[*Translating and the Computer* 29, November 2007]

Lost in specialised translation: the corpus as an inexpensive and under-exploited aid for language service providers

Gloria Corpas Pastor
University of Malaga, Spain

1. Introduction

Most translation activities nowadays involve the rendering of non-literary texts in the context of (semi-)specialised communication. The vast majority of translated texts usually belong to rather specialised and repetitive genres, which makes them particularly suitable for multilingual document management and CAT tools. In fact, translation memory systems (TMS) have proven an indispensable tool for professional translators in that they have been able to enhance the efficiency and cost-effectiveness of translation of voluminous repetitive texts without compromise of quality. Whereas for someone who works on translation of highly repetitive texts from a narrow domain TMS provide a smooth and problem-free solution, translators working on a vast range of different specialised texts may find that things are not as easy as they seem. First, parallel corpora[1] (TM files) may be either nonexistent or difficult to obtain. Secondly, the growth of TMS cannot catch up with the growth of bilingual corpora or bilingual websites. Neither can they catch up with being representative of dynamically developing domains where new terminology is being proposed on a daily basis. This is where other freely available — or relatively inexpensive — resources could come into play.

In this paper we will focus on the compilation of virtual or ad-hoc corpora[2] (i.e. corpora mined from electronic sources for a specific task) and the tools enabling their exploitation by translators. We will demonstrate how a step-by-step approach to building an adequate and/or representative corpus from resources in the Internet works in practice. Corpus design criteria and qualitative issues will be taken into account. For illustrative purposes, we will discuss a translation assignment of a specialised nature. Real examples will be presented that show how to mine a corpus and how to use it in order to meet translators' needs as far as terminology, documentation, target text conventions and other constraints are concerned. We will illustrate how language service providers can use freeware concordances to search for translation equivalents in comparable and parallel corpora.

2. Using corpora in translation practice

Resources and tools for translators lie within the realm of the broader field of translation technologies. As forwarded by standard introductory books on this topic (cf. Austermühl, 2001; Bowker, 2002; Quah, 2005), translation technologies cover machine translation systems, translation memory systems and localisation software, controlled languages, term extraction and terminology management systems, project management systems, globalisation management systems, corpus compilation and exploitation, as well as other electronic resources and information technologies.

[1] Parallel corpora include original texts and their translations into one or more source languages (bitexts).

[2] Henceforth, all occurrences of *corpus* or *corpora* will refer to this type of resources.

So far, however, translators have not fully explored the benefit of the variety of available resources to help them complete or improve their translation assignments. By way of example, whilst TMS are now indispensable in the daily work of professional translators, few translators are aware that in fact translation memories are nothing but samples of parallel corpora and it is the use of the wider availability of parallel corpora that can be beneficial to translators in a variety of additional ways. In addition, monolingual corpora mined from the web can also prove superior to any type of dictionary as they provide examples of how words or expressions are used and translated in context.

Despite the overarching remit of the European project LETRAC *(Language Engineering for Translators Curricula)*,[3] the use of corpora has only relatively recently come to the attention of those working in the field of translation training and translation practice. This is somewhat surprising, as corpora count among the translator's most important aids when faced with a specialised text. As stated in the EN 15038, the professional competences of translators include the "research competence, information acquisition and processing" and the "technical competence" (cf. 3.2.2.). In other words, translators should be able to acquire the additional linguistic and specialised knowledge necessary to understand the source text (ST) and to produce the target text (TT), as well as to use the appropriate research tools, apply the right information extraction strategies and to operate any technical resources required by a given translation project. Research and technical competences include, but are not limited to, the ability to compile, manage and exploit corpora mined from the Internet.

The advantages of using corpora in the work and training of translations have been shown by various studies (cf. Corpas Pastor, 2001; Bowker, 2002; Granger and Petch-Tyson, 2003; Zanettin, Bernardini and Steward, 2003, among others). Some of the principal advantages of using them are their objectivity, their reusability and multiple usage of a single resource. In addition, they allow easy access to and management of huge quantities of information in almost no time. Furthermore, we must consider that the development of our current information society has brought about a new demand for texts written in a variety of languages. Together with economic globalisation, this has resulted in a growing interest in the use of bilingual and multilingual corpora in the fields of machine and computer-assisted translation, language teaching, terminology and specialised language, natural language processing and information retrieval as well as, more recently, in training, practising and documentation as applied to translation (Corpas Pastor, 2002, 2004a, 2004b).

It is a well-known fact that nowadays translators are using corpora increasingly in their daily practice. This is especially true of those language service providers who are regularly translating texts belonging to specialised fields. Highly market-demanded specialised translation poses serious problems which go well beyond finding well-established terminology equivalents. New terms are coined everyday which may require the introduction of neologisms or (semi-)naturalised borrowings in a given target language (TL). Even if the right TL terms or multiword lexical units are identified, there remains the problem to select the right collocational profile. And, then, stylistic, register, textual and cultural problems may arise as well. These are the stumbling block of specialised translation that very rarely can be solved by just browsing dictionaries and standard databases or by looking up for information in a handful of parallel texts.

In the long run, building up target-language and parallel corpora in those specialised fields will improve the quality of the translated texts and significantly

[3] http://www.iai.uni-sb.de/iaien/en/letrac.htm. All URLs in this paper were checked 20 October 2007.

increase translators' productivity. All in all, corpora provide instant access to real usage, depict grammatical and collocational patterns, point to translation equivalents unavailable in existing lexicographic resources, and facilitate guidance to style and text conventions in both source and target languages.

3. Monolingual and Bilingual Concordancers

The term *concordance* or KWIC[4] is used in Corpus Linguistics to refer to the list of all the occurrences of a given *node* (key word) within its corresponding contexts. A *concordancer* is a software tool that queries a corpus in order to locate and display each instance of a given node and the context in which it occurs. Concordancers or concordancing systems are also called *corpus query systems* (CQS). Such systems usually display the search item in the centre of the text line with a few words to the right and to the left. Many Concordancers enable users to look up complex nodes, such as words with wildcats, multiword units, tags, and even regular expressions; sort the contexts by -n positions to the left and to the right; produce lists of words in alphabetical and rank order; and extract collocations, patterns and clusters (lexical bundles or n-grams). Instead of lines, some Concordancers also allow for contexts exceeding the sentence. The output can usually be directed to screen, disk and printers. Concordancers can be:

(a) monolingual or bi-/multilingual corpus-orientated,
(b) commercial or freeware,
(c) Windows-/Mac-orientated or cross-platform (Windows, Mac and Linux),[5]
(d) simple or modular; and
(e) set up and/or web programs.

For example, the *ConcApp Concordancing Programs*[6] is a non-commercial, monolingual concordancer suite for Windows operating systems (98, ME, NT / 2000, XP), that can be freely downloaded as execute or full set up program. It provides basic functionalities (word frequency lists; list of collocates; concordance searchers for words, phrases and derivatives) to process most European languages (English, Spanish, Italian, etc.), as well as Chinese, Japanese, Thai and Russian in Unicode. Another monolingual concordancer for Windows only is the *Multilingual Corpus Toolkit*[7], which supports many European and Asian languages. Freeware Concordancers for Mac only are *Conc* 1.76/1.80[8] and *Concorder X 1*.O.[9]

[4] KWIC is an acronym for "key word in context". On Corpus Linguistic terminology, see Baker et al. (2006).

[5] There are, though, some concordances that still run on MS-DOS, such as *MicroConcord* or *Kontext.* OUP *MicroConcord* can be downloaded from the URL: http://www.liv.ac.uk/~ms2928/software/index.htm. *Kontext* has a German interface, but the documentation is in English. It runs on plain ASCII text, .SGML and HTML formats. Searchers for strings of words and tags can be performed. An executable version can be downloaded from URL:http://www.burkhard-leuschner.de/software/kontext/kontindx.htm#downloadkontext.

[6] ConcApp Concordance and Word Profiler 4. [http://www.edict.com.hk/PUB/concapp/].

[7] http://personalpages.manchester.ac.uk/staff/Scott.Piao/research/DownLoad/download.htm.

[8] http://www.sil.org/computing/conc/.

[9] http://concorder-pro.softonic.com/mac.

Simple Concordance Program 4.0.9.[10] is another freeware, monolingual concordancer, both Windows-orientated (95, 98, ME, NT / 2000, XP) and Mac-orientated (Apple MacOS X on PowerPC and Intel). Apart from usual features such as word lists; search words, derivatives, phrases and patterns; sorting of contexts, frequency counts, etc., this piece of software generates both KWIC and Line-Based concordances, allows stoplists, includes built-in alphabets for many languages (English, Polish, Greek, Russian, etc.) as well as an alphabet editor for other languages. *AntConc 3.2.1*[11] is a non-commercial freely downloadable concordancer for Windows, Mac and Linux. This versatile software features several tools, which display lists of words and keywords (Word List, Keyword List), list, sort and search for lexical bundles (Clusters, N-Grams, Collocates), generate lines in KWIC format (Concordance), indicate the position of the keyword within a given corpus (Concordance Plot), allow the user to have access to the whole source file or corpus (File View). Other freeware cross-platform concordancers are *aConCorde* 4.0.1.[12], with a bilingual English and Arabic interface (full Arabic support); *Dexter,*[13] which also offers the possibility of manually annotating the corpus and later perform tags searchers as well; *TextSTAT*[14] includes a search engine to download webpages and put them in a TextSTAT-corpus.

Commercial monolingual concordances usually offer trial periods or demos that just restrict the number of hits or prevent results to be saved or printed. All in all, they are not superior to non-commercial software. In fact, both feature roughly the same suite of tools and support most European languages. For instance, *AntConc* and *Oxford WordSmith Tools* 3.1. and 4.0.[15] are very similar. *WordSmith Tools* (WS) includes several utilities: (a) WordList, which displays lists of words and clusters in alphabetical or frequency order, and calculates sentence and word length; (b) Concord, which generates KWIC lines which can be sorted by n-left or n-right, centre or tags, as well as patterns, clusters and collocates which can be re-sorted and displayed in dispersion plots; and (c) KeyWords, which extracts key words and key keywords as regards a given reference corpus. In addition, WS can be run on parallel corpora as well, as it contains Viewer and Aligner — a basic utility for producing an aligned version of two or more texts, with alternate sentences or paragraphs from each of them. Another relevant difference is the utility WebGetter which enables the user to compile an a corpus from the Internet, by selecting a search engine which locates the first 100 sources and sends a robot to download each page provided it meets the user's requirements, as defined in settings. This feature turns WS into a modular type of concordancing software, as it allows both for corpus exploitation and compilation/management (see below).

Concordance[16] is a commercial monolingual concordancer for Windows only[17]. It works with nearly all languages supported by Windows, and includes lemmatisation, user-definable alphabet, reference system, contexts and flexible selecting, search and sorting of words, phrases, regular expressions, etc.; it saves and export concordances as

[10] http://www.textworld.com/scp/. (Updated version released on 28th July 2007).

[11] http://www.antlab.sci.waseda.ac.jp/software.html.

[12] http://www.andy-roberts.net/software/aConCorde/index.htm#download.

[13] http://www.dextercoder.org/index.html.

[14] http://www.niederlandistik.fu-berlin.de/textstat/software-en.html.

[15] http://www.lexically.net/wordsmith/index.html. A free, beta version *(WordSmith Tools* 5.0) has been released in June 2007.

[16] http://www.concordancesoftware.co.uk. A 30-day trial can be downloaded.

[17] *Text Analysis Computing Tools* (TACT) is an older commercial concordancer which runs on MS-DOS. It can also work within Windows 3.11 and 95 / 98 but not with NT or beyond. URL: http://www.chass.utoronto.ca/tact/. There is a web version *(TACTweb* 0.5.) which can be accessed at the URL: http://kh.hd.uib.no/tactweb/homeorg.htm.

.txt and .html files or as web concordance. In addition to the usual functionalities, *PhraseContext* 1.0.2.[18] performs nice statistical lexical analysis, such as T-score, Z-score, MI (Mutual Information), Log-likelihood, lexical density, or sentence and words lengths. Results in .txt or .xml formats can be displayed numerically and, in some cases, also graphically. Another well-known commercial software for Windows only is *MonoConc Pro* (MP 2.2.)[19], which supports all European languages plus Chinese, Japanese and Korean. Its distinctive features include Context Search, Regular Expression Search, Part-of-Speech Tag Search, Collocations, and Corpus Comparison.

A bilingual or multilingual concordancer is a program for parallel corpora, i.e. corpora of source texts and their translations into other languages. As a rule, this kind of software requires input aligned at sentence level. Most bi-/multilingual concordances are commercial. A well-known example is *ParaConc* 0.9x[20], the multilingual version of *MonoConc Pro.* It can analyse up to four languages in parallel (one source text corpus and up to three target corpora). *ParaConc* is a comprehensive piece of software that provides customary functionalities plus parallel search options and translation equivalents utilities for words and collocations.[21] *Multiconcorcd*[22], a rather modest program, came out as a by-product of the Lingua project. It can handle one source corpus and one target corpus at a time. It supports most European languages, Venda, Zulu and other alphabetic languages. Words, multiword units and truncated words with wildcats can be used as search items. Although users can upload their own corpora, the program includes a Parallel Texts Library,[23] i.e. a list of downloadable parallel texts from the European Parliament. Another useful tool is the *Translation Corpus Explorer* (WebTCE)[24]. It is a freeware web-based multilingual concordancer which has been designed and implemented within the *English-Norwegian Parallel Corpus* (ENCP) project.[25] The WebTCE can be used to query a subcorpus of ENCP (non-fiction open texts), as well as other sample small, freely accessible multilingual corpora.[26] . The user enters the search string, selects the language, text and data base.

Bilingual and multilingual concordancers are scarce, possibly due to the fact that (a) translation memory systems already integrate alignment, concordancing, and terminology management, among other functionalities; and (b) some word aligners can be used as basic bilingual concordancers as well.[27] However, they can be useful for translators who cannot afford to get a commercial TMS, or, else, want to perform bilingual searches for matches at sub-sentential levels.

[18] http://www.hjkm.dk/PhraseContext/#language. A free 35-day evaluation version with full functionality can be downloaded.

[19] http://athel.com/mono.htm#monopro. A 30-day trial demo with limited functionality can be downloaded.

[20] http://athel.com/para.html. It is possible to download a demo with limited functionality restricted to 150 hits which does not allow to save or print the results.

[21] On ParaConc and bilingual concordancing systems see Barlow (2004).

[22] http://artsweb.bham.ac.uk/pking/multiconc/lingua.htm. A demo can be downloaded from the URL: http://www.copycatchgold.com/MulticoncordDemo.html.

[23] http://artsweb.bham.ac.uk/pking/multiconc/multdata.htm.

[24] http://khnt.hit.uib.no/webtce.htm.

[25] http://www.hf.uio.no/ilos/forskning/forskningsprosjekter/enpc/.

[26] The available and searchable texts are the *WHO Annual Report 1995* (English-German-French), and the EU Multilingual (English-German-French) and Bilingual Corpora (English-German, English-French, English-Spanish). Texts in English are POS tagged.

[27] One such system is *Plug Word Aligner* (PWA). It also requires sentence alignment of the bilingual corpus prior to analysis. A demo with limited functionality can be downloaded from the URL: http://stp.ling.uu.se/plug/pwa/. It runs on Windows and Linux.

Concordancing systems with a modular architecture comprise a corpus management module and a corpus query module, i.e. they have all the options of standard corpus software plus the corpus management functionality: compilation, uploading/downloading, coding, selection and storing tasks. *Corpus Presenter* 10.0. is a Windows-only monolingual concordancer that is commercialised as a companion to a book (Hickey, 2003).[28] The system generates concordances and word lists (.rtf text, line list, single-line and multi-line returns) as well as reverse dictionaries of words in texts. It performs all kinds of text retrieval tasks and exports statistics returns to tables and charts; and it supports most languages and file formats, including multi-media. In addition, this collection of programs is designed to manage corpora by means of several utilities: Corpus Presenter Make Tree, File Manager, Find Text, Text Tool, and List Processor. Corpora can also be easily compiled by using the utility Net Browser which allows to clone any Internet webpage and save it to disk. *Sketch Engine*[29] is a similar modular, web-based system. It includes pre-loaded corpora of Chinese, English, French, German, Italian, Japanese, Portuguese, Spanish and Slovene. In addition, the WebBootCat utility enables users to mine corpora from the Internet, extract key words and domain-specific terminology; and the CorpusBuilder allows corpus uploading. It is a query system for pre-processed, lemmatised and part-of-speech tagged corpora. It contains the customary corpus functionalities, plus word sketches (i.e. one-page automatic, corpus-based summaries of the grammatical and collocational patterns of words), thesaurus and 'sketch differences' between near-synonyms.

Uplug Web 0.1.[30] is a modular monolingual and multilingual web concordancer. Users must register to get free access to the system. This suite of tools comprises two modules. The Corpus Manager allows to upload, update, inspect or remove monolingual and parallel corpora; whereas the Task Manager includes utilities to process the corpora, such as pre-processing tools (sentence splitter, tokeniser and external part-of-speech tagger and shallow parsers), word, phrase and sentence aligner, indexing and corpus query. A local batch system queues the jobs and sends them by e-mail to the users. Similar web-based corpus management and query system for user-defined corpora are *COSMAS II* [31], *Web Concordancer*[32] and *Turbo Lingo*[33].

Finally, there are web concordancing systems for specific corpus query: *BNC Simple Search*[34], *Cobuild Direct Concordance and Collocation Sampler*[35], *LDC Online*[36], *Online Concordancer*[37], *Online KWIC Concordancer*[38], *Corpus de Referencia*

[28] http://www.uni-essen.de/CP/. A restricted version *(Corpus Presenter Lite)* can be freely downloaded. It contains all the functions of the full program except for the third level of text retrieval which is used for advanced tasks.

[29] http://www.sketchengine.co.uk. Users can register for a 30-day free trial account.

[30] http://stp.ling.uu.se/cgi-bin/joerg/Uplug.

[31] http://www.ids-mannheim.de/cosmas2/.

[32] http://www.edict.com.hk/concordance/.

[33] http://www.staff.amu.edu.pl/%7Esipkadan/lingo.htm.

[34] http://thetis.bl.uk/. Free, but limited output.

[35] http://www.collins.co.uk/Corpus/CorpusSearch.aspx. Free, but limited output.

[36] http://www.ldc.upenn.edu/ldc/online/index.html.old. Free service for LDC-members. A guest account allows access to the Brown corpus, the TIMIT speech corpus and the Switchboard corpus.

[37] http://www.lextutor.ca/concordancers/concord e.html. Several corpora are freely searchable: Brown, BNC (Written, Spoken, Med), University Word List, 2000 list corpus, Focus on Vocabulary, US TV Talk, Language & Teaching, Anne-Green Gables, Call of the Wild, Starr Report, Learner (Student, Teacher) and JPU Learner.

[38] http://ysomeya.hp.infoseek.co.jp/. It searches the one-million word Business Letters Corpus (BLC2000).

del Español Actual (CREA)[39] and *Corpus del Español*[40], among others. Other sample freeware monolingual systems exploit the Internet as a gigantic corpus, such as *WebCorp*[41], *KWICfinder*[42] or *TAPoRware* 2.0[43]. Space restrictions prevent us from describing them here.

4. Case Study[44]

In this section we will demonstrate how to use a concordancer for specialised translation. We have chosen the beta version of WordSmith Tools 5.0 that can be downloaded as a demo with full functionality.

The source texts are two new test methods to be included in a draft Commission Directive[45] amending, for the purpose of its adaptation to technical progress, for the 30th time, Council Directive 67/548/EEC of 27 June 1967 on the approximation of the laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances: *B. 44. Skin absorption: in vivo method* and *B. 45. Skin absorption: in vitro method.*

Upon request of the European Commission, the European Centre for the Validation of Alternative Methods (ECVAM) has produced B.44. and B.45 documents to be included as part of Annex V of the future Directive. At present the draft document is under public survey for comments until 04 July 2007. Once the future Directive and annexes are published in the Official Journal of the European Communities, all member states must transpose the Directive into internal law within a certain period of time. Transpositions and implementation of EC legislation into Spanish internal law may differ from EC Spanish translated documents. For this sample translation project, the scopos[46] would be to translate both technical annexes into peninsular Spanish as implementation of the harmonised EC Directive into internal law (Royal or Presidential Decree). Source and target text types belong to the technical-scientific type of specialised communication; the source text form is EC Directive technical annex, whereas the target text form would be a technical annex of a Spanish Royal Decree or Order.

4.1. Corpus compilation

Once the scopos of the translation project is known, and the appropriate target text type and text forms are identified, the next step towards TL corpus design and

[39] http://corpus.rae.es/creanet.html.

[40] http://www.corpusdelespanol.org/.

[41] http://www.webcorp.org.uk/

[42] http://www.kwicfinder.com/.

[43] http://taporware.mcmaster.ca/.

[44] This case study is a partial, revised version of a training seminar given at the European Commission's Directorate-General for Translation (Spanish Translation Unit C 04), in Brussels and Luxembourg, 21-24 February 2007.

[45] http://ec.europa.eu/enterprise/tbt/index.cfm?fuseaction=Search.viewDetail&Country_ID= EEC&num=151&dspLang=EN&nextpage=1&basdatedeb=&basdatefin=&baspays=EEC&baspavs2=&basnotifnum=151&basnotithian2=&bastypepays=ANY&baskeywords=&fromform=viewBasic&FromWhatNew=WhatNew.viewOld&whatNum=36&whatYear=2007. Date arrival: 04 May 2007. Final date for comments: 04 July 2007.

[46] The scopos of a given translation project refers to the target communicative situation as a whole. It is the aim, goals and purposes of the intended translation, its text type and text form, which might differ from the source text, as well as any other client's specifications or requirements (the translation's brief).

compilation is a content analysis of the ST. Next, TL descriptors for indexation have to be identified. After that, appropriate information retrieval systems (IRS) have to be identified and queries evaluated. Relevant documents will be then downloaded, pre-processed and stored as a corpus.

4.1.1. ST content analysis and TL key words

The concordancer can be used for ST content analysis, as it can provide a list of words of the ST ranked by frequency. A cursory look at the most frequent ST content words (see appendix I) will show that they tend to cluster around the following key topics:

(a) Laboratory tests:

- *test, method(s), study(-ies), sampling, testing, analysis, experiment(s), laboratory, protocol, procedure, evaluation, objectives, techniques, validation, values, variance, calculation, dose(s), measure, data, number, rate, percentage, results, conditions, hour(s), intervals, temperature;*

- *exposure, vitro, vivo, absorption, solution, preparation, fluid, diffusion cell(s);*

- *species, animal(s), rat(s), monkeys, pigs, rabbits, cage(s), carcass(es), sex, male(s), mammalian, feeding, food, excreta, urine, faeces, human(s);*

(b) Biochemistry, cell and molecular biology:

- *cell(s), tissue, skin, percutaneous, dermal, dermatome, dermis, epidermis, epidermal, stratum, layer(s), corneum, surface, sebum;*

- *metabolism, metabolite(s), metabolically, metabolic, metabolisation, enzyme, lipophilicity, solubility, pharmacokinetics;*

(c) Ecotoxicity and dangerous substances:

- *chemical, chemical(s), substance(s), physicochemical, radiochemical, radioactive, toxicity, toxicology, corrosive, dioxide, ecotoxicology, enviromental, contaminating;*

On closer inspection the main ST topic can also be deduced from the frequency-based content analysis. The 10 most frequent words by descending order are the following:

1.	skin	6.	preparation
2.	test	7.	exposure
3.	absorption	8.	vitro
4.	method	9.	animals
5.	substance	10.	fluid

Table 1. *10 most frequent ST words.*

A tentative main topic would be “method for test skin absorption of (chemical) substances on animals”.

Taking into account ST content clusters and main topic, possible TL indexation descriptors for documentary search could be:

- “métodos alternativos” + “experimentación animal”

- “ensayo” + “in vivo/vitro”

- “cultivo celular” + “dérmico” + “sustancias químicas” + “toxicidad”

4.1.2. Selection of IRS and relevant documents

After identification of descriptors for searches, general search engines such as *Google, Altavista* or *Yahoo* can be a good starting point to locate useful documents in the surface web. However, relevant information in the invisible web has to be mined by by general metasearch and multisearch engines *(Metacrawler*[47]*, Kartoo*[48]*, Vivísimo*[49]), as well as specialised engines *(Buscasalud*[50] *, Scirus*[51]*),* directories, gateways and portals *(Buscamed*[52]*, Fisterra*[53]*),* forums *(MedTrad*[54]*),* bibliographical databases *(SciELO-Scientific Electronic Library Online*[55]*, Doyma*[56]*, Free Medical Journals*[57]*, La Biblioteca Cochrane Plus*[58]*, Mediagraphic Literatura Biomédica*[59]*)* and legal databases *(Westlaw*[60]*, Aranzadi*[61]*, EurLex*[62]*).*

In addition, various kinds of reliable websites can be also inspected for relevant information in Spanish: (a) official bodies, such as the Spanish Ministry of Health and Consumer Affairs[63] that provides access to relevant technical reports and semi-specialised handbooks, and to national, autonomous and EC legislation on dangerous substances and chemical preparations; (b) scientific networks, such as REMA (Red Española para el Desarrollo de Métodos Alternativos a la Experimentación Animal)[64], 3Erres (Red Iris-Alternativas a la Experimentación Animal)[65] and GTEMA (Grupo de Trabajo Especializado en Métodos Alternativos)[66]; (c) learned societies, e.g. SECAL (Sociedad Española para las Ciencias del Animal de Laboratorio)[67] and AETOX (Asociación Española de Toxicología)[68], which may provide specialised and semi-specialised documents, conference proceedings, free journals, databases, reports and other relevant resources.

Documents are then pre-processed, cleaned of .html and other codes, and converted into .txt format for corpus query. The overall corpus size is 4.545.990 words (types: 35,314; tokens: 708,919; TTR: 4.98; standardised TTR: 33.44). The TL corpus matches the ST as shown by their frequency wordlists (see Appendix I and II). It is further classified into five subcorpora:

[47] http://www.metacrawler.com.
[48] http://www.kartoo.com.
[49] http://vivisimo.com.
[50] http://www.buscasalud.com.
[51] http://www.scirus.com.
[52] http://www.buscamed.com.
[53] http://www.fisterra.com.
[54] http://www.medtrad.org.
[35] http://www.scielo.org.
[56] http://www.doyma.es.
[57] http://www.freemedicaljoumals.com/
[58]http://http://212.169.42.7/newgenClibPlus/WebHelpSpecific/Using.htm.
[59] http://medigraphic.com/inicio.htm.
[60] http://www.westlaw.es. Comercial database.
[61] http://www.aranzadi.es. Comercial database.
[62] http://eur-lex.europa.eu/es/index.htm.
[63] http://www.msc.es/en/home.htm.
[64] http://www.remanet.net/ingles/remaingles.htm.
[65] http://www.rediris.es/list/info/3erres.es.html.
[66] http://tox.umh.es/aetox/Grupos/GTEMA/index.html.
[67] http://www.secal.es/parallel/para-frame.html.
[68] http://www.aetox.es/.

1.- *Community legislation* (CL): Spanish versions of Council Directive 67/548/EEC and 2004/73/CE, its 19th adaptation to technical progress, plus all Directives which have modified Annex V[69]. Size: 60,516 words (types: 15,707; tokens: 26,704; TTR: 6.93; standardised TTR: 31.26).

2.- *National legislation* (NL): Spanish national regulations on new substances and classification, packaging and labelling of dangerous substances which transpose and implement the EC harmonised legislation, i.e. Royal Decrees 363/1995, 700/1998, 507/2001, and 99/2003, plus the Presidential Orders that modify the annexes[70]. Size: 46,488 words (types: 9,312; tokens: 36,641; TTR: 6.81; standardised TTR: 31.66).

3.- *Specialised texts* (SP): Research papers from learned journals on the topics covered by the STs, test methods and protocols, technical reports in peninsular Spanish. Size: 91,289 words (types: 13,486; tokens: 75,108; TTR: 7.70; standardised TTR: 34.20).

4.- *Semi-specialised texts* (SEM): Handbooks, technical papers and notes, conference proceedings, book chapters on toxicity, substances, dermatology, skin tests, etc. Size: Size: 23,294 words (types: 12,679; tokens: 37,636; TTR: 9.21; standardised TTR: 35.87).

5.- *Divulgative texts* (DIV): websites on dermatology, animal experiments, promotional material on alternative test methods, etc. Size: 24,403 words (types: 2,051; tokens: 32,830; TTR: 18.75; standardised TTR: 41.85).

A small, auxiliary parallel text (AUX) has been compiled from EC legislation and bilingual abstracts of scientific papers. Size: 8,134 words (types: 6,155; tokens: 10,943).

4.2. Corpus query

In this section we provide a practical approach to exploiting the corpus for translation. With this aim, we have chosen two excerpts. The first sample comes from the test method B.44. The second sample includes decontextualised sentences that describe figure 1 in B.45. Specific searchers and query strings will be spelt out in order to demonstrate how corpus data can be used for various translation subtasks. Both the comparable TL corpus and the parallel corpus will be consulted. Finally, a tentative translated version for each excerpt will be provided.

[69] Directives 88/302/CE, 1992/69/CE, 1996/54/CE, 1998/73/CE, 2000/32/CE, 2000/33/CE, and 2001/59/CE.

[70] Orders of 13 September 1995, 21 February 1997, 30 Junel998, 11 September 1998, 16 July 1999, 5 October 2000, 5 April 2001.

4.2.1. Excerpt no. 1

ST1

B. 44. SKIN ABSORPTION: *IN VIVO* METHOD

1 METHOD

This testing method is equivalent to the OECD TG 427 (2004).

1.1 INTRODUCTION

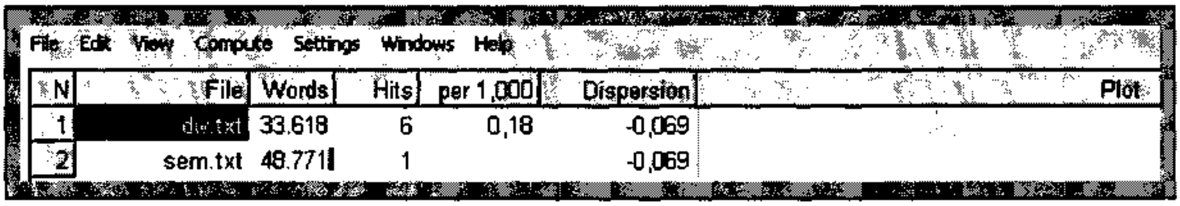
Exposure to many chemicals occurs mainly *via* the skin whilst the majority of toxicological studies performed in laboratory animals use the oral route of administration. The *in vivo* percutaneous absorption study set out in this guideline provides the linkage necessary to extrapolate from oral studies when making safety assessments following dermal exposure.

A substance must cross a large number of cell layers of the skin before it can reach the circulation. The rate-determining layer for most substances is the *stratum corneum* consisting of dead cells. Permeability through the skin depends both on the lipophilicity of the chemical and the thickness of the outer layer of epidermis, as well on factors such as molecular weight and concentration of the substance. In general, the skin of rats and rabbits is more permeable than that of humans, whereas the skin permeability of guinea pigs and monkeys is more similar to that of humans.

The first excerpt comes from the introductory section of the test method. With the utility Viewer & Aligner (option: multiple-text aligned), the AUX corpus is examined for TL discourse conventions and candidate translation equivalents of “this”, “equivalents” and “OECD TG”. Contrary to current TMs, whose matching algorithms operate at sentence or clause level defined by comas, colon or semi-colon, the bilingual concordancer enables the user to identify translation equivalents at sub-sentence levels. So, the AUX corpus reveals that in text openings dealing with test methods, “This” (in capital letter) is usually rendered at “El presente”, “is equivalent” can be translated as “reproduce” and “es análogo”, whereas “the OECD TG” can be alternatively conveyed in Spanish as “la TG de la OECD”, “las directrices de la OECD TG” and “las directrices del documento OECD TG”.

1. Method. **This** testing method **is equivalent** to the **OECD TG** 430 (2004).
   Method. **E1 presente** método de ensayo **reproduce** las **directrices del documento OCDE TG** 430 (2004).
2. **This** bioconcentration test **is equivalent** to the **OCDE TG** 305 (1996). According to the
   El presente método de bioconcentración **reproduce las directrices de la OCDE TG** 305 (1996). Según la
3. **This** testing method for subchronic oral toxicity **is analogous** to **OECD TG** 409.
   El presente método de ensayo de toxicidad oral subcrónica **reproduce** el **documento OCDE TG** 409.
4. **This** testing method is **equivalent** to the **OECD TG** 423**.**
   El presente método es **análogo** a la **TG** 423 **de la OECD**

The use of Latín expressions is a common feature of scientific texts. SL and TL conventions may differ as regards specific domains and genres. In excerpt no. 2 the multiword term *stratum corneum* is used. The TL corpus reveals that the Latin expression is characteristic of semi-specialised and divulgative texts (usually as a technical term in round brackets), as shown in the concordances and dispersion plot.



N	File	Words	Hits	per 1,000	Dispersion	Plot
1	div.txt	33.618	6	0,18	-0,069	
2	sem.txt	48.771	1		-0,069	

Fig. 1. *Dispersion plot for* stratum corneum.

1 Este estrato protege a la piel ante las acciones de las soluciones acuosas. **Stratum corneum** - Estrato córneo *(5)* El estrato córneo está formado por célu c:\aslib sample corpus\sem.txt

2 os constituyen de un 10 a un 30 por ciento del volumen total de la capa córnea **(stratum corneum).** Esto equivale a una proporción en sustancia intercelular sobre c:\aslib sample corpus\div.txt

3 turn granulosum . 3 Exocitosis . 4 Membranas lipídicas dobles . 5 Células del **stratum corneum** . La barrera de permeabilidad . Los lípidos epidérmicos consti c:\aslib sample corpus\div.txt

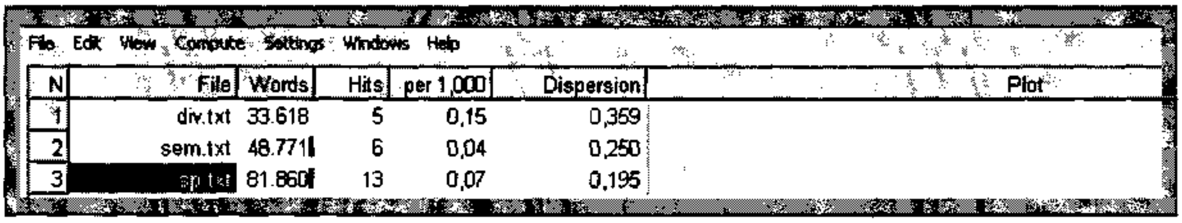
4 Microfotografía electrónica de células comeas defurfuradas. . La capa córnea **(stratum corneum)**.. La capa exterior de la epidermis (cutis superficial), c:\aslib sample corpus\div.txt

5 cuando las células ya no son visibles. Capa córnea (stratum corneum). El **stratum corneum** (del latín cornea = callosidad) es la capa más superficial de la c:\aslib sample corpus\div.txt

6 y entre sí. Los límites entre las células ya no son visibles. Capa córnea **(stratum corneum).** El stratum corneum (del latín cornea = callosidad) es la cap c:\aslib sample corpus\div.txt

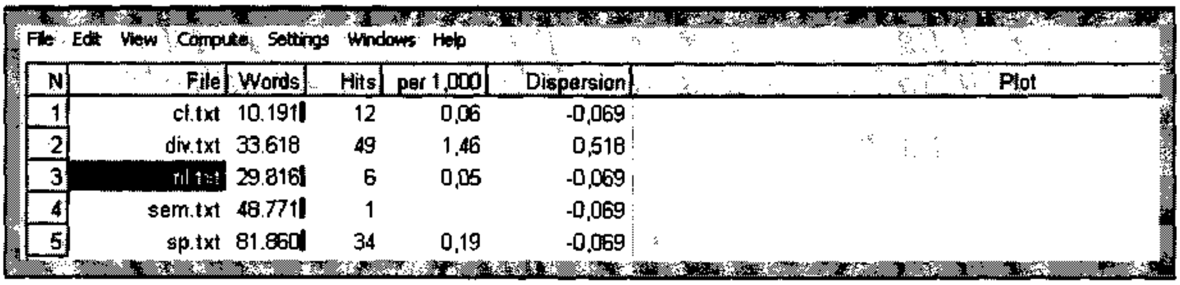
7 llamados desmosomas. Se distinguen en total cinco capas: Capa comea **(stratum corneum).** Capa lúcida (stratum lucidura).. Capa granulosa (stratum c:\aslib sample corpus\div.txt

However, in specialised communication the synonym *estrato córneo* is usually selected, whereas *capa córnea* can be found in all subcorpora, including NL and CL, as can be seen in their distribution along the various subcorpora.



N	File	Words	Hits	per 1,000	Dispersion	Plot
1	div.txt	33.618	5	0,15	0,359	
2	sem.txt	48.771	6	0,04	0,250	
3	sp.txt	81.860	13	0,07	0,195	

Fig. 2. *Dispersion plot for* estrato córneo.



N	File	Words	Hits	per 1,000	Dispersion	Plot
1	cl.txt	10.191	12	0,06	-0,069	
2	div.txt	33.618	49	1,46	0,518	
3	nl.txt	29.816	6	0,05	-0,069	
4	sem.txt	48.771	1		-0,069	
5	sp.txt	81.860	34	0,19	-0,069	

Fig. 3. *Dispersion plot for* capa córnea.

Orthographical variants of scientific terms can also pose problems in specialised translation. The ST word *lipophilicity* has two possible cognates in Spanish: *lipofilia* (609 hits in Google) and *lipofilicidad* (141 hits in Google). A search in the TL corpus with the wordstem plus a wildcat [lipof*] reveals no instances of *lipofilicidad,* contrary to *lipofilia*, that occurs 4 times in the legal subcorpora (CL and NL), twice in each of them.

Finally, KWIC lines can also show cultural differences between the two scientific communities involved in the translation process. For example, whereas in English substances reach the circulation, in scientific Spanish the word *circulación* does not usually appear in isolation, but either as part of the adverbial multiword unit *(en circulación)* or else followed by classifying adjectives to denominate different types of blood circulation: *circulación (sanguínea) sistémica, circulación enterohepática, circulación general,* etc. A common pattern is the colligation[71] with nouns which denote substances or tissues which are permeable *(solución, preparado, soluto, sustancia, dosis, fármaco)* as subject plus verbs of movement *(alcanzar, llegar, llevar, entré)* followed by *circulación sistémica* as direct object or else as prepositional object with *a, hacia, hasta, en,* as appropriate.

1 que pueden biotransformar. los tóxicos.. Es necesario mencionar también la **circulación enterohepática.**. Los tóxicos

2 re y el testículo, algunas generaciones de espermatogenias están expuestas a la **circulación general** y otras no. Si hay

3 os (efecto de primer paso). Los xenobióticos inhalados se distribuyen por la **circulación general** hasta llegar al hígado.

4 porte de los tóxicos y sus metabolitos. La sangre es un órgano líquido **en circulación** que lleva a las. células el oxígeno

5 gadas capas de células y . una distancia de mieras entre el aire alveolar y la **circulación sanguínea sistémica.** Ello hace

6 e . atravesar la barrera **tisular** sino también como su **llegada** ulterior a la **circulación sanguínea.**. Absorción pulmonar.

7 físico-químicas del **soluto.** - la razón por la cual la **circulación sanguínea local** aclara el soluto desde el **tejido** hasta la

8 cual la circulación sanguínea local aclara el **soluto** desde el **tejido** hasta la **circulación sistémica.** La absorción, sin

9 trato córneo hacia la dermis . * difusión del **fármaco** desde la dermis hacia la **circulación sistémica.**.. dQ/dt = D* Kp/h

10 la difusión de la **sustancia** es más importante, pero el **fármaco** no **alcanza** la **circulación sistémica.** El estrato córneo

11 el **fármaco** difundirá desde el estrato córneo hasta la hipodermis y **alcanzará** la **circulación sistémica.** Los parches

12 a biodisponibilidad es la fracción de una **dosis** administrada que entra en la **circulación sistémica.** Cuando no hay

A tentative Spanish version that takes into account the specifications of the scopos and the data obtained from the corpus is provided below:

[71] Colligation is the grammatical counterpart of lexical collocation. Hoey (2006[2005]: 43) defines colligation as:

1 the grammatical company a word or word sequence keeps (or avoids keeping) either within its own group or at a higher rank;

2 the grammatical functions preferred or avoided by the group in which the word or word sequence participates;

3 the place in a sequence that a word or word sequence prefers (or avoids).

TT1

B. 44. ABSORCIÓN DÉRMICA: MÉTODO *IN VIVO*

1 MÉTODO

El presente método de ensayo reproduce las directrices del documento OCDE TG 427 (2004).

1.1 INTRODUCCIÓN

Aunque la exposición a un elevado número de sustancias químicas se produce principalmente por absorción a través de la piel, en la mayoría de los estudios toxicológicos realizados con animales de laboratorio predomina, sin embargo, la administración por vía oral. El estudio de absorción percutánea *in vivo* descrito en esta directriz proporciona el proceso de enlace necesario para extrapolar los resultados obtenidos en los estudios de toxicidad oral a la hora de evaluar la exposición dérmica.

Una sustancia atraviesa muchas capas de células epiteliales antes de alcanzar la circulación sistémica. La capa córnea, compuesta de queratocitos, constituye un índice determinante para la mayoría de las sustancias. La permeabilidad cutánea depende tanto de la lipofilia de la sustancia química como del grosor de la epidermis y de otros factores como son el peso molecular y la concentración de la sustancia. En general, la piel de ratas y conejos es más permeable que la piel humana, mientras que la permeabilidad de la piel de cobayas y monos se asemeja más a la de los seres humanos.

4.2.2. Excerpt no. 2 (B. 45)

ST2

- Glass diffusion cell.

- (Auto)pipette/syringe sampling into scintillation or HPLC vials.

The second sample contains just two isolated clauses from a description of a figure. They have been selected for further illustrative purposes. For instance, the translator may be unsure about the meaning and possible translation equivalent of *cell* within the multiword unit *diffusion cell.* Therefore, a search is performed on the cognate *difusión* in collocation with the stem c*el*:*

1. resultados de **difusión** determinados mediante la **celda** de Franz y los estudios in vivo, permite q

2. Se han descrito en la literatura muchas **celdas de difusión.** Idealmente, un sistema in vitro debe

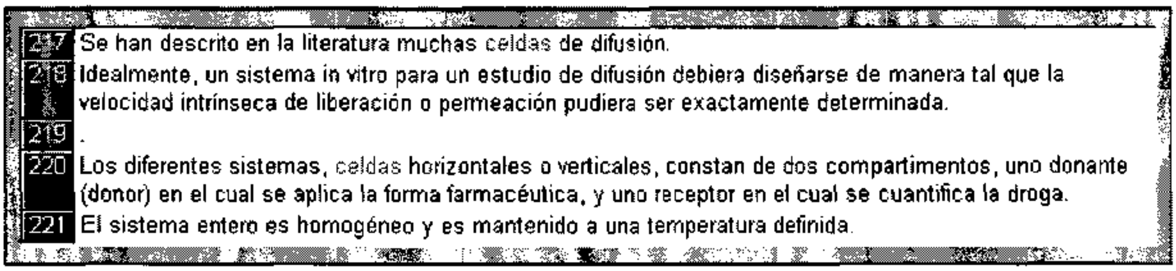
3 de permeación in vitro es el empleo de cultivos celulares como membrana de difusión. Estos c

4. aces de atravesar por difusión las membranas celulares. Unión a la sangre. Las sustancias

5. eso. Dichos estudios se llevan a cabo utilizando **celdas de difusión** ya sea con una solución o sin

6. Sobre la herida se coloca una **celda de difusión** de vidrio, con un diámetro interno de 1,8 cm que se asegura con

The translation equivalent in Spanish would be *celda de difusión* (and not **célula).* The Viewer & Aligner utility enables the user to have access to the full text from where the concordance line has been extracted. This is a very convenient option as it is a source of additional information. In this case, different types of diffusion cells are mentioned (horizontal, vertical), as well as their two compartments: donor and receptor



217 Se han descrito en la literatura muchas celdas de difusión.

218 Idealmente, un sistema in vitro para un estudio de difusión debiera diseñarse de manera tal que la velocidad intrínseca de liberación o permeación pudiera ser exactamente determinada.

219 .

220 Los diferentes sistemas, celdas horizontales o verticales, constan de dos compartimentos, uno donante (donor) en el cual se aplica la forma farmacéutica, y uno receptor en el cual se cuantifica la droga.

221 El sistema entero es homogéneo y es mantenido a una temperatura definida.

Fig. 4. *Access to the corpus from a KWIC line* (celda)

In scientific Spanish there is a tendency to keep English acronyms: either they stand on their own or, else, they appear in brackets after the Spanish translation of the whole multiword units. In this case, a translator may want to check HPLC in the TL corpus. A search with [(HPLC)] or [HPLC*] confirms the use of the English acronym in Spanish, plus, optionally, the full Spanish form. In this case, there is certain terminological instability: *cromatografía líquida de alta eficacia, cromatografía líquida de alta resolución* y *cromatografía líquida de altas características.*

1. CUADRO 3 . Compuestos de referencia recomendados para aplicar el método de HPLC conforme a los datos de adsorción a los suelos... Tabla sólo disponible

2. pos más frecuentes: 14C, 3H, 32P, 35S, etc. • Detector de radiactividad para HPLC. Contador de centelleo acoplable a un HPLC para emisores b.

3 radioactividad Contador de centelleo Cromatografía líquida de alta eficacia (HPLC) Cromatografía en capa fina ---------------------

4 rizantes o muy calientes, residuos líquidos y residuos sólidos. - Un equipo de HPLC., de la marca Hewlett-Packard, con inyector automático, bomba cuaternaria,

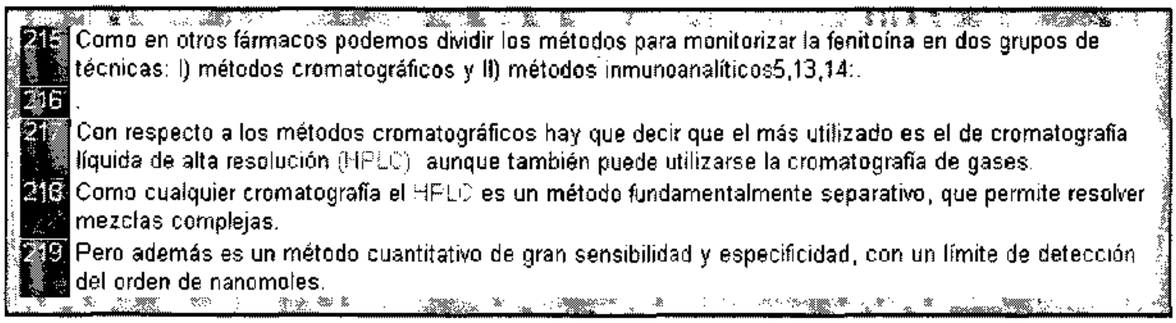
6 Técnicas Instrumentales Cromatografía Líquida de Alta Resolución (HPLC) • Dos bombas isocráticas, tres bombas para gradientes, tres automuest

7 ién puede utilizarse la cromatografía de gases. Como cualquier cromatografía el HPLC es un método fundamentalmente separativo, que permite resolver mezclas comp

8 a cromatografía gas-líquido (GLC), la cromatografía líquida de alta resolución (HPLC), la espectrometría (por ejemplo, cromatografía gaseosa/espectrometría de m

9 tico. ¿Cómo seleccionar el vial adecuado? Inyectores automáticos para GC y HPLC. La mayoría de los inyectores automáticos utilizan uno de los tres tamaños

The concordance lines will also show the actual usage pattern and lexical cohesion of the acronym *(método de, equipo de; el, una; centelleo, vial, inyectores,* etc.), as well as some kind of encyclopaedic information.



215 Como en otros fármacos podemos dividir los métodos para monitorizar la fenitoína en dos grupos de técnicas: I) métodos cromatográficos y II) métodos inmunoanalíticos5,13,14:.

216 .

217 Con respecto a los métodos cromatográficos hay que decir que el más utilizado es el de cromatografía líquida de alta resolución (HPLC) aunque también puede utilizarse la cromatografía de gases.

218 Como cualquier cromatografía el HPLC es un método fundamentalmente separativo, que permite resolver mezclas complejas.

219 Pero además es un método cuantitativo de gran sensibilidad y especificidad, con un límite de detección del orden de nanomoles.

Fig. 5. *Access to the corpus from a KWIC line* (HPLC)

A tentative version into Spanish would read as follows:

TT2

- Celda de difusión de vidrio.

- Recogida de muestras en viales de centelleo o HPLC mediante inyector/pipeta (automáticos).

5. Final remarks

The translation of texts from specialised domains entails very specific documentary and terminological needs. Traditional information sources, such as standard dictionaries, databases or parallel texts fall short of solving all the stumbling blocks posed by this type of translation activity. TMS and other CAT tools prove to be more useful in this context. However, they suffer from serious shortcomings: a steep learning curve, rather expensive for beginning translators, unsuitable for dynamically developing domains, unable to catch up with growing neology, etc. Corpus compilation techniques and concordancing software offer new, exciting opportunities for translators. In addition, these are freely available (or rather inexpensive), updated and user-friendly tools.

In this paper we have demonstrated how concordancers can be used to search corpora which have been tailored to meet the specific needs of a given translation project. Comparable and parallel corpora provide reliable and accurate information about TL text conventions, orthographical variants, colligational patterns and collocational profiles of TL terms, and actual usage and lexical cohesion of translation equivalents. They are also extremely useful for checking terminological instability in the TL, in outlining norms governing the translation of acronyms, for instance, or in digging out conceptual and encyclopaedic information. But this is just the tip of the iceberg. In a near future, practising translators will no doubt become more and more enthusiastic about corpus management and concordancers.

References

Baker, P.; Hardie, A.; McEnery, T. 2006. *A Glossary of Corpus Linguistics.* Edinburgh: Edinburgh University Press.

Barlow, M. 2004. “Parallel Concordancing and Translation”. *Proceedings of the 26th Translating and the Computer Conference.* London: Aslib.

Bowker, L. 2002. *Computer-Aided Translation Technology: A Practical Introduction.* Ottawa: University of Ottawa Press.

Corpas Pastor, G. 2001. “Compilación de un corpus *ad hoc* para la enseñanza de la traducción inversa especializada”. *Trans: revista de traductología.* 5.155-184.

Corpas Pastor, G. 2002. “Traducir con corpus: de la teoría a la práctica”. In J. García Palacios and M. T. Fuentes (eds.) 2002. *Texto, terminología y traducción.* Salamanca: Almar. 189-226.

Corpas Pastor, G. 2004a. “La traducción de textos médicos especializados a través de recursos electrónicos y corpus virtuales”. In L. González and P. Hernúñez. (eds.). 2004. *Las palabras del traductor. Actas del II Congreso Internacional “El español, lengua de traducción”, 20 y 21 de mayo, Toledo 2004.* Bruselas: Comisión Europea/ESLETRA. 137-164.

Corpas Pastor, G. 2004b. “Localización de recursos y compilación de corpus via Internet: aplicaciones para la didáctica de la traducción médica especializada”. In V. García Yebra and C. Gonzalo García (eds.). 2004. *Manual de documentación y terminología para la traducción especializada.* (Colección “Instrumenta Bibliologica”). Madrid: Arco/Libros. 223-506.

EN 15038:2006 *Translation Services-Service Requirements.*

Granger, S.; Petch-Tyson, S. (eds.) 2003. *Extending the scope of corpus-based research. New applications, new challenge.* Amsterdam, New York: Rodopi.

Hickey, R. 2003. *Corpus Presenter. Software for Language Analysis. With a manual and A Corpus of Irish English as sample data.* Amsterdam, Philadelphia: John Benjamins. [+ CD-ROM]

Hoey, M. 2006[2005]. *Lexical Priming. A new theory of words and language.* London, New York: Routledge.

Quah, C. K. 2005. *Translation and Technology.* Palgrave: MacMillan.

Zanettin, F.; Bernardini, S.; Stewart, D. (eds). 2003. *Corpora in Translator Education.* Manchester: St. Jerome.

Appendix I. *SLC 25 Most Frequent Content Words*

RANK	WORD	NUMBER OF OCCURRENCES	%
4	SKIN	137	2,53
8	TEST	93	1,72
13	ABSORPTION	56	1,04
15	METHOD	39	0,72
16	SUBSTANCE	39	0,72
18	PREPARATION	35	0,65
20	EXPOSURE	32	0,59
24	VITRO	27	0,50
25	ANIMALS	25	0,46
28	FLUID	24	0,44
29	RECEPTOR	_ 24	0,44
30	PERCUTANEOUS	22	0,41
32	ANALYSIS	21	0,39
38	DATA	19	0,35
40	APPLICATION	18	0,33
41	DOSE	18	0,33
43	VIVO	18	0,33
44	STUDIES	17	0,31
47	CONDITIONS	16	0,30
49	METHODS	16	0,30
53	SITE	15	0,28
56	CELL	14	0,26
57	CHEMICAL	14	0,26
63	DIFFUSION	12	0,22
66	HUMAN	12	0,22

Appendix II. *TLC 25 Most Frequent Content Words*

RANK	WORD	NUMBER OF OCCURRENCES	%
20	ENSAYO	2.612	0,38
21	SUSTANCIAS	2.548	0,37
25	SUSTANCIA	1.883	0,27
26	ANIMALES	1.577	0,23
39	IMAGEN	1.177	0,17
43	DATOS	1.026	0,15
45	CASO	1.000	0,14
48	CONCENTRACIÓN	960	0,14
51	EXPOSICIÓN	921	0,13
53	MÉTODO	915	0,13
55	ISO	910	0,13
56	EFECTOS	909	0,13
59	ANEXO	892	0,13
60	NUMERO	868	0,13
62	DOSIS	835	0,12
64	CÉLULAS	829	0,12
66	EJEMPLO	818	0,12
69	TOXICIDAD	790	0,11
73	RESULTADOS	725	0,10
74	VERSIÓN	717	0,10
75	PREPARADOS	716	0,10
76	PIEL	711	0,10
78	MÉTODOS	691	0,10
79	PRODUCTOS	690	0,10
80	DOCUMENTO	688	0,10