Corpus-Based Interpreting Studies and Public Service Interpreting and Translation Training Programs: The Case of Interpreters Working in Gender Violence Contexts

Raquel Lázaro Gutiérrez and María del Mar Sánchez Ramos

Abstract The growing popularity of Public Service Interpreting and Translation (PSIT) in different fields, such as healthcare or legal environments, has highlighted the need for interlingual and intercultural communication between public service providers and users who do not have any or sufficient command of the official language of the public authorities. Training is essential in those settings if we want to successfully achieve the appropriate communication. Interpreting and translation training programs are especially useful in the cases of gender violence victims from other countries, with different pragmatic communication strategies. This article explores the use of Corpus-based Interpreting Studies (CIS) as a methodology to train interpreters in gender-based violence context. After a theoretical introduction on CIS, PSIT and interpreting in gender violence contexts a particular emphasis is placed on the design, compilation process and use of a monolingual corpus and concordance software.

Keywords Public Service Interpreting and Translation • Corpus-based Interpreting Studies • Gender-based violence • Interpreting training • Concordance software

1 Introduction

The necessity of communication links between public service providers and the users who do not have a command of the official language of public authorities has developed what is known as a new academic and professional discipline within Translation Studies, namely Public Service Interpreting and Translation (PSIT) or Community Interpreting and Translation, which covers a wide range of fields including, among others, healthcare, educational, legal and administrative settings.

R. Lázaro Gutiérrez (✉) • M.d.M. Sánchez Ramos
Departamento de Filología Moderna, Facultad de Filosofía y Letras,
University of Alcalá/Group FITISPos-UAH, Alcalá de Henares (Madrid), Spain
e-mail: raquel.lazaro@uah.es; mar.sanchezr@uah.es

Due to the growing demand for professional PSIT translators and interpreters specialised in particular fields, different research projects have emerged, such as InterMed,[1] funded by the Spanish Ministry of Economy and Competitiveness and focused on interpreting in medical contexts, Interpreting and Translation in Prison Settings,[2] funded by the University of Alcalá, or SOS-VICS (Speaking Out for Support, co-financed by the EU's Directorate General Justice and nine Spanish universities[3]), whose main goal is to train interpreters to work in gender violence contexts, thus facilitating the assistance to gender violence victims of foreign origin who may not fluently speak the language of the interaction, and, at the same time to contribute to raising awareness of the need to hire professional interpreters during linguistic mediation in such cases. Despite increasing research on PSIT, research based on interpreters working with gender violence victims is still scarce. In this specific case, interpreters need to understand protocols involved in gender violence, as well as key applicable legal terminology and procedures, and the definition of concepts related to gender violence. Apart from that, the pragmatic meaning of the language of the victims is usually hard to render, as people from distant cultures may have different communication styles and may use a great variety of mechanisms to convey a particular meaning.

The Spanish central and regional administrations are putting a lot of effort into combatting gender violence and dealing with victims, who are local. However, foreign victims receive little assistance. For them to be able to access the services provided, an interpreter is essential. If the interpreters are not qualified and specialised enough, the interpretation may not reach the desired quality standards resulting in inaccurate and inefficient communication between victims and service providers. Without qualified and specialised interpreters, the rights of the victims may be violated and their risk of exclusion may thus increase.

Driven by this increasingly demanding necessity of training interpreters involved in gender violence cases, we highlight how Corpus-based Translation Studies (CTS) can be the groundwork for our study to build a monolingual corpus (Spanish) (original oral discourses delivered in similar settings, written texts, such as protocols, leaflets and guides and written spontaneous discourse) to obtain valuable information about specific features of gender violence spoken language and derive pedagogical applications. To this end, we will describe how we have built, compiled and analysed this corpus and discuss the main obstacles related to spoken speech (i.e how to tackle with pragmatics, paralinguistic dimensions of language, copyright issues and ethical implications) that we have overcome. As part of our main tasks, our research investigates the genre of gender violence speech. We will centre our attention on the numerous advantages in the area of interpreting training that our *ad hoc* corpus provides in terms of discourse patterns, pragmatic conventions, lexical clusters and semantic prosody, among others. We will conclude our paper with

---

[1] Ref.: FFI2011-25500.

[2] Ref.: CCG2013/HUM-010.

[3] These are the current research projects in which the Group FITISPos-UAH is currently working. SOS-VICS (Ref.: JUST/2011/JPEN/2912) co-ordinator is the University of Vigo.

some specific examples on the exploitation of our corpus and draw conclusions concerning the advantages of CTS in PSIT training.

This study started in 2013 and is still under development. The authors of this piece of research belong to the Research Group FITISPos-UAH, based in Madrid, and specialised in training and research in public service translation and interpreting. The methodology of this study, if proven useful, will constitute suggestions for trainers in the field, and the final outcome, both scientific and practical, will be particularly useful to train interpreters of the Master's Degree in Public Service Translation and Interpreting at the University of Alcalá, and others wishing to or in need of specialising in gender violence contexts.

2   Corpus-Based Interpreting Studies

The initial steps of corpus linguistics can be traced back to the pre-Chomskian period (McEnery et al. 2006: 3), where followers of the structuralist tradition used a corpus-based methodology to generate 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 and Wilson 2004: 1) has opened many possibilities for the study of language. Similarly, corpora have attracted increasing attention in translation studies over the last years. Depending on the nature of the work carried out, researchers have used corpora to investigate the features of translated texts (Baker 1995, 1996; Kenny 2001; Saldanha 2004), or the possibilities of using corpora as translation and terminology resources (Bowker 2003; Zanettin et al. 2003; Zhu and Wang 2011). There is no doubt that the study of real language in its context can provide valuable information. Calzada Pérez (2007: 216) highlights that:

> […] por su flexibilidad y capacidad de adaptación, los CTS aúnan metodologías descriptivas y lingüísticas; análisis del proceso y producto; exégesis de detalles o amplios patrones de comportamiento de interés tanto por cuestiones formales como por facetas culturales, ideológicas e incluso literarias (Calzada Pérez 2007: 216)
>
>   […] 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. (Calzada Pérez 2007: 216, our translation)

Corpus-based translation studies (CTS) has proved to be a reliable way of collecting data to generalise about the so-called translated linguistic features or universals of translation (Baker 1993). The benefits and the pedagogical implications of using corpora within translation studies have been shown by various researchers. We can take Bowker and Pearson (2002), Corpas Pastor and Seghiri (2009), Lee and Swales (2006), or Sánchez Ramos and Vigier Moreno (forthcoming) as examples of authors that consider using corpus to research and teach specialised translation.

Nevertheless, CIS has not enjoyed the same popularity as CTS. Schlesinger's seminal paper (1998) set the groundwork for a CIS methodology. The use of corpora as a methodology within interpreting studies poses a number of challenges and

opportunities, and a number of difficulties. The main difficulty when dealing with a corpus-driven methodology in the study of interpreting is due to the obstacles involved in the analysis of oral discourse translation (i.e. transcribed speech): "The recording and transcription of unscripted speech events is highly labour intensive in comparison to the work involved in collecting quantities of written text for analysis (Thompson 2005: 254). Additional difficulty can be found in the compilation stage, as stated by Pöchhacker (2008), as it is difficult to obtain data and consent from speakers or service providers. Other types of difficulties to be taken into account have to do with the interpreter-mediated event, the different speakers and their roles, the interpreting mode, and the target audience (Shlesinger 1998).

Despite this challenging background, corpora may constitute the future of interpreting studies (Luzón et al. 2008). Although CIS are small in number if compared with CTS, there is a growing number of interpreting studies based on corpus data. We can take some works as examples. Ryu et al. (2003) focus on the use of corpora and simultaneous interpreting, as well as on the compilation of a bilingual corpus for linguistic and contrastive purposes. Other studies within the field of simultaneous interpreting explore the interpreter's speeches using an aligned simultaneous monologue interpreting corpus in order to research the interpreter's speaking speed and the difference between the beginning time of the speaker's utterance and that of the interpreter's utterance (Takagi et al. 2002). Taking into account different variables (the recording time, the number of utterance units, the speaking time, among others), these authors carry out an exhaustive statistical analysis of their corpus. Other authors like Van Beisen (1999) have used a corpus-driven methodology to study the different techniques involved in interpreting (i.e. anticipation). Lázaro Gutiérrez (2012) carries out a discourse analysis study from a corpus of 75 transcriptions of real doctor-patient conversations. This corpus was not tagged, but was manually processed to find out features of the asymmetry of the encounter.

Apart from studies based on manual corpora, there are few examples of projects based on machine-readable corpora. The University of Bolognia – European Parliamentary Interpreting Corpus (EPIC), 2005 – constitutes one of the first examples of interpreting corpus compilation using a machine-readable methodology. EPIC is the first large-scale interpreting corpus aimed at collecting a "large quantity of authentic simultaneous interpreting data to produce much-needed empirical research on the characteristics of interpreted speeches and to inform and improve training practices" (Russo et al. 2012: 53). EPIC is a trilingual open corpus (English, Italian and Spanish), including source speeches in those three languages and interpreted speeches in all possible combinations and directions (Russo et al. 2012: 53). It consists of nine sub-corpora (177,295 words in total). Recently, community interpreting has been looking into new ways of researching different relationships involved in social discourse. Thus, Angermeyer et al. (2012) offer to the academic community what is called *ComInDat Pilot Corpus*, a collection of two corpora of interpreted doctor-patient communication and a second corpus of interpreted court proceedings. Other authors stimulate the corpus-based interpreting methodology by providing different types of corpora – CoSi, a corpus of consecutive and simultaneous interpreting – in order to encourage the research community to use corpora in inter-

preting studies (House et al. 2012). This corpus was created by using the EXMARaLDA software (Schmidt and Wörner 2009, Schmidt and Kai 2012), which includes the EXMARaLDA Partitur-Editor, a tool for editing transcriptions in musical score notation. These are just some examples of the growing interest of corpora and interpreting and their valuable source for research.

Following in the footsteps of all these studies, our research is meant to contribute to the area of CIS by compiling a monolingual multimodal corpus on gender violence, whose ultimate goal is to train interpreters in this area. Interpreters dealing with gender violence cases have to perform their work in many different public service interpreting settings, such as police offices, medical practices, courts, social work and psychology practices. In what follows, a brief description of public service interpreting will be provided.

3   Public Service Interpreting

Public service interpreting, also called community interpreting, is performed at institutions which offer public services for the general population, as is the case in courts, hospitals, police offices, healthcare centres, schools, public administration offices, and the like. Public service interpreters bridge communication gaps between service providers (lawyers, doctors, teachers, police officers, social workers…) and users. One of the first definitions of public service interpreting is given by Wadensjö (1998: 49):

> Interpreting carried out in face-to-face encounters between officials and laypeople, meeting for a particular purpose at a public institution is (in English-speaking countries) often termed community interpreting.

The areas where public service interpreting is performed are multiple and include a great variety of settings, such as legal (considered by some authors as a distinct variety, apart from public service interpreting (Phelan 2001)), healthcare, educational, administrative, social, police setting, amongst others. Public service interpreters use the bi-directional modality both onsite, over the phone, or through videoconferencing technologies. This area of interpreting has specific characteristics that differentiate it from others. Here we include a compilation of features taken from different sources:

– Interpreters must have a deep knowledge both of the languages they interpret into and from and of the cultures their clients belong to (Valero Garcés 2006). The understanding and expression of concepts related to gender-based violence may vary amongst cultures. Thus, foreign victims may have ways of expressing their problems which may result exotic or even incomprehensible to members of the host culture. Interpreters must not only be familiar with these culturally marked pragmatics patterns of victims, but also with cultural and institutional constructs belonging to the host culture, which will condition the development of interactions between service providers and users.

– The asymmetry between the participants in the conversations (Lázaro Gutiérrez 2012). The characteristic asymmetry of institutional encounters increases when the user of public services does not speak the institutional language (the language used by the service provider, not only at a semantic level, but also at a pragmatic one).
– The tense situations in which these interpreted conversations sometimes take place (Phelan 2001; Valero Garcés 2006). Victims of gender-based violence find themselves in a complicated personal situation. Interpreters usually suffer from the stress generated by having to re-verbalise traumatic events.
– The scarce (though growing) acknowledgement of the profession, which usually leads to the fact that non-professional interpreters undertake this task, or results in poor working conditions for professional interpreters working in this field, who receive low salaries, are assigned tasks other than interpreting, have little support and resources (not receiving previous information about the topic of the conversations to be interpreted or the peculiarities of the interactants, or being called very shortly before the assignment starts (Lázaro Gutiérrez 2014).
– The performance by the interpreter of a much broader task than that of simply interpreting, which includes, among other issues, the weight and responsibility of co-ordinating the turns in the conversation (Wadensjö 1998).

Taking into account the specific characteristics of public service interpreting, according to Inglis (quoted in Iliescu 2001) a public service interpreter should:

– Master a sufficient number of general and specialised terms.
– Be able to remember communicative pragmatic patterns such as greetings, farewells, questions and other ways to obtain information; know how to ask for explanations and repetitions, spell, make remarks about a particular aspect of the conversation, control the sequence of the interaction, express agreement and disagreement, self-repair, apologise; and have the ability to repair a negative impression on the listener or perform other phatic patterns such as compliments and good wishes.
– Be aware of the nature and characteristics of discourse.
– Recognise and transfer register and tenor.
– Have a good command of syntax.
– Have a good command of discursive strategies.
– Be able to recognise and transfer the illocutionary force of the original message.
– Be able to perceive interactants' opinions and their degree of knowledge about the topic of conversation.
– Be able to grasp and transfer the interactants' points of view.
– Be able to improve the structure of the discourse.
– Notice the interactants' cultural differences and have expert knowledge about them.

Thus, interpreters willing to specialise in gender violence contexts should acquire specific vocabulary about this topic, knowledge about the most frequent structures

of this kind of interactions (questions, narratives, explanations, and the like), the nature of the encounters, the pragmatic peculiarities of the discourse (metaphors, use of empathy), amongst many other abilities mentioned by Inglis (Iliescu 2001), such as a good use of syntax, grammar and pragmatics and a deep cultural knowledge. We consider that a multimodal corpus of real conversations, natural oral or written discourse, and written documents will be useful in order to spot and systematise these features. From the findings obtained after the analysis of such a multimodal corpus, it will be possible to elaborate useful training materials for interpreters willing to specialise in gender violence contexts, that will contribute to the acquisition of what has been called "pragmatic competence", which consists of the "the ability to use language effectively in order to achieve a specific purpose and to understand language in context (Thomas 1983: 92).

4   Interpreting in Gender Violence Contexts

The research program which is presented has been developed in Spain, although we believe that its methodology could be extrapolated to many different contexts. According to the most recent data of the European Commission (Eurostat), Spain received the 5th largest number of immigrants in the European Union in 2012 (304,100). Germany reported the largest number of immigrants (592,200), followed by the United Kingdom (498,000), Italy (350,800) and France (327,400).

According to the Spanish Institute of Statistics (*Instituto Nacional de Estadística*), in 2014, the highest number of migrants came from Romania, Morocco and the United Kingdom. These data give us a clue about how necessary interpreters are for foreign languages such as Romanian and Arabic in Spain so that the population from these countries can successfully access public services.
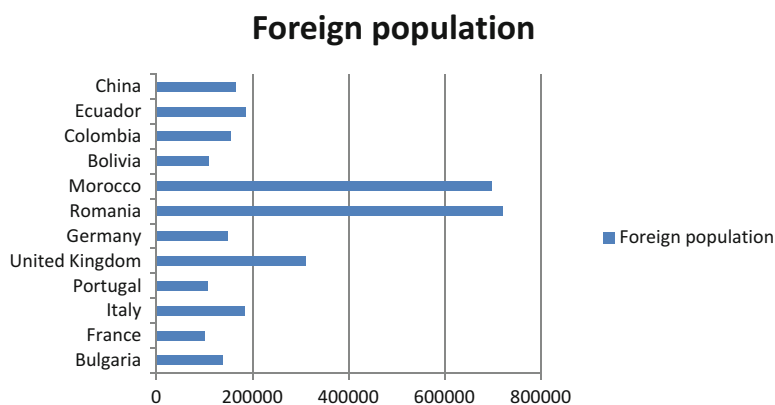
Immigration to Spain according to nationality in 2014. Spanish Institute of Statistics

On the other hand, focusing on gender violence issues, it is worth mentioning that in 2012, the European Institute for Gender Equality (2012) reviewed the implementation by the member states of the measures agreed to at the Beijing Declaration and Platform for Action against violence against women and support to victims and alerted about the fact that no less than 33 % of European women had ever suffered gender violence.

The *Directive 2012/29/EU of the European Parliament and of the Council of 25 October, 2012, establishing minimum standards on the rights, support and protection of victims of crime, and replacing Council Framework Decision 2001/220/JHA* is worth highlighting as it makes it clear that victims of crime (such as victims of gender-based violence), must be provided with the necessary tools to grant them access to legal services, such as the assistance of an interpreter. This is a legal instrument that reinforces national legislation of the member states and is based on other similar proposals, especially the *Directive 2011/36/EU of the European Parliament and of the Council of 5 April 2011 on preventing and combating trafficking in human beings and protecting its victims, and replacing Council Framework Decision 2002/629/JHA,* the *Directive 2011/92/EU of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography, and replacing Council Framework Decision 2004/68/JHA,* or the *Council Framework Decision 2001/220/JHA of 15 March 2001 on the standing of victims in criminal proceedings.* Apart from these, Art. 6 from the *Directive 2012/29/EU of the European Parliament and of the Council of 25 October, 2012, establishing minimum standards on the rights, support and protection of victims of crime, and replacing Council Framework Decision 2001/220/JHA* displays along 6 sub-articles the right to translation and interpreting, which is also recognised in the *Directive 2010/64/EU of the European Parliament and of the Council of 20 October 2010 on the right to interpretation and translation in criminal proceedings.*

In Spain, national authorities have put into practice a set of measures aimed at softening this gender-based violence phenomenon and at contributing to establish a national strategy for the elimination of this social cancer through personalised attention to the victims. The main aim of this strategy is to unite all the institutions and organisations of the country into a national network, and even to duplicate staff and resources in order to fight against gender violence. Amongst these measures is the *Basic Law 1/2004 of 28 December on Integrated Protection Measures Against Gender Violence* (*Ley Orgánica 1/2004 de Medidas de Protección Integral contra la Violencia de Género)* and the *Royal Decree 233/2005 on the Creation and Constitution of Courts on Violence against Women (Real Decreto 233/2005 a través del cual se establece la creación y la constitución de juzgados de violencia sobre la mujer).*

Other measures were the launching of the Observatory against Domestic and Gender Violence (*Observatorio contra la Violencia Doméstica y de Género)* of the General Council of the Judiciary as an instrument of analysis and action within the Spanish Justice Administration that promotes initiatives and measures oriented

towards the elimination of the social problem posed by domestic and gender violence. The National Observatory on Violence against Women (*Observatorio Estatal de Violencia sobre la Mujer*) of the Ministry of Healthcare, Social Affairs and Equality was launched in 2006 and is the institution responsible for the elaboration of reports and research about gender violence, as well as for the evaluation of the impact of the adopted policies and measures through the compilation, analysis and dissemination of materials about gender violence. In the last decade, the Network of Local Points of the Regional Observatory on Gender Violence and the hotline 016, for the assistance of women suffering from gender violence, were launched amongst many other mechanisms and resources.

However, the Macrosurvey on Gender Violence carried out in 2011 by the Centre for Sociological Research (*Centro de Investigaciones Sociológicas*) found out that the prevalence of gender violence against foreign women doubles that against Spanish women. According to this same source, 469,317 foreign women had suffered gender violence at some time in their lives, and 130,241 had suffered it in 2010. On the other hand, data published by the Spanish General Council of the Judiciary (*Consejo General del Poder Judicial*) indicates that in 2010 12 % of the gender violence victims who attended Spanish courts were of foreign origin. Only 3 years later, in 2013, this figure mounted to 35 %. If linguistic assistance is not provided, all these measures and resources are out of reach for these foreign gender-based violence victims who do not speak Spanish (fluently).

These new European and national regulations are reflected on the development of research projects such as "Speak Out for Support – SOS-VICS" (JUST/2011/ JPEN/2912), co-ordinated by the University of Vigo and with the participation of nine Spanish universities. This is a pilot project whose main objective is to train interpreters who want to specialise in gender violence contexts, interpreting for foreign victims. These interpreters, apart from a few lessons they may receive if they follow some Spanish postgraduate studies in public service interpreting and translation, do not usually get any specific training about interpreting in gender-based violence contexts, contrary to what happens with other agents (doctors, police officers, social workers, and so on) who assist victims of gender violence.

We think that, in order to perform accurately, interpreters should acquire thematic knowledge about gender violence, knowledge about institutional procedures and usual communicative events, as well as linguistic, pragmatic, cultural and, particularly, terminological command. Besides, they should be able to manage the stress experimented both by themselves (interpreters) and the other participants in the interaction, following Inglis (1984, quoted in Iliescu 2001). Interpreters working in these contexts need specialisation in linguistic mediation with foreign victims, apart from the command of the languages involved and knowledge about institutional protocols on the one hand, and emotions management on the other. With our proposal we intend to contribute to their training through a corpus-based methodology and, in what follows, we will describe the methodological phases of our project.

5 A Corpus for Interpreters Working in Gender Violence Contexts: Design, Compilation and Use

The main purpose of our project is to compile a monolingual electronic archive (Bowker 2003) on gender violence for pedagogical purposes. Although the potential benefits of applying a systematic corpus-based methodology for research on interpreting have been sufficiently acknowledged in the last few years (Straniero Sergio and Falbo 2012), they are mainly focused on studying interpreting through corpora. Hence our archive has been designed with a pedagogical purpose in mind. This archive is divided into three different, but related, corpora. As stated in Sect. 4, it is of paramount importance to offer high quality training to our students so that they can provide a successful interpreting service in gender violence contexts. We believe that this training gap can be filled with the design and compilation of an archive focused on gender violence to analyse and research the gender violence genre (terminology, spoken and written language, register, pragmatic patterns, etc).

Our final corpus will provide:

1. a comprehensive knowledge of the gender violence genre
2. real language and authentic material to design our interpreting training sessions
3. fairly accurate statistics of word occurrences (essential to design vocabulary acquisition activities)
4. examples of discursive patterns corresponding to different cultural pragmatics
5. quick access to large texts.

From a pedagogical perspective, a corpus-based methodology has proved to be the most adequate. Based on a deductive approach, this methodology will enable us to analyse patterns of use for pre-defined linguistic features (i.e. word frequency, linguistic and register variations of a given category, frequency of pragmatic patterns, and so on).

The different advantages of working with a corpus-based methodology for pedagogical purposes were highlighted by authors such as Flowerdew (1993: 91). According to this author, working with corpus and concordance programs has three main applications for trainers: (i) as a linguistic informant, (ii) as a source of input for training, and, finally, (iii) as input for materials developments. Firstly, as a linguistic informant, the trainer has the possibility to access the corpus in order to corroborate both grammatical and lexical choices, as well as expressions and other pragmatic patterns. Secondly, as a source of input for training, the trainer can use the corpus to generate authentic examples of usage, which would reflect all the levels of language (including the pragmatic level) and the communicative situation. Finally, applying corpus as input for material development can be successful if the following conditions are fulfilled:

1. the trainer is aware of the students' strengths and weaknesses and knows which linguistic points (lexis, grammar, pragmatics…) need to be improved
2. the trainer wishes to design his or her own material

3. the trainer is computer-literate and has the proper software and concordance tools.

Our corpus is made up of three main sources:

1. Texts: manuals and protocols, scientific documents
2. Videos: simulated videos and real conversations
3. Texts: written spontaneous discourse.

It comes as no surprise that our corpus could cover many areas related to gender violence due to the fact that interpreters can work in different settings. As a starting point for our research, we have decided to compile a monolingual corpus on gender violence context in medical and social settings.

5.1 Corpus Design

We were aware that just a compilation of texts (both scientific documents and practical documents such as manuals and guides of practice) was not enough to cater to the training needs of the interpreters. Public service interpreters have to deal with spontaneous oral conversations and interpret both the service provider's and the victim's discourse, each of them with different characteristics, most of them have to do with the pragmatic level of language. Bearing this in mind, our corpus includes spoken discourse (both real and simulated conversations), and written discourse spontaneously produced by victims.

Our corpus contains data from different sources compiled between 1998 and 2014. As stated before, it consists of three sub-corpora. We describe our corpus and the relevant background information contained in our corpus data in the next paragraphs.

(a) Texts: manuals and protocols, scientific documents

Corpus data was collected mainly from the Internet. Official websites of public administrations and NGO associations were especially useful. The usefulness of the project SOS-VICS, which offers a great repository of documents on this field in its public webpage, has to be highlighted. In order to follow a well-designed compilation protocol, we used specific software to automate the downloading process (i.e. HTTrack, GNU Wget or Jdownloader). These tools allow the downloading of websites at one go. Texts can be downloaded automatically. 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. This task is especially necessary in the case of texts retrieved in .pdf format. Finally, all documents were stored in different files.

(b) Videos: simulated videos and real conversations

Another important part of our corpus was a set of videos of medical consultations with victims of gender violence. Here we can distinguish two different

resources: simulated videos, which are accessible via internet and have been published by universities or healthcare organisms; and real conversations, which were recorded in Spanish general practitioner (GP) consultations and hospitals and belong to the FITISPos-UAH Group (1998–2004). Another source of material comes from the InterMed project, which records and analyses real conversations with foreign patients in GP consultations.

Visits to several organisations and associations, such as local points of the Regional Observatory on Gender Violence and several women's associations, were carried out.

Copyright issue was the main problem at this stage. Many associations have shown interest in the project and have contributed to the corpus compilation.

As a future task, and in order to get the most out of this multimodal material, we believe it is important to tag our corpus so that we can analyse it. Tagging will allow us to clearly differentiate the different actors taking part in the discourse (i.e. healthcare providers and patients or victims). Based upon the work done in other projects – European Parliamentary Comparable and Paralles Corpora, ECPC – (Calzada Pérez et al. 2007), our corpus will be annotated in XML (*Extensible Mark-up Language*) with different types of data: linguistic, contextual, and metalinguistic information. XML-tagging will be carried out semi-automatically with the use of regular expressions.

```xml
P<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!DOCTYPE maltrato SYSTEM "maltrato.dtd">
<maltrato>
<doctor>D<text num="1"> a ver Fátima ¿eres alérgica a algo? ¿alergia?</text></doctor>
<paciente>P<text num="2"> no</text></paciente>
<doctor>D<text num="3"> ¿alguna enfermedad importante a parte de lo de la piel?</text></doctor>
<paciente>P<text num="4"> sí pues como... no sé cómo se llama [eh]</text></paciente>
<doctor>D<text num="5"> [esa] enfer[medad] </text></doctor>
<paciente>P<text num="6"> [sí]</text></paciente>
<doctor>D ¿alguna otra?</doctor>
<paciente>P<text num="8"> pues no</text></paciente>
<doctor>D<text num="9"> de corazón [pulmón]</text></doctor>
<paciente>P<text num="10"> [no]     </text></paciente>
<doctor>D<text num="11"> estómago              </text></doctor>
<paciente>P<text num="12"> no</text></paciente>
<doctor>D</doctor>
<paciente>P<text num="14"> no</text></paciente>
<doctor>D<text num="15"> ¿depresión o algo?</text></paciente>
<paciente>P<text num="16"> no [creo]</text></paciente>
<doctor>D<text num="17"> [de] tristeza o</text></doctor>
<paciente>P<text num="18"> no</text></paciente>
<doctor>D<text num="19"> estás bien ¿verdad? ¿tomas algún tratamiento? Ahora lo que te echas para la
<paciente>P<text num="20"> sí</text></paciente>
<doctor>D<text num="21"> ¿pero aparte tomas algo por la boca todos los [días]?</text></doctor>
<paciente>P<text num="20"> sí</text></paciente>
<doctor>D<text num="23"> ¿qué tomas?</text></doctor>
<paciente>P<text num="24"> pues la primera vez es de hoy no sé cómo se llama</text></paciente>
<doctor>D<text num="25"> ah la antibiótico</text></doctor>
<paciente>P<text num="26"> sí</text></paciente>
<doctor>D<text num="27"> vale por la mañana y por la noche ¿no?</text></doctor>
<paciente>P<text num="28"> sí</text></paciente>
<doctor>D<text num="29"> vale entonces esto comenzó ayer ¿no?</text></doctor>
<paciente>P<text num="30"> sí</text></paciente>
<doctor>D<text num="31"> ¿qué pasó ayer?</text></doctor>
<paciente>P<text num="32"> pues ayer están discutiendo mi (xxx) mi hermanos los dos eso sí</text></p
<doctor>D<text num="33"> sus hermanos empezaron a discutir</text></doctor>
<paciente>P<text num="34"> sí</text></paciente>
<doctor>D<text num="35"> ¿y qué pasó?</text></doctor>
<paciente>P<text num="36"> pues me empezó mi mano así me</text></paciente>
<doctor>D<text num="37"> la mano</text></doctor>
<paciente>P<text num="38"> sí</text></paciente>
<doctor>D<text num="39"> ¿pasó algo entre sus hermanos a parte de la de las voces o</text></doctor>
```

XML-tagging

(c) Texts: written spontaneous discourse

In spite of every effort to obtain recordings from which samples of discourse uttered by victims could be extracted, the number of real conversations we were able to gather was scarce. Aware of the great importance of the compilation of spontaneous discourse, it was decided to include in our corpus the contributions that women themselves made to specialised consultation fori about their situations as victims. These contributions are either accessible online, such as the ones sent to the forum of the Region of Castilla-La Mancha, or have been provided by women's associations.

In terms of corpus analysis, there is no doubt that working with corpus requires an efficient use of software. Although we are in a very initial stage of our research, we believe it is relevant to know the tools to be used in order to accomplish our main goal: training interpreters working in gender violence contexts. For instance, concordance software programs will provide the most frequent words service providers and victims use in their communicative interaction. It will enable us to identify appropriate (and inappropriate) terminology, collocations, phraseology, pragmatic patterns, style and register of the gender violence discourse. There are different programs available. One of the most popular is Wordsmith, designed by Mike Scott. It offers a wide range of possibilities for analysing corpora, such as XML reading facilities, wordlists, keywords and concordances. It incorporates follow-up concordance searches, file viewer utility, a corpus corruption detector or a concgram facility. XML reading facilities can be very useful if we want to analyse our second sub-corpus (videos) separately so that we can compare the different speakers' speeches in terms of terminology, for example. This analysis will retrieve the most frequent words to elaborate a wordlist in order to design activities to train our interpreting students. Other functions are particularly useful to grasp the pragmatics of discourse, as they can be used to retrieve utterances in context or collocational and contextual information (concord).

Wordsmith is an indispensable tool but there are other programs that provide reliable information. We can take AntConc as an example, a freeware multiplatform tool created by Laurence Anthony in 2004. Functions are very similar. Wordlist and word frequency proved useful for the three sub-corpora as well as collocates and clusters function.
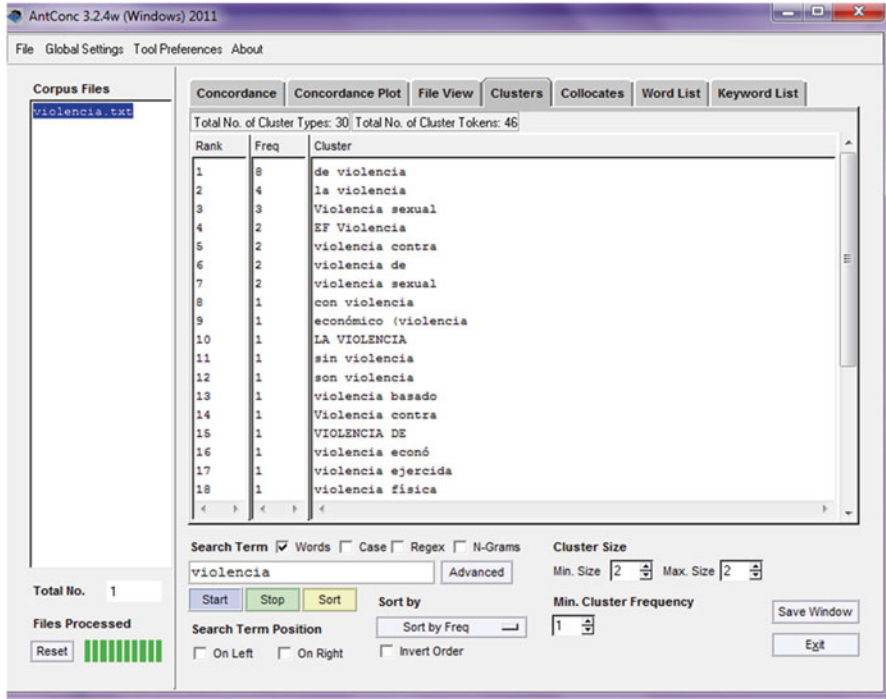
AntConc collocates function

5.2 Some Examples of Practical Application

As previously stated, one of the main challenges for interpreters working in fields related to gender-based violence consists on the rendition of the pragmatic content of messages. They struggle every day to be able to render pragmatically appropriate utterances, matching both the situational context and the intents of the people whose discourse they are interpreting. Interpreter-mediated natural conversations are an example of communication across languages and cultures, which, as Kecskes and Romero-Trillo (2013: 1) point out, has become the new challenge for pragmatics research in the twenty-first century.

Interpreters usually receive training about interpreting techniques (including memorising, diction, making notes, and the like), terminology, and advice on how to prepare for an assignment. One of the most difficult aspects is to obtain contextual information, including details about the communicative event, which is usually structured under institutional constraints, or the culture of the participants, which may determine their communicative styles and their use of language and pragmatic patterns. Training and information gathering about these specific issues must not be taken for granted because, as Kecskes and Romero-Trillo (2013: 1) state, "our individual comprehension of language is dependent upon our biographical socio-cultural experience". Vital experiences might not be enough for interpreters to grasp the

meaning of the utterances of the participants in interpreter-mediated interactions. This is so because of interpreters might belong to the cultural group of one of the interactants or to none of them, making it obvious that they will lack pragmatic information from at least one of the parties.

Once our corpus is fully compiled, we will be able to analyse it taking into account a discourse level, that is, our corpus will allow us to analyse it according to the particular setting of conversations (social, medical, legal, etc. and the different cultures of the victims. Our point of departure is that, as Baider (2013: 8) suggests "words are not culturally neutral and bring with them certain culture-specific ways of thinking". We agree with Wierzbicka (2006) in that we cannot take for granted equivalence between two languages, and go a step further and add that we cannot take for granted equivalence between two cultures, even if people belonging to both of them speak the same language.

Particularly in the field of gender-based violence and precisely when it comes to interpret the discourse of the victims, it is especially challenging for the interpreter to grasp and transmit the meaning of taboo concepts, as they are usually expressed in a very creative way, so as to conceal the taboo part of them. Although it is still soon to advance results of the analysis of our corpus, which is still in a compilation phase, a good example of this can be obtained from Torruella Valverde (2013), when she reports about the experiences of interpreters and tells and gives the following example. One interpreter of Arabic had to render once the meaning of a gesture performed by a gender-based violence victim who was declaring in court. She brought her hand close to her ear, as if she was speaking on the phone, but without touching her head. The interpreter, who shared the culture of the victim, could understand that she meant that her husband had threatened her for death. Imagine how difficult it is for an interpreter who does not know the culture of their interlocutors to grasp the meaning of certain metaphors and uses of language that are the result of the speakers' creativity to avoid taboos.

6   Conclusions

The emergence of new projects focused on CIS has made it necessary to reflect on the need for more research in corpus linguistics and interpreting. Public Service interpreters in particular often lack both general and specialised training. Furthermore, many scholars support the idea that interpreters require a precise understanding of specific discourses, such as the one produced in gender violence contexts, as there are many specific features (particularly pragmatic ones) that are characteristic of specific contexts.

However, training interpreters in specific contexts is challenging because training materials must be consistent with real data. The analysis through CIS of the gender violence discourse produced both by victims and by service providers provides the researcher with valuable linguistic material (i.e. terminology, phraseology, metaphors, pragmatic patterns) that can be transformed into useful training resources.

Although we are aware that CIS has challenges and obstacles that need to be sorted out, we believe that this area of research is the key to accessing real data. We are aware that many methodological questions still remain open and are only likely to be answered once our research has been completed. A closer collaboration between academics, service providers and interpreters could be one of the potential answers, as it may allow us to enlarge our corpus with pieces of real oral discourse (recordings). Although gathering collaboration for this purpose is usually a hard task, it is hoped that, in the future, we will be able to reach our objectives and enlarge our corpus to gradually become a useful resource for the academic community, thus representing what gender violence discourse is and how it should be taught to interpreters in the classroom.

References

Angermeyer, P. S., Meyer, B., & Schmidt, T. (2012). Sharing community interpreting corpora: A pilot study. In T. Schmidt & K. Wörner (Eds.), *Multilingual corpora and multilingual corpus analysis* (pp. 275–294). Amsterdam/Philadelphia: John Benjamins Publishing.

AntConc website http://www.laurenceanthony.net/software.html. Accessed 20, 22 Sept 2014.

Baider, F. (2013). Hate: Saliency features in cross-cultural semantics. In K. Istvan & J. Romero-Trillo (Eds.), *Research trends in intercultural pragmatics* (pp. 7–28). Boston/Berlin: De Gruyter.

Baker, M. (1993). Corpus linguistics and translation studies – Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 233–253). Amsterdam/Philadelphia: John Benjamins.

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, L. (2003). Corpus-based applications for translator training: Exploring possibilities. In S. Granger, J. Lerot, & S. Petch-Tyson (Eds.), *Corpus-based approaches to contrastive linguistics and translation studies* (pp. 169–183). Amsterdam/Philadelphia: Rodopi.

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

Calzada Pérez, M. (2007). *El espejo traductológico. Teorías y didácticas para la formación del traductor*. Madrid: Editorial Octaedro.

Calzada Pérez, M., Cucala, N. M., & Martínez, J. M. (2007). ECPC: European Parliamentary Comparable and Parallel Corpora. *Procesamiento del Lenguaje Natural, 37*, 349–351.

Centro de Investigaciones Sociológicas. (2011). *Macroencuesta sobre Violencia de Género*. Madrid: Ministerio de Sanidad, Ciencias Sociales e Igualdad.

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/Philadelphia: John Benjamins.

EPIC website http://sslmitdev-online.sslmit.unibo.it/corpora/corpora.php. Accessed 20, 22 Sept 2014.

Flowerdew, J. (1993). Concordancing as a tool in course design. *System, 21*(2), 231–244.

House, J., Meyer, B., & Schmidt, T. (2012). CoSi – A corpus of consecutive and simultaneous interpreting. In T. Schmidt & K. Wörner (Eds.), *Multilingual corpora and multilingual corpus analysis* (pp. 295–304). Amsterdam/Philadelphia: John Benjamins Publishing.

Iliescu Gheorghiu, C. (2001). *Introducción a la Interpretación: la modalidad consecutiva*. Alicante: Universidad de Alicante.

Inglis, M. (1984). *A preparatory course in bilateral interpreting*. Edinburgh: University of Edinburgh, Department of Linguistics.

Instituto Asturiano de la Mujer. Disponible en. http://institutoasturianodelamujer.com/iam/category/biblioteca-digital/. Consultado el 15 May 2014.

Instituto Europeo de Igualdad de Género (European Institute for Gender Equality) (2012). *Examen de la aplicación de la Plataforma de Acción de Beijing en los Estados miembros de la Unión Europea sobre violencia contra la mujer y apoyo a las víctimas*. Brussels: EIGE.

Kecskes, I., & Romero-Trillo, J. (2013). Introduction. In I. Kecskes & J. Romero-Trillo (Eds.), *Research trends in intercultural pragmatics* (pp. 1–3). Boston/Berlin: De Gruyter.

Kenny, D. (2001). *Lexis and creativity in translation. A corpus-based study*. Manchester: St. Jerome.

Lázaro Gutiérrez, R. (2012). *La interpretación en el ámbito sanitario. Estudio de la asimetría en consultas médicas*. Saarbrücken: Editorial Académica Española.

Lázaro Gutiérrez, R. (2014). Use and abuse of an interpreter. In C. Valero-Garcés (Ed.), *(RE) Considerando ética e ideología en situaciones de conflicto*. Alcalá de Henares: Servicio de Publicaciones de la Universidad de Alcalá.

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.

Linkterpreting, Universidad de Vigo. http://linkterpreting.uvigo.es/interpretacion-social/material-didactico/. Accessed 15 May 2014.

Luzón, M. J., Campoy, M. C., del Mar Sánchez Ramos, M., & Salazar, P. (2008). Spoken corpora: New perspectives in oral language use and teaching. In M. C. Campoy & M. J. Luzón (Eds.), *Spoken corpora in applied linguistics* (pp. 3–30). Bern: Peter Lang.

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

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies. An advanced resource book*. London/New York: Routledge.

Phelan, M. (2001). *The interpreter's resource*. Manchester: Multilingual Matters.

Pöchhacker, F. (2008). Inside the 'black box'. Can interpreting studies help the profession if access to real life settings is denied? *The Linguist, 48*(2), 22–23.

Russo, M., Bendazzoli, C., Sandrelli, A., & Spinolo, N. (2012). The European parliament interpreting corpus (EPIC): Implementation and developments. In F. S. Straniero & C. Falbo (Eds.), *Breaking the ground in corpus-based interpreting studies* (pp. 53–90). Bern: Peter Lang.

Ryu, K., Matsubara, S., Kawaguchi, N., & Yasuyoshi, I. (2003). Bilingual speech dialogue corpus for simultaneous machine interpretation research. In *Proceedings of the 6th Oriental COCOSDA workshop in Singapore* on Oct 1–3 2003 (pp. 164–168). http://ir2.nul.nagoya-u.ac.jp/jspui/handle/2237/15076. Accessed 15 May 2014.

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 Ramos, María del Mar., & Javier Vigier Moreno, F. (in press). Using corpus management tools in public service translator training: An example of their application in the translation of judgments. In C. C. Antonio Pareja-Lora & P. Rodríguez Arancón (Eds.), *Technological innovation for specialized linguistics domains*. Oxford: Oxford University Press.

Schmidt, T., & Wörner, K. (2009). EXMARaLDA – Creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics, 19*, 565–582.

Schmidt, T., & Kai, W. (2012). *Multilingual corpora and multilingual corpus analysis*. Amsterdam/Philadelphia: John Benjamins Publishing.

Straniero Sergio, F., & Falbo, C. (Eds.). (2012). *Breaking ground in corpus-based interpreting studies*. Bern: Peter Lang.

Shlesinger, M. (1998). Corpus-based interpreting studies as an offshoot of corpus-based translation studies. *Meta, 43*(4), 486–493.

Takagi, A., Matsubara, S., & Inagaki, Y. (2002). A corpus-based analysis of simultaneous interpretation. In *Proceedings of international joint conference of the 5th symposium on natural language processing* (SNLP-2002) (pp. 167–174). http://slp.itc.nagoya-u.ac.jp/web/papers/2002/snlp2002_takagi.pdf. Accessed 15 May 2014.

Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics, 4*(2), 91–112.

Thompson, P. (2005). Spoken language corpora. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 59–70). Oxford: Oxbow Books.

Torruella Valverde, J. (2013). *La adecuación en la interpretación en los SSPP. Comunicación no verbal e implicaciones culturales (árabe-español Alcalá de Henares).* Unpublished master's dissertation.

Valero Garcés, C. (2006/2008). *Formas de Mediación Intercultural e Interpretación en los Servicios Públicos. Conceptos, Datos, Situaciones y Práctica*. Granada: Comares.

Van Besien, F. (1999). Anticipation in simultaneous interpretation. *Meta, 44*(2), 250–259.

Wadensjö, C. (1998). *Interpreting as interaction*. Londres y Nueva York: Longman.

Wierzbicka, A. (2006). *English: Meaning and culture*. Oxford: Oxford University Press.

Wordsmith website http://www.laurenceanthony.net/software.html Accessed 20 Sept 2014.

Zanettin, F., Bernardini, S., & Stewart, D. (Eds.). (2003). *Corpora in translator education*. Manchester: St. Jerome.

Zhu, C., & Hui, W. (2011). A corpus-based, machine-aided mode of translator training. *The Interpreter and the Translator Trainer, 5*(2), 269–291.