

# Automatic Speech Recognition Systems for interpreters: spoken corpora exploitation by interpreter trainers and trainees

*Gaber, Mahmoud  
Corpas Pastor, Gloria  
University of Malaga*

So far spoken corpora do not seem to have attracted enough attention as a documentation resource for interpreters. However, interpreters tend to rely on audiovisual material in the documentation phase to get familiarised with the speaker's accent, the terminology and the phraseologies used in spoken discourse, and so on. Bendazzoli & Sandrelli (2009) point to the challenges posed by the transcription process itself as one of the main causes for the unbalanced advance between the translation and the interpretation corpus-based studies. In this context, Automatic Speech Recognition (ASR) technology appears as a suitable solution to fill in the existing gaps within the field, as well as a powerful resource to support practitioners.

This study presents a novel approach which aims to meet the specific needs of both interpreting trainers and trainees by means of ASR technology (Gaber et al, 2020). On the one hand, the outcome of such approach would enable trainees to carry out the preparation and documentation phase relying on spoken speeches in order to create an ad hoc corpus, extract terms and acquire the subject knowledge. This is expected to contribute to reducing the cognitive load during the interpreting process. On the other hand, interpreter trainers could benefit from this approach, as they could easily compare and analyse trainees' performance against the transcription of the original speech.

In order to assess the appropriate ASR tool for a given interpreting job, we will follow a six-step methodology: collection of material, selection of ASR tools, automatic transcription process, assessment of ASR tools, ad hoc corpus compilation, and term extraction process.

**Keywords:** Automatic speech recognition, terminology extraction, ad hoc corpus.

## References

- Bendazzoli, C., & Sandrelli, A. (2009). Corpus-based Interpreting Studies : early Work and Future Prospects. *Tradumàtica: traducció i tecnologies de la informació i la comunicació*, (7), 1-19.
- Gaber, M., Pastor, G. C., & Omer, A. (2020). Speech-to-Text Technology as a Documentation Tool for Interpreters: a new approach to compiling an ad hoc corpus and extracting terminology from video-recorded speeches. *TRANS: revista de traductologia*, (24), 263-281.