

## **Corpora and Students' Autonomy in Scientific and Technical Translation training<sup>1</sup>**

**Clara Inés López-Rodríguez; María Isabel Tercedor-Sánchez**

**Department of Translation and Interpreting, University of Granada, Spain**

### **ABSTRACT**

The use of corpora in the translation classroom has been explored by several researchers (Bowker 1998, 2000; Faber et al 2001; Zanettin, Bernardini & Stewart 2003). The aims of this paper are to explore the applications of corpora in the teaching of scientific and technical translation at university level, and to design activities that increase learner autonomy while developing translation strategies and evaluation skills. These activities include the analysis of three types of corpora (DIY corpora, learner corpora, and a quality corpus), the introduction of tags for didactic purposes, and the generation of lemmatised wordlists and concordances. We also explore how corpora contribute to the acquisition of knowledge of the subject field and its conventions, and to the production of adequate texts in the target language.

### **KEYWORDS**

Scientific and technical translation, teaching methodology, learner's corpus, corpus annotation, concordances, self-assessment, electronic tools.

## **1. Introduction**

Many studies have applied corpus linguistics to translation didactics (Bowker 1998, 2000; Faber et al 2001; López 2003; Zanettin 1998, 2001; Zanettin, Bernardini & Stewart 2003). The use of comparable and parallel corpora not only contributes to learning autonomy but also helps students find adequate words for a particular context and text type.

This study has been carried out within two research projects aimed at the generation of terminological resources in the domain of coastal engineering: PUERTOTERM and MARCOCOSTA. The methodology and results of these projects are also useful for the design of teaching materials in the context of Scientific and Technical Translation, as well as in Audiovisual Translation-Localisation at university level (Tercedor, López & Robinson 2005). Our students are 3rd and 4th year students working from English to Spanish and from Spanish to English as part of the BA in Translation and Interpreting of the University of Granada.<sup>2</sup>

We take the approach of social constructivism and collaborative learning to translation (Kiraly 2000, González Davies 2004, Robinson, López & Tercedor 2006) for the design of various exercises in contact classes and virtual environments. The methodology of an e-learning course is based on the strategies of proactive learning, focused on the student as the key element in the learning process. As opposed to the traditional translation classroom where knowledge would pass from the teacher to the students, in the collaborative classroom, the teacher acts as facilitator in the

learning process. The translation task is carried out under real-world conditions: there is a client and an expert in the subject field, and students work in groups and rotate their roles.

The aims of this paper are to explore the applications of corpora in the teaching of specialised translation, and to design aids that increase learner autonomy while developing the different skills involved in the translation process and in self-assessment. We also explore how corpora contribute to the acquisition of knowledge of the subject field and its conventions, and to the production of adequate texts in the target language. The use of corpora will be combined with the inclusion of multimedia materials as a valuable means for students to become familiar with the emerging forms of communication that the information society has brought about. The translation aids proposed deal with two subject fields: computing and coastal engineering.

## **2. The use of corpora to increase students' autonomy in the translation classroom**

The combination of corpora and texts including visual material provides a powerful tool for dealing with the tasks of specialised translation. Proactive learning in the translation classroom is facilitated when students carry out learner-centered activities revolving around:

- Getting familiar with corpora and annotation
- Acquiring knowledge about the subject field and its conventions
- Identifying translation problems
- Developing translation strategies
- Evaluating the solutions given by peers and by themselves (peer- and self-assessment)

### **2.1. Types of corpora used in the classroom**

In order to promote autonomy and self-assessment in the classroom, we ask students to compile and use different types of corpora. On the one hand, students compile a DIY corpus, i.e. a collection of Internet documents created ad hoc as a response to a specific text to be translated (Zanettin 2002: 242). This corpus has both a parallel and a comparable component.

In the DIY corpus, we ask students to look for texts that include visual and multimedia materials since the new formats of scientific and technical translation include more than ever this sort of material. Moreover, images and multimedia objects facilitate not only the learning process, but also the translation process. Visualisations may lead to creative translations, or in other words, "translation involves changes when compared with the source text, thereby bringing in something that is novel" (Kussmaul 2005: 380). When students compile a DIY corpus, they read many specialised

texts in electronic format, assess their reliability and organise them in folders.

Secondly, with the original texts and the translations produced by our students, we create a learner corpus of parallel texts that has been tagged to study the solutions to translation problems.

In order to provide our students with more feedback that will help them minimise their mistakes/errors<sup>3</sup> in the future, we take advantage of the two comparable English and Spanish corpora compiled within our research projects on oncology<sup>4</sup> and coastal engineering. The corpus on oncology contains around 32 million words (28, 771, 714 in English, and the rest in Spanish).<sup>5</sup> As to the corpus on coastal engineering, it contains approximately 9.5 million words. The English section includes about 4.5 million words versus the Spanish one, over 5 million words.<sup>6</sup>

These comparable corpora contain, on the one hand, high-quality texts (articles, manuals, textbooks and specialised web sites) in English and Spanish that have been hand-picked because they provide a good explanation of the subject matter. On the other hand, it is a quantity corpus in the sense that it provides a representative sample of the specialised language in question. Therefore, this corpus can be considered an evaluation corpus (Bowker 2001: 352).

## **2.2. Corpus annotation**

Students manipulate the learner corpus, adding tags to the source and translated texts and they search the corpus with WordSmith Tools [<http://www.lexically.net>] in order to see the appropriateness of terms, syntactic combinations and collocations, and to better understand the meaning of certain expressions. As shown in López, Robinson & Tercedor (2007) the manipulation of different types of tags increases students' learning autonomy and their self-assessment strategies. In this study, we proposed four sets of tags to be placed after the problem or mistake/error in question. We decided not to include opening and closing HTML-like tags to avoid visually overloading the text. These tags specified:

- (a) problematic areas both in the source text and in the translation
- (b) type of error/mistake according to criterion descriptors
- (c) adequacy/appropriateness of translated sentences
- (d) first impression of the translation.

Before working with the learner corpus, students read the text to be translated, and identify potential translation problems in the source text described in López & Tercedor (2004: 33). We associate these problems to the following tags (Figure 1):

<CON>	Conceptualization (inability to understand the source text)
<PRO>	Procedural (inability to judge the reliability of documentation sources, reading problems, problems caused by wrong use of dictionaries or terminological databases, etc.)
<TRA>	Transfer not achieved (due to linguistic and cultural differences between the source and target languages )
<QTO>	Lack of source text quality (source text ambiguity, inappropriate style)

**Figure 1.** Tags to identify problematic areas in the source text

Once students are aware of the problematic areas of the source text, they add a number referring to sentence number, followed by a tag referring to potential translation problems (CON, PRO, TRA, QTO), as shown in Figure 2.

```

4 Time exposure images (Timex)<4a><PRO>: These images average the int
ver a time period of 10 minutes<4b><CON>. 5 Typically images are recor
frequently over shallow sanbars<6><QTO><CON> generating foam appear ap
es 7 The wave-breaking patterns<7><PRO> highlighted in the timex image
he surface. 8 Variance images<8><PRO>: These images represent the va
de the waters edge (swash zone)<10><CON><PRO> where the beach is perio
e projected on the ground plane<11a><CON>, resulting in rectified imag
es with real world co-ordinates<11b><CON>. 12 These rectified images

```

**Figure 2.** Concordance displaying the tags that point to translation problems in a source text on coastal engineering.

When there is not sufficient time to tag the source text with numbers and a reference to the type of problem, students are asked at least to tag their translated text with a reference to the problematic expressions in the source text. We call this process ad hoc tagging. One of the source texts used in the classroom is an advertisement aimed at health professionals interested in purchasing Intel™ Wireless Networking Products. Below (Figure 3) we present students' translations of one of the 'problematic' sentences: *The new Rx for tracking meds*. We asked our students to place the tag <Rx> after their translation of the sentence, which includes two translation problems for native speakers of Spanish: the acronym *Rx* (meaning prescription) and the clipping *meds* (meaning medications).

One of the advantages of asking students to use ad hoc tags such as <Rx> is that the tagging process is accelerated. Besides, ad hoc tags enable the teacher to check students' solutions for a particular translation problem very quickly. The teacher can then show these different solutions to students, and can highlight what s/he considers to be the most adequate translations. In our example, using Wordsmith Tools, the teacher searched for the tag <Rx> in students' translations, and generated concordance lines that were subsequently presented to the students. In Figure 3, concordance lines 1, 4, 5 and 10 display viable options as highlighted by the teacher.

N Concordance  
 1 ? <Rx>Una nueva fórmula para controlar la medicación</Rx> Aunque sea el med  
 2 <Rx>El nuevo Rx para seguir la pista a los médicos.</Rx> El medicamento ad  
 3 ico? <Rx>El nuevo Rx para encontrar informes médicos</Rx> El medicamento es  
 4 cables? <Rx>Administración correcta de medicamentos</Rx> Es el medicamento  
 5 ico? <Rx>La nueva forma de administrar medicamentos</Rx> Es el medicamento  
 6 espera. <Rx>Nuevo Rx para comprobar la medicación.</Rx> Es el fármaco cor  
 7 lámbrico? <Rx>Nuevas recetas para nuevas medicinas.</Rx> Es el medicamento  
 8 mundo sin cables? <Rx>The new Rx for tracking meds.</Rx> Es el medicamento  
 9 <Rx>El nuevo receptor para la medicina de seguimiento</Rx> Es la medicina co  
 10 a. <Rx>El nuevo método para prescribir medicamentos</Rx> Es el medicament

**Figure 3.** Concordance from our learner corpus.

With this exercise, students reflect upon the suitability of different translation options in a medical context, something which contrasts with the extended belief that in specialised translation lexical and stylistic variation is not a salient feature.

### 3. Acquiring knowledge about the subject field and textual conventions

In specialised translation it is important that students gain understanding of the main concepts in the subject field. Lemmatised wordlists provide students with a glimpse of the more activated concepts in the texts (López Rodríguez 2001) whereas concordance analysis gives clues about conceptual information as well as co-occurrence patterns of a keyword. Let us illustrate this with two lemmatised frequency lists (Figure 4) generated from advertising texts aimed at convincing two different types of users to buy Intel Wireless Networking Products: on the one hand, health professionals, and on the other hand, managers of small and middle businesses. Both texts were translated from English into Spanish within a unit devoted to the translation of adverts.

If we take a look at lemmatised wordlists in the source text, the most frequent words (i.e. the most representative lexemes) stand out. Therefore, if we compare the lemmatised frequency list of the text for health professionals with the one corresponding to the texts for managers of small and middle businesses, it can be seen that these users have different priorities and interests. Doctors are interested in their *patients* and normally have *laptops*. Business managers are interested in receiving technical *support* and purchasing *products* that will suit their *needs*. Lemmatised wordlists also highlight the most activated vocabulary: *Wireless, Network, Intel, products, access, WLAN, wired, adapters, you, use, connection, easy, support*.

N	Word	Freq.	%	Lemmas
1	NETWORK	31	2.48	networking(11),networked(3),network
2	WIRELESS	30	2.40	
3	INTEL	27	2.16	intel's(1)
4	PATIENT	18	1.44	patient's(5),patients(1)
5	PRODUCTS	11	0.88	product(2)
6	ACCESS	10	0.80	accessible(2)
7	IT'S	10	0.80	
8	LAPTOP	9	0.72	laptops(3)
9	CAN	8	0.64	
10	CONNECTION	8	0.64	connected(2),connecting(1),connect
11	PROVIDED	7	0.56	providers(2),provides(2),provider(1)
12	NEW	6	0.48	
13	PRO	6	0.48	
14	USE	6	0.48	user(1),uses(1)
15	BUILDING	5	0.40	
16	CODE	5	0.40	codes(2)
17	COMPUTERS	5	0.40	computer(1),computerbased(1),com
18	EASY	5	0.40	easily(1)
19	EQUIPPED	5	0.40	equipment(1),equipping(1)
20	FORMS	5	0.40	form(1)
21	INFORMATION	5	0.40	
22	LAN	5	0.40	
23	NEED	5	0.40	needed(1),needs(2)
24	NURSE	5	0.40	nurse's(1),nurses(1),nursing(1)

Word	Freq.	%	Lemmas
INTEL	74	2.95	
WIRELESS	69	2.75	
YOU	59	2.35	your(18)
NETWORK	46	1.83	networking(8),networks(4)
PRODUCTS	41	1.63	product(8)
ACCESS	36	1.43	
WLAN	36	1.43	wlans(8)
LAN	34	1.35	lans(1)
POINT	32	1.27	points(15)
PRO	29	1.15	
CAN	18	0.72	
WIRED	18	0.72	wire(1),wires(1)
ADAPTERS	16	0.64	adapter(6)
SUPPORT	16	0.64	supports(3),supporting(1)
USE	15	0.60	used(1),uses(3),using(3)
USERS	13	0.52	user(3)
GHZ	12	0.48	
FEATURES	11	0.44	feature(3)
OUR	11	0.44	
CONNECTIONS	10	0.40	connection(3),connectivity(2)
OTHER	10	0.40	others(1)
INFORMATION	9	0.36	
NEED	9	0.36	needs(1)

**Figure 4.** Lemmatised frequency list of the most activated lexical words in texts intended for different users: health professionals (left) and managers of small and middle businesses (right).

This vocabulary will be the basis for the glossaries students elaborate for each text to be translated. Besides, as these words are frequent in the texts, students know that they can generate concordance lines around these words, thus familiarising themselves with their use in context. This information is valuable to gain a deeper understanding of the text and to learn current English usage, so it can be included in their glossaries.

Moreover, if students extract lemmatised frequency lists from a specialised corpus in the target language, they will have access to relevant terms in the target language (i.e. *tecnología inalámbrica, conectividad inalámbrica, red de área local inalámbrica (WLAN), red cableada, adaptador Wifi*). This input will help them solve translation problems.

Therefore, the use of lemmatised wordlists sheds light on the most important concepts activated in texts. It also provides students with a list of terms that should be included in their glossaries.

### 3.1. Types of concordance

In our research project, extracting concordances has a fourfold objective (Tercedor & López 2008):

- Extracting conceptual information (*conceptual concordances*): acquiring knowledge about the subject field, its relevant concepts and their relationships in the field
- Knowing co-occurrence patterns in specialised discourse (*structural concordances*)
- Knowing the selection patterns of verbs and pointing to verbs that collocate with certain keywords (*verbal structural concordances*)
- Understanding the different senses of a word: semantic prosody, metaphorical extensions and word sense disambiguation

We can take advantage of the enormous potential of concordance lines in the classroom by offering students a selection of concordances around some of the terms present in the source text. Figure 5 illustrates a selection of concordances around the terms *erosión eólica* (conceptual concordances), *oxidación* (structural concordances) and *marea* (verbal structural concordances). They have been extracted from the PUERTOTERM corpus.

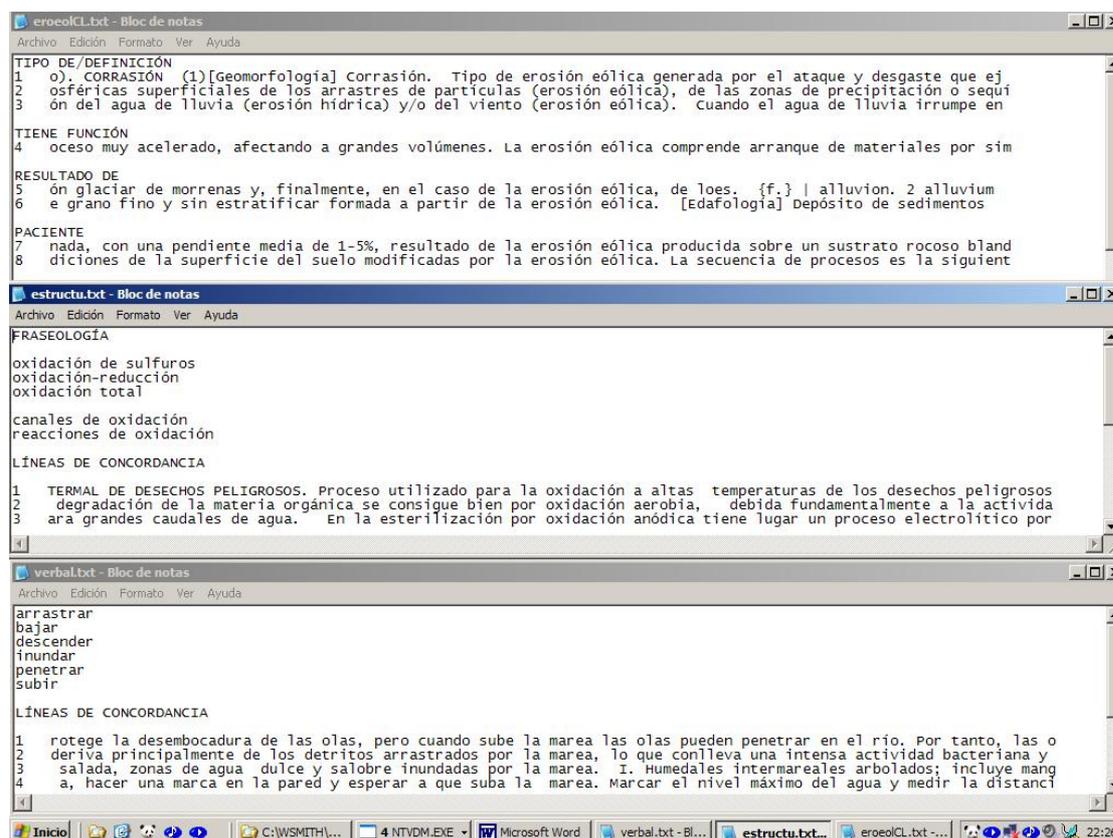


Figure 5. Types of concordance used in the PUERTOTERM project.

From the first type of concordances, conceptual concordances, students can gather concepts which are semantically related to the concept *erosion eólica* 'wind erosion'. The semantic relations between concepts are also made explicit. For example:

IS-A: *Corrasión. TIPO DE erosión eólica*

HAS-FUNCTION: *la erosión eólica COMPRENDE arranque de materiales.*  
 IS-AFFECTED: *superficie del suelo MODIFICADAS por la erosión eólica*

As to the structural concordances, they are useful in becoming familiar with the phraseology of a specialised field. Figure 5 shows some of the phraseological units around the word *oxidación*: *oxidación de sulfuros, oxidación-reducción*, etc.

The generation of verbal structural concordances sheds light not only on the verbs that usually collocate with relevant terms, but also on selection patterns of verbs. The verbal structural concordances in Figure 5 show the verbs that usually collocate with *marea*: *arrastrar, bajar, descender, inundar*, etc.

Finally, concordances provide clues on the different senses and nuances in meaning of a word. They account for a collocational phenomenon called semantic prosody, which is “the consistent aura of meaning with which a form is imbued by its collocates” (Louw 1993: 157). Since non-native speakers of a language find it difficult to identify the attitudinal meaning (positive or negative) associated with certain words, students find in concordances a powerful resource to make semantic prosody explicit. For instance, when translating a medical text containing several instances of the word “effect”, it can be helpful to search our quality corpus on oncology. Concordances around “effect(s)” in our quality corpus show that this word has a negative semantic prosody in texts about cancer. In particular, the word “effect” co-occurs with lexical units indicating negative entities, properties or events (highlighted in bold) as can be seen below:

1 more drugs are needed for maximal effect. Mature results of prospectiv  
 2 om pulmonary and hematologic **toxic** effects; proper administration requ  
 3 , those studies showing a positive effect for combined modality therap  
 4 0-200 cGy to **avoid long-term toxic** effects to normal brain tissue [13,  
 5 because of the **high risk of toxic** effects observed with such treatmen  
 6 acted out in **destructive** ways, the effects can be **devastating**. Likewis  
 7 tively on sleep patterns. **Side** effects of treatment that may affec  
 8 me **anxiety, and serious withdrawal** effects, such as **seizures** [3].The f  
 9 nzodiazepines have an anti-anxiety effect in low doses and a hypnotic  
 10 d that melatonin exerts a **hypnotic** effect through thermoregulatory mec  
 11 nts. **Barbiturates** have a number of **adverse** effects, including the deve  
 12 ize clinical response and minimize **adverse** effects. Some patients may  
 13 een evaluated [34,45-50].The major adverse effects include **headache** (w  
 14 d **moist desquamation** results. This **side** effect increases the **risk** of i  
 15 associated with a wide variety of **side** effects and **toxic** effects. Pru  
 16 **toxic** effects. **Pruritus** has been a side effect associated with several  
 17 effective and does produce gastric **side** effects [17]. Aspirin seems to

Students also learn relevant phraseological units in the field of medicine: “side effect(s)”, “toxic effect(s)” and “adverse effects”.

### 3.2. Search structures to extract conceptual information from corpora

The acquisition of field knowledge necessary in the pre-translation phases is facilitated by the visualisation of conceptual structures and relations. This can be done gathering a set of conceptual concordances to show synonyms, hyperonyms, definitions and other conceptual information, as seen in Figure 5.

Before generating wordlists and concordance lines, students are trained to identify linguistic patterns whose context might be very informative for the identification of superordinate terms, textual and orthographical conventions, and the strategies used in translation. In López Rodríguez (2002), we called these expressions *search structures* following previous research in terminology (Meyer & Mackintosh 1996; Temmerman 2000) aimed at the extraction of superordinate terms, and the formulation of terminographic definitions. In specialised translation, *search structures* such as *is a*, *called*, *kind/kinds of* or *is defined as* help students extract the main concepts in the subject field, and definitions to better understand those superordinate terms:

1 n (Figure 7f-2). **Absorption** is defined as a process in which solar radiati  
 2 discharge. A **recharge area** is defined as a portion of a drainage basin wh  
 3 ter table. A **discharge area** is defined as a portion of the drainage basin  
 4 **channel cover factor**, CCH, is defined as the ratio of degradation from a  
 5 are needed. A **unit channel** is defined as a channel of length  $L = 1$  km and  
 6 The **reflection coefficient** is defined as the reflected wave height divide  
 7 he **transmission coefficient** is defined as the ratio of the transmitted wav  
 9 o « SEDmx (days). A **dry day** is defined as a day with surface runoff less t  
 10 n dry days, where a **dry day** is defined as a day with less than 0.1 mm of s  
 11 the previous day. A **wet day** is defined as a day with 0.1 mm of rain or mor

**Figure 6.** Concordance around the search structure *is defined as*

Following a socioconstructivist approach, students are encouraged to gather their relevant concordance lines for analysis, and are prompted with clues such as: what are the main concepts related to keyword X? Which concepts are hyponyms? Which is the most generic term? How is it defined in the corpus in Language A? How is it defined in Language B? Which is the most salient dimension of concept X in the text?

Search structures make explicit not only paradigmatic relations such as IS-A or PART-OF, but also horizontal relations such as IS-LOCATED, IS-THE RESULT-OF, IS-MADE-OF, HAS-A, HAS-FUNCTION, etc. By identifying relations, students gain a deeper understanding of the domain, and are able to build a conceptual representation of the domain where all terms in the texts are interrelated. Besides, they can write coherent definitions to include in their glossaries, thus acquiring 'terminographer-like' training.

We explain the importance of identifying conceptual relations with examples such as the one below (Figure 7), taken from the researcher interface of the PUERTOTERM database. If we consider one of the concepts in

the database, ABRASIÓN ('ABRASION'), there is a special section called *Relations*, where the concepts and relations around this concept are introduced: EROSIÓN, FRICCIÓN, ABRASIÓN GLACIAR. The definition of ABRASIÓN highlights these semantically-related concepts: EROSIÓN (which is its superordinate term), FRICCIÓN (which causes this process), etc. The definition also codes some important conceptual relations: *produce* (CAUSE) and *resultado de* (RESULT-OF).

Concepto	ABRASIÓN	
Definición	erosión que produce un desgaste mecánico en superficies rocosas o en depósitos sedimentarios consolidados como resultado de la fricción producida por los materiales sueltos (ej.: arena) que transportan los agentes externos como el hielo, las corrientes fluviales o el viento.	
Revisado	No	
Autor	Clara	
Nota	Según el tipo de agente predominante se distingue entre abrasión glaciar, fluvial, litoral o marina, etc.	

Términos	Recursos	Relaciones	Dominios	Mensajes
----------	----------	------------	----------	----------

		<a href="#">+ Nueva relación</a>
Concepto principal:	ABRASIÓN	
Relación:	tipo de	<a href="#">Más información</a>
Concepto secundario:	EROSIÓN	
Concepto principal:	ABRASIÓN	
Relación:	resultado de	<a href="#">Más información</a>
Concepto secundario:	FRICCIÓN	
Concepto principal:	ABRASIÓN GLACIAR	
Relación:	tipo de	<a href="#">Más información</a>
Concepto secundario:	ABRASIÓN	

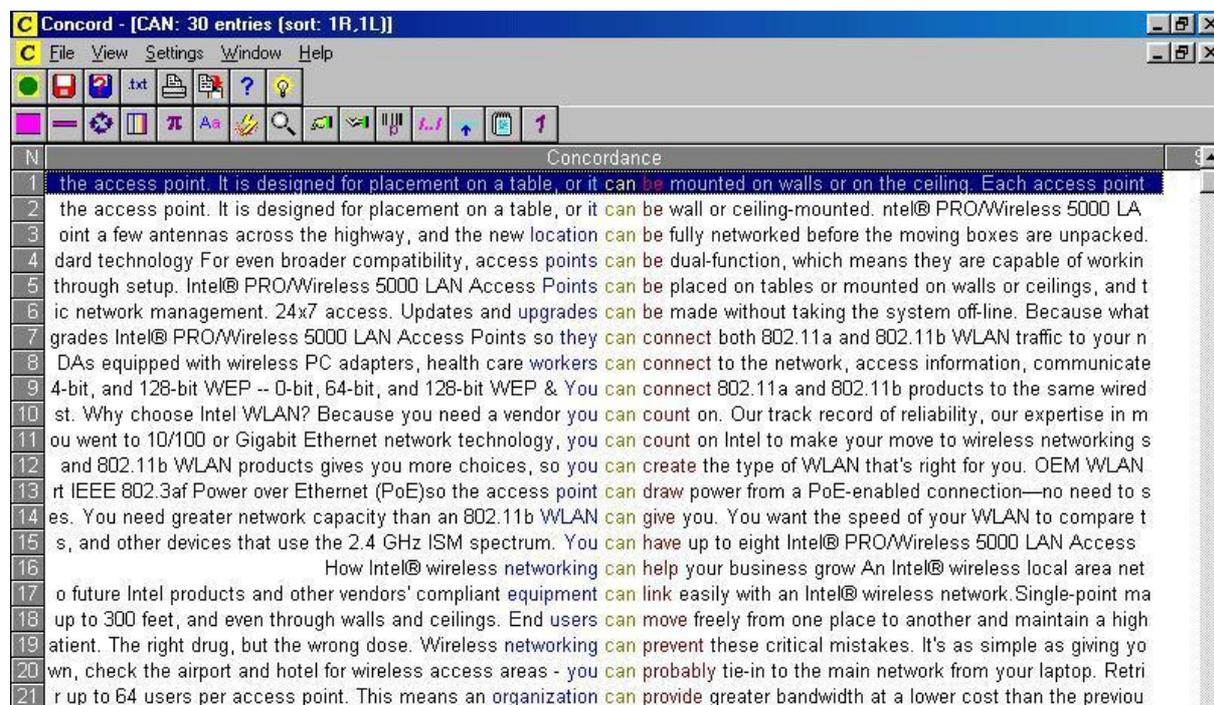
Figure 7. Conceptual relations of the concept ABRASIÓN

### 3.3. Co-occurrence patterns, rhetorical conventions and ideology

Once students are familiar with some of the prototypical concepts in the subject field and their collocational meaning, students can exploit corpora to extract co-occurrence patterns that point to the syntactic behaviour of words. As shown in Figure 5, structural concordances (i.e., those in which the relevant syntagmatic relations of a particular keyword are made explicit) contribute to the acquisition of phraseological and syntactic knowledge.

However, the generation of lemmatised frequency lists, concordance lines and dispersion plots reveals not only linguistic patterns, but also the rhetorical conventions of the genre in question, and some of the values and ideology behind the text. By analysing three print and online advertisements of wireless devices included in the course materials,<sup>7</sup> students get familiar with the textual conventions associated with the genre of advertising texts. These conventions point to the exhortative and informative functions of this type of text. The rhetorical sections of print advertisements can then be related to specific linguistic patterns. For instance, the modal verb *can* is very frequent in two of the sections of these ads: in particular, the section introducing the product and explaining its usefulness for the consumer, and the section presenting the

advantages and features of the product. The analysis of both the collocational span to the right of the word *can*, and lemmatised wordlists (Figure 4) reveals all the capabilities of the wireless device that is being advertised, and to a certain extent some of the values, stereotypes, and desires brought about by information society: *connect, access to information, link, help your business grow, move freely, prevent critical mistakes, etc.*



**Figure 8.** Focussing on textual conventions and ideology: concordance for the modal verb *can*.

Therefore, if we use the source text as a corpus, we can get closer to the areas of interest of potential users of the product, and to the conceptual metaphors which are activated in a subtle way by the images and the words in the text. We can also reflect upon the effect that ads have in the evocation of values, desires and stereotypes of a particular culture.

Advertising for health professionals must, as objectively as possible, present medical, pharmaceutical and pharmacological data, and indicate pros and cons. The affective aspect is nonetheless extensively exploited as medical data are presented with icons and text referring to conceptual metaphors. These metaphors evoke cognitive representations related to collective views of disease and health, pain [...] Vandaele (2002: 329-330)

As we have already seen, the exploitation of corpora makes overt the use of linguistic patterns and ideas interspersed in language. These patterns and ideas are not easily recognised by non-native speakers of a language. As a result, students gain confidence when they can validate in a corpus both their intuitions about the text (such as writer's intention or ideology masked in the text) and their expectations as to how words are used. This is particularly true when translating advertising texts. In this text type,

the subtle introduction of affective elements is achieved by the strategic combination of images and text that evoke the most accepted values in a certain community. In the case of medico-pharmaceutical advertising, pharmaceutical companies are subjected in some countries to regulations that forbid them to explicitly say that the patient should buy their products. As a result, their ads include images, words and expressions which activate in a subtle way some of our metaphors and ideas about health and disease; for example, the DISEASE IS WAR metaphor or the idea that the mental well-being of patients is as important as therapeutic care.

#### 4. Developing translation strategies and self-assessment skills

Now that students have acquired basic knowledge about the subject field and are familiar with typical terms and expressions, students have to translate the texts (trying to overcome translation problems by deploying translation strategies), and to assess the quality of their work. The use of corpora at this stage can also be useful since both the learner corpus and the quality corpus (see section 2.1.) contain many appropriate solutions to translation problems.

In the learner corpus, they can look for the tags that were previously introduced to identify translation problems, whereas in the Spanish quality corpus they can look for terms that activate similar concepts. Then, they 'save as text' some representative concordances of the quality corpus, and they can tag both these concordance lines and the learner corpus according to the adequacy /appropriateness of translated sentences. To that end, these tags are used:

(a) Excellent solution for a translation problem: <AA>

(b) Inappropriate translation: <type of error/mistake>

The type of error/mistake is specified, for example, <f> format, <pr> pragmatic error/mistake, <se> meaning, <chse> lack of cohesion affecting meaning, <rg> inconsistencies in register, etc (See Appendix).

(c) Very serious mistake/error: <type of error/mistake><FF>

Before asking our students to evaluate potentially problematic segments produced by their peers, we show them a problematic segment of the source text and offer them a list of "filtered" concordances with their own rendering of the problematic segments (Figure 9). These concordances have been tagged to indicate the type of mistake/error, and to describe the adequacy or appropriateness of translated sentences. In the example, the best solution is tagged with <AA> (4th concordance line), whereas the 7th line includes an unacceptable translation (<FF>) that includes a lexical mistake (<lx>) and a grammatical mistake (lack of concord, <ccsx>).

**SOURCE TEXT**

Regions where waves break frequently over **shallow sanbars**<6><QTO><CON> generating foam appear as bright white bands in the timex images. The wave-breaking patterns<7><PRO> [...]

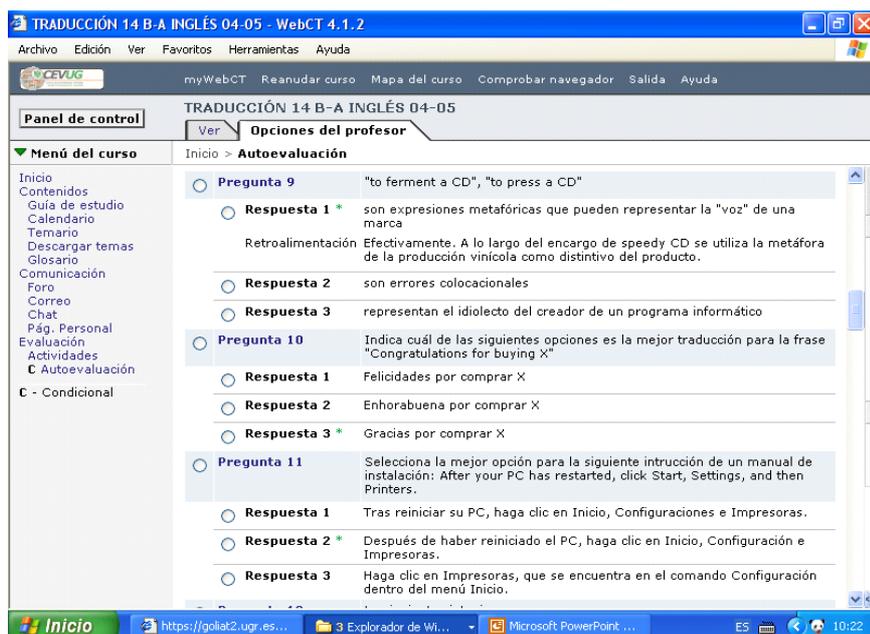
**FILTERED CONCORDANCES**

1 bre bancos de arena poco profundos<6><lx> generando espuma, en las im  
2 en bancos de arena poco profundos<6><lx> generando espuma aparecen c  
3 re barras de arenas poco profundas<6><colx>, aparecen indicadas en l  
4 arras de arena en zonas de poca profundidad<6><AA>, generando espuma,  
5 las barras de arena poco profundas<6><se>, lo que genera espuma, apar  
6 n barras de arenas no muy profunda<6><se>, provocando espuma, se repr  
7 as barreras de arena superficiales<6><lx>, genera<ccsx><FF> la aparic  
8 en una barra de arena superficial<6><lx>, lo que hace que se produzc  
9 bancos de arena son poco profundos<6><lx>. Para inferir la posición y  
10 superficie. Imágenes de varianza<6><lx>: estas imágenes representan  
11 <sx><FF> sobre las barras de arena<6><mise> generando espuma en el ag  
12 arras de arena de poca profundidad<6><se> produciendo espuma aparecen  
13 obre las barras de arena del fondo<6><se>, generando espuma, aparecen

**Figure 9.** Filtered concordances of the translation equivalents for the expression “shallow sanbars”.

Multiple-choice and matching exercises have also been used to foster self-assessment and to enhance microstructural knowledge of particular lexical patterns. The results obtained through the analysis of our two corpora have been tailored to make up this type of exercise with HotPotatoes, an electronic tool to generate matching exercises. In a previous study (Tercedor, Lopez & Robinson 2005), we demonstrated their efficiency when dealing with different theoretical and methodological issues that are idiosyncratic to localization such as:

- Relevant theoretical questions, for instance, ISO standards related to localisation, definitions of basic concepts such as genre, accessibility, internationalisation, etc.
- Standard text structures or formats as a key to identifying the text type and convention (i.e. a particular superstructure such as that of a resource file in localization)
- The identification of particular metaphors in a domain (as a cognitive resource) or a particular text (as a textual element). In question 9 on Figure 10, the use of the metaphor is not a cognitive resource in the domain but responds to a rhetorical device of a particular company aiming at creating a brand image. The metaphor is related to the domain of wine production represented by the expressions "to ferment a CD" and "to press a CD".
- Standard translations in highly technical documents (see question 10 in Figure 10). In English-Spanish translation, the familiar register present in some sections of technical texts with marketing purposes ("Congratulations for buying X") is not seen as acceptable in Spanish, requiring such translation a shift in register and the use of standard options such as the one seen in 10-3: "Gracias por comprar X".
- The identification of the function and focus of an image in relation to a text.



**Figure 10.** Screenshot of a multiple choice questionnaire from the course on Localisation on the WebCT platform.

Self-assessment questionnaires such as the one above are designed by the teacher to point at particular theoretical or methodological issues or to stress translation problems prior to further discussions in the group.

## 5. Conclusions

In this paper we have proposed materials and activities to integrate the use of corpus in the scientific and technical translation classroom. These activities include the analysis of three types of corpora (DIY corpora, learner corpora, and a quality corpus), the introduction of tags for didactic purposes, and the generation of lemmatised frequency lists and concordances. Concordances have been especially useful in the acquisition of expert knowledge (the identification of conceptual structures and relations between concepts), in the identification of syntagmatic structures, and in revealing ideology behind texts.

With these activities students have increased their knowledge about specialised fields and textual conventions, their learning autonomy and their ability to focus on translation problems, mistakes and errors during all the stages of the translation process.

## References

- **Bowker**, Lynne (1998). "Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study." *Meta* 43(4), 631-651.
- **Bowker**, Lynne (2000). "Towards a methodology for exploiting specialized target language corpora as translation resources." *International Journal of Corpus Linguistics* 5(1), 17-52.
- **Bowker**, Lynne (2001). "A Corpus-Based Approach to Translation Evaluation." *Meta* 46(2), 345-363.
- **Corder**, S Pit. 1981. *Error Analysis and Interlanguage*. Oxford: Oxford University Press.
- **Faber** Pamela, Clara Inés López Rodríguez & María Isabel Tercedor Sánchez (2001). "Utilización de técnicas de corpus en la representación del conocimiento médico." *Terminology* 7(2), 167-197.
- **Faber**, Pamela, Carlos Márquez & Miguel Vega (2005). "Framing Terminology: A process-oriented approach." *Meta* 50(4). On line at <http://www.erudit.org/livre/meta/2005/000255co.pdf> (consulted 19.07.2007)
- **González** Davies M. (2004). *Multiple voices in the Translation Classroom*. Amsterdam: John Benjamins.
- **Kiraly**, Don (2000). *A Social Constructivist Approach to Translator Education*. Manchester: St. Jerome.
- **Kussmaul**, Paul (2005). "Translation through Visualisation." *Meta* 50(2), 378-391.
- **López Rodríguez**, Clara Inés (2001). *Tipología textual y cohesión en la traducción biomédica inglés-español: un estudio de corpus*. Granada: Editorial Universidad de Granada. <http://www.ugr.es/~clarailr/tesis.zip> (consulted 19.07.2007).
- **López Rodríguez**, Clara Inés (2002). "Training translators to learn from news report corpora: the case of Anglo-American cultural references." Belinda Maia, J. Haller & M. Ulrych (Eds) (2002). *Training the Language Services Provider for the New Millennium*. Porto: Faculdade de Letras, Universidade do Porto, 213-222).
- **López Rodríguez**, Clara Inés (2003). "Electronic resources and lexical cohesion in the construction of intercultural competence." *Lebende Sprachen* 4/2003, 152-156.
- **López Rodríguez**, Clara & M. Tercedor Sánchez (2004). "Problemas, evaluación y calidad en traducción científica y técnica." *Sendebär* 15 (2004), 29-43.
- **López Rodríguez**, Clara, B. Robinson & M. Tercedor (2007). "A learner generated corpus to direct learner centered courses." Marcel Thelen & Barbara Lewandowska Tomaszczyk (Eds) (2007). *Translation and Meaning Part 7*. Maastricht: Zuyd University, Maastricht School of International Communication, 197-211.

- **Louw**, Bill (1993). "Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies". M. Baker, G. Francis y E. Tognini-Bonelli (Eds) (1993). *Text and Technology*. Amsterdam/Philadelphia: John Benjamins, 157-177.
- **Meyer**, Ingrid & Kristen **Mackintosh** (1996). "Redefining the terminographer's concept-analysis methods: How can phraseology help?". *Terminology* 3 (1), 1-26.
- **Miller**, George A. (1964). "Language and psychology". Eric H. Lenneberg (Ed.) (1964). *New Directions in the Study of Language*. Cambridge: MIT Press, 89-109.
- **Robinson**, Bryan J., Clara I. López Rodríguez & María I. Tercedor (2006). "Self-assessment in translator training." *Perspectives: Studies in Translatology* 14(2), 115-138. <http://www.multilingual-matters.net/pst/014/0115/pst0140115.pdf> (consulted 19.07.2007).
- **Temmerman**, Rita (2000). *Towards new ways of Terminology Description: The sociocognitive approach*. Amsterdam/Philadelphia: John Benjamins.
- **Tercedor Sánchez**, María Isabel, Clara López Rodríguez & Robinson, B. (2005). "Textual and visual aids for e-learning translation courses." *Meta* 50(4). On line at <http://www.erudit.org/livre/meta/2005/000243co.pdf> (consulted 19.07.2007).
- **Tercedor Sánchez**, María Isabel, Clara López **Rodríguez** (2008). "Integrating corpus data in dynamic knowledge bases: the PuertoTerm Project". *Terminology* 14 (2). Forthcoming.
- **Vandaele**, Sylvie (2002). "Traduction publicitaire médico-pharmaceutique et métaphores conceptuelles." Belinda Maia, J. Haller & M. Ulrych (Eds) (2002). *Training the Language Services Provider for the New Millenium*. Oporto: Faculdade de Letras Universidade do Porto, 329-339.
- **Zanettin**, Federico (1998). "Bilingual Comparable Corpora and the Training of Translators." *Meta* 43(4), 616-630.
- **Zanettin**, Federico (2001). "Swimming in words: corpora, language learning and translation." Guy Aston (ed.) (2001). *Learning with corpora*. Houston, Texas: Athelstan, 177-197).
- **Zanettin**, Federico (2002). "DIY Corpora: The WWW and the Translator." Belinda Maia, J. Haller & M. Ulrych (2002), 239-248. <http://www.federicozanettin.net/DIYcorpora.htm> (consulted 19.07.2007).
- **Zanettin**, Federico, Silvia Bernardini & Dominic Stewart (Eds.) (2003). *Corpora in Translator Education*. Manchester: St Jerome Publishing.

## Appendix

### Type of error/mistake according to criterion descriptors

DECODING		ENCODING	
TRANSLATION INTO ENGLISH			
Content	Register, vocabulary, terminology	Translation brief and orientation to target text type	Written expression
<p>&lt;se&gt; meaning</p> <p>We can add nuances to this tag:</p> <p>&lt;chse&gt; Lack of cohesion</p> <p>&lt;mise&gt; less information than ST</p> <p>&lt;pluse&gt; more info than ST</p> <p>&lt;tvse&gt; wrong tense that causes change in meaning</p> <p>&lt;cose&gt; change in meaning due to wrong collocation</p> <p>&lt;dtse&gt; changes in the data</p>	<p>&lt;lx&gt; lexis and terminology</p> <p>&lt;colx&gt; wrong collocation</p> <p>&lt;rglx&gt; term that is not appropriate for the register of the text</p> <p>&lt;rg&gt; register (inconsistencies)</p>	<p>&lt;o&gt; organization</p> <p>&lt;pr&gt; pragmatic mistakes</p> <p>&lt;rtpr&gt; Grammatically correct but it sounds unnatural. The rhetorical effect of the ST is missing. Literal translation.</p> <p>&lt;ist&gt; inappropriate style</p> <p>&lt;ot&gt; orthotypography</p> <p>&lt;f&gt; layout, wrong accomplishment of style sheet or computer requirements</p> <p>&lt;rc&gt; inappropriate rendering of cultural reference</p>	<p>&lt;or&gt; spelling</p> <p>&lt;pt&gt; punctuation</p> <p>&lt;sx&gt; syntax</p> <p>&lt;ccsx&gt; lack of concord</p>
TRANSLATION INTO SPANISH			
Content	Register, vocabulary, terminology	Fluency and orientation to target text type	Translation brief and professional aspects
<p>Same as translation into English</p>	<p>Same as translation into English</p>	<p>&lt;o&gt; organization</p> <p>&lt;pr&gt; pragmatic mistakes</p> <p>&lt;rtpr&gt; Grammatically correct but it sounds unnatural. The rhetorical effect of the ST is missing. Literal translation.</p> <p>&lt;ist&gt; inappropriate style</p> <p>&lt;rc&gt; inappropriate rendering of cultural reference</p>	<p>&lt;f&gt; layout, wrong accomplishment of style sheet or computer requirements</p> <p>&lt;or&gt; spelling</p> <p>&lt;pt&gt; punctuation</p> <p>&lt;sx&gt; syntax</p> <p>&lt;ccsx&gt; lack of concord</p> <p>&lt;ot&gt; orthotypography</p>

(López, Robinson & Tercedor 2007)

## Biographies

Clara Inés López Rodríguez teaches scientific and technical translation at the Faculty of Translation and Interpreting of the University of Granada (Spain), where she is a senior lecturer and Vice-dean of Students and Work placement. She holds degrees from the Universities of Granada and Portsmouth. Her PhD thesis is on the relation between lexical cohesion, text type and medical translation. Her current research deals with scientific translation, terminology, and the application of corpus linguistics and e-learning to translation. She can be reached at [clarailr@ugr.es](mailto:clarailr@ugr.es)

María Isabel Tercedor Sánchez teaches scientific and technical translation, and audiovisual translation (localisation and subtitling) at the Faculty of Translation and Interpreting of the University of Granada, where she is a senior lecturer. Her principal research interests lie in the areas of terminology, scientific and technical translation, audiovisual translation, and the role of new technologies and e-learning in translation. She can be reached at [itercedo@ugr.es](mailto:itercedo@ugr.es)

---

<sup>1</sup> This research has been carried out within the framework of the projects (1) *MARCO COSTA: Marcos de conocimiento multilingüe en la gestión integrada de zonas costeras (P06-HUM-01481)*, funded by the Andalusian Regional government, with a duration of three years from 2007, and (2) *PUERTOTERM: Knowledge representation and the generation of terminological resources in the domain of Coastal Engineering (BFF 2003-04720)*.

<sup>2</sup> These teaching materials have been proposed in an innovative teaching project called *Análisis y desarrollo de la interfaz imagen-texto en traducción científica y técnica*, funded by the University of Granada.

<sup>3</sup> We have kept the distinction between mistake and error following relevant literature in the field of second language teaching (see López, Robinson and Tercedor 2007: 199). In the same way that there is a distinction between performance and competence, we can '...refer to errors of performance as mistakes, reserving the term error to refer to systematic errors of the learner from which we are able to reconstruct his knowledge of the language to date' (Miller 1966, in Corder 1981:10).

<sup>4</sup> The corpus on oncology was compiled within the framework of the project ONCOTERM [<http://www.ugr.es/~oncoterm>], funded by the Spanish Ministry of Education.

<sup>5</sup> The oncology corpus is described in Faber, López and Tercedor (2001: 177-178).

<sup>6</sup> A detailed description of this corpus will appear in Tercedor and López (2008).

<sup>7</sup> The texts were downloaded from the Intel website [<http://www.intel.com>]: (1) The Intel Wireless Networking Products. Fast relief for the Networking Headaches of the Health Care Providers; (2) How Intel wireless networking can help your business grow; and (3) The Intel® PRO/Wireless 5000 LAN Family of Products.