

This is a pre-print copy of a chapter book published by Routledge/CRC Press, available online at <https://www.routledge.com/Research-Methods-in-Legal-Translation-and-Interpreting-Crossing-Methodological/Biel-Engberg-Martin-Ruano-Sosoni/p/book/9781138492103>

Orozco-Jutorán, Mariana (2019). "A mixed-methods approach in corpus-based interpreting studies: quality of interpreting in criminal proceedings in Spain", in: Łucja Biel, Jan Engberg, Rosario Martín Ruano, Vilemini Sosoni (eds.) *Research Methods in Legal Translation and Interpreting: Crossing Methodological Boundaries*. Routledge. p. 152-169.

A Mixed-Methods Approach in Corpus-Based Interpreting Studies: Quality of Interpreting in Criminal Proceedings in Spain

Mariana Orozco-Jutorán

Abstract

This chapter first presents a mixed-methods approach to studying interpreting in criminal proceedings and next shows the application of the method, illustrating it with the project TIPp ("Translation and Interpreting in Criminal Proceedings"). The project aimed at describing and assessing the reality of court interpreting in Spain and at creating a computer application which comprises a complete set of resources to facilitate court interpreters' performance. The researchers compiled, transcribed and analysed a representative oral corpus of real, video-recorded criminal proceedings with interpreting in the three language combinations studied (English, French and Romanian into Spanish).

Regarding the methodology, an attempt was made to operationalise the quality of legal interpreting by trying to measure it with various criteria. Therefore, the research combined a rigorous, qualitative design — ideally suited to describing real practice — with quantitative data analysis techniques and the creation of a measurement instrument. Two direct variables were chosen to describe the quality of court interpreting, namely interaction problems and textual problems, based on Wadensjö's distinction between 'talk-as-activity' and 'talk-as-text' (Wadensjö 1998, p. 21). Then the variables were operationalised into indicators to create a measuring instrument validated through a pilot study and consisting of several interval and categorical scales.

1. Introduction

The social context for our study was the new law passed in Spain in 2015 (*Ley Orgánica 5/2015, de 27 de abril*) to amend Spain's Code of Criminal Procedure. This new law was a result of the transposition of both 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 Directive 2012/13/EU of 22 May 2012 on the right to information in criminal proceedings. As worded in this new Spanish law, it "significantly reinforces procedural guarantees in criminal proceedings, by regulating the right to translation and interpreting in these proceedings and the right of accused persons to be informed of the purpose of the proceedings so that they can exercise their right of defence efficiently"ⁱ. Translation and interpreting have thus become an essential component to ensure the right to effective legal protection in the exercise of lawful rights and interests before the courts.

However, the reality and quality of court interpreting in Spain's criminal courts were never studied in a systematic and rigorous way using a representative oral corpus. This task was undertaken by the research group MIRAS based at Universitat Autònoma de Barcelona, which specialises in public service interpreting, backed by researchers from four Spanish universitiesⁱⁱ. The project was funded by the Spanish Ministry of Economy and Competitivenessⁱⁱⁱ.

2. Research methodology

The study adopted a mixed-methods approach, not only in the sense that it gathered quantitative and qualitative data, but also in that it combined features of different ontological and epistemological positions, usually related to quantitative and qualitative methodologies. On the one hand, it could be classified ontologically as objectivism, which:

assumes a positivist epistemology, which asserts that social phenomena [in this case, court interpreting quality] can be objectively researched, data about the social world can be collected and measured, and the resulting observations must remain independent of the researchers' subjective understanding; that is to say, the researcher remains independent and has no impact in the data. (Saldanha and O'Brien 2013, p. 11, our words between square brackets).

In this sense, the approach was empirical, since it sought new information from the observation of data: the interpreting analysed had already taken place and had been recorded as is the case in all the criminal proceedings in Spain. Thus, the researchers were non-participating observers: they were not present at the trials and had no impact on the data.

On the other hand, the study was not experimental because it did not seek to establish cause-effect relations, but rather to assess the quality of the existing reality. This recalls the ontological and epistemological position of realism, very commonly held in research in social sciences (Ormstorm *et al.* 2014, p. 15).

In respect of the methodology, the study was qualitative. Several features explain this: Firstly, the setting was natural: the researchers did not control any variables or intervene while the interpreting in the trials was taking place, so maximum ecological validity could be claimed. Secondly, once the oral corpus had been compiled and transcribed, a first explorative study was carried out to look at the data and develop and refine techniques to analyse it. Thirdly, the operationalisation of court interpreting quality proceeded in a cyclical fashion: indicators used to measure quality were tested and tailored to the nature of the data being observed, which is typical of qualitative methodology.

However, other features were aligned with a quantitative approach. For instance, the data sample used was not a small or already existing oral corpus, but a large, representative sample of real criminal proceedings that included interpreting. Another component involved creating and validating a measuring instrument, something clearly quantitative since it is deductive, aimed at measuring phenomena and sequential; additionally, there was a validation design which included a pilot study.

Therefore, if we understand methods to be “the practical means by which data are collected” (O’Reilly and Kiyimba 2015, p. 3) or “the practical ‘tools’ to make sense of empirical reality” (Saukko 2003, quoted in Saldhana and O’Brien 2013, p. 13) this was a mixed-methods study, since qualitative methods were used to collect data and quantitative methods were used to analyse them.

There is a growing interest in the use of corpus methods for comparative studies in Translation and Interpreting Studies (TIS) (see, for instance, Calzada Pérez 2017) and this is also true for Legal TIS (see Biel 2017, Portandolfo 2016, Vigier and Sánchez 2017). However, in the case of this project, the term “corpus method” was avoided for two reasons. Firstly, we considered our oral corpus to be “only” the data gathered. While the corpus had to be designed, compiled and transcribed, and it was a time-consuming and complex task^{iv}, this is also the case with other types of data, such as questionnaires or interviews, in any kind of research. Secondly, when corpus-based studies are carried out in TIS, corpus linguistics tools are usually applied to measure frequency, keywords and concordances, including collocates and clusters (Calzada Pérez 2017, p. 236). We did not use any of these, because they were not suitable for our purposes, which were quality-oriented and legal ones, i.e. to describe to what extent the interpreting was fulfilling the right of information of the accused person.

3. Research question and dependent variables

The main objective of the project was to study the quality of court interpreting in criminal proceedings in Spain to determine to what extent the interpreting was fulfilling the right of information of the accused person. Therefore, the main research question was “What is the quality of court interpreting in Spain?”, and the first task was to define and operationalise the theoretical construct of quality so that it could be turned into “tangible” or empirical units that could be observed.

After discussions and a literature review on the quality of public service interpreting and court interpreting^v, we decided to adopt an *ad hoc* definition which was drafted by one of the leading team members, Carmen Bestué, and based partly on a judgment by the Supreme Court of Canada^{vi} mentioned in Roberts-Smith (2009, see Bestué 2018). The definition was as follows:

To guarantee access to justice to the person with limited knowledge of the language being used in the courtroom, court interpreting should be continuous, precise, impartial, competent and contemporaneous. Precision is understood as fidelity, i.e. fulfilling legal equivalence, which means transferring the meaning of the message, including the repetitions, interruptions, errors, etc. and also reflecting the style, tone and register used by the speaker.

Out of the many possible theoretical frameworks in Interpreting Studies (see Pöchhacker 2016 for an overview) we chose Wadensjö’s approach to dialogue interpreting to operationalise the construct. Wadensjö’s approach (1998) goes beyond the monologic view (what she calls “talk as text”) and complements it with the dialogic view (“talk as activity”), understanding interpreting not only as a translation task, but also as mediation and coordination. In this way, she accounts for the double role of dialogue interpreters: relaying original utterances (renditions) and coordinating conversation (non-renditions).

In the words of Pöchhacker (2016, p. 79), Wadensjö launched “a new paradigm for the study of interpreting as dialogic discourse-based interaction”. This paradigm inspired us to create two dependent variables, namely interaction problems and textual problems, based on the distinction between “talk-as-activity” and “talk-as-text” (Wadensjö 1998, p. 21). Textual problems refer specifically to issues regarding the precision or accuracy of the message conveyed, whereas interaction problems encompass all matters related to dialogue-building and multi-party encounters where the interpreter is an active participant in the conversation. These two dependent variables were then operationalised with scales and indicators so that they could be observed and measured, which will be discussed in further sections of this chapter.

4. Data collection and sampling

Random sampling where every subject of the population being studied has an equal chance of being selected as a participant in the study is considered to be an ideal scenario since it is the most reliable sampling technique for ensuring generalisation. In our case, the population consisted of the oral interventions in criminal trials with an interpreter, and therefore the “subjects” were not people but the trials, which were seen as composed of oral interventions.

Since the purpose of the study was to describe and assess court interpreting in criminal proceedings, the decision was made to use a random sampling of criminal proceedings with interpreters taking place in Barcelona. Barcelona was chosen for two reasons. Firstly, it is the place where the research team was based and where contacts with the judiciary were possible to obtain a permission for access to recordings of the trials. Secondly, it is the second biggest city in terms of population in Spain, with 1,620,000 inhabitants^{vii} and it attracts a large number of immigrants and tourists^{viii}, who are the main defendants or witnesses with limited competence in the official languages in trial courts.

Regarding the size of the sampling, there are currently 28 Criminal Courts in Barcelona (known as *Tribunales de lo Penal* in Spanish), but only 24 of them are specifically trial courts that ask for interpreters when needed, since the other four deal with the enforcement of judicial decisions, which involves written proceedings only. Therefore, the question was what sample size should be used out of the total of 24 courts.

According to Creswell (1998, p. 64), there are no specific rules when determining an appropriate sample size in qualitative research, although obviously the bigger and the more representative the better. For phenomenological studies, for instance, Creswell recommends 5 to 25 cases. In TIS, however, corpora have been used long enough to have many scholars pondering on the size of a representative corpus. For example, Baker stresses quality over quantity:

One consideration when building a specialised corpus in order to investigate the discursive construction of a particular subject is perhaps not so much the size of the corpus, but how often we would expect to find the subject mentioned within it (...) Therefore, when building a specialised corpus for the purposes of investigating a particular subject or set of subjects, we may want to be more selective in choosing our texts, meaning that the quality or content of the data takes equal or more precedence over issues of quantity (2006, pp. 28-29, quoted in Corpas Pastor and Seghiri 2010, p. 120)

In this sense, our corpus is specialised and includes full texts rather than extracts. Based on an extensive literature review, Corpas Pastor and Seghiri (2010, p. 122) clarify the concept of representativeness and agree -quoting Wright and Budin (1997)- that the conclusions drawn from a corpus can be claimed to be meaningful, even if the size is not very large, if the subject of corpus is specialized because the vocabulary used in it is restricted. They also quote Ahmad and Rogers (2001: 736) to agree that a specialised corpus requires only tens of thousands of words, compared with the millions of words required for general-language lexicography.

Considering all the above, together with the resources available for transcribing (it takes between 40 and 50 minutes of work to transcribe one minute of trial), a decision was made to request the recordings available from 50% of the criminal courts in Barcelona for a specific period of time (January–June 2015). Therefore, out of the total of 24 such courts where interpreting is used, we requested video-recordings from 12 courts, which were chosen randomly.

Owing to some technical and practical difficulties, which are explained in detail in Orozco-Jutorán (2018), recordings were provided by 10 criminal courts, but the sample was still random. It would have been better to have the recordings of all 12 courts, but we still believe that the sample can be claimed to be representative, since it amounts to 190,000 tokens from 55 hearings, interpreted by 45 different interpreters — or, rather, since it is an oral corpus, 1,116 minutes — and it includes 41.6% of the total “population”. It is also important to add that all the courts are very similar as regards the number of interpreted proceedings, the language combinations concerned and interpreters’ background and experience, with the latter all coming from the same agency, which has an official agreement in place as an interpreting provider.

Naturally, it would be very interesting and necessary in the case of extrapolation to the whole of Spain to replicate the study and see to what extent the results obtained are corroborated. The research team plans to do it in Seville, the fourth Spanish city in terms of population, and is currently taking all the requisite steps to be able to develop another full-scale study using the same procedure to compile, transcribe and analyse the corpus.

Therefore, the final corpus consists of the transcription of all the trials with interpreters that took place between January and June 2015 in three language combinations (English, French and Romanian into Spanish) and in 10 criminal courts in Barcelona. It included 1,116 minutes of trial recordings which, once transcribed, amount to 190,000 tokens, including all the monologic and dialogic parts of the trial that are audible (since some parts were interpreted using the “chuchotage” technique, far from the microphones of the courtroom and were thus not audible enough to allow transcription).

5. Operationalisation of the dependent variable *textual problems*

The first dependent variable was **textual problems** and the phenomenon to be observed and measured was the accuracy of the information transfer, i.e. the message conveyed by the interpreter. By problems, the researchers meant “areas” where interpreters encountered linguistic, cultural, or domain-related (for instance legal) difficulties in the oral discourse. The term “linguistic” is understood here in the broader sense of this word,

including not only textual, syntactic and lexical levels, but also the pragmatic level, i.e. problems posed by register, tone or changes in the discourse and so on.

To measure this variable, two scales were created. The first of these was an interval scale regarding accuracy. Interval scales rank concepts and set the difference between them, just like the grading system that is used to evaluate students. In this case, the interval scale had three categories, shown in Figure 1.

The solution applied by the interpreter when facing a textual problem was:
- (A) Adequate
- (M) Improvable
- (I) Inadequate

Figure 1. The interval scale created to measure accuracy in relation to textual problems.

The definition of the three categories included in the scale is as follows. An “adequate” solution means that the content and the form of the message is conveyed “adequately”, i.e. precisely and accurately, by the interpreter. An “improvable” solution means that the interpreter conveys the message and the basic communicative objective is reached, but not in a complete way, so the solution could be clearly improved either in content or in form. An “inadequate solution” means that there is a serious distortion of meaning in the message conveyed. This may be due to several possible error types, such as serious omissions or additions, shift in meaning and so on, and that is why a second, descriptive scale was created to complement the data obtained from the first scale.

The second scale is categorical. Also called nominal data, categorical scales classify concepts without ranking them, so one category is not better or worse than any other. This scale is related to the usual accuracy scale, and was created to quantify the type of solutions adopted by the interpreters to solve a textual problem. The indicators are shown in Figure 2.

Types of solution applied by interpreters when facing a textual problem
Possible categories for “adequate” solutions: - (EH) Established equivalent - (IM) Making some information implicit - (EX) Making some information explicit
Possible categories for “improvable” solutions: - (CR) Change of register - (NMS) Minor shift in meaning (compared to the source text)
Possible categories for “inadequate” solutions: - (O) Omission - (OG) Serious omission - (NT) Not translated - (AD) Addition of information - (ADG) Serious addition of information - (ITER) Inadequate terminology - (FS) Major shift in meaning (substantial distortion of meaning from that of the original message) - (FSG) Serious major shift in meaning - (SS) Incomprehensible (message is not understandable, does not make sense)

Figure 2. The categorical scale created to quantify types of textual solutions.

These categories are explained and illustrated in detail in Orozco-Jutorán (2017). What should be emphasised is that a distinction was made between “serious” inadequate solutions (i.e. serious addition of information, serious omission, serious major shift in meaning) and other, “less serious” inadequate solutions. By “serious” we mean errors that might affect or interfere with the result of the criminal proceeding, as shown in example 1, where we have included our translation of the Spanish oral interventions between square brackets and where the serious addition of information is underlined.

<p>Judge- ... <i>que si reconoce los hechos y está conforme.</i> [Does he acknowledge the facts and agree?] Interpreter- <i>Do you accept?</i> Defendant- <i>Sí.</i> [Yes] Interpreter- <i>Yeah? And do you agree?</i> Defendant- <i>Yeah.</i> Interpreter- <i>Sí, <u>es culpable.</u></i> [Yes, <u>he is guilty</u>]</p>
--

Example 1. An example of serious addition of information.

6. Operationalisation of the dependent variable *interaction problems*

The second dependent variable was **interaction problems** and the phenomena to be observed were related to the oral interaction conducted by the participants in the criminal proceedings: judge, counsels, prosecutor, interpreter, defendant, witnesses, etc. This variable includes three aspects of the oral interaction: conversation management, non-renditions (as defined by Wadensjö 1998, p.25 and direct or reported speech, corresponding to three different scales created. Figure 3 shows the categorical scale used to observe conversation management.

Types of conversation management problems:
- (S) Overlap
- (I) Interruption
- (DL) Long turn

Figure 3. The categorical scale created to quantify types of conversation management problems

These three categories, based on the existing literature on court interpreting (see Angermeyer 2015), include: (i) overlaps: that is, when two or more members of the judicial staff speak at the same time, or when the interpreter’s voice overlaps with that of the judge or any other judicial staff, causing the latter to stop talking; (ii) interruptions: that is, when the interpreter is interrupted by any member of the judicial staff, leaving his or her rendition unfinished; and (iii) long turns: that is, when a member of the judicial staff speaks for more than two minutes in a single turn.

The second instrument was designed to observe non-renditions, that is, renditions that do not correspond to the original utterance. These can be used to manage turn-taking or dialogue and are then justified, but may also be used for other reasons which are not justified. Therefore, as in the case of textual problems, an initial interval scale measures acceptable non-renditions (justified) and non-acceptable non-renditions (unjustified) and then a second, categorical scale classifies possible types of non-renditions, as shown in Figure 4.

The types of non-renditions by the interpreter (I.)
<p>Possible categories for justified non-renditions:</p> <ul style="list-style-type: none"> - (P) Pause (I. asks for a pause to be able to interpret) - (Cl) Clarification (I. asks for clarification or explains something that was expressed ambiguously) - (Co) Confirmation (I. seeks to confirm that s/he understood or heard the information clearly) - (R) Retrieval (I. is aware that he or she is missing some information and asks to retrieve it)
<p>Possible categories for unjustified non-renditions:</p> <ul style="list-style-type: none"> - (A) Warning (I. gives advice or instructions on how to behave or warns the defendant) - (Res) Answer (I. answers on behalf of the defendant) - (Extra) Extra information (I. gives information to any of the participants or asks questions not posed in the original utterances).

Figure 4. The categorical scale created to quantify types of non-renditions.

Finally, the third instrument regarding interactional problems was designed to observe speech style, both by the judicial staff and the interpreter. The purpose of including this information in the observations was to see if participants were consistent with their own style during a trial. The categorical scale shown in Figure 5 was thus applied twice in every trial observed, once for the judicial staff and a second time for interpreters.

Possible nominal categories for speech style:
<p>(DIR) Direct speech</p> <p>Example: Defendant: No, I wasn't there. Interpreter: No, no estuve allí. [<i>No, I wasn't there.</i>]</p>
<p>(INDIR) Indirect speech</p> <p>Example: Defendant: No, I wasn't there. Interpreter: No, no estuvo allí. [<i>No, he wasn't there.</i>]</p>
<p>(RS) Reported speech</p> <p>Example: Defendant: No, I wasn't there. Interpreter: <u>Dice que</u> no estuvo allí. [<i>He says that he wasn't there.</i>]</p>

Figure 5. The categorical scale created to quantify the speech style.

The categories are explained in detail in Arumí and Vargas-Urpí (2018, *forthcoming*).

7. Validation of measurement instruments through a pilot study

Our intention was to create a measuring instrument for court interpreting quality, with a focus on assessing whether the rights of accused persons are respected. For this purpose, we ensured the measurement validity, which “refers to the techniques we use to acquire our research data and to the appropriateness of the scales we use to measure that data” (Saldanha and O’Brien 2013, p. 33), by creating and describing the indicators with care and ensuring that the scales were consistent and could be applied to real situations (i.e. the oral corpus with interpreters’ intervention in criminal proceedings).

To ensure the validity, after the scales and indicators were created and described, a pilot study was carried out with a subcorpus of 18 trials — out of the 55 trials that made up the whole sample to be analysed — 6 in each of the three language combinations (English,

French and Romanian into Spanish), in which different experts used the same instruments and annotated the same trials to check if the results obtained were the same. Where differences were found, some changes were made to the instruments to ensure consistency in the results. Other pilot study results are explained in Arumí and Vargas-Urpí (2018, *forthcoming*) and Vargas-Urpí (2017).

The measuring instruments take the form of an annotation system that could be used by filling in a grid or, as was done in the TIPp project, annotating the corpus in a software package that allows such annotation and then exporting this information into Excel files in which the indicators can be quantified.

To give an example of what an annotated file looks like, Figure 6 shows a screenshot of the EXMARaLDA software^{ix}, with the transcription and annotation of a small fragment of a trial.

	38 [136.0]	89 [138.0]	90 [140.1]	91 [140.9]	92 [142.0]
J [v]					
I1 [v]	Es mentira. No estaban allí.		The police arrested you that very moment.	The police caught you.	
I1 [MI]					
I1 [Retrad_I1]					
A1 [v]					
A1 [Retrad_A1]					
A2 [v]					
A3 [v]					
A4 [v]					
SPK35 [A5]					
F [v]		Eh, la policía las detuvo en ese momento.			
LD1 [v]					
LD2 [v]					
LD3 [v]					
LD4 [v]					
LAP1 [v]					
[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 6. Fragment of a trial transcribed and annotated using the measuring instruments.

As can be seen in Figure 6, one tier or row is devoted to each type of problems, both textual and interaction problems. In the example, at the top of the screen, there are all the

tiers or rows devoted to the speakers and the transcription of what they said. Below these rows, starting in tier 17, the annotation tiers can be seen, the first of which is called “PROBLEMA”. This tier is where the researchers tag the fragment in which there is a textual or interaction problem. For instance, in the first grey column, below where the interpreter says *Es mentira. No estaban allí* [That is a lie, they weren’t there], there is an “I”, meaning that there is an “interaction problem” in this sentence. Then, some 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 the interpreter does not face any problem in this case. In the next column, the interpreter renders the prosecutor’s words but starts speaking before the prosecutor finishes 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* in the tier belonging to *SOLAPAMIENTO*, which means “overlap” in Spanish. This *SOI* stands for “an overlap caused by the interpreter”, and is differentiated from an overlap between the Judge and the prosecutor or the defence attorney, which would be annotated as *SOJ*. There are two more annotations in the same sentence. The first one is not an interaction problem but an observable phenomenon in the interaction (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 the interpreter is using direct speech, as would be recommended in this case. The second annotation is of a textual nature, which is why, next to the “I” and the “F” in the *PROBLEMA* tier there is also an “S” (meaning “Solution”). In the tier just below this one, there is an annotation “A”, meaning “adequate solution”. Finally, in the tier below the “A”, there is the specification of the type of solution applied by the interpreter to the textual problem, in this case, *EH*, which stands for “Equivalente Habitual” [Established equivalent].

All the information annotated in the corpus as explained above, for the 55 trials, was converted into Excel files, an example of which can be seen in Figure 7.

	A	B	C	D	E	F	G	H	
1	PROBLEMA	SOLUCION	TIPO_SOLUCION	SOLAPAMIE	INTER	DISCURS	VOZ_PROPIA	TIPO_VOZ_PROPIA	ESTILO_OPERAD
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 7. Details of an Excel file with annotations.

As Figure 7 shows, the rows or tiers from the EXMARaLDA software were converted into columns, so that filters and formulas could be applied to obtain quantifiable data like that shown in the Findings section. One Excel sheet was created for each trial and one

Excel file containing all the trials in one language combination. Finally, a “bigger” Excel file was created, linked to the three sheets, containing the total data for each language and combining the results of the three language pairs that were analysed. This system proved to be very useful because it provided researchers with quantifiable data that enabled them to describe real practice systematically rather than anecdotally.

The pilot test and the analysis of the whole corpus with the measuring instruments showed that the differences between researchers when using the measuring instrument would be very small. Of course, there is always some room for interpretation on the part of the person assessing or using the measuring instruments. However, that never happened when dealing with indicators like “adequate” or “improvable”, but rather with allocating items to the different inadequate solution types. Therefore, it would be very useful to have other researchers use the measuring instruments and replicate the study as the way to move forward, both in terms of research methodology and scientific gain for TIS.

8. Results

8.1. How much of the trial is interpreted?

One important finding of the corpus analysis is that a substantial part of the trial is not actually translated for persons with limited competence in the official languages of the court, who is usually the defendant. This is measured by one of the categories created under “inadequate textual solutions”: “not translated” (NT). To be annotated as NT, there needs to be a whole intervention, not only a word or a sentence, by one of the participants which has not been translated at all, so there is an important difference with respect to omissions, which affect only a word or a sentence that has not been translated. Table 1 shows the quantity of NT interventions found per hour and per minute in the corpus.

Language	Total of NT per hour	Total of NT per minute
English	371	1.8
French	190	1.6
Romanian	555	3.7
Mean	372	2.7

Table 1. The number of “Not translated” interventions (NT) in the corpus per hour and per minute of trial.

Other data regarding the parts of the trial which are not interpreted are also alarming, as shown in Table 2 (for more details see Vigier (*forthcoming*)).

Language	% of interpreted time (aloud)	% of interpreted time (chuchotage)	% of time not interpreted in any way
English	33	16	51
French	48	13	39
Romanian	22	17	61
Mean	30	16	54

Table 2. The percentage of time interpreted to defendants in 55 hearings covered by the corpus.

The mean of the percentage of the trial which is not interpreted to the defendant in any way is 54%, and only 33% of the interpretation is recorded, since the other 16% is

whispered to the ear of the defendant and is thus not heard or recorded by the microphones in the courtroom. These findings demonstrate that the defendant's right of information is being violated.

How accurate is the translation of the interpreted part of the trial?

The next variable is the accuracy of interpreting. Table 3 shows the quantity of adequate, improvable and inadequate solutions observed in the corpus per minute of trial.

Language	Adequate solutions	Improvable solutions	Inadequate solutions
English	0.49/min	0.45/min	2.14/min
French	0.07/min	0.36/min	2.07/min
Romanian	0.44/min	2.77/min	3.99/min
Mean	0.3/min	1.2/min	2.7/min

Table 3. The number of textual solutions in the corpus per bilingual minute of trial.

As can be seen in Table 3, the mean of inadequate solutions is almost three per minute while the mean of adequate solutions is one every three minutes, which implies that the interpreting is not accurate. The differences between the three language combinations observed in Table 3 and also in other tables in this section are significant, with interpreting into and from Romanian containing almost double of inadequate solutions that interpreting from and into English or French, for instance, suggesting that this is an important factor to be accounted for in future research.

Table 4 describes the number of serious inadequate solutions in the corpus, which is alarmingly large, because the mean of 21.1 serious errors per hour implies that there is one serious error, which could affect the result of criminal proceedings, every three minutes. The issue is not merely a lack of precision, but serious errors in the translation of the messages, which again violates the defendant's right of information.

Language	Serious omissions per bilingual hour	Serious addition of information per bilingual hour	Serious major shift of meanings per bilingual hour	Incomprehensible sentences (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
Mean	8.5	3.2	7.1	2.3	21.1

Table 4. The number of serious errors in the corpus per bilingual hour of trial.

8.2. How do participants interact during trials?

Table 5 shows the number of conversation management problems during the trial, namely overlaps, interruptions and long turns per hour of trial, while Table 6 shows the number and type of non-renditions per hour of trial.

Language	Judicial staff overlaps per hour	Interpreter overlaps per hour	Long turns per hour	Interruptions per hour
English	18.9	36.1	0.81	24.4
French	17.1	45.7	0.63	48.9
Romanian	11.7	18.4	1.5	6.7
Mean	15.0	32.2	1.1	24.6

Table 5. The number of conversation management problems per hour of trial.

Language	Justified non-renditions	Unjustified non-renditions
English	26.3	50.2
French	14.4	11.9
Romanian	25.6	65.7
Mean	22.8	45.5

Table 6. The number and type of non-renditions per hour of trial.

The number of overlaps between judicial staff, interruptions and long turns confirms the complexity of interpreters' work. This is also supported by other data related to the speech speed of the judicial staff: the researchers counted the trials in which one of the judicial staff exceeded 180 words per minute and it happened in 72% of the trials. However, this complexity does not explain or justify the alarming number of unjustified non-renditions observed, again with considerable differences between the three language combinations, with the interpreter advising, instructing, warning or asking questions of his/her own three times every four minutes.

In respect of the speech style, Table 7 shows the lack of consistency observed in the judicial staff and interpreters when using direct, indirect or reported speech in the same trial. The inconsistencies were widespread in both groups, which adversely affects the clarity of interaction between the participants of criminal proceedings.

Language	Lack of consistency (interpreters)	Lack of consistency (judicial staff)
English	74%	74%
French	67%	67%
Romanian	63%	74%
Mean	67%	73%

Table 7. The percentage of the lack of consistency observed in the speech style of judicial staff and interpreters in each trial.

9. Discussion

The findings indicate that there is an unacceptably low quality of court interpreting in the random sample of 55 criminal proceedings observed in Barcelona's criminal courts in 2015 in three language combinations. The alarming numbers of errors and unjustified

non-renditions suggest that interpreters are not adequately trained and do not follow a code of ethics, and that the judicial staff need to receive awareness-raising information or training to interact better with interpreters. The findings also show that the defendants' right of information is being systematically violated, firstly because less than half of the trial is being translated for them and secondly because the part that is interpreted has an unacceptable number of errors that can affect the result of criminal proceedings.

We hope that these findings shed light on the problems currently facing the quality of court interpreting in criminal proceedings in Spain. We would also like to contribute to the improvement of the quality of court interpreting by creating some tools, which can be found online and include translation-oriented terminological records, thesaurus and recommendations to both court interpreters and judicial staff, based on the observations made on the corpus.

10. Conclusions

This chapter has demonstrated a mixed-methods approach to examining the quality of court interpreting, involving a qualitative overall design with quantitative data analysis techniques and the creation of a measuring instrument. The objective was to operationalise the quality of court interpreting, which was achieved through a range of variables and the creation of measurable, quantifiable indicators for each variable. Departing from the theoretical distinction made by Wadensjö (1998) between 'talk-as-text' and 'talk-as-activity', two dependent variables were defined: textual problems and interaction problems. These variables were then operationalised through a range of interval and categorical scales created *ad hoc* and piloted using the data from the corpus compiled. To be more specific, two scales were created to measure the textual problems variable: one interval scale to measure accuracy of the information transfer and one categorical scale to quantify types of solutions applied by the interpreter. Regarding the interaction problems variable, three categorical scales were created: one to quantify types of conversation management problems, another to quantify types of non-renditions and a third one to quantify the speech style.

The proposed method should be further tested by replication in other contexts (other countries, other language pairs, court types, civil proceedings, etc.) to help us better understand the nature of constraints in court interpreting and, in the long term, to be able to gather comparable data in different contexts.

References

- Ahmad, K., and Rogers, M., 2001. Corpus linguistics and terminology extraction. *In*: S.E. Wright and G. Budin, eds. *Handbook of Terminology Management*. Amsterdam: John Benjamins, 725-760.
- Angermeyer, P.S., 2015. *Speak English or What? Codeswitching and Interpreter Use in New York City Courts*. New York: Oxford University Press.
- Arumí, M., and Vargas-Urpí, M., 2018, *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 (3).

- Bestué, C., 2018. Aproximación empírica a la labor del intérprete en los tribunales de justicia. In: M.J. Ariza Colmenarejo, ed. *Derecho de Traducción Interpretación e Información en el proceso penal*. Valencia: Tirant lo Blanch, 139-175.
- Biel, Ł., 2017. Researching Legal Translation: A multi-perspective and mixed-method framework for legal translation. *Revista de Llengua i Dret, Journal of Language and Law* 68, 76-88.
- Calzada Pérez, M., 2017. Corpus-based methods for Comparative Translation and Interpreting Studies. Mapping differences and similarities with traditional and innovative tools. *Translation and Interpreting Studies* 12 (2), 231-252.
- Corpas Pastor, G., and Seghiri, M., 2010. Size matters: A quantitative approach to corpus representativeness. In: R. Rabadán, T. Guzmán and M. Fernández, eds. *Lengua, traducción, recepción: en honor de Julio César Santoyo / Language, translation, reception : to honor Julio César Santoyo*. León: Universidad de León, 111-145.
- Creswell, J.W., 1998. *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*. Thousand Oaks, CA: Sage.
- Hale, S., 2004. *The Discourse of Court Interpreting*. Amsterdam: John Benjamins.
- Hale, S., 2007. *Community Interpreting*. Hampshire: Palgrave Macmillan.
- Mikkelsen, H., 1998. Towards a redefinition of the role of the court interpreter. *Interpreting* 3 (1), 21-45.
- O'Reilly, M. and Kiyimba, N., 2015. *Advanced Qualitative Research. A Guide to Using Theory*. London: Sage.
- Ormstorm, R., et al., 2014. The foundations of qualitative research. In: J. Ritchie, et al., eds. *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. London, Sage, 1-26.
- 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., 2018. The TIPp project: developing technological resources based on the exploitation of oral corpora to improve court interpreting. *Intralinea* 20. Available from: http://www.intralinea.org/specials/article/the_tipp_project [Accessed 5 Sept 2018].
- Ortega Herráez, J.M., 2011. *Interpretar para la Justicia*. Granada: Comares.
- Pöchhacker, F., ed., 2015. *Routledge Encyclopedia of Interpreting Studies*. London: Routledge.
- Pöchhacker, F., 2016. *Introducing Interpreting Studies*. London: Routledge.
- Pontrandolfo, G., 2016. *Fraseología y lenguaje judicial. Las sentencias penales desde una perspectiva contrastiva*. Roma: Aracne.
- Roberts-Smith, L., 2009. Forensic interpreting. Trial and error. In: S. Hale, U. Ozolins and L. Stern, eds. *The Critical Link 5: Quality in interpreting – a shared responsibility*. Amsterdam: John Benjamins, 13–35.
- Saldanha, G., and O'Brien, S., 2013. *Research Methodologies in Translation Studies*. London: Routledge.
- Saukko, P., 2003. *Doing Research in Cultural Studies: An Introduction to Classical and New Methodological Approaches*. London: Sage.
- Tipton, R., and Furmanek, O., 2016. *Dialogue Interpreting. A Guide to Interpreting in Public Services and the Community*. London: Routledge.
- Vargas-Urpí, M., 2017. Court Interpreting as a Shared Responsibility: Judges and Lawyers in a Corpus of Interpreted Criminal Proceedings. *Revista Canaria de Estudios Ingleses* 75, 139-154.

- Vigier, F.J., *forthcoming*. Interpreting at Spanish criminal courts: preliminary results of the TIPp project. *Translation and Translanguaging in Multilingual Contexts*.
- Vigier, F.J. and Sánchez, María del Mar (2017). Using Parallel Corpora to Study the Translation of Legal System-Bound Terms: The Case of Names of English and Spanish Courts. In: R. Mitkov, ed. *Computational and Corpus-Based Phraseology. Second International Conference, Europhras 2017 Proceedings*. Cham: Springer, 260-273.
- Wadensjö, C., 1998. *Interpreting as Interaction*. New York: Longman.
- Wright, S.E. and Budin, G., 1997. *Handbook of Terminology Management*. Amsterdam: John Benjamins.

ⁱ Our translation of a sentence taken from 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]

ⁱⁱ The research team was composed of the following researchers: Marta Arumí, Anna Gil Bardají (Universitat Autònoma de Barcelona), Anabel Borja (Universitat Jaume I), Mireia Vargas-Urpí (Universitat Pompeu Fabra) Francisco Vigier (Universidad Pablo de Olavide) and two team leaders: Carmen Bestué and Mariana Orozco-Jutorán (Universitat Autònoma de Barcelona).

ⁱⁱⁱ The name of the project was “Quality in translation as an element to safeguard procedural guarantees in criminal proceedings: development of resources to help court interpreters of Spanish - Romanian, Arabic, Chinese, French and English (FFI2014-55029-R). It lasted three years from January 2015 to December 2017 and it received funding from the Spanish State Grant Programme of Research, Development and Innovation related to Social Challenges.

^{iv} The whole process and the difficulties faced by researchers in designing, creating, compiling and transcribing the corpus are explained in Orozco-Jutorán 2018.

^v The list of references would be too long to be included here. See for instance Hale 2004 and 2007, Mikkelsen 1998, Tipton and Furmanek 2016, Ortega Herráez 2011 and Pöchhacker (ed.) 2015.

^{vi} “(...) Second, the claimant of the right must show, assuming it is not a case of a complete denial of an interpreter but one involving some alleged deficiency in the interpretation actually provided, that there has been a departure from the basic, constitutionally guaranteed standard of interpretation. For the purposes of this appeal, I define this standard as one of continuity, precision, impartiality, competency and contemporaneousness” *R. v. Tran*, [1994] 2 S.C.R. 951 [https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/1166/index.do]

^{vii} The first city in terms of population is Madrid, with 3,182,981 inhabitants, followed by Barcelona, with 1,620,809 inhabitants and then Valencia with 78,808 and Seville with 689,434 inhabitants. Source: National Institute of Statistics of Spain [http://www.ine.es/] accessed in January 2018, data from January 2017.

^{viii} The migrant population in Barcelona is 709,557. It is the second largest Spanish city, after Madrid, with 793,513 inhabitants. Source: National Institute of Statistics of Spain [http://www.ine.es/] accessed in January 2018, data from January 2017.

^{ix} <http://www.exmaralda.org/en/>. We would like to express our gratitude to Bernd Meyer, who suggested the use of this software, and Thomas Schmidt, the developer of the software package, who helped us with the technical aspects of the software.