TIPp Traducción e Interpretación en los Procesos penales

'Old' data through new lens

using the ComInDat corpus for new research purposes

MIRAS
Mediació i Interpretació:
Recerca en l'àmbit social

Mireia Vargas-Urpi
Universitat Pompeu Fabra

'Old' data through new lens

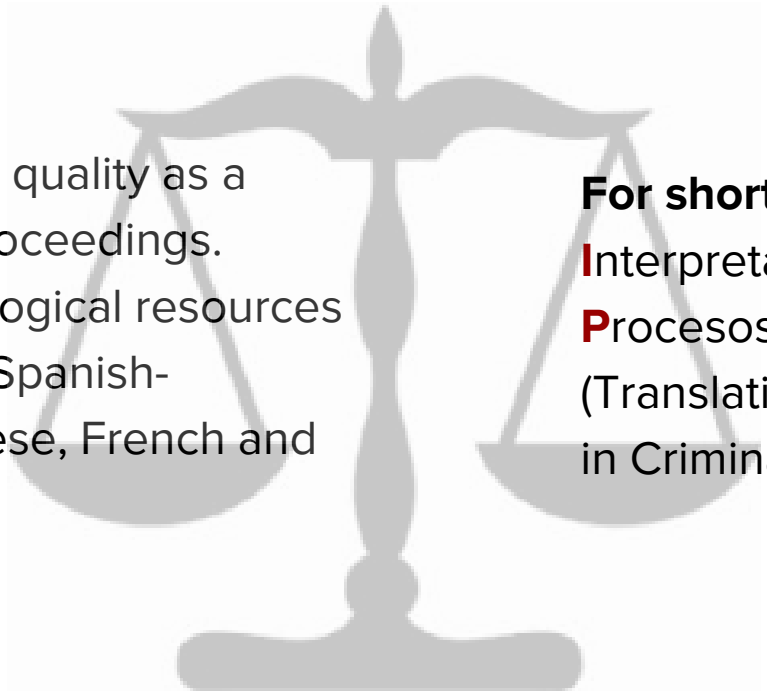
The TIPp project + **The ComInDat corpus**

How can we exploit existing data in the preparation and development of new research projects?

TIPp Traducción e Interpretación en los Procesos penales

The TIPp project

Official title: 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.



For short: **T**raducción e **I**nterpretación en los **P**rocesos **p**enales (Translation and Interpreting in Criminal Proceedings).

http://pagines.uab.cat/tipp/en

Opportunity

a) The Ministry of Economy and Competitiveness has funded our research (FFI2014-55029-R).
b) The Court of Justice of Catalonia (TSJC) has granted us access to recordings of criminal proceedings so we can actually build an oral corpus of real-life interpreting in criminal proceedings.
c) A group of judges is interested in our research and advice.

TIPp objective

To create a computer application which can include in only one interface all the necessary resources to facilitate court interpreters' performance:

1. A set of guidelines to describe which **strategies or translation techniques** can be used in which situations
2. A **protocol for conduct** and behaviour in the most frequent situations for a court interpreter
3. A set of **guidelines for Justice personnel** on interpreters' role and on how to interact with interpreters
4. A **database containing the terms** which are most frequently used in criminal proceedings with comments and two-way translation options in the most frequently translated languages.

Tiers used in the annotation of the transcriptions

"talk as text, talk as interaction"
((Wadensjö, 1998)

1. Problem: textual, interactional or both

Textual problems:

2. Textual solution: adequate, inadequate, improvable
3. Type of textual solution: various possible categories (common equivalent, neutralisation, loan, etc.)

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

Tiers used in the annotation: interactional problems

4. Problems related to conversation management (CM): overlapping, interruptions, long turns, fast speech rate, etc.
5. CM solution: adequate, inadequate, improvable
6. Type of CM strategy: non-rendition, summarised rendition, note-taking, *chuchotage*...

Purpose of 'strategic' non-renditions:

- To ask for a pause
- To ask for repetition
- To ask for clarification
- To seek confirmation of information

Tiers used in the annotation

Description of other relevant aspects of the interaction

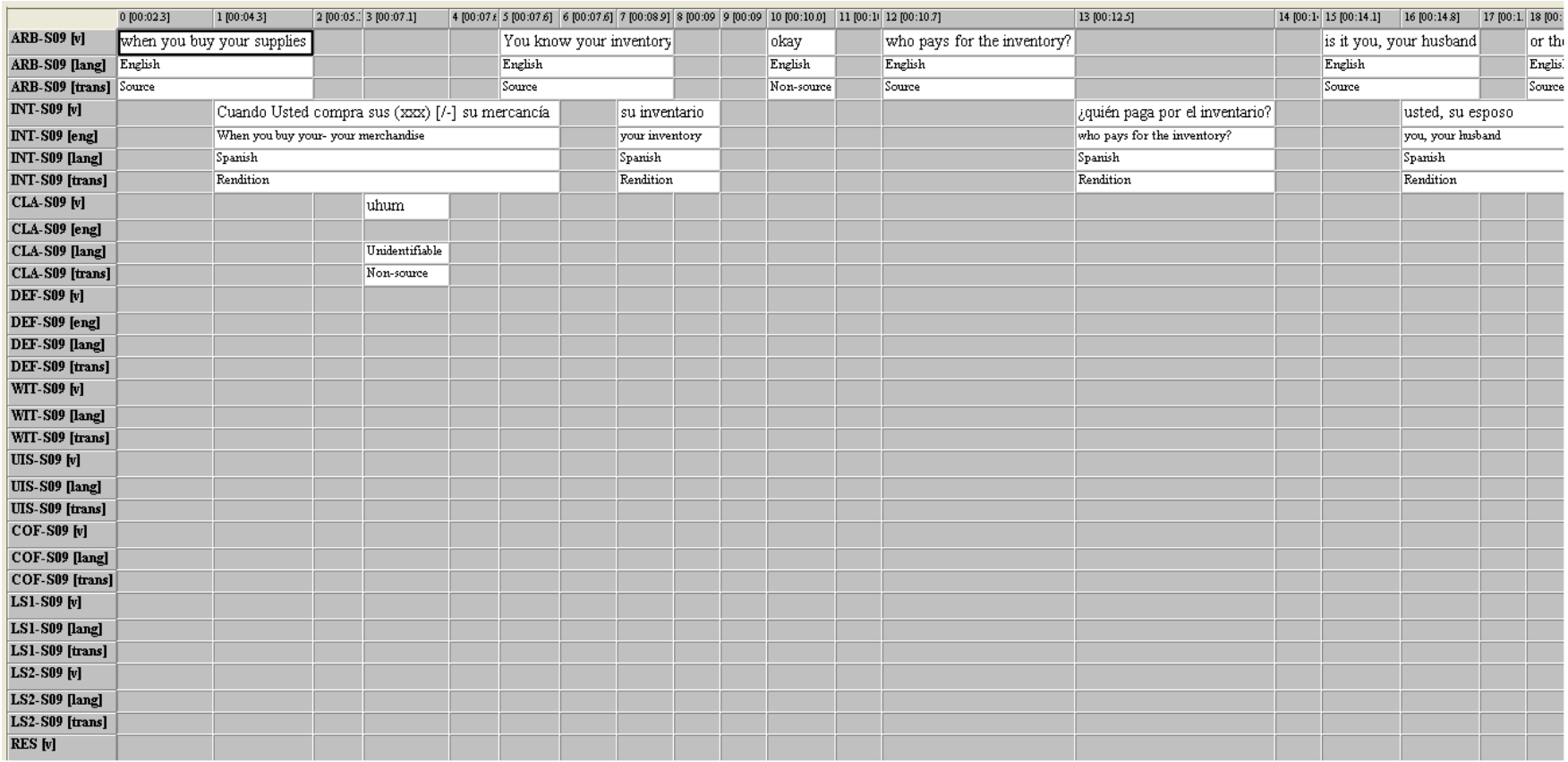
7. Other kinds of non-renditions:
    a. Reactive tokens: Yes, your honour
    b. To give advice to the user (defendant, witness) or warn him/her
    c. To answer on behalf of the user
    d. To ask for 'extra' information to the user
8. Direct or indirect style in judges' and lawyers' turns
9. Interpreters' style (direct, indirect, reported speech)
10. Other problems related to interpreters' code of ethics

The ComInDat pilot corpus

The ComInDat pilot corpus contains sample data from three different projects:

- the DiK corpus of Portuguese/German and Turkish/German interpreted doctor-patient communication in hospitals (Bührig & Meyer 2004),
- the liSCC-corpus, a corpus of interpreted court proceedings in different language constellations (Spanish/English, Russian/English, Haitian Creole/English and Polish/English) (Angermeyer 2006),
- a corpus of simulated interpreted doctor-patient interactions in different language constel-lations (Russian/German, Polish/German and Romanian/German) from a training seminar for bilingual nursing staff ("SimDiK", Bührig, Kliche, Meyer & Pawlack 2012).

Example from the NYSCC subcorpus (Angermeyer, 2006)



	0 [00:02.3]	1 [00:04.3]	2 [00:05..	3 [00:07.1]	4 [00:07.6	5 [00:07.6]	6 [00:07.6]	7 [00:08.9]	8 [00:09	9 [00:09	10 [00:10.0]	11 [00:10	12 [00:10.7]	13 [00:12.5]	14 [00:1	15 [00:14.1]	16 [00:14.8]	17 [00:1.	18 [00:
ARB-S09 [v]	when you buy your supplies					You know your inventory					okay		who pays for the inventory?			is it you, your husband			or th
ARB-S09 [lang]	English					English					English		English			English			Englis
ARB-S09 [trans]	Source					Source					Non-source		Source			Source			Source
INT-S09 [v]		Cuando Usted compra sus (xxx) [/-] su mercancía						su inventario						¿quién paga por el inventario?			usted, su esposo		
INT-S09 [eng]		When you buy your- your merchandise						your inventory						who pays for the inventory?			you, your husband		
INT-S09 [lang]		Spanish						Spanish						Spanish			Spanish		
INT-S09 [trans]		Rendition						Rendition						Rendition			Rendition		
CLA-S09 [v]				uhum															
CLA-S09 [eng]																			
CLA-S09 [lang]				Unidentifiable															
CLA-S09 [trans]				Non-source															
DEF-S09 [v]																			
DEF-S09 [eng]																			
DEF-S09 [lang]																			
DEF-S09 [trans]																			
WIT-S09 [v]																			
WIT-S09 [lang]																			
WIT-S09 [trans]																			
UIS-S09 [v]																			
UIS-S09 [lang]																			
UIS-S09 [trans]																			
COF-S09 [v]																			
COF-S09 [lang]																			
COF-S09 [trans]																			
LS1-S09 [v]																			
LS1-S09 [lang]																			
LS1-S09 [trans]																			
LS2-S09 [v]																			
LS2-S09 [lang]																			
LS2-S09 [trans]																			
RES [v]																			

	1	12 [00:10.7]	13 [00:12.5]
ARB-S09 [v]		who pays for the inventory?	
INT-S09 [v]			¿quién paga por el inventario?
INT-S09 [eng]			who pays for the inventory?
CLA-S09 [v]			
CLA-S09 [eng]			
DEF-S09 [v]			
DEF-S09 [eng]			
WIT-S09 [v]			
UIS-S09 [v]			
COF-S09 [v]			
LS1-S09 [v]			
LS2-S09 [v]			
RES [v]			
[PROBLEM]		T	T
[TEXTUAL SOLUTION]			inadequate
[TYPE OF TEXTUAL SOLUTION]			inaccurate term
[TYPE OF CM PROBLEM]			
[CM SOLUTION]			
[TYPE OF CM STRATEGY]			
[NON-RENDITION]			
[PURPOSE OF NON-RENDITION]			
[STYLE (JUSTICE PERSONNEL)]			
INT-S09 [STYLE]			
[CODE OF ETHICS]			
[INAUDIBLE]			
[COMMENTS]			

annotation tiers that various researchers will use

Conclusions (I): Usefulness of this analysis

Existing data may be useful in pilot studies to inform decisions on different aspects concerning the **method**:

- Is the transcription system chosen suitable for the purposes of the study?
- Are the annotation tiers relevant and feasible considering the objectives of the study?
- Do we need any other kind of information in the metadata of each transcribed interaction?

Conclusions (II): comparison of corpora

The results of the analysis of data of a similar nature (recordings of court interpreted interactions) but collected in different moments and geographical locations (NY - BCN) may be compared in order to draw new conclusions. In this specific case:

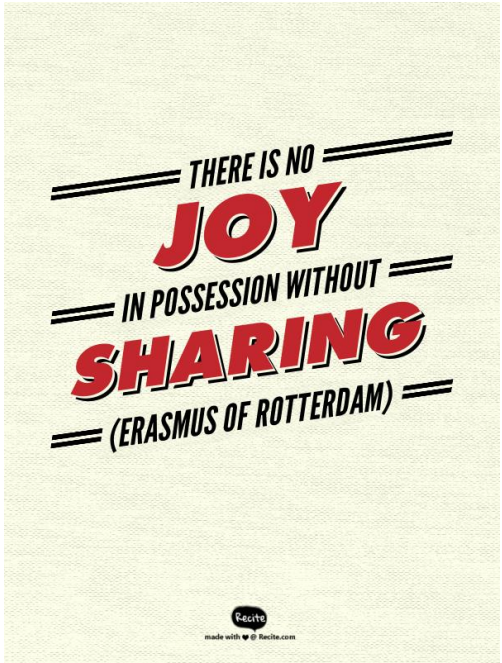
- **NYSCC corpus**: predominance of ‘close renditions’, use of short consecutive (real liaison interpreting), trained and certified interpreters
- **TIPp corpus**: interpreters do not usually interpret everything (‘reduced renditions’, ‘zero renditions’), interpreters often summarise long turns, lack of certification programmes

Conclusions (III): Limitations

- Certain research questions are tightly related to research object/reality, therefore, they are difficult to apply to other contexts or transcription methods.
  - **Contexts**: more overlap in the NYSCC corpus, more long turns in the TIPp corpus; codeswitching related to terminology in NYSCC corpus, less codeswitching in the TIPp project; direct vs. indirect speech [strategy, codes of conduct, guidelines]
  - **Transcription methods**: did the interpreter take notes? [strategy]; how did the interpreter use her nonverbal communication, i.e. gestures, facial expressions, etc.? [strategy, code of conduct]

Conclusions (IV): Sharing corpora

- Transcribing is perhaps the most time-consuming phase in a study. Once the transcription is done, annotation is, in comparison, fairly quick.
- Enormous potential for future comparative studies if more data are made available.



more info coming soon!

http://pagines.uab.cat/tipp/

mireia.vargas@upf.edu

https://twitter.com/miras_uab

https://twitter.com/mireiavu