Исследования языка и современное гуманитарное знание, 2021, т. 3, № 2
Language Studies and Modern Humanities, 2021, vol. 3, no. 2
www.languagestudies.ru

Check for updates

Перевод и переводоведение

UDC 81’25

https://www.doi.org/10.33910/2686-830X-2021-3-2-174-187

Speech translation vs. Interpreting

I. Horváth [✉1]

[1] Eötvös Loránd University, 1–3 Egyetem tér, Budapest H-1053, Hungary

Author
Ildikó Horváth,
e-mail: horvath.ildiko@btk.elte.hu

For citation:
Horváth, I.
(2021) Speech translation vs. Interpreting. *Language Studies and Modern Humanities*, vol. 3, no. 2, pp. 174–187. https://www.doi.org/10.33910/2686-830X-2021-3-2-174-187

Received 18 May 2021; reviewed 12 July 2021; accepted 12 July 2021.

Abstract. Artificial intelligence (AI), deep learning technologies and big data have impacted on the interpretation market and AI-based technologies can be used in automated speech translation. The first experiments to create an automatic interpreter took place at the end of the 1980s and early 1990s. Today, there are several AI-based devices available on the market which attempt to fully automatize the interpreting process, both in the consecutive and in the simultaneous mode in a limited number of specific communication situations. This article first reviews the history and mechanism of automated interpreting and provides a comparison of human and automated interpreting. It also presents the main features and use cases of automated speech translation (AST). By showing that the two activities are intrinsically different, it argues that they need to be distinguished more clearly by defining the speech-to-speech (S2S) language transfer accomplished by computers as automated speech translation (AST) and keeping the term ‘interpreting’ for the human activity. Automated speech translation has an undeniable role and place in today’s world, steeped in technology and AI. However, it needs to be underlined that it is completely different from the complex interpreting service human interpreters provide and the circumstances and contexts in which its use can be advised is intrinsically different from that of human interpreting. Therefore, the real question is how AST and human interpreting can complement each other, in other words, what are the situations and contexts where AST is desired and applicable and when is there a need for human interpreting?

Keywords: automatic speech translation, interpreting, artificial intelligence, speech recognitions, MT.

Technological development gained significant impetus in the mid-2010s. It has not only accelerated but its performance has become more sophisticated. “Automated interpretation” has become a common topic at technological and scientific conferences as well as interpreting studies events and it is also a subject frequently discussed by interpreters. Moreover, it has become more visible in the press as well. Today, the expression “AI interpreter” is used more frequently, even though machine interpretation is still lagging behind machine translation and AI-based technology has not led to a breakthrough in the automatization of the interpreting process.

There are several expressions used by researchers, professionals and developers for the automated interpreting process such as machine interpreting, automated interpreting, translation of speech or speech translation. This article reviews the history and mechanism of automated interpreting and provides a comparison of human and automated interpreting. By showing that the two activities are intrinsically different, it argues that they need to be distinguished more clearly by defining the speech-to-speech (S2S) language transfer accomplished by computers as automated speech translation (AST) and keeping the term “interpreting” for the human activity. Automated speech translation has an undeniable role and place in today’s world, steeped in technology and AI. However, it needs to be underlined that it is completely different from the complex interpreting service human interpreters provide and the circumstances and contexts in which its use can be advised is intrinsically different from that of human interpreting.

The brief history of automated speech translation

For the past 60 years, one of the aims of information technology developments has consisted of the automatization of human speech comprehension and translation. The first experiments to automatize interpreting took place at the end of the 1980s and early 1990s. The concept of speech translation was first presented at the ITU Telekom World (Telecom '83) by NEC Corporation in 1983. Then in 1986 the Advanced Telecommunications Research Institute International (ATR) was set up with the aim of carrying out basic research in the field of speech translation (Nakamura 2009). Another early speech translation system called JANUS dates back to 1991, and it rendered speech from English into German and Japanese (McNair, Waibel, Jain et al. 1991). However, language technology available at that time allowed for only a very basic and limited performance of speech translation tools. Attempts to develop a 'machine interpreter' gained new momentum in the 2010s, when several types of translation software became available on the market, both in the consecutive and in the simultaneous mode (see Figure 1). The recent technological impact on interpreting has been so significant that, as Fantinuoli puts it, the technological change we are witnessing in our profession is irreversible, and interpreting is going through a "technological turn" (Fantinuoli 2018). This turn will bring radical changes in terms of the ecosystem of interpreting concerning the socio-economic status and prestige of the interpreter, the cognitive processes during interpreting, as well as working environments.

A common feature of speech translation tools is the fact that they have been developed for a limited number of specific communication situations. They are used to speech translate the most frequent phrases, questions between different languages in well-defined contexts such as travel, humanitarian missions, medical care, university lectures, wars, and also where human interpreters are not available.

How speech translation devices function

At present, there are two types of speech translation models: (1) the cascade models, and (2) the end-to-end models. The speech translation systems in the first group using the cascade models are composed of several modules (Figure 2) performing the following operations:

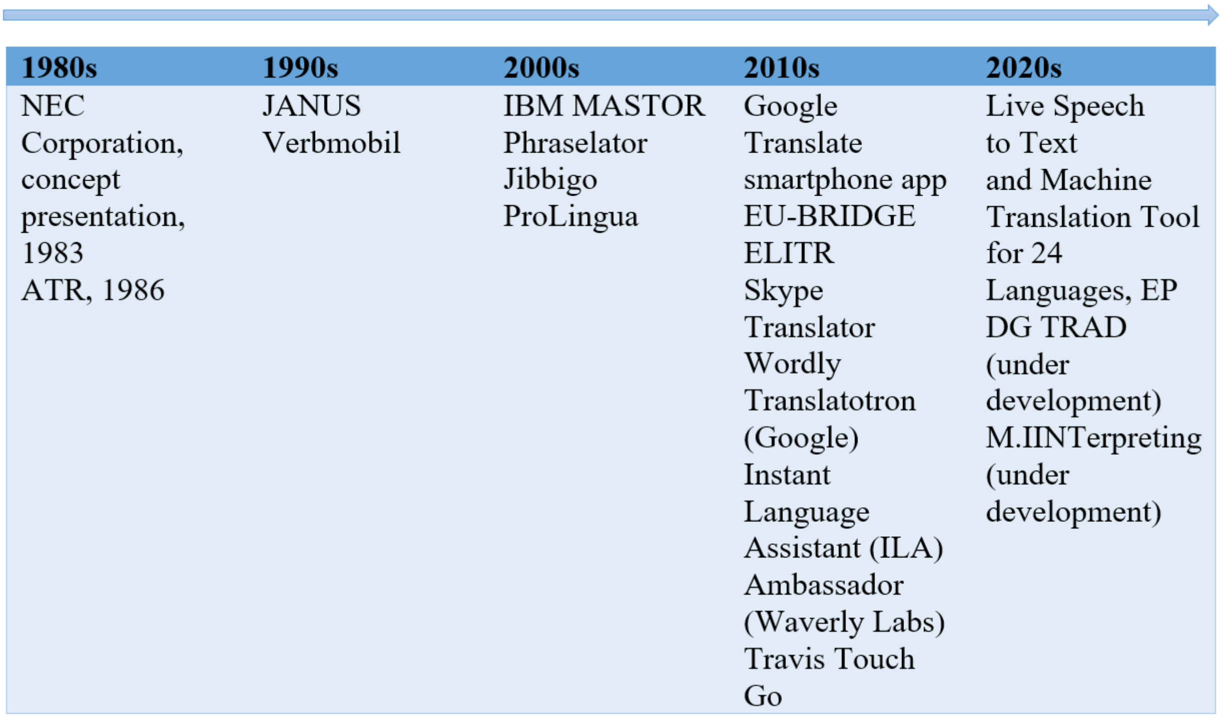
(1) writing down the source language (SL) speech (speech-to-text, STT)

(2) machine translating the SL text into the target language (TL)

(3) synthetizing the TL text to TL speech.

The latest models (Figure 2) insert a fourth component after the STT module to normalize the written SL text and eliminate spoken language phenomena such as repetitions, restarts, disfluencies, hesitations, etc.

The latest models are the so called end-to-end (E2E) or direct models, which leave out the STT



1980s	1990s	2000s	2010s	2020s
NEC Corporation, concept presentation, 1983 ATR, 1986	JANUS Verbmobil	IBM MASTOR Phraselator Jibbigo ProLingua	Google Translate smartphone app EU-BRIDGE ELITR Skype Translator Wordly Translatotron (Google) Instant Language Assistant (ILA) Ambassador (Waverly Labs) Travis Touch Go	Live Speech to Text and Machine Translation Tool for 24 Languages, EP DG TRAD (under development) M.IINTerpreting (under development)

Fig. 1. Speech translation systems development

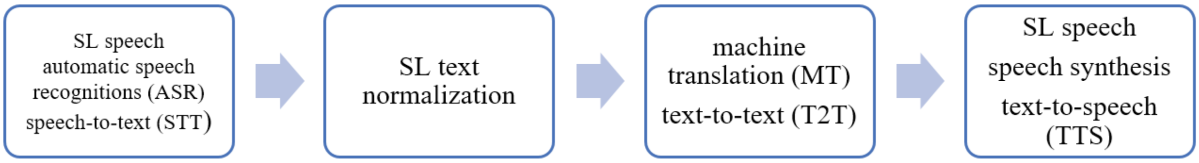


Fig. 2. The cascade model of automated speech translation process

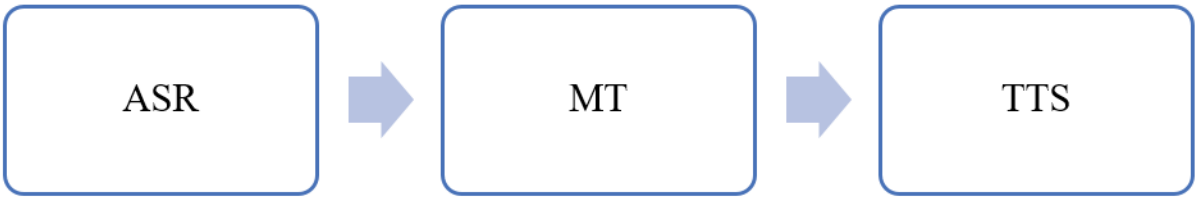


Fig. 3. End-to-end models of automated speech translation

component and go from ASR directly to MT (Figure 3). However, most of the commercialized speech translation devices still use the cascade model (Niehues 2020).

Automatic speech recognition

Good quality and reliable ASR at the beginning of the process is the basis for MT, the next module. The first attempts at ASR date back several decades which consisted of giving simple orders to computers to open files, for example. Then they became more complex and were used for dictation. These systems were taught to recognize the speaker's voice (Yule 1996). However, various problems may arise when using this technology since human speech and speech production are not limited to emitting sounds and words. Such problems may include transforming abbreviations or symbols into words or the lexical elements unknown by the system. Furthermore, possibly the most significant challenge is "automatic speech identification, multilingual acoustic models and multilingual language models" (Jekat 2015, 240). Today, ASR technology is speaker independent and AI-based.

Despite recent progress in technology, for ASR to function well we have to speak languages known by the system and under conditions which enable the speaker's voice to reach the system at a high quality. Another difficulty lies in the fact that the current ASR systems focus on words but ignore paralinguistic features influencing meaning such as intonation, sentence stress or accents. This problem is closely linked to automatic speech segmentation because computers cannot handle these paralinguistic features of speech used for speech segmentation by humans: computers need punctuation. In addition, disfluencies in human speech are similar challenges because speech has to be written down, and written speech needs punctuation marks. Humans are good at handling these phenomena while computer programmes are still incapable of achieving it (Lewis 2020; Niehues 2020).

Machine translation

The second step in automated speech translation is MT. This is the central component on which the current and future success of AST depends to a large extent. It is the technology used by computers to model the human translation process between natural languages. MT is not a recent phenomenon, the first attempts at MT date back as far as to the 1930s (Austermühl 2001) and the first publications on MT in Translation Studies appeared several decades ago (Bar-Hillel 1951; Hutchins 1986; Melby 1981; Sager 1994; Vauquois 1976; Wilss 1993). MT has gone through three stages of evolution so far: (1) rule-based, (2) statistical and error-based, and (3) neural network-based translation (NMT) (Figure 4). In the second period statistical-based machine translation (SMT) became widespread on the translation market. Both SMT and NMT are corpus-based technologies.

Neural network-based research gained momentum in MT around 2007 and penetrated the translation market in 2015, so that by 2017 NMT replaced SMT (Koehn 2017). NMT was a breakthrough in terms of translation quality not only for more widely used language pairs such as English-German, English-Spanish, English-French or English-simplified Chinese (for more detail see Moorkens 2018, 377–378) but also for language pairs where less data is available such as English-Hungarian (Laki 2018).

NMT uses deep learning technology, artificial intelligence and big data. The novelty consists in the fact that NMT does not focus on language structure and does not use only language data and language models. For this reason, it is less dependent

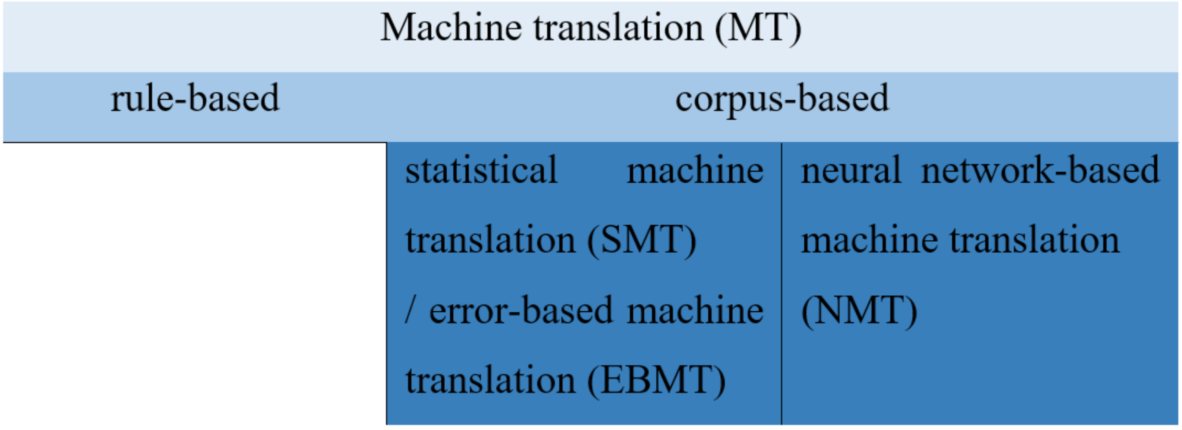

Machine translation (MT)
rule-based
corpus-based
statistical machine translation (SMT) / error-based machine translation (EBMT)
neural network-based machine translation (NMT)

Fig. 4. Technological development of MT

on the SL text than former technologies. Instead, it tries to imitate human thinking and takes into account word context.

The importance of datasets entered into the NMT systems cannot be underestimated for the quality of the text translated into the target language (TL) depends on the quality of the training data. Today, the internet provides developers with a huge amount of data, which is a positive development for automated translation. However, it is also a risk because if low quality data is used for training the MT system, the computer will use that data for the TL text. However, despite the initial enthusiasm over the potentials of NMT, it seems that it has not solved the problem of quality in MT (Le, Schuste 2016; Moorkens 2018).

The issue of quality of MT output is present in automated speech translation to a larger extent for several reasons. First, NMT only takes into consideration the context provided by the word next to a given item and ignores social context or speaker intent. Second, MT uses mathematical algorithms and it is not clear how these can learn in an interpreted situation, since interpreters work with spoken utterances or sign language and not with written texts. Third, interpreting is not only about creating linguistically equivalent texts. Furthermore, the workflow in AST models is a unidirectional mechanical process whereas human interpreting is embedded in a complex communicational situation with the interpreter adapting their behaviour to the requirements and needs of their users.

Speech synthesis

The third step in the AST process involves transforming the written TL text into speech. TTS systems reproduce the acoustic features of sounds and their aim is to create natural or naturally sounding speech (Yule 1996). One of the issues is that written language does not provide clear indication as to how the words written down should be pronounced, or what acronyms stand for. Just like ASR and NMT, speech synthesis technology today is AI-based. However, as Downie notes, progress is still needed for synthesized speech to be not only intelligible but also adequate and appropriate for the context of communication (Downie 2020).

To conclude, it is safe to say that despite recent technological progress, each step in the AST workflow is still problematic and technology has not yet reached the level of development required to provide a human interpreter level of service, consecutive or simultaneous. At this point it needs to be underlined that AST does exist and AST devices are being used every day by humans wishing to understand words, sentences, questions in languages they do not master. However, it should also be mentioned here that AST and human interpreting are two intrinsically different activities, and they should be treated as such for each to find its place on the interpretation market, where AI is gaining ground and, in our technology, oriented world.

Automated speech translation vs. interpreting

AST is language transfer consisting of decoding and encoding linguistic elements and looking for equivalents of translation units. It is a linear process whose main aim is matching data in datasets used as training material fed into language models in MT systems by humans (Table 1). The main aim of this process is therefore not the communication or social intent and fulfilling communication aims. This process is unidirectional from the input at the beginning to the output at the end (see Figures 2 and 3 above). This is not an interactive process, because there is not feedback from the users at the output side that could be built into the ongoing process. Furthermore, because humans do not know exactly how neural networks and such technologies as machine learning function, MI systems are often black boxes (Bird, Fox-Skelly, Jenner et al. 2020; Shaw 2019; Siau, Wang 2020).

The interpreters used to be compared to black boxes and machines in the "conduit model",

Table 1. Automatic speech translation vs interpreting

	Automatic speech translation	Interpreting
Function	artificial language mediation decoding and encoding linguistic elements	natural language mediation, facilitating and supporting the communication process, providing a service
Role	conduit, channel	mediator
Process	no feedback from user linear, unidirectional	feedback from user multidirectional, interactive
Communication	speech recognitions speech synthesis	looking for meaning and sense on speech level situation embedded constructive
Number of languages	limited depending on database	limited only by the number of existing languages
Language use	—	conscious and intentional for supporting the communication intent
Speech	synthetized speech	human speech
Memory	limited only by the size of training datasets	limited
Vocabulary	limited only by the size of training datasets (the Hungarian BERT-large language model's corpus contains 3,67 billion words)	an average educated person's monolingual vocabulary contains 30,000 words (Levelt 1989)
Method	unit matching	creating new TL form context-driven interpretation of meaning
Information-processing	only verbal and not information but matching data	multimodal (verbal, visual)
Knowledge acquisition	—	before, during and after the interpreted event
Professional awareness	—	✓
Soft skills	—	✓

one of the first models of interpreting and interpreter roles in the 1980s and 1990s. In this model the interpreter is described as a channel through which words pass and they are expected to remain completely invisible during the process. However, at the turn of the 2000s, there was a change in the perception of the interpreter and their role in the interpreted communication situation across various types and modes of interpreting (Angelelli 2003; Bischoff, Kurth, Henley 2012; Bot 2003; Diriker 2004; Monacelli 2009; Roy 2002; Tate, Turner 2002; Wadensjö 1998). This new approach in Interpreting Studies takes into account the linguistic, interactional and socio-cultural context of interpreted situations and sees interpreting as a complex linguistic, cognitive and human task. Today, the interpreter is considered a human being and not a non- person, or a communication channel.

AST is an endeavour of a different nature, and it cannot be considered interpreting since it is void of the essence of interpreting, i. e. looking for meaning and making sense of the message. AST functions with text and not speech and it's about mechanically transcoding the linguistic elements of the SL text into the TL.

Human interpreting is a more complex creative and multidimensional task. It is situation-embedded **human communication** developing in a continuously evolving context. As part of the communication situation, interpreters adapt to and depend on the context both in terms of their linguistic and professional behaviour. Human communication is an intrinsically interactive and constructive process, with all the participants of a given communication act contributing to its success or eventual failure. It is safe to say that machines (in this case computers) do not have any communicative competence, and they do not adapt to the communication situation, for example, they do not modify the style or speed or language use according to their users. Today users have to adapt to the needs of the machine, for example, by slowing down their speech, articulating clearly and using accents the computer has been trained on if they want their voice to be recognized by the ASR system. They are also supposed to use vocabulary previously fed into the MT system if they want their words to be transcoded into the TL. Technology provides us with tools, and technological devices created by humans can be expected to be human-centred. Well-functioning devices assist humans and not the other way round.

Human interpreting is also bilingual intercultural communication. Interpreter competences include bilingual competence with well organised and constructed mental lexicons as well as cultural competence, cultural awareness and sensitivity. As Heltai suggests, machines also have some kind of a language competence, but it is different from that of human interpreters in the sense that the machine's language competence is limited: they do not have pragmatic competence therefore, they cannot recognize the contextual meaning of ambiguous expressions. In addition, they do not have discursive competence either, for this reason identifying references is problematic for them (Heltai 2014).

Interpreters as **bilingual language users**, when constructing the TL speech, take into account user needs, their cognitive environment, background and contextual knowledge. An interpreter's speech performance is instrumental, a fundamental feature of human speech. Humans speak because they want to make an impact on their listeners or communication partners by providing information, transferring knowledge, or trying to amuse their audience, just to list a few examples. They are trying to send messages which suit their communication purposes and aims. This also holds true for interpreters, even though they are secondary communicators since the original message is not theirs and they render this message in the TL. In this sense, AST systems do not speak, but rather perform speech recognition and speech synthesis instead of conscious communication and speech behaviour.

Interpreting is a balancing act in the interpreted communication situation. Interpreters' need to find that ideal position where they can facilitate communication without being a natural, primary participant of that communication process. For this, they need creativity, flexibility and adaptation skills. The interpreter's performance requires creative problem-solving, which is closely linked to decision-making and selecting from several possible options as well as divergent thinking, anticipation, imagination, inventiveness and ingenuity. An interpreter's behaviour is characterized by spontaneity, flexibility and a capacity to quickly analyse situations in order to be able to immediately cope with unforeseen events. Another aspect of an interpreter's creativity is finding TL equivalents for new words and expressions. In the case of AST, computers do not possess this capacity because they operate with units in pre-trained datasets and do not create new units even though neural networks can learn from their mistakes to some extent.

Interpreting from a cognitive psychological approach is seen as a process where interpreters are active information users who do not merely take in information in a passive manner but are actively looking for and process information. Thus, interpreting is a **constructive** activity because interpreters contribute actively to the construction of the intended meaning of the message in the SL. AST systems based on neural networks do not operate with words, they operate with word vectors instead. Furthermore, they do not process these vectors but rather match them based on their position in a vector space from training data, which has been previously fed into the system.

Interpreting is an **extremely complex cognitive task**, which is, depending on the interpreting mode, characterized by the partial or complete simultaneity of such mental operations as speech perception and production, attention, memory processes, speech planning and production. For this reason, cognitive skills such as reasoning, attention sharing, fast information processing, task swapping, as well as exercising cognitive control are crucial to successful interpreting. Some of the operations and skills, for example note-taking for consecutive interpreting, language transfer, attention sharing and information processing can be automated to a certain extent. At the same time, because interpreting is carried out in an ever-evolving communication situation, where unexpected events may happen, one of the basic elements of interpreting is **cognitive flexibility**. Cognitive flexibility on behalf of the interpreter is essential for several reasons. First, it is needed for knowledge acquisition during the act of interpreting. Second, cognitive flexibility makes it possible for interpreters to modify their interpreting strategies when needed. Third, it is essential for performing such mental operations as anticipation, inference and creative problem-solving based on divergent thinking.

Interpreters are professionally aware. **Professional awareness** means that they know their profession, they are aware of the requirements in terms of knowledge and skills but also in terms of technology and ethics. They can analyse and evaluate their own performance and collaborate with all stakeholders, including their colleagues. They know the metalanguage used to talk about various modes and types of interpreting and the operations underlying the interpreting process. They are also aware of the cognitive, personal and interpersonal processes which occur during interpreting. They are also familiar with the way interpreting as a profession is practised, the market requirements, players, trends, etc. Another component of professional awareness is familiarity with professional organisations and codes of ethics. Interpreter training programs provide trainees with a comprehensive view of the importance of respecting the ethical principles laid down in codes for professionals and the profession as a whole. This component of the declarative knowledge of the interpreter is completely missing from AST systems.

In addition, interpreting is rendering a **complex service**, which is not restricted to language mediation at the venue of the interpreted communication event. It requires an array of **soft skills**, for example, interpersonal skills, communication and listening skills, empathy, the ability to cooperate and work in a team, etc. Interpreters are usually in contact, before and after the assignment, with their clients, strive to know their expectations and needs as well as the aims and objectives or the role of the interpreted event in the professional life of their users. After the assignment, they consolidate their terminological work and evaluate their performance and the event. An interpreter-client/user relationship based on **trust** is as important for successful interpreting as the interpreter's performance during the interpreted event. One of the reasons for this is that quality in interpreting is a relative concept: it depends to a large extent on the perspective and expectations of those evaluating. Quality criteria may vary, according to whether this person is the interpreter, the event organiser, the user, the client or a colleague.

One area where AST systems have a marked advantage over humans is the amount of data which they can store. This amount is limited only by the size of the training datasets entered into the system. However, as Niehues points out, data efficiency is an issue in the case of NMT systems (Niehues 2020). Data efficiency problems mean that there are still languages for which there is still not enough data available. But it also means that current language models contain considerably more data than a human being comes across during their entire lifetime, but their performance is still lower than human performance.

AST and human interpreting: perspectives

The burning question for interpreters, trainers, all stakeholders in interpreting and laypersons is, of course, whether AST can or will ever replace humans and whether technological advances will reach a level when it can be feasible. Researchers (Downie 2020; Jekat 2015; Jekat, Klein 1996; Pöchhacker 2016) and even developers (Lewis 2020) argue that AST will never fully replace humans, and that it does not aim to do so, it is meant to facilitate bilingual communication to some extent in situations where employing a professional interpreter is financially unaffordable or physically impossible. Ray Kurzweil has introduced several innovations to automate human processes and tasks, but still he is of the opinion that full automatization of translation (and interpreting) will never be achieved (Kelly, Zetzsche 2012). However, some interpretation market players claim that AST is not fiction anymore but has become reality (Nimdzi 2019), and we have seen that there are AST devices already available on the market (Figure 1) which are currently used for supporting bilingual communication. Furthermore, technological advances have not come to an end and will certainly continue.

The typical use cases of AST devices are healthcare, military, travel and tourism, business, public service as well as education (Table 2). These are well defined communication situations, where human interpreters are not often available or cannot be afforded or there is no time to organise human interpretation. Most of these communication situations and the vocabulary and language used are fairly predictable, and AST may serve as a tool to bridge the communication gap if everything goes according to plan. And even in the most frequent AST use cases most of the communication is of an administrative nature. In healthcare, for example, checking in and out of hospitals, routine examinations or re-education exercises are speech translated automatically but human interpreters are asked to interpret more complex and sensitive conversations between doctors and patients. Despite the technological advances since the first AST devices appeared (Verbmobil in the 1990s, or Jibbigo and EU-Bride at the beginning of the 2010s), the use cases, scenarios and aims have remained the same with maybe one exception which is accessibility for persons living with hearing or visual impairment. So far, attempts to automate conference interpreting have failed (Lance 2018), and there are situations such as legal contexts or high-risk meetings in politics or business with classified information, where the use of such devices is atypical.

AST: Cognitive and communicational limitations

The main reasons for this limited number of use cases may be found in the fact that computers today cannot handle unforeseeable events, are unaware of cultural aspects of the communication situation as well as the social and communicational characteristics of the interpreted situation. Furthermore, features linked to the pragmatic and linguistic aspects of oral communication and spontaneity also pose problems. Likewise, language registers, style, idiosyncratic language use, hesitations, ambiguity, humour and irony, too fast or too slow speech, turn-taking, etc. are also areas that humans can handle but are tricky for computers. In addition, natural language is an open and continuously changing system with an infinite number of elements, therefore rare and new expressions are also problematic.

Lewis mentions politeness and gender as specific challenges computers are facing in terms of pragmatics. Another problem is turn-taking in real life conversations, when speakers don't wait for each other to finish their sentences. For such bilingual conversations, when AST is used, "forced turn-taking" takes place, which impacts negatively on user experience by making the communication situation artificial (Lewis 2020).

The current AST devices use the most advanced AI, big data and deep learning technology. However, this technology has not reached the level of human cognitive performance required to fulfil complex cognitive tasks requiring creativity, intuition, cognitive flexibility and judgement which would enable computers to comprehend commu-

Table 2. Examples of AST use cases

Use case	Devices	Mode
healthcare	Prolingua	consecutive
military	IBM Mastor, Phraselator	consecutive
travel and tourism	Jibbigo, ILA, Skype Translator	consecutive
business	Vebmobil, ILA	consecutive, simultaneous
government (e.g. immigration, border patrol)	ILA	simultaneous
university lectures	EU-Bridge	simultaneous
education (e.g. parent-teacher meetings)	Skype Translator	simultaneous
accessibility (deaf/hard of hearing; blind/low vision)	ILA, EP, Skype Translator	simultaneous
conversations	various	consecutive, simultaneous

nication intent, exercise cognitive control over the communication situation and be able to manage and process the communication situation in a holistic manner. Instead, computers work with words and take into account only the immediate context of a given work in the word chain. For this reason, in addition to the problems enumerated above, they cannot handle technical or semantic interferences.

We have seen above that understanding communication intent is a complex process, characterized by the fact that participants in the communication situation actively contribute to meaning construction. For this, they need general and domain specific background knowledge, not only technical vocabulary, since meaning-based interpretation means that it is not words but rather ideas and arguments that we transfer from the SL into the TL. During this process the cultural, social, political, and other contexts as well as the contextual features are as important for the comprehension of the speaker's communication intent as the words they are using.

AST: Technological limitations

There is no doubt that we have been witnessing significant advances in the field of AI-powered technologies for the last decade or so. However, we need to distinguish between two types of artificial intelligence: general AI and AI restricted to certain tasks. General AI is still unavailable and would mean human-like cognitive skills such as thinking, reasoning, autonomous decision-making. What is available, however, are AST devices whose use is restricted to certain areas.

One of the main technological challenges of AST systems is the replication of errors in cascade models. This means that if an error occurs during ASR and a wrong word gets into the SL text to be machine translated, this will not be recognized and filtered by the computer, but will be forwarded to the MT module. Another challenge is posed by simultaneity, which means reducing latency, i. e. the time between the SL speech and the TL speech, to the minimum because it is essential in terms of user experience to produce the SL speech in as little time as possible (Niehues 2020).

Melby notes in his discussion on MT that "there is still a serious flaw in the design", namely the error "to buy into the Black Box Myth of translation, which assumes that each source text has exactly one correct translation". Naturally this is not the case and the way a text is translated "depends not just on the source text itself, and not just on the source texts plus its purpose and intended audience, but also on the purpose and intended audience of the translation, which may differ considerably from the author's purpose and audience" (Melby 2002, 46). Since MT is one of the central modules in the process of AST, the Black Box Myth may be one of the impediments to the full automation of the speech translation process. Language use related decisions, for example, during word retrieval, depends not only on finding the right lexical units but also on taking into consideration the actual users, their needs, background knowledge and also on possibilities offered by the time constraints characterizing interpreted communication situations.

AST: Domain limitations

It is worth noting that there are cases where MT is inapplicable such as literary translation, advertisements or business presentations, where the main aim is to impress the reader. In such cases the form of the text is as important as the information it conveys. Heltai enumerates some other text types which are not suitable for MT at all such as jokes, word games, humorous texts, well formulated editorials, everyday conversations for whose comprehension discursive, pragmatic and socio-cultural competences are needed (Heltai 2014). As for spoken communication, when we speak, our objective is very often to make a positive impression on our communication partners or listeners, and we often use humour to achieve this goal. In addition, our spoken communication often expresses emotions or our attitude towards our subject matter. The information in our spoken communication acts is not merely a dataset, but we aim to achieve something with it, for example, provide information, convince somebody of something, develop somebody's skills or entertain.

In the case of machine translation-based AST, the difference between artificial and natural languages needs to be taken into consideration. Sager considers the language of machine translated texts to be artificial. Contrary to natural languages, artificial languages are of limited nature, meaning that they do not have aesthetic or emotive meaning. This also means that artificial languages do not behave as natural languages "characterized by a maximum freedom of formal variation at all levels of articulation, according to criteria of usage and function". Furthermore, the "growth, diversification and variation of natural language is only limited by the boundaries of mutual comprehensibility among speakers" (Sager 1994, 33). From this it can be deduced that a spoken utterance communicated in a natural language cannot be considered equivalent to its artificial TL form.

However, there are cases when the use of MT is justified. According to Sager, it is suitable to use MT when there is (1) "insufficient human capacity available to translate the considerable volume", or (2) "a very large demand for immediate, very low cost translation which cannot be produced by human translators" (Sager 1994, 261). Varga states that MT "is suitable for the purpose of obtaining information, determining the subject of a text or for determining whether further analysis, processing of a given text requires human input" (Varga 2016, 161). It can also be justified in the case of informative technical texts, where comprehension depends mostly on information expressed explicitly in linguistic form and to a lesser extent on inferences based on context or connotations (Heltai 2014).

For Heltai, "minimal translation" is when a large volume of text needs to be translated urgently, or when low budget translation is needed (Heltai 1999). In such cases, translation quality is lower and the translator, instead of providing a complete version of the SL text in the TL, makes a partial translation of the text, keeping the main elements and messages. The product of MT can often be considered minimal or raw translation, which is then post-edited by humans.

AST's impact on the interpreting professions

Up until this time, machine interpretation's impact on the interpretation profession cannot be felt to the same extent as that of machine translation's on the translation profession. One reason for this lies in the fact that there is more demand for automatically generated written translations. Another reason might be that the automation of interpretation must take into account a number of real-time variables too, which do not arise during translation. In automatizing the human interpretation process, no sub tasks such as text preparation for translation or post-editing that could be carried out independently of the translation itself have evolved. Such tasks will probably never evolve at all, since if computers do replace the human interpreter, post-editing TL spoken utterances would be difficult to carry out because they are intended for immediate use in 'live' communication. Furthermore, preparation or pre-editing for AST does not necessarily have to be carried out by interpreters but rather by language technologists or terminologists.

Developers, however, are of the opinion that this is the way forward because fully automated AST is likely to remain unachievable for some time. In addition to AI-driven transcription and terminology management, a 'hybrid' approach could be applied to AST where the computer's and the interpreter's work is shared. In this setup, the computer would do the boring, repetitive manual work, and the interpreter the creative, real tasks in the interpretation process (Lewis 2020). Research is in its infancy today, however, it is worth noting that the interpretation process is extremely complex, and the interpreter's performance is impacted by various external (the working conditions, the speaker, the SL speech, etc.) and internal factors (interpreting skills, experience and expertise, assignment preparation, stress management, etc.). These factors are interlinked and add up to influence the interpreter's performance.

Transcribing the interpreter's TL output also raises concerns because the TL speech produced by the interpreter is meant to be used in the immediate context of the event, which is not the case of translations disseminated in writing. Due to the nature of the interpreting task, compared to what can be expected in the case of written translations, interpreted TL speeches may be characterized by certain "imperfections", which might make immediate transcription difficult. Another issue to be taken into account here is that the interpreters' consent should be sought if their TL performance was to be transcribed. Furthermore, it is unclear how the interpretation process could be divided into "manual" and more creative work. What can be considered manual work in such a complex bilingual activity? Making terminology work easier and faster, real time help with terminology may be a welcome assistance during the interpretation process. However, we should bear in mind that preparing the terminology of an interpreted event is not only looking up technical terms but also serves the purpose of content preparation, when the interpreter acquires the domain-specific background knowledge needed for successful meaning-based interpreting performance. This is something computers cannot carry out for interpreters.

Technological development will continue and probably even accelerate since our age is obsessed with technology (Besnier 2012). We can presume that attempts to fully automate the interpretation process will not stop either. It is, of course, very difficult to foresee the future of interpreting. However, looking at how technology has impacted on the translation market for the past 20 years, we might predict with reasonable certainty that the interpretation market will be divided into two segments: a market where automated 'minimal interpretation' will be sufficient and available at a lower price or even free of charge; and a premium segment where professional human interpreters will work and provide complex, high-quality services. When the use of MT became widespread on the translation market, translators were afraid that machines would take their jobs and livelihoods. In fact, MT has transformed the translation market but it has not replaced the human translator. Instead, the volume of translation has grown significantly, and new professions have been formed such as pre-editing and post-editing, localization, language engineering, translation project management and vendor management (Horváth 2016).

We have seen that the use of MT is justified in certain cases, and that these translations are often "minimal translations", which provide some basic information on the topic and the main elements of the SL text. In the case of AST, one of the main questions is to what extent such minimal or raw speech translations can be effectively used in an interpreted communication situation, where the participants, users and clients engage in interaction and watch each other's reactions and feedback to spoken utterances. Can a minimal speech translation produced by a computer programme fulfil the role of the interpreter as a language and communication professional?

It depends on several factors. First, it depends greatly on whether AST developers manage to build trust in those requiring spoken language interpretation to use AST tools. It also depends on whether the users of interpretation would wish to enjoy the added value of services and soft skills provided by human interpreters such as empathy, task-oriented approach, managing culturally sensitive situations, gisting or self-correction. It may also depend on whether or not users wish to involve a professional who actually understands what they are saying.

Conclusions

AST tools are being used in well-defined and less complex situations. Accelerating technological development has impacted on the interpretation market and there are numerous attempts to automate the interpretation process. As early as 1994, Sager argued that MT "has a proper place beside human translation as an alternative technique for achieving different communicative objectives" (Sager 1994, 262), which seems to be true for the interpretation market in 2021.

Although today it is safe to say that technological advances have not yet reached the level needed for general fully automated interpretation in all communication situations, it is also obvious that technological development will continue. The modules of the S2S translation process such as ASR, MT and speech synthesis will become increasingly more sophisticated, new modules and technology will become available. In addition, AI-based new technology such as facial expression recognition software, voice imitation and mouth modification to harmonize lip movement with the TL, synthesized video technology or humanoid robotics will offer new possibilities in the automated speech translation market. Although there are various intrinsic differences between AST and human interpreting and reasons why the expertise and competences of a human interpreter cannot be replaced by AI-based tools today, we cannot exclude the possibility of them becoming technologically feasible one day.

However, this issue is not purely of a technological nature, in other words, the use of AI technology in interpreting is not exclusively a question of technological development. It depends to a great extent on user needs and expectations, more precisely on the social impact of technological development on our societies in general. Will the time come when it is trendier to use AST than a human interpreter's services? Will people who use AST devices, avatars or humanoid robots to help them with speech translation be considered progressive and those employing humans outdated? Robotization may reach a level where it impregnates our lives and becomes so widespread that users needing assistance with spoken language translation trust AI-based devices more than human interpreters because they may feel that machines are more discreet and trustworthy. Another question which emerges is whether lower quality TL speeches will be worth cost saving so that AST is chosen. Can we be sure that AI-driven AST devices will come at a lower cost, bearing in mind that maintaining an increasing number of ever more powerful servers has a considerable impact on the environment? Will computers whose performance is restricted to MT provide the same communication experience as highly trained professional intercultural mediators, with whom it is also possible to exchange ideas outside the strict interpreted communication situation?

With the growing automatization of our lives, the services provided by human interpreters may become more appreciated because interpreting is not about finding translation equivalents, it is not only repetitive, routine communication. Human interpreters contribute to successful multilingual communication situations not only in terms of language transfer between source languages and target languages. They also facilitate dialogue, with their expertise they support the organizers of interpreted events, and with their personality, they contribute to successful human communication and relations. This is not to say that human interpreters are perfect. They have limited cognitive, working memory and long-term memory capacity, make mistakes, get tired and feel stressed. However, they are capable of learning from their mistakes, they build newly acquired knowledge into their knowledge base and continuously self-evaluate their performance. They remember past events and such experience, gained through the years, may be priceless for a client who is planning long term. In fact, it is highly typical of event organizers to ask for the same interpreters year after year, provided they were satisfied with their performance. This is more than satisfaction with linguistic performance: this implies trust, loyalty and shared goals.

It also needs to be acknowledged that AI systems have certain advantages over human interpreters, for example, in terms of data volume and vocabulary used to train AST systems, they already considerably outperform humans. Despite this fact, human performance is more sophisticated, multifaceted and adaptable. One of the undeniable values of human performance is that interpreters strive to understand and 'interpret' in the strictest sense of the word what they hear and say. Thus, they can decide what to leave out, what elements need to be explicated, emphasized or reformulated in order to respect cultural, social and personal sensitivities.

The interpreting profession is at an important crossroads. How it will evolve depends on several factors. First, the professional community of interpreters will bear a great responsibility in providing much higher quality and more complex interpreting services than AI-driven systems. For this, they need to know the possibilities offered by technology, and if they are useful, to be able to build them into the interpretation process. Second, interpreting service provider agencies will play an important role in advising clients in which situations AST is sufficient and when human interpreters should be chosen so that they can make an informed decision. Third, interpreter training programmes can also play a significant role in this process by ensuring quality training, thus guaranteeing the supply of highly qualified professional interpreters open to technological advances.

Technological development has brought about unprecedented changes on the interpretation market: new tools and communication situations have emerged where spoken language translation is used. Such a widespread use of technology entails various ethical risks in terms of data security, confidentiality and personal data use. The interpreting community should start preparing for tackling these ethical challenges now before the use of cloud-based AI-systems becomes widespread on the interpretation market.

Abbreviations

AI — artificial intelligence
ASR — automatic speech recognitions
AST — automated speech translation
EBMT — error-based machine translation
E2E — end-to-end
MT — machine translation
NMT — network-based translation
SL — source language
SMT — statistical-based machine translation
S2S — speech-to-speech
STT — speech-to-text
TL — target language
T2T — text-to-text
TTS — text-to-speech

References

Angelelli, C. V. (2003) The interpersonal role of the interpreter in cross-cultural communication. In: L. Brunette, G. L. Bastin, I. Hemlin, H. Clarke (eds.). *The Critical Link 3: Interpreters in the community. Selected papers from the Third International Conference on interpreting in legal, health and social service settings, Montréal, Quebec, Canada, 22–26 May 2001.* Amsterdam; Philadelphia: John Benjamins Publ., pp. 15–26. https://doi.org/10.1075/btl.46.06ang (In English)

Austermühl, F. (2001) *Electronic tools for translators.* London: Routledge Publ., 202 p. https://doi.org/10.4324/9781315760353 (In English)

Bar-Hillel, Y. (1951) The present state of research on mechanical translation. *American Documentation*, vol. 2-4, pp. 229–237. (In English)

Besnier, J.-M. (2012) *L'homme simplifié: Le syndrome de la touche étoile [The simplified man: The star key syndrome].* Paris: Fayard, 201 p. (In French)

Bird, E., Fox-Skelly, J., Jenner, N. et al. (2020) *The ethics of artificial intelligence: Issues and initiatives*. Brussels: European Union Publ. [Online]. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf (accessed 04.05.2021). (In English)

Bischoff, A., Kurth, E., Henley, A. (2012) Staying in the middle: A qualitative survey of health care interpreters' perception of their work. *Interpreting*, vol. 14, no. 1, pp. 1–22. https://www.doi.org/10.1075/intp.14.1.01bis (In English)

Bot, H. (2003) The myth of the uninvolved interpreter. Interpreting in mental health and the development of three-personpsychology. In: L. Brunette, G. L. Bastin, I. Hemlin, H. Clarke (eds.). *The Critical Link 3*: *Interpreters in the community. Selected papers from the Third International Conference on interpreting in legal, health and social service settings, Montréal, Quebec, Canada, 22–26 May 2001.* Amsterdam; Philadelphia: John Benjamins Publ., pp. 27–35. https://doi.org/10.1075/btl.46.07bot (In English)

Diriker, E. (2004) *De-/Re-contextualizing conference interpreting*. Amsterdam; Philadelphia: John Benjamins Publ., 223 p. (In English)

Downie, J. (2020) *Interpreters vs Machines. Can interpreters survive in an AI-dominated world?* New York: Routledge Publ., 176 p. (In English)

Fantinuoli, C. (2018) Interpreting and technology: The upcoming technological turn. In: C. Fantinuoli (ed.). *Interpreting and technology*. Berlin: Language Science Press. pp. 1–12. https://www.doi.org/10.5281/zenodo.1493289 (In English)

Heltai, P. (1999) Minimális fordítás [Minimal translation]. *Fordítástudomány — Translation Science*, vol. 1, no. 2, pp. 22–33. (In Hungary)

Heltai, P. (2014) Gépi fordítás. Mi az, amit a gép nem tud fordítani? [Machine translation. What can't machines translate?] In: *Szaknyelv és szakfordítás. Tanulmányok a szakfordítás és a fordítóképzés aktuális témáiról [Professional language and translation. Research on current issues of professional translation and translator training]*. Gödöllő: Szent István Egyetem Publ., pp. 7–29. [Online]. Available at: http://nyelviintezet.szie.hu/sites/default/files/szie_2014_vegleges_szakfordito.pdf (accessed 04.05.2021). (In Hungary)

Horváth, I. (ed.). (2016) *The modern translator and interpreter*. Budapest: Eötvös University Press. [Online]. Available at: http://www.eltereader.hu/media/2016/04/HorvathTheModernTranslator.pdf (accessed 04.05.2021). (In English)

Hutchins, W. J. (1986) *Machine translation: Past, present, future*. New York: Wiley & Sons Publ., 382 p. (In English)

Jekat, S. J. (2015) Machine interpreting. In: F. Pöchhacker, N. Grbić, P. Mead, R. Setton (eds.). *Routledge encyclopedia of interpreting studies*. London; New York: Routlegde Publ., pp. 239–241. (In English)

Jekat, S. J., Klein, A. (1996) Machine interpretation. Open problems and some solutions. *Interpreting*, vol. 1, no. 1, pp. 7–20. https://doi.org/10.1075/intp.1.1.02jek (In English)

Kelly, N., Zetzsche, J. (2012) *Found in translation. How language shapes our lives and transforms the world*. New York: Penguin Publ., 191 p. (In English)

Koehn, P. (2017) *Statistical machine translation. Draft of Chapter 13: Neural machine translation*. [Online]. Available at: https://arxiv.org/pdf/1709.07809.pdf (accessed 04.05.2021). (In English)

Laki, L. J. (2018) Mesterséges intelligencia a gépi fordításban [Artificial intelligence in machine translation]. In: Tolcsvai Nagy G. (ed.). *A humán tudományok és a gépi intelligencia [Human sciences and machine intelligence].* Budapest: Gondolat Kiadó Publ., pp. 156–183. [Online]. Available at: http://real.mtak.hu/88740/1/Tolcsvai-nyomda_156_laki.pdf (accessed 04.05.2021). (In Hungary)

Lance, Ng. (2018) *AI Interpreter Fail at China Summit Sparks Debate about Future of Profession.* [Online]. Available at: https://slator.com/features/ai-interpreter-fail-at-china-summit-sparks-debate-about-future-of-profession/ (accessed 04.05.2021) (In English)

Le, Q. V., Schuste, M. (2016) A neural network for machine translation, at production scale. *Google Al Blog*. [Online]. Available at: https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html (accessed 04.05.2021). (In English)

Levelt, W. J. M. (1989) *Speaking: From intention to articulation*. Cambridge: MIT Press, 566 p. (In English)

Lewis, W. (2020) AI and interpreting. Will Lewis on automated speech translation: Scale, uses & edge cases. *YouTube*, 13 November. [Online]. Available at: https://www.youtube.com/watch?v=KihPeHh0wyo (accessed 04.05.2021). (In English)

McNair, A., Waibel, A., Jain, A. L. et al. (1991) JANUS: A speech-to-speech translation system using connectionist and symbolic processing strategies. In: *ICASSP 91: International conference on acoustics, speech, and signal processing. Vol. 1*. Washington: IEEE Publ., pp. 793–796. https://doi.ieeecomputersociety.org/10.1109/ICASSP.1991.150456 (In English)

Melby, A. K. (1981) Translators and machines — can they cooperate? *Meta*, vol. 26, no. 1, pp. 23–34. https://doi.org/10.7202/003619ar (In English)

Melby, A. K. (2002) Memory and translation. *Across Languages and Cultures*, vol. 3, no. 1, pp. 45–57. https://doi.org/10.1556/Acr.3.2002.1.3 (In English)

Monacelli, C. (2009) *Self-preservation in simultaneous interpreting. Surviving the role.* Amsterdam; Philadelphia: John Benjamins Publ., 182 p. https://www.doi.org/10.1075/btl.84 (In English)

Moorkens, J. (2018) What to expect from Neural Machine Translation: A practical in-class translation evaluation exercise. *The Interpreter and Translator Trainer*, vol. 12, no. 4, pp. 375–387. https://doi.org/10.1080/1750399X.2018.1501639 (In English)

Nakamura, S. (2009) Overcoming the language barrier with speech translation technology. *Science & Technology Trends — Quarterly Review*, no. 31, pp. 36–49. (In English)

Niehues, J. (2020) AI and interpreting. Jan Niehues on automated speech translation: Challenges and approaches. *YouTube*, 13 November. [Online]. Available at: https://www.youtube.com/watch?v=90E9J1zPxlY (accessed 04.05.2021). (In English)

The Nimdzi Interpreting Index. (2019) *Nimdzi.* [Online]. Available at: https://www.nimdzi.com/the-2019-nimdzi-interpreting-index/ (accessed 04.05.2021). (In English)

Pöchhacker, F. (2016) *Introducing interpreting studies*. 2nd ed. London; New York: Routledge Publ., 280 p. (In English)

Roy, C. (2002) The problem with definitions, descriptions, and the role metaphors of interpreters. In: F. Pöchhacker, M. Shlesinger (eds.). *The interpreting studies reader*. London; New York: Routledge Publ., pp. 344–354. (In English)

Sager, J. C. (1994) *Language engineering and translation: Consequences of automation*. Amsterdam; Philadelphia: John Benjamins Publ., 345 p. https://doi.org/10.1075/btl.1 (In English)

Shaw, J. (2019) Artificial intelligence and ethics. Beyond engineering at the dawn of decision-making machines. *Harvard Magazine*, no. 1, pp. 44–74. [Online]. Available at: https://harvardmagazine.com/sites/default/files/pdf/2019/01-pdfs/0119-44.pdf (accessed 04.05.2021). (In English)

Siau, K., Wang, W. (2020) Artificial Intelligence (AI) ethics: Ethics of AI and ethical AI. *Journal of Database Management*, vol. 31, no. 2, pp. 74–87. http://doi.org/10.4018/JDM.2020040105 (In English)

Tate, G., Turner, G. H. (2002) The code and the culture. Sign language interpreting — in search of the new breed's ethics. In: F. Pöchhacker, M. Shlesinger (eds.). *The interpreting studies reader*. London; New York: Routledge Publ., pp. 374–383. (In English)

Varga, Á. (2016) Machine translation. In: I. Horváth (ed.). *The modern translator and interpreter*. Budapest: Eötvös University Press, pp. 153–165. (In English)

Vauquois, B. (1976) Automatic translation — a survey of different approaches. In: H. Karlgren (ed.). *SMIL: Statistical methods in linguistics*. Stockholm: Språkförlaget Publ., pp. 127–135. (In English)

Wadensjö, C. (1998) *Interpreting as interaction*. London; New York: Longman Publ., 312 p. (In English)

Wilss, W. (1993) Basic concepts of MT. *Meta*, vol. 38, no. 3, pp. 403–413. https://doi.org/10.7202/004608ar (In English)

Yule, G. (1996) *The study of language*. 2nd ed. Cambridge: Cambridge University Press, 294 p. (In English)