

# Using Corpus Linguistics as a research and training tool for Public Service Interpreting (PSI) in the legal sector

CINZIA SPINZI  
University of Palermo (Italy)

## Abstract

*Public Service Translation has for long been the 'forgotten voice' in PSI studies but it is arguably a valuable linguistic support for legal institutions and for training interpreters in the legal sector. Given that interpreters in the legal system in Italy often tend to 'double-up' as legal translators (to make a living) the line between the two is often hazy. Hybrid modalities like sight translation of legal and administrative documents is also a 'borderline' feature of these intertwined professions. The main aim of this paper is to describe how parallel and monolingual corpora can be used to train public service interpreters in double roles (translators, interpreters), namely by using corpora to translate, in multiple community languages. To this purpose, a computerized corpus has been constructed as a representative sample of learners' renditions of legal texts. Then, other two corpora, monolingual and parallel corpora, have been used to verify the stumbling blocks dialogue interpreters struggle with, e.g. discourse markers and phraseological constructions. Corpus data are used descriptively (analyzing data) and prescriptively (providing examples of correct phraseological language usage in the languages at issue). In other words, I will describe how this methodology – through the collection of voice-recorded parallel corpora – is an invaluable tool in the training of legal (dialogue) interpreters. My ultimate aim is to provide concrete tools for legal interpreters and their trainers to facilitate their task primarily by constructing a multilingual parallel corpus as a resource for both academic research and PSIT practitioners.*

Public Service Interpreting, corpus linguistics, training, discourse markers.

## Preamble

Only a society that can provide access to basic services that safeguard health, education and justice is a true democracy. Such access, however, can be sorely tried when the number of citizens in any given country increases rapidly and suddenly. As such, in the multicultural, multiethnic and multilingual project of European integration, translation and interpreting play a fundamental role in ensuring basic human rights for all citizens. As stated by the Vice-President of the European Commission then responsible for Justice, Fundamental Rights and Citizenship (Viviane Reding, October 25<sup>th</sup>, 2010) “A Citizens’ Europe – a Europe for and by the people – can only be built successfully if we ensure that people can exercise their rights, in full knowledge of what those rights are and how they benefit them. [...]”. Furthermore, whether or not interpreting as a profession and discipline is fully recognized and institutionalized will depend in part on the perception of multiculturalism in a given society and whether or not minority communities as members of our society are seen to have the right to full access to public services and the service channels that enable such access (Valero Garcés/Francisco 2012: 13). Equal access to citizens’ legal services is, or should be, provided for at the institutional level, but as the number of foreign citizens increases in any given geographical area, the institutions themselves are often unable to guarantee an adequate provision of services. The geo-political climate of a country at any given time deeply affects how such services are (un)equally distributed and how the provision of services is organized. The wave of populist politics emerging in the second decade of the third millennium and spreading across many Western countries is not benefitting the provision of services and the growth and professionalization of the public service interpreting and translation (PSIT) profession.

These difficulties affect countries in the Americas, Europe, Australia and all those countries where an increasing number of incoming people do not have full command of the language used in public services (PS) and, more specifically, the legal system. As such, not only are translators and interpreters’ crucial for effective communication, but they are a channel through which to safeguard the basic human rights of all citizens. This ‘mission statement’ drives most of the PSI literature, be it in the legal, health, educational, social, refugee, or other broadly humanitarian sectors that touch basic civil rights; it is an underlying ethical

- 1 Although PS interpreting and translation are very different skills, in the PS domain many interpreters also undertake written translation to supplement an often-meagre income. Interpreters in the legal system in Italy, who are grossly underpaid, often tend to ‘double-up’ as legal translators and the line between the two is often hazy. Also, hybrid modalities like sight translation of legal and administrative documents is a ‘borderline’ feature of these intertwined professions.

approach advocated by most interpreting institutions, associations, and broader platforms engaged in this activity, namely that the negation of access to basic citizen's services is a breach of basic rights.

The present study forms part of a wider project initiated and carried out by the author and other colleagues over the last decade, namely LegaII, "Legal Interpreting in Italy" (Rudvin/Spinzi 2013), which is profoundly driven by the 'mission statement' mentioned above. The overall aim of this project is to further, strengthen and promote quality language services in public institutions in my local area, help improve the quality of legal interpreting, and establish and consolidate collaboration with local legal institutions, and undertake activities such as the construction of a multilingual parallel corpus as a translation and interpreter training resource for academic research and PSIT practitioners. One of the more specific aims of LegaII and of this paper is to utilize the tools of corpus linguistics to build a training and research 'package' to be made available to professionals, trainers and students who want to promote and further legal interpreting and translating. More particularly, I will focus on the two main areas of interpreting which need corrective feedback according to my data and that are the use of some pragmatic markers and the phraseology featuring legal language<sup>2</sup>. Both areas will be investigated and discussed in the following sections and sub-sections where my main assumption is that explicit teaching improves accuracy (Robinson 1996: 36-37).

A similar corpus-based PSI project is the one launched and nearly completed at the Autonomous University of Barcelona and financed by the Spanish Ministry of Economy and Competitiveness. This project, called TIPp (*Traducción e Interpretación en los Procesos penales*, i.e. translation and interpreting in criminal court proceedings), investigates quality in court interpreting as an element to safeguard procedural guarantees in criminal proceedings and is developing electronic resources to help court interpreters of five different languages (Orozco-Jutorán 2017; forthcoming). Similar initiatives are emerging in other countries too. One of the ways in which texts collected and compiled through projects such as these can be made more useful and accessible to the public is precisely through Corpus Linguistics (CL). Despite the methodological difficulties of accessing and analyzing data in PSI settings, corpus-based studies are thus potentially valuable bringing these resources into the public domain as concrete findings of analyzed data or, at a further step, as training tools.

- 2 Another area, which I have not addressed here and which falls outside the legal domain, is idiomatic expressions in English, which proved a true challenge for Italian trainees. Paradoxically, the technical language in the medical, and to some extent business sectors were less of a challenge because the etymological roots of the technical terms derive from Greek and Latin and were familiar to Italian students – indeed medical terminology is very often similar in the English-Italian language-pair. Counter-intuitively, it is frequently the idiomatic expressions and words that English speakers find easy and natural – at the level of every-day expressions – that pose a true challenge to Italian trainees at the level of comprehension.

1. Corpus-based Interpreting Studies as a methodological tool for PSI practitioners, researchers and trainers: opportunities and limitations.

The application of CL, a visual-text based medium, to interpreting is problematic for various reasons: technical (creating a spoken discourse corpus), institutional (confidentiality) and interactional (dialogic and human/social features) (see also Cencini 2002 and more recently Lázaro Gutiérrez/Sánchez Ramos 2015<sup>3</sup>). Accessing the domain of PS oral interactions, that is of a highly restrictive and private nature is inevitably a frustrating, arduous and lengthy process that requires perseverance – to get the required permission to record data in confidential medical, social or legal settings – and patience, insofar as it requires the physical presence of the researcher (or person collecting the data). Furthermore, these interactions often engage a very small group of people and may not last very long, so the collection of large amounts of data is much more complicated and time-consuming than with more technologized and accessible conference interpreting or interpreting in international organizations where large amounts of data can be accessed. This is one of the inherent limitations of most corpus-based PSI projects and greatly restricts the construction and analysis of large datasets and the subsequent ability to draw more general conclusions regarding social, pragmatic, discursual and textual features of PSI interactions, even at a local level. Acquiring data that is in any way ‘representative’ of a larger reality thus becomes very difficult indeed<sup>4</sup>. Nevertheless, collecting and constructing a broad range of analyzable datasets locally is one way of contributing to build a ‘bigger picture’ of interpreter-mediated discourse in the PS setting.

The problematic issues of privacy and permission, recording, transcription of oral data that beset corpus analysis of interpreter-mediated interactions do not apply to the same extent to written translation, which lends itself easily to this methodology. Its usefulness lies in the easier access of data, i.e. easily ‘manageable’ in corpus linguistic terms. Even with large datasets it is a reasonably easy process of analysis, and also in the creation of standardized and accessible multilingual datasets and other training materials. Multilingual corpora can then be used to assist translation in multiple community languages, and to update administrative documents, especially those that are used daily or regularly by police and other legal institutions. The emerging branch of Translation and

- 3 Cencini (2002) identifies these same practical, technological and institutional limitations to the collection of oral data for CL purposes; more than a decade later we see that these same challenges have not yet been overcome and are also described by Lázaro Gutiérrez/Sánchez Ramos (2015) regarding the construction of their PS spoken corpus at the University of Alcalá.
- 4 Individual case studies in Discourse Analysis sometimes give the impression of being representative of a larger reality when in fact it is only a very small sample of one or a few encounters and can only be illustrative of discourse features in that specific event. The two main reasons for this is the lack of accessibility to private and/or confidential data, and secondly, to the highly time-consuming transcription process of even small amounts of spoken data. This may change in the future with advances in data-processing technology.

Interpreting Studies known as *Translation and Interpreting in Public Services* (TIPS) also covers written translation, which has for long been the ‘forgotten voice’ in PSI studies. The works of Valero Garcés (2014), Valero Garcés/Francisco (2012) and Vargas Urpi (2011) in Spain represent the first in-depth studies on linguistic/ translational and policy aspects in this area in Europe.

## 2. Data and methods

### 2.1 Corpus description

The data used in this paper were chosen from three sub-corpora of an open miscellaneous legal corpus (LegalI corpus, under construction). The LegalI corpus can be considered a composite corpus, namely a set of several individual corpora, each of which encompasses a specific sub-field and can be used independently to investigate spoken and written (legal) language. The compilation of the corpus reflects the complexity of legal discourse.

The design of the miscellaneous corpus, shown in Table 1, includes a bilingual (Italian/English) Learner Corpus, a monolingual spoken corpus of legal English (MonoLegalI), and a written parallel corpus of legal Italian texts and their English translations (ParaLegalI).

The Learner Corpus is based on 30 hours of voice-recorded data of simulated interpreted interactions between Italian service providers and non-Italian speakers and has helped to identify some of those stumbling blocks dialogue interpreters struggle with and which were then incorporated into our training programmes. The audio-recordings stretch over a ten-year period of an oral exam that is part of the final assessment of two MA programmes. These include a course on ‘language mediation’ (in Italy broadly synonymous with ‘PSI interpreting’) and a Continuous Development course in legal interpreting at the University of Bologna where students are required to interpret a dialogue from and into English and Italian<sup>5</sup>. It should be noted that the MA students are C1+ learners of English (based on the Common European Framework of Reference for languages), and that English is the medium of instruction. Due to time constraints, only some interactions have been transcribed amounting to a total of 11,075 running words which have been used as an ‘exploratory channel’ to highlight students’ uncertainties and difficulties when interpreting from English into Italian and vice versa.

MonoLegalI contains samples of the English (spoken) language as it is used in different legal settings such as police stations and courtrooms; the latter being of a complex hybrid nature (dialogic/monologic). The data constituting the Monolingual Police Interview Corpus (MoPICo) come from two different legal cases (for a total of four police interviews and two hearings) involving criminal offences<sup>6</sup>. In order to manage these police interview transcripts more efficiently

5 The author has access to the data as one of the trainers involved in the Intercultural Communication module.

6 The cases considered are the following: the first is known as the Steven Avery case, a

and to allow more accurate searches, the corpus has been annotated using XML mark-up, according to the TEI Guidelines (Sperberg-McQueen/Burnard 2007), so that specific contextual information could be retrieved during the analysis; it incorporates information about individual speakers and their roles, date of the interview and text structure.

ParaLegallII comprises two sub-corpora, i.e. the Parallel Corpus of Administrative Documents (PaCad), including source and target texts which mainly perform a bureaucratic function such as identification of a person, renewal of residence permit and so forth. The data contained in this sub-corpus come from official Italian judiciary police websites where the need for publishing multilingual documentation makes it easier for non-Italians to access vital information in their own language. A bilingual parallel corpus like PaCad is a translation corpus in the strictest sense (Biel 2010: 4). The other sub-corpus contains precautionary measures and injunctions granted by the Court of Palermo. These documents are written orders requiring a party to take certain steps or refrain from them. They have been translated from Italian into English by professional translators. This parallel sub-corpus has not been considered for the analysis. The text retrieval software used for quantitative observations and for extracting phraseology is AntPConc<sup>7</sup>.

For the purposes of the present study I will rely on the Learner Corpus along with two sub-corpora, MoPICo and PaCad.

Corpus	Sub-corpora	Corpus type	Language mode	Total running words
Learner Corpus		Bilingual - dialogic	Spoken (IT > EN) (EN > IT)	30 hours recorded (from the MA course and the Continuous Development Course)
MonoLegall	Police interviews (MoPICo)	Monolingual	Spoken (EN)	8,035 (English, 7 hours)
	Court hearings	Monolingual	Spoken (EN)	4,917 (English, 3 hours)
ParaLegall	Administrative documents (PaCad)	Parallel	Written (IT > EN)	16,345 (Italian source texts) 14,322 (English target texts)
	Precautionary measures and injunctions	Parallel	Written (IT > EN)	18,677 (Italian source texts) 16,323 (English target texts)

Table 1. Miscellaneous Legal Corpus (LegaII).

man from Wisconsin wrongfully convicted of attempted murder in 1985. Only after 18 years of a 32-year sentence, was he released to be charged with another murder two years later; the second case included in the corpus is the case of English DJ Sir James Vincent Savile, accused of child sexual abuse in 2009.

7 <http://www.laurenceanthony.net/software/antpconc/>.

## 2.2 Analytical steps

The corpus-based methodological focus of the current investigation is manifold, including three distinct phases here termed as Identification, Verification, and Activation. CL functions foremost as a data-collecting research tool (Identification and Verification) and secondly as a tool through which to illustrate and demonstrate ‘correct’ language-specific phraseology (Activation).

The first step of this three-pronged study is a preliminary investigatory deductive phase in which problems of translation accuracy in PSI learners have been searched for (the Identification Phase). The data for the Identification Phase come from the Learner Corpus. During the course of the training and assessment period, problem areas that were noticed were (predictably) terminological in legal system-bound technical phrases (but for the purposes of the MA exam, many terms were provided by the trainer at the beginning of the exam) and discourse markers (DMs). The former emerged through phenomena such as simpler structures and less ambiguous expression. The latter, unless they contained a strong propositional content, were often sacrificed, especially when students were under stress.

Apart from being intuitively plausible, the terminological/phraseological challenge for my students is confirmed by 20+ years of experience in the classroom and by consulting with colleagues with similar didactic experience. In the Continuous Development course data, the challenges were predominantly terminological and the role plays included few DMs. Stretching across all professional domains DMs are of course general, ubiquitous discourse facilitators, and it is also for this reason that I chose to focus on them, preliminary in a Verification Phase in which they appeared specifically in the field at issue – the legal domain.

In the second phase (the Verification Phase) I investigated challenging features in a monolingual English corpus of police interviews, looking at the more formal and writing-based features of legal discourse that can also be present in an interpreter-mediated setting.

Lastly, the parallel bilingual Italian-English specialized corpus (in progress) of legal documents of a procedural nature (reports and summary notices) was used to see which terminological and phraseological, or indeed other features emerged as potential trouble areas. In the last step of this methodology, I show phraseological examples to be adopted to provide input for PSIT trainees (the Activation Phase). The primary aim of the present study is thus to gain a clearer insight into those stumbling blocks observed in the Learner Corpus (i.e. discourse markers and also more specialized phraseologies) and then to use CL tools to investigate occurrences and contexts in which these occur in order to assist us in formulating more targeted tools – e.g. parallel multilingual specialized corpora – to be used for training purposes.



### 3. Discourse markers and phraseology

Discourse markers (DMs) have been the object of numerous studies in a vast range of language domains. As noticed in the literature, DMs do not impinge on the propositional content as such but create communication at the level of rapport and beyond the immediate co-text:

[...] the truth condition of the utterance is not affected by them, the propositional content of the utterance is not altered, they relate to the speech situation and their function is emotive and connotative rather than denotative or referential. (Hale 2004: 61 referring to Hölker 1991: 78-79)

DMs have numerous functions, but generally speaking they could be said to facilitate discourse coherence between interlocutors in a conversation (Hale 2004: 62) and they may fulfill an interactional function. Schiffrin (1987: 326) has highlighted their relevant role in building up coherence by “locating utterances on particular planes of talk” and claims that these words contribute to making the speech more understandable and coherent. Other scholars have stressed the pragmatic function of these ‘smallwords’ (Östman 1981; Blakemore/Gallai 2014). For instance, the most frequent functions of *well* include: to preface or mark disagreements or divergence and dissatisfaction, to request clarification, and elaboration (Hale 2004: 63).

A number of scholars, most notably Berk-Seligson (1990), Hale (2004), and Szczyrbak (2014), have drawn attention to the importance of discourse markers in the legal setting. More recently, other, equally important, areas of the legal setting have come under scrutiny, namely the police setting (Blakemore/Gallai 2014; Tipton/Furmanek 2016; Nakane 2014). As Nakane points out:

This is a sensitive and highly important aspect of legal interpreting, since difficulties in collecting evidence – such as lack of coherence, hesitations, and (un)willingness to provide relevant information – are themselves also part of the evidence, especially in cases in which two competing stories are being told. (*ibid.*: 80)

Such studies point to the need to include such aspects in the training of PSI and legal interpreters, and illustrate how corpus linguistics tools can be used profitably to create training material for interpreter students in the legal field. CL is also a valid instrument through which to study phraseological occurrences and patterns not only in written texts, but also in both spontaneous/conversational and formulaic/institutional/domain-specific oral language to train dialogue interpreters. A large body of research on phraseology (see for example, Ellis 2008; Meunier/Granger 2008) has clearly demonstrated that languages are based on regular patterns, associations, formulaic structures, etc. – in short, recurrent lexical and grammatical combinations that also facilitate second-language acquisition due to its formulaic and mnemonic features. This motivated me to search primarily for terms at the phraseological level.

Another point can be made regarding the usefulness of phraseology in a PSIT perspective: Colson (2008) draws a connection between the intrinsically for-



mulaic, phraseological nature of natural language and translation. During the process of translation, the formulaic aspect of language emerges powerfully, precisely because translation functions at the phraseological level (holographs, collocations, phrasal verbs, etc.) rather than at the semantic (individual word) level (Colson 2008: 199). Furthermore, research in the field of collocations and phraseology (Prodromou 2005) has shown that the speech of native speakers can be distinguished from that of non-native speakers by the presence of strings of language. Learning chunks of language, rather than isolated words, improves fluency because usually chunks occupy a single intonation unit: “Choral or private repetition, increasing the speed at each repetition [...] can be a useful way of drilling chunks so that they become imprinted in the memory as ‘musical’ items” (O’Keeffe *et al.* 2007: 77). Presentation of specialized chunks, as those extracted from my sub-corpus PaCad, can raise awareness of them through ‘noticing activities’. Furthermore, some studies in this field have demonstrated that a phraseology-based learning process “‘frees up’ the cognitive processing load so that mental effort can be allocated to other aspects of production such as discourse organization and successful interaction” (Girdard/Sionis 2004, in O’Keeffe *et al.* 2007).

## 4. Analysis and results

### 4.1. Learner Corpus

The scripted nature of the students’ dialogues clearly is not ideal for representational purposes, they do not exemplify a real-life situation, but they do illustrate discourse cohesion and facilitation. The stress factor in an exam-situation also matches, at least to some degree, the stress level of a real-life interaction. Although some discourse markers can be sacrificed without significantly affecting the propositional or pragmatic meaning of the utterance, the significance of DMs was underestimated by students who were not always alert to their pragmatic importance in a dialogic interaction or in the power relations that are sometimes played out in institutional dialogues. I believe that bringing greater awareness of the function of DMs and other pragmatic features to the students will improve their overall renditions.

Table 2 shows some of the most frequently recurring translations of ‘well’, the most frequent DM in the spoken Learner Corpus in both source and target texts. ‘Well’ occurred predictably mainly in initial position, but unlike the data described in the monolingual English corpus (see §4.2 below), not in a time-management or defensive capacity (i.e. in the defendant’s replies in the Police Interviews described in §4.2). In the table below the renditions of ‘well’ into Italian – it was translated in 6 different ways – have been listed according to the frequency of use by the students.

RENDITIONS IN ITALIAN  ENGLISH ST	It1	It2	It3	It4	It5	It6
well	Ø	<i>uhh</i>	<i>allora</i>	<i>sa</i>	<i>ok</i>	<i>niente</i>
very well	Ø	<i>uh, beh</i>				
yes, well						
well actually	<i>in realtà</i>					
well, you see	Ø					

Table 2. Renditions of ‘well’ in Italian translations in the Learner Corpus (Ø = zero rendition).

Vice versa, when translating from Italian into English, ‘well’ was used to render five different discourse markers: *bene, ma, però, adesso, ho capito*.

The dialogues were scripted by the author and other colleagues but were created to emulate a ‘real-life’ situation as closely as possible. Nevertheless, the simulated dialogues were guided by the trainers to test precisely these and other features. One of the easiest ways to guide and assist students is to decrease the pace and increase quality (articulation) of speech when it is clear that a student is having difficulties with the rendition. Slowing down and articulating more clearly significantly increases the ability of the student to maintain ‘extra-propositional’ features such as pragmatic markers. By the same token, when increasing the pace of speech and decreasing the phonetic quality, including the discourse markers, a drop in the maintenance of discourse markers and an increased focus and attention to the propositional content were observed. Only the best students were able to maintain both with a high degree of accuracy in their rendition. This is an excellent training tool to calibrate the level of difficulty of a dialogue to the student’s individual needs.

The following are the main functions for the DM ‘well’ that emerged in the source texts in the Learner Corpus in the Identification as well as in the Verification Phase:

1. an *acknowledgment* token, a *continuer*, above all when followed by ‘yeah’;
2. opens the next turn thus performing a *time-managing* function;
3. *changing* a topic;
4. *avoiding* a topic;
5. *playing for time*;
6. marker of *self-repair* to try to make utterances *clearer* (paraphrase);
7. *requests clarification* and/or *elaboration* (intonation-dependent);
8. prefaces or marks *approval* or *agreement* with the interviewer;
9. prefaces or marks *disapproval*, *disagreement*, dissatisfaction or divergence with the interviewer (intonation dependent).

I suggest that the DM ‘well’ can be seen on a continuum from primarily Relational / Interactional, to Organizational, and primarily Propositional.



Figure 1. Continuum of the functions of the DM ‘well’.

In the Learner Corpus the students were under pressure to perform in the exam, rather than being in a real-life interpreting situation and their focus was primarily on two aspects: remember each utterance (and not be marked down for omissions) and translate as accurately as possible primarily at the propositional level; translating pragmatically adds a higher stress factor. Thus, the didactic translation function was uppermost in their minds, rather than the interactional function (especially given that the other speakers were the trainer(s) and were perfectly able to understand them despite any omissions; in a real-life situation their management of DMs might have been slightly different. When DMs were jettisoned for reasons of time, memory or focus/concentration, students did tend to compensate for this by signaling cohesion (causal-logical and textual-syntactical) through other discourse elements or through intonation.

In general, I observed that the closer a discourse marker was to the propositional end of the scale, as an explicit acknowledgement, the more students would incorporate it, translating it with an affirmative or negative, for example. Again, intonation would be used to emphasize a strong affirmative or negative function, or intonation would be ‘flattened’, the more redundant the DM was. Intonation was crucial in distinguishing agreement from disagreement, but strong (dis)agreement would usually have a propositional content-based follow-up in the following sentence/utterance that would make the DM ‘well’ redundant.

In turns of the functions of time-management (function no. 2) and topic-avoidance/evasion (function no. 4) ‘well’ was often jettisoned, because the pace of the situation had already changed by the time the turn came to the interpreter (student) and rendered it redundant. As shown in §4.2, in the Police Interview data, these two functions appeared more frequently, probably because of the tense and conflictual nature of the setting (police interviews, the police officer’s more aggressive questioning and defendant’s more defensive and strategic responses, also logistical/organizational). For self-repair motives (function no. 6), the DM would be jettisoned because the weight of a successful rendition (the paraphrasing of the previously unclear utterance) took precedence over signaling it (i.e. signaling that it is unclear), whilst as a signal of a topic shift (function no. 3) there would generally be some indication to substitute this even if with just a pause, a gaze or change of body position or DM like ‘ecco’ or ‘allora’ in a very weak form.

When the trainer was using ‘well’ to simulate that she was thinking out loud and delaying the beginning of the utterance (function no. 5), ‘well’ was, under-

standably, rendered void – a zero rendition. The pragmatic effect of rendering void this discourse marker was to make the utterance and the trainer's attitude seem more definite and less hesitant. Whether or not this impacts on the overall effect of the utterance into the TL is, however, debatable. Nevertheless, students should also be made aware of the subtler pragmatic functions of DMs. Also, when the student asked for repetition and the trainer repeated an utterance, DMs were dropped, unless they were propositionally or pragmatically significant. When the student was playing for time in Italian, the use of the DM *'allora'*, articulated in a long drawn-out fashion, was often resorted to. 'Ok' and 'so' were used frequently in English both at the more propositional end of the continuum but also as gap fillers and the playing-for-time function.

I observed that students and trainees generally aimed to maintain discourse coherence and would attempt to assess the significance of DMs on the spot, but were not always successful. Omission was one frequent option, or substitution for interjections such as *'uhh'*, or *'beh'*. Sometimes a non-lexical empty DM (*'beh'*, *'uh'*) would signal a forward-looking glance at the next words in the utterance if the words were challenging at the propositional level – i.e. the students were already 'attacking' and focused on the next (difficult) part of the utterance and recalling specific and non-specific memory in order to solve a task two or three steps further down the sentence. In those moments (seconds) of activating short-term memory, 'empty' DMs would sometimes, but not always, efficiently and often successfully function as pause-filler to cover up while the students were trying to recall the first part of the utterance (function no. 5, playing for time).

Where the DM *'well'* had simply an interactional and weaker propositional or pragmatic function students seemed to be more relaxed about accuracy (a correct judgment in my view) and used Italian renditions such as *'ma'*, *'sì'*, *'allora'*, *'bene'*. DMs in scripts that were as simple as my corpus was (predominantly main clauses rather than subordinate clauses) had less of a textual cohesive function; forward looking text-cohesive discourse markers are arguably less frequent in spontaneous conversation and difficult to manage for interpreter trainees.

The few times DMs were used to organize discourse in terms of distinguishing between speakers (especially with more than 3 interlocutors) (*'ecco'*, *'così'*), they were rarely jettisoned, and also accompanied by spatial-organizational body language and signs – as such it was the entire discourse event rather than just the spoken utterances that required 'organization' at the cohesive level.

Clearly, at a micro-level, the significance of DMs can be crucial in the legal setting, as Hale (1999), Blakemore/Gallai (2014), Szczyrbak (2014), and others have shown because they can skew not just the development of the interview or conversation, but even the propositional content as well as the perception of a truth condition.

#### 4.2 MoPICO (monolingual spoken corpus): the case of 'well' (Verification Phase)

The aim of the last two sub-sections is to show that both spoken and written corpora can be used to investigate specific translation problems that arise in the translation of legal documentation at various levels of language use and,

more precisely, to emphasise the link between them and interpreter training. This sub-section, in particular, looks at ‘well’ in the transcribed video recordings of four police interviews (two conducted in the USA and the other two in the UK) dealing with murder-related criminal cases. Seen as a sub-genre of legal discourse, the broad purpose of the police interview (PI)<sup>8</sup> is both to elicit information and to establish whether or not the allegation being made is true or credible. The interview is also a *product*, as Nakane notes (2014: 8), in that it is used as evidence if the case goes to trial. Both process and product are characterized by being a synthesis of two competing narratives, the police officer’s and the suspect’s, at least when the suspect denies the allegation. If the *purpose of police questioning* can be classified as “*elicitation of information*” and “*confirmation of a particular version of events*” (*ibid.*: 33), interviewers will direct their discourse through strategies and ‘interactional resources’ that achieve these broader aims. These strategies are primarily control of topic shifts and of turn-taking as well as question types (information-seeking or confirmation questions). Against this backdrop, DMs are used strategically to perform the above functions.

In MoPiCo, *well* (4%)<sup>9</sup> comes after the most frequent markers *you know* (6%) and *yeah* (6.7%) and is followed by other discourse devices such as *I mean* (1.1%) and *actually* (1.01%). Syntactically speaking, in my sub-corpus, *well* occupies its proto-typical initial position (Urgelles-Coll 2010: 23), usually after the officer’s question. Thus, compared with the interviewee’s talk, the incidence of *well* in the interviewer is more marginal. Among the functions reported in the literature (Sidnell 2010), in MoPiCo *well*, as used by the interviewee, mainly performs the task of a continuer as in the following examples:

*Example 1*

OFFICER: And then what did he do to her?

INTERVIEWEE: **Well** after he was done, that’s when he put her back in the jeep...

This “acknowledgment token” function may be easily explained by the fact that most of the interview is made up by the interviewee’s narrative where the use of this marker is primarily a way to accept what had been previously stated and

8 Although this genre – police interviews – has not been studied nearly as much as court interpreting (also because access to empirical data is a severe limitation), it is a crucial phase in the legal process. The importance of DMs in this semi-spontaneous discourse in the police interview (a spontaneous narrative of the event being described and investigated in the framework of a standardized Q/A format) is significant, and may channel the course of the investigations in one direction or another, especially when mediated through interpreting. Accurate interpreting of police interviews is crucial because they are fundamental in potential future court proceedings. Accurate translation/interpreting safeguards the procedural value of the interview as evidence in subsequent phases (see Nakane 2014; Pöchhacker/Kolb 2009). Because the interview is written down and kept as a written police report, all procedural aspects must be maintained, but may also lead to a stylistic ‘mismatch’ between the oral and written versions describing the same episode.

9 The percentages indicated refer to the relative frequency of the DMs in the Monolingual Corpus.

to move on with the narration. In other cases, even though less frequently, the interviewee resorts to *well* as a marker of disapproval with what the officer has said as shown by example 2:

*Example 2*

OFFICER: And you were dropped off, it's such an event that someone's standing in your field taking a picture of that van, that you remember that too don't you? The bus driver remembers it, the kids on the school bus remember it. The girl taking pictures, you remember that?

INTERVIEWEE: **Well** I wasn't lookin' at the...

OFFICER: Huh

INTERVIEWEE: **Well** sometimes I'm talkin' to Blaine.

Conversely, when used by the interviewer, the marker takes on a time-managing function which is a device to control discourse:

*Example 3*

OFFICER: You give us permission to go in your house and get the jeans?

INTERVIEWEE: Yeah

OFFICER: Ok. **Well** I'm just gonna make a phone call quick...

As reported in the previous section, in my Learner Corpus, numerous instances of  $\emptyset$  (non-rendition) for DMs were found, but it was also noticed that generally speaking the students were able to judge the importance of a specific DM (as they had been trained to do so in class) and non-render or substitute it with propositional (or non-verbal) content. CL can thus be used to raise trainees' awareness of the importance of DMs, but also of what other students and practitioners tend to do in similar situations (substitutions). Also, to show them that in a real-life situation they are not 'helped' by a familiar trainer who will be reading the scripted dialogue<sup>10</sup>, but they will have a real service provider who depends completely on them for successful communication. If used judiciously in a classroom setting and guided by experienced trainers, the didactic usefulness emerges clearly when using the Learner Corpus and PI data to direct and guide students in their classroom work with peers as well as in the assessment.

#### 4.3. PaCad (parallel written corpus): administrative legal texts and their phraseology

The collection and translation of recurrent and standard documentation (notices, forms, reports, summons, etc.) used by the police forces and the courts is one way to facilitate the work of legal interpreters. Frequently, the translated versions of these administrative texts serve as documentary evidence in subsequent court

10 Although we are mindful of the fact that scripted dialogues are not ideal, the sheer size of my MA classes (often 100+) and having only one trainer, makes it difficult to use more lifelike training and assessment alternatives.

proceedings and their importance should not be underestimated (see Nakane 2014). Administrative texts form a part of our daily lives, but are not ‘trivial’ as they reflect and construct our attitudes towards the worlds which surround us. As well as familiarity with legal formulae, as Pontrandolfo notes, translations of this typology also require familiarity with the “genre structures through which legal institutions conduct their affairs” (Pontrandolfo 2015: 138). The documentation collected is the one used most frequently by the police and courts in their dealings with people involved in criminal procedures, be they suspect, defendant, witness or victim. The communicative function of such texts is arguably descriptive and procedural rather than informative, and thus phraseology will tend to be predominant and of a very fixed nature, more so than an informative typology<sup>11</sup>. Administrative documents are of different types and usually initiate a legal procedure. Their structure depends on the authorities involved which vary across cultures and they are also a reflection of the institutional role they play (Charrow 1982). It goes without saying that these cultural differences bring about multiple translation problems.

The phraseological constructions which surfaced from the parallel corpus may be considered as underlying structural features of the subsequent or parallel phases in the legal/judiciary process, from police interviews to courtroom discourse. What is argued here is that a hybrid discourse form emerges with a combination of fixed phraseology (of a written nature coming from administrative documents) and pragmatic discourse features pertaining to oral discourse. Indeed, the subgenres of police interviews as well as courtroom discourse contain both written phraseological elements in the administrative and ‘ritual’ language of legal procedures (fixed terminology as well as fixed form – in particular the Q/A format of interviewing and cross-examination) and the dialogic features of spoken interaction.

In my analysis I have adopted the phraseological approach which is described in the literature as being more empirically-based, rather than a lexicographical approach, despite the fact that legal phraseology, at least in Italy, has received scant attention in the literature. Even though the formulaic nature of legal language has been highlighted by many scholars (Crystal/Davy 1969), it is only with the advent of computerized corpora techniques that the landscape of contemporary legal phraseology is beginning to change (Pontrandolfo 2015). In order to investigate the phraseology/fixed collocations of my parallel texts two pieces of software were used: AntConc, which allows the extraction of multiword units that are potential candidates for being considered phraseologies, and AntPConc, which aligns parallel texts automatically and displays the source text (ST) phrases aligned with the target text (TT) translation equivalents.

The investigation started with the extraction of the multi-word units in Italian with minimum 3 units as a baseline parameter to search for phraseological syntagms. I used the n-grams/cluster utility of the software AntConc. All those

11 The notion of phraseology used in this paper refers to the co-occurrence of more than one lexical item with other linguistic elements and which functions as one semantic unit (see Gries 2008: 6).



instances which were not candidate to be phrases (e.g. *art cp* (3), *persona offesa che* (5), *non conosce la* (35)) were disregarded.

Secondly, I looked at all the cases of what I deemed to be translational correspondence and non-correspondence and I classified both cases of phrases according to the following categories:

- a) legal system-bound technical collocations/phrases (e.g. *l'incidente probatorio*/special evidentiary hearing)
- b) semi-technical collocations (e.g. *la persona offesa*/injured party);
- c) shared general language phrases (e.g. *forze dell'ordine*/law enforcement authorities; *i presidi sanitari*/health care establishments).

The first group of fixed expressions, the category which causes difficulties of the cultural type, includes phrases that might refer either to juridical organs (e.g. *la Polizia giudiziaria*) or to procedures (e.g. *l'incidente probatorio*); they all have a precise legal meaning in Italy. As far as *la Polizia Giudiziaria* is concerned, Italy has various law enforcement agencies, each with a different status and structure. The most important law enforcement agencies are the State Police (*Polizia di Stato*), the *Arma dei Carabinieri* and the Customs and Excise Police (*Guardia di finanza*). Any member of any law enforcement agency can be given the task of carrying out investigations into a criminal offence by the public prosecutor and, in this case, they come under the umbrella term of *la polizia giudiziaria*, a term used to indicate a function rather than a specific law enforcement agency. There are particularly close ties between the judicial police and the public prosecutor, with special judicial police sections in all of the Public Prosecutor's Offices. The work of the judicial police starts from the beginning of investigations, when the authorities are first aware of the possibility of a criminal offence. It goes without saying that translators are involved in those communicative acts which include a variety of agents, from authorities to receivers; they are involved in and constitutive of an essential socio-juridical and cultural process that is fundamental for the safeguarding of justice and non-Italian speaking citizens' basic civil rights. During the simulations with interpreter trainees, these terminological items were sometimes either simplified (i.e. using 'police' as a substitute for the other institutions, or disambiguated (i.e. *Arma dei Carabinieri*: the institution that carries out preliminary investigations).

Table 3 shows phraseological units retrieved from the parallel written corpus. What is argued here is that, during their training, interpreters should be provided with these system-bound collocations to be used in their simulations in the classroom and, consequently, in their professional performances. This type of training would accommodate the acquisition of those terminological collocations that were problematic in the Identification Phase and that play an important cultural role in specialized communication across languages and cultures.

Type of phrase	Parallel Corpus	
	Italian	English
a) Legal system-bound technical collocations / phrases	<ol style="list-style-type: none"> <li>1. <i>Ufficio di Polizia Giudiziaria</i></li> <li>2. <i>presso la Procura della Repubblica competente</i></li> <li>3. <i>Ufficio dei Carabinieri</i></li> <li>4. <i>registro delle notizie di reato</i></li> <li>5. <i>Giudice di Pace</i></li> <li>6. <i>richiedere un incidente probatorio</i></li> <li>7. <i>Questore</i></li> </ol>	<ol style="list-style-type: none"> <li>1. Office of the Judicial Police</li> <li>2. at the competent Public Prosecutor's Office</li> <li>3. Carabinieri Office</li> <li>4. register of <i>notitiae criminis</i></li> <li>5. <i>Giudice di Pace</i> [Justice of the Peace]</li> <li>6. request a special evidentiary hearing</li> <li>7. <i>Questore</i> [Provincial Police Chief]</li> </ol>
b) Semi-technical collocations	<ol style="list-style-type: none"> <li>1. <i>infondatezza della notizia di reato</i></li> <li>2. <i>articoli (numero) del cpp</i></li> <li>3. <i>richiesta di archiviazione</i></li> <li>4. <i>richiedere il patrocinio gratuito</i></li> <li>5. <i>cessazione delle misure cautelari</i></li> <li>6. <i>in assenza d'apposita richiesta</i></li> <li>7. <i>remissione di querela</i></li> <li>8. <i>per i reati perseguibili a querela di parte</i></li> <li>9. <i>impugnare l'ordinanza del giudice</i></li> <li>10. <i>a mezzo del procuratore speciale</i></li> <li>11. <i>sentenza di non luogo a procedere</i></li> </ol>	<ol style="list-style-type: none"> <li>1. groundlessness of the <i>notitia criminis</i></li> <li>2. articles (number) of the Italian Code of Criminal Procedure</li> <li>3. request to the judge to drop the case</li> <li>4. request of legal aid</li> <li>5. termination of the protective measures</li> <li>6. without explicit request</li> <li>7. withdrawal of complaint</li> <li>8. in case of a criminal offence persecuted with a complaint of the victim</li> <li>9. challenge the judge's order</li> <li>10. specially appointed representative</li> <li>11. judgement of no grounds to proceed</li> </ol>
c) Shared general language phrases	<ol style="list-style-type: none"> <li>1. <i>con il presente avviso</i></li> <li>2. <i>traduzione gratuita degli atti</i></li> <li>3. <i>permesso di soggiorno</i></li> <li>4. <i>crimini previsti dalla legge</i></li> </ol>	<ol style="list-style-type: none"> <li>1. with this notice</li> <li>2. free of charge translation of all documents</li> <li>3. residence permit</li> <li>4. crimes specifically referred to in the law</li> </ol>

Table 3. Examples from the Parallel Italian-English Corpus.

## 5. Discussion and final remarks

The primary aim of the present study was to look at corpus linguistics techniques for PSIT pedagogical applications and professional practice, to gain a clearer insight into those stumbling blocks that trainee interpreters tend to encounter, and to appreciate how that knowledge could be used to tackle lexical, pragmatic or interactional challenges.

First, I assembled a miscellaneous legal corpus containing different types of corpora: a bilingual (Italian/English) Learner Corpus of simulated interactions, a monolingual spoken corpus of legal English (MonoLegalII), and a written parallel corpus of legal Italian texts and their English translations (ParaLegalII). These language resources were exploited in the different methodological steps, developed under the labels of Identification, Verification and Activation phases.

In the first step of the analysis, the Learner Corpus was used to analyse learners' obstacles both from a lexical and/or pragmatic perspective and from the point of view of translational transfer in a dialogic interpreting situation (PSI). Before even beginning the analysis, I had a robust vision of what would emerge because I had done the data recording and I was also directly involved as one of the trainers, examiners and data researcher; moreover, my experience in training as English language instructor and as language mediation instructor had already given me a solid knowledge basis of trainees' errors in this sector. Then, the main advantage of using a Learner Corpus is that data can be supplemented and replicated for further research. Furthermore, computer-aided analysis of my corpora through wordlisting and concordances allowed a more systematic analysis of DMs and their translations. Evidence from the Learner Corpus showed regularities in students' deviations from the standard norms. These standard norms were then showed in the Verification Phase through the use of the monolingual corpus MoPiCo. In the third phase, I searched for ways in which I could activate (Activation Phase) what I had unearthed so far, render the data and findings available (e.g. technical phraseology), and model them in such a way as to be useful for interpreter trainees, as a concrete output.

This three-pronged process/approach, that is Identification, Verification and Activation can be broadly used as a model for PSI trainers. The regular access to legal corpora, that is large sets of authentic data, might improve the training of mediators/interpreters with more attention to aspects such as simulation, identification of learners' stumbling blocks, and then, verification against evidence from the corpus. In other words, interpreting output from simulations lends itself to be collected as quantitative representation of students' non-conformity to frequent patterns of usage in a language. This non-conformity becomes more visible when verified against authentic materials from monolingual and parallel corpora in the same area. An example was given by the use of the monolingual corpus of spoken Police Interviews to show the regular behaviour of DMs (e.g. well) in naturally occurring language and the difference between the students' use and the corpus evidence.

Corpus-based Translation approaches with the compilation of a parallel corpus can also be used to show specialized phraseologies which are usually found in spoken legal discourse. Due to the discussed limitations that this type of studies entails, the compilation of corpora of the monolingual type, even though in the written modality, may be a valuable instrument for the training of legal (dialogue) interpreters in that they provide students with pre-packaged or formulaic expressions that are part and parcel of the specialized language and are also used orally.

Another didactic application that will come in useful is to compile specialized phraseological glossaries that provide Public Service Interpreting with technical culture-bound legal phrases, above all when working with minority languages. Greater familiarity with the relevant legal phraseology in Italian documents, for example, should be a fundamental part of interpreter training. Voice-recorded police interviews or courtroom examinations provide examples of set, recurrent legal phraseology, but also of important pragmatic features of spoken discourse, especially in the Q/A format where questions are often imbued with an asym-

metric power relationship through the use of subtly coercive strategies (especially tag questions and silence). The immediate nature of oral interpreting does not (unlike translation) permit the perusal of dictionaries, although a personal glossary, possibly in electronic format on a tablet computer, is feasible. Learning such phraseology is enormously useful, indeed essential. Public Service Translation is, therefore, a valuable linguistic support for legal institutions that could ultimately reduce the need for interpreting. If the non Italian-speaking public is provided with information, material and administrative documents in languages they understand and can access easily, this will reduce contact with institutions.

If used properly – both as a tool to help interpreter trainees learn and memorize phraseological correspondences, and as a reference material to use on the job – multilingual corpora can save PSI trainees a great deal of work. Lastly, although it is not the focus of this paper, the use of multilingual specialized corpora processed through CL is invaluable in the time-consuming and expensive process of written translation in PS and in the public sector generally. That is because multilingual documentation and information reduce the need for person-mediated information.

Generally speaking, the main aim of the Activation Phase is in line with pedagogic approaches to the study of a foreign language (see Seidlhofer 2002, based on Swain 1985), in that the use of corpus-based output increases linguistic knowledge and contributes to accuracy. As such, students are helped to notice the discrepancy between what they want to say and what they are linguistically able to say (Seidlhofer 2002: 218). The collection and analysis of authentic data serve the purpose of identifying those areas – at the linguistic, pragmatic and institutional and socio-cultural level – that lead to miscommunication, misunderstandings, damaging power imbalances, and thus thwart communication and the resulting access to services.

## Acknowledgments

The author would like to thank the editors and Mette Rudvin for providing insights throughout the paper. Any shortcomings and all opinions are entirely my own.

## References

- AntConc (s.d.) <<http://www.laurencethony.net/software.html>>.
- Berk-Seligson S. (1990) *The Bilingual Courtroom. Court Interpreters in the Judicial Process*, Chicago, University of Chicago Press.
- Biel L. (2010) “Corpus-based studies of legal language for translation purposes: methodological and practical potential”, in C. Heine / J. Engberg (eds) *Reconceptualizing LSP. Online Proceedings of the XVII European LSP Symposium 2009*, Aarhus, Aarhus School of Business, Aarhus University.
- Blakemore D. / Gallai F. (2014) “Discourse markers in free indirect style and interpreting”, *Journal of Pragmatics* 60, 106-120.

- Cencini M. (2002) "On the importance of an encoding standard for corpus-based interpreting studies. Extending the TEI scheme", *CULT2K Special Issue of InTRAlinea*, <<http://www.intralinea.org/specials/article/1678>>.
- Charrow V.R. (1982) "Language in the bureaucracy", in R. di Pietro (ed.) *Linguistics and the Professions: Proceedings of the Second Annual Delaware Symposium on Language Studies*, Norwood NJ, Ablex, 173-188.
- Colson J.P. (2008) "Cross-linguistic phraseological studies", in S. Granger / F. Meunier (eds) *Phraseology in Language Learning and Teaching*, Amsterdam/Philadelphia, John Benjamins, 191-206.
- Crystal D. / Davy D. (1969) *Investigating English Style*, Harlow, Longman.
- Ellis N. (2008) *Phraseology in Foreign Language Learning and Teaching*, Amsterdam/Philadelphia, John Benjamins.
- Gries S.Th. (2008) "Phraseology and linguistic theory: a brief survey", in S. Granger / F. Meunier (eds) *Phraseology: An Interdisciplinary Perspective*, Amsterdam/Philadelphia, John Benjamins, 3-25.
- Hale S. (1999) "Interpreters' treatment of discourse markers in courtroom questions", *International Journal of Speech Language and the Law* 6/1, 57-82.
- Hale S. (2004) *The Discourse of Court Interpreting. Discourse Practices of the Law, the Witness and the Interpreter*, Amsterdam/Philadelphia, John Benjamins.
- Lázaro Gutiérrez R. / Sánchez Ramos M. del Mar (2015) "Corpus-Based Interpreting Studies and Public Service Interpreting and Translation training programs: the case of interpreters working in gender violence contexts", in J. Romero-Trillo (ed.) *Yearbook of Corpus Linguistics and Pragmatics 2015. Current Approaches to Discourse and Translation Studies*, Cham CH, Springer, 275-292.
- Meunier F. / Granger S. (eds) (2008) *Phraseology in Foreign Language Learning and Teaching*, Amsterdam/Philadelphia, John Benjamins.
- Nakane I. (2014) *Interpreter-Mediated Police Interviews. A Discourse-Pragmatic Approach*, London, Palgrave Macmillan.
- O' Keffe A. / McCarthy M. / Carter R. (2007) *From Corpus to Classroom. Language Use and Language Teaching*, Cambridge, Cambridge University Press.
- Orozco-Jutorán M. (2017) "Anotación textual de un corpus multilingüe de interpretación judicial a partir de grabaciones de procesos penales reales", *Revista de Llengua i Dret. Journal of Language and Law* 68, 33-56.
- Orozco-Jutorán M. (forthcoming) "The TIPp project: developing technological resources based on the exploitation of oral corpora to improve court interpreting", in C. Bendazzoli / M. Russo / B. Defrancq (eds) *Corpus-based Interpreting Studies: The State-of-the-Art*, Special issue of *InTRAlinea* 20.
- Östman J.O. (1981) *'You Know': A Discourse Functional Approach*, Amsterdam/Philadelphia, John Benjamins.
- Pöschhacker F. / Kolb W. (2009) "Interpreting for the record. A case study of asylum review hearings", in S. Hale / U. Ozolins / L. Stern (eds) *The Critical Link 5: Quality in Interpreting – A Shared Responsibility*, Amsterdam/Philadelphia, John Benjamins, 119-134.
- Pontrandolfo G. (2015) "Investigating judicial phraseology with COSPE: a contrastive corpus-based study", in C. Fantinuoli / F. Zanettin (eds) *New*

- Directions in Corpus-based Translation Studies, Berlin, Language Science Press, 137-160.
- Prodromou L. (2005) 'You see, it's sort of tricky for the L2-user': the puzzle of Idiomaticity in English as a Lingua Franca, PhD thesis, University of Nottingham.
- Robinson P. (1996) "Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions", *Studies in Second Language Acquisition* 18, 27-68.
- Rudvin M. / Spinzi C. (2013) *Mediazione linguistica e interpretariato. Regolamentazione, problematiche presenti e prospettive future in ambito giuridico*, Bologna, Clueb.
- Schiffrin D. (1987) *Discourse Markers*, Cambridge, Cambridge University Press.
- Seidlhofer B. (2002) "The shape of things to come? Some basic questions about English as a lingua franca", in K. Knapp / C. Meierkord (eds) *Lingua Franca Communication*, Frankfurt am Main, Peter Lang, 269-302.
- Sidnell J. (2010) *Conversation Analysis. An Introduction*, London, Blackwell.
- Sperberg-McQueen C.M. / Burnard L. (eds) (2007) *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, <<http://www.tei-c.org/Guidelines/P5/>> (last accessed 20 December 2016).
- Swain M. (1985) "Communicative competence: some roles of comprehensible input and comprehensible output in its development", in S. Gass / C. Madden (eds) *Input in Second Language Acquisition*, Rowley MA, Newbury House, 235-253.
- Szczyrbak M. (2014) "Pragmatic marker use in police interviews: the case of I mean and you know (Part 1)", *Studia Linguistica Universitatis Iagellonica e Cracoviensis* 131, 287-297.
- Tipton R. / Furmanek O. (2016) *Dialogue Interpreting. A Guide to Interpreting in Public Services and the Community*, London, Routledge.
- Urgelles-Coll M. (2010) *The Syntax and Semantics of Discourse Markers*, London, Continuum.
- Valero Garcés C. (2014) *Communicating across Cultures. A Coursebook on Interpreting and Translating in Public Services and Institutions*, Lanham, Maryland/ Plymouth, UK, University Press of America, (Traducción de A. Tobin).
- Valero Garcés C. / Francisco R. (eds) (2012) *Retos del siglo XXI en comunicación intercultural*, Saarbrücken, Editorial Académica Española, LAP LAMBERT Academic Publishing.
- Vargas Urpi M. (2011) "Traducir para la comunidad china en Catalunya. El ejemplo de los materiales de acogida disponibles online", *Journal of Cultural Mediation* 1, 24-35.