Investigating the use of speech technologies in consecutive interpreting:

A pilot study on ASR-enhanced CAI tool prototype ‘Sight-Terp’

Cihan Ünlü
İstanbul Yeni Yüzyıl University
Hacettepe University

Prof. Aymil Doğan
Hacettepe University

25.11.2022

Why this study?

- The growing impact of ICT and the technological turn in interpreting
- Necessity to enrich empirical research on computer-aided interpreting (CAI) tools
- *Lack of empirical data on technology-mediated consecutive interpreting*

> *[...] in order to shed light onto the advantages and disadvantages of CAI tools, the way they are affecting the interpreting process and the tasks interpreters can perform better with their help and those which [they] cannot, research on new technologies needs to be performed not only on the basis of naturalistic methods (such as corpus analysis), but empirical experiments should be conducted also in stringently controlled experimental conditions (Fantinuoli, 2018, p. 170)*

Computer-Aided Interpreting

- CAI tools are software solutions specifically designed to support interpreting sub-processes such as “knowledge acquisition and management, lexicographic memorization, terminology access”. (Fantinuoli, 2018; 2022).

- Product-oriented CAI tools draw on automatic speech recognition (ASR) and artificial intelligence (AI) technology to provide interpreters with real-time content-related information and enhance the look-up mechanism.

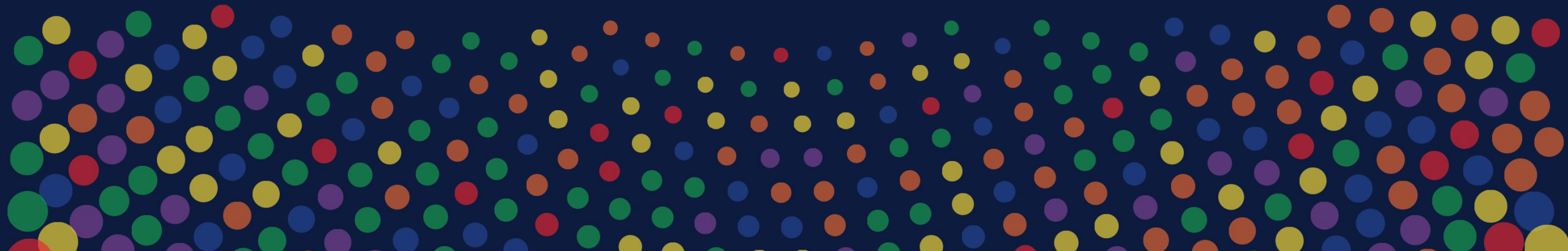
Automatic Speech Recognition

in CAI

“

ASR has a considerable potential for changing the way interpreting is practiced.

(Pöchhacker, 2016, p. 188)



ASR in Interpreting

Preparation

- Helping interpreters in a wide range of productivity and quality-related tasks.
- For corpus building and term extraction.
- For conference preparation and terminology organization.

See

- ❑ (Gaber et al., 2020)

In-booth support

- Automated querying system for information retrieval (Hansen-Schirra, 2012; Fantinuoli, 2017; 2016).
- Offering selected information real-time (terms, numbers, acronyms etc.) (Frittella & Rodríguez, 2022; Fantinuoli et.al, 2021)
  - 22.5-39% error reduction in rendition of numbers in SI (Defranq & Fantinuoli, 2020), %25 gain in accuracy of numbers (Fantinuoli & Pisani, 2021),
- Running transcription (Cheung and Tianyun, 2018)

See

- ❑ InterpretBank (Fantinuoli, 2016; Defranq and Fantinuoli, 2021)
- ❑ The Ergonomics for the Artificial Booth Mate (EABM) project by Ghent Uni.
- ❑ SmarTerp (Rodruigez et al., 2021)
- ❑ VIP System (Corpas-Pastor, 2021)
- ❑ KUDO Interpreter Assist (Fantinuoli et al., 2021)

Can ASR create accurate reference texts for consecutive interpreting action?

Note-Taking

Reading from the notes

Studies on tablet interpreting:

(Goldsmith and Holley 2015)
(Paone, 2016)
(Oceguera López, 2017)
(Goldsmith, 2017)
(Dreschel and Goldsmith, 2016)
(Altieri, 2020)

Features

Named entity highlighting

Automatic segmentation in enumerated style

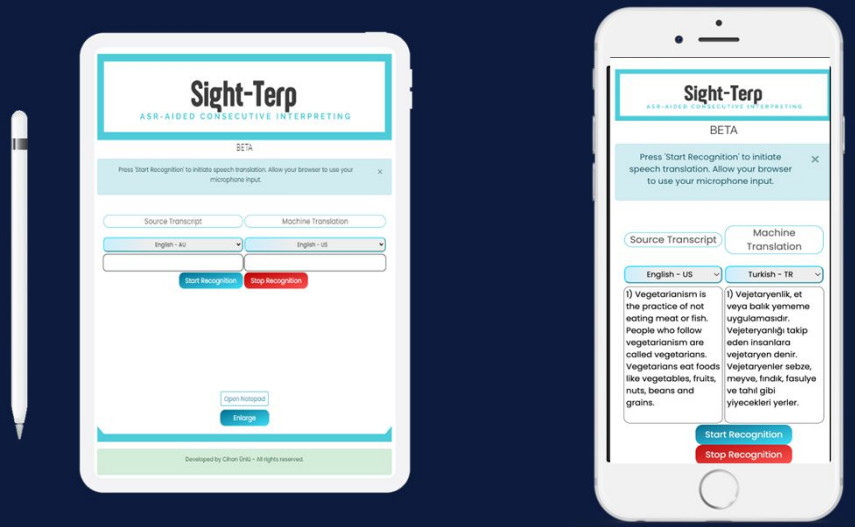
Continuous Speech Translation

Digital Notepad

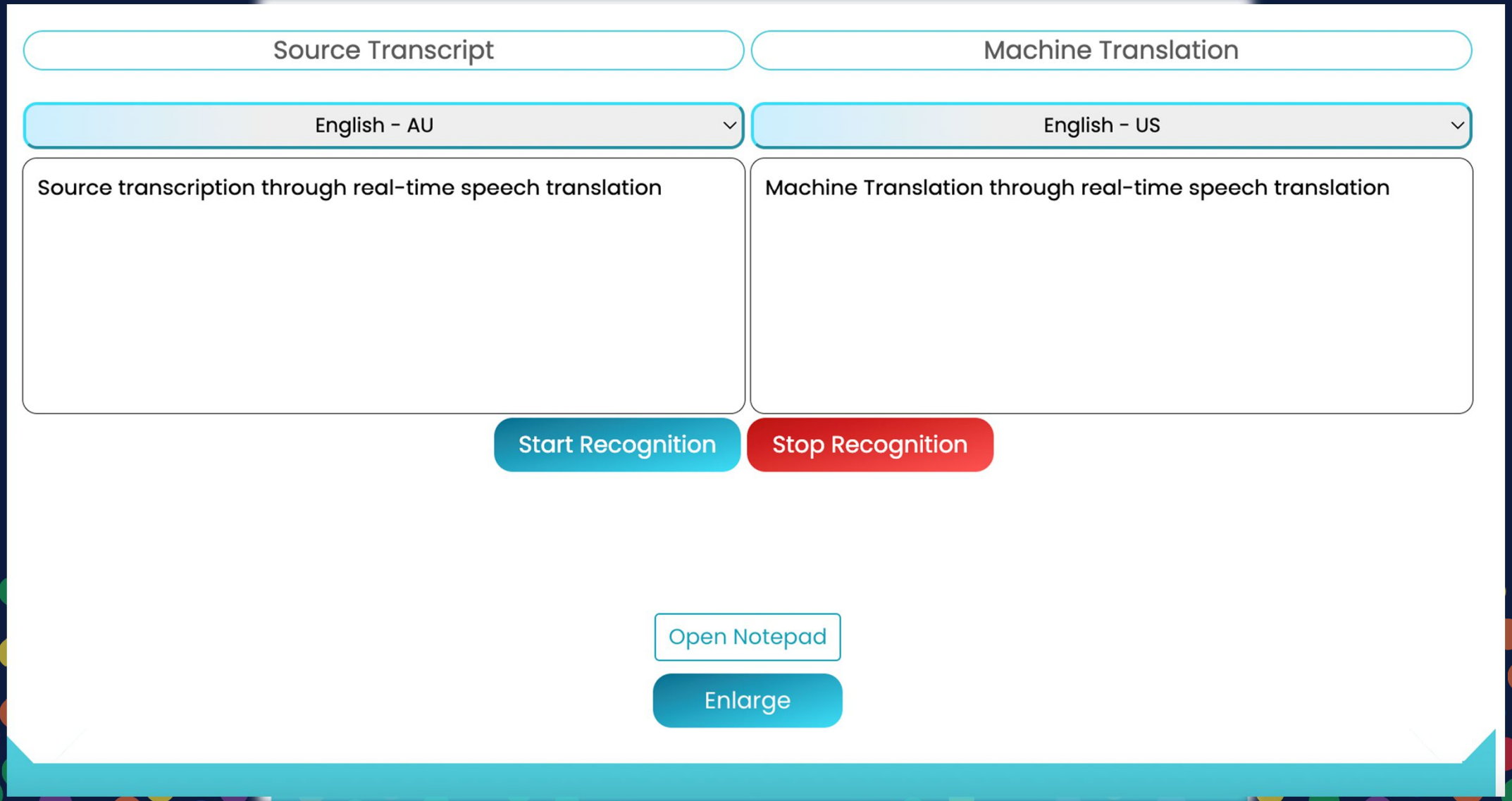
A tool with plain, fast, user-friendly interface

Sight-Terp

Sight-Terp is a non-commercial, experimental web-based ASR-enhanced computer-aided interpreting tool mainly designed for consecutive interpreting scenarios.

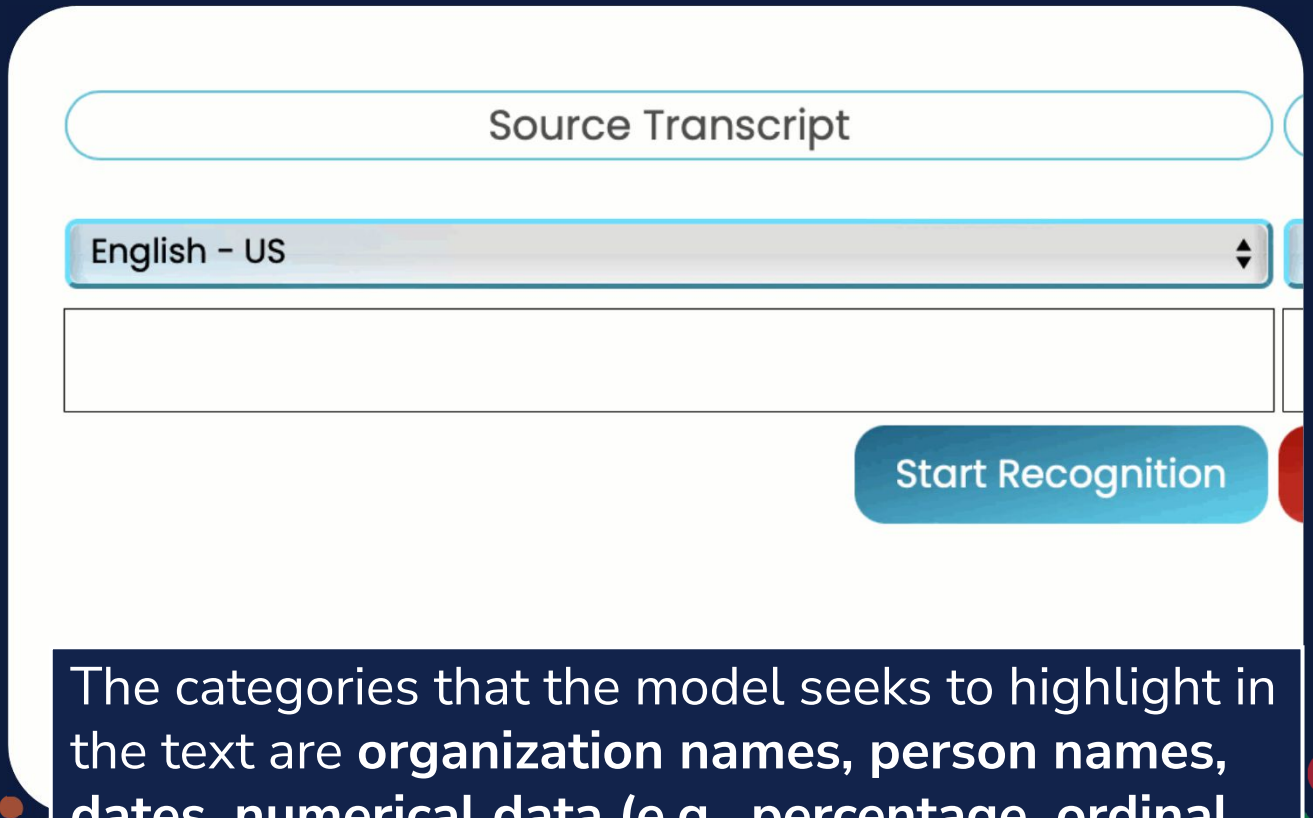


1) Continuous Speech Translation



2) Named Entity Highlighting

- Named entity recognition (NER) is a computational sub-task used for information extraction from a raw text, which identifies the pre-defined entity categories from the text spans (Kim Sang et. al, 2003; Cui et. al, 2021).
- Sight-Terp uses a NER model from Microsoft Cognitive Services Text Analytics API and highlights them.

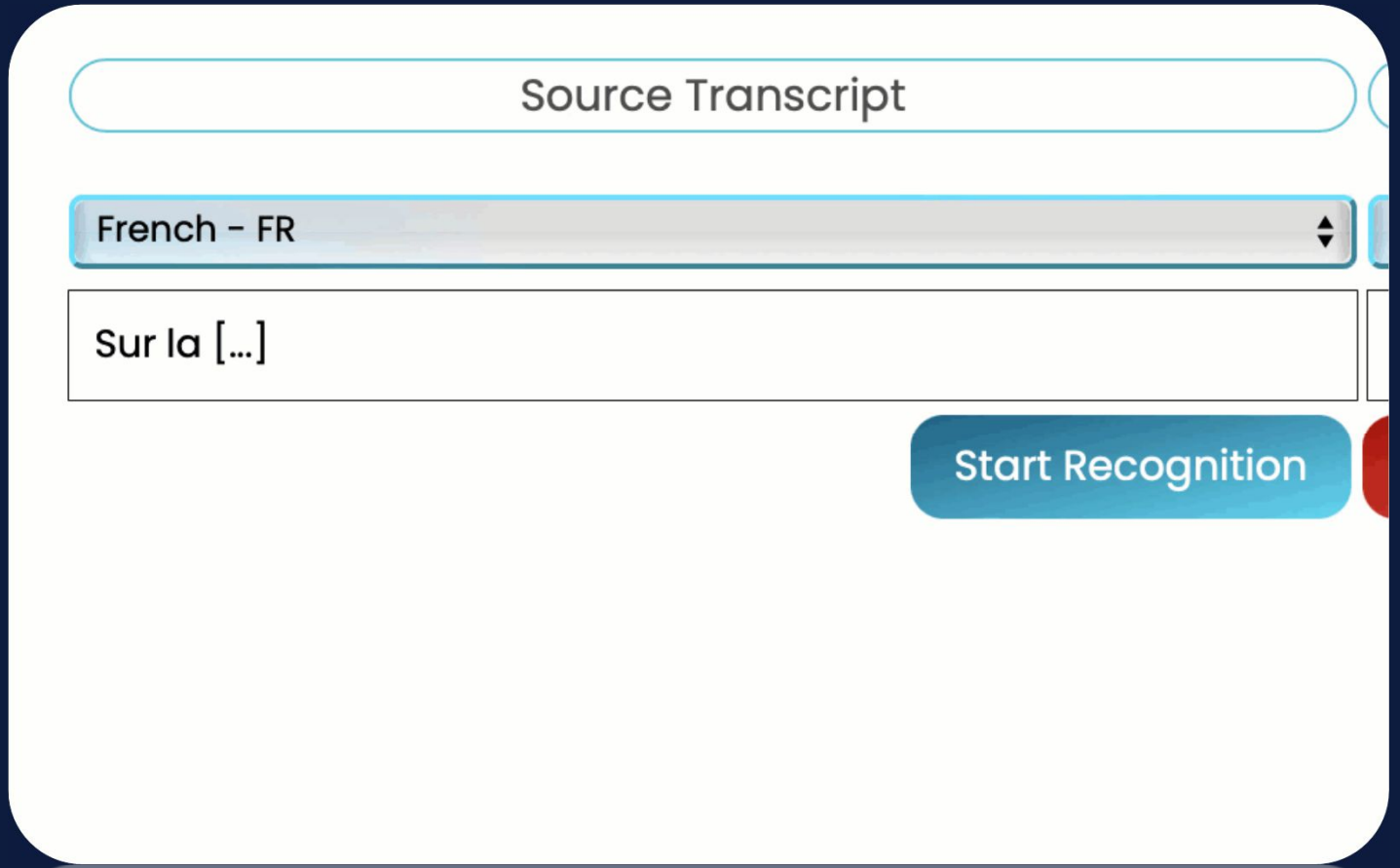


The categories that the model seeks to highlight in the text are **organization names, person names, dates, numerical data (e.g., percentage, ordinal numbers, temperature), location names, and currency data (e.g., two million $)**.

2) Named Entity Highlighting

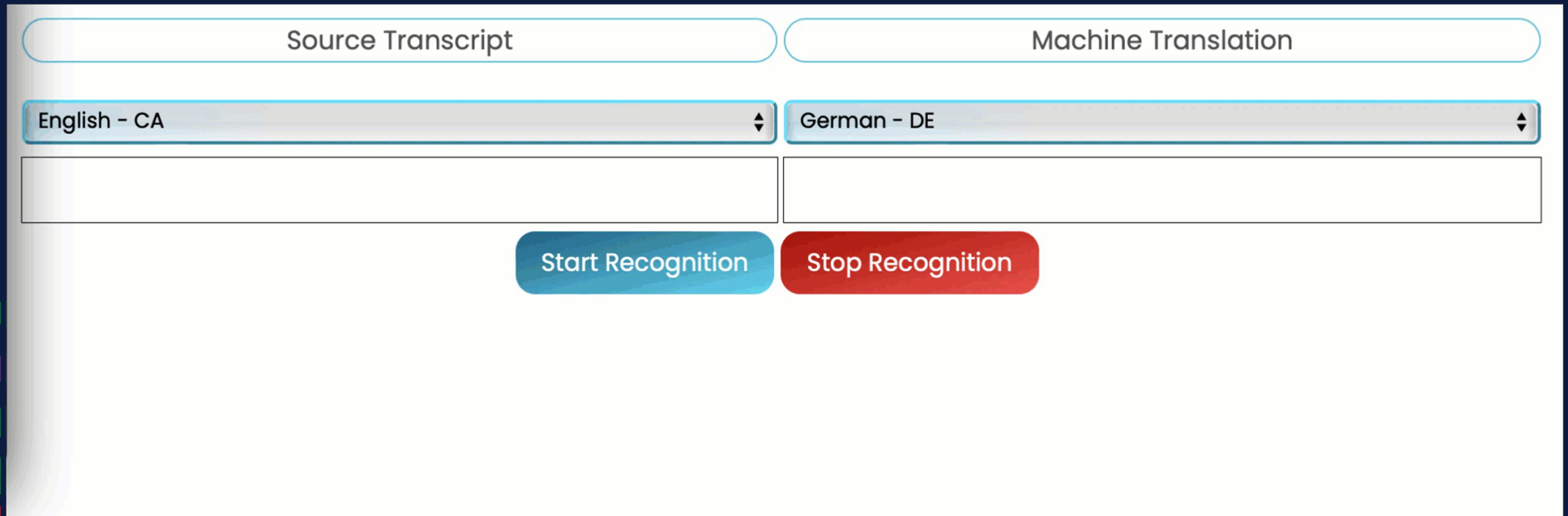
The categories to be recognized by the model (organization names, person names, proper names, dates, numerical data etc.) are the units that have critical technical and contextual information.

The highlighting feature is implemented to ease ‘reading from notes’ effort by facilitating the detection of content words for a faster reformulation of the message while reading.



3) Segmentation in enumerated style

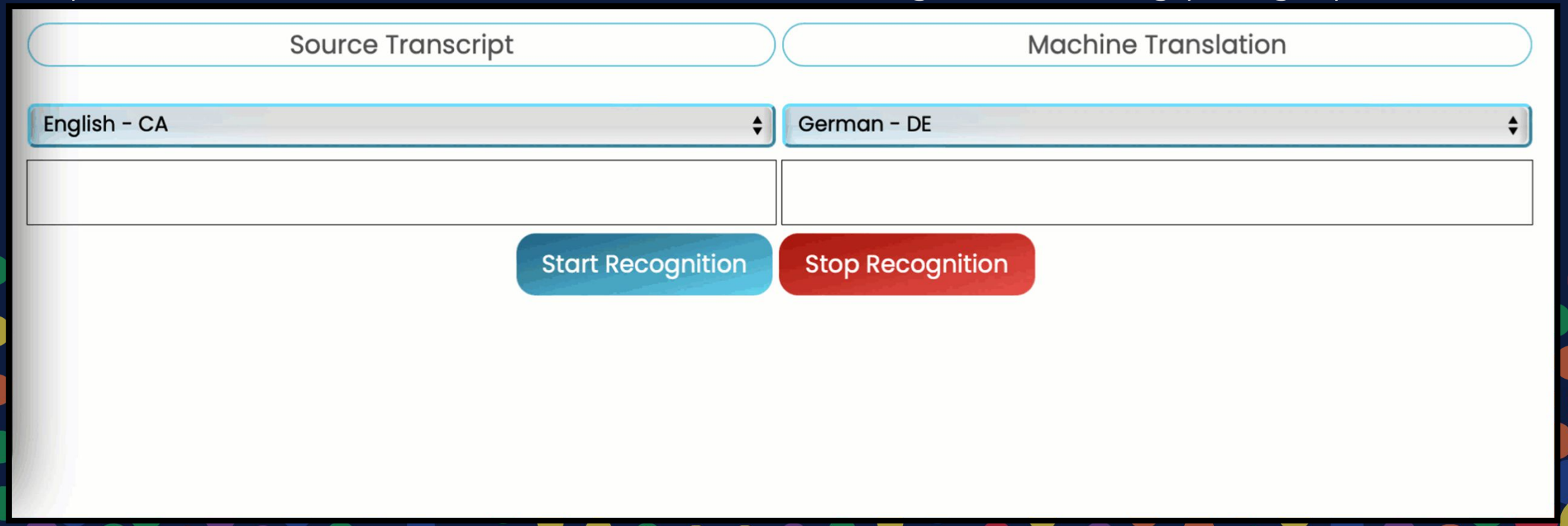
- Automatic text segmentation allows both source and machine translation texts to be displayed concurrently in a vertical form in the adjacent text boxes during continuous speech recognition.
- Verticality and sectioning with lines are common practices in traditional consecutive interpreting advised by other scholars such as Roderick Jones (2002), Dörte Andres (2002) Christopher Thiéry (1981).



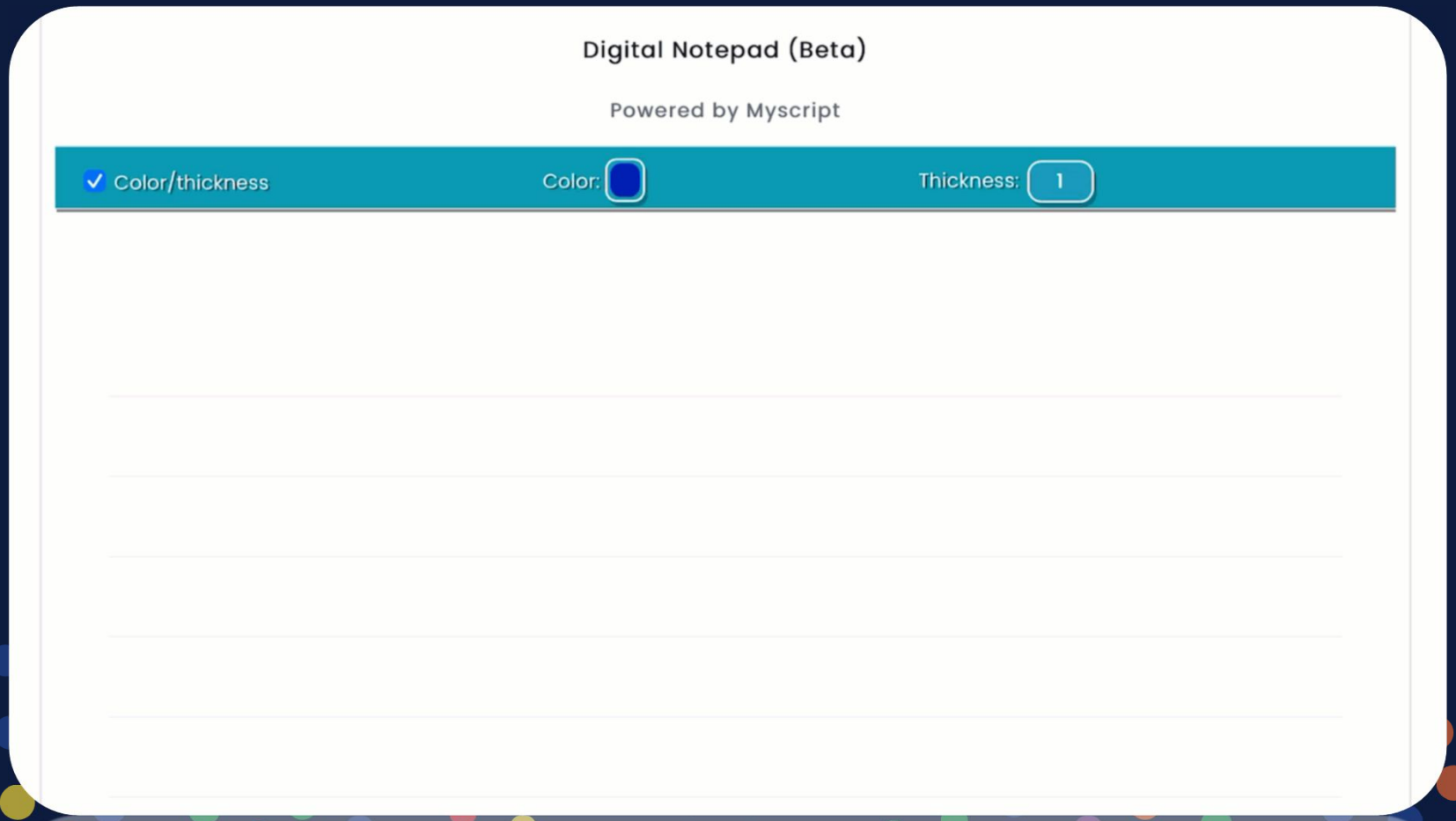
3) Segmentation in enumerated style

Chunk segmentation in enumerated style

- This feature is designed to display the reference text in an easy-to-read fashion and allow the user (interpreter) to follow up the source segment with its target MT output thanks to the enumerated style.
- Xinyu Wang and Caiwen Wang (2018) investigated whether a possible MT reference in consecutive interpreting might boost interpreting accuracy.
- In the post-experiment questionnaire, 9 out of 10 participants in study (2018) reported that they failed to locate needed information in unsegmented long paragraph.



4) Digital Notepad (Nebo)



Preliminary Test Methodology

Procedure

- The participants were asked to interpret two speeches in traditional consecutive mode and two speeches using Sight-Terp, delivering in *sight-consecutive* modality. (From English into Turkish)
- All participants were trained on how to use the tool before the experiment. (using iPad Pro)
- Tokyo itself was left relatively unharmed by the earthquake last year. but the size of that earthquake and the devastation that it caused in the north of Japan was a reminder to Tokyo that Tokyo needs to be ready for a similar disaster.

Fluency

- Examining the speech rates.
- **Examining the disfluency markers** (overall frequency of disfluencies; false starts; frequency of filled pauses, filler words, whole-word repetitions, broken words, incomplete phrases) (Lickley, 2015)

Research Question

Does the provision of ST-generated reference texts together with NER and automatic vertical segmentation improve interpreters' interpreting accuracy thereby enhancing performance in a consecutive interpreting task?

Do participants sacrifice fluency while using Sight-Terp?

Limitations

Samples

Participants are not professional interpreters. Main test --> 20+ participants

Language Pairs

The pilot experiment is conducted for Turkish-English language pairs.

Training

Training is provided for once and not more than one hour.

ASR Engine

Microsoft Speech Translation API only.

Samples

The participants are non-professional TIS graduates.

Preliminary Test

Materials

- Four speeches (4-5 min each) to be delivered in English.
- All speeches have close lexical density levels and readability index show similar results

	Speech Materials Used			
	Topic: Earthquakes in Japan		Topic: Violence against Women	
Reading Index	Speech 1	Speech 2	Speech 3	Speech 4
Automated Readability Index	9,47	10,75	9,06	9,56
SMOG	10,91	11,13	11,15	11,71
Flesch–Kincaid Grade Level	8,88	9,24	8,5	9,66
Coleman-Liau Index	10,61	12,11	11,08	12,46
Gunning-Fog Index	11,12	11,40	11,24	12,14
Average Grade Level	**10,2**	**10,93**	**10,21**	**11,67**
Median Grade Level	**10,61**	**11,12**	**11,08**	**12,06**
Flesch Reading Ease	60,207	58,298	56,084	40,906
Lexical Density	51.57%	56.09%	50,00%	54,93%

Readability index results and lexical density ratios of the materials

Preliminary Test

Materials

- All speeches are delivered in a moderate pace.
- The meaning units (Seleskovitch, 1989) are calculated for each speech.

Material Name	Duration	Length (Words)	Number of Meaning Units
Speech1/1 Earthquakes in Japan	03:57	465	109
Speech1/2 Earthquakes in Japan	04:14	452	127
Speech 2 /1 Violence against Women	04:34	513	159
Speech 2 /2 Violence against Women	03:32	404	125

Duration, length (words) and the total number of units of meaning of each speech

Material Name	Word-Error-Rate (WER) by ASR	Named Entity Precision by ASR
Speech 1/1 Earthquakes in Japan	N/A	N/A
Speech 1/2 Earthquakes in Japan	9.7%	30/30
Speech 2 /1 Violence against Women	N/A	N/A
Speech 2 /2 Violence against Women	7.4%	30/32

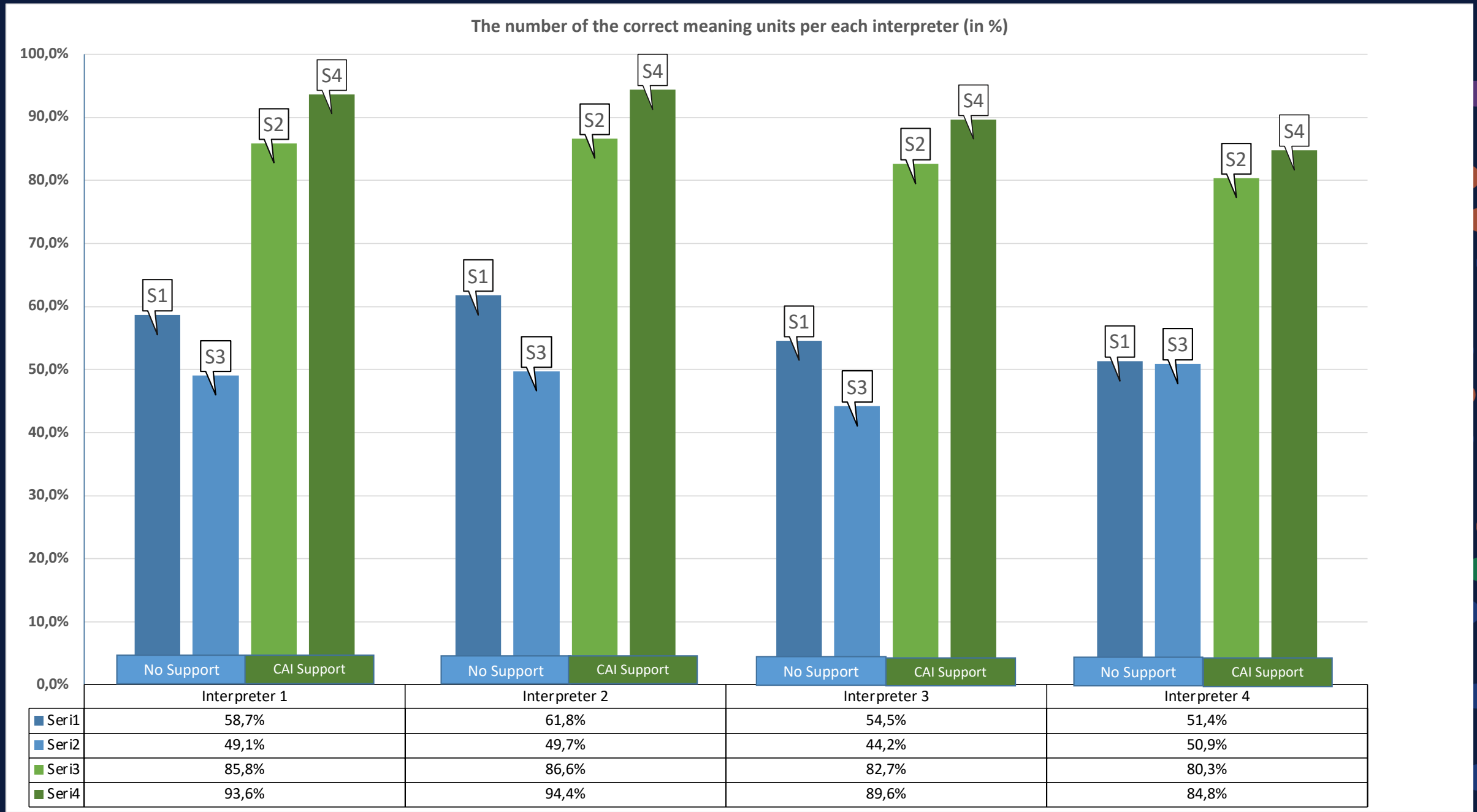
Preliminary Test

Participants

- Participants are fresh graduates of TIS with special focus on interpreting.
- All of them have successfully passed consecutive interpreting and note-taking courses during their undergraduate education.

Participants	Earthquakes in Japan		Violence against Women	
	Speech1/1 Earthquakes in Japan	Speech1/2 Earthquakes in Japan	Speech 2/1 Violence against W.	Speech 2/2 Violence against W.
Interpreter 1	No Support	CAI Support	No Support	CAI Support
Interpreter 2	No Support	CAI Support	No Support	CAI Support
Interpreter 3	No Support	CAI Support	No Support	CAI Support
Interpreter 4	No Support	CAI Support	No Support	CAI Support

Distribution of the speeches for each participant



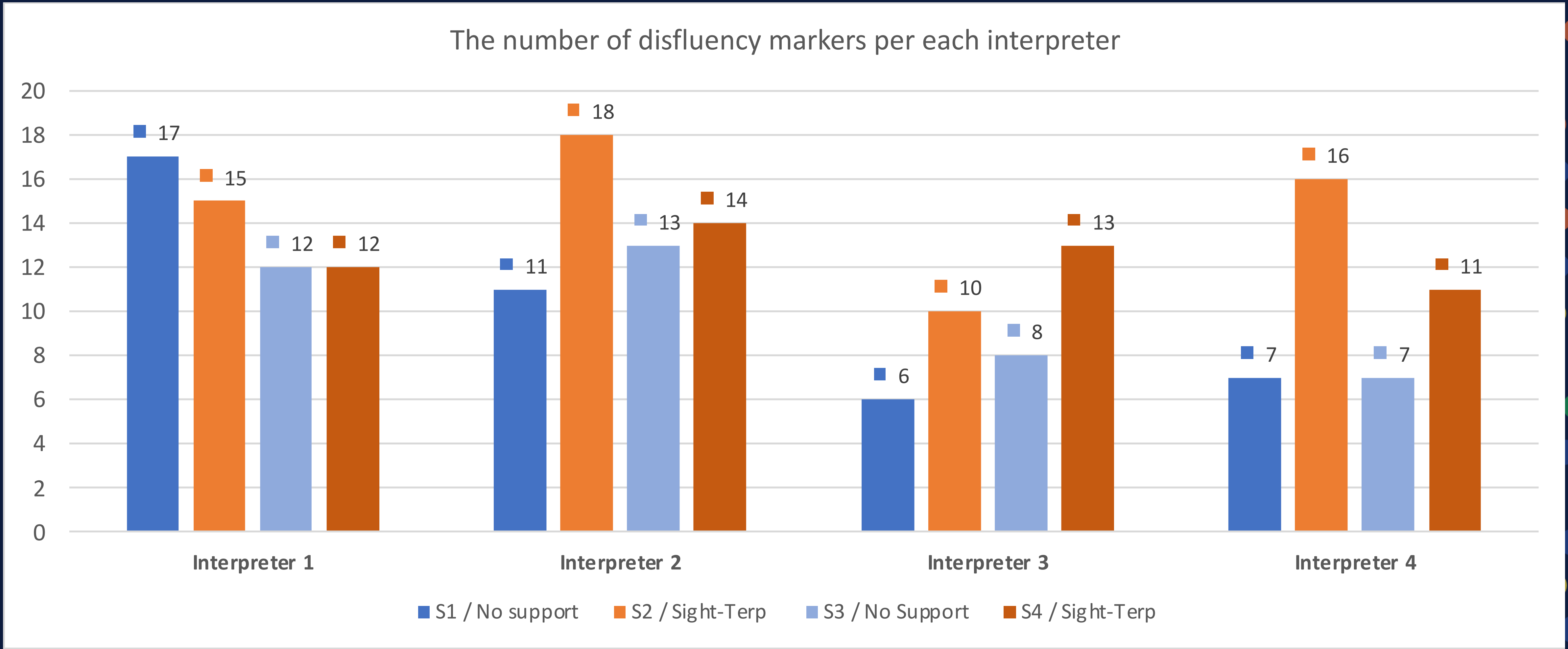
	Interpreter 1	Interpreter 2	Interpreter 3	Interpreter 4
Seri1	58,7%	61,8%	54,5%	51,4%
Seri2	49,1%	49,7%	44,2%	50,9%
Seri3	85,8%	86,6%	82,7%	80,3%
Seri4	93,6%	94,4%	89,6%	84,8%

Results and Discussion

DĬ IÉÍĜĦĤĂĞĂFIGĦIHÉĦĘFIJ

Participants	Speech Topic 1		Speech Topic 2	
	Speech1/1 Earthquakes in Japan	Speech1/2 Earthquakes in Japan	Speech 2/1 Violence against W.	Speech 2/2 Violence against W.
	Pen&Paper	Sight-Terp	Pen&Paper	Sight-Terp
Interpreter 1	05:25	05:04	04:33	04:29
Interpreter 2	04:00	05:08	03:40	05:10
Interpreter 3	03:44	04:45	03:46	04:36
Interpreter 4	03:50	06:34	03:45	06:10

Results and Discussion



Results and Discussion

Participants using the CAI tool show **higher rates** in accuracy (units of meaning)...

...but have more disfluency markers, which results in ĒFIJIJÄĞĞ FĦíĂ delivery.

Compared to the conventional method, the participants in the preliminary experiment made more repetitions, incomplete sentences, hesitations when they used CAI tool.

Results and Discussion

Small-scale post-exp. questionnaire results

- All participants think Sigh-Terp's layout and design are comfortable as they report no problem while being adapted to the tool.
- 3/4 report that two reference texts shown (speech translation bound with source transcript) are not a hindrance but a useful aid. When they face difficulty in remembering equivalent the linguistic units in the target text they can get help from MT.
- 3/4 think that seeing the ongoing real-time speech translation allows them to plan and organize their delivery while speech is being delivered. But being able to take notes at the same time would be better for them not to forget.
- 4/4 want to see the entity highlighting in the target text as well.

Results and Discussion

Small-scale post-exp. questionnaire results

➢ Which one did you get the most help from? MT or source transcript? Or both?

Interpreter 1	I sought support from **source transcription** most and used MT for terms and longer, complex sentences.
Interpreter 2	I looked at the machine translation more and got help from the source text for some parts that I missed or couldn't hear.
Interpreter 3	I delivered my interpretation mostly using the machine translation. I got support from the original text for the sentences that I found that the application gave wrong results.
Interpreter 4	I mostly used the **source transcription** except when I need to look up some words.

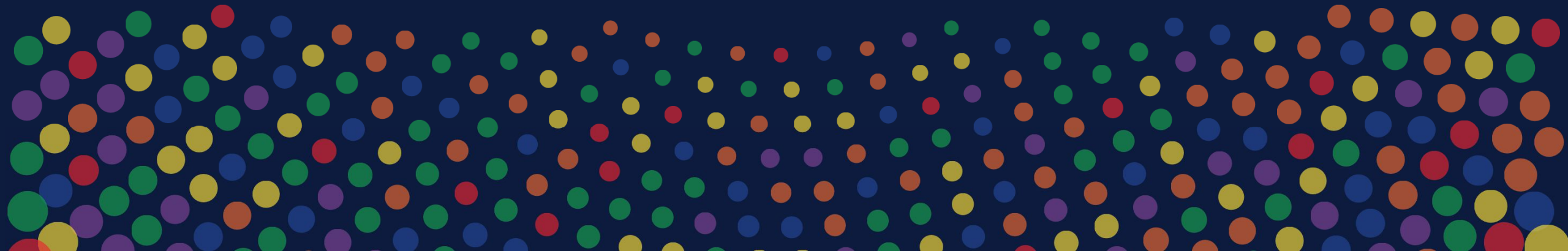
Future Research

- An eye-tracking study to see the feasibility of named entity highlighting during sight interpreting.
- Testing the tool for different directionalities (retour) especially for unsymmetrical language pairs.
- Testing the tool for semi-formal dialogue interpreting scenarios.
- An experimental study using digital notepad and ST together. (usable or more cognitive load?)

Thanks!

Any questions?

unlucihan@yahoo.com
aymildogan@gmail.com



References

- Andres, Dörte (2000) *Konsekutivdolmetschen und Notation*, Mainz: FASK.
- Cheung, A.& Tianyun, L. (2018). Automatic speech recognition in simultaneous interpreting: A new approach to computer-aided interpreting, Proceedings of Ewha Research Institute for Translation Studies International Conference, Ewha Womans University.
- Corpas Pastor, G. (2021). Interpreting and Technology: Is the Sky Really the Limit? *Proceedings of the Translation and Interpreting Technology Online Conference TRITON 2021*.
- Defrancq, B. & Fantinuoli, C. (2020). Automatic speech recognition in the booth Assessment of system performance, interpreters' performances and interactions in the context of numbers. Target. doi: https://doi.org/10.1075/target.19166.def
- Drechsel A. & Goldsmith J. (2016). Tablet Interpreting: the evolution and uses of mobile devices in interpreting. Proocedings of CUITI Forum 2016
- Fantinuoli C., Marchesini G., Landan D. (2021). Interpreter Assist: Fully-automated real-time support for Remote Interpretation". 2022 In: *Proceedings of Translator and Computer 53 Conference*
- Fantinuoli, C. (2016). Interpretbank: Redefining computer-assisted interpreting tools. Proceedings of the 38th Conference Translating and the Computer, 42–55.
- Fantinuoli, C. (2017): Speech Recognition in the Interpreter Workstation. *Proceedings of the Translating and the Computer 39*, Geneva: Editions Tradulex, p. 25–34.
- Fantinuoli, C. (2018a): Computer-assisted interpreting: challenges and future perspectives. In Durán, I. / Corpas, G. (eds.), *Trends in e-tools and resources for translators and interpreters*. Leiden: Brill, p. 153–174.
- Fantinuoli, C. (2018b). Interpreting and technology: The upcoming technological turn. In *Interpreting and Technology*. pp. 1–12. Zenodo.
- Fantinuoli, C., & Pisani, E. (2021). Measuring the Impact of Automatic SpeechRecognition on Number Rendition in Simultaneous Interpreting. In *Empirical Studies of Translation and Interpreting* (1st ed.). Routledge.
- Frittella, F. M., & Rodríguez, S. (2022). Putting SmarTerp to Test: A tool for the challenges of remote interpreting. *INContext: Studies in Translation and Interculturalism*, *2*(2). https://doi.org/10.54754/incontext.v2i2.21
- Gaber, M., Corpas Pastor, G., & 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 Traductología*, (24), 263-281. https://doi.org/10.24310/TRANS.2020.v0i24.7876
- Goldsmith J. / Holley J. (2015*) Consecutive Interpreting 2.0: The Tablet Interpreting Experience*, unpublished MA Thesis, University of Geneva.

References

- Goldsmith, J. (2017). A Comparative User Evaluation of Tablets and Tools for Consecutive Interpreters. *Proceedings of Translation and the Computer* 39, 39, 40–50.
- Goldsmith, J. (2018). Tablet interpreting Consecutive Interpreting 2.0. Translation and Interpreting Studies. *The Journal of the American Translation and Interpreting Studies Association*, 13(3), 342–365. https://doi.org/10.1075/tis.00020.gol
- Hansen-Schirra, S. (2012). Nutzbarkeit von Sprachtechnologien fu¨r die Translation. *trans-kom* 5(2):211–226.
- Jones, Roderick (2002) *Conference Interpreting Explained,* Manchester: St Jerome
- Lickley, R. J. (2015). "Fluency and Disfluency." In Redford, M. (Eds.) *The Handbook of Speech Production*, pp. 445-474. Hoboken, NJ: John Wiley and Sons
- Mirko, A. (2020). Tablet interpreting: étude expérimentale de l'interprétation consécutive sur tablette. *The Interpreters' Newsletter*. 25, 19–35. http://doi.org/10.13137/2421-714X/31235
- Orlando, M. (2014). A study on the amenability of digital pen technology in a hybrid mode of interpreting: Consec-simul with notes. Translation & Interpreting, 6(2). 39–54.
- Paone, M. (2016). Mobile *Gerate beim Simultandolmetschen mit besonderem Bezug auf Tablets* (Unpublished MA thesis.) University of Vienna, Austria
- Pöchhacker, F. (2016). *Introducing Interpreting Studies*, Abingdon: Routledge.
- Rodriguez S., Gretter R., Matassoni M., Alonso A., Corcho O., Rico M.,Falavigna D. (2021). SmarTerp: A CAI System to Support Simultaneous Interpreters in Real-Time. In Proceedings of the Translation and Interpreting Technology Online Conference, pages 102–109, Held Online. INCOMA Ltd..
- Seleskovitch, D., & Lederer, M. (1989). *Pédagogie raisonnée de l'interprétation*. Brussels-Luxembourg: OPOCE/ Didier Erudition.
- Thiéry, Christopher (1981) "L'enseignement de la prise de notes en interprétation consécutive: un faux problème?", in Jean Delisle (ed), *L'enseignement de l'interprétation et de la traduction – de la theorie a la pédagogie*. Cahiers de traductologie 4, Ottawa: Editions de l'Universite d'Ottawa, 99–112.
- Wang, X., and Wang, C. (2019). "Can Computer-Assisted Interpreting Tools Assist Interpreting?" Transletters. *International Journal of Translation and Interpreting*, 3, 109–39.