Content list available at http://ijltr.urmia.ac.ir

Iranian Journal of Language Teaching Research

Urmia University

Teaching English for Medical Translation: A Corpus-based Approach

María del Mar Sánchez Ramos [a, *]

[a] *University of Alcalá, Spain*

ABSTRACT

Developing the appropriate linguistic and documentation skills in specialised domains is one of the greatest challenges in English for Specific Purposes (ESP) courses. Among other tools, corpora are intended as efficient resources students can use to improve and develop their thematic, terminological, and phraseological knowledge (Sánchez Ramos, 2020). Based on a non-experimental mixed-methods study, this paper reports the feasibility of a corpus-based approach to teaching medical translation at a postgraduate level. It analyses how a group of 40 postgraduate students at the University of Alcalá (Madrid, Spain) perceive the compilation of their own corpus and the usefulness of corpus management tools (CMTs) in solving thematic, terminological, and phraseological issues. The software LancsBox (Lancaster University) was used, which is a recently created CMT that incorporates automatic tagging, cutting-edge statistics information, and innovative visualization options. The quantitative and qualitative analyses of this data suggest that the ESP students were highly satisfied with the corpus-based approach. Our findings also reveal that postgraduate students incorporated CMTs during the whole course as documentation resources for their translation tasks, and they are willing to use them in their near future professional career. These results show the potential usefulness of CMTs and offer a way of integrating them into the ESP curriculum.

Keywords: ESP; mixed-methods methodology; corpus-based approach; students' perception; medical translation

© Urmia University Press

ARTICLE HISTORY

Received: 29 July 2019 **Revised version received:** 9 Jan. 2020

Accepted: 11 June 2020 **Available online:** 1 July 2020

* Corresponding author: FITISPOS Research Group, Department of Modern Philology, University of Alcalá, Madrid, Spain
Email address: mar.sanchezr@uah.es
© Urmia University Press

10.30466/ijltr.2020.120887

Introduction

Using English for specialised domains (e.g. English for business, English for medical settings, etc.) requires a high level of accuracy. In translation practices, for example, students need to develop linguistic, documentation, and technological skills, among other abilities (EMT, 2017; Sánchez Ramos, 2020). As a result of the expansion of the Internet and the technological era, along with traditional electronic documentation tools (e.g. dictionaries, terminological databases, and glossaries), corpora and terminology extraction tools have also become important resources in the specialised domain. Compiling a corpus often becomes necessary in specialised contexts such as medical communication, since the use of dictionaries or glossaries is insufficient due to the highly specific nature of the subject. Whereas the process of compiling a corpus often proved arduous and tedious even a few years ago, today's digital panorama is such that a specialised corpus can be compiled in a matter of hours, depending on its size and intended purpose. Within Translation Studies, different researchers have highlighted the enormous benefits corpora have in specialised domains (Buendía-Castro & López-Rodríguez, 2013, Vigier Moreno, 2016). However, despite the potential advantages, many students state they hardly know how to apply them in real language and specialised translation tasks (Sánchez Ramos & Lázaro Gutiérrez, 2019). Given this situation, we feel that it is of paramount importance to carry out a research to determine if a corpus-based methodology is useful in translation tasks. In other words, we will determine if our corpus-based instruction was useful for the language and translation development of the participants involved in the research.

Accordingly, the results of a non-experimental mixed-methods study are hereby presented with the aim of exploring the use of specialised corpora within English for Specific Purposes (ESP) courses. The research addresses the following questions:

- What is the perception of a group of postgraduate students towards the use of specialised corpora?
- Are corpus management tools appropriate documentation resources for these students' professional development?

The remainder of the article presents the existing literature and the underlying theoretical approach of our research, mainly focusing on the principles of corpus-based studies. Next, the methodology is described, and quantitative and qualitative instruments are used to elicit the main research data from the participants. The paper continues with the main results of the research, which suggest that the ESP students were highly satisfied with the corpus-based approach. The discussion elucidates how this demonstrates the potential of using CMTs and offers a way of integrating them into the ESP curriculum.

Literature review

Specialised vocabulary forms the core content of any ESP course. Together with the well-known classification of receptive and productive vocabulary (Graves, 2006), specialised vocabulary comprises professional language, including technical, semi-technical, and general vocabulary. Technical vocabulary is defined as "content words whose meaning is restricted to the specific subject, characterizes the specific language as an individual area of the global language and constitutes the terminology of the domain" (Rea Rizzo, 2009, p. 162). Semi-technical vocabulary refers to general content words whose meaning becomes specialised in a particular domain, but which is understandable in general contexts (Rea Rizzo, 2009). Finally, general vocabulary "refers

to those words that are known to the general speaker and coexist with technical and sub-technical vocabulary in a specialised text" (Pérea-Barberá & Bocanegra del Valle, 2014, p. 132).

Specialised vocabulary is also closely related to terminology. As Cabré (2004) observes, the subject matter is not sufficient to classify a text as specialised; rather, such classification is dictated by the way in which the subject is dealt with. The terms, understood as lexical units transmitting knowledge, establish and represent the degree of specialisation of a text. Terminological variants are, therefore, the key elements in teaching medical translation (Calonge Prieto, 2009), and these include morphological variants, orthographic variants, elliptical forms, and abbreviations. Calonge Prieto (2009) is categorical when she states that elaborating terminological repertoires that include these variants is a useful practice in teaching medical translation. For this purpose, students must follow the procedures of corpus linguistics.

The origins of corpus linguistics can be traced back to the pre-Chomskian period (McEnery, Xiao, & Tono, 2006), where followers of the structuralist tradition used a corpus-based methodology to provide empirical results based on observed data. This area of research, defined as "the study of language based on examples of 'real life' language use" (McEnery & Wilson, 2004, p. 1) has opened many possibilities for the study of language. Specifically, corpora have attracted increasing attention in translation studies in recent years. Depending on the nature of their work, researchers have used corpora to investigate the features of translated texts (Baker, 1995), or the possibilities of using corpora as translation and terminology resources (Arce Romeral & Seghiri, 2018; Toledo Baéz, 2019; Zhu & Wang, 2011). There is thus no doubt that the study of real language in its proper context can provide valuable information. In this regard, Calzada Pérez (2005, p. 216) highlights that:

> […] due to their flexibility and adaptation possibilities, corpus-based translation studies merge descriptive and linguistic methodologies, process and product analysis, display of details or wide behaviour patterns, which are interesting both because of formal issues and cultural, ideological and even literary aspects.

The variety of corpus types reflects the diversity of their applications in language study, and different studies describe a wide range of corpora (Faya, 2016; Kübler & Zinsmeister, 2015). One notable contribution is that of Laviosa (2002), who classifies the main corpora used in translation studies. First, she divides monolingual corpora into simple (collection of texts in a single language) and comparable corpora (collection of original texts in language A and translated texts in that same language A); bilingual corpora are then classified as parallel (original texts in language A and their translations in language B) or comparable corpora (original texts in language A and original texts in language B); finally, for multilingual corpora, a distinction is made between parallel corpora (original texts in various languages with their respective translations) and comparable corpora ("bi/multilingual corpora made up of two or more sets from the same subject domain(s)"; Laviosa, 2002, p. 36).

In addition to these types, attention has recently turned toward what are called virtual corpora, also known as *ad hoc*, disposable, or do-it-yourself (DIY) corpora (Varantola, 2003). A virtual corpus is a collection of texts developed from electronic resources and compiled for the purpose of providing linguistic or field-specific information. In comparison with other types of large corpora that comprise millions of words (for example, the *British National Corpus* or the *Brown Corpus*), virtual corpora are notable for their small size (Seghiri, 2011). Virtual corpora are much smaller, with specific and extremely precise subject matter, and quality takes precedence over quantity. These corpora are employed to fulfil particular documentation needs or as a pedagogical resource for students in the elaboration of didactic material (Charles, 2012). Virtual corpora can also serve as documentation resources during different language learning tasks (e.g. EAP writing,

ESP) and prove useful for "earning specialised terms, identifying words that appear together, assimilating grammatical uses, deciphering stylistic and generic issues, and delving more deeply into the conceptual subject matter of the texts to be translated" (Calzada Pérez, 2005, p. 170). As also noted by Charles (2012), virtual corpora allow students to take more responsibility and boost their autonomy.

The benefits and pedagogical implications of the use of corpora in ESP contexts have already been highlighted by various scholars (Boulton, 2010; Bondi, 2001; Charles, 2012; Gaviolli, 2005). The use of corpora in the ESP classroom is important (and arguably essential) because the learner needs to be able to utilize different types of information sources in order to acquire and maintain expert knowledge in a particular subject area. There are occasions when the use of traditional lexicographical tools (e.g., specialised dictionaries, glossaries, and databases) is insufficient, due to the specific nature of the classroom assignments. From a pedagogical point of view, the design of *ad hoc* lexicographical resources will lead to a higher degree of specialisation and a much quicker and more effective resolution of linguistic problems, such as the identification of phraseological patterns, terminology, word frequency, and discursive elements.

Corpora are also part of students' documentation repertoire when learning a language. They are examples of real and authentic language use, which makes corpora valuable tools for language learning and acquisition. The arrival of the Internet and the generalised use of technology in various spheres have radically changed the way in which documentation tasks and the documentation process are carried out. As Pinto Molina (2005, p. 19) states:

> Information is an ever-growing resource and it is moving towards uncontrolled over-abundance. Due to its exponential growth, half of the information available today has been accumulated in the last fifty years. But it is not only the quantity and enormous supply of data that drives the constant evolution of our society: the rational, ordered, productive and intelligent use of information is an essential factor in our development.

Studies confirm the consolidation of electronic sources as documentation resources for different tasks, such as Durán Muñoz's (2010) study on translation professionals. She concludes that translators prefer to solve their doubts – in this case, terminological doubts – by consulting terminological sources online, because they generally find these quick and easy to access. Another example may be found in the work by Cid-Leal and Perpinyà-Morera (2015) on the use of translation students' electronic documentation tools. They confirm the findings of Durán Muñoz (2010) and – quite notably – they observe that training in understanding these types of tools, in addition to making efficient searches with them, should be included in the training of translators and interpreters.

When dealing with specialised translation, the student must turn to documentation sources to glean information about the source text, solve terminological problems that may arise due to the use of specialised languages, and gain phraseological information. Of the different documentation resources, and on account of the specialised subject matter inherent in medical translation, corpora play an essential role in acquiring precise terminological and phraseological knowledge in specialised contexts (Rodríguez Martínez & Ortega Arjonilla, 2018).

For all the above reasons, the postgraduate program Master in Intercultural Communication, Public Service Interpreting and Translation (University of Alcalá, Madrid, Spain) offers a compulsory course called *Medical and Healthcare Resources*, which includes three main modules to develop students' linguistic and documentation skills: a) corpus-based approaches, b) technology-mediated learning, and c) terminology. The next section will focus on the two 4-hour sessions devoted to corpus-based approaches. The main objectives of this module are: a) to search for, select, and evaluate documentation resources for the scientific field; b) to gain a command of

resources for making linguistic and terminological inquiries on the Internet; and c) to gain a command of query tools (corpora and concordance programs) as documentation resources. The decision to include CMTs has been made for several reasons. First, students are not usually familiar with these tools, or if they are, they are unaware of their practical applications. Furthermore, besides solving problems of a terminological, phraseological, or conceptual nature, these tools develop more technical skills, since students will work with new formats and specific programs. And finally, they are tools that can help them both professionally and in their research.

Method

The present article describes the research carried out as part of the required postgraduate module *Medical and Healthcare Resources* (Master in Intercultural Communication and Public Service Interpreting and Translation, University of Alcalá, Madrid, Spain). A group of 40 students took part in this study, compiling their own specialised corpus and analysing it with LancsBox1 (Brezina, Timperley, & McEnery, 2018), a corpus management tool (CMT) used as a documentation resource throughout the semester.

Our study follows a non-experimental mixed-methods approach (Creswell, 2013), designed to answer the following research questions using quantitative and qualitative approaches.

- RQ1: What is the perception of a group of postgraduate students towards the use of specialised corpora?
- RQ2: Are corpus management tools appropriate documentation resources for these students' professional development?

To answer RQ1 and RQ2, we administered two questionnaires (a quantitative approach), which were completed by students in class at the end of the module and at the end of the semester, respectively. These questionnaires aimed to get insight about the usefulness of corpora after students completed the module on corpus management tools. A qualitative approach based on focus groups was also utilized at the end of the module.

Participants, Materials, and Procedure

A total of 40 students (13 males and 27 females) participated in the study. They all signed up for the compulsory course *Medical and Healthcare Resources* during the first half of the 2018–2019 academic year. All of them were native Spanish speakers and had completed a degree in Linguistics or Translation Studies, except for two postgraduates in Sciences and three postgraduates in Law.

During the whole semester, the course covered the following topics:

- Weeks 1, 2, 3, and 4: Main issues in documentation resources for medical translation and interpreting (an online module that lasted one month)
- Week 5: Corpus-based approaches to medical translation and interpreting (4 hours)
- Week 6: Corpus-based approaches to medical translation and interpreting (4 hours)
- Week 7: Computer-assisted translation tools (4 hours)

- Week 8: Computer-assisted translation tools (2 hours) + Terminological tools for medical translation (2 hours)
- Week 9: Terminological tools for medical translation (4 hours)

Our main concern was to offer a theoretically and pedagogically sound module on CMTs. This specific module took place over two weeks, in one four-hour session per week. In the first session, they learnt about the theoretical concepts of corpus-based approaches, such as corpus typology following the classification provided by Laviosa (2002) and Zanettin (2012). We also paid special attention to virtual corpora, as we have seen in the Literature Review section that they have been proved to be useful for acquiring specialised terminology and enhancing students' learning responsibility (Charles, 2012). This first session also covered the main electronic documentation resources for specialised domains (specialised dictionaries, terminology glossaries, visual dictionaries, terminology databases, etc.) and the differences between the use of the web for compiling a corpus (Web for Corpus, WfC) and the use of the web as a corpus (Web as Corpus, WaC). Due to the different participants' backgrounds, this first approach was explained in the introductory session. The participants also learnt to use CMTs such as LancsBox, AntConc,[2] and Sketch Engine[3], becoming familiar with their main functions (for word frequency generation, the use of concordances, and the retrieval of collocations and lexical packages). In the second session, they were given the task of compiling a bilingual comparable corpus (English-Spanish) formed of specialised texts about the different topics they were going to work on for the rest of the course (comprised of informed consents, pharmaceutical leaflets or specialised articles published in journals). The compilation of a virtual corpus can be done either manually or semi-automatically; the latter method is quicker, since programs such as BootCaT and WebBootCaT make the task easier (Costa, Durán Muñoz & Corpas Pastor, 2016; Seghiri, 2017).

However, as some of the participants were not familiar with CMTs, we decided to propose an activity in which they had to manually compile their own corpus. The class was divided into five groups of eight, and the activity consisted of compiling and analysing a comparable virtual specialised corpus. As previously mentioned, the participants had to bear in mind that the corpus compiled would serve as a documentation resource for familiarising themselves with some of the subject areas with which they would work in subsequent classes, most of them implying translation tasks. In short, the activity involved building a comparable virtual specialised corpus as a documentation resource, so that participants might become aware of (1) the importance of documentation requirements in the field of medical translation and (2) the creation of quality material that could be consulted quickly and effectively, in order to acquire conceptual, terminological, and phraseological knowledge in subsequent translation tasks.

The corpus design process starts with the documentation stage. Before compilation, students need to select texts for their corpus. At this stage, we encouraged our students to read texts on the different topics on which they would focus during the course (e.g. academic articles on myocardial infarction). Once they had read extensively about the topic, students had to locate different Internet-based texts to be included in their own virtual corpus. To accomplish this, they searched for information using keywords (e.g. 'heart attack', 'coronary artery disease', 'coronary angiogram', etc.). It is of paramount importance at this stage to use very precise keywords – *seed words* – as filters, in order to exclude irrelevant information or 'noise'. For instance, the operator *site:* restricts the search to a specific website or domain. The search string *myocardial infarction site:www.imedpub.com* (without a space after the colon) would only yield results from the website *www.imedpub.com*, which provides access to a very comprehensive set of medical texts, whereas the search string *myocardial infarction:.co.uk* would restrict the search to all websites under the domain .co.uk. Participants were also encouraged to use free software (i.e. HTTrack, GNU Wget, or Jdownloader) so that they could automate the download process. Once the documents had been found and downloaded, the texts had to be converted to .txt files to be processed by corpus

analysis software like LancsBox. This task is especially necessary in the case of texts retrieved in .html and .pdf format. For this purpose, our students used HTMLasTExt (http://www.nirsoft.net/utils/htmlastext.html) and pdf2txt (www.pdf2txt.com). Finally, all documents were stored and named.

Participants were then able to initiate an analysis of their materials. In this second stage, the participants begin to use their corpus as a documentation resource that would enable them to acquire linguistic knowledge, such as terminological and phraseological knowledge about a specific subject. The participants used their virtual corpus as a terminological tool, and they generated a list of terms ordered alphabetically and by frequency of occurrence. Secondly, participants created a list of Ngrams, i.e. sequences of words (1-Ngram, one-word phrases; 2-Ngram two-word phrases; 3-Ngram, three-word phrases, etc.). This is a very useful activity for the creation of glossaries that would help participants in the preliminary stage of text comprehension (Figure 1).

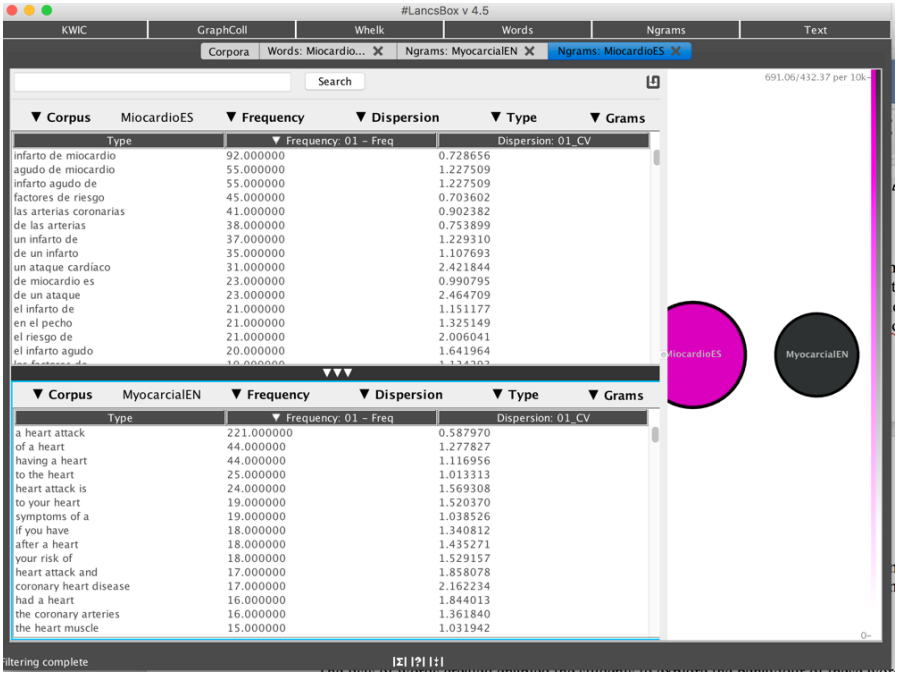

Figure 1. LancsBox: a 3Ngram

Other functions that the participants used were collocates for the creation of collocation lists and collocational patterns, in addition to lexical clusters, through which the corpus created could be exploited in more detail. LancsBox also offers the opportunity to consult grammatical information and look for specific grammatical patterns by creating a lemmatised bigram list. For this purpose, the participants used different filters, which refined the searches. Having opened the Wordlist section, the participants entered different search patterns to find the specific grammatical pattern **_n *_n* or **_adj *_n *_n*. As a result, patterns such as 'heart attack' or 'coronary heart attack' were retrieved.

Another function that the participants used was the collocation network function to visualise a list of collocations and collocational patterns. LancsBox allows users to select the range of words to be considered in the collocational pattern and the statistics of the collocational pattern. For instance, a range of -3 to 3 will show the three words to the left of the word and the three words to the right as the collocational range, in addition to a minimum frequency of appearance of three occurrences in the corpus (Figure 2).

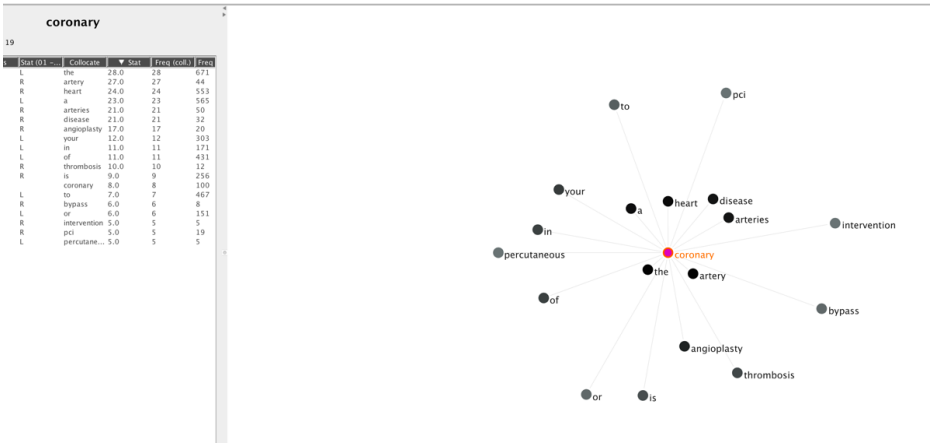

Figure 2. Collocation network for the word coronary

By the end of the second stage of analysis, the participants had compiled a specialised comparable virtual corpus as a documentation resource. Following this compilation, they analysed the corpus to gain a better understanding of LancsBox software and the specialised lexis related to the subsequent topics that the course would cover. To achieve this aim, the participants analysed their corpus as a source of: (a) terminological information (the extraction of terminological units); (b) grammatical information (the use of filters and specific search patterns); and (c) collocational and grammatical information (use of the collocations network function).

Quantitative Research

As the research design was non-experimental, we used descriptive statistics to report the quantitative results. In the analysis and interpretation of our results, the aim was to understand participants' perceptions of CMTs.

Bearing in mind that a questionnaire is an imperfect but practical measurement method whose suitability is enhanced when complemented by other sources of information, we also included qualitative instruments in our research to triangulate data. The questionnaire used a 5-point Likert-type scale reflecting agreement or disagreement. As recommended by Edley & Litoselitti, (2018), all the variables were coded analogically, that is, each variable was assigned a numerical value as follows: (1) Strongly disagree, (2) Disagree, (3) Neutral, (4) Agree, (5) Strongly agree. The first questionnaire included a total of 10 items covering issues related to the CMT used (LancsBox) and issues related to the participants' experiences. The questionnaire is reproduced in Table 1 along with its results, included in the following section.

The second questionnaire was administered at the end of the semester to elicit information about the usefulness of corpus management tools during the other course subjects. The final questionnaire comprised the following five open-ended questions:

1. Have you used your own corpus and CMT during the course? Yes/No

2. What has been the main purpose of using corpus and corpus-management tools during the course?

a) Translation

b) Revising

c) Creating other documentation resources (i.e. glossaries)

d) Others

3. Do you consider corpus and corpus management tools be worth incorporating into your everyday working tools? Yes/No

4. Do you consider corpus and corpus management tools module to be compulsory for the postgraduate course? Yes/No

5. I consider my documentation skills have been improved after working with corpus and CMT and using in real language projects. Yes/No.

Both questionnaires were piloted among colleagues and students to produce consistent, valid and reliable data collection instruments.

Qualitative Research

Qualitative methods offer various instruments that can be adapted to multiple types of research, such as interviews, focus groups, and direct observation by the researcher. Specifically, we used semi-structured focus group interviews, which were recorded and transcribed for content analysis (Ravitch & Mittenfelner, 2015). This qualitative tool is considered an efficient method for obtaining data from multiple participants (Edley & Litoselitti, 2018; Rahmati, Sadeghi & Ghaderi, 2019; Yazdanmehr & Akbari, 2015) and can also boost spontaneous responses and interaction among the participants. To guide the interviews, we developed the following questions:

- What are the main difficulties you found when using CMTs during your postgraduate course?
- Do you consider CMTs useful tools for the development of your professional career?
- Do you consider CMTs useful in specialised domains such as medicine and healthcare?

To ensure the validity and reliability of the data, all the interviews followed the proposed structure, took place in similar conditions, and had the same duration (30-35 minutes). Eight interviews were conducted – one with each group – after completion of the activity.

Results

Quantitative Results

In the analysis and interpretation of our results, the aim was to understand participants' perceptions of CMTs as documentation resources in order to improve instruction in this field (RQ1).

As can be seen from the results (Table 1), the participants had no major difficulties using the platform (Item 1). LancsBox was found to have a simple, intuitive interface that enabled the corpus compilation to be easily uploaded (Item 2). Approximately 85% of the respondents agreed that LancsBox was an appropriate software to create bilingual glossaries (Item 3), as it offers the option of comparing the different corpora uploaded. This function is especially interesting as glossaries are useful when preparing translation tasks. LancsBox was also found to be helpful in dealing with terminology and grammatical patterns. A high number of respondents indicated their satisfaction with the Wordlist and GraphColl functions (85% for Item 4 and 90% for Item 5).

Results regarding student perceptions about benefits of the activity were highly positive. After using the CMTs, participants' confidence in looking for specific terms or patterns had improved (Item 6). The answers also reflected participants' positive attitude towards the fact that CMTs offer more specific linguistic information (Item 7). They also considered the module to be closely related to the postgraduate course's aims (Item 8), i.e. to provide specific training in specialised domains. Finally, the module increased their interest in CMTs (Item 9), probably because they created their own documentation material, and it is hoped that CMTs will indeed be used as part of their professional development (Item 10).

Table 1
Results of the Questionnaire Administered After the Corpus-based Module

Item	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
1. It was easy to use LancsBox.			13%	32%	55%
2. It was easy to compile my own corpus.		8%	13%	16%	63%
3. Building other documentation resources (e.g. glossaries) has been easy with LancsBox.			5%	10%	85%
4. I find LancsBox useful for exploring *specialised terminology.*				15%	85%
5.- I find LancsBox useful for exploring *collocations and grammatical patterns.*				10%	90%
6. I feel more confident about my final linguistic decisions after consulting my own corpus.				20%	80%
7. Searching in a CMT provides me more linguistic information.					100%
8. I consider the module to be *closely related to the postgraduate course aims.*					100%
9. The module has increased my interest in corpora *and CMT*s.				8%	92%
10. I believe I will use my corpus and CMT during the course.			10%	8%	82%

The final questionnaire also provided very encouraging results. All the respondents used their own corpus and CMTs during the course (Item 1). This positive finding can be related to the results of Items 6-10 on the first questionnaire, where the respondents indicated that they used their corpus and CMTs mainly for the purposes of translation (65%), revising translations (15%) and creating other documentary resources (21%). As with the results of the first questionnaire, participants strongly agreed that CMTs should be used as an additional documentation resource

when dealing with specialised domains (Item 3). In response to whether education on CMTs should be compulsory in courses on medical and healthcare resources (Item 4), all the respondents agreed. Finally, all respondents had no doubt that their documentation skills had improved (Item 5).

Qualitative results

Qualitative results derived from the semi-structured interviews provided insight into the advantages and disadvantages of using CMTs during the course, as well as the usefulness of CMTs (RQ2). To this end, participants were asked the following questions.

- What are the main difficulties you found when using CMTs during your postgraduate course?
- Do you consider CMTs useful tools for the development of your professional career?
- Do you consider CMTs to be useful in specialised domains such as medicine and healthcare?

In general, the participants had a positive attitude towards CMTs. Stated advantages included the opportunity to work with new software, which they were previously unfamiliar with. As the following interview excerpt shows, they were able to appreciate the vast number of possibilities that CMTs offer.

> *Student (S) 1: 'I had never worked with CMT. For instance, LancsBox is just amazing. There are so many things I have learnt during the course!'*

As expected, when considering the technical issues related to compiling a corpus, the participants also discussed the difficulties of working with specialised software, especially those participants who were not familiar with linguistic issues. They commented on the necessity of receiving instruction on corpus typology, different CMTs, and their main functions, in order to utilize the software effectively.

> *S2: 'I was a bit lost at the beginning of the module, but using LancsBox was not very difficult. Maybe I just needed more sessions to learn more functionalities'.*

The benefits of working with CMTs were also highlighted when participants discussed their professional development. Participants indicated that they were going to use this type of documentation resources, as dictionaries would not be enough. Using CMTs and their own corpora was also stressed as an advantage due to the subject specificity of the tasks carried out during the course.

The participants also highlighted disadvantages of CMTs; for instance, the time invested in compiling a corpus or the size of the corpus.

Participants generally expressed their satisfaction with the activity. Regarding knowledge acquisition, and corroborating answers to the questionnaire, the participants believed that the activity had helped them consolidate their documentation skills and raised their awareness of documentation tools. The different groups confirmed their belief that specific instruction needed to be provided. They also repeatedly mentioned that corpus-based approaches should be part of the course *Medical and Healthcare Resources*, since they are connected to real language use. All groups

stated explicitly that they had profited from the module on corpus-based approaches, as the content was new to them, and believed that they would be able to apply what they had learnt in the near future.

Discussion and Conclusion

In this article, our aim has been to show how corpus methodology (the compilation and analysis of a specialised comparable virtual corpus) can be used in an ESP classroom and its usefulness in the subsequent translation tasks. Results drawn from the quantitative and qualitative methodology have demonstrated that a virtual corpus is a useful documentation resource in specialized domains (e.g. medical translation) and can help prepare the student for specialised language tasks. Also, mainly due to the technical expertise required to perform effective searches, our study highlights that educators should take into consideration to include a corpus-based methodology in the specialised translation curriculum.

Findings indicated that the students were generally positive about using CMTs in medical translation prior and after the treatment. These findings encourage us to state that the corpus-assisted methodology was appropriate and can be interpreted as a success of the module. Our findings concur with Sánchez Ramos & Lázaro Gutiérrez (2019), who carried out a non-experimental research to analyze the success of implementing a corpus methodology in a postgraduate program to translate specialised texts (informed consent forms). Their results show that students were highly positive about using comparable corpus and CMSs during their translation tasks and were willing to incorporate them in their future translation tasks as documentation resources. Our research is also in line with Frankenberg-Garcia (2015); her study discusses how a group of thirteen postgraduate translation students perceived a learning module based on corpora and CMTs (i.e. SketchEngine). Mainly drawn from a questionnaire given at the end of students' period of instruction and students' reports, Frankenberg-García (2015) concluded that students considered corpus to be very helpful in terms of the concordance, word list and collocations information.

According to the students, the corpus-assisted methodology applied during the postgraduate course provided some short and long-terms benefits. First, both quantitative and qualitative results show that CTMs were useful in order to solve linguistic difficulties, which seems a general finding in most literature related to using corpora in translator training (Bermúdez Bausela, 2016; Vigier Moreno, 2016). Perhaps one of the most outstanding results is that some translation students have never heard about CMT (Frankengerg-García, 2015; Sánchez Ramos & Lázaro Gutiérrez, 2019). An important achievement of our corpus-assisted methodology was that students fully used corpora and CTS during the postgraduate course, as the second questionnaire's findings showed, which supports the integration of CMTs in specialised translation curriculum design.

The instructors may not forget the challenges of introducing a corpus methodology. As our qualitative results showed, some of the participants had never worked with any CMS or even heard about corpus methodology. This is even more challenging when students' backgrounds are different (e.g. postgraduate students from different fields such as Sciences, Law, and Translation Studies). However, the quantitative results obtained after introducing a corpus methodology into this postgraduate medical translation course show a positive student perception towards corpora as documentation resources and LancsBox, the CMT used during the course. The participants found LancsBox useful in helping them acquire existing knowledge to assist with future translation tasks. Qualitative and quantitative data demonstrate that participants acquired basic knowledge of corpora and that they feel confident about their linguistic choices. It is also interesting to note that they perceived these documentation resources as useful for the course and

considered them an essential part of the syllabus. The implementation of this type of experimental module also served as a point of reflection for the participants, since they had the opportunity to work with (and evaluate) another documentation tool to add to traditional tools (i.e. dictionaries, glossaries, and parallel texts). As other studies suggest, (Charles, 2012; Frankenberg-Garcia, 2015), although students' perception of corpora was generally very positive, our results confirm that new tools can pose some initial difficulties.

Although we are aware that our study is limited, and its results are applicable to a sample of a group of students, we believe our findings can be considered as an evidence of the success of a corpus-based approach in ESP. What this work reports may also be taken as a first step of an ongoing research to carry out an experimental study to explore the performance and different language acquisition before and after applying a corpus-based methodology.

References

Arce Romeral, L., & Seghiri, M. (2018). Booth-friendly term extraction methodology based on parallel corpora for training medical interpreters. *Current Trends in Translation Teaching and Learning, 5*, 1-46.

Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target, 7*(2), 223-243.

Bermúdez Bausela, M. (2016). Applying corpora-based translation studies to the classroom: language for specific purposes acquisition. In E. Martin-Monje, I. Elorza, & B. García Riaza (eds.), *Technology-enhanced language learning for specialized domains. Practical applications and mobility* (pp. 255-265). London: Routledge.

Bondi, M. (2001). Small corpora and language variation. In M. Ghdessy, A. Herny & R. Roseberry (Eds), *Small corpus studies and ELT* (pp. 135-175). Amsterdam: Benjamins.

Boulton, A. (2010). Data-driven learning: taking the computer out of the equation. *Language Learning, 60*(3), 534-572.

Brezina, V., Timperley, M., & McEnery, T. (2018). #LancsBox v. 4.x [computer software]. Retrieved from http://corpora.lancs.ac.uk/lancsbox.

Buendía-Castro, M. & López-Rodríguez, C. (2013). The web for corpus and the web as corpus in translator training. *New Voices in Translation Studies*, *10*, 54-71.

Cabré, M. T. (2004). ¿Lenguajes especializados o lenguajes para propósitos específicos?, *Foro Hispánico-Revista Hispánica de Flandes y Holanda*, *26*(1), 19-34.

Calonge Prieto, M. (2009). La complejidad del lenguaje de los textos médicos y la terminología especializada. In M. J. Varela Salinas (Ed.), *Panorama actual del estudio y la enseñanza de discursos especializados* (pp. 97-121). Bern: Peter Lang.

Calzada Pérez, M. (2005). Corpus electrónicos como herramientas de documentación y formación de traductores. In D. Sales Salvador (Ed.), *La biblioteca de Babel. Documentarse para traducir* (pp. 163-199). Granada: Comares.

Charles, M. (2012). Proper vocabulary and juicy collocations: EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes*, *31*, 93-102.

Cid-Leal, P., & Perpinyá-Morera, R. (2015). Competencia informacional en traducción: análisis de los hábitos de los estudiantes. *Textos universitarios de biblioteconomía y documentación*, *34*. Retrieved from http://bid.ub.edu/pdf/34/es/cid.pdf

Costa, H., Durán Múñoz, I., & Corpas Pastor, G. (2016). Compiling specialised comparable corpora. Should we always trust (semi) automatic compilation tools? *Linguamática*, *8*(1), 3-19.

Creswell, J. (2013). *Research design: qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.

Durán Muñoz, I. (2010). Specialized lexicographical resources: A survey of translators' needs. In S.Granger & M. Paquot (Eds.), *ELexicography in the 21st century: New challenges, new applications* (pp. 55-66). Lovaina-La-Nueva: Presses Universitaires de Louvain.

Edley, N., & Litoselitti, L. (2018). Critical perspectives on using interviews and focus groups. In L. Litoselitti (ed.), *Research methods in linguistics* (pp. 195–225). London: Bloomsbury.

EMT group (2017). Competence framework. Retrieved from https://ec.europa.eu/info/sites/info/files/emt_competence_fwk_2017_en_web.pdf

Faya, G. (2016). *Medical brochure as a textual genre*. Cambridge: Cambridge Scholars Publishing.

Frankenberg-Garcia, A. (2015). Training translators to use corpora hands-on: challenges and reactions by a group of thirteen students at a UK university. *Corpora*, *10*(3), 351-380.

Gaviolli, L. (2005). *Exploring corpora for ESP Learning*. Amsterdam: Benjamins.

Graves, M. F. (2006). *The vocabulary book: Learning and instruction*. New York: Teachers College Press.

Kübler, S., & Zinsmeister, H. (2015). *Corpus linguistics and linguistically annotated corpora*. London: Bloomsbury.

LAVIOSA, S. (2002). *Corpus-based translation studies. Theory, findings, applications*. Amsterdam/New York: Rodopi.

McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-based language studies*. London: Routledge.

McEnery, T., & Wilson, A. (2004). *Corpus linguistics*. Edinburgh: Edinburgh University Press.

Perea-Barberá, M. D., & Bocanegra-Valle, A. (2014). Promoting specialised vocabulary learning through computer-assisted instruction. In E. Barcena, T. Read, & J. Arús (eds.), *Languages for specific purposes in the digital area* (pp. 129-154). Switzerland: Springer International Publishing.

Pinto Molina, M. (2005). Alfabetización en información para traductores: propuesta del modelo ALFINTRA. In D. Sales Salvador (ed.), *La biblioteca de Babel* (pp. 19-32). Granada: Comares.

Rahmati, T., Sadeghi, K., & Ghaderi, F. (2019). English as a foreign language teacher inmmunity: an integrated reflective practice. *Iranian Journal of Language Teaching Research*, 7(3), 91-107.

Ravitch, S., & Mittenfelner, N. (2015). *Qualitative research. Bridging the conceptual, theoretical, and methodological*. Thousand Oaks: Sage.

Rea Rizzo, C. (2009). A first approach to the lexical profile of telecommunication English: frequency, distribution, restriction and keyness. *Revista de Lingüística y Lenguas Aplicadas*, *4*, 161-173.

Rodríguez Martínez, M. C., & Ortega Arjonilla, E. (2018). El corpus de prospectos farmacéuticos como recurso didáctico en el aula de traducción especializada francés-español. *MonTi, Monografía de Traducción e Interpretación*, *10*, 117-140.

Sales Salvador, D. (2006). *Documentación aplicada a la traducción: presente y futuro de una disciplina*. Gijón: Ediciones Trea.

Sánchez Ramos, M. M. (2020). Documentación digital y léxico en la traducción e interpretación en los servicios públicos (TISP): fundamentos teóricos y prácticos. Bern: Peter Lang.

Sánchez Ramos, M. M., & Lázaro Gutiérrez, R. (2019). Traducción de consentimientos informados y lingüística de corpus: una propuesta metodológica para el aprendizaje de la traducción de textos médico-jurídicos. In M. Seghiri (ed.), *La lingüística de corpus aplicada al desarrollo de la competencia tecnológica en los estudios de traducción e interpretación y la enseñanza de segundas lenguas* (pp. 49-67). Bern: Peter Lang.

Seghiri, M. (2011). Metodología protocolizada de compilación de un corpus de seguros de viajes: aspectos de diseño y representatividad. *RLA. Revista de Lingüística Teórica y Aplicada*, *49*, 2, 13-30.

Seghiri, M. (2017). Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores. *Babel*, *63*(1), 40-62.

Toledo Báez, C. (2019): Sketch Engine en traducción científico-técnica (francés-español): creación y explotación del corpus ad hoc comparable GeneCorp. In M. Seghiri (ed.), *La lingüística de corpus aplicada al desarrollo de la competencia tecnológica*. Bern: Peter Lang.

Varantola, K. (2003). Translators and disposable corpora. In F. Zanettin, S. Bernardini, & D. Stewart (Eds.), *Corpora in translator education* (pp. 55-70). Manchester: St. Jerome.

Vigier Moreno, F. (2016). Teaching the use of ad hoc corpora in the translation of legal texts into the second language. *Language and Law*, *3*(1), 100-113.

Yazdanmehr, E., & Akbari, R. (2015). An expert EFL teacher's class management. *Iranian Journal of Language Teaching Research, 3*(2), 1-13.

Zanettin, F. (2012). *Translation-driven corpora*. Manchester. St. Jerome.

Zhu, Ch. & Wang, H. (2011). A corpus-based, machine-aided mode of translator training. In *The Interpreter and the Translator Trainer, 5*(2), 269-291.

María del Mar Sánchez Ramos is an associate professor in translation in the Department of Modern Philology at the University of Alcalá (Madrid, Spain). Her research has focused on specialised translation, translation technology and language acquisition. She is an active member of different national and international research projects and has published extensively in indexed journals.

[1] LancsBox is a corpus management software developed at Lancaster University (United Kingdom). It can be downloaded from http://corpora.lancs.ac.uk/lancsbox/.
[2] AntConc can be downloaded at https://www.laurenceanthony.net/software.html.
[3] Sketch Engine can be accessed at https://www.sketchengine.eu.