Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.

# COMMUNITY HEALTHCARE TRANSLATOR TRAINING AND *AD HOC* CORPORA

MARÍA DEL MAR SÁNCHEZ RAMOS

University of Alcalá

FITISPos Research Group

## Abstract

This article deals with the exploitation of a monolingual specialised *ad hoc* corpus as a documentation resource in the community healthcare translation classroom. We start with a general overview of using a corpus-based methodology in the Community Translation context, an emerging branch within Translation Studies. We follow with a description of the main steps involved in the compilation and exploitation of the corpus in a postgraduate Community Translation environment, focusing on corpus linguistics as a documentation tool for the translation of healthcare texts, and the use of a corpus-based methodology.

Keywords: specialised corpus, community translation, postgraduate translator training, corpus compilation

Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.

# 1. INTRODUCTION

In the last twenty years, there has been a growing demand for cultural mediation, due to the ever-increasing mobility of people. As stated by Valero-Garcés (2006), Spain has had an influx of immigrants (and tourists), posing challenges which require adequate responses to ensure balanced co-existence. As a consequence, there is a need for translators and interpreters in different public services environments (hospitals, police stations, administration offices, and other settings), and successful communication is a must in these contexts (Creeze, Mikkelson and Monzon-Storey, 2015). This has created a need for professionals who can deal with linguistic barriers. Translation Studies has seen the emergence of a new academic branch called Community Translation (Taibi and Ozolins, 2016), also known as Public Service Interpreting and Translation (PSIT).

Nowadays, many translator training programmes incorporate specific modules focused on PSIT as undergraduate or postgraduate courses. The University of Alcalá is an example. Its Master`s

Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.

Degree in Intercultural Communication, Public Service Interpreting and Translation presents a wide variety of language pairs, namely Spanish and English, French, Arabic, Chinese, Russian or Romanian, and comprises specific modules on intercultural communication, healthcare translation, healthcare interpreting and legal and administrative translation and interpreting. There are also specific modules focused on the development of documentation and technical skills, such as *Resources for PSIT*. The main goal of this module is to offer PSIT trainees adequate tools and resources to be able to accomplish the interpreting and translation tasks they will carry out in the other modules (legal and administrative translation and interpreting, healthcare translation, healthcare interpreting).

Our paper highlights how monolingual virtual corpora and concordance software can be used as tools for translator training within a PSIT syllabus to develop documentation skills and the ability to use corpora to engender a better understanding of specialised text types as well as phraseological and terminological information.

## 2. *AD HOC* CORPORA AND TRANSLATION

Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.

The use of corpora and corpus analysis tools has attracted the attention of different scholars in Translation Studies. Depending on the nature of the different tasks carried out, researchers have used corpora to investigate different translation issues, from the features of translated texts (Baker, 1995, 1996; Fantinuoli and Zanettin, 2015; Kenny, 2001; Saldanha, 2004) to the possibilities of using corpora as translation and terminological resources (Bowker, 2003; Zanettin et al., 2003; Zhu and Wang, 2011). Other academics have highlighted the numerous advantages that using corpora in translator training can offer (Fantinuoli, 2013; Krüger, 2012; Kübler, 2011; Zanettin, 2012). A corpus-based methodology has proved useful in specialised translation learning environments, mainly due to the valuable terminological and phraseological information it offers (Corpas Pastor and Seghiri, 2009; Sánchez Gijón, 2009; Vila Barbosa, 2013).

In terms of text corpora typology, there is a general distinction between three major corpus types (Zanettin et al., 2003): monolingual corpora, comparable corpora and parallel corpora. Monolingual corpora usually contain texts in the target language. They can be efficient linguistic instruments offering relevant information on idiomatic use, collocations, syntactic constructions or genre conventions. Comparable corpora consist

of original source and target language texts with similar content. They can be divided into different subtypes based on the intended purpose or function of texts. Parallel corpora contain source language texts and their translations.

Nevertheless, other types of corpora are becoming more popular in the specialised translation area. Compiling small corpora – monolingual, comparable or parallel – as documentation resources for a specific translation task or for translator training purposes is gaining popularity among translation scholars. Such corpora are known as *ad hoc* corpora or virtual corpora. A virtual corpus is a collection of texts developed from electronic resources by the translator and compiled "for the sole purpose of providing information – either factual, linguistic or field-specific – for use in completing a translation task" (Sánchez Gijón, 2009, p. 115). The final goal is to gather as much information as possible in a very short period of time. The main concern when compiling an *ad hoc* corpus is not size; rather, it is the appropriate selection and evaluation of text, thematic and textual typology, compared with the source text. Zanettin (2012, p. 64) also highlights the advantages of virtual corpora, pointing out that virtual corpora are created in response to specific translation problems. Due to this information need, these corpora are

usually very precise and content-specific, and can be extended at any time.

In general terms, the benefits and the pedagogical implications of using corpora within translation studies have also been shown in various examples of research (Bowker and Pearson, 2002, p. 10; Corpas Pastor & Seghiri, 2009, p. 102; Lee & Swales, 2006, p. 74). Some of the main advantages identified are related to the development of instrumental sub-competence (PACTE, 2003, p. 53) or so-called information mining competence (EMT Expert Group, 2009). The need to know and use different electronic corpora and concordancing tools is also illustrated by Varantola (2003) and Rodríguez Inés (2010, p. 253), who identifies a further sub-competence within the instrumental sub-competence of the PACTE model, namely "the ability to meet a number of learning outcomes: identifying the principles that lie at the basis of the use of corpora; creating corpora; using corpus-related software; and solving translation problems by using corpora" (Rodríguez Inés, 2010, p. 253). Using corpora as translation tools, translators can verify lexical, phraseological and textual patterns beyond their intuition or previous knowledge and thus make more informed, justified translation choices (Neunzig, 2003; Rodríguez Inés, 2008). The use of corpora in training students to translate into

their L2 also boosts trainees' self-confidence and autonomous learning (Ulrych, 2000). Unsurprisingly, using corpora, especially virtual corpora, is part of the daily practice of professional translation, due to the fact that they are typically collected to accomplish a single translation assignment or to satisfy a transitory need (Varantola 2003, p. 55).

## 3. COMPILING AN *AD HOC* CORPUS FOR THE COMMUNITY TRANSLATION CLASSROOM: SOME PEDAGOGICAL IMPLICATIONS

As previously mentioned, the development of instrumental competence, including the use of documentation sources and electronic tools, is particularly relevant in PSIT, where translators need to manage different information sources in order to acquire sufficient understanding of the subject of a text and thus enable the accurate transfer of information. Building an *ad hoc* corpus can satisfy specific needs. From a pedagogic perspective, using this type of corpus allows trainee translators to identify phraseology, specific use patterns, terminological variants, frequency of words and cohesive features, among other things. Hence, given the importance of documentation in PSIT training to

Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.

ensure production of a functionally adequate and acceptable target language text, we used a corpus-based methodology with a twofold objective: first, to make our students aware of the usefulness of corpus as a documentation resource in Community Translation; and second, to help our trainees develop their instrumental competence by using corpus management tools. Of the different major corpus types (Zanettin et al, 2003, p. 6), we found monolingual corpora especially useful for our documentation task, as the main goal was to retrieve information (conceptual, terminological and phraseological) to understand a specific domain in the healthcare context. We will describe in the following paragraphs how we implemented a corpus-based methodology and how our postgraduate students compiled and analysed an *ad hoc* corpus focused on overcoming the challenges and difficulties they face when translating healthcare texts.

## 3.1. Training to use corpus in translation

Previous to the documentation stage and corpus analysis, attention should be focused on the design criteria. Our students were asked to design a monolingual *ad hoc* corpus of original specialised texts on sickle-cell anaemia written in the target

language (British English). Our students attended two training sessions of eight hours in total. In the first session, they were introduced to the main theoretical concepts in Corpus-based Translation Studies (CTS), the main documentation resources for PSIT (lexicographical databases, specialised lexicographical resources and specialised websites) and different word search strategies needed to take advantage of search engines and Boolean operators. This introductory session was relevant, as searching for keywords would allow students to move into more specific areas or domains (Austhermühl, 2001, p. 54). Although our time was reduced, we considered it relevant to provide training translators with a practical approach to CTS, through a hands-on session on the Web for Corpus (WfC) and Web as Corpus (WaC) approaches, which focus on the potential of the Web as a corpus in itself (De Schryver, 2002, p. 272). The students were also introduced to specific computer software and shown how to use retrieval information software such as AntConc. They also learned the basic functions of both software programs (generating and sorting concordances, identifying language patterns, retrieving collocations and collocations clusters, and other features). In the second session, we moved into the practical stage: the corpus design. We summarise the corpus design process in three words: documentation, compilation and analysis.

Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.

## 3.2. Corpus design stages: documentation, compilation and analysis

The corpus design process starts with documentation. Before compilation, students need to select texts for their corpus. At this stage, we encouraged our students to read texts similar to the source text. Once they had read extensively on the topic, students had to locate different Internet-based texts to be included in their own virtual corpus. To do so, they put into practice what they had learned about Boolean operators in previous sessions, that is, searching for information using keywords (i.e., "sickle-cell anaemia", "haemoglobin", "erythrocytes", "Hb SS disease"). It was 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, using "site:" restricts the search to a specific website or domain. The search string *sickle cell site:www.nhlbi.nih.gov* only yields results from the website www.nhlbi.nih.gov, which provides access to a very comprehensive set of medical texts, whereas the search string *sickle cell site:.co.uk* restricts the search to websites under the .co.uk domain. Students were also encouraged to use free software (i.e., HTTrack, GNU Wget, Jdownloader) so that they could automate the downloading

process. Once the documents had been found and downloaded, the texts had to be converted to .txt files in order to be processed by corpus analysis software like AntConc. This task is especially necessary in the case of texts retrieved in .html and .pdf format. Students used HTMLasTExt (www.nirsoft.net/utils/htmlastext.html) and pdf2txt (www.pdf2txt.com) for this purpose. Finally, all documents were stored and named. Students were able to initiate an analysis of their materials.

The main aim of compiling this corpus was to use it as a documentation resource – that is, to make students aware of corpus as a source of conceptual, terminological and phraseological information. In terms of terminology, students generated a wordlist of monolexical terms ordered by alphabet and frequency of appearance. They also created what is known as a stop list or negative list in order to filter words, such as articles, adverbs and prepositions, so that irrelevant words or 'noise' could be avoided. AntConc offers the possibility of creating a stop list either manually, adding the words to be excluded from the corpus, or automatically. Towards this end, pre-built stop lists that can be found on the Internet can be used.

Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.

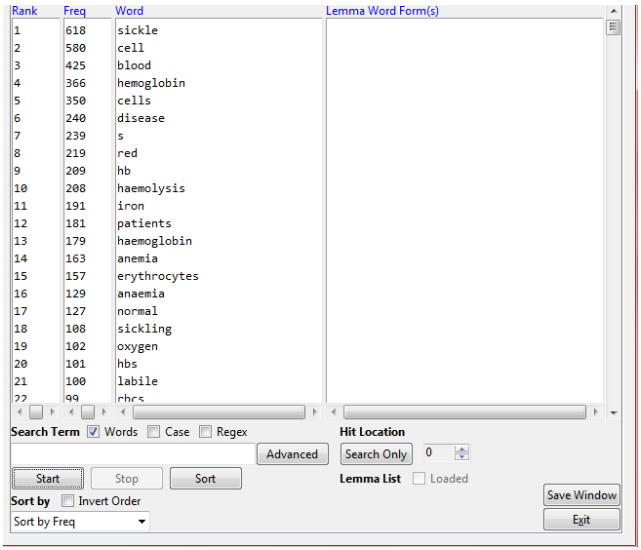


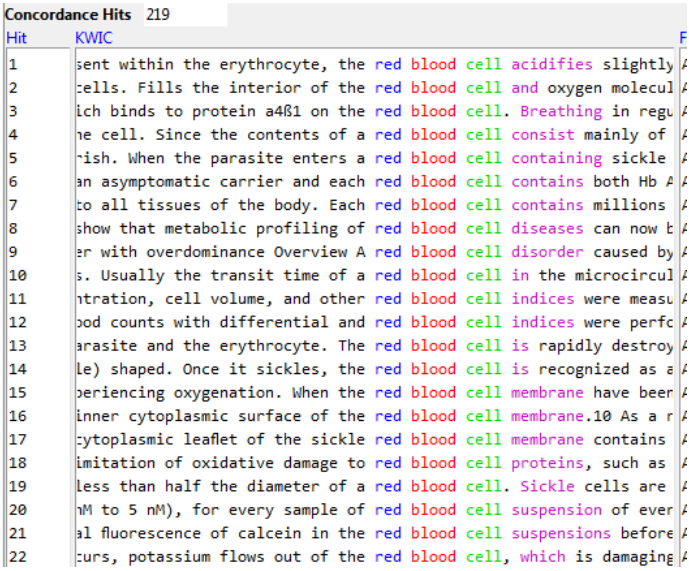

Figure 1. Frequency word list generated by AntConc

The analysis of monolexical units revealed the use of general words acquiring a specialised meaning in the corpus (“cell”, “patient”, “anaemia”, “disease”, “red”, “blood”, “treatment” “results”) and specialised vocabulary related to the disease of sickle-cell anaemia (“oxygen”, “haemoglobin”, “haemolysis”, “erythrocytes”, “bilirubin”). The wordlist was also useful for working with acronyms (DNA, “deoxyribonucleic acid”; SCD, “sickle-cell anaemia”; RCB, “red blood cells”) and

abbreviations (ml, "millilitre"; hb, "haemoglobin"; db, "decibel").

The concordance function was also a very attractive resource for our students. A simple query generated concordance lines listed in keyword in context (KWIC) format, where the software displays all occurrences of the searched word (or pattern) in the corpus. The most outstanding feature is that the concordance function presents the search pattern surrounded by its immediate context. For instance, students looked up the word "red", one of the most frequent words. They could check that some of the general words (e.g., "red") were the base for more specific terms ("red blood cell"). Very interesting for them was the "sort" function, as the user can highlight words to the left or right of the keyword in order to check syntactical patterns.

Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.



**Concordance Hits** 219

| Hit | KWIC | F |
|---|---|---|
| 1 | sent within the erythrocyte, the red blood cell acidifies slightly | A |
| 2 | :ells. Fills the interior of the red blood cell and oxygen molecul | A |
| 3 | ich binds to protein a4ß1 on the red blood cell. Breathing in regu | A |
| 4 | ne cell. Since the contents of a red blood cell consist mainly of | A |
| 5 | ish. When the parasite enters a red blood cell containing sickle | A |
| 6 | an asymptomatic carrier and each red blood cell contains both Hb A | A |
| 7 | to all tissues of the body. Each red blood cell contains millions | A |
| 8 | show that metabolic profiling of red blood cell diseases can now b | A |
| 9 | er with overdominance Overview A red blood cell disorder caused by | A |
| 10 | s. Usually the transit time of a red blood cell in the microcircul | A |
| 11 | ntration, cell volume, and other red blood cell indices were measu | A |
| 12 | ood counts with differential and red blood cell indices were perfo | A |
| 13 | arasite and the erythrocyte. The red blood cell is rapidly destroy | A |
| 14 | le) shaped. Once it sickles, the red blood cell is recognized as a | A |
| 15 | periencing oxygenation. When the red blood cell membrane have beer | A |
| 16 | inner cytoplasmic surface of the red blood cell membrane.10 As a r | A |
| 17 | :ytoplasmic leaflet of the sickle red blood cell membrane contains | A |
| 18 | imitation of oxidative damage to red blood cell proteins, such as | A |
| 19 | less than half the diameter of a red blood cell. Sickle cells are | A |
| 20 | nM to 5 nM), for every sample of red blood cell suspension of ever | A |
| 21 | al fluorescence of calcein in the red blood cell suspensions before | A |
| 22 | :urs, potassium flows out of the red blood cell, which is damaging | A |

Figure 2. Concordance function

In terms of phraseology, looking up collocations and groups of words (collocation clusters) was a valuable way to identify the frequency of appearance of more specific lexical patterns. Collocations are sequences of words that occur together more often that we would expect. AntConc offers the possibility of sorting the context words to the right and/or left of the keyword. It also gives the option of extracting all n-grams – sequences of words of a certain length – from the text. The cluster/Ngram function was particularly useful for checking the collocational patterns of some words, (e.g., “AA erythrocytes”, “haemoglobin

concentrations", "cell dehydration"), where students positively evaluated the contextual information their virtual corpora offered.

AntConc also offers the possibility of using wildcards to refine a search and look for definitions and synonyms. For instance, in order to find this information, students used the following search patterns: "sickle* is", "sickle-cell anemia* or", "sickle-cell disease*", "sickle* is called", "sickle* is known as". As result, definitions of this term and synonyms were found. Some examples of synonyms were "meniscocytosis", "sicklemia", or "drepanocytosis".

Apart from the extraction of specific terms, AntConc offers direct access to the individual corpus files, allowing students to investigate in more detail a searched pattern in its context.

Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.

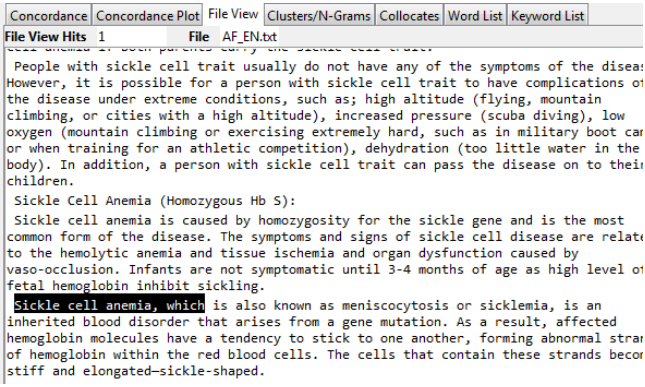



Figure 3. File view function

Student feedback was largely positive. They appreciated the usefulness of the different functions that a software tool such as AntConc can provide, such as frequency lists, collocates, clusters/N-grams and concordances. Compiling and using corpora made the students feel more confident in their technical skills and translation solutions. All in all, corpus use was evaluated by our students as a valuable tool for developing instrumental competence.

## 4. CONCLUSION

We have shown how we developed and exploited a specialised *ad hoc* corpus as a resource in the Community Translation training environment. As has been highlighted, there are multiple advantages

of using virtual corpora in this setting, particularly for the retrieval of very useful information about terminology, collocations, formulaic language and phraseological units. Trainee translators were able to grasp this information, providing a sound understanding of the linguistic specificities of the topic.

The use of *ad hoc* corpora as a documentation resource has been positively evaluated by our postgraduate students. They have been able to look for individual and specific texts for their corpus by using searching skills and to analyse material in a different way to that they are used to. After compiling and analysing their corpus, our students were aware that information provided by the *ad hoc* corpus could be extremely helpful when translating healthcare texts, from a linguistic and functional point of view. *Ad hoc* corpus and concordancing programs are additional computer-aided translation tools to be used along with translation memories or lexicographical resources.

We must not forget that training is essential, as compiling and analysing requires a variety of competences, not only linguistic and instrumental but also thematic. As we have seen, a well-planned corpus methodology can contribute to the development of these competences, by improving

knowledge of specialised terminology and phraseology and by developing instrumental and research skills. This scenario fits perfectly with the general demands of today's professional practice and therefore must be considered in specialised translator training, if we aim to equip our students with the skills required in an increasingly demanding and competitive translation market.

Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.

**References**

Austhermühl, F. (2001). *Electronic Tools for Translators*. Manchester: St. Jerome

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

Baker, M. (1996). Corpus-based translation-studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP, and translation: Studies in language engineering in honor of Juan C. Sager*, (pp. 175-186). Amsterdam/Philadelphia: John Benjamins.

Bowker, J., & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. London: Routledge.

Bowker, L. (2003). Corpus-based applications for translator training: Exploring possibilities. In S. Granger, J. Lerot & S. Petch-Tyson (Eds.), *Corpus-based Approaches to Contrastive*

Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.

*Linguistics and Translation Studies*, (pp. 169-183). Amsterdam/Philadelphia: Rodopi.

Crezee, I., Mikkelson, I. & Monzon-Storey, L. (2015). *Introduction to healthcare for Spanish-speaking interpreters and translators*. Amsterdam/Philadelphia: John Benjamins Publishing.

Corpas Pastor, G. & Seghiri, M. (2009). Virtual corpora as documentation Resources: Translating travel insurance documents (English–Spanish). In A. Beeby, P. Rodríguez Inés, & P. Sánchez-Gijón (Eds.), *Corpus use and translating. Corpus use for learning to translate and learning corpus use to translate* (pp. 75–107). Amsterdam: John Benjamins.

De Schryver, G. (2002). Web for/as corpus: A perspective for the African languages. *Nordic Journal of African Studies*, *11*, 2: 266-282.

EMT Expert Group. (2009). Competences for professional translators, experts in multilingual and multimedia communication. Retrieved March 28, 2014 from the European Master's in Translation Website of the DG Translation of the European Commission: http://ec.europa.eu/dgs/translation/programmes/emt/key_documents/emt_competences_translators_en.pdf.

Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.

Fantinuoli, C. & Zanettin, F. (Eds). (2015). *New directions in corpus-based translation studies*. Berlin: Language Science Press.

Kenny, D. (2001). *Lexis and Creativity in Translation. A Corpus-based Study*. Manchester: St. Jerome.

Kübler, N. (2011). New trends in corpora and language learning. In A. Frankenberg-García, L. Flowerdew, & G. Aston (Eds), *New trends in Corpora and Language Learning* (pp. 62-80). Continuum

Krüger, R. (2012). Working with corpora in the translation classroom. Languages in the globalised world: Interpreting multilingual literacy and cross-cultural encounters. *Special issue of Studies in Second Language Learning and Teaching*, *2, 4*, 505–525.

Lee, D., & Swales. J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, *25*, 56–75.

Neunzig, W. (2003). Tecnologías de la información y traducción especializada inversa. In D. Kelly et al (Eds.) *La direccionalidad en Traducción e Interpretación. Perspectivas teóricas, profesionales y didácticas* (pp. 189-205). Granada: Atrio.

Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.

PACTE Group. (2003). Building a translation competence model. In F. Alves (Ed.) *Triangulating translation: Perspectives in process oriented research* (pp. 43–66). Amsterdam: John Benjamins.

Rodríguez Inés, P. (2008). Translation into L2 in education and in real life, and how electronic corpora can help. Paper presented at the *Eighth Translation Conference: The Changing Face of Translation,* University of Portsmouth. Retrieved March 20, 2014, from http://www.academia.edu/1688485/_Translation_into_L2_in_education_and_in_real_life_and_how_electronic_corpora_can_help_

Rodríguez Inés, P. (2010). Electronic corpora and other information and communication technology tools. An integrated approach to translation teaching. *The Interpreter and Translator Trainer, 4, 2*, 251–282.

Saldanha, G. (2004). Accounting for the exception to the norm: A study of split infinitives in translated English. *Language Matters, Studies in the Languages of Africa, 35*, 1, 39-53.

Sánchez Gijón, P. (2009). Developing documentation skills to build do-it-yourself corpora in the specialised translation course. In A. Beeby, P. Rodríguez-Inés, & P. Sánchez-Gijón (Eds), *Corpus use and translating.*

Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.

*Corpus use for learning to translate and learning corpus use to translate* (pp. 109–127). Amsterdam: John Benjamins.

Taibi, M. & Ozolins, U. (2016). *Community Transaltion*. London: Boomsbury

Ulrych, M. (2000). Teaching Translation into L2 with the Aid of Multilingual Parallel Corpora: Issues and Trends. *Miscellanea, 4*, 58-80.

Valero-Garcés, C. (2006). *Formas de mediación intercultural: Traducción e Interpretación en los Servicios Públicos.* Granada: Comares.

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.

Vila Barbosa, M. (2013). Corpus especializados como recurso para la traducción: análisis de los marcadores de la cadena temática en artículos científicos sobre enfermedades neuromusculares en pediatría. *Onomazéin*, 27, 78-100.

Zanettin, F., Bernardini, S. & Dominic, S. (Eds.) (2003). *Corpora in Translator Education*. Manchester: St. Jerome.

Zanettin, F. (2012). *Translation-Driven Corpora*. Manchester: St. Jerome.

Zhu, Ch. & Wang, H. (2011). A corpus-based, machine-aided mode of translator training.

Sánchez Ramos, M. M. (2016). Community Healthcare Translator Training and Ad Hoc Corpora. *Current Trends in Translation Teaching and Learning E, 3*, 119–139.

*The Interpreter and the Translator Trainer*, 5, 2: 269-291.