

Para citar este artículo:

Orozco-Jutorán, M. (2018). "The TIPp project: developing technological resources based on the exploitation of oral corpora to improve court interpreting", en *Intralinea*, 20. Special Issue: New Findings in Corpus-based Interpreting Studies

# The TIPp project

## Developing technological resources based on the exploitation of oral corpora to improve court interpreting

By Mariana Orozco-Jutorán (Universitat Autònoma de Barcelona, Spain)

### Abstract & Keywords

In Spain, new laws have been passed that significantly reinforce procedural guarantees in criminal proceedings, as they provide regulation on the right to translation and interpreting in criminal proceedings as well as on the right to information of an accused person in relation to the subject of the criminal proceedings, so that they can exercise efficiently their right to self-defence. Translation and interpreting thus become an essential element in the right to effective legal protection in the exercise of lawful rights and interests before the courts in order to avoid any state of defencelessness.

In the light of this new situation, the research group MIRAS, of the Universitat Autònoma de Barcelona, launched a research project called TIPp (Translation and Interpreting in Criminal Proceedings) aimed at describing the reality of court interpreting and at creating a computer application which comprises all the necessary resources to facilitate court interpreters' performance. These include recommendations for court interpreters and courtroom personnel on interpreters' role and on how to interact with interpreters, monolingual glossaries in Spanish for different contexts, such as certain type of crimes or 'general vocabulary' for criminal trials and a pilot sample of five databases – one in each language combination (English, French, Romanian, Arabic and Chinese from and into Spanish) – containing the problematic units most frequently encountered by court interpreters, as observed in the TIPp corpus.

This article explains the design and the methodology used to compile and exploit the corpus, which will be made publicly available, as well as some of the results from the outcomes of this project.

**Keywords:** court interpreting, corpora, quality, detainee rights, ICT resources

©inTRAlinea & Mariana Orozco-Jutorán (2018).

"The TIPp project Developing technological resources based on the exploitation of oral corpora to improve court interpreting"

*inTRAlinea* Special Issue: New Findings in Corpus-based Interpreting Studies

Edited by: Claudio Bendazzoli, Mariachiara Russo & Bart Defrancq

This article can be freely reproduced under [Creative Commons License](#).

**Stable URL:** <http://www.intralinea.org/specials/article/2316>



## Introduction

In Spain, court interpreting has been an under-researched area until recently. Academic contributions started barely a decade ago (Ortega Herráez 2006; Del Pozo Triviño et al. 2014; Onos 2014) and the descriptions of the current situation of court interpreting in Spain are not based on authentic, representative data. In the last two decades, however, research in court interpreting has emerged as a major topic in Europe. In fact, within the Horizon 2020 Programme, the Directorate-General for Justice of the European Commission, through the Justice Programme 2014-2020, is offering grants to undertake research in this area.

As a result of the transposition of two European Directives<sup>[1]</sup>, a new law was passed by the Spanish Parliament in April 2015 (*Ley Orgánica 5/2015, de 27 de abril*) amending Spain's Code of Criminal Procedure. As stated therein, this new legislation 'significantly reinforces procedural guarantees in criminal proceedings, as it provides regulation on the right to translation and interpreting in criminal proceedings as well as on the right to information of an accused person in relation to the subject of the criminal proceedings so that they can exercise efficiently their right to self-defence'<sup>[2]</sup>. Translation and interpreting thus become an essential element in the right to effective legal protection in the exercise of lawful rights and interests before the courts in order to avoid any state of defencelessness. This law is referring to the right to be informed of the accusation against a subject and the right to a public process with all procedural guarantees, as enshrined in Section 24 of the Spanish Constitution.

The research group MIRAS, of the Universitat Autònoma de Barcelona, is specialised in Public Service Interpreting, and the research projects previously undertaken by this group in Barcelona, Spain (for instance, see Onos 2014) revealed that court interpreters currently lack the required technological and research resources to carry out their tasks with accuracy, rigour and diligence. Furthermore, a review of the current literature, that is detailed in section 1, shows the absence of a description of reality using sufficiently representative data, that is, there are many assumptions and hypotheses being made about court interpreting

but there is a lack of authentic and representative data to know what is actually happening.

### **1. The TIPp project**

Given these needs, the research group MIRAS decided to launch a research project called TIPp[3] (Translation and Interpreting in Criminal Proceedings) aimed at compiling and analysing a representative oral corpus of trials in order to be able to describe the reality of court interpreting and at creating a computer application which comprises all the necessary resources to facilitate court interpreters' performance.

The researchers involved in the project had prior experience of projects using oral corpora from public service interpreting (see Arumí *et al.* 2011, 2012; Vargas-Urpi and Arumí Ribas 2014; Vargas-Urpi 2012) but TIPp is the first project based on collection and use of authentic oral corpora in court interpreting settings. There are four features of the TIPp project that make it unique.

The first is the novelty of being able to access real, video-recorded criminal proceedings. This is a breakthrough for court interpreting research, and has required a great deal of effort because, as Angermeyer, Meyer and Schmidt (2012: 276) point out:

Permissions for tape-recording sensitive data from medical or juridical communication can usually be obtained only after long, strenuous negotiations with the respective institutional bodies, and it surely can be assumed that many research projects have turned out to be not feasible simply because of bureaucratic hindrances.

The second feature is the size and representativeness of the oral corpus. The literature available (Berk-Seligson 1987, 1988, 1989 and 1999; Cooke 1995; Goldflam 1995; Hale 1997a, 1997b, 1997c, 1999, 2002 and 2008; Kadric 1999; Lane, McKenzie-Bridle and Curtis 1999; Mikkelson 1998; Montalvo 2001; Morris 1999; Nicholson and Martinsen 1997; Niska 1995; Ortega 2006 and 2011; Rigney 1999; Stern 1995, to name a few) shows that the studies on court interpreting based on oral corpora conducted so far have yielded very interesting and insightful data, such as the role usually played by the interpreter in a courtroom, that goes from mere conductor to “assistant” of the courtroom personal or even “mediator”. However, they have mostly been based on corpora that are either simulated – and thus cannot be claimed to describe reality – or relatively small – and thus cannot be used to extrapolate results or claim significance from the point of view of research methodology.

There are only two known exceptions to this, the first is a study conducted by Berk-Seligson (1990/2002), where the author investigated how Spanish-English interpreters faced the challenges of legal discourse in 114 hours of interaction in US courtrooms, highlighting the influence on the receivers' perceptions of the way in which people spoke and were interpreted. The second exception is Angermeyer's study (Angermeyer 2015: 6), where the researcher observed over 200 court proceedings and tape-recorded 60 hearings and transcribed them. The main difference between this study and TIPp is that Angermeyer observed small claims courts, so the cases studied were mostly arbitration hearings, whereas the TIPp corpus is based on criminal proceedings in criminal courts.

Since one of the TIPp project's declared aims is to describe reality using representative data, researchers chose to create and exploit a significant, representative corpus of real criminal proceedings that had very recently taken place. TIPp has accessed the video-recordings of criminal trials where interpreting took place in almost half of the criminal courts in Barcelona from 2010 to 2015. The corpus is described in depth below, in section 3.

Due to the importance and the difficulty of having access to a representative oral corpus of real criminal proceedings, the corpus compiled, transcribed and annotated will be made available for researchers so that it can be used in the future.

The third feature of the project consists of the systems used for the transcription and annotation of the corpus. In order to obtain quantifiable data to be able to describe reality in a systematic rather than an anecdotic way, the research team chose to use one tool that not only facilitated the transcription and annotation of the corpus but also allowed the creation of ad hoc categories for the annotations. This all-inclusive tool is a software package called EXMARaLDA, a system for the computer-assisted creation and analysis of spoken language corpora[4]. This tool enables the user to compile and manage a corpus, transcribe videos and, most importantly, it facilitates the type of ad hoc annotation created as well as its conversion into quantifiable data. Details of the transcription and annotation are explained below, in sections 4 and 5.

Finally, the fourth feature is the number of resources created. As well as describing reality, the project aims to provide support for users of interpreting in court settings by creating a computer application which includes resources to improve court interpreters' performance. These resources include (i) a set of recommendations for court interpreters, (ii) a set of recommendations for courtroom personnel regarding the role of the interpreter and how to interact with

interpreters, (iii) monolingual glossaries in Spanish for different contexts, such as certain type of crimes or “general vocabulary” for criminal trials, for instance and (iv) a pilot sample of five databases -one in each language combination (English, French, Romanian, Arabic and Chinese from and into Spanish)- containing the problematic units most frequently encountered by court interpreters, as observed in the TIPp corpus. This freely accessible resource is described below, in section 6.

## **2. Corpus compilation**

After a long process of interaction with the judicial institutions involved, researchers were able to request access to the video-recordings of criminal trials where interpreting took place in criminal courts in Barcelona. Criminal proceedings have been video-recorded in Barcelona’s criminal courts since 2009 and these recordings constitute the official records of the proceedings. Obtaining permission to access the video recordings involved several meetings and submission of written documents explaining very clearly the interests of the researchers and the use that would be made of the oral corpus to be created, as well as the commitment to anonymise the corpus by signing a strict confidentiality agreement. Attention was focused on a specific criminal summary procedure known in Spanish as *procedimiento abreviado* and specifically the cases tried in courts known as *Tribunales de lo Penal* (Criminal Courts).

Once permission to access video-recordings was granted, the researchers first studied the listings of all the trials that had taken place in the last seven years (2009-2015) as provided by the service of translation and interpreting of the Justice Department. They then selected those in which interpreting in the five working language combinations of the research team (Romanian, English, Arabic, Chinese and French) had supposedly taken place. Finally, the decision was made to request the recordings available from 50% of the *Tribunales de lo Penal* in Barcelona, so that the corpus compiled would be representative. There are currently 28 such courts in Barcelona, but only 24 of them are specifically trial courts, since four of them are devoted to enforcement of judicial resolutions which are solved through written proceedings. Therefore, out of a total of 24 such courts where interpreting is used, the researchers requested the video-recordings of 12 criminal courts, which were chosen randomly.

### **2.1. Unexpected events**

In principle, the methodology for the corpus design thus satisfied all research requirements for representative data collection in time. However, unexpected events modified this initial situation, and although researchers can still claim to

have a representative corpus that can describe reality, because it includes 50% of all data available, the size of the corpus was considerably diminished as a result of the following circumstances.

Firstly, although permission had been granted for access to video-recordings of the 12 courts chosen randomly, each court had to be provided with a specific list of the videos requested. Each court had its own clerks and its own method of dealing with the working processes for which they were responsible. Consequently, some of the courts were very quick to provide the videos requested whilst in others the process took several months. In fact, two of the selected courts were finally eliminated from the list because, after waiting for over one year, they were unable to deliver the recordings requested, due to administrative and bureaucratic problems.

Secondly, after receiving the video recordings from each selected court and checking them against the list of recordings that had been requested, the researchers found that some videos were missing, so that each remaining court was sent a second request for the missing videos. Finally, only after a further year was it possible to complete an electronic folder including all the videos received.

Thirdly, given the late reception of the video-recordings, the process of transcription was delayed until June 2015. Moreover, when researchers began studying the videos, they discovered that quality of sound and image of many of them, especially the oldest ones, made it very difficult to transcribe them and to create a corpus that could then be used for the proposed description of reality. The decision was thus made to use only the most recent recordings (2014-2015) which were of a better quality.

A further problem related to the transcription phase arose when the large number of video-recordings received was checked against the funds available to pay technicians to transcribe the recordings. This led to the difficult decision to start by transcribing the 2015 videos and to transcribe only three language combinations (English, French and Romanian) instead of the five initially conceived[5].

### **3. Corpus description**

Although it was not possible to transcribe many of the video-recordings obtained because of lack of time and funds, there was nevertheless some very interesting metadata that could be obtained from them. Therefore, a list of 20 items was created and the TIPp project, besides transcribing the 2015 trials, is also extracting this metadata from all the video recordings received, from 2010 to 2015. The metadata includes items such as the quality of sound and image, the interpreting techniques used (*chuchotage*, notetaking) if the interpreter is introduced by the

judge or not, and other data that could be of interest for further studies or for deciding which trials to transcribe in possible future projects.

The transcribed corpus, thus, includes the first six months' recordings from 2015. This, however, does not mean that more trials will not be transcribed in the future, if more funds are found to enlarge the corpus.

Therefore, the final, transcribed corpus includes all the videos obtained from trials where interpreting took place in 10 criminal courts of Barcelona for three language combinations (English, French and Romanian into Spanish). Table 1 illustrates the characteristics of the corpus that has been actually transcribed. The researchers hope that this work in progress can evolve in the future to include the transcription of recordings in the other two language combinations (Arabic and Chinese into Spanish) in the corpus.

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>	<b>H</b>
<b>2015 (January -June)</b>	Trials where an interpreter was requested	Missing video recordings	Video recordings obtained	Trials with no actual interpreting	Trials with interpreter	Transcribed Trials	Bilingual minutes transcribed	Total minutes of trial transcribed
French	52	9	43	32	11	9	92	190
English	65	10	55	33	22	19	123	371
Romanian	114	37	77	45	32	27	124	555
<b>TOTAL</b>	231	56	175	110	65	55	335	1116

**Table 1. Transcribed 2015 corpus description**

The first column of table 1 shows the period in which the trials were video-recorded and the linguistic interpreting combination, in all cases into and from Spanish. Column A shows the number of hearings in which, according to the listings provided by the service of translation and interpreting of the Justice Department, an interpreter was requested (a total of 231). However, when recordings were requested, even the second time, many were missing, and column B shows the number of final missing recordings (a total of 56). We cannot be absolutely sure about the reasons for these missing video recordings, but it is very likely that this may be due to trials that were suspended, for example, because the defendant or his/her lawyer did not show up. The result of subtracting the missing recordings (column B) from the original list of potential recordings (column A) is column C, which shows the actual number of videos obtained. These vary from 43 in the French-Spanish language combination to 77 in the Romanian-Spanish language combination, a total of 175 trials.

A further number has to be subtracted (column D) from these 175 recorded trials because the researchers noticed that in effect there was no intervention of an

interpreter. The reasons in this case are known and vary from cases where the witness who was going to be interpreted did not appear in court, to cases where a plea bargain agreement was reached between the parties before the trial started and therefore the intervention of the interpreter was unnecessary. The resulting number is surprising since it represents almost two thirds of the trials, so that once subtracted from the initial number, only 65 include the intervention of an interpreter (column E). The marked difference between the official data made available in the list of trials in which an interpreter was requested (231) and the trials in which an interpreter was actually involved (65) should be taken into account when describing reality. This article does not aim to discuss the results of the data obtained, but we believe they are worthy of note, since they have clear implications that will be discussed in further articles.

Finally, of the 65 recordings in which an interpreter was involved, some were not transcribed because either the interpreter did not have to speak -because the accused or the witness said that s/he could speak in Spanish and did not need an interpreter- or the only interpreting taking place during the trial was *chuchotage*, which is not recorded in the video because the volume is too low to be recorded. Ultimately, therefore, the corpus gathered consists of the transcription of 55 trials (column F) which altogether last for 1116 minutes (column H).

Column G shows the difference between the total duration of the trials (that amount to 1116 minutes of oral interventions that have been transcribed) and the total minutes interpreted, which only amount to 339. If we add the number of minutes where there has been *chuchotage*, this figure grows to 513 minutes, which is 46% of the total minutes of the trials. This data is of interest because it means that less than one half of the trial is actually interpreted to the defendant, a finding that implies a clear violation of the defendant's right of information according to both European and Spanish laws.

In sum, the TIPp transcribed corpus consists of 55 trials and 1116 minutes of oral interventions in three language combinations: English, French and Romanian from and into Spanish.

Given the amount of material gathered and the impossibility of transcribing all the 2014 video-recordings, the researchers decided to leave the transcription of the 2014 videos for further research projects. Nevertheless, the main data in these videos, as displayed in Table 2, was extracted and proved to be very helpful when determining whether the 2015 corpus was really representative and significant in terms of the description of reality. A comparison was made between the number of trials that finally did not take place for whatever reason in 2014 and 2015 as well as



the numbers of trials in which an interpreter was requested but finally was not needed.

<b>2014 (January- December)</b>	Trials where an interpreter was requested	Missing video recordings	Video recordings obtained	Trials with no interpreting	Trials with interpreter
Arabic	258	97	161	77	84
Chinese	97	36	61	17	44
French	75	44	31	18	13
English	77	19	58	31	27
Romanian	206	89	117	68	49
<b>TOTAL</b>	<b>713</b>	<b>285</b>	<b>428</b>	<b>211</b>	<b>217</b>

**Table 2. Not transcribed 2014 corpus description**

Table 2 shows that the data obtained for 2014 supports the reliability of the data obtained in 2015, since the proportion between the total figures for both items is very similar.

#### **4. Corpus transcription**

There are several transcription systems available to researchers, and the differences between them can be very small regarding, for instance, how to represent a pause or a fragment that cannot be understood inside the transcription, but there can also be important theoretical variations[6].

Researchers in the TIPp project, after considering the different possibilities, decided to transcribe in the simplest and most straightforward way possible, since the main interest was in the annotation of a corpus that reflected reality. This means writing what is said exactly as it is heard. Thus, for instance, grammar mistakes, incorrect pronunciation or hesitations are transcribed without amendments or comments. There is only one exception to this rule, when incorrect pronunciation causes the reader of the transcription to misunderstand what is being said. For example, in one case in which the accused says what sounds like “aquachis” meaning “aquagym”, the transcriber has to write “aquagym” but also to include the word as it was pronounced between square brackets. When a word is incomprehensible, due to problems of the sound recording –for instance during *chuchotage* if the interpreter talked at a distance from the microphone- then the transcriber marks it with three points inside round brackets: (...).

Another important decision researchers made was not to include any reference to nonverbal communication, unless it was completely necessary in order to understand the message. An example of this necessary comment would be the case of the accused shaking his head to say “no” but not saying anything and the interpreter saying “no”. In this case, the transcriber includes a comment explaining

in a very simple way what has happened, between double round brackets, for example ((the accused moves his head from right to left meaning “no”)).

Finally, the most important difference between the chosen transcription system and other possible options is that all the TIPp corpus is fully anonymised, for confidentiality reasons. This is also important in order to be able to make the corpus publicly available. Therefore, all references to names of people, streets, recognisable places such as restaurants, badge numbers of policemen, telephone numbers and so on have been substituted by a list of fake, previously accorded names and numbers.

Regarding the software used, only one tool was used to transcribe, annotate and retrieve the data desired: EXMARaLDA[7]. EXMARaLDA was originally developed in the project “Computer-assisted methods for the creation and analysis of multilingual data” at the Collaborative Research Center “Multilingualism” (Sonderforschungsbereich “Mehrsprachigkeit” – SFB 538) at the University of Hamburg and since 2011, the development of EXMARaLDA continues at the Hamburg Centre for Language Corpora in cooperation with the Archive for Spoken German at the Institute for the German Language in Mannheim. It consists of a transcription and annotation tool (Partitur-Editor), a tool for managing corpora (Corpus-Manager) and a query and analysis tool (EXAKT). It works with XML based data formats which interoperate with one another and enables a flexible processing and sustainable usage of the data. This was a major finding for the TIPp project, since the tool allows researchers to create, annotate and exploit the corpus at one and the same time, and even facilitates the extraction of specialized terminology, which is also important for one of the outputs of the TIPp project: the creation of terminological records.

Figure 1 shows an example of a fragment of a trial transcribed, using the EXMARaLDA software.

The screenshot shows the EXMARaLDA software interface. The title bar reads 'EXMARaLDA Partitur-Editor 1.6 [C:\Users\2041860\Desktop\PO2\_0000088\_2015\_01\_EN\_COMPLETA (2)-L.T.erb]'. The menu bar includes 'File', 'Edit', 'View', 'Transcription', 'Tier', 'Event', 'Timeline', 'Format', 'CLARIN', and 'Help'. The toolbar contains various icons for file operations and editing. The main window displays a transcription table with the following structure:

	00 [034.0]	00 [038.0]	00 [040.0]	01 [040.0]	01 [042.0]	01 [044.0]
J [J]						
I1 [I]	Es mentira. No estaba allí		The police arrested you that very moment.		The police caught you.	
I2 [I2]						
I1 [Retrad_I1]						
A1 [A]						I'm just walking on my own, I don't even know what is going on.
A1 [Retrad_A1]						
A2 [A2]						
A3 [A3]						
A4 [A4]						
SPK35 [AS]						
F [F]			Eh, la policía las detuvo en ese momento.			
F [F2]						

**Figure 1. Fragment of a trial transcription using the EXMARaLDA software.**

The example transcribed in Figure 1 shows how a different colour has been assigned to every speaker, so the green colour refers to the interpreter, who says ‘Es mentira. No estaba allí’ [*That is not true, I wasn’t there*], then the prosecutor, in blue colour, says ‘Eh, la policía las detuvo en ese momento’ [*Eh, the police arrested them at that moment*] and then the interpreter, before the prosecutor finishes the sentence, starts speaking again to translate what the prosecutor just said. The overlap between the speakers can be seen thanks to the timeline provided by the EXMARaLDA software. Then, the accused person, in red colour, says ‘I’m just walking on my own, I don’t even know what’s going on’.

As can be also seen in Figure 1, there is one tier or row devoted to each speaker, so that all trials consulted can be easily analysed, because they always follow the same order: the first tier or row is devoted to the Judge, the second to the interpreter, the third to a second interpreter in case there are two interpreters in the room, the fourth to the retranslation of the interpreter into Spanish (this is only used when needed, for instance in the case of Romanian, Chinese and Arabic, which are languages not so well known for all the researchers, but it is not used when the language to which the interpreter is translating is English or French), the fifth to the accused, the sixth and the seventh to other possible accused people, the eighth to the prosecutor, the ninth to the defence lawyer, and so on.

## 5. Corpus annotation

Regarding annotation, the researchers first checked many different annotation systems, such as part-of-speech, lemmatization, syntactical (parsing), semantic (domain classifications), coreference (discourse), pragmatic (speech acts – dialogue) and stylistic[8], and then also considered other qualitative content analysis annotation systems, but found that, although some of the latter systems were close to the needs of the TIPp project, none was suitable for the study purposes.

Therefore, an ad hoc annotation system for this research was created from zero. The main goal of the project of describing reality was operationalised into categories or indicators that can be observed and marked in the corpus, and a whole classification system was created. This system includes, first of all, two main categories, namely interaction and textual problems, based on Wadensjö's distinction between 'talk-as-activity' and 'talk-as-text' (Wadensjö 1998: 21).

The textual problems annotated assess the fidelity of the message conveyed by the interpreter and signal the places in which the interpreter has found linguistic, cultural, or domain-related (for instance legal) problems in the oral discourse. Here, linguistic is understood in the wider sense of the term, including not only textual, syntactic and lexical levels, but also the pragmatic level, so that it would include, for example, problems of register or changes in the discourse.

The textual problems are firstly tagged and then two different annotations are marked and stored in the corpus for each element; the first one, shown in Table 3, assesses if the solution to the textual problem found has been (i) adequate, that is, conveying the message adequately, (ii) inadequate, that is, not conveying the message adequately or (iii) improvable, that is, the interpreter conveys the message roughly but the solution could be clearly improved.

---

**Textual annotation:**

1. Indicator of fidelity, that is the solution applied by the interpreter when facing a textual problem was:

.....  
- (A) Adequate.  
.....

- (M) Improvable  
.....

- (I) Inadequate.  
.....

**Table 3. Scale created and used to annotate in the corpus the solution applied by the interpreter when facing a textual problem.**

The second annotation for textual problems signals the type of solution adopted by the interpreter and the possible categories are shown in Table 4.

---

**Textual annotation:**

2. Indicator of the type of solution applied by the interpreter when facing a textual problem:

.....  
**Possible categories when the solution applied has been marked in the previous textual indicator as 'adequate':**  
.....

- (EH) Usual equivalent.
- (IM) Making some information implicit.
- (EX) Making some information explicit.

---

**Possible categories when the solution applied has been marked in the previous textual indicator as ‘improvable’:**

- (CR) Change of register
  - (NMS) Slightly different meaning (from that of the original message).
- 

**Possible categories when the solution applied has been marked in the previous textual indicator as ‘Inadequate’:**

- (O) Omission.
  - (OG) Dangerous omission.
  - (NT) Not translated.
  - (AD) Addition of information.
  - (ADG) Dangerous addition of information.
  - (ITER) Inadequate terminology.
  - (FS) Wrong meaning (a very different meaning from that of the original message).
  - (FSG) Dangerous wrong meaning.
  - (CS) Opposite meaning (saying the opposite of what was conveyed in the original message).
  - (SS) Sentence with no meaning (message is not understandable, does not make sense).
- 

**Table 4. Scale created and used to annotate in the corpus the type of textual solution applied by the interpreter when facing a textual problem.**

As shown in Table 4, there are many possible solution types that have been annotated and stored in the corpus. Unfortunately, we cannot describe them thoroughly here, since a whole article is needed to do that, so in order to see a thorough explanation of these categories and examples of each of them see Orozco-Jutorán (2017b).

However, we would like to point out that there has been a distinction made between ‘serious errors’ (which include four of the categories listed in Table 4 as inadequate types of solutions: dangerous addition of information, dangerous omission, sentence with no meaning and dangerous wrong meaning) and other, ‘less serious’ type of errors. By serious errors, we mean errors that might affect or interfere with the result of the proceeding, as shown in the following example, where we have included our translation of the Spanish oral interventions between square brackets and where the dangerous addition of information is underlined:

Judge: ... *que si reconoce los hechos y está conforme.*  
 [Does he acknowledge the facts and agrees?]  
 Interpreter: *Do you accept?*  
 Defendant: *Sí.*  
 [Yes]  
 Interpreter: *Yeah? And do you agree?*  
 Defendant: *Yeah.*  
 Interpreter: *Sí, es culpable.* [Yes, he is guilty]

Although there is no space here to make an analysis of the results found, we would like to signal that the amount of serious errors found in the corpus is alarmingly large, as Table 5 shows.

Language	Dangerous omissions per bilingual hour	Dangerous addition of information per bilingual hour	Dangerous wrong meanings per bilingual hour	Sentences with no meaning (SS) per bilingual hour	Total of serious errors per bilingual hour
English	6,3	2,6	7,3	4,4	20,6
French	5,9	1,3	6,5	1,3	15,0
Romanian	12,6	4,8	7,3	1,0	25,7
<b>Mean</b>	<b>8,5</b>	<b>3,2</b>	<b>7,1</b>	<b>2,3</b>	<b>21,1</b>

**Table 5. Number of serious errors found in the corpus per bilingual hour of trial.**

We would also like to mention that one of the findings yielded by the analysis of the annotated corpus is that most of the trial is not actually translated for the user, who is usually the defendant. This is measured by one of the categories created under “inadequate textual solutions”: “not translated” (NT). In order to be marked as NT, there needs to be a whole intervention (therefore, a whole speech act) by the judge, the defence lawyer, the public prosecutor or a witness which has not been translated, so there is an important difference with the omissions, which affect only a word or a sentence which has not been translated. Table 6 shows the amount of NT found per hour and per minute in the corpus, which, again, is alarmingly large.

Language	Total of NT per hour	Total of NT per minute
English	371	1,8
French	190	1,6
Romanian	555	3,7
<b>Mean</b>	<b>372</b>	<b>2,7</b>

**Table 6. Number of “Not translated” interventions (NT) found in the corpus per hour and per minute of trial.**

The interaction problems annotated signal the moments in the oral interaction in court where any one of the participants (judge, lawyers, interpreter, defendant, witnesses, and so on) has had a problem. These problems include those relating to conversation management, non-renditions (Wadensjö 1998) and speech style. In order to annotate each of these types of problems, several categories were created, as Table 7 shows.

**Interaction annotation:**

**Possible categories regarding conversation management problems:**

- (S) overlap
- (I) Interruption
- (DL) long turns (that is when a member of the judicial staff speaks for more than two minutes in a single turn)

**Possible categories regarding conversation non renditions:**

- (J) Justified (that is pause, clarification, confirmation or retrieval)
- (NJ) Unjustified (that is warning, instructions, advice, answering on behalf of the defendant or adding extra information)
- (RT) Reactive tokens (that is when the interpreter's non-rendition merely acknowledges that he or she received the information in the original utterance)

**Possible categories regarding speech style, by both the interpreter and the courtroom personnel:**

- Direct speech
- Indirect speech
- Reported speech

**Table 7. Scale created and used to annotate interaction problems in the corpus.**

Again, we cannot describe the categories thoroughly here, since a whole article is needed to do that, so in order to see a thorough explanation of these categories and examples of each of them, see Arumí and Vargas-Urpí (forthcoming).

Figure 3 shows what the annotations look like in the corpus. As can be seen, one tier or row is devoted to each of the types of problems mentioned, both textual and interaction problems. In the example, on top of Figure 3, there are all the tiers or rows devoted to the speakers and the transcription of what they said at the fragment previously shown in Figure 1. Then, below those rows, starting in tier 17, the annotation tiers can be seen, the first one called 'PROBLEMA'. This tier is where the researchers tag the fragment where there is a textual or interaction problem. For instance, on the first grey column in Figure 3, below where the interpreter says 'Es mentira. No estaba allí', there is an 'I' meaning that there is an 'interaction problem' in that sentence, and then, a few rows below, in the tier devoted to speech style, there is the annotation *INDIR*, meaning that the interpreter is using indirect speech (saying '*They* were not there' instead of using the same speech style used in the original sentence by the defendant, which would be '*We* were not there').

In the next column, to the right, the prosecutor speaks, saying 'Eh, la policía las detuvo en ese momento' [*Eh, the police arrested them at that moment*] with no annotation or tag below, because there is nothing to be annotated in that sentence, since there is not any problem faced by the interpreter there, and then in the next column, the interpreter translates the prosecutor but starts speaking before the prosecutor finished his sentence. This overlap between the speakers is marked by the tag "I" in the tier '*PROBLEMA*', since there is an interaction problem, and then there is the tag *SOI* at the tier belonging to *SOLAPAMIENTO*, which means "overlap" in Spanish. This *SOI* stands for "overlap with the interpreter", and is differentiated from an overlap between the Judge and the prosecutor of the defence attorney, which would be annotated as *SOJ*. In this same sentence there are two

more annotations. The first one is not an interaction problem but an interaction observable phenomenon (and that is why, in the tier for *PROBLEMA*, next to the “I”, there is an “F”, which stands for *Fenómeno*, which is the Spanish word for “phenomenon”). The observable phenomenon is then annotated in the style tier, tagged as *DIR*, because here the interpreter is not using indirect speech but direct speech, as would be recommended in this case. The second annotation is of textual nature, that is why next the “I” and the “F” at the *PROBLEMA* tier there is also an “S” (meaning “Solution”). In the tier right below this one, there is the annotation “A”, meaning “adequate solution” and in the tier below the type of solution applied by the interpreter to the textual problem is tagged as EH, which stands for “Equivalente Habitual” [*usual equivalent*].



EXMARaLDA Partitur-Editor 1.6 [C:\Users\2041860\Desktop\PO2\_0000088\_2015\_01\_EN\_COMPLETA (2)-1\_T.exb]

File Edit View Transcription Tier Event Timeline Format CLARIN Help

	88 [136.0]	89 [138.0]	90 [140.1]	91 [140.9]	92 [142.0]
J [t]					
II [t]	Es mentira. No estaban allí		The police arrested you that very moment.	The police caught you	
II [AD]					
II [Retrad_II]					
A1 [t]					
A1 [Retrad_A1]					
A2 [t]					
A3 [t]					
A4 [t]					
SPKS [A5]					
F [t]		Eh, la policía las detuvo en ese momento.			
LD1 [t]					
LD2 [t]					
LD3 [t]					
LD4 [t]					
LAPI [t]					
[PROBLEMA]	I		I/F/S		
[SOLUCION TEXTUAL]			A		
[TIPO SOLUCION TEXTUAL]			EH		
[SOLAPAMIENTO]			SOI		
[INTERRUPCION]					
[DISCURSO LARGO]					
[VOZ PROPIA]					
[TIPO VOZ PROPIA]					
[ESTILO (OPERADORES JUDICIALES)]					
[ESTILO (INTERPRETE)]	INDIR		DIR		
[INAUDIBLE]					

**Figure 3. Fragment of a trial transcribed and annotated.**

All the information annotated in the corpus in the way that has just been explained is then converted or transformed into excel files, an example of which can be seen at Figure 3.

	A	B	C	D	E	F	G	H	
1	PROBLEMA	SOLUCION	TIPO_SOLUCION	SOLAPAMIENTOS	INTERPRETACION	DISCURSOS	VOZ_PROPIA	TIPO_VOZ_PROPIA	ESTILO_OPERACION
2	F							DIR	
3	I							INDIR (FP)	
4	I								
5	I							INDIR (FP)	
6	F								
7	I							INDIR (FP)	
8	I/T	I	NT				NJ	INFOEXTRA	
9	T	I	NT						
10	F					J		ACLAR	
11	I				IR				

**Figure 3. Detail of an excel file that includes the annotations.**

There is an excel sheet for each trial and then one excel book or file containing all the trials in one language combination. Then, there is a “bigger” excel file, linked to the three sheets containing total data for each language, that combines the results of the three language pairs that have been analysed. As can be seen in Figure 3, the rows or tiers from the EXMARaLDA software are converted here in columns and allow the application of filters and formulas to obtain quantifiable data as the one shown in tables 5 and 6. This system has proved to be very useful because it allows researchers to obtain quantifiable data to be able to describe reality in a systematic rather than an anecdotic way.

## 6. Resources

As has been already mentioned, the TIPp project aims at describing reality but also wishes to contribute to improving court interpreters’ performance by creating a series of resources directed to interpreters and courtroom personnel. TIPp has created a free, accessible website designed to be used from any mobile device that includes four resources.

Firstly, a set of recommendations for court interpreters, which could be considered a code of good practice, that is, a protocol for professional conduct in the most frequent situations for a court interpreter. The difference between the already existing codes and this resource is that TIPp’s intention is to give focused, practical advice that can be applied by interpreters in their daily performances and that all advice given is based on irregular or difficult situations observed in the corpus compiled and analysed, and therefore respond to real court interpreters’ needs. The recommendations are specific, written suggestions or videos.

Secondly, the same procedure has been followed to write a set of recommendations for courtroom personnel regarding the role of the interpreter and how to interact with interpreters.

Thirdly, the website contains Spanish monolingual glossaries for different contexts, such as certain type of crimes or “general vocabulary” for criminal trials for example. Each of these glossaries includes lists of terms found in the corpus and examples of use for each term, so that collocations and context can be seen by the interpreter in order to help him/her when preparing for court interpreting.

Lastly, the application includes a pilot sample of five databases -one in each language combination (English, French, Romanian, Arabic and Chinese from and into Spanish)- containing the problematic units most frequently encountered by court interpreters, as observed in the TIPp corpus. The databases include, for every term or unit, a translation-oriented terminological record which includes potential solutions, comments and translation options, following the structure of the translation-oriented record created for a previous research project[9] (for further explanations on this type of record, see Prieto and Orozco-Jutorán 2015 and Orozco-Jutorán 2017a). Although researchers initially intended to create exhaustive databases in five language combinations, the decisions made by researchers (explained in sections 2 and 3) have meant that the current corpus is only exhaustive for three language pairs (English, French and Romanian into Spanish). Therefore, the terms included for Chinese and Arabic are only a small sample, taken from transcription of case studies. The researchers aim to enhance the databases by adding more transcriptions to the corpus in the future, provided more funds are made available for transcribing a larger number of trials.

## **5. Conclusions**

The TIPp project has used a pioneering methodology in the field of court interpreting research in Spain, as it is based on authentic materials extracted from real criminal proceedings. These materials have allowed the researchers to create and exploit a representative oral corpus that can be further extended in the future. On the basis of what has been observed in the corpus, through the use of an ad hoc annotation system that marks both textual and interaction problems found by the court interpreters, the researchers have reached important and alarming conclusions, such as that less than half of the hearing is actually interpreted to the defendant, that only 30% of the interpretation is audible and is properly recorded, or that there are too many serious errors in the translated part of the trials for it to be considered acceptable, which actually means that there is a violation of the defendant’s rights.

Besides this descriptive data, the researchers have used the information obtained from the corpus to create a computer application, accessible through any mobile device, that includes resources directed to both interpreters and courtroom personnel which aims at helping court interpreters to perform their tasks more accurately and efficiently.

We hope that this will subsequently have an impact on the main users, namely defendants, usually from migrant communities, who could be left defenceless unless provided with effective legal protection through the services of good quality, professional interpreting in the courtroom setting. Furthermore, other secondary users of interpreting during trials, such as witnesses and victims will also benefit from the improved quality of court interpreting.

## References

- Angermeyer, Philipp S. (2015) *Speak English or What? Codeswitching and Interpreter Use in New York City Courts*, New York, Oxford University Press.
- Angermeyer, Philip S., Bernd Meyer, and Thomas Schmidt (2012) “Sharing Community Interpreting Corpora. A Pilot Study” in *Multilingual Corpora and Multilingual Corpus Analysis*, Thomas Schmidt and Kai Wörner (eds), Amsterdam, John Benjamins: 275–294.
- Arumí, Marta, Carmen Bestué, Sofía García-Beyaert, Anna Gil-Bardají, Jacqueline Minett, Liudmila Onos, Begoña Ruiz de Infante, Xus Ugarte, and Mireia Vargas-Urpi (2011) *Comunicar en la diversitat. Intèrprets, traductors i mediadors als serveis públics*, Barcelona, Linguamón-Casa de les Llengües.  
URL: [http://grupsderecerca.uab.cat/miras/sites/grupsderecerca.uab.cat/miras/files/informe\\_miras\\_ispc\\_2011\\_0.pdf](http://grupsderecerca.uab.cat/miras/sites/grupsderecerca.uab.cat/miras/files/informe_miras_ispc_2011_0.pdf)(accessed 10 June 2016).
- Arumí, Marta, Carmen Bestué, Sofía García-Beyaert, Anna Gil-Bardají, Jacqueline Minett, Miren Olaciregui, Liudmila Onos, Begoña Ruiz de Infante, Xus Ugarte, and Mireia Vargas-Urpi (2012) “Traducció i immigració: La formació de traductors i intèrprets als serveis públics, noves solucions per a noves realitats” in *Recerca i immigració IV*. Barcelona, Generalitat de Catalunya: 157–183.
- Arumí, Mireia and Marta Vargas-Urpi (forthcoming) “Annotation of Interpreters’ Conversation Management Problems and Strategies in a Corpus of Criminal Trials in Spain: The Case of Non-renditions”, *Translation and Interpreting Studies* 13, no. 3.
- Bendazzoli, Claudio (2010) *Corpora e interpretazione simultanea*, Bologna, Asterisco. URL: <http://amsacta.unibo.it/2897/>(accessed 10 June 2016).

- Berk-Seligson, Susan (1987) "The Intersection of Testimony Styles in Interpreted Judicial Proceedings: Pragmatic Alterations in Spanish Testimony", *Linguistics* 25: 1087–1125.
- (1988) "The Impact of Politeness in Witness Testimony: The Influence of the Court Interpreter", *Multilingua* 7, no. 4: 411–439.
- (1989) "The Role of Register in the Bilingual Courtroom: Evaluative Reactions to Interpreted Testimony" in *U.S. Spanish: The Language of Latinos. Special issue of the International Journal of the Sociology of Language*, Irene Wherritt and Ofelia Garcia (eds) 79, no. 5: 79–91.
- (1999) "The Impact of Court Interpreting on the Coerciveness of Leading Questions", *Forensic Linguistics* 6, no.1: 30–56.
- (1990/2002) *The Bilingual Courtroom: Court Interpreters in the Judicial Process*, Chicago and London, University of Chicago Press.
- Cooke, Michael (1995) "Interpreting in a Cross-cultural Cross-examination: An Aboriginal Case Study", *International Journal of the Sociology of Language* 113, no. 1: 99–111.
- Del Pozo Triviño, Maribel, Antonio Vaamonde List, David Casado-Neira, Silvia Pérez Freire, Alba Vaamonde Paniagua, Doris Fernandes del Pozo, and Rut Guinarte Mencía (2014) *Communication Between Professionals Providing Attention and Gender Violence Victims/Survivors Who Do Not Speak the Language: A Report on the Survey Carried Out on Agents During the Speak Out for Support (SOS-VICS) Project*. Vigo: Servizo de Publicacións da Universidade de Vigo.
- Edwards, A. Jane, and Martin D. Lampert (eds) (1993) *Talking Data: Transcription and Coding in Discourse Research*, Hillsdale NJ, Lawrence Erlbaum Associates.
- Edwards, A. Jane (2001) "The Transcription of Discourse", in *The Handbook of Discourse Analysis*, Deborah Schiffrin, Deborah Tannen and Heidi E. Hamilton (eds), Malden MA, Blackwell: 321–348.
- Emerson Crooker, Constance (1996) *The Art of Legal Interpretation. A guide for court interpreters*, Portland OR, Portland State University.
- Fairclough, Norman (2001). "Critical Discourse Analysis", in *How to Analyse Talk in Institutional Settings: A Casebook of Methods*, Alec McHoul and Mark Rapley (eds), London, Continuum: 25–40.

Falbo, Caterina (2005) "La transcription: une tâche paradoxale", *The Interpreters' Newsletter* 13: 25–38.

Fowler, Yvonne (2007) "Interpreting into the Ether: Interpreting for Prison/Court Video Link Hearings". Presentation at the *Critical Link 5 – Quality in Interpreting: A Shared Responsibility*, 11-15 April 2007 Parramatta – Sydney (Australia).

URL: [http://static1.squarespace.com/static/52d566cbe4b0002632d34367/t/5347f7e7e4b0b891fcd56cee/1397225447306/CL5Ellam\\_Fowler.pdf](http://static1.squarespace.com/static/52d566cbe4b0002632d34367/t/5347f7e7e4b0b891fcd56cee/1397225447306/CL5Ellam_Fowler.pdf)(accessed 10 June 2016)

Goldflam, Russell (1995) "Silence in Court! Problems and Prospects in Aboriginal Legal Interpreting" in *Language in Evidence: Issues confronting Aboriginal and multicultural Australia*, Diana Eades (ed), Sydney, University of New South Wales Press: 28–54.

Hale, Sandra (1997a) "Interpreting Politeness in Court. A Study of Spanish-English Interpreted Proceedings" in *Research, Training and Practice. Proceedings of the 2nd Annual Macarthur Interpreting and Translation Conference*, Stuart Campbell and Sandra Hale (eds), Milperra, UWS Macarthur/LARC.

---- (1997b) "Clash of World Perspectives: The Discursive Practices of the Law, the Witness and the Interpreter", *Forensic Linguistics* 4, no. 2: 197–209.

---- (1997c) "The Treatment of Register Variation in Court Interpreting", *The Translator* 3, no. 1: 39–54.

---- (1999) "Interpreters' Treatment of Discourse Markers in Courtroom Questions", *Forensic Linguistics* 6, no. 1: 57–82.

---- (2002) "How Faithfully Do Court Interpreters Render the Style of Non-English Speaking Witnesses' Testimonies? A Data Based Study of Spanish-English Bilingual Proceedings", *Discourse Studies* 4, no. 1: 25–48.

---- (2008) "Controversies over the Role of the Court Interpreter" In *Crossing Borders in Community Interpreting. Definitions and Dilemmas*, Carmen Valero-Garcés and Anne Martin (eds), Amsterdam, John Benjamins: 99–122.

Halverson, Sandra (1998) "Translation Studies and Representative Corpora: Establishing Links between Translation Corpora, Theoretical/Descriptive Categories and a Conception of the Object of Study", *Meta* 43, no. 4: 494–514/1–22.

- Heritage, John (1997) "Conversation Analysis and Institutional Talk: Analyzing Data" in *Qualitative Research: Theory, Method and Practice*, David Silverman (ed), London, SAGE: 161–182.
- Kadric, Mira (1999) "Interpreting in the Austrian Courtroom", in *The Critical Link 2: Interpreters in the Community*, Roda P. Roberts, Silvana E. Carr, Diana Abraham and Aideen Dufour (eds), Amsterdam, John Benjamins: 154–164.
- Lane, Chris, Katherine McKenzie-Bridle, and Lucille Curtis (1999) "The Right to Interpreting and Translation Services in New Zealand Courts", *Forensic Linguistics* 6, no.1: 115–136.
- Mikkelsen, Holly (1998) "Towards a Redefinition of the Role of the Court Interpreter", *Interpreting* 3, no. 1: 21–45.
- Montalvo, Margarita (2001) "Interpreting for Non-English-speaking Jurors: Analysis of a New and Complex Responsibility", in *ATA Proceedings for the 42nd Annual Conference*: 167–176.
- Moreno Sandoval, Antonio, and José María Guirao (2006) "Morphosyntactic Tagging of the Spanish C-ORAL-ROM Corpus: Methodology, Tools and Evaluation", in *Spoken Language Corpus and Linguistic Informatics*, Yuji Kawaguchi, Susumu Zaima and Takagaki Toshihiro (eds), Amsterdam, John Benjamins: 199–218.
- Morris, Ruth (1999) "The Gum Syndrome: Predicaments in Court Interpreting", *Forensic Linguistics* 6, no.1: 6–29.
- Nicholson, S. Nancy, and Bodil Martinsen (1997) "Court Interpretation in Denmark", in *The Critical Link: Interpreters in the community*, Silvana Carr, Roda P. Roberts, Aideen Dufour and Ludmila Stern (eds), Amsterdam, John Benjamins: 259–270.
- Niska, Helge (1995) "Just Interpreting: Role Conflicts and Discourse Types in Court Interpreting", in *Translation and the Law*, Marshall Morris (ed), Amsterdam, John Benjamins: 293–316.
- O'Connell, C. Daniel, and Kowal Sabine (1994) "Some Current Transcription Systems for Spoken Discourse: A Critical Analysis", *Pragmatics* 4, no.1: 81–107.
- Onos, Liudmila (2014) *La interpretación en el ámbito judicial: el caso del rumano en los tribunales de Barcelona*, PhD diss., Universitat Autònoma de Barcelona. URL: <http://hdl.handle.net/10803/285160> (accessed 9 June 2016).

- Orozco-Jutorán, Mariana (2017a) "Efficient Equivalent Search at Your Fingertips – The Specialized Translator's Dream", *Meta*62, no. 1: 137–154.
- (2017b) "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.
- Ortega Herráez, Juan Miguel (2006) *Análisis de la práctica de la interpretación judicial en España. El intérprete frente a su papel profesional*. PhD diss., Universidad de Granada. URL: <http://hdl.handle.net/10481/977> (accessed 9 June 2016).
- (2011) *Interpretar para la justicia*, Granada, Comares.
- Prieto Ramos, Fernando, and Mariana Orozco-Jutorán (2015). "De la ficha terminológica a la ficha traductológica: hacia una lexicografía al servicio de la traducción jurídica", *Babel* 61, no. 1: 110–130.
- Rapley, Tim (2007) *Doing Conversation, Discourse and Document Analysis*, London, Sage.
- Rigney, C. Azucena (1999) "Questioning in Interpreted Testimony", *Forensic Linguistics* 6, no.1: 83–108.
- Schmidt, Thomas (2011) "A TEI-based Approach to Standardising Spoken Language Transcription", *Journal of the Text Encoding Initiative* 1: 1–22.
- Schmidt, Thomas, and Kai Wörner (2009) "EXMARaLDA – Creating, Analysing and Sharing Spoken Language Corpora for Pragmatic Research", *Pragmatics* 19, no. 4: 565–582.
- (2012). "Introduction" in *Multilingual Corpora and Multilingual Corpus Analysis*, Thomas Schmidt and Kai Wörner (eds), Amsterdam, John Benjamins: ix–xi.
- (2014) EXMARaLDA. In *Handbook on Corpus Phonology*, Ulrike Gut Jacques Durand and Gjert Kristoffersen, (eds), Oxford, Oxford University Press: 402–419.
- Stern, Ludmila (1995) "Non-English Speaking Witnesses in the Australian Legal Context: The War Crimes Prosecution as a Case Study", *Law/Text/Culture* 2: 6–31.
- Tusón, Amparo (1997) *Análisis de la conversación*, Barcelona, Ariel.



Vargas-Urpi, Mireia (2012) *La interpretació als serveis públics i la mediació intercultural amb el col·lectiu xinès a Catalunya*. PhD diss., Universitat Autònoma de Barcelona. URL: <http://hdl.handle.net/10803/96486> (accessed 9 June 2016).

Vargas-Urpi, Mireia, and Marta Arumí (2014) “Estrategias de interpretación en los servicios públicos en el ámbito educativo: estudio de caso en la combinación chino-catalán” *InTRAlinea*, Vol. 16.

URL: [http://www.intralinea.org/current/article/estrategias\\_de\\_interpretacion\\_en\\_los\\_servicios\\_publicos\\_en\\_el\\_ambito\\_edu](http://www.intralinea.org/current/article/estrategias_de_interpretacion_en_los_servicios_publicos_en_el_ambito_edu)(accessed 10 June 2016).

Wadensjö, Cecilia (1998) *Interpreting as Interaction*, New York, Longman.

## Notes

[1] 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 and the Directive 2012/13/EU of the European Parliament and of the Council of 22 May 2012 on the right to information in criminal proceedings

[2] Our translation, taken from the text of the law: *Ley Orgánica 5/2015, de 27 de abril por la que se modifican la Ley de Enjuiciamiento Criminal y la Ley Orgánica 6/1985, de 1 de julio, del Poder Judicial, para transponer la Directiva 2010/64/UE, de 20 de octubre de 2010, relativa al derecho a interpretación y a traducción en los procesos penales y la Directiva 2012/13/UE, de 22 de mayo de 2012, relativa al derecho a la información en los procesos penales*. [<https://www.boe.es/boe/dias/2015/04/28/pdfs/BOE-A-2015-4605.pdf>]

[3] The official name of the project is ‘Translation quality as a guarantee of criminal proceedings. Development of technological resources for court interpreters in Spanish-Romanian, Arabic, Chinese, French and English language pairs’ and it has been funded by the Spanish Ministry of Economy and Competitiveness (FFI2014-55029-R). Seven researchers make up the research team: Dr. Marta Arumí, Dr. Anna Gil Bardají (Universitat Autònoma de Barcelona), Dr. Anabel Borja (Universitat Jaume I), Dr. Mireia Vargas-Urpí (Universitat Pompeu Fabra) and Dr. Francisco Vigier (Universidad Pablo de Olavide) and the two main researchers who lead the team are Dr. Carmen Bestué and Dr. Mariana Orozco-Jutorán (Universitat Autònoma de Barcelona).

[4] For a thorough description of the tools, see Schmidt and Wörner (2009, 2012 and 2014).

[5] To fully understand this decision one has to bear in mind that transcribing one minute of live trial involves at least 30 minutes of work for a trained transcriber.

[6] In this respect, see, for instance, Bendazzoli (2010), Edwards (2001), Edwards and Lampert (1993), Emerson (1996), Fairclough (2001), Falbo (2005), Fowler (2007), Halverson (1998), Heritage (1997), Moreno and Guirao (2006), O'Connell and Kowal (1994), Rapley (2007), Schmidt (2011), Tusón (1997).

[7] <http://www.exmaralda.org/en/>

[8] To see a quick review of these and other annotation systems, visit, for instance, <http://ucrel.lancs.ac.uk/annotation.html>

[9] The results of the Law10n research project, funded by the Spanish Ministry of Economy and Competitiveness, can be accessed at <http://lawcalisation.com/>

#### **About the author(s)**

Mariana OROZCO JUTORÁN is a professional translator and a Senior Lecturer in Translation from English into Spanish at the Autonomous University of Barcelona since 1996. She holds a PhD in Translation Theory, for which she received the Best Doctoral Thesis Award in 2005. She has been involved in several funded research projects and is currently a member of the following research groups: TRAFIL (Translating Remote Philosophies to Facilitate Understanding), MIRAS (Intercultural Mediation), and the LEXTRA (Legal Translation Studies) research network. She has published a book and more than 30 articles – in translation journals such as *Meta*, *Perspectives*, *Jostrans*, *Sendebarr*, *Trans*, *Quaderns*, *Hermeneus* – and book chapters in publications by John Benjamins, *Multilingual Matters*, Gunter Narr and others. She has been a member of the advisory boards of the *Quaderns (Revista de Traducció)* and *Tradumatica* journals, and is a reviewer for several international journals. As a professional translator, she has worked in a wide range of areas over the last twenty years.

Email: [please login or register to view author's email address]

©inTRAlinea & Mariana Orozco-Jutorán (2018).

"The TIPp project Developing technological resources based on the exploitation of oral corpora to improve court interpreting"

*inTRAlinea* Special Issue: New Findings in Corpus-based Interpreting Studies

Edited by: Claudio Bendazzoli, Mariachiara Russo & Bart Defrancq

This article can be freely reproduced under [Creative Commons License](#).

**Stable URL:** <http://www.intralinea.org/specials/article/2316>