



---

# Proceedings of Machine Translation Summit XVIII

<https://mtsummit2021.amtaweb.org>

---

## 1st Workshop on Automatic Spoken Language Translation in Real-World Settings

Organizers:  
Claudio Fantinuoli and Marco Turchi





### **Bio**

Silvia Hansen-Schirra is Professor for English Linguistics and Translation Studies and Director of the Translation Cognition (TraCo) Center at Johannes Gutenberg University Mainz in Gernersheim. She is the co-editor of the book series "Translation and Multilingual Natural Language Processing" and "Easy – Plain – Accessible". Her research interests include machine translation, accessible communication and translation process research.

### **2.3 Juan Pino, Facebook AI**

#### **End-to-end Speech Translation at Facebook**

End-to-end speech translation, the task of directly modeling translation from audio in one language to text or speech in another language, presents advantages such as lower inference latency but faces a data scarcity challenge, including for high resource languages. In this talk, various data and modeling solutions are presented in order to overcome this challenge. Similar to the textual machine translation case, multilingual speech translation provides maintainability and quality improvements for lower resource language pairs. We present our initial efforts on this topic. As simultaneous speech translation is a prerequisite for practical applications such as simultaneous interpretation, we also give an overview of our investigations into end-to-end simultaneous speech translation. Finally, we describe initial work on speech translation modeling for speech output.

**Bio** Juan Pino is a Research Scientist at Facebook, currently working on speech translation. He received his PhD in machine translation from the University of Cambridge under the supervision of Prof. Bill Byrne.

### **2.4 Prof. Bart Defrancq, Ghent University**

#### **Will it take another 19 years? Cognitive Ergonomics of Computer-Assisted Interpreting (CAI)**

In 1926 the first experiments were held where interpreters were required to interpret diplomatic speeches (semi)- simultaneously. Different experimental setups were put to the test to study interpreters' performances and simultaneous interpreting was successfully carried on from 1928 on in different diplomatic contexts (Baigorri-Jalón 2014). However, the real breakthrough only came in 1945 with the Nüremberg trials, where simultaneous interpreting was offered for weeks in a row and served as a model for the organisation of standing diplomatic conferences. Recent years have seen the development of the first usable CAI-tools for simultaneous interpreters, based on automatic speech recognition (ASR) technologies. These tools provide interpreters not with full transcripts of speeches but rather with lists of specific target items that pose problems, such as numbers, terms and named entities. Full transcripts are of little use for simultaneous interpreters as they are working with extremely narrow time frames with regard to the source text and combine several cognitive, language-related tasks. Adding the (language-related) task of consulting a running transcript of the source speech would probably over-burden cognitive processing in interpreters. Experiments with simulated ASR and ASR prototypes have shown that the provision of targeted information improves interpreters' performances on the accuracy dimension with regard to the rendition of the target items (Desmet et al. 2018, Fantinuoli Defrancq 2021). The first analyses of cognitive load associated with consulting ASR while interpreting suggest that no additional cognitive load is involved with the use of the prototype ASR. However, all aforementioned studies were conducted in quasi-experimental settings, with carefully presented speeches by native and near-native speakers, in physical interpreting booths and using prototypes whose features are based on intuition rather than on ergonomic analysis. There is a real risk that in the absence of systematic ergonomic analysis, CAI-tools will face the same fate as simultaneous interpreting technology. In my contribution I will apply Cañas' (2008) principles of cognitive ergonomics to the integration of ASR in interpreting booths or

remote simultaneous interpreting (RSI) platforms. According to Cañas, successful integration of software in the human workflow relies on 4 requirements: it should (1) shorten the time to accomplish interaction tasks; (2) reduce the number of mistakes made by humans; (3) reduce learning time; and (4) improve people's satisfaction with a system. Cognitive ergonomics seeks improvement in those areas to make the execution of the overall task assigned to what is called the "Joint Cognitive System", i.e. the joint processing by humans and devices involved in that task (Woods Hollnager 2006), more successful. I will argue that although the first research results based on data from physical booths are encouraging, the integration of ASR in the interpreters' workflow on RSI platforms will face particular challenges.

#### References

Baigorri-Jalón, J. (2014). From Paris to Nuremberg. The Birth of Conference Interpreting. Amsterdam: Benjamins. Cañas, J. (2008). Cognitive Ergonomics in Interface Development Evaluation. *Journal of Universal Computer Science*, 14 (16): 2630-2649.

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.

Desmet, B., Vandierendonck, M., Defrancq, B. (2018). Simultaneous interpretation of numbers and the impact of technological support. In *Interpreting and technology* (pp. 13–27). Language Science Press.

Woods, D. Hollnager, E. (2006). *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering*. Boca Raton: CRC Press.

#### Bio

Born in 1970, studied Romance Philology at Ghent University (1987-1991) and was granted a PhD in Linguistics at the same University in 2002. Worked at the College of Europe as a French lecturer from 1992 until 1995, as a researcher at Ghent University from 1995 until 2007, as a visiting professor at the Université Catholique de Louvain-la-Neuve from 2004 until 2009 and as a postdoctoral researcher at Hogeschool from 2007 until 2010. Trained as a conference interpreter in 2010 and was appointed as an assistant professor of interpreting and translation the same year. Has been head of interpreter training both at the masters' and at the postgraduate levels since 2010, both at Hogeschool Gent and University Ghent (since 2013, when the department was moved from the Hogeschool to the University in the framework of an institutional reform). Is a member of the Department Board, the Faculty Board, the Research Commission of the alpha-Faculties, the Doctoral School Board and of the CIUTI Board.

### 3 Scientific committee

- Nguyen Bach Alibaba US
- Laurent Besacier University of Grenoble
- Dragos Ciobanu University of Vienna
- Jorge Civera Universitat Politècnica de València
- Marta R. Costa-jussà Universitat Politècnica de Catalunya
- Bart Defrancq Ghent University
- Marco Gaido Fondazione Bruno Kessler
- Hirofumi Inaguma University of Kyoto
- Alfons Juan Universitat Politècnica de València

- Alina Karakanta Fondazione Bruno Kessler
- Evgeny Matusov AppTek
- Jan Niehues University of Maastricht
- Sara Papi Fondazione Bruno Kessler
- Franz Pöchhacker University of Vienna
- Bianca Prandi Johannes Gutenberg-Universität Mainz
- Pablo Romero-Fresco Universidade de Vigo
- Juan Pino Facebook
- Claudio Russello UNINT - Rome
- Matthias Sperber Apple
- Sebastian Stueker Karlsruhe Institute of Technology S
- hinji Watanabe Johns Hopkins University

# Contents

- 1 Seed Words Based Data Selection for Language  
Roberto Gretter, Marco Matassoni and Daniele Falavigna
- 13 Post-Editing Job Profiles for Subtitlers  
Anke Tardel, Silvia Hansen-Schirra and Jean Nitzke
- 23 Operating a Complex SLT System with Speakers and Human Interpreters  
Ondřej Bojar, Vojtěch Srdečný, Rishu Kumar, Otakar Smrž, Felix Schneider, Barry Haddow, Phil Williams and Chiara Canton
- 35 Simultaneous Speech Translation for Live Subtitling: from Delay to Display  
Alina Karakanta, Sara Papi, Matteo Negri and Marco Turchi
- 49 Technology-Augmented Multilingual Communication Models: New Interaction Paradigms, Shifts in the Language Services Industry, and Implications for Training Programs  
Francesco Saina