PARALLEL CORPORA IN TRANSLATOR EDUCATION

GUADALUPE RUIZ YEPES
UNIVERSITÄT HILDESHEIM

ABSTRACT

Mainly four kinds of corpora have been used in the field of translator education: comparable corpora, parallel corpora, parallel learner corpora and 'do it yourself' or 'disposable' corpora. This article focuses on the use of parallel corpora in translator education. As a starting point different approaches to translator education with corpora will be described in order to figure out which one has made use of a methodological background in translation didactics to carry out their research. Secondly different parallel corpora will be presented and the strengths and limitations of different tools, such as bilingual concordancers and translation memories, will be analysed. To round off this article different translation alternatives to the same source text will be explored using EMEA in the context of translators' training.

KEYWORDS: corpus linguistics, translator education

CONTENTS:

1. Introduction

The use of corpora has become very popular during the last decades in translation education. The most used so far are comparable, parallel, parallel learner and 'do it yourself' or 'disposable' corpora.

Comparable corpora[1] are the most used in the field of translator education simply because the accessibility to monolingual texts in the web makes their compilation less time-consuming than the compilation of parallel corpora[2]. Many Translation lecturers encourage their students to use comparable corpora as they would use dictionaries; to look up terminology or lexicogrammatical phenomena such as collocations and semantic prosodies. Kübler (cf. 2003:35), for instance, explains how students can use comparable corpora in order to figure out which adjectives modify certain verbs in the field of cameras. Cosme, on her part, applies comparable corpora for English and French to make students analyse and compare hypotaxis and parataxis in these two languages (cf. 2006: 97).

As for parallel corpora, due to the difficulty of finding texts with their translations in the web, most lecturers use existing corpora in their translation courses instead of trying to compile their own. Marco and Van Lawick (2009), for instance, use COVALT (Corpus Valenciano de Literatura Traducida) to design translation exercises (cf. 2009: 13-16). Kübler's students had access to a parallel corpus composed by *Linux HowTo's* original texts with their translations into French. This corpus allowed them to figure out that some English terms have to be translated into French in a different way depending on the text genre. In a French specialized journal, for instance, the word *firewall* is supposed to be left in English while in an operating manual intended for laymen the term would be translated into French as *garde-barrière* or *coupe-feu* (cf. 2003:39).

There are many reasons that evidence the usefulness of parallel learner corpora for Translation lecturers. By comparing translations of the same original text by different students in the same course the lecturer can detect what translation problems caused them more difficulties. By comparing the translations of different original texts by a single student the lecturer can observe the learning progression of that particular student. Since learner corpora have been used in the context of second language acquisition with great success (cf. Granger 1993, 1998), Bowker and Bennison decided to compile a corpus of digitalized translations by students and called it *Student Translation Archive* (STA). They developed for that special corpus a query language; *Student Translation Tracking System* (STTS) (cf. Bowker and Bennison 2003:104). Their main motive being to build a corpus that is able to "provide a means of identifying areas of difficulty that could then be integrated into the curriculum and discussed in class" (cf. 2003:103). Following this line, the MeLLANGE (Multilingual eLearning in LANGuage Engineering) project was launched in order to create a multilingual annotated aligned corpus of translations produced by translation students and professionals. This Learner Translator Corpus (LTC) was annotated with part-of-speech tags and lemma information, as well as error annotations to identify and classify types of errors. The meta-data concerning the translator is also recorded because this

---

[1] "What links the collections of texts in comparable corpora is that they have been put together according to the same type of criteria (texts of a certain size, on a set topic, from a given period, etc.)"(Kenning 2010: 487).

[2] "The prototypical parallel corpus consists of a set of texts in language A and their translations in language B (e.g. the works of Dickens and their translations into French)." (Kenning 2010: 487).

information may help lecturers to associate types of errors with text production circumstances. A similar translation collection mechanism to the one designed by Bowker and Bennison (cf. 2003: 103-118) was applied in order to collect the meta-data.

Last but not least, disposable corpora ought to be mentioned. They are also known as DIY (do it yourself) corpora and are usually compiled for the fulfillment of one translation assignment (cf. Varantola 2003:55). Disposable corpora are usually comparable corpora with a simple structure composed of texts downloaded from the internet and designed to solve the translation problems of one particular translation assignment. According to Varantola, the know-how about compiling a disposable corpus is part of the translation competence and should be regarded as fundamental in the training of translators:

> As an umbrella term, we could use *corpus management in translation* and
> by this refer to the knowledge and skills needed in the compilation and use
> of corpus information for individual translation assignments. (Varantola
> 2003:57).

In this sense the group *WaCky*[3] has been carrying out significant research. Their main purpose is to deal with the topic Web as Corpus (WaC) and to find out how the web can help us compiling a corpus. Baroni and Bernardini (cf. 2006:10) focus on four different ways of using the web as corpus: the *Web as corpus surrogate*, the *Web as a corpus shop*, the *Web as corpus proper* and the *Web as mega-corpus/mini-web*.

The studies mentioned so far are extremely practice-oriented. They describe situations and exercises carried out in class without referring to any theories or methodological backgrounds although there are at least 7 different major methodological approaches to translator education (cf. Kelly 2005: 11-18):

- Objectives-centred approach: Delisle
- Profession-based and learner-centred approach: Nord
- The socioconstructive approach: Kiraly
- The psycholinguistic approach: Hönig and Kussmaul
- Process-centred approach: Gile
- Task-based approach: Hurtado, Gónzalez Davies
- The situational approach: Gouadec

The PACTE group is an exception in this respect as they carry out research applying corpora in the teaching of translation within the *task-oriented approach* based on Delisle's postulates. Delisle was probably the first translation scholar to establish *teaching objectives* for translator education (cf. Kelly 2005: 11). In his book *L'analyse du discours comme méthode de traduction* (1980) he suggests 23 teaching objectives which can be general or specific. Hurtado Albir (cf. 1999) and González Davies (cf. 2004) follow this research stream and develop translation curricula that are based on teaching objectives. Marco Borrillo and Van Lawik (cf. 2009: 16) design, within the PACTE group, corpus based and corpus driven multiple choice and cloze test exercises for translator education using mainly comparable and parallel corpora.

In the next chapters of this article the use of parallel corpora in the context of translator education will be discussed. Different corpora such as the OPUS corpora collection and BancTrad will be presented. The strengths and limitations of different tools, such as bilingual concordancers and translation memories, will be analysed and the usefulness of parallel corpora in translators' education will be highlighted.

---

[3] *Web-as-Corpus kool yniative.*

2. Parallel corpora

2.1. Alignment

While the comparison between corpora of original texts in different languages has been a very important domain of contrastive linguistics, the comparison of translated texts in different languages and their originals has been favoured in the field of corpus-based translations studies. Nevertheless, comparable corpora have been the most used in translator education, which has to do with the fact that parallel corpora are much more expensive to build. They must be aligned in order to allow observing the existence of equivalent text segments in the languages being analysed. Although there are three levels of alignment: paragraph, sentence and word, the sentence unit is the most used as the basic alignment segment. That does not mean that each sentence in language A is aligned with one sentence in language B throughout the texts, since two sentences in language A could have been translated into just one sentence in language B or vice versa. If the paragraph is taken as the basic alignment unit the results of a search are more difficult to analyse because the searched words and their translations will be embedded within a much larger amount of text. The usual alignment procedure involves using alignment software to align texts automatically and inspecting the output manually in order to correct misalignments. The wide range of alignment software includes tools integrated in translation memory systems such as *SDL Trados WinAlign* and in parallel concordancers such as *Multiconcord*. However, there are also some 'independent' aligners such as Danielsson and Ridings' (1997) 'vanilla' aligner or the Microsoft bilingual sentence aligner (cf. Kenning 2010: 490-491).

Table 1 shows two examples of sentence alignment from the EMEA corpus for the search of the German word 'Blut' with their equivalents in English and Spanish.

	de	en	es
3015064	Daher werden Episoden mit hohen Kalziumwerten im **Blut** gegenüber der alleinigen Einnahme von Kalziumtabletten zur Regulierung des Phosphatspiegels im Blut reduziert	So the number of episodes of high calcium levels in your blood will be reduced as compared to when you are taking calcium tablets alone to regulate the phosphate levels in the blood.	Por lo tanto, el número de episodios de niveles de calcio altos en sangre se verá reducido en comparación con el número de episodios cuando se toman comprimidos de calcio por sí solos, utilizados para regular los niveles de fosfato en sangre.
9075905	Die häufigsten Nebenwirkungen (beobachtet bei 1 bis 10	The most common side-effects with Sebivo (seen in between 1 and 10	Los efectos secundarios más frecuentes de Sebivo (observados entre 1 y 10

von 100 Patienten) sind Schwindel, Kopfschmerzen, Husten, erhöhte Konzentrationen einziger Enzyme im **Blut** (Amylase, Lipase, Kreatin-Phosphokinase, Alanin-Aminotransferase), Durchfall, Übelkeit (Unwohlsein), Bauchschmerzen, Hautausschlag und Erschöpfung (Müdigkeit).	patients in 100) are dizziness, headache, cough, increased levels of some enzymes in the blood (amylase, lipase, creatine phosphokinase, alanine aminotransferase), diarrhoea, nausea (feeling sick), abdominal (tummy) pain, skin rash and fatigue (tiredness).	pacientes de cada 100) son mareo, tos, incremento de los niveles séricos (en sangre) de algunas enzimas (amilasa, lipasa, creatinina, fosfoquinasa, alanina aminotransferasa ALT), diarrea, náuseas, dolor abdominal, prurito cutáneo y fatiga (cansancio).

Table 1: An example of sentence alignment from the EMEA corpus.

2.2. Parallel corpora available in the web

Existing parallel corpora tend to be domain-specific (mainly parliamentary proceedings and technical manuals), considerably small and are scarcely annotated. It is for these reasons that the following corpora deserve special consideration (cf. Xiao 2009):

- *JRC-ACQUIS* (Multilingual Parallel Corpus) comprises texts in 22 languages about the European legislation (over one billion tokens in total). 231 language combinations are aligned. These corpora are accessible free of charges for researchers over the Acquis website.
- *JOC* (Official Journal of European Community) comprises annotated texts (pos-tags) in five languages (English, German, Italian, Spanish and French). 4 language combinations are aligned (English-German, English-Italian, English-Spanish and English-French). In this case only some corpora are accessible free of charge.
- *MLCC* (Multilingual Corpora for Cooperation) comprises a comparable (*Comparable Polylingual Document collection*) and a parallel corpus (Multilingual Parallel Corpus). The parallel corpus is composed of texts in nine languages from the *Official Journal of European Community*. This corpus is accessible through ELRA.
- *OPUS parallel corpus* is composed of European and Assian translated texts from the internet. It comprises hitherto 352 millions of words, but it is growing incessantly because it is being fed constantly with new texts. OPUS is accessible free of charges and allows access to corpora such as OO (the OpenOffice), EUconst (The European constitution) EMEA (European Medicines Agency documents) and EUROPARL (European Parliament Proceedings).
- *CRATER Corpus* (The Corpus Resources and Terminology Extraction project) comprises annotated and aligned texts about international telecommunications in three languages: Spanish, French and English (one million tokens per language).
- *BancTrad* (BT) is a web interface for access to annotated monolingual and parallel corpora. The languages included in the parallel corpora are Catalan, Spanish, English, German and French. The language combinations of Catalan and Spanish with the other three languages are aligned.

Other examples are COMPARA (Portuguese, English), INTERSECT (English, French and German), etc. To mention them all would go beyond the scope of this paper. The Linguistic Data Consortium, hosted by the University of Pennsylvania, offers a catalogue of ready-made parallel corpora available for researchers liable to pay charges (see www.ldc.upenn.edu).

The efficiency of automated retrieval from a corpus depends greatly on the levels of annotations included and, therefore, the types of queries allowed. The possible linguistic levels of annotation in corpora so far are the orthographic, the phonetic, the prosodic, the semantic, the morphologic, the syntactic and the pragmatic levels. However, ideally a corpus should include at least these four levels of annotation: words (tokenization), lemmata (lemmatization), morphologic (Part-of-speech tags) and syntantic function (parsing).

Corpora interfaces also influence information retrieval from corpora. One of the main hazards impeding the quick propagation of work with corpora among translation professionals is the fact that too much time has to be spent in order to become familiar with the different interfaces and query languages (cf. Colominas and Badia 2008: 76). Translation training institutions are aware of this problem and are implementing their systems by creating uniform interface platforms from which researchers, lecturers and students can access more than one corpus applying just one query language. This is the case for several parallel corpora; for instance, the OPUS website puts CQP (Corpus Query Processor) at the disposal of its visitors for the retrieval from its subcorpora.

2.3. Tools

Acccording to Bowker and Barlow there are two main types of tools "that can be used to search for and retrieve information from a bilingual parallel corpus: a bilingual concordancer (BC) and a translation memory (TM)" (2008: 2). While TMs have enjoyed always widespread recognition within the translation industry, BCs don't seem to be found useful by professional translators and are used rather in academic circles. Bernardini and Castagnoli (cf. 2008: 48) admit that corpus use is not only time-consuming but also cognitively demanding and professional translators always work under time constraints and are glad to apply user-friendlier resources. Nevertheless, it is indisputable that both tools have strengths and limitations, so that in certain situations the use of a BC could be more enlightening for a translator trying to figure out, for instance, possible semantic prosodies of words. It is not the aim of this article to compare BCs with TMs since Bowker and Barlow (2008: 1-22) have done so already with *ParaConc* and *SDL Trados*. The main goal of this chapter is, firstly, to present briefly different BCs and TMs in order to gain an overview of available types of tools to retrieve information from parallel corpora and, secondly, to describe to what extent have they been applied in the context of translator education.

2.3.1. Bilingual/Multilingual Concordancers

Most concordancers are not tied to particular languages. They are hence multilingual rather than bilingual, since they are designed to work with a pair of languages (corpora composed by texts in a source language with its translation into one language), or with

various languages, (corpora composed by texts in a source language with its translations into several languages). They should be called, therefore, Multilingual Concordancers (MCs). MCs allow translators to search through parallel corpora "to find information that might help them to complete a new translation" (ibd. 2008: 3). But as already mentioned, professional translation is not the only field of application of MCs; they can also be applied in the fields of contrastive analyses, language learning and translator training/education. These are the most well-known MCs on the market:

- ParaConc: its alignment takes the sentence as basic unit and is based on the Gale-Church algorithm. ParaConc offers a search feature for the analysis of parallel texts, which allows observing the search results in a KWIC layout in both languages, to sort the concordance lines and to move from one language to another. Other utilities of ParaConc are 'hotwords', 'frequency information' and 'advanced search'.
- TransSearch: is a web-based service which also takes the sentence as basic alignment unit. It provides access to a translation database containing millions of sentences translated into English, French and Spanish. TranSearch arranges search results into two columns: the first containing sentences of the source texts in which the word or expression occurs, and the second containing the translation of those sentences. It is a search engine that could be compared to Google, but it is applied to source texts with its translations.
- Find Bi-Text Advantage: Beetext Find has put on the market three versions of this concordancer: for freelancers (Desktop Find Bi-Text Advantage - DBA), for translators employed by small businesses (Small Business Find Bi-Text Advantage - SBBA) and for translators employed by large enterprises (Enterprise Find Bi-Text Advantage - EBA). SBBA and EBA can be networked to allow several members of a team work with the same concordancer and EBA allows users to see documents in their original layout with their graphics and not just as plain texts.
- MultiConcord: is a multilingual concordancer for Windows by Lingua, which combines a parallel concordancer with facilities for the semi-automatic generation of texts. Users are able to add their own pairs of texts to the corpus, using simple mark-up conventions based on SGML.

2.3.2. Translation Memories

The reuse of text segments that have been previously translated in the past is the main function of translation memory systems. The underlying principle is to save every translated sentence in a database creating a parallel corpus of aligned translation units and to compare each new sentence that needs to be translated with this database in order to reuse existing translations. There are many different TMs on the market and the recycling of existing translations is their common feature, but we are going to focus now on what differentiates them:

- SDL Trados Studio: offers, apart from the usual translation memories features, the AutoSuggest (smart suggestions as you type) and the Automated Translation feature (if the translation memory does not provide a match for a particular phrase, this feature helps to translate it).

- Déjà Vu X2: includes different features such as the DeepMiner (carries out cross-analyses of the databases you have built to "*mine*" translations and phrase segments in them), AutoWrite (an interactive predictive translation system even for texts with few or no database segment matches that proposes phrases and sentences mined from your database as you translate), Guaranteed Matches (takes the context of each segment into account to select the most appropriate match), Project Lexicon (offers temporary glossaries) and AutoSearch (shows not just relevant matches but also similar phrases and segments) among other things.
- STAR Transit: In contrast to other TM systems, Transit does not use an exclusively sentence based database. Instead, the context contained in a document remains available in the reference material.
- WordFast: Has been designed to run on multiple platforms, including Windows, Linux, and Mac. Wordfast is filled with features such as real time quality assurance and the ability to hook into multiple glossaries of any format.
- Across: As SDL Trados Studio and Déjà Vu X2, Across offers a feature for the intelligent automatic completion of sentences during translation, based on the source text and the keystrokes entered by the translator.

2.3.3. TMs and BCs/MCs in translator training/education

Most TM providers are interested in becoming popular within the educational community, as they are aware of the fact that a high percentage of translation students will be professional translators in a near future and possible customers for them. Providers try to promote their TMs as the ideal product in the learning environment and some of them even list in their websites the Educational Institutions that buy their products. In the last decades it has become evident that most translators need certain technological tools to practise their work. This is especially the case for those working in the technical field and therefore translating operating manuals, software-surfaces, online-helps and so forth. In order to avoid having discrepancies between what students learn at universities and the requirements of the profession they are being trained for, universities are adapting their curricula to the demands of the market. Some institutions dedicated to the training of translators like the University of Mainz/Germersheim, the School for Applied Linguistics and Translatology (Institut für angewandte Linguistik und Translatologie) at the University of Leipzig or the University Pompeu Fabra in Spain followed the LETRAC (*Language Engineering for Translator Curricula, LE4-8324*) project, which was financed by the EU and created in order to design a common language engineering curricula for Universities. Translation curricula at universities focus therefore more and more on 'training' students for their future as professional translators and since TMs are the most used tools by this community, they are imposing themselves over BCs/MCs. As Bernardini and Castagnoli state

> The term "training" implies that the abilities and competences to be learned are acquirable through practice with the kinds of tools and tasks one will be faced with during one's future professional career, in an environment that reproduces as closely as possible the future work environment. (2008: 41).

There seems to be a vicious circle in this respect; the more lecturers use TMs in 'training' institutions for translators, the more students will use TMs in their future day to day as professional translators, and TM providers can continue selling them to these institutions reassuring them that they are the tools most used in the translation profession. It is in the hands of translation lecturers to change the course of these events in favour of BCs/MCs by using them more often in their modules. Making students familiar with the tools they will have to use in their future professional practice is very important and no one would want to discuss the importance of TMs in this regard. But there are translation problems related to differences between the pair of languages involved in the translation process that must be dealt with using other kind of tools such as BCs/MCs. According to Bernardini and Castagnoli while the term "training" refers to a situation where students are supposed to work in a simulation of real settings in professional translation, like working under time constraints and using TMs, the term "education" refers to a situation where students would have time for activities and reflections related to cultural and linguistic differences of the pair of languages involved in the translation process that future professional translators will not have time for (cf. 2008:42). For this kind of activities and reflections BCs/MCs are a more suitable tool since they allow the student/researcher to observe and analyse aspects of the language that are beyond the scope of TMs. Lecturers should focus therefore more on 'educating' translators, as proposed by Bernardini and Castagnoli (cf. 2008: 39-55), rather than 'training' them.

On the other hand, translators don't really have to choose between a TM or a MC for their daily work since these two tools can be used simultaneously complementing each other. Both articles by Bowker/Barlow and Bernardini/Castagnoli quoted above admit that the use of corpora and corpus search facilities among translators in the long term is "likely to depend on their integration with Computer-Aided Translation (CAT) technology" (Bernardini/Castagnoli 2008: 52). In this sense the authors mentioned above plead for a 'hybrid tool' (Bowker/Barlow 2003: 20) that combines the strengths of TMs and BCs/MCs. While there is no tool currently that combines all these features, efforts in this direction have been undertaken. *MultiTrans*, for instance, is the result of implementing Computer Aided Translation Technologies (CAT) with multilingual publishing and document management operations, which allow users to benefit from both systems and has been referred to as "next generation TM" (Gervais 2006: 48).

TMs and BCs/MCs also complement each other in the sense that a TM can be converted any time into a parallel corpus to be searched by means of a BC/MC.

2.4. Learning to observe and translate patterns with the help of a parallel corpus: EMEA (European Medicines Agency documents)

The fact that corpora can be very helpful during the translational process is something that has been said very often. The question that matters now is: How? Observing patterns in one language and how they have been translated into other languages is a very useful starting point. Sinclair (cf. 2003: xvi-xvii) suggests a seven-stage procedure to read concordance lines for this purpose:
1. Initiate - Search for patterns to the right and left of the node.
2. Interpret - Form a hypothesis that may link these patterns. They may be from the same part of speech or mean something similar.

3. Consolidate - Look further away from the node to figure out if there are variations in the patterns found or additional patterns.
4. Report - Write out your hypothesis to use it in future searches.
5. Recycle - Search again the extended context of the node, to find further examples.
6. Result - Record the results for further studies.
7. Repeat - Repeat the process with more data!

Hunston, on her part, uses the technique of 'accumulative collocation', which "can be used to perform recursive searches that gradually refine what is observed" (Hunston 2010: 163). For example:

The most frequent adjacent-word collocate of *distinguishing* is *between*, so the string *distinguishing between* is then taken as the starting point for further search. The most frequent adjacent collocate of *distinguishing between* is *of*. Taking *of distinguishing between* as the node, the words which most frequently precede this string are: *way, capable, importance, difficulty, means, incapable, task, point, method* and *ways*. (Hunston 2010: 163).

These procedures have been widely applied by many corpus linguists working with monolingual corpora. If used with parallel corpora they enable to compare patterns in different languages and can be very enlightening for the translation process. Searches for the verb *to react* in the EMEA (European Medicines Agency documents) parallel corpus based upon the aforementioned procedures will be presented in order to see, which are the collocates and typical patterns it appears with, and which are the collocates and typical patterns its equivalents for German and Spanish appear with.

1794576	The patient 's ability to concentrate and	**react**	may be impaired as a result of hypoglycaemia.
de	Eine Hypoglykämie kann die Konzentrations- und Reaktionsfähigkeit eines Patienten herabsetzen.		
es	La capacidad de concentración y de reacción de los pacientes diabéticos puede verse afectada por una hipoglucemia.		
609968	As rodents do not	**react**	to human interferon beta , repeated dose studies were carried out with rhesus monkeys .
de	Da Nager nicht auf humanes Interferon beta reagieren, wurden Untersuchungen mit wiederholter Verabreichung an Rhesus-Affen durchgeführt.		
es	Puesto que el Interferón beta humano no es activo en los roedores, los estudios de administración repetida se efectuaron en monos rhesus.		
909354	Check carefully how you	**react**	to the medicines before driving or using any machinery.
de	Überprüfen Sie sorgfältig , wie Sie auf das Arzneimittel reagieren, bevor Sie Auto fahren oder Maschinen bedienen.		
es	Compruebe cuidadosamente cómo reacciona cuando toma medicamentos antes de conducir o utilizar máquinas.		
9786195	A slower injection is preferable in patients who	**react**	to the treatment with " flu-like " symptoms.
de	Bei Patienten , die auf die Behandlung mit grippeähnlichen Symptomen		

es	reagieren , ist eine langsamere Injektion vorzuziehen.		
	Es preferible una inyección más lenta en los pacientes que reaccionan al tratamiento con síntomas "de tipo gripal".		
909354	You should be aware of how you	**react**	to VIAGRA before you drive or use machinery.
de	Achten Sie darauf , wie Sie auf die Einnahme von VIAGRA reagieren , bevor Sie Auto fahren oder Maschinen bedienen .		
es	Debe saber cómo reacciona usted a VIAGRA antes de conducir vehículos o utilizar maquinaria.		
2954425	In clinical studies , antibodies that cross-	**react**	with human insulin and insulin glargine were observed
de	In klinischen Studien wurden mit Humaninsulin und Insulin glargin kreuzreagierende Antikörper mit der gleichen Häufigkeit in den NPH-Insulin- und Insulin-glargin-Behandlungsgruppen beobachtet .		
es	En ciertos estudios clínicos se observaron anticuerpos que presentaban reacción cruzada con la insulina humana y la insulina glargina , con la misma frecuencia tanto en los grupos de tratamiento con insulina-NPH como con insulina glargina.		
371780	In the case of side effects such as dizziness the patient´s ability to concentrate and to	**react**	properly may be impaired.
de	Beim Auftreten von Nebenwirkungen wie Schwindel kann die Fähigkeit des Patienten sich zu konzentrieren und richtig zu reagieren, beeinträchtigt sein.		
es	En caso de que se produzcan efectos adversos como el mareo, puede verse afectada la capacidad de concentración y reacción del paciente..		

Table 2: An example of collocates of the verb "to react" from the EMEA corpus.

The concordance lines in table 2 show that the verb *to react* appears in the EMEA English corpus aligned with its German and Spanish translations mainly followed by the prepositions *with* (example 2954425), *to* (examples 609968, 909354, 9786195 and 909354) and in the multi-word unit *ability to…* (examples 1794576 and 371780). If the search was done only for one of these languages, without the aligned units, the EMEA interface allows more complicated queries using the CQP query syntax. For example, if we wanted to see all the instances of the verb *react* plus a preposition, we would type in *"react" [pos="IN"];* and would get *react with, react against, react in, react within* and *react on*. Since EMEA does not allow this kind of query for several languages simultaneously, because it does not have the same annotation for all languages, we will stick to the examples given in table 2 for English, German and Spanish and will analyse their collocational behaviour and patterns. Some observers might argue that 7 examples are too few to draw conclusions, however, in this study 100 examples were observed and no other instances were found. The examples found kept repeating in slightly different ways, which is a proof for the fact that the compilation of the EMEA corpus takes place automatically and nobody seems to check if texts in the corpus appear repeatedly. Hunston (cf. 2010: 152-166) made a similar search for the verb *react* in the Bank of English and found hundreds of different examples, which is an unequivocal evidence for the diversity of texts available in this corpus.

A. *React* in the noun phrase *ability to…*:

In table 3 students can observe that the English noun phrase *the patient's ability to concentrate and react* has been translated into German using two different grammatical structures, which differ quite from the English one, whereas it has been translated almost straightforwardly into Spanish. The first German translation *Konzentrations- und Reaktionsfähigkeit* makes use of the German's language facility of producing compounds. This is a very typical case for German: when two compounds in succession end with the same radical word the first is substituted by a hyphen for the sake of the economy principle in language. The second translation *die Fähigkeit des Patienten sich zu konzentrieren und richtig zu reagieren*, puts to use the "zu + infinitive" structure in the infinitive clause, which is a quite common structure in German. Both instances have been translated into Spanish as *la capacidad de concentración y de reacción del/de los paciente/s*, which is a similar structure to the English but substitutes the infinitives by nouns. However the Spanish language offers also the possibility of using infinitives to translate this noun phrase: *la capacidad de concentrarse y reaccionar del/de los paciente/s*, which can be considered the parallel structure to the English one.

	en	de	es
1794576	The patient 's ability to concentrate and **react** may be impaired as a result of hypoglycaemia.	Eine Hypoglykämie kann die Konzentrations- und Reaktionsfähigkeit eines Patienten herabsetzen.	La capacidad de concentración y reacción de los pacientes diabéticos puede verse afectada por una hipoglucemia.
371780	In the case of side effects such as dizziness the patient´s ability to concentrate and to **react** properly may be impaired.	Beim Auftreten von Nebenwirkungen wie Schwindel kann die Fähigkeit des Patienten sich zu konzentrieren und richtig zu reagieren, beeinträchtigt sein.	En caso de que se produzcan efectos adversos como el mareo, puede verse afectada la capacidad de concentración y reacción del paciente.

Table 3: 'React' in the multiword unit 'ability to....'

B. *React* followed by the preposition *to*:

The translation into German and Spanish of the preposition *to* when it appears with *react* does not pose major difficulties because these two languages have prepositions at their disposal to express equivalent meanings (*auf* and *a*). Nevertheless, the instances in table 4 show that there are many other alternatives to translate this English verb with its preposition into German and Spanish. The compliance of these alternatives will permit students to learn how to write stylistically qualitative texts, instead of repeating the same syntactical structures throughout their translations.

Examples 609968 and 909354 were translated into German very literally from the English, but quite freely into Spanish. The sentence *Puesto que el Interferón beta humano no es activo en los roedores…* differs from the English by swapping subject and direct object, and translating the verb *to react* by *ser activo (to be active)*. The sentence *Compruebe cuidadosamente cómo reacciona cuando toma medicamentos…* is

a conditional clause and a more natural sounding alternative than *Compruebe cuidadosamente cómo reacciona a los medicamentos…*

Example 17744110 is one of many examples in the corpus where *react to + the name of a medicament* was translated into German as *reagieren auf + die Behandlung mit/die Einnahme von (the treatment/taking) + Name des Medikaments (name of the medicament)*. To translate *reagieren auf + Name des Medikaments* without the noun phrase *die Behandlung mit/die Einnahme von* is not grammatically incorrect nor does it sound unnatural. The inclusion of this noun phrase is probably due to the German striving after precision, and a proof for the theory of explicitation as a common strategy in translation and, therefore, of explicitness as a typical feature of translated texts (cf. Olohan and Baker 2000).

	en	de	es
9786195	A slower injection is preferable in patients who **react** to the treatment with "flu-like" symptoms.	Bei Patienten, die auf die Behandlung mit grippeähnlichen Symptomen reagieren, ist eine langsamere Injektion vorzuziehen.	Es preferible una inyección más lenta en los pacientes que reaccionan al tratamiento con síntomas "de tipo gripal".
609968	As rodents do not **react** to human interferon beta, repeated dose studies were carried out with rhesus monkeys.	Da Nager nicht auf humanes Interferon beta reagieren , wurden Untersuchungen mit wiederholter Verabreichung an Rhesus-Affen durchgeführt.	Puesto que el Interferón beta humano no es activo en los roedores , los estudios de administración repetida se efectuaron en monos rhesus.
909354	Check carefully how you **react** to the medicines before driving or using any machinery.	Überprüfen Sie sorgfältig, wie Sie auf das Arzneimittel reagieren, bevor Sie Auto fahren oder Maschinen bedienen.	Compruebe cuidadosamente cómo reacciona cuando toma medicamentos antes de conducir o utilizar máquinas.
17744110	You should be aware of how you **react** to VIAGRA before you drive or use machinery.	Achten Sie darauf, wie Sie auf die Einnahme von VIAGRA reagieren , bevor Sie Auto fahren oder Maschinen bedienen.	Debe saber cómo reacciona usted a VIAGRA antes de conducir vehículos o utilizar maquinaria.

Table 4: 'React' followed by the preposition 'to'.

C. *React* followed by the preposition *with*:

The translation into German and Spanish of the preposition *with* when it appears with *react* does not pose major difficulties and no interesting alternatives were found in the corpus.

3. Conclusions

My experience as a lecturer for Translation has shown me that a considerable amount of translation errors made by students are related to phenomena, such as collocations and semantic prosodies, that can be explored successfully with the use of parallel and comparable corpora. This is especially the case for those students translating from their mother tongue into a foreign language, but also for those translating into their mother tongue, or for those translating from one foreign language into another foreign language. The student body at European universities is becoming day by day more international. A German-Spanish translation class at a German university, for instance, will include Erasmus students from Spain, who will be translating into their mother tongue, but also German fulltime students translating into a foreign language, and non-German fulltime students, mostly from East European countries, translating from one foreign language into another. It is in the hands of lecturers to facilitate the access to corpora of all kinds to the students to show how to solve translation problems.

Furthermore, if students become familiar with corpora during their stay at their universities, it is very probable that they will continue using them in their future as professional translators. However, in the long term, widespread use of parallel corpora and multilingual concordancers among professional translators will probably depend on their integration with translation memories creating a 'hybrid tool' as suggested by Bowker and Barlow (cf. 2003: 20). Such a tool would improve the performance of CAT systems by allowing two different kinds of searches: "automatic search for full and fuzzy matches in gold standard TMs, and manual concordancing of comparable and parallel texts for hypothesis development and testing" (Bernardini and Castagnoli 2008: 52).

4. Bibliography

Baroni, Marco and Silvia Bernardini (ed.). 2006. *Wacky! Working papers on the Web as Corpus*. Bologna: Gedit.

Bernardini, Laura and Sara Castagnoli. 2008. Corpora for translator education and translation practice. In *Topics in Language Resources for Translation and Localisation,* ed. by Elia Yuste Rodrigo, 39-56. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Bowker, Lynne and Peter Bennison. 2003. Student Translation Archive and Student Translation Tracking System. Design, Development and Application. In *Corpora in translator education,* ed. by Federico Zanettin; Silvia Bernardini und Dominic Stewart, 103-118. Manchester: St. Jerome Publishing.

Bowker, Lynne and Michael Barlow. 2008. A comparative evaluation of bilingual concordances and translation memory systems. In *Topics in Language Resources for Translation and Localisation,* ed. by Elia Yuste Rodrigo, 1-22. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Colominas, Carme and Toni Badia. 2008. The real use of corpora in teaching and research contexts. In *Topics in Language Resources for Translation and Localisation,* ed. by Elia Yuste Rodrigo, 71-88. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Cosme, Christelle. 2006. Clause combining across languages. A corpus-based study of English-French translation shifts. In *Languages in Contrast* 6(1): 71-108.

Danielsson, Pernilla and Daniel Ridings. (1997). Practical Presentation of a «Vanilla Aligner», in *Proceedings of the TELRI Workshop on Alignment and Exploitation of Texts*, Ljubljana, available at http://nl.ijs.si/telri/Vanilla/doc/ljubljana.

Delisle, Jean. 1980. *L'analyse du discourse comme méthode de traduction : Initiation à la traduction française de textes pragmatiques anglais, théorie et pratique.* Ottawa : Presses de l'Université d'Ottawa.

Gervais, Daniel. 2006. Product profile : A corpus-based approach. In *Multilingual 77,* vol. 17, issue 1, 47-49.

González Davies, Maria. 2004. *Multiple voices in the translation classroom: activities, tasks and projects.* Amsterdam: John Benjamins Publishing Company.

Granger, Sylviane. 1993. The International Corpus of Learner English. In *Englisch Language Corpora: Design, Analysis and Exploitation,* ed. by Jan Aarts, Pieter de Haan and Nelleke Oostdijk, 57-69. Rodokopi: Amsterdam.

Granger, Sylviane (ed.). 1998. *Learner English on Computer.* London: Longman.

Hunston, Susan. 2010. Using a corpus to explore patterns. In *The Routledge Handbook of Corpus Linguistics,* ed. by Anne O'Keeffe and Michael McCarthy. 152-166. London/New York: Routledge.

Hurtado Albir, Amparo (dir.). 1999. *Enseñar a traducir. Metodología en la formación de traductores e intérpretes.* Madrid : Edelsa.

Kenning, Marie-Madeleine. 2010. What are parallel and comparable corpora and how can we use them? In *The Routledge Handbook of Corpus Linguistics,* ed. by Anne O'Keeffe and Michael McCarthy. 487-500. London/New York: Routledge.

Kelly, Dorothy. 2005. *A Handbook for Translator Trainers.* Manchester: St. Jerome Publishing.

Kübler, Natalie. 2003. Corpora and LSP Translation. In *Corpora in translator education,* ed. by Federico Zanettin, Silvia Bernardini and Dominic Stewart, 25-42. Manchester: St. Jerome Publishing.

Marco Borrillo, Josep and Heike Van Lawick. 2009. Using corpora and retrieval software as a source of materials for the translation classroom. In *Corpus Use and*

Translating, ed. by Allison Beeby, Patricia Rodríguez Inés and Pilar Sánchez-Gijón, 9-28. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Olohan, Maeve and Mona Baker. 2000. Reporting *that* in translated English: evidence for subconscious processes of explications? *Across languages and cultures* 1(2): 141-158.

Sinclair, John. 2003. *Reading Concordances*. Harlow: Pearson Longman.
Varantola, Krista. 2003. Translators and Disposable Corpora. In *Corpora in translator education,* ed. by Federico Zanettin, Silvia Bernardini and Dominic Stewart, 55-70. Manchester: St. Jerome Publishing.

Xiao, Richard. 2009. Well-known and influential corpora. In *Corpus linguistics: an international handbook*, ed. by Lüdeling, Anke and Merja Kytö, 383-457. Berlin: De Gruyter.