



UNIVERSIDAD  
DE MÁLAGA

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

Mahmoud Gaber

Gloria Corpas Pastor

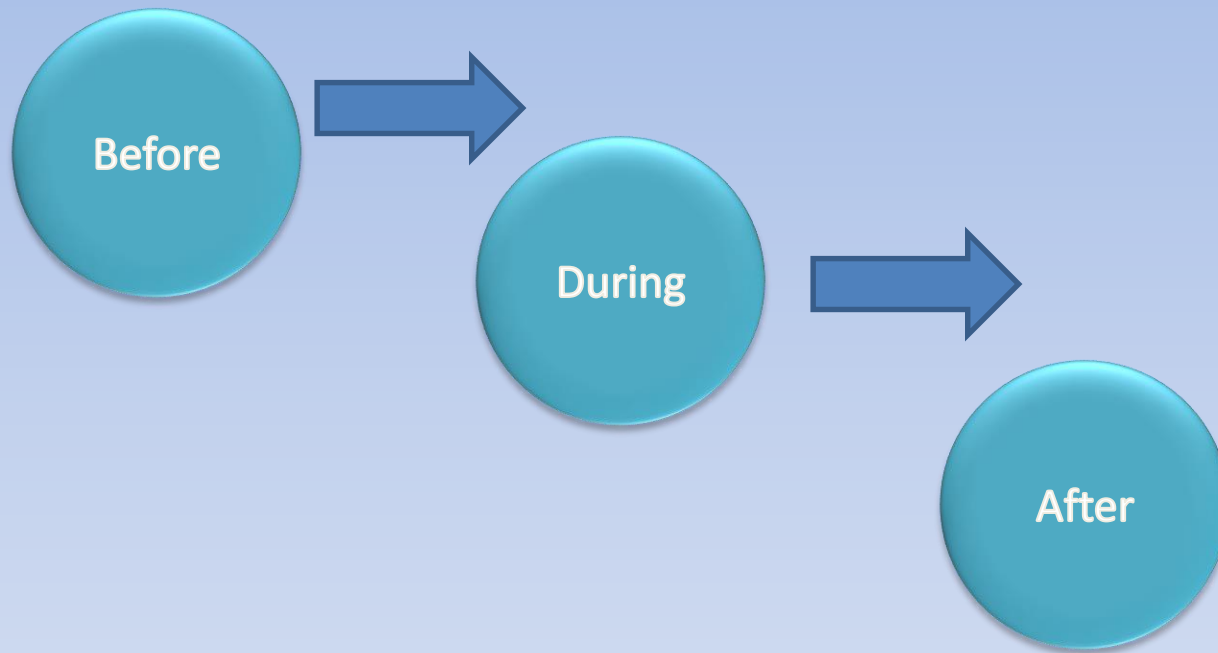
University of Malaga, Spain

Bologna, 3rd November 2021

# Content

- CAI tools
- Abstract
- ASR Literature review
- Hypothesis
- Objectives
- Methodology
- Findings
- Conclusion

# Computer-Aided Interpreting (CAI) tools



# Abstract

This study presents a novel approach which aims to meet the specific needs of both interpreting trainers and trainees by means of ASR technology. 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.

# Literature review

**Fantinuoli (2017): introduced ASR as querying system** during simultaneous interpreting, establishing several requirements for a successful integration of ASR into a CAI tool, such as being speaker-independent, having the capacity to operate on continuous speech, supporting large-vocabulary recognition, detecting specialised terms, and having high accuracy and speed.

**Desmet, Vandierendonck & Defrancq (2018):** conducted an experimental study to **evaluate the feasibility of using ASR systems** (specifically **automatic number recognition**) to determine whether or not it is helpful for interpreters in-booth. The study concluded that technological support was able to reduce the cognitive loads and improve interpreting quality from 56.5 to 86.5 per cent.

**Cheung & Tianyun (2018):** carried out a pilot experiment providing the interpreters with the transcription of speeches delivered in a non-standard accent. The study reported that the fluency score improved when using the transcriptions generated by the ASR during the interpreting process.

**Wang & Wang (2019):** ASR combined with machine translation (better performance in consecutive interpreting).

**Defranco & Fantinuoli (2021):** highlighting figures in the original text (better performance and psychological benefits).

# Objectives:

This study pursues three aims:

- (i) to establish the **most suitable ASR tool** for building ad hoc corpus by comparing several ASR tools and assessing their performance;
- (ii) to **use ASR in order to extract terminology** from the transcriptions obtained from video-recorded speeches; and
- (iii) to **promote the adoption of ASR** among interpreter trainers so they may easily compare and analyse trainees' performance against the transcription of the original speech.



# Speech to Text

## Why?

- **We might not have at our disposal enough resources and material of written texts for certain languages or communication settings.**
- **Even if we do have written material, the spoken language differs from the written one.**
- **Professional interpreters are always keen to listen to spoken speeches during the documentation phase to familiarise themselves with the speaker's accent, common expressions, specific formulae, etc.**

## **Hypothesis:**

- (H1) ASR technology is mature enough to generate a reliable transcription
- (H2) ASR outcome can be useful to build ad hoc corpus and extract terminology to help interpreters during the preparation phase

# Methodology:

(1) Selection  
of videos

(2) Selection  
of ASR tools

(3) Automatic  
transcription

(4) Evaluation  
process

(5) Ad hoc  
corpus

(6) Term  
extraction

# Data collection

## Criteria (ASR assesment Model):

- Videos length
- Gradual difficulties
- Specialised language degree
- Setting
- Accent
- Sound background

## Data source

- Intergovernmental Panel on Climate Change (IPCC),
- University of British Columbia (UBC),
- United Nations Framework Convention on Climate Change (UNFCCC) and
- The Obama White House.

# ASR tools

## Criteria:

We only looked for free or, at least, semi-free tools that do not require any training or optimising to transcribe the audiovisual material.

## Tools requirements:

File format, transcription method, etc.

Tool	License type	Import video/Audio (Speech to text)	VR/Dictation	Insert web link	Supported languages	Pros	Cons
<b>Otter AI</b>	Free 600 minutes/Month	✓	✓	✗	English	Speaker identification. Punctuation. Keywords extraction. Export output to various formats: txt, pdf, srt, web link and copy to clipboard. Click any word through the transcript to listen to it again.	Supports only English
<b>YouTube</b>	Free	✓	✗	✗	English, Dutch, French, German, Italian, Japanese, Korean, Portuguese, Russian and Spanish	Supports 10 languages. Click any word through the transcript to listen to it again	No punctuation.
<b>IBM's Watson</b>	Free	✓ only audio	✓	✗	Arabic, English, Spanish, French, Brazilian Portuguese, Japanese, Korean, German, and Mandarin	Supports 9 languages. Speaker identification. Keywords to spot. Punctuation.	File format limitation: mp3, .mpeg, .wav, .flac, or .opus only
<b>Google Docs</b>	Free	✗	✓	✗	Arabic, Chinese, German, English, French, Italian, Romanian, Spanish, Portugues, Ruso etc.	Supports more than 60 languages and dialects	No punctuation. Disconnection.
<b>SpeechTexter</b>	Free	✗	✓	✗	German, Greek, Hebrew, Hindi, Hungarian, Italian, Japanese, Persian, Polish, Portuguese, Romanian, Russian, Spanish, Sundanese, Turkish, Ukrainian, Urdu, etc.	Supports 44 languages. Export output to various formats: txt, doc and copy to clipboard.	No punctuation
<b>Speechnotes</b>	Free	✗	✓	✗	English. Frensh, Dutch, Deutch, Spanish, Italian, Portugues, Rumanian, Bulgarian,Turkish, Arabic, etc.	Supports more than 40 languages and dialects. Export output to various formats: txt, doc, upload to google drive and copy to clipboard.	No punctuation
<b>Textfromtospeech</b>	Free	✓	✓	✗	Arabic, English, Deutch, French, Italian, Japanese, Ukrainian, Russian, and Spanish	Supports 9 languages. Export output with various formats: txt, doc, copy to clipboard or email the dictated text	No punctuation
<b>SpeechPal</b>	Free 120 miuntes	✓	✗	✓	English	Punctuation. Export output to txt. or email the dictated text. Click any word through the transcript to listen to it again.	Supports only English
<b>Dictation</b>	Free	✗	✓	✗	Arabic, Chinese, English. Frensh, Deutch, Spanish, Italian, etc.	Supports more than 40 languages and dialects. Export output to txt. or email the dictated text.	No punctuation

# Evaluation and results

Hypothesis text vs reference text

WER, Wacc, Bleu score

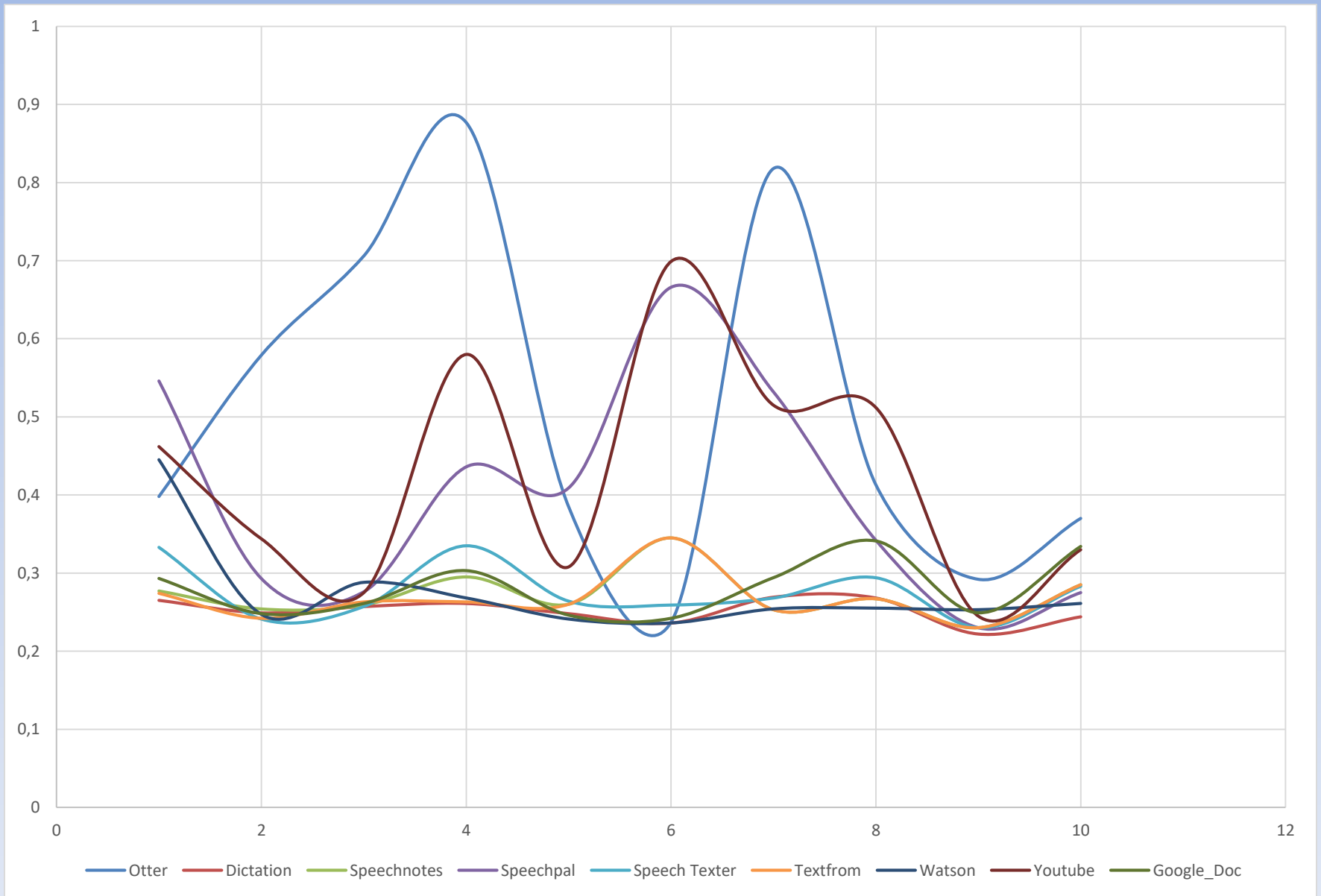
***BLEU score***: BLEU score can range from 0 to 1, where higher scores indicate closer matches to the human transcription.



# Results

<b>Tools/ Videos</b>	<b>V1</b>	<b>V2</b>	<b>V3</b>	<b>V4</b>	<b>V5</b>
<b>Otter AI</b>	0.398	0.579	0.706	0.877	0.385
<b>Dictation</b>	0.265	0.249	0.257	0.261	0.248
<b>Speechnotes</b>	0.277	0.254	0.259	0.295	0.26
<b>SpeechPal</b>	0.546	0.293	0.276	0.436	0.409
<b>SpeechTexter</b>	0.333	0.241	0.257	0.335	0.264
<b>Textfromtospeech</b>	0.274	0.242	0.263	0.263	0.26
<b>IBM's Watson</b>	0.445	0.247	0.288	0.268	0.241
<b>YouTube</b>	0.462	0.344	0.275	0.58	0.308
<b>Google Docs</b>	0.293	0.248	0.261	0.303	0.246

# BLEU score for ASR applications across all videos



# Ad hoc corpus

Corpus Driven Interpreters Preparation (CDIP) and what is called also corpus-based terminology preparation can improve the interpreter's performance on specialized topics (Fantinouli 2006; Castillo Rodríguez 2009; Bale 2013; Xu 2015; Perez-Perez 2018).

Concordance | Sketch Engine

https://app.sketchengine.eu/#concordance?corpname=user%2FMahmoud1983%2Fs2t&tab=advanced&queryselector=cql&cql=%5Blemma%3D%22greenhouse%22%5D&select...

# CONCORDANCE

S2T

**SUBSCRIBE**

CQL [lemma="greenhouse"] 59 (2,233.07 per million)

Details Left context KWIC Right context

1	doc#0	havior change so that we get a fuller picture of why efforts to reduce	<b>greenhouse</b>	gas emissions might have taken off in some places and stalled and o
2	doc#0	</s><s> Others may disagree with proposals to significantly reduce	<b>greenhouse</b>	gas emissions or feel that's not their responsibility to act. </s><s> Gc
3	doc#0	also clearly have an important role to play in implementing legislated	<b>greenhouse</b>	gas reduction policies, sustainable land use plans, and by supporting
4	doc#0	rts to respond to climate change have often been limited to reducing	<b>greenhouse</b>	gas emissions by a few percent here or there. </s><s> And in some
5	doc#1	serve the conditions in which the current crops thrive, either through	<b>greenhouses</b>	, or changing irrigation practices. </s><s> For instance, the same go
6	doc#1	s clearly a role here for wealthier nations to Ironically, the majority of	<b>greenhouse</b>	gas emissions have historically come from the industrialized West. <
7	doc#1	tems, not air conditioning, because this would ironically increase our	<b>greenhouse</b>	gas emissions and create a trade off for climate change mitigation. <
8	doc#2	is unprecedented in human history. </s><s> Due to the emission of	<b>greenhouse</b>	gases that trap heat in the atmosphere, we are experiencing the war
9	doc#2	of us who are alive right now, to mitigate enough that means to bring	<b>greenhouse</b>	gas emissions down enough to allow that adequate, in particular in th
10	doc#3	lions 25 years ago. </s><s> Since then, the United Kingdom has cut	<b>greenhouse</b>	gas emissions by one quarter, we have created the world's first clima
11	doc#4	hank you. </s><s> Despite existing mitigation policies, the growth of	<b>greenhouse</b>	gas emissions has increased over the last decades. </s><s> To limit
12	doc#4	er the last decades. </s><s> To limit global warming to two degrees,	<b>greenhouse</b>	gas emissions need to be substantially reduced in all sectors of the e

Back to the [original interface](#)

Corpus size: total of 170 minutes and 23 seconds, 23, 757 words

# Terminology extraction

Terminology extraction aims to "identify a list of monolingual specialised terms and phrases from the collected corpus that can be used by the interpreter to create a conference glossary as well as to start the learning process" (Fantinuoli 2017a: 33).

<input type="checkbox"/>	171	climate change		▼	▲
<input type="checkbox"/>	33	we' ve		▼	
<input type="checkbox"/>	30	greenhouse gas		▼	
<input type="checkbox"/>	30	carbon pollution		▼	
<input type="checkbox"/>	26	greenhouse gases		▼	
<input type="checkbox"/>	22	change impacts		▼	
<input type="checkbox"/>	22	climate change impacts		▼	
<input type="checkbox"/>	21	gas emissions		▼	
<input type="checkbox"/>	21	greenhouse gas emissions		▼	
<input type="checkbox"/>	15	impacts of climate		▼	
<input type="checkbox"/>	14	carbon capture		▼	
<input type="checkbox"/>	13	natural environment		▼	
<input type="checkbox"/>	12	clean energy		▼	
<input type="checkbox"/>	12	capture and storage		▼	
<input type="checkbox"/>	12	years ago		▼	
<input type="checkbox"/>	11	natural gas		▼	
<input type="checkbox"/>	10	economic growth		▼	
<input type="checkbox"/>	10	adaptive capacity		▼	
<input type="checkbox"/>	10	carbon dioxide		▼	
<input type="checkbox"/>	9	natural systems		▼	
<input type="checkbox"/>	9	they' ve		▼	
<input type="checkbox"/>	9	We' ve		▼	
<input type="checkbox"/>	8	power plants		▼	
<input type="checkbox"/>	8	young people		▼	
<input type="checkbox"/>	8	rising sea		▼	
<input type="checkbox"/>	8	changing climate		▼	
<input type="checkbox"/>	7	reduce carbon pollution		▼	
<input type="checkbox"/>	7	based approaches		▼	
<input type="checkbox"/>	7	global warming		▼	
<input type="checkbox"/>	7	climate change mitigation		▼	▼

Terminology Extraction Suite (TES) (Oliver and Vázquez, 2007)

## SST vs ITST

(Source Speech Transcription vs Interpreted Target Speech Transcription)

- **Trainees** may spend some time **self-assessing** their work and their peers' work by comparing the transcription of the original speech against their interpretation.
- **Trainers** may carry out deeper **analysis**:  
Quality of expression; Grammatical errors;  
Syntactical errors; Lexical errors; False starts ;  
Repetitions; strategies to cope with  
difficulties, etc.

# Findings

- Some tools were able to perform the transcription in less time than the original video duration. For instance, video 9 is 48:46 minutes length and was transcribed in only 21 minutes by Otter AI.
- Otter AI: best performance, keywords extraction feature .
- Otter is limited to English language, and not totally free software
- YouTube: good accuracy rate, with the advantage of supporting 10 languages
- But provides the text without punctuation marks.



# Keyword extraction on Otter AI

The screenshot displays the Otter AI web interface. On the left is a dark sidebar with the Otter AI logo and navigation options: 'My Conversations', 'Shared with Me', 'GROUPS', 'FOLDERS', 'All Conversations', and 'Trash'. The main content area shows a conversation titled '8.1. What is climate change mitigation.mp4' with a search bar and an 'Edit' button. Below the title, the date and time are 'Wed, 6/5 - 2:45 PM' and '9:32'. A section labeled 'SUMMARY KEYWORDS' contains a grid of extracted terms: greenhouse gases, climate change, atmosphere, climate change mitigation, geoengineering, targets, carbon sequestration, greenhouse gas reduction, greenhouse gas emissions, mitigation, emissions, climate, quantity, intensity, carbon capture, adaptation, strategies, result, concentrations, and reducing. Below the keywords is a transcript snippet from 'Speaker 1' at 0:03, starting with 'Welcome to Module eight. In this part of the course, we're moving on from our exploration of the causes and the impacts of climate change to discussing possible solutions. In this lecture will equip ourselves with the basic concepts, we need to explore the first major response type at our disposal, as we attempt to address climate change, which is mitigation. Following this lecture will delve into the specific strategies that we can use to mitigate climate change, exploring the controversy, the potential of these strategies, and some examples associated with each. But first, let's make sure we're clear on just the basic concepts and definitions. During this lecture will learn the widely held definitions of Of course climate change mitigation, greenhouse gas emissions and concentrations, which is a refresher from'. A progress bar at the bottom of the transcript shows the current position at 0:00 of a 9:32 audio clip. The Windows taskbar at the bottom shows the system tray with the date and time '23:37 7/28/2019'.

# Conclusion

- We have been able to establish the most accurate ASR tool
- We managed to compile a monolingual ad hoc corpus and extract candidate terms from spoken speech by means of S2T technology.
- Although the ASR technology is still far from being perfect, the results reveal a great advance which allowed us to obtain a valuable resource through the automatic transcriptions.

## References

CHEUNG, Andrew KF, & Li Tianyun (2018): “Automatic speech recognition in simultaneous interpreting: A new approach to computer-aided interpreting”, Proceedings of Ewha Research Institute for Translation Studies International Conference, At Ewha Womans University.

Defrancq, B., & Fantinuoli, C. (2021). Automatic speech recognition in the booth : assessment of system performance, interpreters’ performances and interactions in the context of numbers. TARGET-INTERNATIONAL JOURNAL OF TRANSLATION STUDIES, 33(1), 73–102. <https://doi.org/10.1075/target.19166.def>

DESMET, Bart, Mieke Vandierendonck, and Bart Defrancq (2018): «Simultaneous interpretation of numbers and the impact of technological support», in Claudio FANTINUOLI (ed.), *Interpreting and technology*, Berlin: Language Science Press, 13–27, doi:10.5281/zenodo.1493291.

FANTINUOLI, Claudio (2017b): “Speech recognition in the interpreter workstation”, Proceedings of the Translating and the Computer 39 Conference, London: Editions Tradulex, 367–377.

OLIVER, Antoni; Mercè Vázquez, & Joaquim Moré (2007): «Linguoc LexTerm: una herramienta de extracción automática de terminología gratuita», *Translation Journal*, 4/11.

Wang, X. & Wang, C. (2019). Can Computer-assisted Interpreting Tools Assist Interpreting?. *Transletters. International Journal of Translation and Interpreting*, (2), 109-139. Recuperado de <https://core.ac.uk/download/pdf/195265773.pdf>

YU, Dong, & Li Deng. (2015): *Automatic speech recognition: A deep learning approach*. London: Springer.

**THANK YOU**

**Mahmoud Gaber**

mahmoudgaber@uma.es