminological resources are often unable to keep up with language, technology and terminology change in a particular area.

Including translations in prestigious reference books, such as dictionaries or term banks, also validates translators’ work. It is vital to develop tools to assist translators and one alternative is to draw on their own work to create such tools, hence rejecting the insidious assumption that translation is a “second-class” text that carries traces from both source and target texts/cultures. After all, as Teubert wittingly states: “The meaning of the translation unit in the source language is its equivalent in the target language. [...] It is the target language that determines the unit of meaning” (2002: 212). If the meaning of words derives from the linguistic context in which they occur, translation is the ideal place to expose cross-linguistic lexical correspondences, as Teubert said:

The core issue of translation is meaning. For each semantic unit of the source text, there has to be an equivalent in the target text. Therefore cross-linguistic lexicography in quest of meaning must pay close attention to the practice of translators. It is they who invent the translation equivalents for lexical expressions. For these translation equivalents are not discovered, they are invented (2002: 191).

Thus, the use of parallel corpora in bilingual terminological products means recognising the creative vitality of translators, who produce many of the texts that circulate in our society.

From the economic point of view, it also makes sense to reuse terminology of translated texts in order to meet companies' growing need for term banks that describe their products (Teubert 2005: 97, 101). *Turigal* aims to address this necessity in the tourism marketing of Portugal and to fulfil some translation needs, namely the translation of terms associated with cultural features, whose translation is particularly difficult. One has come to realize that translators rarely have access to the translation of these words in both general and specialized dictionaries, thus parallel corpora are a valid instrument to identify translation equivalents.

#### 4. *Turigal* Corpus – collection and encoding

The corpus on which the term bank is based consists of texts (printed brochures/leaflets and websites) in Portuguese and their translations into English, all of which were sourced from the nineteen Portuguese Tourism Regions and Regional Tourism Promotion Agencies, and stored as plain text. Tourism Regions are public organizations which are responsible for promoting regional and national tourist activity. Their aim is to contribute to the development of a region's historical, cultural and natural heritage. According to Kotler (2004: 482), Tourism Regions are mainly responsible for designing, developing and promoting destination products in adequate markets. Regional Tourism Promotion Agencies are non-profit associations constituted by representatives of economic agents from the tourism sector, responsible for creating and executing the regional external tourism promotion, in accordance with the national tourism strategy defined by the Portuguese Tourism Board.

Our study started out with the compilation of the *Turigal* corpus, which for the moment contains a total of 1.285.764 words (469.873 words in the leaflets and 815.891 words in the web pages; 632,193 words in Portuguese and 653,571 in English). The corpus is included in the *Linguistic Corpus of the University of Vigo* (CLUVI), which has been built by the *Computational Linguistics Group of the University of Vigo* (SLI). Like Pearson (1998: 57), we believe a special purpose corpus does not have to be as big as a general purpose corpus. *Turigal* is considered to be sufficiently representative of all bilingual promotional materials published and distributed by Tourism Regions and Regional Tourism Promotion Agencies in 2007, the year the texts were collected.

The corpus contains complete written informative/promotional texts, of different size, in Portuguese and their translations into English. All texts come from brochures, tour guides and websites and are freely available. Most brochures and tour guides have no publishing date, but the ones which do are mostly from 2005 and 2006.

The case of authorship is particularly complex in our corpus, since most websites, brochures and tour guides do not mention the authors of texts or the translators. Websites frequently give the name of the company responsible for creating the websites, but do not indicate the name of people responsible for creating their texts. All texts are edited by Tourism Regions and Regional Tourism Promotion Agencies, the official entities responsible for internal and external tourism promotion of Portugal. This fact validates these texts as sources of terminological information. One assumes their authors are experts in tourism or someone with technical qualifications in the area of tourism.

All graphs, addresses and pictures have been removed from texts. Some information considered irrelevant for our terminographical purposes, such as proper names (people and companies) and addresses, were not included in the corpus. As far as hypertexts are concerned, these have been saved sequentially, according to the "site map", whenever this was available. To allow for the alignment of texts – the process of matching each phrase in Portuguese to its English translation – all texts which have originally been saved individually were joined together in a single text, creating a larger text or "super text".

The format chosen for storing the aligned parallel texts is an adaptation of the TMX format (Translation Memory eXchange), as this is the XML encoding standard for translation memories and parallel corpora, regardless of the application used (Savourel 1995). A translation memory is a term bank which holds the original and translated version for each sentence translated in the framework of a computer-aided translation system. All the aligned parallel texts in the CLUVI Corpus, including *Turigal*, are stored in TMX format and three translation strategies – omission, addition and reordering – have been encoded. The encoding procedures will be discussed in more detail shortly.

Due to the texts' format – brochures/leaflets and hypertext – the apparently simple task of storing them as plain text turned out to be time-consuming. On the one hand, the printed brochures had different formatting types (size, font, image and text layout, page configuration), different colours and quite often texts in Portuguese, English and other languages were kept side by side on the same page. On the other hand, working with web pages, though more productive in terms of the quantity of texts obtained, also required substantial post-processing, since many web pages had formatting codes which prevented easy access. Web pages had multiple input formats and in the process of text conversion some chunks of text would sometimes disappear and hence had to be manually typed. Moreover, texts which were not translated or were only in English had to be disregarded. Finally, some newly implemented sites were extremely slow and one could open only a page at a time, which slowed down the storage process.

Both types of texts – brochures/leaflets and hypertext – were then submitted to an Optical Character Recognition (OCR) programme and then to a spelling correction programme, in order to check the text generated by the OCR. The brochures/leaflets which remained practically illegible after the OCR was applied to them were manually typed.

The texts were then aligned in the TMX format. Each text has a header with information about the type of text (website or brochure/leaflet), its title in Portuguese and English, author, editor, year, translator and date of access to the website and to the brochure, whenever the latter had no indication of publishing date. As for the alignment itself, although a source sentence usually corresponds to a sentence in the translation, on some occasions one source sentence corresponds to two or more translated sentences or vice-versa, i.e., two or more source sentences correspond to one sentence in the translation. The alignment always starts with the source sentence, which means that the translation sentences were split or joined together to match the source sentence.

Thus, aligning a parallel corpus also entails its manual tagging, since translating is not a linear task. Translators can omit words, phrases or sentences from the source text, insert new ones as well as reorder segments or whole sentences in the translation. Encoding omissions, additions and reorderings in the corpus allows for the automatic search of these translation strategies and

facilitates their analysis. However, the purpose of the present research is not the study of translation strategies, but the use of an aligned parallel corpus for term extraction.

## 5. Methodological approach to the compilation of bilingual terminology

Research on the *Turigal* corpus went through the following steps. First, some background reading was done in order to become familiar with the subject area. Diverse introductory textbooks, encyclopaedia articles, glossaries and thesaurus, namely the *Eurovoc* thesaurus, were consulted. The Portuguese official website on tourism (*Turismo de Portugal*), the inventory of tourist offer, published by the Brazilian tourist authorities (Ministério do Turismo 2006), and the Universal Decimal Classification (UDC) were also checked.

Secondly, a thematic map, i.e. a semantic organization of the area of tourism was devised. This hierarchical ascription of terms to a specific branch of the thematic tree was done with a view to systematizing the subject area as well as clarifying the sense of each term in relation to each other. In the tourism term bank whenever there are different terms referring to a single concept, these are all grouped in the same term record.

At this stage, it is important to mention that the tourism industry is a fragmented one, for it overlaps with a series of other sectors (such as accommodation, transport, public services and food) which can subsequently be divided in other sectors. This interdisciplinary nature of tourism created problems in setting limits to the subject field, when it came to building the thematic tree. This tree had to reflect the fragmentary nature of subject fields; therefore experts on tourism were consulted.

The extraction of Portuguese term candidates from the *Turigal* corpus was done with the help of *kfNgram*, a computer programme which produced a list of the most frequent words in the corpus (essentially nouns), as well as the most frequent sequences of compound nouns, up to five words. Though extremely useful, the automatic extraction of term candidates still requires time-consuming post-processing. Terms were also identified with the aid of statistical term-extraction lists (which present terms with a higher degree of association) and, finally, reference works on tourism (which assist in the identification of key terms that were not found through the previous methods). Thus, the terminologist also selects terms which clearly belong to the area of tourism, even though they have a low frequency in the corpus. Hence the usefulness of doing a previous thorough analysis of reference books on the subject field, to check if some terms contained in those reference books also appear in our term bank.

In the tourism term bank, term entries group together equivalent terms in Portuguese and English. It is semantic content, defined in a single entry, which establishes a bridge between both languages. Each entry also groups all inter and intra linguistic synonyms. From a practical point of view, the awareness of linguistic and cultural differences between languages led us to accept, for example, loan words and paraphrases as inter linguistic equivalents, whenever faced with the inexistence of English terms.

Our research also involves the extraction of semantic relations. It is possible to identify semantic relations with the help of a pre-defined set of patterns, namely those of inclusion (hyponym-hyperonym) and meronymy-holonymy. The term record used in the *Termoteca* and in our term bank on tourism has the following XML structure:

each concept is tagged as *cc* and identified with a unique number (*ic*);

the semantic relationships with other concepts are tagged under *rs*;

the branch of the thematic tree each term refers to is identified with the tag *ct*;

the terminological information for each language is grouped with the tag *lg*, followed by the two letter ISO code for each language: *pt* (Portuguese) and *en* (English);

*texto\_def* stands for the definition of the term;

each term in a language represents a linguistic variant of the concept in that language (*var*), thus a single concept can have several variants;

the variants can be intralinguistic (synonymous terms) and interlinguistic (translated terms);

each variant includes the lemma of the term (*lema*), its grammatical category (*cat val*) and a context of use (*texto\_ex*) documented in the corpus (*obra; num*).

The XML example that follows is the term record 2220194, which refers to the term *paço* (palace).

```
<cc>
<ic>2220194</ic>
<rs tipo-rs="hiper">2220195</rs>
<rs tipo-rs="hiper">2220196</rs>
<rs tipo-rs="hiper">2220197</rs>
<rs tipo-rs="hiper">2220240</rs>
<def xml:lang="pt">
<texto_def>No definition</texto_def>
<fonto_def>No definition</fonto_def>
</def>
<ct st="tuvi">TURIGAL.B.1.2.2.1</ct>
<lg xml:lang="pt">
<var norm="s" tipo="com">
<lema>paço</lema>
<cat valor="m"/>
<ex>
<texto_ex>É nesta época que nascem ao lado do |Paço# novas estruturas: a Biblioteca, a Torre, a Via Latina e a Porta Férrea.
</texto_ex>
<fonto_ex>
<obra>CEN</obra>
<num>989</num>
</fonto_ex>
</ex>
<frec>
<fab>169</fab>
```

```

<vcorpus>17.11.08</vcorpus>
<palcorpus>1285764</palcorpus>
</frec>
</var>
</lg>
<lg xml:lang="en">
<var norm="s" tipo="com">
<lema>palace</lema>
<cat valor="com"/>
<ex>
<texto_ex>Some new buildings were built close to the |palace# during this period : the Library, the Tower, the Via Latina and the
Iron Gate.</texto_ex>
<fonte_ex>
<obra>CEN</obra>
<num>989</num>
</fonte_ex>
</ex>
<frec>
<fab>259</fab>
<vcorpus>17.11.08</vcorpus>
<palcorpus>1285764</palcorpus>
</frec>
</var>
</lg>
</cc>

```

Naturally, the interface of the term bank is more accessible to the term bank user, as shown in the following term record.

<i>Ref.</i> <u>2220194</u>	
<i>Campo temático:</i>	< <a href="#">TURIGAL.B.1.2.2.1/B.1.2.2.1.Arquitectura civil</a> >
<i>Relacións semánticas:</i>	[ <a href="#">hiper =&gt;2220999</a> ] [ <a href="#">hiper =&gt;2220195</a> ] [ <a href="#">hiper =&gt;2220196</a> ] [ <a href="#">hiper =&gt;2220197</a> ] [ <a href="#">hiper =&gt;2220240</a> ]
<i>Termo PT:</i>	<b>paço</b>
<i>Categoría:</i>	m
<i>Variante:</i>	com
<i>Frecuencia relativa:</i>	131.439362122
<i>Contexto de uso:</i>	É nesta época que nascem ao lado do <i>Paço</i> novas estruturas: a Biblioteca, a Torre, a Via Latina e a Porta Férrea. [CEN]
<i>Máis exemplos:</i>	<a href="#">No CLUVI</a> (algún dos exemplos do termo no corpus poden non corresponder ao concepto da ficha)
<i>Termo EN:</i>	<b>palace</b>
<i>Categoría:</i>	com
<i>Variante:</i>	com
<i>Frecuencia relativa:</i>	201.43665556
<i>Contexto de uso:</i>	Some new buildings were built close to the <i>palace</i> during this period : the Library, the Tower, the Via Latina and the Iron Gate. [CEN]

The terminological treatment of terms drawn from the *Turigal* parallel corpus includes three types of information: conceptual– the thematic tree (*Campo temático*) and the semantic relationships (*Relacións semánticas*); linguistic – gender, grammatical category (*Categoría*), lemma (*paço/palace*) and types of variants (*Variante*); and pragmatic – context of use (*Contexto de uso*) and relative frequency (*Frecuencia relativa*).

## 6. Conclusions

The main goal of this paper was to provide insight on our parallel corpus-based approach to the creation of a term bank in the subject field of tourism. The term bank is conceived as a tool for all those interested in finding information on the terminology of tourism, whether this information is a translation equivalent or simply the context of use for that term. Thus, it can become particularly useful for translators or other communication mediators who need to master the specialized lexical items and find their appropriate foreign language equivalents.

Theoretically, this project is firmly grounded on Teresa Cabré's Communicative Theory of Terminology, according to which terms, as a part of natural language, may acquire a specialized value in a particular context or communicative setting. It is context that creates the specialized value, hence our emphasis on a linguistic-textual theoretical and methodological approach. It is fundamentally a descriptive model that records language in use and therefore acknowledges the principle of conceptual variation.

Terminographical products, such as dictionaries, glossaries or term banks, are essentially ways to organize and make public the terminology of a particular subject area. Thus, we still feel the need to establish a link between terms and concepts and to structure the terminological information around concepts, not holding a normalizing intent, but with the aim of creating user-oriented products.

Our methodology seems in our opinion to successfully allow for the creation of multi-dimensional conceptual/semantic networks in a given subject area (Sager 1990: 160), while simultaneously recognizing the importance of the communicative setting in which terms are used, i.e. it embraces the concept of variation that characterizes all languages in use.

## References

- Biber, Douglas, Susan Conrad and Randi Reppen (2004) *Corpus Linguistics. Investigating Language Structure and Use*, Cambridge, Cambridge University Press.
- Bowker, Lynne and Jennifer Pearson (2002) *Working with Specialized Language: A Practical Guide to Using Corpora*, London and New York, Routledge.
- Cabré, Castellví M. Teresa (1999) *La Terminología: Representación y Comunicación. Elementos para una teoría de base comunicativa y otros artículos*, Barcelona, Institut Universitari de Lingüística Aplicada.
- Cabré Castellví, M. Teresa (2003) "Theories of Terminology. Their Description, Prescription and Explanation", *Terminology*, 9, no. 2: 163-199.
- Cabré Castellví, M. Teresa and Rosa Estopà Bagot (2005) "Unidades de Conocimiento Especializado: Caracterización Y Tipología" in *Coneixement, Llenguatge i Discurs Especialitzat*, M. T. Cabré and C. Bach (eds), Barcelona, IULA, Documenta Universitaria: 71-95.
- Cabré Castellví, M. Teresa, Anne Condamines and Fidelia Ibekwe-SanJuan (2007) "Introduction: Application-driven terminology engineering" in *Application-Driven Terminology Engineering*, Fidelia Ibekwe-SanJuan, Anne Condamines and M. Teresa Cabré Castellví (eds), Amsterdam – Philadelphia, John Benjamins Publishing Company: 1-17.
- Domingues, Celestino (1990) *Dicionário Técnico do Turismo*, Lisboa, Dom Quixote.
- Domingues, Celestino (1997) *Prontuário Turístico*, Lisboa, Instituto Nacional de Formação Turística.
- Eurovoc Thesaurus*, <http://europa.eu/eurovoc/> (accessed 1 March 2012).
- Gómez Clemente, Xosé María and Xavier Gómez Guinovart (dir.) (2006-) *Termoteca - Banco de Datos Terminológico da Universidade de Vigo*, Vigo, Universidade de Vigo, <http://sli.uvigo.es/termoteca/>.
- Gómez Guinovart, Xavier (dir.) (2003-) *Corpus CLUVI - Corpus Lingüístico da Universidade de Vigo*, Vigo, Universidade de Vigo, <http://sli.uvigo.es/CLUVI/>.
- Gómez Guinovart, Xavier and Elena Sacau Fontenla (2004) "Parallel corpora for the Galician language: building and processing of the CLUVI (Linguistic Corpus of the University of Vigo)", in *Proceedings of the 4th International Conference on Language Resources and Evaluation*, Teresa Lino et al. (eds): 1179-1182.
- Gómez Guinovart, Xavier and Ánxeles Torres Padín (2005) *Vocabulario Xurídico-Administrativo Galego-Castelán. Baileirado Terminológico Documentado no Corpus Lingüístico da Universidade de Vigo*, Vigo, Universidade de Vigo (Servicio de Normalización Lingüística).
- Gómez Guinovart, Xavier (2008) "A investigación en lexicografía e terminoloxía no Corpus Lingüístico da Universidade de Vigo (CLUVI) e no Corpus Técnico do Galego (CTG)" in *A lexicografía galega moderna. Recursos e perspectivas*, Ernesto González Seoane, Antón Santamarina and Xavier Varela Barreiro (eds), Santiago de Compostela, Consello da Cultura Galega / Instituto da Lingua Galega: 209-228, [http://webs.uvigo.es/sli/arquivos/sli\\_ilgo7.pdf](http://webs.uvigo.es/sli/arquivos/sli_ilgo7.pdf) (accessed 1 July 2012).
- González-Jover, Adelina Gómez and Chelo Vargas Sierra (2002) "Córpora comparables y paralelos para la detección de terminología bilingüe: su explotación y uso con herramientas informáticas" in *Actas del VIII Simposio Iberoamericano de Terminología*, CD-ROM, ISBN: 958-33-4022-7, Cartagena, Colombia, [http://www.riterm.net/actes/8simposio/gomez\\_vargas.htm](http://www.riterm.net/actes/8simposio/gomez_vargas.htm) (accessed 1 July 2012).
- Instituto Camões* (2010) <http://cvc.instituto-camoes.pt/traduzir/lextec.html> (accessed 1 March 2012).
- Kotler, Philip, John Bowen and James Makens (2004) *Marketing para Turismo*, trans. Cristina Belló, Elena Blásquez, Raquel Nieto, Madrid, Pearson Educación.
- McEnery, Tony and Andrew Wilson (2004) *Corpus Linguistics: An Introduction*, 2nd ed., Edinburgh, Edinburgh University Press.
- Ministério do Turismo (2006) *Manual do Pesquisador – Inventário da Oferta Turística: instrumento de pesquisa*, Brasília, Ministério do Turismo.

- Pearson, Jennifer (1998) *Terms in Context*, Amsterdam – Philadelphia, John Benjamins Publishing Company.
- Sager, Juan C. (1990) *A Practical Course in Terminology Processing*, Amsterdam, John Benjamins Publishing Company.
- Savourel, Yves (2005) *TMX 1.4b Specification*, Localisation Industry Standards Association.  
<http://www.lisa.org/standards/tmx/specification.html> (accessed 1 March 2012).
- Stubbs, Michael (1996) *Text and Corpus Analysis: Corpus-Assisted Studies of Language and Culture*, Oxford, Blackwell.
- Temmerman, Rita (2000) *Towards New Ways of Terminology Description: the Sociocognitive-Approach*, Amsterdam – Philadelphia, John Benjamins Publishing Company.
- Teubert, Wolfgang (2002) "The Role of Parallel Corpora in Translation and Multilingual Lexicography", in *Lexis in Contrast: Corpus-Based Approaches in Corpus Linguistics*, Bengt Altenberg and Sylviane Granger (eds), Amsterdam – Philadelphia, John Benjamins Publishing Company: 189-214.
- Teubert, Wolfgang (2005) "Language as an Economic Factor: The Importance of Terminology" in *Meaningful Texts: The Extraction of Semantic Information from Monolingual and Multilingual Corpora*, Geoff Barnbrook, Pernilla Danielsson and Michaela Mahlberg (eds), London – New York, Continuum: 96-187.
- Turismo de Portugal*, <http://www.turismodeportugal.pt> (accessed 1 March 2012).
- UDC: Universal Decimal Classification* (1993) 2nd edition, International medium edition, Milton Keynes, BSI Standards.
- Vintar, Spela (2001) "Using Parallel Corpora for Translation-Oriented Term Extraction", *Babel Journal*, 47:2, 121-132.  
<http://www2.arnes.si/~svinta/babel.rtf> (accessed 1 March 2012).

## Notas

[1] Within this research, a corpus is a set of texts of a given field, which have been written and used by specific groups of people and selected according to a specific purpose. In this case, the purpose is the extraction of bilingual terminology.

---

©inTRAlinea & Adonay Moreira (2014).

"A methodology for building a translator- and translation-oriented terminological resource", *inTRAlinea* Special Issue: Translation & Lexicography.

**Stable URL:** <https://www.intralinea.org/specials/article/2032>