

COMINDAT

The Community Interpreting Database

Philipp Sebastian Angermeyer
York University

Bernd Meyer
Mainz University

Thomas Schmidt
Leibniz-Institut für Deutsche Sprache

1. General info

The Community Interpreting Database provides transcripts of audio and video recordings of various types of community interpreted discourse (doctor-patient communication, simulated doctor-patient communication, courtroom communication) in German (simulated and authentic doctor-patient communication) and US (courtroom communication) institutions with varying community languages. Video recordings only exist for the simulated communication. For authentic interpreted doctor-patient communication, no audio files will be made available. Languages used in the data: German, English, Spanish, Turkish, Polish, Portuguese, Romanian, Russian, Haitian Creole. Corpus access is provided for academic purposes.

2. Community Interpreting

Community Interpreting is bilateral consecutive interpreting, done by professional or lay interpreters in conversations between migrant clients and agents from governmental, social service, or medical institutions.

3. Linguistic approaches to the study of community interpreting

Linguistic approaches often focus on different participation formats due to the presence of a third person (the interpreter). Other research topics include the translation of specialized language, code-switching, and the handling of language barriers.

4. Data collection and transcription

The data were recorded from authentic interactions in legal and medical settings. A video-taped sub-corpus stems from role plays on interpreter-mediated doctor-patient communication. Data were originally transcribed using different digital tools (ELAN, EXMARALDA) and transcription conventions (CHAT, HIAT).

	3 [00:18.6]	4 [00:21.6]	5 [00:22.3]	6 [00:22.8]	7 [00:24.0]	8 [00:24.5]	9 [00:25.5]	10 [00:26.3]
A-20 [v]	Nein, • darüber sprechen, das Ergebnis • besprechen.	((0,4 s)) Kannst du	schon...					
A-20 [eng]	No, we just want to talk about it and discuss the results.	Go ahead...						
A-20 [lang]	German	German						
A-20 [trans]	Source	Non-source						
D-20 [v]			„Ela,	ela tá a dizer ((takes breath))	eh	que quer •	ehm	dizer • o
D-20 [deu]			Sie, sie sagt, äh dass sie ähm sagen will, was...					
D-20 [eng]			She is saying eh she wants to say that what is...					
D-20 [lang]			Portuguese					
D-20 [trans]			Rendition					

5. Metadata

Transcripts in the corpus come with a set of different metadata on participants and institutional settings. However, some info has to be blocked due to privacy issues.

6. Annotation layers

Transcription of verbal and, in some cases, non-verbal communication is enriched by annotations of different (binary) classifications: language of utterance and translation status.

- Language of utterance distinguishes between monolingual and mixed language utterances.
- Translation status classifies utterances according to whether they are translated later on by the interpreter (“source” or “non-source”), or whether they are based on some previous utterance of primary participants (“rendition” vs “non-rendition”).

7. Access to COMINDAT

The corpus is hosted at the Hamburg Center on Language Corpora (HZSK). The center is part of the CLARIN-infrastructure to provide long-term access and persistent identifiers. Users from European academic institutions need a valid account from their institution plus permission from the owners of the corpus. Access is granted to users with scientific interests. For details, see <https://corpora.uni-hamburg.de>

References

- Angermeyer, P.; Meyer, B. & Schmidt, T. (2012) Sharing community interpreting corpora – a pilot study. In: Schmidt, T. & Wörner, K. (ed.): Multilingual Corpora and Multilingual Corpus Analysis, Hamburg Studies in Multilingualism 14, 275-294. Amsterdam: John Benjamins.
- Meyer, B. (2019) Corpus-based studies on interpreting and pragmatics. In: R. Tipton & L. Desilla (eds.) The Routledge Handbook on Pragmatics and Translation, 75-92. Abingdon & New York: Routledge.