
THE WEB AS A CORPUS AND FOR BUILDING CORPORA IN THE TEACHING OF SPECIALISED TRANSLATION: THE EXAMPLE OF TEXTS IN HEALTHCARE/ LA WEB COMO CORPUS Y PARA LA CONSTRUCCIÓN DE CORPUS EN LA ENSEÑANZA DE LA TRADUCCIÓN ESPECIALIZADA: EL EJEMPLO DE LOS TEXTOS SOBRE SALUD

Elina Symseridou

Aristotle University of Thessaloniki
symseridou@hotmail.com

Abstract: One of the key issues faced by translators and translation students of specialised texts is finding the equivalents of terms in L2 of the field in question. A greater challenge, however, is the formation of the textual environment with the appropriate collocations (adjectives, nouns, verbs) for those terms in the language for special purposes (LSP). The web offers the most convenient and immediate solution by providing access to updated language data presenting the terms in original contexts that help overcome the shortcomings of hard copy lexicographic resources. Taking into account the importance of documentation skills in the training of translators of specialised texts, this paper examines the use of the Web as a Mega Corpus that can be read directly with Google and as a means for constructing corpora automatically with the help of the WebBootCat software. The texts dealt with in this paper are from the healthcare field, which is an important sector of the public service.

Keywords: Ad hoc specialised corpora; WebBootCat; Specialised translation; Translator training.

Resumen : Uno de los retos clave a que se enfrentan los traductores de textos especializados y los estudiantes de traducción es encontrar los equivalentes de términos en la L2 del área en cuestión. Sin embargo, aún mayor resulta el reto de conformar el ambiente textual con las colocaciones apropiadas (adjetivos, sustantivos, verbos) alrededor de esos términos. La red ofrece la solución más conveniente e inmediata al otorgar acceso a datos lingüísticos actualizados que presentan los términos en contextos originales que ayudan a pasarse de las deficiencias de los recursos lexicográficos en forma de libro. Tomando en consideración la importancia de las capacidades de documentarse en la formación de traductores de textos especializados, en este artículo se examinará el uso de la Red como un Mega Corpus que se puede leer directamente con Google y como medio de construcción de cópura de manera automática con la ayuda del soporte WebBootCat. Los textos tratados en este trabajo provienen del área de la salud, que es un sector importante de los servicios públicos.

Keywords: Corpus especializado ad hoc; WebBootCat; Traducción especializada; Formación del traductor.

1. Introduction

The advent of the Internet is perhaps the most revolutionary aspect of the development of written language corpora. The web has made it possible to gain access to thousands of on-line documents and highly specialised texts, consult comparable texts on the same subject in several languages, locate translated texts, and build ad hoc corpora using Google. For these reasons, a growing number of researchers view the Web as a Mega Corpus (see Rundell, 2000; Kilgarriff and Grefenstette, 2003; McCarthy, 2008; Crystal, 2011; Gatto, 2014), the largest ever and the broadest in scope (Borja, 2008).

The aim of this paper is to explore the use of the Internet as a source to collect linguistic data for the purposes of specialised translation. More specifically, the main objectives that are examined are how the Internet can be used as:

- 1) A corpus that can be read directly with Google.
- 2) A means for constructing Corpora automatically with the help of the WebBootCat software.

In order for the above goals to be achieved, the following will be carried out:

- a) Analyze the methods of using Google to investigate the search results as in a concordancer.
- b) Present the results of a case study of automatic construction of a medical comparable corpus in Greek and English on the subject of childhood acute lymphoblastic leukemia with WebBootCat.
- c) Present the results of a pilot study conducted on MA students who used WebBootCat and Trados to translate a specialised text from French into Greek on the mental health issue of schizophrenia; they were then interviewed using a semi-structured oral questionnaire to examine their attitudes towards such a Web documentation tool on the one hand, and the combination of this source with a CAT tool on the other.

Taking into account the importance of documentation and electronic-tools literacy in Public Services Interpreting and Translation (PSIT) training (Sánchez Ramos and Vigier Moreno, 2016: 378), referred to as information mining and technological competences by the EMT Expert Group (2009), I designed a translation activity combining WebBootCat and Trados. Based on the results of the research work described in section 8, I considered that a module comprising ad hoc corpus building and analysis, along with the creation of translation memories on texts for specific subject domains, which current and future students can consult, would bring trainees within a PSIT syllabus closer to real-world translation conditions.

Corpora have long been recognized as a valuable source of documentation and information mining, helping translators become familiar with the subject area of the texts to be translated. Despite the growing interest in academic circles and the proven usefulness and efficiency of corpora for translational purposes, it appears that they have not yet received the recognition they deserve from professional translators (Frankenberg-Garcia, 2015; Picton et al., 2015; Gallego-Hernández, 2015; Carratalá-Puertas, 2015), including translation trainers and trainees (Bernardini and Castagnoli, 2008; Kübler, 2011; Frankenberg-Garcia, 2015; Frérot and Karagouch, 2016). Concerning Greek Universities, corpus-use in specialised translation teaching does not seem to be implemented in any organized or methodical manner. This is based on a review I conducted (unpublished work) in 2017 of the course

catalogues of the four Universities offering postgraduate studies in translation (Aristotle University of Thessaloniki, National and Kapodistrian University of Athens, Democritus University of Thrace, and the Ionian University). Although much discussion has been generated about introducing corpora in translation training syllabuses there is no evidence that these resources are systematically used in pedagogic settings. However, in the European Master's in Translation (EMT) meeting held in March 2015, members of the Working Group on Tools and Translation Technologies featured the use of corpora as a translation resource in training and professional contexts as amongst the most salient themes to be dealt with in the near future (Frérot, 2016: 38). This is because some of the main advantages of corpus use are related to the EMT translator competence model described in section 6.

The literature seems to indicate that in regards to teaching environments, there are only a few curricula that include the use of corpora in translation teaching and production, as opposed to translation memories, which are a prerequisite for the labour market. As far as professionals are concerned, a recent research conducted by Frankenberg-Garcia (2015) in two international translator forums, ProZ.com and Translator's Cafe, reveals that there are no references to corpora in contrast to the questions raised daily in relation to translation memories and CAT tools. The fact that skills related to terminology documentation using corpora are not mentioned nor required in job offers is, perhaps, one of the main reasons why translators and translation trainers remain indifferent towards this tool. Another significant reason why corpora are ignored as sources of documentation (terminology, phraseology and informational content) is their limited or non-existent availability for many specialised subjects and language pairs (Kübler, 2011: 66).

Regarding specialised translation, which is the subject of interest in this paper, “small” specialised corpora, are considered adequate to extract domain-specific terminology. In the absence of printed specialised dictionaries and bilingual parallel texts for many subject areas, the monolingual texts found on the Internet constitute the main source for direct consultation or building text collections.

Research works on the use of corpora have demonstrated that these text collections can be significant educational resources, not only in language learning (Leech, 1997; Granger et al., 2002; Sinclair, 2004; Frankenberg-Garcia et al., 2011; Kilgarriff et al., 2015) –where they were originally used– but also in translator training (Baker, 1995; Aston, 1999; Zanettin et al., 2003; Olohan, 2004; Kunz et al., 2010; Bernardini and Castagnoli, 2008; Beeby et al., 2009; Frérot, 2016). In this paper, the improvement of the working languages, i.e., foreign *and* native is emphasised, both quantitatively (e.g., learning new vocabulary) and qualitatively (e.g., reinforcing syntax and grammar), which should be a parallel process, complementary to translation training. Therefore, this dual use of corpora, as tools for language improvement and as a source of documentation during the translation process, is treated as one and the same activity that contributes to enhancing the knowledge and skills of a future professional translator.

2. Specialised Corpora: Process Facilitation and Quality Improvement of the Translation

Robinson points out (1998: 114) that, under real working conditions, the translator needs to “fake” specific knowledge. Since they will never be able to reach the level of a specialist in a particular field, they need to acquaint themselves well with the thematic subject of the text to be translated. Thus, translators can become mini-experts by consulting specialised corpora, which are a proven wealth of information for locating terms, studying collocations, studying

the grammatical and the syntactic structure of the specialised text as well as for incidental learning.

Locating terms: The large numbers of online texts available in most languages is a major reason why printed dictionaries can no longer be considered the primary source of translation-related information (Vintar, 2008: 153). The web as corpus and for creating corpora provides updated textual material where standard terms, as well as neologisms appear in context.

Studying collocations: Collocations are typical word combinations or put simply, they are words that are found “in each other’s company” (Bowker and Pearson, 2002: 32) and are common in a LSP. One of the challenges in translation practice is the use of a specialised term in a way to create meaning in a LSP. Using a concordancer to study the context surrounding a term, one can find information, such as which adjective precedes or which verb follows, especially when they translate towards a foreign language.

Studying the grammatical and syntactic structure of the specialised text: Some types of texts are intertwined with specific grammatical-syntactic and stylistic features. Studying and analysing a corpus can reveal the syntactic structure that the translator needs to follow to create the appropriate style, in a way that sounds natural or idiomatic in the LSP depending on the communication situation.

Incidental learning: Researchers (Bernardini, 2000, 2001; Varantola, 2003) have pointed out that the use of corpora and, more precisely, of concordancers for text analysis offers a wide range of opportunities for unpredictable, incidental learning. The user can observe, discover unknown uses of terms and expressions, and then verify them.

3. The World Wide Web as Corpus

The World Wide Web (WWW) as corpus is a new, evolving field of research. According to Gatto (2014: 7), the enormous potential of the web as a linguistic resource has been addressed under the umbrella term “Web as Corpus” to designate several methods that treat the Internet as their primary source for the implementation of the corpus-linguistics approach. The various methods used to exploit the possibilities offered by the Internet should not be considered as competing with each other or with more traditional methods of using corpora; rather, they should be seen as a useful addition for establishing practices in corpus-linguistics. The main objection to the use of WWW as a corpus concerns data representativeness. Kilgarriff and Grefenstette (2003: 8) state that “representativeness” is a fuzzy notion and that outside very narrow, specialised domains, we do not know with any certainty what existing corpora might be representative of. Fletcher’s point of view (2001) appears to be more diplomatic and well founded. Although he does not consider the web to be a corpus, he acknowledges that it can be used as such. He points out that the Internet is clearly a gold mine for translators, as it contains updated documents on almost all issues and in almost all languages. This confronts us with the question of what constitutes or will constitute third-generation mega-corpora. Flowerdew (2012: 39) believes that it is, undoubtedly, the Web, that is constantly evolving and offers access to vast numbers of texts, of any content, directly and in many cases, still free of charge. As it can be seen from the research, developments and supporting evidence make a strong case for the fact that –at least for specialised subjects– the Web can be considered a huge corpus or can be used as a source for building corpora.

3.1 Example of Using the WWW as a Dictionary and a Concordancer

A survey based on the MeLLANGE questionnaire (2005) stated that the use of the Internet as a virtual corpus and search engines as “concordancers” is a common practice for translators and translation students (94.4%). In effect, this means that, since they are already using a search engine with which they are familiar and which gives them access to the largest amount of information possible, there is no need to attempt to learn new software.

In the example below, I looked for the multi-word term *Σύνδρομο ΔΕΠ* (*Sýndromo DEP*), which was encountered in a Greek medical report to be translated into English, for a patient seeking treatment in leukaemia abroad. First, I located the meaning of the acronym in Greek, and then I reintroduced it in the words *σύνδρομο διαχ...* (Figure 1). As shown below, Google stemming technology displayed the term in the drop-down menu, before all the words had been entered. The first, third and fourth results provide the equivalent in English (*disseminated intravascular coagulation* and *DIC*). By choosing the multi-word term given by the stemming system, the term appears in the Google search results page in bold within a context, as in a concordancer. There are specific Google search techniques with wildcards, operators, file type etc., which although they cannot be discussed here due to space constraints, make it relatively easy to find relevant information on the web.

Google

σύνδρομο διάχυτης ενδοαγγειακής πήξης

σύνδρομο διάχυτης ενδοαγγειακής πήξης

Αναζήτηση

Όλα Χάρτες Εικόνες Βίντεο Ειδήσεις Περισσότερα ▾ Εργαλεία αναζήτησης

Περίπου 1.260 αποτελέσματα (0,38 δευτερόλεπτα)

Διάχυτη ενδοαγγειακή πήξη - disseminated intravascular coagulation ...
www.iatronet.gr/iatriko-lexiko/diaxyti-endaggeiaki-pixi.html ▾
iatronet.gr - Παθολογική κατάσταση στην οποία τα μονοπάτια της πήξης έχουν υπερδιεγερθεί, με αποτέλεσμα μια διάχυτη παρά εντοπισμένη ενεργοποίηση των ...

Διάχυτη ενδοαγγειακή πήξη - Care
www.care.gr/post/742/diaxyti-endoaggeiaki-piksi ▾
Φυσιολογικά υπάρχει δηλαδή μία ισορροπία μεταξύ των συστημάτων πήξης και ινωδολυσης. Το σύνδρομο της διάχυτης ενδοαγγειακής πήξης είναι αποτέλεσμα ...

[PDF] Πήξη II Ενεργοποίηση της Πήξης στους Βαρέως Πάσχοντες Ασθενείς ...
anesthesia.gr/download/TOMOS_18_19/tefhos_39/225.pdf ▾
κατάσταση η οποία κλινικά εμφανίζεται ως διάχυτη ενδοαγγειακή πήξη (DIC). ... "Η DIC είναι ένα επίκτητο σύνδρομο που χαρακτηρίζεται από την γενικευμένη ...

Τα κριτήρια για τη διάγνωση της DIC - Αιμοστατικό σύστημα ...
omedicine.info/el/kriterii-diagnostiki-dvs-sindroma.html ▾
Διάχυτη ενδοαγγειακή πήξη (DIC) - Μία από τις πιο κοινές επίκτητες ... Όταν σηπτικό σοκ Σοβαρό Οξύ DIC- σύνδρομο ανιχνεύεται σε 100 % περιπτώσεις.

Figure 1. The web as a dictionary and a concordance

4. Specialised Corpora and New Technologies: The WebBootCat Tool on the SketchEngine Platform

The most common problem faced by translators is finding correct terminology for translating specialised texts from a specific field. Traditional hard copy resources, such as dictionaries – even specialised ones– have proven to be insufficient (Burgos Herrera, 2006: 358) as well as costly and time-consuming concerning looking up words (Dziemianko, 2012: 334).

It appears that the solution might lie in the Internet itself, which is a store of copious data on language that is easily accessible and whose popularity as a tool seems to be on the rise with a growing number of professional translators (Enríquez Raído, 2013: 83-84). The updated specialised texts that can be found online are essential resources for language professionals who routinely work with LSPs.

Baroni and Bernardini (2004: 1-4) responded to this challenge by creating the BootCat¹ tool, aimed at building ad hoc corpora within minutes. Its basic method of operation is as follows:

- 1) The user chooses a few representative seed words, i.e., terms that are expected to be typical of the domain of interest.
- 2) The server sends queries with the seed words to Google.
- 3) The server collects the pages retrieved by Google formed from the seed words (Baroni et al., 2006: 1).

SketchEngine will display the relevant web pages on a list and the user can exclude the links that seem unreliable (such as public blogs, newspaper articles, news websites, etc.) or inadequate (such as literature), by removing the ticks. On completing these steps, a first-pass specialised corpus is created. Using the option “Keywords/Terms”, the single- and multi-word terms of the corpus will automatically be extracted and displayed in two columns side by side. The process can also be iterated choosing new terms as seeds to give a “purer” specialist corpus. According to the creators, there is no need for the user to repeat this process more than three times (Baroni and Bernardini, 2004: 2).

Cross-border healthcare is increasing among EU citizens and residents who seek care under Directive 2011/24/EU or Regulation (EC) N° 883/2004². In multilingual Europe, cross-linguistic communication is increasingly frequent, especially when it relates to accessing services. The development of effective public services such as cross-border healthcare has contributed to the need for public service translation and interpreting as a means to access healthcare services. Quality healthcare means that providers of treatment need access to the relevant information for patients (e.g. insurance policy, medical records or previous prescriptions) in their own language, as well as producing content to be shared with other professionals (Angelelli, 2015: 2-5). Medicine is a highly specialised field with a wide variety of materials and one of the topics treated within the discipline of PSIT. To explore the function of WebBootCat as a corpus-building tool, I chose the topic of childhood leukaemia as a medical text. The screenshot below (Figure 2) shows the seed words from the first-pass English medical corpus of 266,176 words that was built using only five multi-words and three single-words.

¹ Bootstrapping Corpora and Terms

² Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32011L0024>

Sketch Engine

Medcorpus: Extracted keywords / terms

Change extraction options Download singlewords: [TBX](#) [CSV](#). Download multiwords: [TBX](#) [CSV](#).

Singlewords and multiwords are ordered by [keyness score](#). The score and corpus frequency (leading to the respective concordance) are displayed in parentheses. Highlighted words were used as seeds in a previous WebBootCaT run within this corpus.

<< Back to corpus files Use WebBootCaT with selected words

Single-word	Score	F	RefF	Multi-word	Score	F	RefF
<input type="checkbox"/> medulloblastoma	W 2,745.88	995	1,171	<input type="checkbox"/> pubmed abstract	W 705.91	243	0
<input type="checkbox"/> oncol	W 1,047.22	370	588	<input type="checkbox"/> radiation therapy	W 676.56	266	16
<input type="checkbox"/> clin	W 650.89	280	5,669	<input type="checkbox"/> palliative care	W 445.77	167	10
<input type="checkbox"/> embryonal	W 600.40	216	1,029	<input type="checkbox"/> posterior fossa	W 340.57	118	1
<input type="checkbox"/> cns	W 577.11	486	32,834	<input type="checkbox"/> advanced cancer	W 314.94	113	5
<input type="checkbox"/> radiotherapy	W 470.40	344	25,536	<input type="checkbox"/> brain tumor	W 249.78	88	3
<input type="checkbox"/> neuroectodermal	W 462.76	161	259	<input type="checkbox"/> bone marrow	W 245.74	205	160
<input type="checkbox"/> medulloblastomas	W 448.64	156	248	<input type="checkbox"/> residual disease	W 244.67	84	0
<input type="checkbox"/> pubmed	W 429.31	245	14,951	<input type="checkbox"/> stem cell	W 236.00	102	29
<input type="checkbox"/> resection	W 427.50	266	18,349	<input type="checkbox"/> cell rescue	W 227.27	78	0
<input type="checkbox"/> tumors	W 426.88	1,057	140,581	<input type="checkbox"/> lymphoblastic leukemia	W 201.16	69	0
<input type="checkbox"/> gy	W 393.41	175	6,658	<input type="checkbox"/> stem cell rescue	W 198.26	68	0
<input type="checkbox"/> supratentorial	W 390.70	136	280	<input type="checkbox"/> acute lymphoblastic leukemia	W 195.36	67	0
<input type="checkbox"/> atra	W 387.76	149	2,665	<input type="checkbox"/> high-dose chemotherapy	W 185.17	64	1
<input type="checkbox"/> chemotherapy	W 370.80	991	153,544	<input type="checkbox"/> adjuvant chemotherapy	W 164.90	59	5
<input type="checkbox"/> craniospinal	W 355.65	123	138	<input type="checkbox"/> overall survival	W 158.08	59	10
<input type="checkbox"/> prognostic	W 332.30	169	10,872	<input type="checkbox"/> central nervous system	W 156.23	112	122

Figure 2. Extracted keywords/terms from an ad hoc medical corpus

As shown above, WebBootCat cannot translate terms or locate equivalents; the user needs to know what the equivalent of a term is in order to search for it in the corpus and retrieve examples of the word node with the help of the concordancer. One may ask the following: “How useful is this tool if we have to know what we are searching for?” On the one hand, the terms to be selected and translated into the language of the corpus to be compiled can be the easiest to locate on an online dictionary or the Web. For the said corpus I have entered the terms “acute lymphoblastic leukaemia”, “red blood cells”, “white blood cells”, “bone marrow”, “petechiae”, “platelets” and “chemotherapy”. These terms were identified looking at only a few articles regarding childhood leukaemia on the Internet. As displayed in Figure 3, only eight terms were needed with WebBootCat to generate a corpus of hundreds of thousands of words, where other terms are “hiding” and are discovered through incidental learning. For example, in the ad hoc medical corpus created with WebBootCat, we searched the term “immature cells” and the concordancer generated, among others, the following result:

www.LLS.org The identification of mature and immature blood cells in peripheral blood smears and bone

The term “peripheral blood smears” constitutes an element of incidental learning. A quick search on Google gives the explanation in English: “laboratory work-up that involves cytology of peripheral blood cells smeared on a slide” and further search with Greek keywords on Google reveals the equivalents “επίχρισμα αίματος” (*epíchrismá aimatos*) for “blood smear” and “αντικειμενοφόρος πλάκα” (*antikeimenofóros pláka* - official) or “πλακάκι” (*plakáki* - colloquial) for “slide”. Thus, this is how two more terms often found in medical texts on leukaemia have been incidentally located and identified.

On the other hand, an ad hoc corpus is not a source to be used alone and it cannot, at least for the time being, totally replace other sources of terminology documentation in the translation process. However, it must be kept in mind that the main aim of this paper is the study of the contexts surrounding terms and not how to translate terms *per se*.

The software was initially freely available for download and was widely used to produce both special and large general-language corpora (Sharoff, 2005; Baroni et al., 2006). However, since the software had to be downloaded and installed, this presented a barrier for people with no or few computer systems skills (Baroni et al., 2006: 1). Thus, its designers presented a new version of the web service, namely, the WebBootCat (Baroni et al., 2006). In addition to its availability as standalone software, BootCat is now available as an online service via the Sketch Engine website (www.sketchengine.co.uk). The user does not need to install the software on their PC, but he/she can directly use the program installed on a remote server. The server stores a copy of the built-in corpus and the user can either upload it in a query tool like the Sketch Engine or export it as a full or indexed “vertical” text and save it to their PC for off-line analysis with any other tool (e.g. Wordsmith Tools or AntConc).

4.1. Innovative SketchEngine tools at the translator's service

In addition to the concordancer and the frequency list, SketchEngine offers innovative tools to facilitate, accelerate and improve translators' and trainees' work.

4.1.1. Word Sketch

As stated on the site:

“A word sketch is a one-page summary of the word’s grammatical and collocational behaviour. It shows the words’ collocates categorised by grammatical relations such as words that serve as an object of the verb, words that serve as a subject of the verb, words that modify the word etc.”³

It often happens that the translator finds an equivalent term in L2 but does not know how to fit it in a prepositional structure; which verb follows, which adjective precedes; to put it differently, which words surround the term. The text material found in corpora is a very appropriate source for context information, and the Word Sketch negates the need for reading thousands of concordance lines to draw conclusions. All the information is displayed in a compact format, as shown in Figure 3, saving the user valuable time.

³ <https://www.sketchengine.co.uk/user-guide/user-manual/word-sketch/>

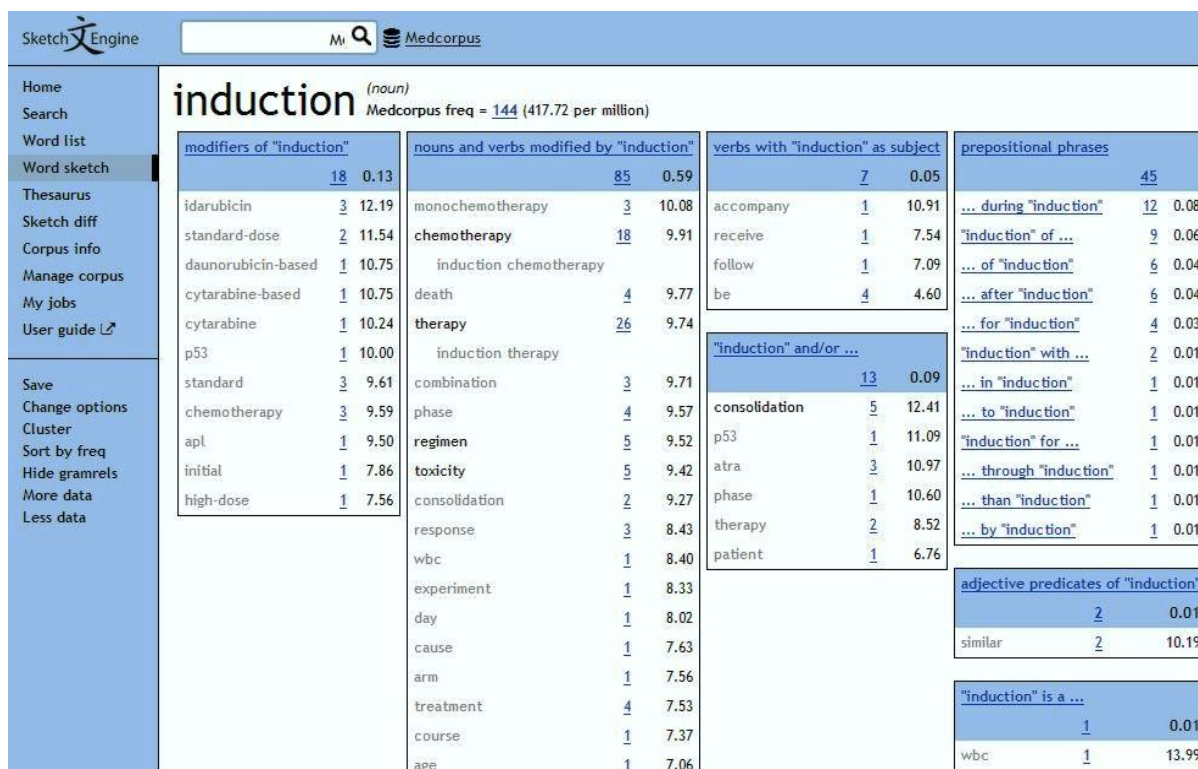


Figure 3. Word Sketch results for the term induction

4.1.2. Word Sketch Difference

This is an extension of the Word Sketch that allows one to study the uses of two synonyms or antonyms comparatively: “Word sketch difference is used to compare and contrast two words by analysing their collocations and by displaying the collocates divided into categories based on grammatical relations”.⁴

Results appear in green or red for every lemma. In the screenshot below (Figure 4), Sketch Engine assigned the green colour to “therapy” and the red colour to “chemotherapy”. Green collocates are more closely related to “therapy”, red collocates to “chemotherapy”. Stronger colour indicates stronger collocations, white means similarity.

⁴ <https://www.sketchengine.co.uk/user-guide/user-manual/word-sketch-difference/>



Figure 4. Sketch Difference for the comparative study of the terms therapy and chemotherapy

4.1.3. Thesaurus

A Thesaurus, often referred to as a dictionary of synonyms, is a reference work that categorizes words into groups according to their conceptual similarity⁵. The Sketch Engine Thesaurus is automatically generated using algorithms that analyze the corpus⁶. In Figure 5, the special term “methotrexate” has been introduced, and it can be seen that for this term Thesaurus generates a frequency list and a word cloud with lemmas that can be found in its contextual environment. By selecting any word from either the list or the cloud we are directed back to Sketch Difference to compare the search word and the selected word.

⁵ <https://www.sketchengine.co.uk/automatic-thesaurus-synonyms-for-all-words/>

⁶ <https://www.sketchengine.co.uk/quick-start-guide/thesaurus/>

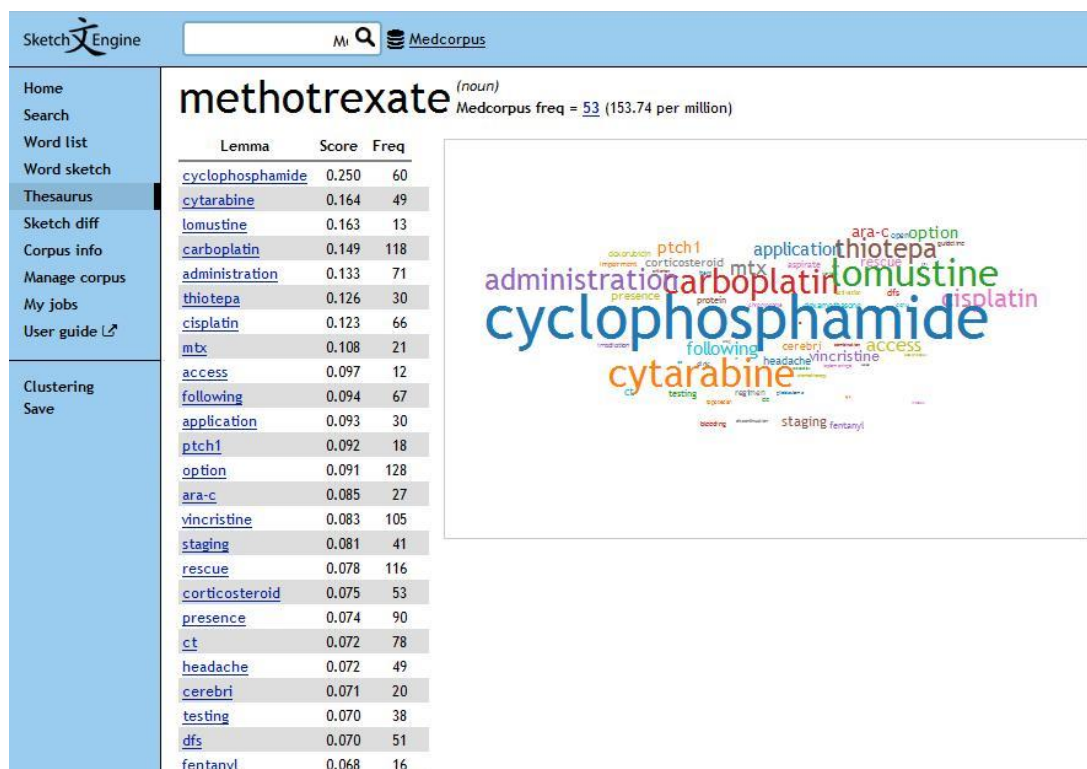


Figure 5. Frequency list and word cloud generated by Thesaurus

4.2. Construction of a comparable corpus with WebBootCat

The basic method for constructing a comparable corpus (CC) is to create two monolingual text collections, on the same subject in two different languages, in the following sequence:

- 1) Enter seed words for a specific subject in L1,
- 2) Construction of a special corpus in L1,
- 3) Enter equivalent seed words in L2,
- 4) Construction of a special corpus in L2 (Kilgarriff et al., 2011: 2).

These steps result in building a comparable corpus. To study the functions of SketchEngine exploring an ad hoc CC built with WebBootCat, for the same subject matter, childhood acute lymphoblastic leukaemia as the aforementioned medical English corpus, the comparable Greek corpus, of 247,236 words was created, entering the equivalents of the terms used in the former. To harvest more texts and create a purer corpus, with only six new terms, the software collected texts of another 236,927 words, that is, a total of 477,326 words for the Greek corpus.

With the creation of a specialised medical comparable corpus in English and Greek, SketchEngine provides an additional service, which is an extension of the above mentioned Word Sketch. With the Bilingual Word Sketch tool, which is an extension of the Word Sketch tool, the user can insert one lemma in L1 and its equivalent in L2 and observe the context surrounding them. In the example below (Figure 6) the word “leukemia” and its equivalent in Greek (λευχαιμία) were chosen, restricting the search to words that appear before this term. It is observed that among the results, the most frequent types of leukaemia in

the two languages of the comparable corpus appear (e.g. *acute* – *οξεία* - *οξία*, *lymphoblastic* – *λεμφοβλαστική* - *lemphovlastikí*, *myeloid* – *μυεολογενής* - *myelogenís*, etc.)

modifiers of "leukemia"		words before "leukemia"		words just before "leukemia"				
	445	1.51	1,420	2.37	530	0.89		
acute	164	12.94	οξεία	124	10.73	λεμφοβλαστική	71	11.91
acute lymphoblastic leukemia			οξεία λεμφοβλαστική λευχαιμία			οξεία λεμφοβλαστική λευχαιμία		
lymphoblastic	66	11.98	λεμφοβλαστική	67	10.44	μυελογενής	60	11.70
of acute lymphoblastic leukemia			οξεία λεμφοβλαστική λευχαιμία			Χρόνια μυελογενής λευχαιμία		
promyelocytic	53	11.75	Οξεία	78	10.24	λεμφοκυτταρική	36	11.02
newly diagnosed acute promyelocytic leukemia			μυελογενής	55	10.20	μυελογενής λευχαιμία (
myeloid	20	10.41	Χρόνια	40	9.57	Χρόνια λεμφοκυτταρική λευχαιμία		
acute myeloid leukemia .			λεμφοκυτταρική	34	9.56	μυελογενή	27	10.63
childhood	25	10.28	Χρόνια μυελογενής λευχαιμία			οξεία μυελογενή λευχαιμία		
childhood acute lymphoblastic leukemia			λεμφοκυτταρική			οξεία λευχαιμία		
chronic	8	9.05	χρόνια λεμφοκυτταρική λευχαιμία			οξεία	32	10.24
lymphocytic	6	8.76	χρόνια	47	9.52	προμυελοκυτταρική	14	9.72
myeloid	5	8.50	χρόνια μυελογενής λευχαιμία			οξεία προμυελοκυτταρική λευχαιμία		
acute	5	8.49	μυελογενή	24	9.08	μυελομονοκυτταρική	14	9.72
myelomonocytic	4	8.19	οξεία μυελογενή λευχαιμία			μυελομονοκυτταρική λευχαιμία		
refractory	4	8.12	προμυελοκυτταρική	14	8.30	μυελοβλαστική	11	9.38
plasma	4	8.06	οξεία προμυελοκυτταρική λευχαιμία					
monocytic	3	7.78						

Figure 6. Bilingual Word Sketch presentation of the term leukemia

5. The Pros and Cons of Building Corpora from the Web or Consulting it Directly

According to Fletcher (2007: 27), the multitude of online texts is a challenge for linguists and other language professionals: the self-renewing, machine-readable multilingual text corpus of the Internet is readily accessible, but it is difficult to evaluate its content and use it efficiently. However, strong arguments are being put forward regarding renewing existing corpora or creating new ones with online content. The WWW ensures:

- **Updated data:** Soon after the compilation of standard reference corpora –let alone of dictionaries– their content becomes outdated, whereas the Internet is an inexhaustible reservoir of machine-readable texts on contemporary issues (Fletcher, 2007: 25).
- **Terminology and neologisms.** A characteristic of our modern world is the rapid development of technology and the sciences, and with it, the influx of technological and scientific terms into the common core of the language is continuously increasing (Stein, 2002: 2). The translator cannot access libraries or all of the online specialised magazines and, since the Internet plays an increasingly important role in the lives of an ever-growing number of people and is becoming more and more interactive, the general mechanisms and principles of new-word

developments may not be too different from what goes on outside the Web⁷ after all (Kerremans et al., 2012: 61). A research conducted by Kristiansen (2013: 134) on researchers' blogs to detect specialised neologisms in economic-administrative domains indicated that a high degree of disciplinary relevant neologisms were detected (71.56%).

- **Personalization:** for a great number of specialised subjects and language pairs there are hardly any specialised corpora. Using the web as a corpus or for corpus building, translators and trainees can collect text material according to their field of interest and needs.
- **Representativeness and sampling:** representativeness and sampling could be two key points for a linguist to support that online texts are not a reliable source. However, specialised texts, such as articles and papers written in a LSP by field experts are representative of the language that a community of scientists uses to communicate information. Nonetheless "representativeness", as already mentioned, is considered by many scholars a next-to-impossible goal for many corpora (Gatto, 2014: 146).

Besides the benefits, there are shortcomings:

- **Availability:** as Zanettin (2002: 5) points out not all topics, not all text types, not all languages are equally suitable or available,
- **Reliability:** Before using information one finds on the Internet for assignments and research, it is important to check its accuracy and to establish that the information comes from a reliable and appropriate source. Considering specific word search strategies (boolean operators, wildcards, file extensions, site: edu, site: gov, etc.) to evaluate web content is a practice that should be learnt in academic environments.

Working with corpora offers authentic material and empirical data in language research, language teaching and translation. More than ever, the adoption of a corpus-based methodology can divert the focus away from the teacher (as a repository of answers) and place it onto the students' needs, as well as on the translation process and the sources used to complete this process. Corpora (as sources) and corpus linguistics (as a methodology) promote a sense of discovery that increases students' motivation and autonomy. Furthermore, such a methodology encourages the use of IT tools and the processing of information in electronic form.

6. The Importance of Students' Active Participation in Building Corpora: Autonomy Granting and the Exemption of the Teacher from the Role of Authority

As shown by the research works presented in section 8, when the teacher him/herself does not run a corpus-based translation class, it is not easy to persuade the students to use corpora only by suggesting ready-made collections among other documentation sources (online dictionaries, encyclopaedias, glossaries, etc.). In contrast, instructors who are experienced in the use of corpora can guide students through building ad hoc specialised corpora to meet specific needs. Student participation in assessing and collecting material from the web (manually or automatically using software, such as WebBootCat), reveals the principles that

⁷ The Indexed Web contains at least 4.51 billion pages (<http://www.worldwidewebsize.com/>)

lie at the basis of the creation and use of corpora and gives them the incentive to use such tools.

Thus, they become familiar with the electronic medium and learn how to conduct research, evaluate online sources, and, collect information for terminology extraction. In other words, they learn how to solve translation problems in practice (active learning, as suggested by Kiraly, 2000). In a corpus-based translation class, the teacher is not a repository of answers, but acts as a guide showing students how to use a corpus to get answers to their questions, how to participate in the learning process, and how to act independently. Moreover, the teacher no longer feels the pressure to guide the correction of translations, to answer questions about a specific field he/she is not familiar with, or to provide the final “correct” translation as expected by the students. The teacher can use the information found in a corpus as tangible proof to support his/her or their students’ translational choices considered to be more appropriate.

The act of making students aware of the process and possibilities of building corpora equips them with the professional competences they should acquire according to the EMT framework (2009), such as:

- Information-mining competence, requiring the skills and ability to search for information by looking at the various sources in a critical way.
- Domain-specific knowledge, which includes information on specialist fields comprising the knowledge to be used in professional translation practice.
- Language competence, by observing the style, phraseology, collocations and idioms used in the LSPs.
- Technological competence, especially in handling translation related software and terminology management.

Regarding the translation programme, the adoption of a corpus-based teaching methodology allows for the inclusion of more specialised texts in the curriculum, even if the teacher is not acquainted with a discipline, as well as the creation of a collaborative learning environment.

It is worth noting that the roles that both teachers and students undertake in a corpus-based module are consistent with Kiraly’s (2000: 184-185) social-constructivist approach to translator training, which puts the emphasis on students’ autonomy and cooperation. What Kiraly proposes is translation courses, during which the teacher helps students learn through practice. In the period that he refers to, his estimation of the education system was based on the transmissionist approach, that is, the active transmission of knowledge by the teacher and passive listening on the part of the students. In contrast, social constructivism sets the epistemological basis for creating knowledge and aims to encourage the learner to act responsibly, independently and effectively. The fundamental principle of socio-constructivist education is active participation in authentic and empirical learning by assigning real or at least simulated translation tasks with the complexity that characterizes them (Kiraly, 2000).

Cognitive theories, and in particular constructivism, attach great importance to the individual’s inner, mental processes. According to these theories, learning is not transmitted but it is a process of the personal construction of knowledge, which is based on previous knowledge (and has been accordingly modified to be coupled with new knowledge). Learning requires rearrangement and reconstruction of the individual's mental structures so that they adapt to new knowledge, but also “adapt” the new knowledge to the existing mental structures. By adopting the inquiry-based method and the development of internal learning stimuli, i.e. practices that are part of the theory of constructivism, the student learns to

construct the knowledge he/she needs and to use it according to the requirements of the tasks to be accomplished.

The use of educational software, such as an automated corpus building tool, supports the idea of building knowledge by the learners themselves, as they attempt to solve problems and in their effort, interact with the material environment (which includes the educational software), their fellow students and the teacher. The student explores, discovers gradually, makes assumptions that he/she verifies or contradicts, in an educational environment that supports this process.

7. Educational Model to Develop Student Translation Competence

As Rodríguez Inés (2009: 130) suggests, with technological developments revolutionizing the translation profession, translator trainers should draw on the new pedagogical approaches that are available to train translators for the 21st century.

This paper argues in favour of the need to design corpus-based translation courses, considering the working environment of professional translators and market requirements. The design of translation tasks using corpora is necessary in reflecting the problems faced by professional translators and for students to be trained in real working conditions. Corpus-based work offers a wide range of possibilities in the translation classroom and can be easily adapted to competence-based training, focusing on “learning how to learn” and professional requirements. Work with corpora is based on activities that involve searching and analyzing data, and, therefore, strengthens the sense of learning through discovery, as well as through reorganizing and building upon previous knowledge (Rodríguez Inés, 2009: 130).

Given that this paper focuses on the translation of specialised texts, the contribution of existing corpora is considered to be insufficient. Taking into account that we are now in an era of pervasive computing, in conjunction with the Internet as an inexhaustible source of data, I propose the following educational model.

Aiming at developing student translation competence (information-mining, domain-specific, language, technological competence), it is necessary for a translator training programme to include the use of the Web as corpus and as a source for creating corpora, provide tips on how to collect specialised texts and data from the web, promote the building of ad hoc specialised-corpora with WebBootCat, and the creation of Translation Memories and glossaries on domain-specific subjects.

The methodology proposed here is to move beyond standard –and inadequate– corpora and create ad hoc text collections with the help of WebBootCat; then use the linguistic information found in these corpora to translate texts, treated within the syllabus, with the help of Trados in order to create a Translation Memory (TM) and glossary for future reference. Thus, students develop skills and competences related to translation practice, such as information-mining and terminology documentation, as well as computer and software literacy. Furthermore, this activity helps create a tank of texts and glossaries for current and future trainees of MA programmes to consult. All of the above practices are essential in real working conditions so it is imperative that they are included in training programmes.

8. Research Work on Corpora Introduction to Specialised Translation Modules of the MA Translation Programme of the Aristotle University of Thessaloniki (AUTH)

To examine students' familiarisation with corpora and technology and to record their attitude towards a tool for automatic corpus-building, it was deemed necessary to conduct a field observation, on top of reviewing course catalogues of the Greek Universities mentioned at the beginning. Thus, in the academic year 2016-2017, I was invited to conduct a standalone detailed two-hour training session on the use of SketchEngine and the WebBootCat to the class "Translation of Specialised Texts and Terminology documentation". The class was attended by nine out of the twenty-five students on the Interdepartmental Programme Translation and Interpretation, who had French to Greek as a second working language (the first one being English to Greek for all students). This training session, divided into a theoretical and practical section, comprised the pilot study for a larger research work that will follow. The aim of the pilot study was to pin-point possible drawbacks in the methodology (analysed in more detail below) in order to make the necessary adaptations to the research study that will follow. The theoretical part focused on ad hoc specialised corpora and the functions of SketchEngine, whereas the practical part included the translation of a French specialised text with the help of WebBootCat and Trados. Since the class teacher (who was not part of the research team) was working with texts on the subject of healthcare, it was predetermined that we work with a text on schizophrenia.

As already mentioned, the class was attended by nine out of the twenty-five students on the programme, of which seven held a BA in English or French Language and Literature and had little theoretical and practical experience in translation, while each of the two remaining came from the School of Primary Education and the School of Drama, respectively, having had no previous knowledge in translation studies or translation practice. To have students study how to render part of the French text into Greek, I asked them to build an ad hoc specialised corpus in the target language that would provide them with linguistic information about terminology, collocations and phraseology, as well as scientific information to acquire subject field knowledge. The methodology included the following stages: assessment of retrieved sources (links), compilation and corpus analysis, creation of translation memory (with Trados), translation memory uploading to SketchEngine, and, terminology extraction (with SketchEngine Keyword/Term extraction tool).

Before starting the corpus-building process, in the theoretical section, I made a brief introduction on how to locate PDFs using the file type extension right after entering domain-specific terms on Google. Since the text on schizophrenia was a medical topic, it was more suitable to search for PDFs rather than random sources. This is because PDF content on specialised subjects are usually published articles, conference papers, dissertations written by field specialists.

Once the Greek monolingual corpus was created, I uploaded the French text to the Moodle e-learning platform of AUTH and asked the students to translate two pre-defined paragraphs into Greek with the help of the monolingual Greek corpus that they had built with WebBootCat, as comparable linguistic material. It should be noted that the free trial version of the tool was used in the pilot study; the subscription version could not be used due to lack of funding. Students were introduced to the construction of corpora by inserting either seed-words or the URLs of PDF files in WebBootCat but they were free to choose which way suited them better. Thus, seven out of nine students created two corpora on the subject of schizophrenia; the first using Greek seed words and the second by locating PDF files on the web and inserting their URLs in the said tool. One student used only PDFs, while another used only seed words. Although the majority of students were not able to find many PDFs on the subject, one student was able to locate quite a significant number that helped her build a

representative corpus. In conjunction with their corpus, the students also used the web as corpus for additional help. As already mentioned above, the user must enter a few representative terms on WebBootCat or locate PDF files entering these terms on Google (with PDF extension) and then inserting their URLs on the tool.

Moreover, during the assessment phase all the students who used seed words commented that a great number of sources seemed to be inappropriate or unreliable. These were public blogs, literature, newspaper articles, as well as other types of texts written by non-specialists. By contrast, those who used PDFs to create their corpus noticed that the sources were more reliable and the content of the texts more targeted, since they were published works written by specialists and located on health care sites.

As a secondary exercise, the students were asked to upload the French text to Trados and translate the pre-defined excerpt of text into Greek using this CAT tool. The reason I thought it was interesting to go through this process is because after completing the translation they could upload their translation memory (French-Greek) to Sketch Engine and use the Keyword/Term option to extract terminology and save a glossary in txt for future reference.

During the translation process, one student came across the term “*déficits cognitifs*”. Since she was not sure whether to translate “*cognitive*” as “*γνωστικά*” (*gnostikà/cognition*) or “*γνωσιακά*” (*gnosiakà/knowledge*) she used Sketch Difference to study the contexts of the two notions in the Greek corpus. While all students wondered whether “*hallucinations*” had to be translated as “*παραισθήσεις*” (*paresthísis/delusions*) or “*ψευδαισθήσεις*” (*pseudaesthísis/hallucinations*) another student found in Sketch Differences that “*ψευδαισθήσεις*” (*pseudaesthísis/hallucinations*) can be either “*ακουστικές*” (*akoustikés/auditory*), “*οσφρητικές*” (*osfritikés/olfactory*) or “*γευστικές*” (*yefstikés/gustatory*) and that this was the correct translation of the term “*hallucinations*”. In addition, another multi-word term “*Programme Intégréatif de Thérapies*” that was difficult to locate in Greek was identified by 3 students who found the equivalent “*Απαρτιωτικό Θεραπευτικό Πρόγραμμα*” (*Apartiotikó Therapeftikó Próγραμμα/Integrated Psychological Therapy*) using the web as corpus and then inserting the Greek translation as a search-word to the Greek corpus to further investigate the context surrounding it.

A few days after the field observation, the participants responded to an oral interview based on a semi-structured questionnaire including 19 questions. The aim of the questionnaire was to record the following:

- 1) Whether they were aware of corpora and more specifically of Sketch Engine and WebBootCat.
- 2) Their views on the use of the web as corpus and ad hoc corpora in translator training,
- 3) Their views on the use of technological tools in translator training.
- 4) Whether they would continue to use WebBootCat as translators.

Regarding the awareness on corpora in general, eight students responded that reference had been made to corpora in one or two modules. One stated that she had had absolutely no contact with corpora, while another had scarcely engaged in corpora in the module “*Translation and Technology*” during Erasmus. A point of concern is that only one student could recall having once used the BNC to study English collocations in the classroom, but none of the others could name the ready-made corpora referred to them in the modules. They commented that the approach had been theoretical and that they had been given no literature; although some ready-made corpora had been presented, they did not move on to practice.

Regarding the different types of corpora, the interviews revealed that students could not tell the difference between Parallel and Comparable Corpora despite having been familiarised with Eur-Lex, Europarl, Linguee, and Glosbe in the context of their graduate and postgraduate studies. None of them knew what an ad hoc corpus was, and one student confused it with AntConc, which they had once used in a class. Only two out of the nine knew Sketch Engine before the presentation but had never used it; however, all of them considered it a very useful tool for the purposes of specialised translation when the content of the corpora created is reliable, that is, PDF-based.

Specifically, the study participants recognised that a tool of automatic creation of corpora could help trainees to:

- a) Become familiar with the subject matter.
- b) Study collocations and the linguistic environment of a term.
- c) Locate terms.
- d) Improve their information-mining competence.

As far as the use of the Web as corpus and as a dictionary is concerned, they all granted that the Internet alone is not a sufficient source of documentation, especially when the texts to be translated are large; they agreed that corpora can be more targeted as to their content regarding a specific subject and, they all supported the inclusion of technological tools in a specialised translation module as necessary in facilitating their work.

When asked about the use of Trados and the degree of familiarisation with one of the most in-demand CAT tools on the translation market –with the exception of one student who owned the tool– all the students felt confident about the technical know-how but not about application in the translation practice and production⁸. To the question: “What is your opinion about the combined use of technological tools, such as WebBootCat-Sketch Engine?” one student replied that the said combination was not needed, as she found the use of WebBootCat efficient. The rest of them considered the teaching and their learning of how to use multiple tools necessary for their development as translators, and some even found this combination interesting.

Finally, when asked if they would continue to use such a tool for building corpora, two students replied positively, whereas the others expressed reservations about the time-consuming practice of analysing corpora under real working conditions where time is limited. Nevertheless, they recognised the utility of such a tool for the translation of large texts, and especially when one specializes in a subject area, as it offers a better perspective on the topic and the related linguistic information, in contrast to the content displayed on the Google search results page. For small texts, they believed that searching the web is faster and more immediate.

Considering that the aim of this pilot study is to measure students’ familiarisation with corpora and technology and to record their attitude towards a tool they had not used before, the main results are summarized as follows:

- 1) Students had no explicit knowledge of corpora nor had used SketchEngine and the WebBootCat before.
- 2) They already used the web as a source of documentation but they were not aware of specific techniques to make their searches more efficient; moreover, they had never created ad hoc corpora before, which they found quite useful.

⁸ The fact that masters’ programmes offer core courses on TM technology with an emphasis on technical know-how rather than translation competence has also been criticized in research works (Chung-ling 2006; Sauron 2007; Kenny 2007).

- 3) They were concerned about their technological knowledge and they were keen on the inclusion of more tools in translation training.
- 4) As far as building corpora in real working conditions is concerned, they had their reservations because of time constraints, but they would try it for large specialised texts.

The reason why I interviewed the students after exposing them to the resources used is because the results of a semi-structured questionnaire given to MA students to measure their knowledge on corpora were not convincing. More specifically, starting back in 2015, as a pilot study, I distributed a semi-structured questionnaire of 12 questions to 9 second-year students of the interdepartmental programme in translation at AUTH. The results clearly showed that students had no explicit knowledge about the different types of corpora and how they could be exploited. However, it was obvious that they wanted to give the impression that they were not totally ignorant in this subject. More specifically, seven out of nine participants responded that they included corpora to the sources they used for documentation. Two of them could not name which corpora they used while four said they used Eur-Lex, the Corpus of Greek Texts and the BNC (suggested by some teachers), the contents of which are either irrelevant or insufficient for the purposes of specialised translation. Although, there was a question on which online dictionaries they used, two students included Linguee (which is an online dictionary based on parallel texts), among corpora. Only two admitted they do not use corpora at all, whereas none of the students referred to ad hoc corpora. This is why suggesting a tool or a source to students is almost worthless, unless they are well informed, and trained to use it, as well as assigned tasks incorporating that tool.

Distributing a questionnaire alone was deemed fruitless and it was necessary to modify the methodology to draw conclusions based on more qualitative results. Thus, it was decided to first make an introduction on corpora and the tools to be used; then replace the written questionnaire with an interview in an effort to make a distinction between what they thought they knew (before the introduction) and what they learnt about corpora (after the introduction).

Qualitative results regarding translation could not be produced due to time limitations. The measurement of students' efficiency and improvement of their translations using such a tool for creating corpora requires systematic classroom observation. The methodological tools were considered effective and offered interesting data. Undoubtedly, these findings can by no means be universalized, since they are based on a limited sample. However, they are indicative of a trend. The same research will be conducted on the other sixteen students of the MA programme in the English to Greek language pair.

Overall, the findings of the pilot study indicate that MA students interested in becoming professional translators, on the one hand, were not aware of this valuable tool or the different types of corpora and their possibilities and, on the other, seemed to have little or no idea of what an ad hoc corpus is and how they could create one automatically to serve specialised-translation related needs, using an online tool. These findings lend support to the importance of knowing how to use online tools for the creation of specialised corpora and to the necessity of incorporating such training in the relevant MA programmes at Greek universities. Having said that, more extensive research needs to be carried out in all the Greek universities offering translation MA courses.

9. Conclusion

More than ever, the use of corpora has diverted the focus away from the teacher (as a repository of answers) and place it onto the students' needs, as well as on the translation process and the sources used to complete this process. Corpora (as sources) and corpus linguistics (as a methodology) promote a sense of discovery that increases both student motivation and autonomy. Moreover, a corpus-based teaching methodology equips students with the professional competences (information-mining, thematic, language and technological competence) provided for in the reference framework of the European Master's in Translation (EMT Expert Group, 2009: 3). Searching the web is not an uncommon practice for most users; however, in a translator training programme such practice needs to be put under a more organized framework, providing the necessary information for more effective search techniques and web content evaluation. Student participation in corpus-building encourages the use of such source, allows the inclusion of more specialised texts in the curriculum, and releases the teacher from the role of authority. Finally, a methodological framework combining learning objectives with competence development sets the foundations for an understanding of translational reality. However, despite evidence of the manifold advantages of corpus use, their inclusion in academic environments remains scarce and students are unaware of the possibilities this tool can offer them.

References

- Angelelli, C. V. 2015. Study on Public Service Translation in Cross-border Healthcare. *Final Report for the European Commission*. Directorate-General for Translation, Luxembourg: Publications Office of the European Union. [Available at: <https://publications.europa.eu/en/publication-detail/-/publication/6382fb66-8387-11e5-b8b7-01aa75ed71a1>].
- Aston, G. 1999. *Corpus Use and Learning to Translate*. Textus, 12: 289-313.
- Baker, M. 1995. *Corpora in Translation Studies. An Overview and Suggestions for Future Research*, Target, 7 (2): 223-243.
- Baroni, M. and Bernardini, S. 2004. *BootCaT: Bootstrapping Corpora and Terms from the Web*. In Proceedings of LREC 2004. Lisbon: ELDA: 1313–1316. [Available at: http://sslmit.unibo.it/~baroni/publications/lrec2004/bootcat_lrec_2004.pdf].
- Baroni, M., Kilgarriff, A., Pomikálek, J. and Rychl, P. 2006. *WebBootCat: Instant Domain-Specific Corpora to Support Human Translators*. [Available at: <http://www.mt-archive.info/EAMT-2006-Baroni.pdf>].
- Beeby, A., Rodríguez Inés, P., Sánchez-Gijón P. 2009. *Corpus Use and Translating: Corpus Use for Learning to Translate and Learning Corpus Use to Translate*. Amsterdam and Philadelphia: John Benjamins.
- Bernardini, S. 2000. "Systematising Serendipity: Proposals for Concordancing Large Corpora with Language Learners". Burnard, Lou and McEnery, Tony (eds.) *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang: 225-234.
- Bernardini, S. 2001. "Spoilt for Choice': A Learner Explores General Language Corpora". Aston, Guy (ed.) *Learning with Corpora*. Houston TX: Athelstan: 220-249.
- Bernardini, S., Castagnoli, S. 2008. "Corpora for Translator Education and Translation Practice". E. Yuste Rodrigo (ed). *Topics in Language Resources for Translation and Localisation*. Amsterdam/Philadelphia: John Benjamins: 39-55.
- Borja, A. (2008). "Corpora for Translators in Spain". Anderman, G. and Rogers, M. (eds.)

- Incorporating Corpora: The Linguist and the Translator*. Multilingual Matters Ltd: 243-252.
- Bowker, L., Pearson, J. 2002. *Working with Specialised Language, A practical Guide to Using Corpora*. London and NY: Routledge.
- Burgos Herrera, D-A. 2006. "Concept and Usage-Based Approach for Highly Specialized Technical Term Translation". Gotti, M. and Sarcevic, S. (eds.) *Insights into Specialized Translation*. Peter Lang: 347-366.
- Carratalà-Puertas, I. 2015. *Corpus y traducción profesional, una relación tan omnipresente como invisible, IV Congreso Internacional CULT Corpus Use and Learning to Translate*, mayo de 2015, Alicante.
- Chung-ling, S. 2006. *Using Trados's WinAlign Tool to Teach the Translation Equivalence Concept*. *Translation journal*, 10 (2).
- Crystal, D. 2011. *Internet Linguistics*. London: Routledge.
- Dziemianko, A. 2012. "On the Usefulness of Paper and Electronic Dictionaries". Granger, S. and Paquot, M. (eds.) *Electronic Lexicography*. OUP Oxford: 319-337.
- EMT Expert Group 2009. *Competences for Professional Translators, Experts in Multilingual and Multimedia Communication*. [Available at: http://ec.europa.eu/dgs/translation/programmes/emt/key_documents/emt_competences_translators_en.pdf].
- Enríquez Raído, V. 2013. *Translation and Web Searching*. Routledge.
- European Parliament and the Council 2011. *Directive on the Application of Patients' Rights in Cross-Border Healthcare*. [Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32011L0024&from=EN>]
- Fletcher, W. H. 2001. "Concordancing the Web with KWIC Finder". American Association for Applied Corpus Linguistics. Third North American Symposium on Corpus Linguistics and Language Teaching, Boston, MA, 23-25 March. [Available at: <http://www.kwicfinder.com/FletcherCLLT2001.pdf>].
- Fletcher, W. 2007. "Concordancing the Web: Promise and Problems, Tools and Techniques". Hundt, M., Nesselhauf, N., Biewer, C. (eds.) *Corpus Linguistics and the Web*. Rodopi, Amsterdam and New York: 25-46.
- Flowerdew, L. 2012. *Corpora and Language Education*. UK: Palgrave Macmillan.
- Frankenberg-Garcia, A.; Flowerdew, L. and Aston, G. 2011 *New Trends in Corpora and Language Learning*. London: Bloomsbury: 62-80.
- Frankenberg-Garcia, A. 2015. *Training Translators to Use Corpora Hands-On: Challenges and Reactions by a Group of Thirteen Students at a UK University*. Edinburgh University Press, 10 (3): 351-380.
[Available at: <http://www.eupublishing.com/doi/abs/10.3366/cor.2015.0081>].
- Frérot, C. 2016. "Corpora and Corpus Technology for Translation Purposes in Professional and Academic Environments. Major Achievements and New Perspectives". *Cadernos de Tradução*, On-line version ISSN 2175-7968. [Available at: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2175-79682016000500036].
- Frérot, C. and Karagouch, L. 2016. "Outils d'aide à la traduction et formation de traducteurs: vers une adéquation des contenus pédagogiques avec la réalité technologique des traducteurs". *ILCEA, Revue de l'institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*. [Available at : <http://ilcea.revues.org/3849>].
- Gallego-Hernandez, D. 2015. "The use of Corpora as translation resources: a study based on a survey of Spanish professional translators". *Perspectives: Studies on Translatology*: 375-391.
[Available at: www.tandfonline.com/doi/abs/10.1080/0907676X.2014.964269].
- Gatto, M. 2014. *Web as Corpus: Theory and Practice*. A&C Black.

- Granger, S.; Lerot, Hung, J. and Petch-Tyson, S. 2002. *Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.
- Kenny, D. 2007. "Translation memories and parallel corpora: Challenges for the translator Trainer". Kenny, D. and Ryou, K. (eds.) *Across Boundaries: International Perspectives on Translation Studies*. Newcastle: Cambridge Scholars Publishing: 192-208.
- Kerremans, D.; Stegmayr, S. and Schmid, H-J. 2012. "The NeoCrawler: Identifying and Retrieving Neologisms from the Internet and Monitoring Ongoing Change". Allan, K., Robinson J. (eds.) *Current methods in historical semantics*. Walter de Gruyter: 59-96.
- Kilgarriff, A., Grefenstette, G. 2003. "Introduction to the Special Issue on the Web as Corpus". *Journal Computational Linguistics*, 29 (3). [Available at: <http://www.kilgarriff.co.uk/Publications/2003-KilgGrefenstette-WACIntro.pdf>].
- Kilgarriff, A., PVS, A., Pomikálek, J. 2011. *Comparable Corpora BootCat*. [Available at: [https://www.sketchengine.co.uk/wp-content/uploads/Comparable Corpora BootCaT 2011.pdf](https://www.sketchengine.co.uk/wp-content/uploads/Comparable_Corpora_BootCaT_2011.pdf)].
- Kilgarriff, A.; Marcowitz, F., Smith, S. and Thomas, J. 2015. *Corpora and Language Learning with SketchEngine and SKELL*. [Available at: <https://www.sketchengine.co.uk/wp-content/uploads/2015/04/2015-corpora-submitted.pdf>].
- Kiraly, D. 2000. *A Social Constructivist Approach to Translator Education: Empowerment from Theory to Practice*. NY: Routledge.
- Kristiansen, M. 2013. *Detecting Specialised Neologisms in Researchers' Blogs* [Available at: <https://bells.uib.no/index.php/bells/article/view/366>].
- Kübler, N. 2011. "Working with Corpora for Translation Teaching". Frankenberg-Garcia, A., Flowerdew, L. and Aston, G. (eds.) *New Trends in Corpora and Language Learning*. London: Bloomsbury: 62-80.
- Kunz, K., Castagnoli, S. and Kübler, N. 2010. "Corpora in Translator Training: A Program for an eLearning Course". Gile, D., Gyde, Hansen, G., Pokorn, N. (eds.) *Why Translation Studies Matters?* Amsterdam and Philadelphia: John Benjamins: 195-208.
- Leech, G. 1997. "Teaching and Language Corpora: A Convergence". Wichmann, A., Fligelstone, S., McEnery, A. and Knowles, G. (eds.) *Teaching and Language Corpora*. London: Longman: 1-23.
- McCarthy, M. 2008. "Accessing and Interpreting Corpus Information in the Teacher Education". *Language Teaching*, 41 (4): 563-574.
- MeLLANGE 2007. [Available at: <http://corpus.leeds.ac.uk/mellange/ltc.html>].
- Olohan, M. 2004. *Introducing Corpora in Translation Studies*. London and NY: Routledge.
- Picton, A.; Fontanet, M.; Pulitano, D. and Maradan, M. 2015. "Corpora in Translation: Addressing the Gap between the Scholars" and the Translators' Point of View". CULT Conference, 26-29 May, Alicante.
- Rodríguez Inés, P. 2009. "Evaluating the Process and not just the Product when Using Corpora in Translator Education". Beeby, A.; Rodríguez Inés, P. and Sánchez-Gijón P. (eds.) *Corpus Use and Translating: Corpus Use for Learning to Translate and Learning Corpus Use to Translate*. Amsterdam and Philadelphia: John Benjamins: 129-150.
- Robinson, D. 1998. *Becoming a Translator: An Accelerated Course*. Routledge. [Available at: <http://ilts.ir/Content/ilts.ir/Page/142/ContentImage/Becoming%20a%20Translator%201st%20Ed.pdf>].
- Rundell, M. 2000. "The Biggest Corpus of All". *Humanising Language Teaching*, 23. [Available at: <http://www.hltmag.co.uk/may00/idea.htm>].

- Sánchez Ramos, M. and Vigier Moreno, F. 2016. *Using Corpus Management Tools in Public Service Translator Training: An Example of its Application in the Translation of Judgments*.
 [Available at : <https://reference.research-publishing.net/publication/chapters/978-1-908416-35-3/449.pdf>].
- Sauron, V. 2007. “Les Nouvelles Technologies dans l’Enseignement de la Traduction : L’Exemple de la Traduction Juridique”. Lavault, E. (ed.) *Traduction Spécialisée: Pratiques, Théories, Formations*. Bern: Peter Lang: 207-224.
- Sharoff, S. 2005. “Open-Source Corpora: Using the Net to Fish for Linguistic Data”. *International Journal of Corpus Linguistics*, 114: 435-46.
- Sinclair, J. 2004. *How to Use Corpora in Language Teaching*. Amsterdam and Philadelphia: John Benjamins: 125-152.
- Stein, G. 2002. *Better Words: Evaluating EFL Dictionaries*. University of Exeter Press.
- Varantola, K. 2003. “Translators and Disposable Corpora”. Zanettin, F.; Bernardini, S. and Stewart, D. (eds.) *Corpora in Translator Education*. UK: St. Jerome Publishing.
- Vintar, Š. 2008. “Corpora in Translator Training and Practice”. Anderman, G. and Rogers, M. (eds.) *Incorporating Corpora: The Linguist and the Translator*. Multilingual Matters Ltd: 153-163.
- Zanettin, F. 2002. *Corpora in Translation Practice*. [Available at: https://www.academia.edu/30887110/Corpora_in_translation_practice].
- Zanettin, F., Bernardini, S., Stewart, D. 2003. *Corpora in Translator Education*. UK: St. Jerome Publishing.