

BOOK OF ABSTRACTS

100 Years of Technology in Conference Interpreting

INTTECH100

1–2 October 2026

International Labour Organization · 4 Rue des Morillons · Geneva, Switzerland

Co-hosted by the Faculty of Translation and Interpreting (FTI) of the University of Geneva and the International Labour Organization (ILO)

Contents

Keynotes	4
Reimagining Conference Interpreting: Technology, Technocuriosity, and Human Agency	4
Is automating communication progress?	5
Talks.....	6
Distance Interpreting	6
Psychophysiological responses of interpreters in remote interpreting: The effect of boothmate presence and visual support	6
Hearing the difference: sound artifacts, frequency and the psychophysiology of remote interpreting	8
Computer-Assisted Interpreting	10
The impact of real-time CAI support in simultaneous interpreting on listeners' retention of the original content and intention	10
The impact of ASR and AI-based CAI tools in RSI to support human interpreting: Performance, cognitive load and stress levels.....	12
Artificial Intelligence in Interpreting	14
How good is AI? A quality evaluation of FAST output in seven languages	14
From Error Analysis to Audience Reception: Assessing Selected Aspects of Machine Interpreting Quality A.D. 2026 in the Polish–English Language Pair	16
Algorithms in the booth: why and how to automate interpreting quality assessment.....	18
Posters	21
Poster 1 · CAI in the Booth – Is It Worth the Risk? A Study on Quality and Risk Perception among Professionals	21
Poster 2 · Interpretación con IA versus interpretación humana de discursos con dimensión de género en contextos institucionales. Un análisis comparativo de la calidad	24
Poster 3 · Risk and Reliability in AI-Assisted Interpreting: Types of Errors in ASR and Speech to Speech Systems	26
Poster 4 · Challenges in Contemporary Interpreting Practice: A Comparison of Onsite and Remote Modes in the Ukrainian Context	29
Poster 5 · 40 years of interpreting quality: A systematic literature review	31
Poster 6 · A scoping review of computer-assisted interpreter training (1995–2026): research foci, gaps, and a pedagogy-driven multi-component interaction framework.....	33
Poster 7 · Evaluation of Live Translation: A 3D Approach	35
Poster 8 · Beyond “Just” Interpreters: Fostering Interdisciplinary Interpreter Training for Broader Competency Frameworks in Future Professionals	37
Poster 9 · Interpreting in Transition: How Training Institutions Adapt Curricula to Technological Change	39
Poster 10 · More than Just Listening: How Visual Supports Shape Simultaneous Interpreting	41

Poster 11 · VKD Market Snapshot: German Conference Interpreters' Perceptions of Market Change in the Context of AI and Economic Uncertainty	44
Poster 12 · The Project INNOVATRAD-CM: Research and Innovation in Real-Time Speech Generation and Interpretation	46
Poster 13 · Outil d'auto-entraînement des apprentis-interprètes assisté par l'IA.....	48
Poster 14 · CALIPSO: A Programmable Environment for Research and Training in AI-Assisted Interpreting	50
Poster 15 · Effective theory-based AI prompting for in-class speech material: A solid point of departure.....	52

Keynotes

Reimagining Conference Interpreting: Technology, Technocuriosity, and Human Agency

Keynote by Dr. Özüm Arzık-Erzurumlu · Thursday 1 October 2026, 9:30–10:15

Program in Translation and Interpreting Studies, University of Illinois Urbana-Champaign, USA



ABOUT THE SPEAKER

Özüm Arzık-Erzurumlu, PhD, is a full-time faculty member in the Program in Translation and Interpreting Studies at the University of Illinois Urbana-Champaign. She is the author of *Reimagining Conference Interpreting in the Age of AI* (Routledge) and co-editor of the forthcoming special issue *Paradoxes in Translation and Interpreting Technology* in *Translation and Interpreting Studies*. Her publications have appeared in journals including *Interpreting*, *Translation and Interpreting Studies*, and *Perspectives*. She has delivered invited talks and conference presentations internationally, including at ATISA, AIIC USA, EST, and IATIS. Arzık-Erzurumlu is a professional accredited conference interpreter with Turkish (A), English (B), and Spanish (C). She has interpreted speeches by American presidents for major Turkish broadcasters since 2006 and has pursued her professional interpreting career for international organizations, political summits, and media broadcasts. She is a board member of ATISA and serves on the advisory boards of *Encounters in Translation* and the *Journal of Translation Studies*.

Abstract

Conference interpreting has been shaped by technological innovation throughout its history, from the introduction of simultaneous interpreting equipment to videoconferencing, remote interpreting, CAI tools, and, more recently, AI. Yet technological change has never been a one-way process; interpreters have continually negotiated, resisted, adapted to, and shaped the technologies they use.

Drawing on insights from *Reimagining Conference Interpreting in the Age of AI* and building on themes explored in *100 Years of Technology in Conference Interpreting*, this keynote examines how the profession has evolved in response to rapid technological transformation. It explores the acceleration of remote interpreting, the ways interpreters adapted familiar practices to new environments, and the opportunities and challenges that emerged in the process. Particular attention is given to the profession's gains and losses, highlighting how some interpreters moved from feeling constrained by technology to becoming empowered through it.

Viewing interpreting as a networked process involving both human and non-human actors, the presentation emphasizes interpreters' agency and technocuriosity as drivers of professional change. Looking ahead, it argues that the future of conference interpreting will depend not only on technological innovation but also on the profession's ability to preserve the social, emotional, ethical, and communicative dimensions that remain central to human-mediated multilingual communication.

Is automating communication progress?

Keynote by Dr. Laura Keller · Friday 2 October 2026, 9:15–10:00

Freelance conference interpreter, Geneva, Switzerland



ABOUT THE SPEAKER

Laura Keller bridges multilingual communication and digital ethics. Her expertise in interpreting is rooted in an MA and a PhD in interpreting from the University of Geneva and her work as a freelance conference interpreter on the private and institutional markets in Geneva for going on 15 years. She is accredited to the European Institutions and is an active member of AIIC. Laura is currently pursuing an MA in Ethics at the University of Lucerne, specializing in the ethics of the digital transformation and is a supporting member of the Association of AI Ethicists. In her work in AI literacy advocacy she focuses on the ethical dimension of automating multilingual communication and the questions that arise regarding biases and power imbalances that are highlighted and exacerbated by AI-driven technologies. She has published and regularly intervenes in panels and workshops on the subject.

Abstract

Technological advances are the contemporary epitome of progress. However, when priority is given to technical feasibility, questions relating to what is desirable and just may fail to be adequately addressed. When the kind of progress prioritized leaves no room for ethical deliberation, ethical safeguards are easily framed as obstacles to progress.

Using AI to automate communication within and across languages is a prime example of the tension between technical feasibility and ethical considerations. Communication is ideally in service of relationship building and of negotiating a shared reality. Technologies that allow for scaling communication easily give the impression that they also automatically exponentiate underlying values such as inclusion. But addressing more stakeholders in more instances of communication over more channels and in more languages without engagement likely amounts to scaling without direction.

While it is illusory to believe that all ethical issues can be anticipated prior to the deployment of new technologies, positioning ethical deliberation as part of the innovation process may help resolve the perceived tension between progress and ethics. This is particularly relevant today, when the understanding of progress determines institutional and policy choices and drives resource allocation.

Talks

Distance Interpreting

Thursday 1 October 2026, 10:15–11:30

Psychophysiological responses of interpreters in remote interpreting: The effect of boothmate presence and visual support

Paweł Korpai, Agnieszka Chmiel, Paulina Rozkrut & Marta Kajzer-Wietrzny

Adam Mickiewicz University, Poznań, Poland

Recent years have seen an increasing interest in psychophysiological research on remote simultaneous interpreting (RSI) including the analysis of stress experienced by interpreters in onsite and remote settings (e.g., Moser-Mercer, 2005; Roziner & Shlesinger, 2010). Some of this research suggests that while interpreters perceive remote interpreting as more cognitively taxing, this relationship is not reflected in objective psychophysiological measures (Roziner & Shlesinger, 2010). Studies on the role of visual support in simultaneous interpreting have also gained some ground, with some research pointing to the facilitatory role of access to visual materials in interpreters' performance (Stachowiak-Szymczak & Korpai, 2019). Limited access to visuals may be a challenge in RSI because interpreters typically use available visual information to support their task (Seubert, 2019). On the other hand, the multimodality involved in processing visual materials has been shown to increase the interpreters' cognitive load (Gieshoff, 2018).

This study is part of a larger project examining the impact of RSI setups and boothmate location on both interpreting performance and interpreter experience (see Chmiel et al., 2025). Here, we aimed to complement previous research by studying the impact of boothmate presence (co-located, not co-located, and virtual) and visual support (slides provided, slides not provided) on interpreters' physiological stress measured by electrodermal activity (EDA), heart rate, and heart rate variability (HRV). In the co-located condition, the interpreter and the participant were in the same room; in the not co-located condition, the interpreter was available remotely via chat; while in the virtual condition, the interpreter provided support through a virtual booth (Chmiel et al., 2025, p. 218). We also calculated the correlations between self-reported state anxiety levels, measured with STAI-X2 (Spielberger et al., 1970) scores (reported in Chmiel et al., 2025) and all physiological measures used in the study. The overall prediction was that stress would be the highest in the not co-located condition and the lowest in the co-located condition. We also hypothesised that the presence of slides would decrease the level of interpreters' stress. Additionally, we expected positive correlations between psychophysiological measures and STAI-X2 scores (Spielberger et al., 1970).

Our participant sample included twenty-one professional interpreters with Polish as their native language (L1) and English as their working language (L2) – a subset of the sample recruited for a larger project reported in Chmiel et al. (2025). Participants were asked to interpret three speeches from English into Polish, each presenting one country based on reports prepared by the European Commission (Environmental Implementation Review, 2022) and matched for readability. BIOPAC's MP160 data acquisition unit with EDA and ECG (electrocardiogram) amplifiers was used to collect physiological data. The EDA metrics included mean skin conductance level, conductance level range; skin conductance response frequency, and skin conductance response amplitude. We used heart rate and two HRV measures: standard deviation of normal-to-normal intervals and root mean square successive differences.

For the EDA measures, we observed the expected effect of access to visual materials: arousal was higher in the no-slides condition for three of the four measures (mean skin conductance level, skin conductance level range, and skin conductance response frequency). However, arousal was not

higher in the not co-located condition. Interestingly, and contrary to expectations, the frequency of skin conductance responses showed the reverse pattern, with higher arousal in the co-located condition. Heart rate was higher in the condition without slides as compared to the condition with slides and, at a variance with our predictions, significantly higher in the co-located condition than in the not co-located and virtual condition. The analysis of HRV measures brought no effects of boothmate presence or access to visual materials. Additionally, we observed weak positive correlations between STAI-X2 scores and three physiological measures: skin conductance response frequency, skin conductance response amplitude, and heart rate.

The present study points to a few aspects to consider when applying physiological methods, such as EDA, heart rate and heart rate variability, in interpreting studies. First, psychophysiology is often triangulated with self-report measurements to provide greater insight into emotional valence. However, due to intrinsic characteristics of physiological and self-report measures, correlations may not always be observed. Second, using physiological measures in a within-subject design may involve a risk of habituation, which can manifest as reduced arousal in conditions presented later in the experiment. In addition, practice or fatigue effects may be present in designs where participants complete a few interpreting tasks within a single session (Korpál & Mellinger, 2025). Finally, it is important to consider the ecological validity of studies employing physiological methods in interpreting, ensuring that these methods do not interfere with the interpreting process itself.

The study examined interpreters' stress levels while performing RSI under different conditions, varying in boothmate location and visual support. While the results concerning the boothmate location were not consistent across all measures in the study, two of them – the frequency of skin conductance responses and heart rate – indicate that interpreters are more stressed with the boothmate present in the same room. In addition, our analysis shows that access to slides decreases stress levels during RSI, which might be linked to the facilitative role of visual input in lowering interpreters' cognitive load. The study results may enhance our understanding of how visual support and boothmate presence influence stress and performance in RSI.

References

- Chmiel, A., Spinolo, N., Korpál, P., Olalla-Soler, C., Rozkrut, P., Kajzer-Wietrzny, M., & Ghiselli, S. (2025). The impact of remote interpreting settings on interpreter experience and performance. *Translation and Interpreting Studies*, 20(2), 212–243. <https://doi.org/10.1075/tis.24125.chm>
- Environmental Implementation Review. (2022). Environmental implementation review. https://environment.ec.europa.eu/law-and-governance/environmental-implementation-review_en#country-reports
- Gieshoff, A. C. (2018). The impact of audio-visual speech input on work-load in simultaneous interpreting (Unpublished PhD dissertation). Universität Mainz.
- Korpál, P., & Mellinger, C. D. (2025). Stress and accent in community interpreting. *Poznań Studies in Contemporary Linguistics*, 61(4), 521–542. <https://doi.org/10.1515/psic-2025-0014>
- Moser-Mercer, B. (2005). Remote interpreting: The crucial role of presence. *Bulletin Suisse de Linguistique Appliquée*, 81, 73–97.
- Roziner, I., & Shlesinger, M. (2010). Much ado about something remote: Stress and performance in remote interpreting. *Interpreting*, 12(2), 214–247. <https://doi.org/10.1075/intp.12.2.05roz>
- Seubert, S. (2019). Visuelle Informationen beim Simultandolmetschen. Frank & Timme.
- Spielberger, C. D., Gorsuch, R., Lushene, R. E., Vagg, P. R., & Jacobs, G. A. (1970). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press.
- Stachowiak-Szymczak, K., & Korpál, P. (2019). Interpreting accuracy and visual processing of numbers in professional and student interpreters: An eye-tracking study. *Across Languages and Cultures*, 20(2), 235–251. <https://doi.org/10.1556/084.2019.20.2.5>

Hearing the difference: sound artifacts, frequency and the psychophysiology of remote interpreting

Dongpeng Pan & Kilian G. Seeber

Faculty of Translation and Interpreting, University of Geneva, Switzerland

Remote interpreting is now a routine part of professional practice, and the audio an interpreter receives is whatever the videoconferencing platform delivers. Interpreters have complained about that audio for years, but the evidence behind the complaints is thin. The speech codecs most platforms rely on restrict the transmitted signal to a narrow band of roughly 200–3,400 Hz (Gibson, 2005), and an AIIC evaluation found that eight of nine platforms failed the frequency-response requirements set for interpreting equipment (AIIC, 2020). Earlier studies documented fatigue and stress in remote settings (Moser-Mercer, 2005; Roziner & Shlesinger, 2010) but could not separate sound from the other differences between remote and on-site work. We designed an experiment that manipulates a single acoustic parameter and records what interpreters report, produce and do, and how their bodies respond. This presentation gives its first, preliminary results.

The overall aim is to establish whether restricting audio bandwidth, as narrowband codecs do, changes interpreters' perceived workload, output quality, compensatory behaviour and physiological arousal during simultaneous interpreting, and whether this differs between audio-only (ARI) and video (VRI) remote interpreting. The present report covers the three measures for which analysis is complete: perceived workload, rated output quality and electrodermal activity.

RQ1. Does restricted audio bandwidth raise perceived workload (NASA-TLX)?; RQ2. Does it lower output quality as judged by blind raters?; RQ3. Does it change tonic or phasic electrodermal activity during the task?; RQ4. Do these effects differ between ARI and VRI?

Seventy-eight professional conference interpreters took part: 45 with French as their A language and 33 with Spanish, German, Russian, Italian or Dutch. Each interpreted four ten-minute English scenes simultaneously in a 2×2 within-participant design that crossed modality (ARI vs VRI) with sound quality (HQ vs LQ). Scene and order were counterbalanced across four lists in a Latin square. HQ met the ISO frequency-response standard for interpreting equipment (125 Hz–15 kHz). LQ was band-pass filtered to 300–3,400 Hz, the telephone band that a norming study showed interpreters encounter most often. Electrodermal activity, photoplethysmographic heart rate and pupil size were recorded throughout each scene, together with every adjustment of a volume dial; participants completed the NASA-TLX (Hart & Staveland, 1988) after each scene. Recordings were scored blind for content, style and presentation by four to six trained raters per language in five languages (25 raters; 69 interpreters, eight excerpts each). Dutch renditions were not evaluated because suitable raters were unavailable at the time of analysis. Heart rate, pupil size, blink rate and volume adjustments are still being analysed and are not reported here.

Workload was analysed for all 78 interpreters with factorial linear mixed models (sound $\times$ modality, participant as a random effect), Holm-corrected across the seven TLX measures. Effect sizes are model contrasts divided by the total (random-effect plus residual) standard deviation. Ratings were analysed with cumulative-link mixed models with raters as random effects; coefficients are cumulative log-odds and are reported with odds ratios. Electrodermal activity, recorded at 128 Hz from fingertip electrodes (Shimmer3; Boucsein et al., 2012), was decomposed with continuous decomposition analysis (Benedek & Kaernbach, 2010) into tonic level and phasic responses. Trial usability was decided from signal quality alone, with criteria from the literature fixed and documented before any condition effect was examined, leaving 287 trials from 75 interpreters. The primary electrodermal analysis used participant-level condition contrasts (paired tests on one value per interpreter per contrast). Mixed and generalised additive models served as robustness checks, cluster-based permutation tested the time course, and JZS Bayes factors (Rouder et al., 2009) quantified evidence for the null.

Perceived workload rose under degraded sound. Across all 78 interpreters, LQ raised overall NASA-TLX workload by 4.5 points on the 0–100 scale ($d = 0.34$, Holm-corrected $p = .002$). The largest effects were on frustration ($d = 0.42$), self-rated performance ($d = 0.32$), physical demand ($d = 0.30$) and effort ($d = 0.25$). Temporal demand was lower under LQ ($d = -0.25$), and mental demand did not differ. ARI was rated more demanding than VRI (5.7 points, $d = 0.44$, $p < .001$), and sound quality and modality did not interact ($p = .80$).

Rated quality fell. Across the five evaluated languages, blind raters were less likely to give LQ renditions a higher content score (cumulative-logit coefficient -0.28 , odds ratio 0.75, $p = .004$) or a higher overall score (-0.25 , odds ratio 0.78, $p = .015$). There was no evidence that the penalty differed between languages, and style and presentation were not affected.

Electrodermal activity, so far, did not follow. In the 75 interpreters with usable recordings, tonic skin-conductance level, phasic response rate and integrated response amplitude did not differ between LQ and HQ (tonic: -0.04 z, $p = .71$; response rate: log-ratio -0.05 , $p = .50$). Nor did they differ between VRI and ARI (tonic: -0.10 z, $p = .41$) or in the interaction ($p = .70$). No cell comparison reached a standardised difference of 0.1. Cluster-based permutation found no window of difference over the ten minutes, and the mixed and generalised additive models agreed. Bayes factors of 5.6 to 7.7 in favour of the null provide moderate support for the absence of an effect on these particular measures. A strong presentation-order effect in the same recordings shows that the measures did register change over the session.

So far, narrowband audio increased interpreters' perceived workload and reduced rated content quality, while the electrodermal measures examined showed no detectable change during the ten-minute tasks. These results are consistent with the case for enforcing audio-frequency standards in remote interpreting, with workload assessed across six languages and output quality across five. They also raise the possibility that subjective workload, product quality and electrodermal arousal do not move together under degraded sound. Whether the same holds for heart rate, pupil size, blinks and behaviour is being tested, and any conclusion about the physiological cost of degraded audio must wait for those results.

References

- AiIC. (2020). Evaluation of simultaneous interpreting delivery platforms for ISO compliance. International Association of Conference Interpreters.
- Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1), 80–91. <https://doi.org/10.1016/j.jneumeth.2010.04.028>
- Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, 49(8), 1017–1034. <https://doi.org/10.1111/j.1469-8986.2012.01384.x>
- Gibson, J. D. (2005). Speech coding methods, standards, and applications. *IEEE Circuits and Systems Magazine*, 5(4), 30–49. <https://doi.org/10.1109/MCAS.2005.1550167>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North-Holland.
- Moser-Mercer, B. (2005). Remote interpreting: Issues of multi-sensory integration in a multilingual task. *Meta*, 50(2), 727–738. <https://doi.org/10.7202/011014ar>
- Roziner, I., & Shlesinger, M. (2010). Much ado about something remote: Stress and performance in remote interpreting. *Interpreting*, 12(2), 214–247. <https://doi.org/10.1075/intp.12.2.05roz>
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2), 225–237.
- Seeber, K. G., & Fox, B. (2021). Distance conference interpreting. In M. Albl-Mikasa & E. Tiselius (Eds.), *The Routledge handbook of conference interpreting* (pp. 491–507). Routledge. <https://doi.org/10.4324/9780429297878-43>

Computer-Assisted Interpreting

Thursday 1 October 2026, 14:00–15:15

The impact of real-time CAI support in simultaneous interpreting on listeners' retention of the original content and intention

Nele Kirstein, Anja Rütten & Magdalena Lindner-Juhnke

TH Köln – University of Applied Sciences, Cologne, Germany

The use of Computer-Aided Interpreting tools providing live prompting of terminology, numbers, or a full running transcript in simultaneous interpreting has so far mainly been discussed under the aspect of cognitive load or attention for the interpreters, or its effect on accuracy and precision (Su and Li 2025; Prandi 2025; Chan 2025; Cheung and Li 2023; Defrancq and Fantinuoli 2020, Du and Wang 2025, Frittella et al. 2025, Li and Chmiel 2024). While research points at CAI tools potentially contributing to a more accurate and precise rendition especially in numbers and terminology, the question whether such a CAI-supported rendition leads to a better understanding and especially retention of the overall message of the source text will be a perspective that is both relevant and enriching to the discussion. The ultimate motivation is to gain insights on the possible added value of CAI tools for the audience.

The proposed research aims at comparing how much listeners understand and remember of the content of a speech interpreted with or without CAI support.

Are there measurable differences between the knowledge retention after listening to an interpreted target text with vs. without CAI support? If so, are there correlations between the characteristics of the respective target texts, or particular parts thereof, and the respective knowledge retention results?

Four English source texts were generated for the purpose of studying the effects of CAI support. They were all approx. 16-18 minutes long and comparable in terms of topic, degree of technicality and structure. Complicated terminology, proper names, and numbers were incorporated. All four texts were read by the same native speaker and video-recorded. They were interpreted into German by several interpreters with significant professional experience. Both test subjects interpreted two speeches without CAI support, one with well-functioning CAI support and one with dysfunctional CAI support, i.e. which unexpectedly occurred due to a technical glitch, but constitutes an additional interesting set of data due to the additional distraction. Optionally, a machine-interpreted version of the text will be generated. While this is not the primary focus of the study, it may provide interesting insights complementing Matsushita's (2026) study on human vs. machine interpreting.

In a first step, an analysis of the different target texts will be performed using the AMI methodology (Rütten 2026), which provides a segment-wise categorisation of different types of terminological, semantic, pragmatic and delivery-related deviations between source and target, providing a granular basis to identify potential correlations between such deviations and the knowledge retention of the original text.

In a second step, a questionnaire will be elaborated to test the retention of factual information as well as the speaker's intention. Several listeners of one CAI and one non-CAI target text will answer the questionnaire. The methodology will be based on Reithofer's (2013) and Matsushita's (2026) measurements of comprehension-based evaluations of interpreting vs. English as a lingua franca and human vs. machine interpreting respectively. The questions will refer to the content of the original text, even if this is altered or omitted in the target text. In case listeners seem to remember content from the original that wasn't explicitly rendered in the target text this might provide insights on the question whether listeners infer on missing parts of a speech if the surrounding content is rendered. It might point towards further research on the merits of completeness vs. coherence,

delivery and plausibility and other typical strategies in simultaneous interpreting (Kohn and Kalina 2002) in cases where details like complicated numbers are clearly rendered on the basis of the CAI support, it will be interesting to see if such information is later remembered by listeners.

The final step will be matching the results from step 1 and 2, mapping existing or non existing correlations between deviations from source to target and the results of the retention questionnaire.

The first, “intermediate” product will be a qualitative and quantitative assessment of the deviations of the different target texts (with and without CAI) from the source text. Deviations in terms of different aspects of terminology, content, language, context and delivery will be rated and a statistical overview provided. On this basis, a comparison between the different target texts can be made. The second step will provide the listeners’ perspective and provide insights on how much of the source text content and speaker’s intentions are understood and remembered by listeners. Based on the matching between the quality analysis and the retention analysis, insights can be gained on the kinds of deviations that do or do not lead to a loss of information in the chain of interlingual communication and whether CAI support on one hand and human interpreting strategies on the other hand have an impact.

This work aims at broadening the perspective in research on CAI tools by looking at the ultimate goal of interpreting, i.e. bringing the message across.

References

- 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*, 33(1), 73–102. <https://doi.org/10.1075/target.19166.def>
- Du, M., & Wang, B. (2025). Cognitive processing in simultaneous interpreting with synchronous captioning: An eye-tracking study of Chinese head-final noun phrases with varied dependency distances. *System*, 133, 103761. <https://doi.org/10.1016/j.system.2025.103761>
- Frittella, F. M., Prandi, B., & Rodríguez, S. (2025). How usable are CAI tools in professional interpreting settings? Insights from a highly technical assignment. *The Interpreters’ Newsletter*, 30. <https://www.openstarts.units.it/handle/10077/37794>
- Kohn, K., & Kalina, S. (2002). The Strategic Dimension of Interpreting. *Meta*, 41(1), 118–138. <https://doi.org/10.7202/003333ar>
- Li, T., & Chmiel, A. (2024). Automatic subtitles increase accuracy and decrease cognitive load in simultaneous interpreting. *Interpreting. International Journal of Research and Practice in Interpreting*, 26(2), 253–281. <https://doi.org/10.1075/intp.00111.li>
- Matsushita, K. (2026). Understanding AI interpreting in context A comprehension-based evaluation of human vs. Machine-generated interpretations in a real-world setting. *International Journal of Language, Translation and Intercultural Communication*, 11, 71–85. <https://doi.org/10.12681/ijltic.44192>
- Prandi, B. (2025). Computer-assisted interpreting (CAI) tools and CAI tools training. In E. Davitti, T. Korybski, & S. Braun, *The Routledge Handbook of Interpreting, Technology and AI* (1st edn, pp. 123–144). Routledge. <https://doi.org/10.4324/9781003053248-11>
- Reithofer, K. (2013). Comparing modes of communication: The effect of English as a lingua franca vs. interpreting. *Interpreting. International Journal of Research and Practice in Interpreting*, 15(1), 48–73. <https://doi.org/10.1075/intp.15.1.03rei>
- Rütten, A. (2026). Can automatic quality estimation help to pre-assess the quality of human simultaneous interpretation and automatic speech translation? An exploratory case study based on MQM-inspired Assessment Metrics for Interpreting (AMI) and COMET. *Lebende Sprachen*. <https://doi.org/10.1515/les-2025-0046>
- Su, W., & Li, D. (2025). Eye-tracking visual attention of professional interpreters during technology-assisted simultaneous interpreting. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, 24. <https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/818>

The impact of ASR and AI-based CAI tools in RSI to support human interpreting: Performance, cognitive load and stress levels

Valentina Baselli

IULM University / SSML Carlo Bo, Milan, Italy

Over the last few years, the technological turn (Fantinuoli 2018a) has had a profound impact on interpreters' work due to the spread of remote simultaneous interpreting (RSI) and new technologies, e.g., automatic speech recognition (ASR) and AI-based computer-assisted interpreting (CAI) tools. Interpreters use CAI tools to support and facilitate specific aspects of the interpreting process, with the aim of improving quality and productivity (Fantinuoli 2018b). With the integration of Artificial Intelligence (AI), third-generation CAI tools can make use of advanced features such as automatic extraction of terminology and automatic speech recognition (ASR), and they can automatically display on the screen the so-called problem triggers of simultaneous interpreting already identified in the literature (Gile 2009). These problem triggers include numbers, named entities and specialised terminology that may increase cognitive load (Cheung 2024; Korpál & Stachowiak-Szymczak 2020; Prandi 2023). However, studies have found that CAI tools and the source language speech transcript may lead to an increase in cognitive load (Prandi 2023; Seeber 2017), even though recent findings report lower theta power in EEG during SI with subtitles – with 100% and 80% subtitle precision – while no significant differences were observed for 95% and 90% subtitle precision compared to the baseline (Li & Chmiel 2024). These findings seem to suggest that visual prompts can improve accuracy and reduce cognitive load by easing listening and memory demands. Against this backdrop, this proposal aims to investigate the impact of a new AI-based CAI tool for RSI integrated into the Converso Education platform, a Simultaneous Interpreting Delivery Platform (SIDP) specifically developed for interpreter training, with a particular focus on:

- assessing the accuracy and quality of advanced trainee interpreters' performance under three different situations: with CAI tool support, with the aid of an ASR transcript and with a traditional Microsoft Word glossary;
- investigating the cognitive load and stress levels in RSI under the three above-mentioned conditions (CAI tool, ASR transcript and Microsoft Word glossary) with neuroscientific methods.

To answer these research questions, an experimental study was conducted at the Brain&Behaviour Lab at IULM University, with the aim of investigating the cognitive impact of the ASR transcript and AI-based CAI tool on English-Italian RSI by measuring the cognitive load and stress levels during RSI under the three above-mentioned conditions. To assess these variables, the results of electroencephalogram (EEG) in the form of Workload Index (WLI), eye-tracking, Heart Rate Variability (HRV) and Skin Conductance (SC) were triangulated with the subjective method findings of the NASA Task Load Index, the short version of the State-Trait Anxiety Inventory (STAI) and the Short Stress State Questionnaire (SSSQ), as well as interpreters' performance. The participants were 23 conference interpreting trainees in their last semester (end of second year) of the master's degree course in Conference Interpreting at IULM University, the final users of the Converso platform. The task required connecting to the RSI platform and interpreting from English into Italian three speeches within the same domain with the CAI tool short prompts, with ASR transcript and without any AI-based tool – only a Microsoft Word Glossary – in a randomized way. The speeches on the same domain had a similar difficulty level, contained a similar number of problem triggers and were recorded with a standard accent to evaluate the neurophysiological response related to RSI from English to Italian under the three conditions. Based on the literature (Seeber 2017; Prandi 2023), it could be hypothesized that WLI and perceived workload, as indicators of cognitive load, would be higher for the ASR Transcript condition, intermediate in the CAI condition and lower for the condition without AI-based tools, although it is worth considering the potential for computer-assisted interpreting technologies to have a general mitigating effect obtained by augmenting interpreters' cognitive resources by relying on external support (Mellinger 2023). As far as stress levels are

concerned, a mitigating effect of the AI-based tools on SC, HRV and perceived stress, according to Olalla-Soler, Spinolo and Muñoz Martín (Olalla-Soler et al. 2023), was expected.

Data from WLI, HRV and SC – triangulated with eye-tracking, self-reported findings and performance analysis – has been analysed using inferential statistics, revealing specific patterns, particularly when interpreters have to simultaneously interpret English noun phrases (NPs) into Italian.

References

- Cheung, Andrew K. F. (2024). "Cognitive load in remote simultaneous interpreting: place name translation in two Mandarin variants", *Humanities and Social Sciences Communications* volume 11, Article number, 1238.
- Fantinuoli, Claudio (ed.) (2018a). *Interpreting and technology: the upcoming technological turn*, Berlin, Language Science Press, 1-12.
- Fantinuoli, Claudio (2018b). "Computer-assisted interpreting: challenges and future perspectives", in: Isabel Durán and Gloria Corpas Pastor (eds.), *Trends in e-tools and resources for translators and interpreters*, Leiden, Brill, 153-174.
- Gile, Daniel (2009), *Basic Concepts and Models for Interpreter and Translator Training*, Amsterdam/Philadelphia, John Benjamins Publishing Company.
- Korpál, Paweł & Stachowiak-Szymczak, Katarzyna (2020). "Combined problem triggers in simultaneous interpreting: exploring the effect of delivery rate on processing and rendering numbers", *Perspectives* 28(1):126-143.
- Li, Tianyun & Chmiel, Agnieszka (2024). Automatic subtitles increase accuracy and decrease cognitive load in simultaneous interpreting. *Interpreting*, 26(2), 253–281.
- Mellinger, Christopher D. (2023). "Embedding, extending, and distributing interpreter cognition with technology", in G. Corpas Pastor and B. Defrancq, *Interpreting Technologies – Current and Future Trends*, Amsterdam/Philadelphia, John Benjamins Publishing Company.
- Olalla-Soler, Christian, Spinolo Nicoletta, Muñoz Martín Ricardo (2023). "Under Pressure? A Study of Heart Rate and Heart-Rate Variability Using the SmarTerp CAI tool", *Hermes – Journal of Language and Communication in Business* 23.
- Prandi, Bianca (2023). *Computer-assisted simultaneous interpreting: A cognitive-experimental study on terminology*. (Translation and Multilingual Natural Language Processing 22), Berlin, Language Science Press.
- Seeber, Kilian G. (2017). "Multimodal Processing in Simultaneous Interpreting", in J. W. Schwieter, A. Ferreira (eds.), *The Handbook of Translation and Cognition*, John Wiley & Sons, Inc., 461-475.

Artificial Intelligence in Interpreting

Friday 2 October 2026, 10:30–12:00

How good is AI? A quality evaluation of FAST output in seven languages

Anja Rütten, Anne Catherine Gieshoff, Bianca Prandi, Agnieszka Chmiel, Csilla Szabó, Maria Abad Colom, Roxane Zuccherro, Natacha Body, Elisabet Tiselius & Lorenz Mohler

TH Köln, Germany; ZHAW Zurich University of Applied Sciences, Switzerland; Università di Bologna, Italy; Adam Mickiewicz University, Poznań, Poland; Budapest University of Technology and Economics, Hungary; OsloMet – Oslo Metropolitan University, Norway; Université de Mons, Belgium; Stockholm University, Sweden

Fully automatic speech to speech translation systems (FAST; Pöchhacker, 2024) are designed to provide the target speech simultaneously with the source speech. Such solutions have the potential to offer real time renditions during work meetings, plenary sessions, or other gatherings in international organizations, such as the EU or the UN, which have so far relied on conference interpreting services. One prerequisite, among others, for deploying such tools in diplomatic settings is that all delegates must be able to participate in debates and discussions on equal terms, irrespective of their language. Systematic disparities could limit access to information for some delegations and may consequently compromise diplomatic negotiations. This implies that FAST should deliver acceptable and comparable output quality in all languages.

Both the EU and the UN have already begun piloting FAST systems. A study based on speeches delivered by members of the European Parliament in 2022 reports human assessments of FAST renditions. The rating categories were taken from the Multidimensional Quality Metrics (MQM) framework (Lommel et al., 2024). Although small, the sample suggests that output quality depends not only on the target language but also on the source language. Translations from French or Spanish into Italian, for instance, contained fewer errors than translations from Romanian into Italian (Di Nuovo, 2023). A similar investigation was conducted by the WHO interpreting team (WHO Interpretation Team, 2025). They rated the machine translated speeches in their six official languages on content, delivery, and form. A 75% accuracy threshold was defined as the minimum for an acceptable translation. Unsurprisingly, the best score was achieved for English (51%) and the worst for translations into Chinese (35%). Still, the average score was unacceptable for all six languages, including English.

These two pilot studies relied on expert ratings for their data analysis, that is, ratings produced by conference interpreters using scales designed to assess the quality of interpretations. A more user centered perspective was taken by Fantinuoli and Prandi (2021) and by Matsushita (2026). Fantinuoli and Prandi (2021) compared the intelligibility and informativeness—e.g., sense consistency—of AI translations and human interpretations and found that AI translations were less intelligible but comparable in informativeness. Similar results were obtained in a reception study by Matsushita (2026). She reported lower listener comprehension after listening to the AI translation compared to the interpretation, which she explains mainly by the lack of prosodic salience (p. 74), that is, the lack of meaning-based stress and pauses, increasing the listener's cognitive load.

The present study contributes to the question of quality in FAST output by combining error based and user centred assessments of AI translations in seven European languages. The source and target speech material was taken from a mock conference organized at ZHAW and translated by a commercial FAST provider. The source speech was delivered in English and translated into seven languages: French, German, Hungarian, Italian, Norwegian, Polish, and Swedish. Participants accessed the AI translations individually on their smartphones or laptops by scanning a QR code or following a link. The AI translations were provided in spoken form (speech translation) and written

form (live captions), meaning that participants could either read the live captions or use earphones to listen to the speech translation.

The quality assessment is currently ongoing. For the user centered assessment, we apply the methodology developed by Tiselius (2009) and later used by Fantinuoli and Prandi (2021). The source speech was automatically transcribed, checked against delivery, and divided into idea units by the first author. Each idea unit consists of three to four sentences. The segmentation was based on semantic content, slide changes or bullet points, prosody (pauses, falling intonation), typical discourse markers indicating the start of a new idea (“Now”, “So”, etc.), and the use of pronouns. The segmentation was discussed with an experienced conference interpreter, and uncertainties were resolved. Each target speech will be assessed by two raters who first rate intelligibility and then informativeness on a scale from 1 (totally unintelligible/completely different target speech) to 6 (completely intelligible/completely consistent). For the error based assessment, we use the framework developed by Rütten (2026), which is inspired by the MQM framework and quality assessment criteria from interpreting studies. For this comprehensive analysis, we will select three idea units based on the informativeness scores: 1) the unit with the highest score across languages, 2) the unit with the lowest score across languages, and 3) the unit with the highest variance across languages. The rating is done for each sentence of these idea units, resulting in approximately twelve to sixteen assessments per language. We expect informativeness and intelligibility scores and error profiles to vary between languages such that languages of lesser diffusion such as Hungarian or Swedish will be rated lower in intelligibility and informativeness scores. This study is among the first to systematically compare intelligibility, informativeness and error profiles of FAST renditions of the same source text across different European languages.

References

- Di Nuovo, E. (2023). An experiment in error analysis of real-time speech machine translation using the example of the European Parliament’s Innovation Partnership. *Proceedings of the Ninth Italian Conference on Computational Linguistics*, 185–195.
- Fantinuoli, C., & Prandi, B. (2021). Towards the evaluation of automatic simultaneous speech translation from a communicative perspective. *Proceedings of the 18th International Conference on Spoken Language Translation (IWSLT 2021)*, 245–254. <https://doi.org/10.18653/v1/2021.iwslt-1.29>
- Gillies, A. (2025, June 2). WHO report on AI interpretation [LinkedIn]. https://www.linkedin.com/posts/andy-gillies-interpreter-french-german-polish-english_who-report-on-ai-interpretation-activity-7335194739979272192-gldS/
- Lommel, A., Gladkoff, S., Melby, A., Wright, S. E., Strandvik, I., Gasova, K., Vaasa, A., Benzo, A., Sparano, R. M., Foresi, M., Innis, J., Han, L., & Nenadic, G. (2024). The Multi-Range Theory of Translation Quality Measurement: MQM scoring models and Statistical Quality Control (arXiv:2405.16969). *arXiv*. <https://doi.org/10.48550/arXiv.2405.16969>
- Matsushita, K. (2026). Understanding AI interpreting in context: A comprehension-based evaluation of human vs. Machine-generated interpretations in a real-world setting. *International Journal of Language Translation and Intercultural Communication*, 11, 71–85. <https://doi.org/10.12681/ijltic.44192>
- Rütten, A. (2026). Can automatic quality estimation help to pre-assess the quality of human simultaneous interpretation and automatic speech translation? An exploratory case study based on MQM-inspired Assessment Metrics for Interpreting (AMI) and COMET. *Lebende Sprachen*. <https://doi.org/10.1515/les-2025-0046>
- Tiselius, E. (2009). Revisiting Carroll’s scales. In C. V. Angelelli & H. E. Jacobson (Eds), *Testing and Assessment in Translation and Interpreting Studies: A call for dialogue between research and practice*. John Benjamins Publishing Company.
- WHO interpretation team. (2025). Report on Wordly AI Interpretation. <https://www.atanet.org/wp-content/uploads/2025/06/Report-on-WORDLY-AI-Interpretation.pdf>

From Error Analysis to Audience Reception: Assessing Selected Aspects of Machine Interpreting Quality A.D. 2026 in the Polish–English Language Pair

Tomasz Korybski, Wojciech Figiel & Małgorzata Tryuk

University of Warsaw, Poland

Despite growing commercial claims that machine interpreting (MI) has reached or is approaching near-human performance, empirically grounded, ecologically valid assessments of this technology remain scarce. This presentation builds on a study conducted in June 2025 at the Institute of Applied Linguistics, University of Warsaw, in which human simultaneous interpreting was compared against output from two leading MI service providers in the Polish–English language pair (Korybski et al., 2026). The original experiment simulated a real-life conference setting in which EU-accredited interpreters, MA-level interpreting students, and two MI systems worked in parallel during a live event comprising a Polish introductory speech, a 40-minute English lecture, and a bidirectional Q&A session. Error analysis using a simplified approach adapted from Barik’s (1971) typology and subsequent frameworks — including the NER and NTR (Romero-Fresco and Pöschhacker, 2017) and their adaptations for interpreting research (Davitti and Sandrelli, 2020; Korybski and Davitti, 2024) — revealed a clear accuracy hierarchy: accredited interpreters outperformed students, who in turn outperformed both MI systems. MI output was characterised by a disproportionately high share of major, meaning-distorting errors, including ASR-related misrecognitions, overly literal renditions, voiced punctuation, random language switches, gender bias, and failures attributable to limited contextual memory.

The present paper extends this research in two significant directions. First, it reports on a controlled reception study (implemented in April and May 2026) designed to complement the error-based accuracy assessment with audience perspective data. In total, 48 respondents — the majority of whom were non-interpreters and non-translators — completed a specially designed questionnaire in which they were exposed to a sequence of two audio recordings: one excerpt from MI output and one from human interpreting output, presented in chronological order and randomised across eight questionnaire versions, four per each language. The randomisation was applied separately to both languages included in the study (Polish and English), ensuring that the sequence of MI and human interpreting stimuli varied across respondents and that order effects could be controlled for. Participants assessed each recording by responding to three targeted statements on a Likert scale and answering two open-ended questions, allowing them to articulate their subjective impressions and indicate which performance they considered superior overall.

The findings of the reception study offer a lay audience perspective that enriches the expert error-analysis results. While detailed outcomes will be presented and discussed at the conference, the data allow for a direct comparison between objective quality metrics derived from error counts and the subjective quality judgements of non-specialist listeners — an audience type that is the primary end-user of MI services in real conference settings. Unlike expert evaluators, the lay respondents had no access to the source speech; their judgements were instead grounded in perceived logical coherence, linguistic correctness, and the prosodic qualities of the output — how natural, fluent, and pleasant to listen to each interpretation sounded. The data thus shed light on a relatively underexplored dimension of machine interpreting quality assessment: what non-specialist audiences actually notice, and how far this maps onto the product-focused and error-based criteria central to scholarly evaluation.

Second, the presentation addresses the methodological challenges inherent in designing and conducting a controlled reception study of this kind. Comparing machine-generated and human interpreting output in an experimental setting raises a range of issues that have not yet been systematically discussed in the literature: selecting audio stimuli that are representative yet comparable in length and difficulty; managing the risk of listener fatigue and order effects; determining an appropriate sample of respondents, ensuring sufficient ecological validity while

maintaining experimental control; and, crucially, dealing with the fundamental asymmetry between MI and human output in terms of voice, prosody, and fluency — features that may influence listener perception independently of semantic accuracy. The presentation reflects on the decisions made in the course of designing the current study and proposes a preliminary framework for reception studies in this rapidly evolving area of research.

Third, the presentation reports on a follow-up experiment planned for May 2026, which replicates the core design of the original June 2025 study. This replication is intended to track any measurable progress in MI quality over a twelve-month period, providing longitudinal data on the pace and nature of technological development in this domain. Given the rapid release cycles characteristic of the MI industry, even a twelve-month window may capture meaningful shifts in system performance — or, conversely, demonstrate that certain quality limitations are persistent and structural rather than incidental. The follow-up study thus serves a dual function: it contributes to the empirical record on MI quality and provides a methodological test of the replicability and sensitivity of the original study design.

Taken together, the three strands of this presentation — error analysis, reception study with its challenges, and longitudinal replication — contribute to a more comprehensive, multi-method picture of MI quality than a single study approach could provide. In a field where rapid technological claims often outpace independent evaluation, this research advocates for speedy, highly replicable and empirically grounded assessment that considers both expert and lay perspectives, and that is designed to keep pace with the technology it seeks to evaluate.

References

- Barik, H. C. (1971). A description of various types of omissions, additions and errors of translation encountered in simultaneous interpretation. *Meta*, 16(4), 199–210. <https://doi.org/10.7202/001972ar>
- Davitti, E., & Sandrelli, A. (2020). Embracing the complexity: A pilot study on interlingual respeaking. *Journal of Audiovisual Translation*, 3(2), 103–139. <https://doi.org/10.47476/jat.v3i2.2020.135>
- Korybski, T., & Davitti, E. (2024). Human agency in live subtitling through respeaking: Towards a taxonomy of effective editing. *Journal of Audiovisual Translation*, 7(2), 1–22. <https://doi.org/10.47476/jat.v7i2.2024.302>
- Korybski, T., Figiel, W., Tryuk, M., & Górník, M. (2026). Experimenting with machine interpreting in the PL-EN language pair: Are we (getting) close to “human-like” quality? *International Journal of Language, Translation and Intercultural Communication*, 11, 1–14.
- Romero-Fresco, P., & Pöchhacker, F. (2017). Quality assessment in interlingual live subtitling: The NTR model. *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 16, 149–167.

Algorithms in the booth: why and how to automate interpreting quality assessment

Agnieszka Chmiel, Danijel Koržinek, Marta Kajzer-Wietrzny & Paulina Rozkrut

Adam Mickiewicz University, Poznań, Poland; Polish-Japanese Academy of Information Technology, Warsaw, Poland

Imagine that you are an interpreting trainee and you get detailed written feedback about each of your in-class interpretations, with specific recommendations. You also get further exercises and personalised exercise speeches that include your weak points so that you can more efficiently work on improving your overall interpreting skill. Or imagine you are using automatic speech translation when speaking to someone with whom you do not share a language. You can see a quality indicator that makes you doubt the machine output and immediately ask your interlocutor for confirmation or clarification to make sure the communicative process is successful. These scenarios might soon become reality as recent technological advances facilitate automation of interpreting quality assessment. Such developments could potentially streamline training of human interpreters and offer guardrails when using automatic speech translation.

Previous studies on automatic quality assessment focused on comparing various automatic metrics with human judgment. The analyses showed moderate-to-strong correlations (Wang & Fantinuoli, 2024; Wang & Wang, 2025b). Additionally, such fluency indicators as filled pauses used by interpreters were shown to be associated with human quality scores (Han et al., 2025). LLMs have recently been used successfully to assess language quality in interpreting (Wang & Wang, 2025a). However, these studies have largely focused on evaluations of interpretations performed by trainee interpreters, rather than professionals, and predominantly in the English-Chinese language combination (Han & Lu, 2025; Wang & Wang, 2025b). As a notable exception, Wang and Fantinuoli (2024) focused on the English-Spanish language pair but used interpretations by three professionals only.

This is an initial exploratory study in the AQUAPINC project that aims to develop a set of automatic quality evaluation procedures to best suit interpreting quality assessment. We aim to extend the approach by Han et al. (2025) and Wang and Wang (2026) who suggest that aggregation of various automatic metrics can efficiently approximate human judgment. We used the Polish Interpreting Corpus (PINC) (Chmiel et al., 2022) as our dataset. PINC is a 220,000-token corpus of professional interpretations from the European Parliament with a rich set of metadata. We analysed professional interpretations from 39 interpreters with the Polish-English language combination, a language pair previously unexplored in the studies of automatic quality assessment. The choice of this language pair seems well justified because Polish ranks first as the best performing language for complex AI tasks and large language model evaluation (Kim et al., 2025) and English is the top high-resource language. In our analysis, we first looked at correlations of human and automatic quality assessment. We then explored source and text characteristics that can help predict automatic quality scores. Finally, we evaluated to what extent reference translations can finetune automatic quality evaluation of interpreting.

In the first stage, we used two automatic quality metrics, COMET and GEMBA, to estimate quality of these interpretations. We employed well-established reference-free models - specifically, the wmt23-cometkiwi-da-xl model (Rei et al., 2022) for COMET. In parallel, we used GEMBA-DA (Kocmi & Federmann, 2023), which has been shown to outperform other reference-free metrics and to achieve performance close to that of its reference-based counterpart. We updated the existing GEMBA metric by using the newer model (GPT 5.5 instead of GPT 4.0 used in the original metric). We then performed a series of expert evaluations. First, two experts provided independent quality scores on a subset of data from the PINC corpus. Their inter-rater reliability was low (0.2) and correlations with automatic scores were moderate ($r=0.6$). Then, three expert raters participated in a calibration session and assessed a subset of interpretations with a validated rubric scale (Han et

al., 2025), achieving an inter-rater reliability of 0.4 and correlations for information completeness reaching $r=.77$ for COMET and $r=.63$ for GEMBA-DA.

In the second stage, we focused on source and target speech characteristics and their associations with automatic scores. We fitted two linear mixed-effects models with PINC corpus metadata as potential predictors of quality for the two automatic quality scores. The predictors included source speech mode of delivery (read-out vs. impromptu), delivery rate and syntactic complexity (operationalised as mean dependency distance), as well as target speech fluency (operationalised as mean duration of silent and filled pauses) and interpreters stress (operationalised as fundamental frequency). Our analyses showed that mode of delivery and interpreter fluency predicted both automatic scores. Read-out speeches and disfluent interpretations received lower automatic quality scores for COMET and GEMBA-DA. Higher syntactic complexity of the source text led to lower GEMBA scores, while higher fundamental frequency of the interpreter's voice in return (used as an index of stress) was associated with higher quality scores.

In the third stage, we evaluated the same dataset from PINC with reference-based COMET and updated GEMBA metrics. Reference-based metrics (which compare to-be-evaluated target text with a model reference translation) are said to perform better than reference-free metrics (Kocmi & Federmann, 2023). We utilised translated verbatim reports which used to be produced for the European Parliament plenaries and controlled for the differences between edited verbatim reports and actual transcriptions of the source texts with WER, a word error rate typically used to evaluate speech recognition. We found high correlations between reference-based and reference-free metrics ($r=0.97$ for GEMBA and $r=.71$ for COMET). We also found that reference-based GEMBA scores but not reference-based COMET scores significantly predict human evaluations.

Our findings suggest that automatic quality assessment, particularly LLM-based approaches such as GEMBA, shows promise for approximating expert judgments of professional interpreting quality. Additionally, including source-text and target-text characteristics in quality evaluation may finetune alignment with human ratings. As an exploratory contribution, this work expands automatic quality assessment to professional interpreting in the Polish-English pair and provides a foundation for developing more robust, multi-dimensional evaluation procedures. Subsequent stages of the project will include correlations of automatic metrics with user evaluation scores and finetuning GEMBA prompts to arrive at the most comprehensive quality evaluation scores. By testing various approaches and tools, we aim to provide guidelines for constructing the most efficient automatic quality evaluation procedure suited for both human interpreting and automatic speech translation.

References

- Chmiel, A., Kajzer-Wietrzny, M., Koržinek, D., Janikowski, P., Jakubowski, D., & Polakowska, D. (2022). Fluency parameters in the Polish Interpreting Corpus (PINC). In M. Kajzer-Wietrzny, S. Bernardini, A. Ferraresi, & I. Ivaska (Eds.), *Empirical investigations into the forms of mediated discourse at the European Parliament*. Language Science Press.
- Han, C., & Lu, X. (2025). Beyond BLEU: Repurposing neural-based metrics to assess interlingual interpreting in tertiary-level language learning settings. *Research Methods in Applied Linguistics*, 4(1), 100184. <https://doi.org/10.1016/j.rmal.2025.100184>
- Han, C., Lu, X., & Chen, S. (2025). Modeling rater judgments of interpreting quality: Ordinal logistic regression using neural-based evaluation metrics, acoustic fluency measures, and computational linguistic indices. *Research Methods in Applied Linguistics*, 4(1), 100194. <https://doi.org/10.1016/j.rmal.2025.100194>
- Kim, Y., Russell, J., Karpinska, M., & Iyer, M. (2025). One ruler to measure them all: Benchmarking multilingual long-context language models (arXiv:2503.01996). arXiv. <https://doi.org/10.48550/arXiv.2503.01996>
- Kocmi, T., & Federmann, C. (2023). Large language models are state-of-the-art evaluators of translation quality. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation* (pp. 193–203). European Association for Machine Translation. <https://aclanthology.org/2023.eamt-1.19>

Rei, R., Guerreiro, N. M., Van Stigt, D., Treviso, M., Coheur, L., de Souza, J. G., & Martins, A. F. (2023). Scaling up CometKiwi: Unbabel-IST 2023 submission for the quality estimation shared task. In Proceedings of the Eighth Conference on Machine Translation (pp. 841–848).

Wang, X., & Fantinuoli, C. (2024). Exploring the correlation between human and machine evaluation of simultaneous speech translation (arXiv:2406.10091). arXiv. <https://doi.org/10.48550/arXiv.2406.10091>

Wang, X., & Wang, B. (2025a). Advancing automatic assessment of target-language quality in interpreter training with large language models: Insights from explainable AI. *The Interpreter and Translator Trainer*, 19(3–4), 465–485. <https://doi.org/10.1080/1750399X.2025.2533015>

Wang, X., & Wang, B. (2025b). Speech-to-speech assessment vs text-based assessment of simultaneous interpreting: Tapping into the potential of large language models and the metrics of machine translation quality estimation in enabling automatic assessment. *New Series*.

Wang, X., & Wang, B. (2026). Neural network models vs. MT evaluation metrics: A comparison between two approaches to automated assessment of information fidelity in consecutive interpreting. *Humanities and Social Sciences Communications*. <https://doi.org/10.1057/s41599-026-06562-z>

Posters

Poster session · Friday 2 October 2026, 13:30–14:15

Poster 1 · CAI in the Booth – Is It Worth the Risk? A Study on Quality and Risk Perception among Professionals

Cinzia Hirschvogel & Katja Voncina

University of Vienna, Vienna, Austria

The rise of computer-assisted interpreting (CAI) tools has sparked intense debate within the interpreting profession. Advocates highlight their potential to support accuracy, consistency, and efficiency in high-pressure conference settings (Defrancq & Fantinuoli 2020; Fantinuoli & Pisani 2021; Prandi 2023), while critics warn of risks to data protection, quality, concentration, and professional identity (Horváth 2022; Peng and Dai 2023; Guo and Li 2025). Despite growing technological developments and increasing exposure to CAI in the booth, little is known about how these tools are perceived by interpreters.

The theoretical framework of this study integrates cognitive, risk-oriented, and quality-based approaches within interpreting studies. It draws on Seeber's (2011) Cognitive Load Model, which conceptualizes simultaneous interpreting as the coordination of competing processing efforts under limited cognitive resources. Risk is conceptualized in line with Pym's (2025) notion of translation and interpreting as risk management, complemented by process-oriented perspectives on interpreter decision-making (Kalina 2015). The quality dimension is grounded in the classical criteria established by Bühler (1986) and further contextualized by empirical investigations about interpreters' self-perception (Zwischenberger & Pöschhacker 2010). By combining cognitive load theory, risk management, and multidimensional quality models, the framework enables a structured analysis of interpreters' perceptions and strategic behaviour in ASR-supported CAI environments.

The aim of the study is to investigate how professional conference interpreters perceive their own output quality and assess the associated cognitive and professional risks when using a third-generation, ASR-based CAI tool during simultaneous interpreting.

By focusing on practitioners' own evaluations, this study addresses the following research questions: RQ1: How do professional conference interpreters subjectively evaluate their output quality when interpreting simultaneously with the support of a CAI tool (TerpMate), compared to interpreting without it?; RQ2: How do interpreters experience and assess the cognitive demands and professional risks of using a third-generation, ASR-based CAI tool in the booth?; RQ3: What specific risks do professional interpreters identify when using CAI tools, and how do they strategically manage those risks during the interpreting task?

Adopting a mixed-methods design, the study involves ten professional interpreters in a three-phase procedure: (1) a pre-task questionnaire assessing anticipated risks and expected benefits of ASR support; (2) an ASR-supported simultaneous interpreting task using the CAI tool TerpMate; and (3) retrospective interviews after the interpreting task.

The interpreting performance in the second phase was not evaluated; rather, the experimental task served solely as a basis for eliciting participants' perceptions of working with a CAI tool.

The results of the questionnaire and the interviews with professional conference interpreters reveal a differentiated and largely critical picture of CAI tool use, organized around several key findings. Across all participants, the tool's strengths were specific and well-defined: numbers, proper names, and glossary terms – precisely the elements that prior ASR research has shown to benefit most from visual support. Outside this domain, participants were largely unconvinced, though most were quick to note that their familiarity with the tool was limited.

The central challenge was attention management. Most participants described the transcription as an additional distractor alongside the audio and presentation slides, finding it impossible to process more than two channels simultaneously. A selective strategy – looking at the screen only for numbers or anticipated terminology – was reported by more experienced participants, while less experienced ones tended to be pulled reflexively toward continuous reading.

Self-perceived quality suffered across multiple dimensions: fluency and prosody were disrupted by the act of consulting the screen, coherence was frequently lost when attention shifted to the transcription, and completeness showed a paradoxical redistribution: fewer failures at the level of isolated numbers, but more failures at the level of discourse segments and logical connectives.

The interviews confirmed risks which were anticipated by the participants in the pre-task questionnaire such as cognitive overload and ASR inaccuracy, while also revealing unanticipated ones: tool suggestions conflicting with the interpreter's own analytical processes and the risk of slipping from simultaneous interpreting into a sight translation mode from a flawed transcript. On latency, participants with a faster processing tempo found TerpMate too slow to be useful, while those with a longer ear-voice span found it more manageable, suggesting that latency tolerance is substantially person-dependent rather than a fixed system parameter. Regarding booth collaboration, all participants expressed concern that the tool's visual demands would erode traditional partner support. On adoption, most participants were conditionally open to future use, with at least five independently requesting a selective display mode showing only numbers and glossary hits. The least experienced participant was notably more positive overall, suggesting the cost-benefit ratio of CAI tools may be more favourable earlier in professional development.

The findings imply that effective integration of CAI tools requires targeted user training, careful consideration of cognitive load, and improvements in tool reliability to ensure that potential benefits for interpreting quality can be realized without compromising core interpreting processes.

References

- Bühler, H. (1986). Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua*, 5(4), 231–235.
- Defrancq, B. & Fantinuoli, C. (2020). Automatic speech recognition in the booth. In: *Target-international Journal of Translation Studies*, Volume 33, Issue 1, Mar 2021. 73 – 102. DOI: <https://doi.org/10.1075/target.19166.def>.
- Fantinuoli, C. & Pisani, E. (2021). Measuring the impact of automatic speech recognition on interpreter's performances in simultaneous interpreting. In: Wang, Caiwen & Zheng, Binghan (eds.): *Empirical studies of translation and interpreting: the post-structuralist approach*. London/New York: Routledge, 181–196.
- Gile, D. (2021). Risk management in translation: How much does it really explain? *International Journal of Interpreter Education*, 13(1), 56–65.
- Guo, Z. & Li, K. (2025). Limitations and Best Practices for the Design of Computer-Aided Interpretation. In: Degen, Helmut & Ntoa, Stavroula (Hg.): *Artificial Intelligence in HCI. 6th International Conference*. Switzerland: Springer Nature.
- Horváth, I. (2022). AI in interpreting: Ethical considerations. In: *Across Languages and Cultures* 23(1). Budapest: Akadémiai Kiadó Zrt, 1–13.
- Kalina, S. (2015). Measure for measure – Comparing speeches with their interpreted versions. In M. Behr & C. Zwischenberger (Eds.), *Quality in interpreting*. Frankfurt & Timme, 15–34.
- Peng, K., & Dai, H. (2023). A model of risk factors in remote interpreting. In Kubincová, Z., Hao, T., Capuano, N., Temperini, M., Ge, S., Mu, Y., Fantozzi, P., & Yang, J. (Eds.), *Emerging technologies for education: 8th International Symposium, SETE 2023, Sydney, Australia, November 26–28, 2023, revised selected papers, part II*. Springer, 3–15.
- Prandi, B. (2023). *Computer-assisted simultaneous interpreting: A cognitive-experimental study on terminology (Translation and Multilingual Natural Language Processing 22)*. Berlin: Language Science Press.
- Pym, A. (2025). *Risk Management in Translation*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781009546836>.

Zwischenberger, C., & Pöchhacker, F. (2010). Survey on quality and role: Conference interpreters' expectations and self-perceptions. <https://aiic.org/document/9646> (accessed on 18.02.2026).

Poster 2 · Interpretación con IA versus interpretación humana de discursos con dimensión de género en contextos institucionales. Un análisis comparativo de la calidad

Lucía Garbín Batanero & Elisa Alonso Jiménez

Universidad Pablo de Olavide, Seville, Spain

Antecedentes

La irrupción de sistemas de inteligencia artificial (IA) con capacidades speech-to-speech en tiempo real, en particular modelos de lenguaje de gran escala como GPT Realtime (OpenAI), ha abierto un debate urgente en los estudios de interpretación: ¿en qué medida pueden estos sistemas operar con garantías en entornos institucionales de alta responsabilidad comunicativa? A pesar del creciente interés académico por la calidad en la interpretación mediada por la tecnología (Fantinuoli, 2018; Corpas Pastor, 2021; Prandi, 2020), la literatura carece aún de marcos evaluativos consolidados que permitan comparar sistemáticamente outputs de interpretación humana e IA bajo parámetros equivalentes.

Este vacío metodológico se sitúa en la intersección de tres líneas de investigación escasamente exploradas de forma conjunta: los estudios de calidad en interpretación de conferencias, las tecnologías del lenguaje natural aplicadas a la interpretación simultánea y el análisis del discurso institucional con perspectiva de género. El corpus seleccionado, compuesto por discursos del Parlamento Europeo relativos a la Gender Equality Strategy 2025, introduce además el lenguaje de género. Su elevada carga ideológica y terminológica añade una dimensión adicional de complejidad, lo que constituye un terreno especialmente exigente para cualquier sistema de interpretación, humano o automatizado.

El estudio tiene como objetivo principal evaluar comparativamente la calidad de la interpretación simultánea humana y la producida por GPT Realtime sobre un corpus de discursos parlamentarios, con el fin de determinar en qué medida el sistema puede considerarse una alternativa funcional a la interpretación profesional en contextos institucionales de alta exigencia.

¿En qué medida es posible diseñar un marco evaluativo mixto que permita comparar sistemáticamente outputs de distinta naturaleza (interpretación simultánea humana e interpretación consecutiva automatizada) sin perder validez analítica?

¿Existen diferencias cuantitativas y cualitativas significativas entre los errores cometidos por intérpretes humanos y los producidos por GPT Realtime?

¿Qué tipos de errores son más frecuentes y graves en cada agente (interpretación humana e interpretación con IA)?

Ante la ausencia de un protocolo evaluativo consolidado para este tipo de comparación, se diseña un marco evaluativo mixto ad hoc estructurado en cinco dimensiones de calidad: (D1) precisión y fidelidad del contenido; (D2) calidad lingüística y gramatical; (D3) coherencia discursiva y pragmática; (D4) presentación y delivery; y (D5) estrategias de procesamiento y gestión de la información, esta última dimensión adaptada a los parámetros técnicos propios de los sistemas de IA, como la latencia de respuesta. Cada error identificado fue clasificado según una escala tripartita de gravedad (leve, grave, crítico), inspirada en los modelos de evaluación de traducción automática y adaptada a las especificidades de la interpretación de conferencias (Koponen, 2016; Lommel et al., 2014).

El corpus está compuesto por cuatro fragmentos de discursos parlamentarios en inglés, sus correspondientes interpretaciones simultáneas al español realizadas por intérpretes profesionales acreditados del Parlamento Europeo, y los outputs generados por GPT Realtime en modalidad consecutiva.

Se anotaron un total de 119 errores mediante un protocolo sistemático aplicado a cada segmento. El análisis revela que la diferencia cuantitativa entre agentes es reducida: 64 errores en la interpretación humana (53,78%) frente a 55 en la IA (46,22%). Sin embargo, el análisis cualitativo muestra divergencias sustanciales en el perfil de los errores. La interpretación humana concentra sus fallos en el nivel leve (52 de 64, el 81,25%), con únicamente dos errores críticos en todo el corpus. GPT Realtime, en cambio, incurre en siete errores críticos (12,73% de sus errores totales), de naturaleza semántica y pragmática.

Es decir, la IA comete alucinaciones como la sustitución del nombre del orador por un nombre inexistente; realiza sustituciones semánticas en terminología política sensible, como la traducción de Islamists por misóginos; y presenta cierta incapacidad para gestionar la dimensión retórica e ideológica del discurso parlamentario. Adicionalmente, el sistema presenta tres errores críticos de latencia, con tiempos de respuesta incompatibles con los estándares de comunicación institucional en tiempo real.

El estudio realiza tres contribuciones principales. En primer lugar, propone un marco evaluativo mixto ad hoc replicable que integra dimensiones lingüísticas, discursivas y técnicas, lo que cubre un vacío metodológico relevante en la literatura especializada. En segundo lugar, los resultados demuestran que la diferencia cuantitativa entre agentes no es el indicador más pertinente para evaluar la idoneidad de la IA en entornos institucionales. Es, por el contrario, el perfil cualitativo del error, y en particular la tasa de errores críticos, el que determina si un sistema puede operar con garantías en contextos de alta responsabilidad. En tercer lugar, el estudio evidencia que el lenguaje de género en contextos parlamentarios constituye un dominio de especial riesgo para los sistemas de IA actuales, con implicaciones directas para la política lingüística institucional, el uso responsable, seguro y transparente de la IA y los marcos éticos de su adopción en entornos multilingües.

Los hallazgos sugieren que GPT Realtime, en su versión beta actual, no constituye una alternativa funcional a la interpretación simultánea humana en estos contextos. No obstante, el marco evaluativo propuesto sienta las bases para investigaciones comparativas futuras a medida que los modelos evolucionen, y contribuye al debate sobre los criterios y métricas necesarios para una adopción responsable de la IA en interpretación de conferencias.

Referencias

- Corpas Pastor, G. (2021). Interpreting and technology: Is the sky really the limit? Translation and Interpreting Technology Online Conference, 15–24.
- Fantinuoli, C. (2018). Interpreting and technology: The upcoming technological turn. En C. Fantinuoli (Ed.), *Interpreting and technology* (pp. 1–12). Language Science Press.
- Gile, D. (2009). *Basic concepts and models for interpreter and translator training* (2nd ed.). John Benjamins.
- Koponen, M. (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation*, 25, 131–148.
- Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). Multidimensional Quality Metrics (MQM): A framework for declaring and describing translation quality metrics. *Tradumàtica*, 12, 455–463.
- Moser-Mercer, B. (1978). Simultaneous interpretation: A hypothetical model and its practical application. En D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication* (pp. 353–368). Plenum Press.
- Pöchhacker, F. (2016). *Introducing interpreting studies* (2nd ed.). Routledge.
- Prandi, B. (2020). An exploratory study on CAI tools in simultaneous interpreting: Theoretical framework and key concepts. *Interpreting*, 22(1), 139–160.
- Stryker, C., Liu, X., & Koehn, P. (2026). *Speech-to-speech translation benchmarks for low-resource language pairs*. Johns Hopkins University.

Poster 3 · Risk and Reliability in AI-Assisted Interpreting: Types of Errors in ASR and Speech to Speech Systems

Eyhab Bader Eddin

Dhofar University, Salalah, Oman

Background

AI assisted interpreting is moving at a fast pace from the lab to the real world in multilingual live events, remote meetings, public sector communication and platform based speech to speech mediation. Contemporary systems combine automatic speech recognition (ASR), speech translation and speech synthesis into an ever-increasing set of streamlined workflows, while multilingual models (SeamlessM4T) enable a significant expansion of the technical boundaries of speech-to-speech and speech-to-text translation across multiple languages (Barrault et al., 2023). Nevertheless, it is becoming widely acknowledged by the interpreting profession and the wider speech technology research community that raw system performance does not per se answer the more important question of reliability in high stakes situations. Imperceptibly small errors in spoken mediation may have disproportionately large effects if they're numbers, dates, names, institutional titles, dosage data or negation. The risk profile of AI assisted interpreting is not determined only by aggregate rates of error, but also by the type and consequence of errors. Empirical work has shown that named entities continue to be particularly challenging in ASR and speech translation and that this often leads to serious meaning distortion, making the usage of such systems in contexts close to professional interpreting difficult (Gaido et al., 2022). Similarly, research done on ASR combined with downstream entity recognition shows that word error rate is too coarse to be considered very useful for assessing the impact of transcription mistakes in practice, especially in spontaneous or noisy speech (Szymanski et al., 2023). More recent research on speech translation quality estimation suggests that speech translation is a separate problem that needs a separate evaluation framework: segmentation, speech specific noise and audio to text mismatch make both automatic and human evaluation difficult (Han et al., 2016; Sperber et al., 2014). In parallel, the Multidimensional Quality Metrics, or MQM, framework provides a good starting point for analytic error annotation, such as omissions, additions, mistranslations, entity errors and machine translation hallucinations - all of which are directly relevant for AI-mediated interpreting risk (MQM, n.d.).

This paper aims at developing and empirically validating a risk-oriented typology of errors in interpreting outputs produced by AI-aided interpreting with ASR-supported and speech-to-speech pipelines. Rather than focusing on inquiry about average system accuracy, the study asks questions about what types of errors have the most influence, in what circumstances they occur, and how they affect the reliability of AI-supported interpreting in real-world live environments.

Research questions:

Which types of errors are most common in AI assisted interpreting pipelines based on ASR and speech-to-speech translation?

Where is the greatest discrepancy between reliability (i.e., the proportion of times the model correctly identifies the subject being asked about) and effectiveness?

What are the related risk management implications for the use of AI assisted interpreting in a range of communicative situations that occur in low-, medium-, and high-stakes situations?"

The research design is of mixed method and is a blend of analytic error annotation and risk oriented interpretation. The proposed corpus consists of around six to eight hours of multilingual spoken material typical of the use cases of interpreting - adjacent use cases such as conference-like speeches, institutional briefings, question-and-answer. Recordings will be treated in two technology configurations: (1) a cascaded pipeline, where the ASR output goes to machine translation component and (2) an end-to-end or integrated speech translation/speech-to-speech system. Where

possible, outputs from a powerful multilingual model such as SeamlessM4T or from a related streaming variant will be incorporated as a benchmark condition (Barrault et al., 2023; Papi et al., 2024). The analysis is done in three stages. First, source speech, ASR transcripts and target outputs are matched segment by segment. Second, there is manual output annotation based on a specific typology of error which was inspired by MQM but tailored to spoken mediation, where core categories include omission, addition, mistranslation, hallucination-like generation, numeral error, named entity error, segmentation error, temporal misalignment and disfluency-induced distortion. This adaptation expands upon existing MQM and speech - translation evaluation work towards an interpreting - oriented risk-framework (MQM, n.d.; Sperber et al., 2024). Third, each error is not only given a linguistic label, but also assigned a risk - severity score: low (stylistic inconvenience), moderate (meaning degradation), high (decision relevant distortion) or critical (potential safety, reputational, legal or operational harm). To complement the manual analysis, a comparison of the manual annotation of the audio and text with standard automatic analysis such as WER for ASR and a selection of MT/speech translation metrics, where it applies, is performed in this study. This comparison is meant not to deny automatic metrics, but to test their sensitivity to errors that have a disproportionate impact on the interpretation of and in environments. Prior work has suggested that named entities are notably fragile in speech systems (Gaido et al., 2022; Zhou et al., 2024) and speech translation evaluation is underdeveloped in comparison to text translation (Han et al., 2024).

The expected result will be that overall system scores will mask a relatively small but highly consequential set of errors. In particular, numeral errors, named-entity substitutions, and omissions are likely to be responsible for a high percentage of high- and critical-risk failures, even where wider output quality seems acceptable. Existing research shows that named entities are likely to be one of the major weaknesses of both ASR and speech translation (Gaido et al., 2022; Luo et al., 2025; Zhou et al., 2024). Omissions are also expected to be particularly relevant in dense spoken discourse, where segmentation issues, latency limits and disfluencies can cause information loss that can be hard to detect using sentence level automatic metrics alone (Sperber et al., 2024). A second thing that we expect to see is that end-to-end speech translation and cascaded ASR+MT pipelines will have different signatures of error and not just different aggregate accuracies. Cascaded systems may spread ASR distortions in translation whereas integrated systems may produce more opaque or hard to trace mistranslations. A third expected result is that automatic metrics that are commonly used will correlate only partially with human judgements of reliability in interpreting-like conditions. Consequently, it is possible for a system to have reasonable overall scores and yet a few critical failures, which would make unsupervised use unacceptable in many professional situations.

The paper contributes to current debates on AI and interpreting in 3 main ways. First of all, it changes the evaluation question from “How good is the system?” to “How risky is the system, and why?” Second, it suggests a pragmatic error typology that links speech technology evaluation and interpreting studies by putting consequence-sensitive categories such as hallucinations, omissions, numerals, and named entities into the foreground. Third, it provides an evidence-based framework for understanding the suitability of AI-assisted interpreting, in which the need for interpreter-in-the-loop oversight is identified, and when unsupervised deployment is still too risky. The larger implication of this is that there is no way to assume from average performance metrics that AI-assisted interpreting is reliable. For policymakers, event organizers and language service providers the study helps make more complex deployment choices based on communicative stakes and consequences of errors. For researchers, it makes the case for evaluation frameworks that include a combination of analytic error annotation and speech specific quality estimation and risk severity analysis. For those of us who work with interpretation, I think it offers a vocabulary for talking about more than whether AI “works,” but whether it can be trusted when the cost of getting it wrong is high.

References

Barrault, L., Chung, Y.-A., Meglioli, M. C., Dale, D., Dong, N., Duquenne, P.-A., Elsahar, H., Gong, H., Heffernan, K., Hoffman, J., Klaiber, C., Li, P., Licht, D., Maillard, J., Rakotoarison, A., Sadagopan, K. R., Wenzek, G., Ye, E., Akula,

B., ... Seamless Communication. (2023). SeamlessM4T: Massively multilingual & multimodal machine translation (arXiv:2308.11596). arXiv.

Gaido, M., Negri, M., & Turchi, M. (2022). Who are we talking about? Handling person names in speech translation. In Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022) (pp. 62–73). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.iwslt-1.6>

Han, H., Duh, K., & Carpuat, M. (2024). SpeechQE: Estimating the quality of direct speech translation. In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (pp. 21852–21867). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.1218>

Luo, Y., Wei, D., Li, S., Shang, H., Guo, J., Li, Z., Wu, Z., Chen, X., Rao, Z., Yang, J., & Yang, H. (2025). Generative annotation for ASR named entity correction. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (pp. 20824–20835). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.emnlp-main.1052>

MQM. (n.d.). The MQM error typology. <https://themqm.org/error-types-2/typology/>

Papi, S., Gaido, M., Negri, M., & Bentivogli, L. (2024). SimulSeamless: FBK at IWSLT 2024 simultaneous speech translation (arXiv:2406.14177). arXiv.

Sperber, M., Bojar, O., Haddow, B., Javorský, D., Ma, X., Negri, M., Niehues, J., Polák, P., Salesky, E., Sudoh, K., & Turchi, M. (2024). Evaluating the IWSLT2023 speech translation tasks: Human annotations, automatic metrics, and segmentation. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (pp. 6559–6571). ELRA and ICCL.

Szymański, P., Augustyniak, Ł., Morzy, M., Szymczak, A., Surdyk, K., & Żelasko, P. (2023). Why aren't we NER yet? Artifacts of ASR errors in named entity recognition in spontaneous speech transcripts. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1760–1774). Association for Computational Linguistics.

Zhou, S., Li, Z., Hong, Y., Zhang, M., Wang, Z., & Huai, B. (2024). CopyNE: Better contextual ASR by copying named entities. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics.

Poster 4 · Challenges in Contemporary Interpreting Practice: A Comparison of Onsite and Remote Modes in the Ukrainian Context

Yuriy Velykoroda & Oleksandra Litvinyak

Vasyl Stefanyk Precarpathian National University, Ivano-Frankivsk, Ukraine; Ivan Franko National University of Lviv, Ukraine

Remote and distance interpreting have been discussed in research prior to the COVID-19 pandemic, but the global shift to online communication significantly intensified scholarly attention to this modality. Bachir Mahyub Rayaa and Anne Martin (2022) define remote interpreting (RI/RSI) as interpreting performed when the interpreter is physically separated from speakers and/or listeners, regardless of participants' locations.

Professional bodies such as AIIC (2025) emphasize that distance interpreting introduces additional cognitive strain due to reduced sensory input, degraded sound quality, prolonged screen exposure, and the need to manage technology. Andrew Constable (2015) highlights the importance of multisensory input for meaning construction; when visual and contextual cues are limited, interpreters must make additional effort, increasing fatigue and the risk of error. Health-related concerns, including auditory strain and eye fatigue, have also been documented. Survey-based research confirms substantial changes in professional practice during the pandemic. Danuta Przepiórkowska (2021) reports that most interpreters invested in new equipment and widely adopted platforms such as Zoom and Microsoft Teams. Studies by Agnieszka Chmiel and Nicoletta Spinolo (2022) show that many interpreters work from home and often collaborate with booth partners located remotely.

Beyond technical factors, researchers note relational changes: reduced interaction with speakers and colleagues, more complex coordination, and feelings of isolation (Rayaa & Martin, 2022). Clare Donovan and Cecilia Manzoni (2024) argue that remote modalities may amplify stress, fatigue, and cognitive overload inherent in interpreting. The aim of this study is to conduct an in-depth overview of the challenges associated with onsite and online interpreting in Ukraine, determine the severity of each of the challenges, and identify the key similarities and differences between the challenges in these two modalities. The study seeks to explore interpreters' perceptions of modality-specific difficulties and discuss whether online interpreting introduces fundamentally new challenges or amplifies existing ones present in onsite settings.

The research questions that underpin this study are: What challenges do interpreters working with Ukrainian report in onsite interpreting settings? What challenges do interpreters working with Ukrainian report in online interpreting settings? How do the perceived challenges differ between onsite and online modalities? Which challenges are modality-specific, and which are common to both settings?

This study is based on survey data collected from professional interpreters. The data were gathered through a structured questionnaire consisting of multiple-choice questions and optional open-ended responses. Questions on onsite and online interpreting and their challenges were part of a larger survey of the situation on the Ukrainian interpreting market. This part of the survey aimed to capture the interpreters' experiences with both onsite and online interpreting, focusing on perceived challenges, working conditions, technological factors, cognitive demands, and interactional dynamics.

Participants were recruited through professional association, social media groups, and personal networks of the researchers.

The questionnaire included two parallel sections designed to examine challenges associated with onsite and online interpreting. In both sections respondents were asked to assess how frequently they encounter specific difficulties in each modality.

Based on the responses we determined the most problematic issues in each of the modes. Thus, for onsite interpreting the respondents most commonly complained of organizational and logistical issues relating to scheduling and prior preparation for the event, while technical and physical conditions caused least concern.

For the online interpreting mode there was some overlap with the onsite mode specifically in the sphere of adequate preparation materials. Yet, the physical and audio conditions of the speakers and the quality of the source text verbal presentation were reported as the most problematic issues.

From the results of the survey we can draw some implications that might need further verification via personal interviews or focus group discussions with the respondents. Thus, insufficient or inadequate technical support was hardly ever identified as a challenge in the onsite mode, but it looks like a major challenge in interpreting online. Although deeper research is needed to draw definitive conclusions, we can assume that while the technical support role in onsite interpreting is provided by a group of specialized technicians, this function is usually left unfulfilled in the online mode. The physical separation of the speakers and the interpreters (which is the foundation for the definition of distance/remote interpreting) often means that the task of providing technical support falls respectively on the speakers and the interpreters themselves, as technical support teams are often unavailable, not planned or less efficient in managing challenges in online events. This underlines the importance of the work of technical support teams present during onsite events, as they facilitate the flow of the event smoothly on the spot. This challenge is clearly one that is modality specific.

Another implication relates to the overlap between both modes in the challenge of insufficient materials for preparation. While this is a dominant problem in both modes, it is defined as significantly more problematic in the onsite mode of interpreting.

The online-specific challenges, such as remote coordination with the online boothmate or physical and emotional exhaustion from working in isolation do not seem to be a major problem in the online mode of interpreting.

Overall, the results of the survey demonstrate that online interpreting presents an additional strain on the interpreters primarily in terms of the quality of the source speech that they receive, while in onsite interpreting the main challenges relate to poor scheduling and insufficient preparation options. These findings serve as a first step toward identifying structurally distinct challenges across interpreting modes and their implications for professional practice

References

- AIIC. (2025). AIIC guidelines on distance interpreting (v1, May 2025). International Association of Conference Interpreters.
- Chmiel, A., & Spinolo, N. (2022). Testing the impact of remote interpreting settings on interpreter experience and performance: Methodological challenges inside the virtual booth. *Translation, Cognition & Behavior*, 5(2), 250–274.
- Constable, A. (2015). Distance interpreting: A Nuremberg moment for our time. AIIC Assembly. https://aiic.org/document/9692/Constable_Distance%20Interpreting_A%20Nuremberg
- Donovan, C., & Manzoni, C. (2024). Remote interpreting: A game-changer for interpreters? *The Interpreters' Newsletter*, 29(Special Issue), 1–29.
- Przepiórkowska, D. (2021). Adapt or perish: How forced transition to remote simultaneous interpreting during the COVID-19 pandemic affected interpreters' professional practices. *Między Oryginałem a Przekładem*, 27(4), 137–159. <https://doi.org/10.12797/MOaP.27.2021.54.08>
- Rayaa, B. M., & Martin, A. (2022). Remote simultaneous interpreting: Perceptions, practices and developments. *The Interpreters' Newsletter*, 27, 21–42. <https://doi.org/10.13137/2421-714X/34390>

Poster 5 · 40 years of interpreting quality: A systematic literature review

Francesco Pelizza & Sara Pasini

University of Vienna, Vienna, Austria

The technological turn has reshaped interpreting practices through both interpreter-oriented technologies, such as CAI, RSI and VMI, and the rapid development of automated speech translation systems. While both interpreters and automated speech translation enable real-time communication, they differ fundamentally in their underlying processes. The term “live translation” is therefore used as an umbrella concept for temporally immediate translation performance.

Evaluating the product of live translation remains a central challenge in interpreting studies. Despite numerous frameworks proposed since the end of the 20th century, no consensus has been reached, partly because of differing needs and expectations of stakeholders.

In order to work towards a comprehensive framework, this contribution presents a systematic literature review of interpreting quality assessment with the aim of identifying and comparing existing approaches and their underlying assumptions. The results of this review provide a basis for discussing current limitations and outlining requirements for an evaluation framework for live translation.

The primary aim of this study is to map the development of interpreting assessment over the past four decades (1986–2026). To this end, a systematic literature review (SLR) is conducted to identify the criteria, stakeholders and methodological approaches currently employed in the evaluation of interpreting quality.

Based on this review, the study seeks to synthesize existing approaches in order to lay the groundwork for designing systematic approaches to the evaluation of live translation, encompassing both human-produced and automated speech-to-speech outputs. The focus is placed on evaluation contexts beyond educational assessment and system-oriented performance evaluation, with particular attention to real-world communicative settings and user reception.

This study is guided by the following research questions:

RQ1. How are the foundational concepts of “evaluation,” “assessment,” and “quality” defined and operationalized within interpreting studies?

RQ2. What is the current state of research regarding the criteria and methods to assess live translation?

This study adheres to a rigid structured protocol (Wetterich & Plänitz 2021) for all screening decisions. Eligibility was restricted to records published between the seminal work of Bühler on quality (1986) and the time of research (February 2026), and accessible via online repositories or library systems. Inclusion criteria were formulated as follows: a) the research topic is interpreting or speech-to-speech translation, b) publications must be peer-reviewed, c) acceptable formats are original primary research papers from journals and collective volumes, d) only records in English are considered, e) both empirical studies and theoretical contributions are eligible.

Searches were conducted across the following electronic databases: Web of Science (Core Collection), Scopus, Bibliography of Interpreting and Translation (BITRA), and Translation Studies Bibliography (TSB). Search strings incorporated Boolean operators and terms for interpreting (e.g., interpretation, speech translation, machine interpreting, simultaneous translation, consecutive translation) with quality (quality, assessment, evaluation). To ensure comprehensive coverage, additional searches were conducted on AI-powered platforms including Consensus, SciSpace, Semantic Scholar, and Google Scholar Labs using natural language prompts.

Using the software Catchii, a two-stage screening process (title/abstract followed by full-text) was implemented, checking for inclusion criteria and removing duplicates. Two raters assessed the

records independently. Of the 1,646 initially screened records, 354 underwent full-text assessment with high inter-reviewer reliability (90.11% agreement; $k = 0.80$; 95% CI [0.73, 0.86]; SE = 0.03), ultimately yielding 183 sources for final data extraction.

Data were extracted for the following variables: the nature of the research (empirical or conceptual), interpreting mode and setting, language combination, number and type of assessor and assessed populations, underlying models, assessment methods, assessment criteria, and measurement scales.

A preliminary analysis of 183 records (73% journal articles, 27% book chapters) reveals a shift in research dissemination, from foundational contributions (1980s–90s) and growth phase based on book publication (2000s–10s) toward recent scholarship published in journals.

Methodologically, research is predominantly of empirical nature (90.7%), and evenly distributed between consecutive (41.4%) and simultaneous (41.9%) interpreting modes. While conference settings constitute the largest specified context (29.5%), a majority of studies (56.3%) do not report a specific setting, indicating research conducted in controlled experimental or educational environments. Participant profiles reveal that assessees are predominantly students (50%) and professionals (38%), with assessment conducted by a mix of evaluators and researchers (22.8%), practitioners (16.4%), and students (16%).

Regarding assessment methods, the study identifies rating scales and questionnaires as the most prevalent instruments. These are followed by rubrics and error analysis, as well as automated metrics and qualitative, criteria-based comments. Furthermore, the review highlights an increasing reliance on AI evaluation.

Despite heterogeneity in assessment criteria and their terminology, a recurring pattern can be identified across studies. Three core dimensions of live translation assessment emerge: content, form, and presentation. At the same time, the coexistence of human-centred evaluation approaches and automated metrics poses an ongoing challenge for the development of a unified assessment framework.

This study contributes to a comprehensive state-of-the-art map of research on interpreting assessment, offering a more rigorous, multi-dimensional overview of quality assessment that can serve as the basis for a future unified framework. The results offer both practitioners and institutions a clear set of criteria to ensure both human and automatic live translation meet established professional standards.

References

- Bühler, H. (1986). Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua*, 5 (4), 231–235.
- Wetterich, C., & Plänitz, E. (2021). *Systematische Literaturanalysen in den Sozialwissenschaften*. Leverkusen: Verlag Barbara Budrich.

Poster 6 · A scoping review of computer-assisted interpreter training (1995–2026): research foci, gaps, and a pedagogy-driven multi-component interaction framework

Nan Zhao & Liping Wei

Hong Kong Baptist University, Hong Kong

Interpreter education has expanded substantially since the 1960s as professionalisation and university programmes have grown globally. Yet persistent pedagogical challenges remain: securing authentic and graded resources; managing high stress learning environments; addressing diverse student needs; and fostering continuous self-regulated learning. To respond to these challenges, successive technologies from online repositories to the most recent GenAI have been explored to enhance the teaching and learning of interpreting, consolidating computer-assisted interpreter training as a robust area of scholarly inquiry. Following this development, review studies have emerged to take stock of CAIT developments and research outputs. Early research has established a solid foundation of the conceptual development and typology of CAIT (e.g., Sandrelli & de Manuel Jerez, 2007; Sandrelli, 2015; Frittella, 2021), while the recent studies (Hu, 2024; Guo, 2025) show a trend of increased methodological rigour and a focus on specific technology. However, none of them is updated to include GenAI technologies, and there is still a lack of data synthesis reviews that comprehensively map empirical evidence of CAIT to afford a data-informed understanding of the field, including how technologies are integrated into training, what evidence has been collected, what pedagogical impact has been achieved, and where the research gaps are. Responding to this gap, the present study conducts a scoping review systematically mapping empirical CAIT studies published from 2000 to 2026 to explore four questions related to publication trend, technology integration into training, methodological choices, and research foci.

To answer these questions, the study followed rigorous scoping review procedures (Arksey & O'Malley, 2005) and PRISMA-ScR guidance (Tricco et al., 2018), searching Scopus and Web of Science for English-language empirical studies addressing pedagogical uses of technology in interpreter training. After deduplication and screening against predefined inclusion/exclusion criteria, 75 empirical articles were included in the review database. For each study, bibliographic features, technologies employed, and data on pedagogical integration (the role/affordance of technology in training, interpreting modes/settings, targeted skills, in/out of class use, its use in interpreting preparation, skill training, and assessment), research foci (variables/constructs from research questions or research aims, data collected), and methodological choices (methodological approach, data collection method, participant, sample size, intervention duration) are charted and analysed via descriptive analysis, thematic analysis (Braun & Clarke, 2006) and network analysis to reveal meaningful findings.

Major findings via thematical analysis reveal five major research dimensions: 1) technological feasibility and access, concerning assessing technical performance, feasibility, limitations, and access; 2) user experience and perception, focusing on trainees subjective experience and perception of the training with educational technologies; 3) learning outcomes, concerning assessing interpreting performance, skill acquisition, and metacognition; 4) behaviour, examining frequency and quality of interaction and communication with peers and trainers in a CAIT environment, along with the use and engagement with the technology; 5) pedagogy, related to teaching adaptation/design, instructional challenges, teaching experience, trainers' role, workload, and support. Further, network analysis of these research dimensions shows that user perception is the largest and most connected dimension in the network, while pedagogy and behaviour are peripheral. Moreover, except for the more frequent co-occurrence of perception and outcomes, there are weak links between pedagogy and other research dimensions. Even when pedagogy is mentioned in studies, it is a peripheral finding, with fewer reports on how pedagogy relates to other research dimensions they studied. This systemic marginalisation might suggest that existing research is

conceived predominantly from a techno-centric perspective, treating CAIT as an isolated intervention to be tested, rather than a dynamic multi-component ecosystem where the synergy among technologies, stakeholders, and pedagogical orchestration is essential for success.

To move research away from a techno-centric driven perspective, the study proposes a pedagogy-driven, multi-component interaction framework which includes macro-institutional and social support, technological affordances and constraints, pedagogical orchestration/trainers, and learner component. In this framework, pedagogy is discussed with evidence as the central pivot that actively adapts to technological constraints/affordances while systematically shaping learner behaviours, experiences, and learning outcomes through instructional design.

This framework has practical implications for major stakeholders in practice and guides future research. Practically, institutions should invest in trainer professional development, IT infrastructure, and clear guidelines for technology, especially GenAI use. Trainers should improve technology competence and enhance TPACK-informed pedagogy design (Mishra & Koehler, 2015) and shift their role from a knowledge transmitter to an architect who skillfully integrates technology in pedagogy, provides interaction & feedback, and scaffolds learners' self-regulated practices. Trainees should embrace technologies in their learning environment and improve technology competence to have meaningful interaction with technology, peers, and trainers in a CAIT environment and enhance meta-cognition to become self-regulated learners. As for future research, besides exploring basic trainer perceptions and experience, there is a critical need to focus on trainers and pedagogy and investigate the underlying mechanism of how pedagogy interacts with other components in the framework via more robust process-related long-term studies. The study contribution is twofold: practically, it provides trainers with an evidence-based foundation for optimising instructional design in computer-assisted interpreter training. Conceptually, it proposes a pedagogy-driven multi-component interaction framework to guide practical CAIT implementation and future research.

References

- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19-32. <https://doi.org/10.1080/1364557032000119616>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Frittella, F. M. (2021). Computer-assisted conference interpreter training: Limitations and future directions. *Journal of Translation Studies*, 1(2), 103-142. <https://doi.org/10.3726/JTS022021.6>
- Hu, P., Gao, B., & Li, K. (2024). Learning interpreting in virtual reality: A scoping review. *Interactive Learning Environments*, 33(1), 347-366. <https://doi.org/10.1080/10494820.2024.2347304>
- Guo, W., & Cui, F. (2025). A systematic review of technology integration in interpreter education: Current applications, challenges, opportunities and regional variations. In A. K. F. Cheung, D. Li, K. Liu, & R. Moratto (Eds.), *Technology and interpreting: Navigating the digital age* (pp. 57-84). Routledge.
- Mishra, P., & Koehler, M. (2015). TPACK (Technological pedagogical content knowledge). In J. M. Spector (Ed.), *The SAGE encyclopedia of educational technology* (Vol. 2, pp. 782-785). SAGE Publications, Inc. <https://doi.org/10.4135/9781483346397.n318>
- Sandrelli, A. (2015). Becoming an interpreter: The role of computer technology. *MonTI*, 111-138. <https://doi.org/10.6035/MonTI.2015.ne2.4>
- Sandrelli, A., & De Manuel Jerez, J. (2007). The impact of information and communication technology on interpreter training: State-of-the-art and prospects. *Interpreter and Translator Trainer*, 1(2), 269-303.
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garrity, C., ... Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, 169(7), 467. <https://doi.org/10.7326/M18-0850>

Poster 7 · Evaluation of Live Translation: A 3D Approach

Franz Pöchhacker, Claudio Fantinuoli, Chao Han, Tomasz Korybski & Min-hua Liu

University of Vienna, Vienna, Austria; University of Mainz, Mainz, Germany; National University of Singapore, Singapore; University of Warsaw, Warsaw, Poland; National Taiwan University, Taipei

This paper presents the results of a collaborative effort by researchers from different areas of expertise but with a shared interest in what is loosely called “quality” in “interpreting” – two basic notions in need of further conceptual development. The assessment of interpreters’ performance, for training, research and quality assurance purposes, has long been a major concern in the field of interpreting studies, given its tradition of close links to the interpreting profession and its institutions. Nevertheless, the concept of quality remains poorly defined, and there is little consensus on the criteria and methods to be adopted in evaluating interpreting as a service to enable communication in real time. This lack of shared standards has made itself felt most acutely through the rapid emergence of AI-powered automatic speech translation (AST) systems capable of real-time performance in a growing range of contexts (Fantinuoli, 2025). There is an urgent need to assess machine performance, for example, to decide if a system might be fit for adoption in a particular setting, but also to clarify more generally how such performance can be evaluated in ways that are meaningful for researchers, users, developers, institutions and policy makers. The widely discussed prediction that AST systems may sooner or later perform as well as professional interpreters is difficult to test in the absence of established standards for evaluating the human as well as the automatic performance of what we refer to as “live translation” – translation produced in real time for immediate use within a specific communicative situation, covering both interpreting (when produced by a human) and its automatic counterpart.

Drawing on state-of-the-art approaches in interpreting studies (e.g. Davitti et al., 2025; Han, 2022; Winston et al., 2023) as well as on methodologies developed in other fields, notably computer science (e.g. Cettolo et al., 2026; Mieno et al., 2015), we therefore seek to address this need by proposing a set of general principles for a comprehensive approach to the evaluation of live translation across modes, linguistic modalities and settings, including spoken and written output and both human and automated performance. Rather than proposing a ready-made assessment method for a given type of live translation, the paper sets out axiomatic assumptions that may guide the future design of evaluation frameworks and campaigns. For example, it argues that professional interpreting needs not be viewed as the unquestioned “gold standard” from which other forms of live translation will deviate. Rather, evaluation should be fair and broad enough to account also for characteristics and affordances that may not be found in human performance. The paper further argues that evaluation cannot be separated from its purpose. Whether the aim is to compare AST systems to interpreting or to other systems, improve performance, support certification, or inform procurement and deployment decisions will shape both what is evaluated and how. The evaluation of live translation must therefore be explicitly sensitive to its function and scope while also remaining theoretically grounded and practically operational.

An important premise underlying our approach to evaluation is that live translation must be viewed as a service to enable communication in a given institutional and situational context. Therefore, evaluation must not stop at an examination of the translational product (i.e., output, performance) but must also cover user reception as the ultimate purpose for which the service is offered. Moreover, we posit that the combined focus on production and reception must be widened to also account for the dynamics of interaction between the social agents and technical affordances in a given communicative event.

On this basis, the paper introduces a three-dimensional (3D) approach to the development of a comprehensive evaluation framework “TRI approach” to the evaluation of live translation which comprises three interrelated dimensions: Translational performance, user Reception, and ecology of Interaction. The T dimension concerns the output of live translation, which is in turn made up of

three components: content, target-language form (correctness, appropriateness), and presentation (fluency of delivery). Content refers to the substantive relationship between output and input, particularly in terms of criteria such as fidelity, accuracy, and completeness. Its evaluation may draw on existing assessment methods, such as error analysis (e.g., Barik, 1971), rubric scoring (e.g., Lee, 2008), and comparative judgment (e.g., Wu, 2010), as well as automated evaluation metrics. The R dimension concerns how the translational output is received, processed, and evaluated by its intended or actual users. It may include such aspects as comprehensibility, cognitive processing demands, perceived quality, and user satisfaction. The I dimension concerns the broader communicative and technical ecology in which live translation is embedded and includes such features as turn-taking, latency, usability, interface design, and related elements of the sociotechnical system.

By situating the evaluation of live translation within the broader study of communicative services and socio-technical systems, the paper highlights both the possibilities and the limits of evaluation. It advocates a pragmatic approach that reduces complexity where necessary, selects features according to context and purpose, and still seeks results that retain broader validity. In doing so, it offers a conceptual foundation for the development of evaluation practices capable of responding to the current and future realities of live translation in both its human and automatic forms.

References

- Barik, H. C. (1971). A description of various types of omissions, additions and errors of translation encountered in simultaneous interpretation. *Meta*, 16(4), 199–210. <https://doi.org/10.7202/001972ar>
- Cettolo, M., Gaido, M., Negri, M., Papi, S., & Bentivogli, L. (2026). How to evaluate speech translation with source-aware neural MT metrics. <https://doi.org/10.48550/arXiv.2511.03295>
- Davitti, E., Korybski, T., Orăsan, C., & Braun, S. (2025). Quality-related aspects, in E. Davitti, T. Korybski, & S. Braun (eds), *The Routledge handbook of interpreting, technology and AI*. Routledge, 305–326.
- Fantinuoli, C. (2025). Machine interpreting, in E. Davitti, T. Korybski, & S. Braun (eds), *The Routledge handbook of interpreting, technology and AI*. Routledge, 209–226.
- Han, C., 2022. Interpreting testing and assessment: A state-of-the-art review. *Language Testing*, 39, 30–55. <https://doi.org/10.1177/02655322211036100>
- Lee, J. (2008). Rating scales for interpreting performance assessment. *The Interpreter and Translator Trainer*, 2(2), 165–184. <https://doi.org/10.1080/1750399X.2008.10798772>
- Mieno, T., Neubig, G., Sakti, S., Toda, T., & Nakamura, S. (2015). Speed or accuracy? A study in evaluation of simultaneous speech translation, in: *Interspeech 2015, ISCA*, pp. 2267–2271. <https://doi.org/10.21437/Interspeech.2015-498>.
- Winston, E. A., Lee, R. G., Monikowski, C., Peterson, R., & Swabey, L. (2023). *Beyond equivalence: Reconceptualizing interpreting performance assessment*. Gallaudet University Press.
- Wu, S. C. (2010). *Assessing simultaneous interpreting: A study on test reliability and examiners' assessment behavior*. Doctoral dissertation, Newcastle University. <https://theses.ncl.ac.uk/jspui/bitstream/10443/1122/1/Wu%2011.pdf>

Poster 8 · Beyond “Just” Interpreters: Fostering Interdisciplinary Interpreter Training for Broader Competency Frameworks in Future Professionals

María Victoria Troyano Fernández

Universidad Pontificia Comillas, Madrid, Spain

Artificial Intelligence (AI) is rapidly transforming how professions are conceptualised, reshaping required skill sets, professional profiles, and the expectations of clients and organisations across sectors, thereby prompting the development of learning and educational models responding to these challenges. Within the field of Translation and Interpreting (T&I), the development of technology and the irruption of AI-driven systems —particularly in speech processing, machine translation, and speech-to-speech models—has sparked an urgent debate on the future of the profession. In this context of change and uncertainty, T&I professionals face a working environment that requires constant adaptation, where traditional competencies must be complemented by advanced technological proficiency. In response to the need for upskilling and reskilling, educational institutions, particularly universities that have traditionally offered training in Translation and Interpreting, mediation, languages, and multilingual communication, face increasing pressure to provide forward-looking, adaptable training that aligns with evolving market demands.

This study builds on the premise that Conference Interpreting training programmes must undergo a meaningful reorientation in order to remain relevant and aligned with evolving professional demands. Beyond incorporating technological tools into the classroom, it may be necessary to revisit certain foundational aspects of interpreter education: what competencies are essential, how they should be taught, and how they relate to a rapidly changing professional ecosystem. In particular, this analysis focuses on the necessity to redefine competency frameworks to include not only traditional interpreting skills and techniques but also a broader form of technical literacy, critical thinking and interdisciplinary knowledge.

The main research question guiding this study is: How should conference interpreter training programmes be reoriented to meet the demands of the current and future professional market? To address this overarching question, the study also explores several sub-questions: How do students perceive their preparedness for a technology-driven market that is not only increasingly unstable, but also characterised by the growing presence of technological solutions performing tasks traditionally carried out by humans? How do active professional interpreters, who also serve as trainers, conceptualise the skills required in this new landscape? What pedagogical gaps can be identified between current training practices and emerging professional needs?

To answer these questions, the study adopts a qualitative methodology based on focus group discussions, given the inherently subjective nature of the topic and the potential of peer interaction to stimulate creativity and generate more nuanced and higher-quality data. Three main data sources are considered. First, focus groups will be conducted with students enrolled in the 2025–2026 cohort of a Master's programme in Conference Interpreting. These discussions aim to capture students' perceptions of their training, their awareness of technological developments and their expectations regarding employability. Second, another focus group will be organised with trainers, many of whom are also active professional interpreters. This dual perspective allows for an in-depth exploration of how practitioners interpret current market transformations and how these perceptions influence their teaching practices. Third, insights will be drawn from internal discussions held within a Master's programme ad hoc committee, where curriculum design and adaptation to AI-related challenges are actively debated.

The focus group discussions will address, among other issues, how interpreter training can adapt to the integration of AI technologies, what types of modules or courses could be introduced to better prepare students, and how these additions should balance technical skills with broader cognitive and analytical competencies. Particular attention will be paid to identifying areas where current

curricula may fall short, especially in terms of fostering technical literacy and interdisciplinary awareness.

The expected results point towards a clear need to update existing Master's programme in conference interpreting. However, this redesign should not be limited to the integration of AI tools or the teaching of how to interact with such technologies. Rather, the findings are likely to highlight the importance of a more comprehensive transformation. This includes strengthening student's analytical abilities, enhancing critical thinking skills, and promoting domain-specific expertise. In other words, the "future interpreter" will extend beyond the traditional role of linguistic and cultural mediator to become a highly adaptable and interdisciplinary professional, combining strong multilingual competences with the ability to engage critically and effectively with technology and AI in dynamic complex professional settings.

Furthermore, the study anticipates identifying a growing demand for more flexible and modular forms of education, such as micro-credentials. These shorter, targeted learning units could allow both students and professionals to continuously update their skills in response to rapid technological change and market trends. Such an approach would complement traditional Master programmes and contribute to a more dynamic and lifelong learning-oriented model of interpreter education.

From a broader perspective, this paper contributes to ongoing discussions on the role of universities in shaping the future of professions affected by AI. It argues that higher education institutions should actively lead the process of curricular innovation. In the case of conference interpreting, this means taking responsibility for defining new competency frameworks that integrate technological literacy with the core values and skills of the profession. The practical implications of this study include generating concrete ideas for curriculum development, including the potential introduction of new modules or seminars focused on AI literacy and interdisciplinary knowledge.

Ultimately, the paper advocates for a reflective approach to interpreter training, one that recognises the transformative impact of AI while reaffirming the importance of human judgement and expertise. By combining insights from students and professionals with a critical analysis of current pedagogical practices, this study seeks to contribute to the design of more resilient, future-oriented training programmes capable of preparing interpreters for a rapidly evolving professional landscape.

Poster 9 · Interpreting in Transition: How Training Institutions Adapt Curricula to Technological Change

Tatiana Kaplun & Camille Collard

University of Strasbourg, France; ESIT, Université Sorbonne Nouvelle, Paris, France

Since the establishment of specialised conference interpreting (CI) training programmes, a shared consensus has emerged regarding interpreter training in postgraduate university settings, that is grounded in the assumption that students must develop a complex set of cognitive skills (Kalina & Barranco-Droege, 2021). While some degree of mastery is expected at entry, the systematic development of these competences lies at the core of interpreter training, allowing students to acquire the techniques, strategies and coping mechanisms required for professional practices. Despite some differences between individual programmes, and the evolution of conference interpreting practices over the last decades, this pedagogical framework seems to have remained largely intact and unchallenged, putting primary emphasis on extensive general culture and cognitive processes such as analysis, anticipation, concentration, memory, decision making and problem solving under pressure (Bailey-Ravet, 2015; Setton & Dawrant, 2016).

Beyond these essential cognitive competences, interpreter training curricula also encompass more specialised professional skills. Some, such as ethics, meeting preparation strategies, and note-taking, are widely considered essential, while others, including marketing and negotiation, stress and voice management, etc., have traditionally been viewed as ancillary.

Until recently, technological competences were largely assigned to the latter category (Seeber, 2021). As technological solutions become increasingly present in the profession, training institutions have progressively revised their curricula to integrate these solutions more systematically. Nevertheless, such integration remains relatively limited, focusing primarily on preparation techniques and distance interpreting (DI), and specifically remote simultaneous interpreting (RSI), while more advanced applications are often relegated to continuing professional development (CPD).

This is reflected in the updated Core Curriculum of the European Masters in Conference Interpreting (EMCI) network, as the updated framework requires members to provide training in remote interpreting using RSI platforms and to introduce Computer-Assisted Interpreting (CAI) tools for preparation. By contrast, a more cautious approach is applied with regards to a broader integration of CAI tools and AI-based solutions, with an emphasis on the importance of developing critical awareness. This mindset mirrors current market realities, where technology is more often used for RSI and to enhance meeting preparation capabilities, especially in contexts with heavy documentation volumes (Defrancq, 2023; Pöchhacker & Liu, 2024). More advanced CAI tools, such as artificial boothmates and sophisticated terminology management systems, as well as AI-driven applications, remain far from widespread adoption.

The further integration of technology into interpreter training raises several pedagogical and strategic questions, to ensure that new technical skills do not undermine the acquisition of underlying competences. These concerns are not unprecedented, having accompanied earlier technological shifts in the profession, including the shift to the use of simultaneous itself, yet they remain especially relevant today (Orlando, Liao & Kruger, 2024).

Moreover, training institutions face a complex strategic decision regarding whether to anticipate and integrate technologies that have not yet been fully adopted by the profession in order to train future-proof interpreters. Similarly, the inclusion of RSI in curricula raises concerns regarding the potential cognitive and physiological effects of poor sound quality on students (Braun, 2024).

At the same time, the growing presence of generative AI in educational and professional contexts has sparked a broader debate around confidentiality, ethics, sustainability, and pedagogy (Šveda & Poláček, 2025). While collaboration with generative AI holds many promises and has been shown

to enhance productivity across a wide range of cognitive tasks, concerns persist regarding over-reliance and the possible erosion of cognitive effort and skill development (Chen et al., 2025). Which leads to the question of whether technological literacy should be considered a pre-requisite and is forcing many curricula designers to debate the extent to which it should be taught, and how it should be assessed.

The many questions and challenges that course directors and interpreter trainers currently have to face make the findings of this study, to be shared for the first time during the Conference, a timely and much needed contribution, that the authors hope will offer some reassurance and guidance to training programmes currently at a crossroads.

This project aims to provide an updated overview of conference interpreting training in an increasingly technology-driven professional environment, with a view to outlining best practices and informing future curricular decisions.

Research questions: How are training institutions adapting their curricula to the technological demands of the current era? What constraints shape course directors' curricular decisions, and what strategies have emerged in response? What adjustments and innovations are expected in conference interpreting curricula in the coming years?

The study adopts a mixed methodological approach combining the use of a targeted questionnaire designed to collect both qualitative and quantitative data, and a follow-up semi-directed interview to boost the qualitative data with nuanced answers to these complex questions, that will allow us to explore potentially overlooked factors and issues. The questionnaires exploring institutional strategies, pedagogical philosophies, and future perspectives, alongside detailed questions about present and projected curricula, will be distributed to course directors of conference interpreting programmes worldwide.

Participating programmes have been selected on the basis of their affiliation with recognised professional and academic networks (AIIC, CIUTI, EMCI).

Preliminary evidence (including the EMCI revised curriculum framework) suggests that conference interpreting programmes currently tend to adopt a pragmatic yet cautious approach to integrating new technological solutions into their curricula.

The study seeks to provide factual data about current curricula with a focus on RSI, AI, and CAI tools (teaching hours, types of platforms and tools), and to identify to which extent the aforementioned cautious model is shared internationally. It also explores key changes implemented in recent years, as well as the overall strategy and challenges in curriculum design, including the main drivers of decision-making, external pressures, institutional constraints, and potential responses to these challenges. In addition, it analyses specific approaches to integrating technology into training, particularly in relation to assessment, trainer recruitment, varying student digital literacy and ethical considerations. Finally, the study considers prospects and anticipated developments in this area.

More broadly, the findings are expected to clarify whether technological integration is being developed as a complementary layer to traditional interpreter training or whether it is gradually reshaping the hierarchy of competences within interpreter education.

The study findings will be presented for the first time at this conference.

The findings will contribute to the development of evidence-based recommendations for CI course directors, trainers, and students. By identifying prevailing practices, institutional constraints, and innovative curricular solutions, the study will support the decision-making process with regard to ensuring a balanced integration of technology in interpreter training. Ultimately, the project aims to support the development of resilient, sustainable and future-oriented interpreter training models.

Poster 10 · More than Just Listening: How Visual Supports Shape Simultaneous Interpreting

Camille Breuskin

University of Antwerp, Antwerp, Belgium

Simultaneous interpreting (SI) is a form of oral language mediation in which interpreters render a speech into another language while the speaker is still delivering the original message, typically from a soundproof booth. Because comprehension, production, and monitoring processes must be carried out concurrently and under severe time constraints, SI is widely considered one of the most cognitively demanding language activities. Contemporary SI is furthermore performed in an increasingly technologized and multimodal working environment (Mellinger, 2023), in which interpreters must process information originating from multiple channels simultaneously, i.e., multimodally. In addition to verbal and nonverbal cues produced by speakers, interpreters are frequently exposed to a variety of visual supports, including PowerPoint slides, digital terminology resources, live captions, and computer-assisted interpreting (CAI) tools.

While these visual supports are generally assumed to facilitate comprehension and accuracy, they may also compete for attentional resources and increase cognitive load. The relationship between the potential benefits of multimodal support to performance and the additional cognitive demands associated with processing multiple streams of information remains insufficiently explored empirically (Fantinuoli, 2023). This issue has become particularly relevant with the rapid development of artificial intelligence (AI)-based CAI tools.

CAI tools can be defined as “all sorts of computer programs and mobile applications specifically designed and developed to assist interpreters in at least one of the different subprocesses of interpretation” (Fantinuoli, 2021, p. 513). Their emergence is part of what Fantinuoli (2018) describes as the “technological turn” in interpreting, often considered the third major technological transformation of the profession after the introduction of wired systems for speech transmission in the 1920s and the rise of the Internet in the 1990s. Recent CAI systems integrate AI technologies such as automatic speech recognition (ASR), machine translation, terminology extraction, and real-time captioning, thereby fundamentally reshaping interpreters’ interaction with information during SI.

Despite the increasing availability and adoption of these technologies, empirical evidence regarding their impact on interpreters’ cognitive processing, stress, visual attention, and performance remains limited. Existing studies have often focused on isolated tools or laboratory-based settings, while relatively little research has investigated multimodal support configurations in ecologically valid interpreting conditions involving professional interpreters. A handful of studies based on experiments (e.g., Defrancq & Fantinuoli, 2021; Desmet et al., 2018; Frittella, 2023; Prandi, 2023) indicate positive effects on performance. However, negative trends based on qualitative data and anecdotal observations are also noted and deserve further exploration (Prandi, 2025).

The present study addresses this gap by investigating the effects of different types of visual support on SI performance and cognitive processing in simulated real-life conditions. More specifically, the study examines how combinations of slides, automatic captioning, and AI-based CAI tools influence interpreters’ performance, visual attention allocation, cognitive load, stress, and user experience.

To this end, a within-subject experimental design was implemented with 18 professional conference interpreters. Participants interpreted six speeches of approximately twelve minutes each under six experimental conditions: (1) SI without visual support, (2) SI with slides, (3) SI with slides and automatic captioning, (4) SI with CAI support, (5) SI with CAI support and slides, and (6) SI with CAI support, slides, and automatic captioning. Speech rate, widely recognized as a major problem trigger in SI, was systematically manipulated as an additional experimental factor. The experiment was conducted in a professional interpreting booth in order to maximize ecological validity and approximate authentic working conditions. Participants wore mobile eye-tracking glasses throughout

the tasks, enabling the collection of gaze behavior data in a dynamic and realistic interpreting environment.

The study combines objective and subjective measures to provide a comprehensive account of multimodal SI processing. The resulting dataset includes:

Performance data: audio recordings of interpreted output;

Eye-tracking data: mobile gaze recordings, fixations, saccades, and gaze transitions between modalities;

Physiological data: wristband-based heart-rate measurements used as indicators of stress and cognitive effort;

Self-reported data: post-task questionnaires and semi-structured interviews assessing perceived effort, stress, usability, and user experience.

Performance analysis will adopt a mixed-method approach. First, problem-trigger accuracy will be evaluated by categorizing renditions as “correct rendition,” “good approximation,” “incorrect rendition,” or “omission.” Second, propositional accuracy and fluency will be assessed using an adapted version of the NTR model (Romero-Fresco & Pöchhacker, 2017), which allows for weighted error analysis by distinguishing between minor, major, and critical errors.

Eye-tracking analysis will make it possible to examine how interpreters distribute visual attention across modalities, how often and under which circumstances they shift attention between visual supports, and whether visual competition emerges between concurrent information sources. Particular attention will be paid to allocation strategies across different speech rates and support configurations. Physiological and self-reported measures will further contribute to understanding how interpreters experience multimodal SI environments in terms of perceived effort, stress, and usability.

The study aims to contribute to current debates on technology integration in conference interpreting by providing empirical evidence on the cognitive implications of multimodal support environments. Beyond assessing whether visual supports are beneficial or detrimental overall, the study seeks to identify under which conditions particular combinations of tools may support or hinder interpreters’ performance and attentional management.

The paper will present the experimental design, methodological framework, and preliminary findings. Expected results include identifiable patterns of visual attention allocation across modalities, differences in interpreters’ interaction with visual supports depending on speech rate, and potential trade-offs between improved terminological accuracy and increased cognitive load. The study is also expected to shed light on interpreters’ subjective perceptions of AI-based CAI tools and their usability in realistic SI contexts.

By combining performance analysis, eye-tracking, physiological measures, and user experience data within a single ecologically valid experimental framework, the study contributes to ongoing research on multimodality, cognitive ergonomics, and technological change in interpreting. These findings are expected to contribute to a better understanding of the cognitive implications of AI-assisted interpreting environments and to inform both interpreter training and the future design of CAI technologies. More broadly, the study addresses how emerging AI-supported tools are reshaping interpreters’ working environments one hundred years after the technological innovations that first enabled simultaneous interpreting.

References

Defrancq, B., & Fantinuoli, C. (2021). ASR in the booth: assessment of system performance, interpreters’ performances and interactions in the context of numbers. *Target*, 33(1), 73–102. <https://doi.org/10.1075/target.19166.def>

Desmet, B. et al. (2018). Simultaneous interpretation of numbers and the impact of technological support. In C. Fantinuoli (Ed.), *Interpreting and technology* (pp. 13–27). Language Science Press. <https://doi.org/10.5281/zenodo.1493281>

Fantinuoli, C. (2021). Conference interpreting and new technologies. In M. Albl-Mikasa, & E. Tiselius (Eds.), *The Routledge handbook of conference interpreting* (pp. 508–522). Routledge. <https://doi.org/10.4324/9780429297878-44>

Frittella, F. M. (2023). Usability research for interpreter-centred technology: The case study of SmarTerp. Language Science Press. <https://doi.org/10.5281/zenodo.7376351>

Mellinger, C. D. (2023). Embedding, extending, and distributing interpreter cognition with technology. In G. Corpas Pastor & B. Defrancq, *Interpreting Technologies – Current and Future Trends* (pp. 195–216). Benjamins. <https://doi.org/10.1075/ivitra.37.08mel>

Prandi, B., 2023. Computer-assisted simultaneous interpreting: A cognitive-experimental study on terminology. Language Science Press, Berlin (*Translation and Multilingual Natural Language Processing*, 22). <https://doi.org/10.5281/zenodo.7143055>

Prandi, B., 2025. Computer-assisted interpreting (CAI) tools and CAI tools training. In, E. Davitti, T. Korybski, S. Braun (Eds), *The Routledge Handbook of Conference Interpreting* (pp. 123–144). Routledge. <https://doi.org/10.4324/9781003053248>

Romero-Fresco, P. & Pöchhacker, F., 2017. Quality assessment in interlingual live subtitling: The NTR Model. *Linguistica Antverpiensia New Series*, 15, pp. 149–167. <https://doi.org/10.52034/lanstts.v16i0.438>

Poster 11 · VKD Market Snapshot: German Conference Interpreters' Perceptions of Market Change in the Context of AI and Economic Uncertainty

Nele Kirstein & Sabine Seubert

TH Köln – University of Applied Sciences, Cologne, Germany; VKD im BDÜ e. V., Cologne, Germany

Artificial Intelligence is rapidly changing jobs and workplaces across industries, including translation and interpreting. In 2025, a study by Tomlinson et al., conducted on behalf of Microsoft Research, has caught the attention of language professionals around the world, stating that “[I]nterpreters and Translators are at the top of the list [of occupations most affected by AI], with 98% of their work activities overlapping with frequent Copilot tasks with fairly high completion rates and scope scores” (Tomlinson et al., 2025, p. 11). This headline-grabbing statement requires further attention. Under closer scrutiny, the logic of the study is flawed, given that it only tracked what specific tasks individuals asked Copilot to carry out and how successfully they felt these tasks were completed, and not whether these individuals would have potentially paid translators and interpreters for these services, or whether indeed some of these translation tasks were carried out by professional translators themselves. However, flawed or not, it is a prominent example of a long series of studies and news articles predicting the extinction of the translation and interpreting professions as we have known them (e.g. Frey & Llanos-Paredes, 2025; Vivian, 2025) and many translators and interpreters have voiced concern about the future of their profession (e.g. Rösener et al., 2024; Rivas Ginel et al. 2024; ELIS 2024; ELIS 2025; ELIS 2026).

The 2025 and 2026 ELIS reports indicate that some of the prevailing fears have started to turn into reality: Both independent professionals and language companies reported a decrease in revenue and/or growth in 2024 (ELIS, 2025, p. 12) and a continued decline in market activity in 2025/26 (ELIS, 2026, p. 12). Interestingly, while in 2025 23% of independent language professionals were considering ending their freelance career (ELIS, 2025, p. 19), that number dropped to 17% in the 2026 report (ELIS, 2026, p. 18). At the same time, the data also suggests that individual language professionals are already leaving the industry, stating that the number of professionals in the market has dropped by 11% (ibid, p. 15). The most recent ELIS report has also shown that “almost one in two [language professionals] are acquiring new skills or developing another main activity to cope with decreasing activity” (ibid). Overall, the 2026 survey establishes a quite pessimistic mood among language professionals who “do not see any signs of improvement for this year” (ibid, p. 12). They report on declining market rates (although 25 % of independent language professionals have increased their rates in 2025, ibid, p. 13), price pressure, increased competition and lost confidence in the sustainability of a freelancing career in the language industry (ibid, pp. 12ff.).

Despite the scepticism and what the ELIS study authors in 2025 referred to as “alarming sentiments” and “a serious wakeup call” (ELIS, 2025 p.19), there are also more positive attitudes, particularly in (conference) interpreting: “Many believe that AI cannot replace human interpreters due to its limitations in understanding meaning, emotion, cultural differences, and body language.” (Pielmeier, 2024, p. 16) Instead, there are calls for AI to be used to complement human interpreters (ibid, p. 17). In addition, the ELIS 2026 survey found that, while income satisfaction has declined among language professionals in general, it has actually risen among those who are only active as interpreters (ELIS, 2026, p. 39). Although much research has been conducted into translators’ and interpreters’ perceptions of AI, their fears and concerns, and the potential impact of new technologies on the job market, recent studies have focused much less on the current state of the language markets. Apart from the cited ELIS surveys, there is little evidence of the impact of AI disruption and other factors, such as the ongoing economic crisis and geopolitical tensions, on the job market in the language industry. Furthermore, the ELIS surveys provide an overview of the language industry as a whole and do not distinguish between translators and conference interpreters. Given that previous research has shown more positive attitudes among conference interpreters (see above), a study focusing only on this subgroup of language professionals would be highly interesting and could contribute to the current debate.

Building on this, we propose analysing local data from the German market and contextualising it against the backdrop of the aforementioned broader studies and surveys. The German professional association of conference interpreters, VKD, has recently conducted a market survey among its members, the results of which will be presented at our poster session. Our poster session aims to contribute a local market perspective focusing only on conference interpreters to the ongoing debate around the disruption of the language industry. Instead of just perceptions and fears, we will analyse actual market developments.

Our research should answer the following main research questions: What is the current state of the German conference interpreting market? Did prices, revenue, and the number of interpreting assignments decrease in 2024/2025? How do German conference interpreters view current developments around AI and the economy? What do they see as the biggest challenges, and how confident are they about the future of their profession?

The study is based on secondary analysis of survey data collected by the German “Verband der Konferenzdolmetscher” between 19 December 2025 and 18 January 2026. The survey targeted professional conference interpreters (members of both VKD and AIIC Germany) and explored perceptions of recent market developments, challenges and opportunities, and changes in the profession. Due to the non-probability sampling design and professional association distribution channels, the dataset is considered exploratory rather than statistically representative. The analysis therefore focuses on descriptive trends, cross-tabulations and thematic patterns, rather than generalisation at the population level. Quantitative responses will be analysed using descriptive statistics and comparative subgroup analysis where appropriate.

The survey is intended to be repeated and further refined in subsequent rounds to improve methodological rigor and expand the empirical evidence base.

References

- European Commission, Directorate-General for Interpretation (DG SCIC) (2024). Summary of the Interpreting Europe Conference 2023: Attracting talent for an upskilled profession (Virtual event, 30 November–1 December 2023). European Commission. URL: https://commission.europa.eu/document/download/5ba45f4b-3119-4cae-9df7-3081e83cc514_en
- European Language Industry Survey 2024. Trends, Expectations and Concerns of the European Language Industry. (2024). ELIS Research.
- European Language Industry Survey 2025. Trends, Expectations and Concerns of the European Language Industry. (2025). ELIS Research.
- European Language Industry Survey 2026. Trends, Expectations and Concerns of the European Language Industry. (2026). ELIS Research.
- Frey, C. B., & Llanos-Paredes, P. (2025). Lost in Translation: Artificial Intelligence and the Demand for Foreign Language Skills. (Working Paper). Oxford Martin School, University of Oxford. https://oms-www.files.svdcn.com/production/downloads/academic/Frey_LlanosParedes_2025_LostInTranslation.pdf
- Pielmeier, H. (2024). Summary of “Perceptions on Automated Interpreting” - Results of a Large-Scale Study of End-Users, Requestors, and Providers of Interpreting Services and Technology. CSA Research. <https://insights.csa-research.com/chapters/305013619/Introduction>
- Rivas Ginel, M. I., Sader Feghali, L., & Accogli, F. (2024). Exploring Translators’ Perceptions of AI. ELC Survey.
- Rösener, C., Canfora, C., Dörflinger, T., Hoberg, F., & Varga, S. (Eds.). (2024). Übersetzen im Wandel: Wie Technologisierung, Automatisierung und Künstliche Intelligenz das Übersetzen verändern. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-42903-4>
- Tomlinson, K., Jaffe, S., Wang, W., Counts, S., & Suri, S. (2025). Working with AI: Measuring the Occupational Implications of Generative AI. arXiv. <https://doi.org/10.48550/ARXIV.2507.07935>
- Vivian, I. (2025, August 4). ‘I wouldn’t recommend anyone in their 20s study interpreting.’ The Brussels Times. <https://www.brusselstimes.com/belgium/1689271/the-age-of-cost-cutting-is-ai-stealing-interpreters-jobs>

Poster 12 · The Project INNOVATRAD-CM: Research and Innovation in Real-Time Speech Generation and Interpretation

Raquel Lázaro Gutiérrez & Elena Alcalde Peñalver

Universidad de Alcalá, Alcalá de Henares, Spain

In the field of computer-assisted simultaneous interpretation, various tools have been developed to support interpreters throughout all stages of the interpretation process. These tools are designed to optimize the workflow, from preparation to interpretation in the booth. They offer features such as the creation of custom glossaries, the development and management of specialized corpora, the transcription of figures, and the accurate interpretation of specific terms and named entities, real-time transcription of conferences, and automated note-taking; leveraging advanced technologies such as automatic speech recognition and text-to-speech synthesis algorithms. These technological advancements aim to increase the efficiency and accuracy of interpreters in both traditional and remote settings.

However, the adoption of these tools faces significant challenges. One of the main issues is the generalizability of their effectiveness, as studies conducted to date often involve small, standardized samples that do not reflect the diversity of real-world simultaneous interpretation situations and exclude modalities such as bilateral interpretation. Furthermore, the integration of emerging technologies such as automatic speech recognition has not yet been fully implemented in these workflows, which limits these tools' ability to consistently extract and process terms in real time—a crucial aspect for improving the accuracy and effectiveness of interpretation.

Interpreters' relationship with technology has traditionally been complex. Unlike translators, interpreters have been slower to adopt automation, partly due to the cognitively demanding nature of their work, which requires quick and accurate responses in real time. This slow adoption is influenced by the perception that technology may not meet all of interpreters' specific needs and by the fear that it could increase the cognitive load during interpretation, potentially impairing performance rather than improving it.

To address these challenges, solutions such as the creation of specialized interpreting labs and the inclusion of specific modules in interpreter training curricula have been proposed. These initiatives aim not only to facilitate the adoption of new technologies but also to ensure that future interpreters are adequately equipped with the skills needed to effectively use these tools in their professional practice.

One of the main drawbacks of CAI tools is that most of them are paid and require subscriptions, which can limit their accessibility. Furthermore, their source code is not available as open source, since they are proprietary software. This complicates their integration into academic or professional processes for conducting controlled studies of their impact on interpreters' workflows. Another significant problem is that many of these tools do not integrate all the necessary functionalities into a single product, forcing interpreters to use multiple tools to fully meet their needs during an assignment. Furthermore, current tools often do not integrate well with the equipment used in simultaneous interpreting. On the other hand, CAI tools have not yet been adapted for other interpreting modalities such as consecutive interpreting. Recent studies on the use of technologies by bilateral interpreters indicate high levels of intended use, particularly in telephone interpreting (OPI) and video-based bilateral interpreting (VRI) settings. However, acceptance of these technologies is low due to the lack of adaptation of CAI tools for this type of interpreting.

As can be seen, advances in the field of CAI offer significant opportunities to improve professional interpreting practice. However, there are still major obstacles that must be overcome to maximize its potential. In this context, it is necessary to develop and integrate technologies that bring together all the necessary functionalities into a single, comprehensive, and efficient tool. This tool must be perfectly aligned with the practical needs of interpreters, and its rigorous evaluation in controlled

real-world environments is essential to ensure effective and useful integration into the field of interpreting. Furthermore, it is essential that the developers of these tools actively incorporate feedback from interpreter users to ensure the practicality and relevance of the final product.

In response to the specific needs of interpreters a multidisciplinary team has gathered that comprises experts in software development, artificial intelligence, semantics, lexicography, ontology engineering, and translation and interpretation. This team develops the project INNOVATRAD, aimed at not only developing advanced technological tools that transform the practice of interpreting, but also their implementation, evaluation, and continuous monitoring to ensure their ongoing refinement.

The project is based on the hypothesis that it is possible to develop a comprehensive and user-friendly CAI tool, designed using advanced AI, ASR, NMT, TTS, and NLP technologies based on a language model, which incorporates continuous user feedback and overcomes the limitations of current CAI tools. It is hypothesized that such integration, within a single user-centered and accessible product, will facilitate its adoption and optimize interpreters' performance. Furthermore, the importance of rigorous evaluation in controlled real-world environments is emphasized to validate the effectiveness of this tool in the practice of interpreting.

The novelty of the project lies in the integration of all these advanced features into a single, accessible product, combining various technologies into a single tool, with a user-centered development approach, and an evaluation based on Learning Analytics and Eye Tracking to monitor whether the tool's use in a controlled academic environment allows for its refinement during the different phases of development, ensuring that advanced AI technologies, ASR, NMT, TTS, and NLP based on an LM effectively adapt to the needs of interpreters. This approach will ensure that the tool is practical, efficient, and aligned with the real needs of interpreters, promoting faster adoption and effective use in educational and professional settings in the future.

Poster 13 · Outil d'auto-entraînement des apprentis-interprètes assisté par l'IA

Lina Sader Feghali & Wadad Wazen Gergy

Saint-Joseph University of Beirut, Beirut, Lebanon

Le projet qui fera l'objet de cette présentation a pour objectif de vérifier la faisabilité d'un pipeline combinant plusieurs grands modèles de langage (GML) (Fantinuoli 2023) pour assister l'apprenti-interprète dans son auto-entraînement et éventuellement l'interprète débutant lors de la préparation d'une conférence.

Ce projet intervient au moment où l'École de traducteurs et d'interprètes de Beyrouth (ETIB) entreprend une refonte du programme de Master en interprétation qui sera lancé à la prochaine rentrée universitaire 2026-2027. L'auto-entraînement étant une composante essentielle de la formation (initiale et continue) des interprètes, il a fallu lui accorder une place encore plus importante dans la formation afin d'assurer un meilleur encadrement aux étudiants dont le rapport au savoir et à la maîtrise des langues a changé. Il s'agit d'un encadrement méthodique du travail autonome qui offre aux apprentis-interprètes un cadre structuré propice au développement de pratiques d'apprentissage réflexives et durables. En favorisant l'autonomisation des étudiants, cet encadrement contribue ensuite au développement de compétences transversales telles que la planification du travail, l'auto-évaluation et la régulation des apprentissages. Il permet enfin la consolidation des stratégies d'entraînement spécifiques tout en préparant les étudiants à s'inscrire dans une logique de développement professionnel continu.

Ce nouveau format d'encadrement se présente sous deux formes : la première forme est un espace d'auto-entraînement dans Moodle conçu spécialement pour l'Université de Bologne (Bertozzi 2024) dont l'approche didactique indépendante des langues de travail visent à consolider les compétences stratégiques et technologiques des futurs interprètes. Composante structurante du dispositif de formation en interprétation de conférence, cet espace est constitué d'unités intégrées à la plateforme Moodle que les étudiants sont amenés à compléter de manière progressive, selon leur propre rythme, tout en respectant un échéancier prédéfini. La deuxième forme est l'entraînement supervisé dans le cadre duquel un enseignant assure le suivi en groupe des étudiants pour faire le point sur les activités effectuées dans Moodle et pour combler les lacunes éventuelles en relation notamment avec les compétences linguistiques des trois langues de travail.

Quoiqu'utiles voire incontournables, les deux formats d'encadrement demeurent insuffisants pour combler les besoins spécifiques des apprentis-interprètes dont la principale langue de travail est la langue arabe, une langue qualifiée de moins dotée compte tenu du manque de données d'entraînement fiables de bonne qualité notamment dans les domaines de spécialité. D'où l'initiative prise par l'ETIB, en collaboration avec le Centre de l'innovation numérique et de l'intelligence artificielle (CINIA) de l'USJ et le programme Génie Informatique et Communications – option Intelligence Artificielle de l'École supérieure d'ingénieurs de Beyrouth, d'explorer la possibilité de concevoir un outil convivial pour et par les interprètes qui intègre les avancées récentes en LLM dans une plateforme IA modulaire combinant génération automatique de discours dans les trois langues de travail, synthèse vocale, transcription automatique, reconnaissance automatique de la parole et analyse de performance, l'outil devrait permettre aux apprentis interprètes de s'auto-entraîner de manière autonome, avec, si possible, un feedback adaptatif personnalisé.

L'outil à développer permettra à l'apprenti-interprète de réaliser, via l'IA, sans pour autant recourir à la rédaction de prompts (Wiedenmayer, 2026) ni à de multiples itérations, les tâches suivantes, entre autres :

Se documenter sur le thème d'une conférence (avec ou sans documents) ;

Générer un résumé des principales idées/thématiques/problématiques ;

Générer un QCM pour tester les connaissances sur le sujet ; repérer les termes/notions de base ; élaborer un glossaire dans les langues de travail (arabe, français, anglais) ;

S'auto-entraîner en générant des discours pour l'entraînement avec des scénarios réalistes (conférences, interviews, débats, réunions diplomatiques) à partir d'un corpus avec des paramètres variables : longueur, langues de travail (arabe/français/anglais), différents domaines de spécialité, complexité lexicale variable, structure (logique ou désorganisée), divers registres, comportant beaucoup de nombres, riches en terminologie et en noms propres, etc.

Générer des sorties orales et écrites pour auto-évaluer sa prestation et éventuellement obtenir un feedback adaptatif personnalisé.

Le projet est certes ambitieux d'autant plus qu'il est légitime de se demander jusqu'à quel point l'IA est capable, par exemple, de produire des discours de bonne qualité notamment en arabe, de simuler les accents, de relever les erreurs (hésitations, longs silences, auto-corrections, répétitions, lapsus linguae, faux départs, erreurs de reproduction des chiffres, de dates, etc.) voire d'identifier les erreurs récurrentes (par exemple, dans la vocalisation d'un texte en langue arabe) et de proposer des corrections.

À part sa fonction principale qui est celle de soutenir l'auto-entraînement, cet outil, en cours de développement grâce à la technique du vibe coding, sert deux principaux objectifs pédagogiques : d'une part, il permet aux étudiants d'interagir avec un outil afin d'en identifier les limites et les potentialités ; d'autre part — et surtout — il leur donne éventuellement les moyens de concevoir leurs propres applications.

À l'heure où l'avenir de la profession d'interprète est remis en question, il devient de plus important de former de futurs professionnels capables d'interagir avec l'IA de manière créative, flexible et éclairée (Wiedenmayer, 2026, p.110). C'est d'ailleurs l'un des résultats d'apprentissage du nouveau programme de l'ETIB, à savoir former des interprètes professionnels réflexifs, capables d'utiliser de manière critique, éthique et efficace les nouvelles technologies et de s'adapter de manière proactive à l'évolution des environnements technologiques professionnels.

Références

Bertozzi, M. (2024). Continuous self-learning for conference interpreting trainees: The case of the University of Bologna. *The Interpreters' Newsletter*, 29 (Special Issue), 19–39. <https://doi.org/10.13137/2421-714X/36971>

Fantinuoli, C. (2023, 18 novembre). Empowering autonomous learning through AI. Claudio Fantinuoli Official Website. <https://www.claudiofantinuoli.org/2023/11/18/how-ai-can-empower-learners/>

Fantinuoli, C. (2026). Reflections on technology and interpreter training. *REDIT*, XX, 21–24.

Wiedenmayer, A. (2026). Artificial Intelligence as a Pedagogical Tool for Speech Generation in Conference Interpreter Training. *International Journal of Language, Translation and Intercultural Communication*, 11, 100–113. <https://doi.org/10.12681/ijltic.44846>

Poster 14 · CALIPSO: A Programmable Environment for Research and Training in AI-Assisted Interpreting

Geoffrey Tissier & Leila Cherrouk

Université Paris 8 Vincennes-Saint-Denis, France; Universidad de Alicante, Spain

The growing presence of AI-assisted support systems in professional conference interpreting (including real-time captioning, automatic speech recognition and machine translation tools) is reshaping the attentional and cognitive demands of the interpreting booth in ways that remain largely understudied (Fantinuoli & Dastyar, 2022). While systems such as SmarTerp or InterpretBank are increasingly used in operational contexts, researchers still lack configurable infrastructures for studying interpreter-AI interaction under controlled conditions. Most existing CAI tools are designed as operational support environments rather than experimental or pedagogical platforms, affording limited control over the timing, reliability and congruency of the visual information presented to interpreters (Fantinuoli, 2018). This limits both the systematic investigation of cognitive ergonomics in AI-assisted interpreting and the development of training scenarios based on controlled exposure to machine-generated interference.

Beyond research, interpreter education faces a structural challenge: AI-assisted tools are entering professional practice faster than curricula are evolving to address them. While interpreters are increasingly expected to engage critically with machine-generated assistance, few training programmes currently offer controlled environments in which students can develop the skills needed to evaluate, resist or strategically deploy AI-generated suggestions under realistic production conditions.

The present project addresses this gap through the development of CALIPSO (Controlled AI Lab for Interpreting Performance and Support Operations), a programmable simulation environment designed for both research and pedagogical applications in AI-assisted conference interpreting.

CALIPSO aims to provide a configurable infrastructure allowing researchers and educators to design, manipulate and observe interpreter interaction with AI-generated assistance under controlled multimodal conditions. Rather than functioning as a fixed support tool, the platform operates as a programmable simulation environment in which the cognitive scenario is defined in advance while the system dynamically generates assistant behaviour during the interpreting task. The project pursues dual objectives: advancing experimental research on human-AI interaction in simultaneous interpreting, and supporting simulation-based interpreter training involving varying levels of AI reliability and multimodal interference. The poster will present the conceptual architecture of the platform, the scenario editor interface, and illustrative examples of simulated interpreting conditions across varying levels of congruency and interference.

The development of CALIPSO is guided by the following questions: How can AI-assisted interpreting environments be simulated under controlled yet ecologically plausible conditions? Which interface and timing parameters influence interpreters' attentional allocation and cognitive effort during interaction with AI-generated suggestions? How do interpreters regulate trust when confronted with varying levels of system reliability and plausible ASR- or MT-like interference? To what extent can programmable simulation contribute to the development of technology-related competencies within interpreter education?

CALIPSO was developed within the LUTIN Userlab at the Cité des Sciences et de l'Industrie in Paris. The platform is implemented in Unity and designed to support synchronized multimodal experimentation. A central design principle is accessibility: trainers configure simulations through a visual scenario editor rather than modifying the underlying code architecture, thereby separating scenario design from software programming.

The scenario editor allows users to define the categories and timing of information displayed during the task, configure interface properties, determine the temporal behaviour of visual events, and

manipulate congruency between the source speech and the visual suggestions. Controlled ASR- or MT-like interference can be introduced directly into the simulation, affecting numerals, terminology and named entities with parametrised control over type, strength and plausibility. Unlike static CAI systems, CALIPSO accommodates a wide range of conditions without requiring modifications to the software architecture itself. These range from highly reliable assistance suited to novice learners to cognitively dense multimodal conflict scenarios designed to probe attentional regulation and trust calibration.

A distinctive feature of the platform is its synchronization with multimodal observation streams, including eye-tracking data, physiological signals, behavioural recordings and interpreting output. This integration enables frame-accurate analysis of interpreters' responses to specific visual events, supporting investigation of attentional allocation, cognitive load and adaptive trust during AI-assisted interpreting (Seeber, 2011).

CALIPSO addresses a methodological gap in research on AI-assisted interpreting by proposing an infrastructure that allows for the systematic manipulation of assistance reliability, visual timing and multimodal interference under ecologically plausible conditions. By enabling synchronized observation of interpreter behaviour across eye-tracking, physiological and linguistic channels, the platform may support more fine-grained investigation of phenomena such as attentional capture, cognitive load and trust calibration during human-AI interaction in interpreting contexts (Risko & Gilbert, 2016).

From a pedagogical perspective, the project suggests that programmable simulation environments could usefully complement existing CAI training approaches by offering structured, progressive exposure to unreliable assistance. Such environments may help interpreters develop critical engagement with AI-generated input and begin calibrating their reliance on assistive systems before encountering them under the full constraints of professional practice.

Taken together, these contributions suggest that CALIPSO may offer a productive shift from viewing CAI systems as static support tools toward programmable environments for studying and teaching human-AI interaction in conference interpreting.

References

- Fantinuoli, C. (2018). Interpreting and technology. In C. Fantinuoli (Ed.), *Interpreting and technology* (pp. 1–12). Language Science Press.
- Fantinuoli, C., & Dastyar, V. (2022). Interpreting and the emerging augmented paradigm. *Interpreting and Society*, 2(2), 185–194.
- Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676–688.
- Seeber, K. G. (2011). Cognitive load in simultaneous interpreting: Existing theories - new models. *Interpreting*, 13(2), 176–204.

Poster 15 · Effective theory-based AI prompting for in-class speech material: A solid point of departure

Lazaros Ouzounoglou

Independent researcher, Greece

Interpreting has changed considerably over the years, and so has interpreter training. Be that as it may, some elements of training remain as indispensable as they were a hundred years ago: the teacher, the student and the speech (Kautz, 2000). However, as succinctly put by Setton & Dawrant (2016), “teachers make or break an interpreter training course” (p. 11). This, in turn, largely depends on the trainer’s ability to select and/or adapt practice material to the student’s level, needs and weaknesses. Whereas technology seems to be calling the tune in the language industry and, by extension, interpreter training, opening up new fields of research, speech selection has yet to receive the attention it deserves, in literature and training alike.

This may seem too rudimentary in view of the rapid changes we are witnessing, prompting us to search for ways to stay relevant. Nevertheless, it is of the highest relevance if high-quality training is what interpreting schools are still aiming at in order to give budding interpreters a competitive edge over laypeople offering interpreting services (Kalina, 2011; Kautz, 2000). Given the limited amount of class time, it is advisable that in-class speech material embody the principles of deliberate practice, allowing for a steady progression and increase in the degree of difficulty.

Largely informed by my master’s thesis (Freri & Ouzounoglou, 2021), the present paper addresses the topic of selection of appropriate in-class material oriented towards deliberate practice. Firstly, it provides a concise overview of the appropriateness criteria that can be used to systematise speech selection and/or adaption for didactic purposes. Secondly, it introduces the concept of problem-strategy units and shows how it can be utilised to prepare deliberate practice sessions. Finally, it illustrates how AI can be leveraged to this end based on interpreting didactics. It is, therefore, hoped that this contribution will provide a solid point of departure for interpreter trainers in a world of sensory, emotional and cognitive overload, where virtually everything has been said, where tools and information are at our fingertips, yet compasses are somehow missing.

As far as appropriateness criteria are concerned, I have adapted Kautz’s (2000) text selection criteria for translation to interpreting didactics. They are intended to serve as general guidelines to aid trainers in making informed choices about the speech material they expose their students to. For reasons of space, they are summarized as follows:

1. Didactical suitability: the speech offers ample opportunity to practise and/or consolidate interpreting-relevant skills;
2. Authenticity: the speech is derived from a real-life communication situation where interpreting services are used;
3. Subject matter: the speech is relevant to the target-language culture and professional practice;
4. Degree of difficulty: the degree difficulty of the subject matter, content, language and interpreting skills corresponds to the student’s competence level, thereby facilitating a steady learning progress;
5. Text type: the speech type occurs frequently in professional practice and addresses a more or less clearly defined target-language audience;
6. Length: the speech allows for the usual didactic processes within class time;
7. Interestingness: working on the speech is designed in an interesting way, even when the text is not interesting in itself; motivating and challenging;
8. Topicality: the speech is of current relevance and serves as a means of assessing whether students keep abreast of current affairs; (Kautz, 2000; Ouzounoglou, in Freri & Ouzounoglou, 2021, pp. 188–189).

Against this backdrop, speech selection and evaluation of appropriateness for classroom use will inevitably take into account the respective language pair and direction, the interpretation mode, and other objective and subjective factors. As they are easier to identify and measure, this paper focuses solely on objective factors of difficulty, commonly referred to as problem triggers, in simultaneous interpreting from German, an SOV language, into an SVO target language (e.g. English or Greek), in particular verb-final sentences and numbers.

It is worth recalling that interpreting is an inherently demanding cognitive task, which can be viewed as “a problem of capacity availability and allocation” (Chabasse & Dingfelder Stone, 2015, p. 75). Some processes are controlled and attention-intensive, while others are automatic or highly automatable (Gile, 2009). Problem triggers add to the difficulty of this effortful task and are considered objective indicators of an increase in the degree of difficulty (Andres, 2015), or, more precisely, of the cognitive load expected to accumulate during the interpretation of a speech (Ouzounoglou, in Freri & Ouzounoglou, 2021).

In contrast to speeches in real-life communication settings, in-class speeches should fulfil a didactic-pedagogical function, namely the development of interpreting competence (Andres, 2015). Arbitrarily selected speeches are not only counterproductive but can also compromise student performance, with long-term consequences (Hönig, 2002). In-class speech material must, therefore, provide the framework for carefully planned deliberate practice sessions targeted at achieving specific learning outcomes, along with the development of automatisms.

This is where problem-strategy-units can prove useful. After typologising language-pair and mode-specific problem triggers, trainers should match them with suitable strategies that students are expected to develop and employ efficiently. Consequently, trainers can select their speeches based on these criteria and didactise their material more purposefully using problem-strategy units (Freri & Ouzounoglou, 2021).

Picking up on the aforementioned problem-triggers, the following problem-strategy units have been identified in the relevant literature:

Verb-final sentences – Chunking, anticipation (Liontou, 2011)

Numbers – Flexible décalage, approximation (Kader & Seubert, 2015)

Having established a solid foundation for speech selection and didactisation, trainers could use artificial intelligence to produce or adapt speeches that focus more specifically on certain problem areas. Using the above guidelines, they can refine the prompts they enter into the AI chatbot of their preference to obtain better results. However, it remains to be tried and tested whether these results are reliable. Mastering the art of effective prompting may well prove useful in this regard. Thus, as a closing note—not least because I find it amusing—I will hazard a prompt to make this more tangible.

Dear AI chat bot, I am a conference interpreter trainer, teaching simultaneous interpreting from German into English to students in their second semester of a two-year master's degree programme. I want you to shorten the following speech to 1,000 words. Please include 7 numbers, not so close to each other, ranging from 100 to 5,000,000, and 5 verb-final sentences of 20 words each. Please allow for redundancies after every fifth sentence. And 5 common idioms would be just great. Pretend you are the speaker. And please do not switch to English only because my prompt is in English. Remember, you are a German speaker! Here is the speech: ...

References

- Andres, D. (2015). Easy? Medium? Hard? The Importance of Text Selection in Interpreter Training. In D. Andres & M. Behr (Eds.), *To Know How to Suggest...: Approaches to Teaching Conference Interpreting* (pp. 103-124). Frank & Timme.
- Chabasse, C., & Dingfelder Stone, M. (2015). Capacity Management in Interpretation: Efforts, Directionality, and Language Pair Considerations. In D. Andres & M. Behr (Eds.), *To Know How to Suggest...: Approaches to Teaching Conference Interpreting* (pp. 75-102). Frank & Timme.

Freri, M., & Ouzounoglou, L. (2021). Optimale Herausforderung im Konsekutiv- und Simultandolmetschunterricht am Beispiel didaktisch kommentierter Übungsreden mit Deutsch als Ausgangssprache [Unpublished master's thesis]. Johannes Gutenberg University Mainz.

Gile, D. (2009). Basic Concepts and Models for Interpreter and Translator Training. John Benjamins.

Hönig, H. G. (2002). Piece of Cake – or Hard to Take? Objective Grades of Difficulty of Speeches Used in Interpreting Training. In Teaching Simultaneous Interpretation into a “B” Language, EMCI Workshop 20–21 September, 38-50.

Kader, S., & Seubert, S. (2015). Anticipation, Segmentation ... Stalling? How to Teach Interpreting Strategies?. In D. Andres & M. Behr (Eds.), To Know How to Suggest...: Approaches to Teaching Conference Interpreting (pp. 125-144). Frank & Timme.

Kalina, S. (2011). Maß für Maß: Eine Vergleichende Profilanalyse von Diskursen beim Dolmetschen. trans-kom, 4(2), 161-175. http://www.trans-kom.eu/bd04nr02/transkom_04_02_03_Kalina_Profilanalyse.20111205.pdf.

Kautz, U. (2000). Handbuch Didaktik des Übersetzens und Dolmetschens. Iudicium.

Liontou, K. (2011). Strategies in German-to-Greek Simultaneous Interpreting: A Corpus-Based Approach. Gramma, 19, 37-56. <https://www.enl.auth.gr/gramma/gramma11/Liontou.pdf>.

Setton, R., & Dawrant, A. (2016). Conference Interpreting: A Trainer's Guide. John Benjamins.