

# Speech Recognition in the Interpreter Workstation

**Claudio Fantinuoli**

Johannes Gutenberg Universität Mainz/Germersheim

fantinuoli@uni-mainz.de

## Abstract

In recent years, computer-assisted interpreting (CAI) programs have been used by professional interpreters to prepare for assignments, to organize terminological data, and to share event-related information with colleagues. One of the key features of such tools is the ability to support users in accessing terminology during simultaneous interpretation. The main drawback is that the database is queried manually, adding an additional cognitive effort to the interpreting process. This disadvantage could be addressed by automating the querying system through the use of Automatic Speech Recognition (ASR), as recent advances in Artificial Intelligence have considerably increased the quality of this technology. In order to be successfully integrated in an interpreter workstation, however, both ASR and CAI tools must fulfil a series of specific requirements. For example, ASR must be truly speaker-independent, have a short reaction time, and be accurate in the recognition of specialized vocabulary. On the other hand, CAI tools face some challenges regarding current implementations, and need to support the handling of morphological variants and to offer new ways to present the extracted data. In this paper we define and analyse a framework for ASR-CAI integration, present a prototype and discuss prospective developments.

## 1 Introduction

In recent years, computer-assisted interpreting (CAI) programs have been used by professional interpreters to prepare for assignments, to organize terminological data, and to share event-related information among colleagues (Corpas Pastor and May Fern, 2016; Fantinuoli, 2016, 2017a). One of the main features of such tools is the ability to support users in accessing multilingual terminology during simultaneous interpretation (SI). With state-of-the-art CAI tools, interpreters need to manually input a term, or part of one, in order to query the database and retrieve useful information. This manual lookup mechanism is considered the primary drawback of this approach, as it appears time-consuming and distracting to search for terminological data while interpreters are performing an activity that requires concentration and rapid information processing. Although initial empirical studies on the use of CAI tools seem to support the idea that interpreters in the booth may have the time and the cognitive ability to manually look up specialized terms (Prandi, 2015; Biagini, 2016), an automated querying system would undoubtedly represent a step forward in reducing the additional cognitive effort needed to perform this human-machine interaction. With this in mind, it is reasonable to assume that a CAI tool equipped with an automatic lookup system may have the potential to improve the interpreters' performance during the simultaneous interpretation of specialized texts.

Automatic speech recognition (ASR) has been proposed as a form of technology to automate the querying system of CAI tools (Hansen-Schirra, 2012; Fantinuoli, 2016). In the past, the difficulty of building ASR systems accurate enough to be useful outside of a carefully controlled environment hindered its deployment in the interpreting setting. However, recent advances in Artificial Intelligence, especially since the dissemination of deep learning and neural networks, have considerably increased the quality of ASR (Yu and Deng, 2015). With systems that achieve a 5.5 percent word error rate<sup>1</sup>, the deployment of ASR in the context of

<sup>1</sup><https://www.ibm.com/blogs/watson/2017/03/reaching-new-records-in-speech-recognition> [last access 28

interpretation appears conceivable nowadays. Some scholars regard ASR as a technology “with considerable potential for changing the way interpreting is practiced” (Pöchhacker, 2016, p. 188). For example, it has the potential to dramatically change the way consecutive interpreting is usually performed (through note-taking with pen and paper) and may outcome alternative technology-based methods recently proposed, such as the digital pen (Orlando, 2014). With ASR, the consecutive interpreter may use the transcription of the spoken word to sight-translate the speech segment, with obvious advantages in terms of precision and completeness. In the context of simultaneous interpreting, ASR can be used not only to query the interpreter’s glossary, as pointed out above, but also to implement innovative features that aim at facilitating the processing of typical “problem triggers” in interpretation, such as numbers, acronyms and proper names. In order to be successfully integrated in an interpreter workstation, however, both ASR and CAI tools must fulfill a series of specific requirements. For example, ASR must be truly speaker-independent, have a short reaction time, and be accurate in the recognition of specialized vocabulary. On the other hand, CAI tools need to overcome some challenges of current implementations. For instance, they must be able to handle morphological variants and offer ergonomic ways to present extracted information.

This paper is organized as follows: Section 2 describes computer-assisted interpreting tools and the unique features and limitations of their use in the booth as a terminology lookup system; Section 3 gives an overview of the potential shortcomings of ASR systems that may arise from their integration into an interpreter workstation and summarizes the requirements that ASR systems and CAI tools need to meet for successful integration; Sections 4 and 5 present a prototype of ASR-CAI integration and the results of an empirical test on the ability of the tool to identify relevant information from three English specialized speeches; finally, Section 6 summarises the topics introduced in this paper and presents some future perspectives.

## **2 Computer-assisted interpreting tools**

Technology is growing as an important aspect of the interpreting profession. There is general consensus that some of the recent advances in information and communication technology have facilitated some aspects of the profession (Tripepi Winteringham, 2010; Fantinuoli, 2016, 2017a). Suffice it to say how easy it is today to find domain-related texts on a large variety of subjects and languages and to consult the plethora of terminological resources available on the Web. Advance preparation is considered one of the most important activities to ensure quality, especially in the interpretation of highly specialized domains (Kalina, 2005; Gile, 2009), and the use of correct and precise terminology can facilitate communication and increase the perceived professionalism of interpreters (Xu, 2015). Hence, it is not surprising that the introduction of technological advances is favoured by the interpreting community (Fantinuoli, 2017b) considering the evident improvement of preparation and assignment management.

Among the different kinds of technology used by interpreters, computer-assisted interpreting tools have emerged as the most distinctive development in recent years. CAI tools are computer programs designed to support interpreters during different phases of an assignment, from the preparation stage to accessing information in the booth. In the last decade, various CAI tools have been designed and used by practitioners with the goal of rationalizing and optimizing some steps of the interpreting workflow<sup>2</sup>. CAI tools generally focus on the lexical and terminological aspect of the profession. They aim at supporting the user in acquiring and managing linguistic information, creating multilingual glossaries, and accessing them during the preparation stage.

---

September 2017]

<sup>2</sup>For a classification of CAI tools, see Fantinuoli (2017b); for a tentative evaluation of available terminology solutions for interpreters, see Will (2015).

This is true particularly when interpreters learn and memorize event-related terminology or when they follow up on the completed terminology work.

CAI tools have also been proposed as a means to access target language equivalents (specialized terminology) in the booth whenever interpreters are not able to retrieve them from their long-term memory, and alternative strategies, such as the use of paraphrasing, hypernyms, etc. are not possible or, if used, would lead to a loss of quality or compromise the complete and accurate rendition of the original. While working in the booth, however, the idea of being supported by a computer program has been perceived by practitioners with mixed feelings. Some seem to be enthusiastic and appreciate the possibility of accessing subject-related translations in real time, while others are reluctant and consider it unnatural (cf. Tripepi Winteringham, 2010; Berber-Irabiien, 2010; Corpas Pastor and May Fern, 2016).

Although first empirical experiments suggest an improvement of terminological rendition in highly specialized conferences if a CAI tool is used (cf. Biagini, 2016; Prandi, 2015), there are objective constraints in interpretation that make the use of such tools in the booth less straightforward than during preparation or follow-up work. In the case of simultaneous interpretation, such constraints are primarily related to the time pressure and the cognitive load involved in this activity. Since interpreters often work on the edge of saturation (Gile, 2009), as many concurring activities are taking place at the same time, including listening, comprehension, translation, text production and monitoring, the use of a tool for terminology search adds further cognitive load to an already precarious balance. For this reason, the interpreter controlling even a carefully designed lookup solution (i.e. inputting a term, searching for the most adequate result, etc.) may experience a cognitive overload with following deterioration of the quality of interpretation.

There is no doubt that the limitations of state-of-the-art term search mechanisms adopted by CAI tools can benefit from recent advances in artificial intelligence. One of the most promising developments has been indicated in the integration of automatic speech recognition. Automating the lookup mechanism by means of ASR can not only reduce the additional cognitive effort needed to perform human-machine interaction for terminology lookup, but the integration of ASR can also allow the implementation of other innovative features, such as automatic transcription of numbers, abbreviations, acronyms, and proper names. Since these linguistic forms are generally considered to be potential problems for interpreters because of heavy processing costs on cognitive resources<sup>3</sup> – with severe errors and disfluencies as a consequence (Gile, 2009, cf.) – being prompted with a transcription of this information may alleviate the work load during simultaneous interpretation.

In light of the preceding considerations, it is reasonable to suggest that the integration of ASR and algorithms to identify specialized terms as well as numbers, proper names and abbreviations in a transcribed speech would contribute to further increase the usability of CAI tools, leading to an improvement in the terminological rendition and in the overall performance of interpreters during the simultaneous interpretation of specialized texts. A CAI tool with ASR integration could act like an electronic boothmate, providing useful information to the colleague whenever necessary. Since the cooperation between boothmates (writing down numbers, terms, etc.) is generally seen as positive among interpreters (Setton and Dawrant, 2016) and – when silent and discrete – not considered a source of distraction, this development may lead to an increase in the acceptance of CAI tools in the booth.

---

<sup>3</sup>According to the “effort model”, names and numbers tend to increase the effort of the interpreter and may lead to cognitive saturation.

### 3 Speech Recognition and CAI integration

Speech recognition or automatic speech recognition (ASR) is the process of converting human speech signals to a sequence of words by means of a computer program (Jurafsky and Martin, 2009). ASR has been around for more than three decades and has been used in many areas, such as human-machine interface or for dictation purposes, but only recently has there been a renewed interest for this technology. There are several reasons for this. On the one hand, new computational approaches, especially Neural Networks and Deep Learning, have significantly improved the quality of ASR systems. On the other, the commercial interest for ASR nowadays is on its verge, with global players such as Microsoft, Amazon and Apple investing significant funding and research in improving their commercial products Cortana, Alexa and Siri, just to name a few. Such improvements are expected to continue in the years to come.

Yet, ASR is far from perfect. Language is a complex system and language comprehension consists of more than simply listening and decoding sounds. Humans use acoustic signals together with background information, such as information about the speaker, world knowledge, subject knowledge, as well as grammatical structures, redundancies in speech, etc. to predict and complete what has been said. All of these features are difficult to model in a computer program. As a consequence, the problems that ASR systems have been pressed to solve are many. In connection with the integration in an interpreter's workstation, the following issues for ASR can be identified:

- **Use of spoken language** - Speakers may use a variety of styles (e.g. careful vs. casual speech). In formal contexts, such as conference venues, political meetings, speakers use spontaneous language, read aloud prepared texts, or use a mixture of both. The correct transcription of casual speech represents a big challenge for ASR. Especially in spontaneous speech, humans make performance errors while speaking, i.e. disfluences such as hesitations, repetitions, changes of subject in the middle of an utterance, mispronunciations, etc. The presence of such elements of spoken language poses a serious problem for ASR and generally leads to poor system performance.
- **Speaker variability** - Speakers have different voices due to their unique physical features and personality. Characteristics like rendering, speaking style, and speaker gender influence the speech signal and consequently require great adaptation capabilities by the ASR. Regional and social dialects are problematic for speaker-independent ASR systems. They represent an important aspect in the interpreting setting both for widely spoken and less spoken languages considering the variability of pronunciation is vast. Furthermore, in the context of English as *lingua franca* ASR should be able to cope with both native and foreign accents as well as mispronunciations.
- **Ambiguity** - Natural language has an inherent ambiguity, i.e. it is not easy to decide which of a set of words is actually intended. Typical examples are homophones, such as “cite” vs. “sight” vs. “site” or word boundary ambiguity, such as “nitrate” vs. “night-rate”.
- **Continuous speech** - One of the main problems of ASR is the recognition of word boundaries. Besides the problem of word boundary ambiguity, speech has no natural pauses between words, as pauses mainly appear on a syntactic level. This may compromise the quality of a database querying mechanism, as this relies on the correct identification of word units.
- **Background noise** - A speech is typically uttered in an environment with the presence of other sounds, such as a video projector humming or other human speakers in the

background. This is unwanted information in the speech signal and needs to be identified and filtered out. In the context of simultaneous interpretation, the restrictive standards for the audio signal in the booth<sup>4</sup> offer the best setting for good quality transcription. In other settings, however, such as face-to-face meetings, noise is expected to pose a problem for the quality of the ASR output.

- **Speed of speech** Speeches can be uttered at different paces, from slow to very high. This represents a problem both for human interpreters, as they need sufficient time to correctly process the information, and for ASR. One reason is that speakers may articulate words poorly when speaking fast.
- **Body language** - Human speakers do not only communicate with speech, but also with non verbal signals, such as posture, hand gestures, and facial expressions. This information is completely absent with standard ASR system and could only be taken into consideration by more complex, multimodal systems. However, for the integration of ASR in CAI tools, this shortcoming does not seem to play an important role, as the ultimate goal is to trigger a database search for terminology units, and not to semantically “complete” the oral message uttered by the speaker.

There are different applications for speech recognition depending on the constraints that need to be addressed, i.e. the type of utterances that can be recognized. ASR solutions are typically divided into systems that recognize *isolated words*, where single words are preceded and followed by a pause (e.g. to command digital devices in Human-Machine interface), and systems that recognize *continuous speech*, where utterances are pronounced naturally and the tool has to recognize word boundaries. These two basic classes can be further divided, on the basis of vocabulary size, spontaneity of speech, etc. Integration of CAI with ASR is a special case of human-computer interaction and automatic transcription of speech. The whole talk needs to be transcribed for the CAI tool to select pertinent chunks of text to start the database query algorithm and to identify entities, such as numerals and proper names.

To be used with a CAI tool, an ASR system needs to satisfy the following criteria at minimum:

- be speaker-independent
- be able to manage continuous speech
- support large-vocabulary recognition
- support vocabulary customisation for the recognition of specialized terms
- have high performance accuracy, i.e. a low word error rate (WER)
- be high speed, i.e have a low real-time factor (RTF)<sup>5</sup>

ASR systems can be both stand-alone applications installed on the interpreter’s computer, such as Dragon Naturally Speaking<sup>6</sup> or cloud services, such as the Bing Speech API<sup>7</sup>. For privacy reasons, it seems more advisable to prefer stand-alone ASR systems for integration into CAI tools, as they do not require the user to send (confidential) audio signals to an external service provider.

---

<sup>4</sup>See for example the norm ISO 20109, Simultaneous interpreting — Equipment — Requirements.

<sup>5</sup>RFT is the metric that measures the speed of an automatic speech recognition system.

<sup>6</sup><https://www.nuance.com/dragon.html>

<sup>7</sup><https://azure.microsoft.com/en-us/services/cognitive-services/speech/>

As for CAI tools, in order to successfully support the integration of a ASR system, the tool needs to satisfy the following requirements:

- high precision, precision being the fraction of relevant instances among the retrieved instances
- high recall, recall being the fraction of relevant instances that have been retrieved over the total amount of relevant instances present in the speech
- if a priority has to be set, precision has priority over recall, in order to avoid producing results that are not useful and may distract the interpreter
- deal with morphological variations between transcription and database entries without increasing the number of results
- have a simple and distraction-free graphical user interface to present the results

In the next sections the implemented prototype will be briefly presented together with the results of an experimental test designed to test the quality of the CAI implementation.

#### 4 Prototype

The prototype described in this study was designed and implemented within the framework of InterpretBank<sup>8</sup>, a CAI tool developed to create assignment-related glossaries accessible in a booth-friendly way (Fantinuoli, 2016). The tool reads the transcription provided by an ASR system and automatically provides the interpreter with the following set of information:

- entries from the terminology database
- numerals

The tool has been designed with an open interface between the CAI tool and the ASR system of choice, provided the ASR system meets the features described in the previous section. The specially designed open structure allows users to choose the ASR engine with the best quality output for the source language, domain, and operative system without having to change or adapt the CAI interface. Since the tool is based mostly on language-independent algorithms, for example to deal with morphological variants (database query) and to identify numbers and acronyms, the prototype supports the integration of ASR for any input language.

The acoustic input signal required by the system is the same that interpreters receive in their headset. Since most standard booth consoles have more than one audio output for headphones<sup>9</sup>, one of these can be connected to the audio input of the computer equipped with the ASR-CAI tool. If a second audio output is not available, a headphone splitter can be used to provide an audio signal both to the interpreter's headphones and to the computer audio card.

The working procedure can be divided into two main phases: the tool first reads the provided transcription and pre-processes the text. It then queries the terminological database and identifies the entities from the text flow, visualizing the results in an interpreter-tailored graphical user interface. The algorithms are triggered any time a

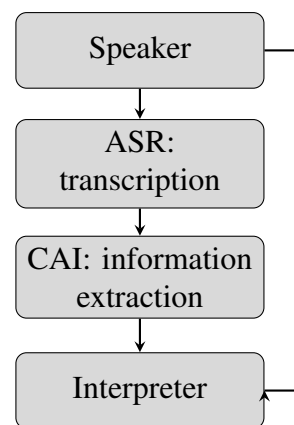


Figure 1: Workflow

<sup>8</sup>[www.interpretbank.com](http://www.interpretbank.com)

<sup>9</sup>Like the Sennheiser or the BOSCH interpreter console

new piece of text is automatically provided by the ASR. For this reason, the tool needs the transcription to be provided by the ASR system in chunks of text. The chunks are the final version of the portion of text that has been transcribed (in contrast to a “live” temporary version that will be changed by the ASR system during the further elaboration of the acoustic signal). Whenever a chunk of text is provided by the ASR system, the text is normalized. A series of rules have been written to take into account the different ways SR systems may enrich the transcribed text, for example by adding capitalization and punctuation, or converting numerals.

The provided normalized text is tokenized. The tokenization aims at distinguishing sections of a string of characters and producing a list of words (tokens) contained in the transcribed text. The tokens are used to identify numerals and as a query string to match entries in the terminological database. The approach used is based on n-gram matching. From the tokenized input, n-grams are created and matched against one-word and multi-word units previously saved in the database. The algorithm needs to match also n-grams that may appear in different forms between the glossary and the transcription (for example plurals: fuel vs. fuels). In order to do so, the prototype implements a fuzzy match approach which should produce good results with most European languages. This approach is based on string metrics for measuring the difference between two sequences. The tool uses the Levenshtein distance (the distance between two words being the minimum number of single-character edits, such as insertions, deletions or substitutions, required to exchange one word with another) and computes a percentage of variation. N-grams are matched if they differ less than a given percentage. In order to reduce the number of potential results retrieved by the tool (which is a prerequisite for its usability), a series of heuristics are applied that aim at identifying the most probable term given the various results. There are limitations with this approach, for example in its use with agglutinative languages. In the future, to extend the querying function to such languages, other language-dependent matching approaches to term recognition should be analysed, such as stemming or inflection analysis (Porter, 2001).

In order to take into account the specific constraints of interpreting, not only does the tool need to achieve a high precision and recall, but it also needs to minimize the visual impact of the extracted data. For this reason, the interface is kept as clean as possible. Information is divided into three sections, one for the transcription, one for terminology, and one for numerals. The visualisation appears in chronological order. Among other things, the user is able to set background color, font size and color and to influence some extra parameters, such as the possibility to suppress the repetition of the same terms, etc. Terminology and entity data are visualized essentially in real time. The time span between the moment an utterance has been said and the data visualization depends on the speed of the ASR system and its latency.

## 5 Evaluation

The overall quality of a CAI system with an integrated ASR engine depends on two factors: the quality of the transcription provided by the ASR system (low word error rate) and the ability of the CAI tool to retrieve and identify useful information. For the purpose of this paper, the integrated system<sup>10</sup> has been empirically evaluated by measuring the precision and recall scores for the identification of terminology and numerals.

The test has been conducted using three speeches in English which are rich in terminological units. The speeches are the same used by Prandi (2017) in a pilot study designed to empirically test the use of CAI tools during simultaneous interpretation. All three texts are on the subject of renewable energy. The bilingual glossary used to test the terminology retrieval quality

---

<sup>10</sup>CAI: InterpretBank 4; ASR: Dragon Naturally Speaking 13.

comprises 421 entries and has the size of a typical glossary compiled by interpreters. For this experiment, the terminological units under investigation are defined as the one- and multi-word terms that are present both in the speeches and in the glossary. The system will be tested on this set of terms (119) as well as on the numerals (11) contained in the texts. The latency of the ASR system was not object of testing. Table 1 reports the metrics of the texts.

	Tokens	Terms	Numerals
Text 1	1533	39	7
Text 2	1513	40	2
Text 3	1512	40	2
Total	4558	119	11

Table 1: Text metrics

Ideally, the system should reach a high *recall*. This would mean that it is able to recognize all terminological units of the transcribed speech, independent of the presence of orthographical differences. It should also have high *precision*, i.e. present a low number of undesired or erroneous results. This ensures that interpreters are not prompted with superfluous results which may cause distraction.

Table 2 summarizes the ASR performance on the set of stimuli defined above. This result is obtained after importing the list of English specialized terms contained in the glossary. With a word error rate (WER) of 5.04% on the terminology list, the ASR system performs well in recognizing the terminological units. It is worth mentioning that importing the list of specialized words from the glossary contributed to decrease the WER from the initial value of 10.92%. The transcription of numerals was completed without errors.

	Terms	Numerals
Text 1	38 (of 39)	7 (of 7)
Text 2	39 (of 40)	2 (of 2)
Text 3	36 (of 40)	2 (of 2)
Total	113 (of 119)	11 (of 11)

Table 2: Correctly transcribed terms and numerals

Table 3 summarizes the results of the terminology retrieving algorithms on the transcription delivered by the ASR engine. The system was able to retrieve and visualize 112 terminological units out of the 119 contained in the texts, which corresponds to 94.11%, while the number of terms erroneously retrieved was 3. With an F1 score<sup>11</sup> of 0.97, the overall quality of the identified terminology seems to be satisfying. Among the missing terms, there are complex plural forms (nucleus vs. nuclei) and quasi-synonyms (“coal-fired plants” and “coal-fired power plants”). Among the erroneously retrieved terms there are phrases such as *save energy* that was matched against the terminological unit *wave energy*. It is worth noting that the fuzzy searching algorithm implemented in the CAI tool was able to “correct” terms wrongly transcribed by the ASR system, such as *malting* which was transcribed as *moulding*, and was able to identify and visualize the correct term.

<sup>11</sup>F1 score considers the precision  $p$  and the recall  $r$  of the test to compute the score, being  $p$  the number of correct positive results divided by the number of all positive results and  $r$  the number of correct positive results divided by the number of positive results that should have been returned. An F1 score reaches its best value at 1 and worst at 0.



	Visualized	Recognized	Missing	Errors
Text 1	38	37 (of 39)	2 (of 39)	1
Text 2	40	39 (of 40)	1 (of 40)	1
Text 3	37	36 (of 40)	4 (of 40)	1
Total	115	112 (of 119)	7 (of 119)	3

Table 3: Performance of CAI terminology retrieval

The identification of numerals does not represent a problem for the ASR system and the CAI retrieving algorithm. The system reaches an F1 score of 1, meaning no number has been left out and no wrong numbers have been retrieved and presented to the user.

## 6 Conclusions

In this paper, we have proposed the integration of automatic speech recognition in computer-assisted interpreting tools as a means to improve their lookup mechanism. A prototype of ASR-CAI integration has been presented and its output tested in terms of precision and recall of terminology retrieval and numbers identification. Although available ASR engines are still not perfect and fail under certain circumstances (non native accents, unknown words, etc.), they already reach high precision values in standard conditions, even within specialized domains. The ASR-CAI integration tested in our experimental setting reaches an F1 value of 0.97 for terminology and 1 for numerals. This value is quite promising and seems to suggest that the use of this technology is – at least in “standard” interpreting settings – already possible. In the future, with the expected increase of ASR quality, the proposed technology may be good enough to be also used in more difficult settings, with mispronunciations, background noise, etc.

The proposed technology has the potential to change the way interpreting will be performed in the future. However, further investigation would be necessary to evaluate its impact on the interpreting process and product. For example, it has to reveal whether the interpreter may experience a visual (and cognitive) overload when working with ASR-CAI tools or if their use may lead to the expected quality increase in the interpretation of specialized texts.

## References

- Berber-Irabien, Diana-Cristina. 2010. *Information and Communication Technologies in Conference Interpreting*. Lambert Academic Publishing.
- Biagini, Giulio. 2016. *Printed glossary and electronic glossary in simultaneous interpretation: a comparative study*. Master’s thesis, Università degli studi di Trieste.
- Corpas Pastor, Gloria and Lily May Fern. 2016. A survey of interpreters’ needs and their practices related to language technology. Technical report, Universidad de Málaga.
- Fantinuoli, Claudio. 2016. InterpretBank. Redefining computer-assisted interpreting tools. In *Proceedings of the Translating and the Computer 38 Conference*. Editions Tradulex, London, pages 42–52.
- Fantinuoli, Claudio. 2017a. Computer-assisted interpreting: challenges and future perspectives. In Isabel Durán Muñoz and Gloria Corpas Pastor, editors, *Trends in e-tools and resources for translators and interpreters*, Brill, Leiden.
- Fantinuoli, Claudio. 2017b. Computer-assisted preparation in conference interpreting. *Translation & Interpreting* 9(2).
- Gile, Daniel. 2009. *Basic Concepts and Models for Interpreter and Translator Training: Revised edition*. John Benjamins Publishing Company, Amsterdam, 2nd edition.
- Hansen-Schirra, Silvia. 2012. Nutzbarkeit von Sprachtechnologien für die Translation. *trans-kom* 5(2):211–226.

- Jurafsky, Dan and James H. Martin. 2009. *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall series in artificial intelligence. Pearson Prentice Hall, Upper Saddle River, N.J, 2nd edition.
- Kalina, Sylvia. 2005. Quality Assurance for Interpreting Processes. *Meta: Journal des traducteurs* 50(2):768.
- Orlando, Marc. 2014. A study on the amenability of digital pen technology in a hybrid mode of interpreting: Consec-simul with notes. *Translation & Interpreting* 6(2).
- Prandi, Bianca. 2015. The use of CAI tools in interpreters' training: a pilot study. In *Proceedings of the 37 conference Translating and the Computer*. Editions Tradulex, London, pages 48–57.
- Prandi, Bianca. 2017. Designing a multimethod study on the use of cai tools during simultaneous interpreting. In *Translating and the Computer* 39. London.
- Pöchhacker, Franz. 2016. *Introducing Interpreting Studies*. Routledge, 2nd edition.
- Setton, Robin and Andrew Dawrant. 2016. *Conference interpreting: a complete course*. Number volume 120 in Benjamins translation library (BTL). John Benjamins Publishing Company, Amsterdam & Philadelphia.
- Tripepi Winteringham, Sarah. 2010. The usefulness of ICTs in interpreting practice. *The Interpreters' Newsletter* 15:87–99.
- Will, Martin. 2015. Zur Eignung simultanfähiger Terminologiesysteme für das Konferenzdolmetschen. *trans-kom* 8(1):179–201.
- Xu, Ran. 2015. *Terminology Preparation for Simultaneous Interpreters*. PhD thesis, University of Leeds.
- Yu, Dong and Li Deng. 2015. *Automatic speech recognition: a deep learning approach*. Springer, London.