User perspectives on developing technology-assisted access services in public broadcasting

Maarit Koponen, University of Helsinki, Finland,
maarit.koponen@helsinki.fi
Tiina Tuominen, Yleisradio Oy, Finland,
tiina.tuominen@yle.fi
Maija Hirvonen, Tampere University & University of Helsinki, Finland,
maija.hirvonen@tuni.fi
Kaisa Vitikainen, Yleisradio Oy, Finland,
kaisa.vitikainen@yle.fi
Liisa Tiittula, University of Helsinki, Finland,
liisa.tiittula@helsinki.fi

Abstract

The growing demand for accessible media content in the Creative Industries and increased pressure to produce more content at lower costs has led the industry to look for technological support for creating and managing audiovisual content. In order to design truly accessible technology-assisted solutions and services, it is crucial to center the knowledge and experience of the intended users – both the consumers and the professionals involved in producing the content and services. This article explores potential technological solutions for audiovisual media access services addressing intralingual, interlingual and intersemiotic access in public service television broadcasting. We introduce an ongoing research project taking a user-centered approach, and present work carried out on automatic and semi-automatic methods involving intralingual and interlingual subtitling, and description of visual content. Based on themes raised in individual and focus group interviews, we discuss how potential user groups respond to technological solutions at differing levels of maturity. We examine conceptualizations of quality, trust, and accessibility emerging in the users' comments, and chart the differences and similarities between different stakeholders. The article demonstrates how diverse user perspectives can inform research and development and enhance our understanding of the role of technology in promoting media accessibility.

1. Introduction

Recent years have seen increased interest in accessibility of various types of products, content and services. Accessibility is regarded as vital for all humans (Greco 2016, 2018): it can be defined as "enabling

participation of all people" (Hirvonen and Kinnunen 2021, 470). Accessibility as such is an abstract experience. In more concrete terms, what is provided by institutions aiming to promote access are specific *access services* (Jankowska 2020, 243). Accessibility measures and access services help overcome sensory, physical, mental and linguistic barriers (see e.g. Matamala and Ortiz-Boix 2016; Hirvonen and Kinnunen 2021 for discussion). Translation and interpreting have an important role in accessibility, in that crossing (linguistic) barriers is central to translation and simultaneously translation is an accessibility measure.

The present article focuses on media accessibility, which covers practices aiming to provide access to audiovisual media for people who would not be able to (fully) access that content in its original form (Greco 2016, 23). Audiovisual media access services include content-based methods where translation practices such as intralingual and interlingual subtitling and audio description (AD) are used to produce new texts to reach specific audiences, or technology-based methods where existing content is processed digitally (Jankowska 2020). We discuss four use cases which involve intralingual and interlingual access as well as access to visual information, and explore the potential of technological solutions (automatic speech recognition, machine translation and automatic video description) to support the production of access services in media, particularly in public television broadcasting. The discussion draws upon user studies conducted as part of a European research project MeMAD (Methods for Managing Audiovisual Data, see Section 3). The use cases that form the background of our discussion arise from the pragmatic context and needs of the Finnish public broadcasting company Yle, which is one of the project partners. The aim of this article is not to provide a detailed description of the empirical studies of the project. Rather, it is intended as a discussion paper exploring how various user groups in these studies respond to the introduction of technological solutions at different stages of maturity, and how the concepts of usability and user-centeredness may support technological advances in access services.

At the same time as demand for accessible content and access services is growing, Creative Industries are facing pressure to produce more content with shortening turnaround times and lower costs. Most audiovisual content is not currently made accessible, and tackling the sheer volume by human effort alone may not even be possible, which has led the audiovisual translation field to turn to technological solutions for support (see Georgakopoulou 2012; Remael *et al.* 2016). Discussing audio description, Remael, Reviers and Vandekerckhove (2016, 257) argue that high priority should be placed on "increased insight and approaches that promote interaction between human-driven and technology-driven solutions". The need for further research and insight

into technological and semi-automatic solutions similarly applies to other accessibility approaches.

The need for emphasizing the knowledge and experience of the intended users is increasingly being acknowledged in the design of services and technologies. As explained by Greco (2018, 212), a traditional view was to design solutions or products according to the maker's perspective or interpretation of the users' needs. However, for developing truly accessible solutions, the users must be central to the design process (*ibid*.). In fact, access services have been at the forefront of user-oriented thinking, and continue to focus on how best to account for the users' perspective (e.g. Di Giovanni 2018). For audiovisual media access services, the "users" include not only consumers, but also professionals who are involved in the production of the content and services and use various technologies to make media accessible. They represent various fields from translators and subtitlers to technology professionals (Jankowska 2020).

Furthermore, as Greco and Jankowska (2019, 8) point out, it is important to account for the differences between user groups and their perspectives on access services, particularly their considerations of quality. For this, user-centered design (UCD; see e.g. User-Centered Design Basics) offers a framework to make users a crucial part of the design process. User-Centered Design has previously been used as a model for the concept of User-Centered Translation (UCT; Suojanen *et al.* 2015), where the principles of user-centeredness have been applied to interlingual communication and linguistic accessibility. UCT provides tools and approaches that incorporate users' views into the translation process. These tools have been used in the MeMAD project in interlingual, intralingual as well as intersemiotic contexts. The underlying purpose in both UCD and UCT is to make products and services usable, i.e. easy and pleasant to use when completing a specific task. For the end users, the task could be to follow and enjoy media content with the help of the access service. For the professionals involved, the task is to produce that content efficiently and to a high standard. This article introduces the user-centered approach employed by the MeMAD project, which has included different stakeholder groups. We discuss how charting the differences and similarities between different stakeholders can inform decision-making and development.

In this article, we bring together these perspectives of media accessibility, specifically in the context of public service television broadcasting, technology and user-centeredness. Based on individual and focus group interviews conducted as part of the user studies in the MeMAD project, we discuss the varying experiences and themes that emerge as important for the different user groups. In Section 2, we give a brief overview of other research and projects focused on technological

solutions for accessibility. In Section 3, we present the MeMAD project, and Section 4 outlines work conducted with different user groups as part of the project. Section 5 discusses the conceptualizations of quality, trust, and accessibility expressed by the different users. Our final remarks and reflections on the potential of technological innovation to facilitate accessibility are presented in Section 6.

2. Research on technology and access services

Technological solutions for increasing accessibility to audiovisual content in the Creative Industries have been investigated in various projects. Research appears to have mainly focused on automatic and semi-automatic solutions for producing intralingual subtitles with automatic speech recognition (ASR) and translating content (subtitles, AD and voice-overs) with machine translation (MT). An early example of using ASR and MT for creating fully automated subtitles is the MUSA project[1], which focused on subtitles for documentaries and "current affairs" programs (Piperidis *et al.* 2004). The eTITLE project (Melero *et al.* 2006) developed an online subtitling platform with MT and translation memories to support subtitle translators' work, and assessed the usability of the system for movie subtitling. Automated intralingual subtitling through ASR was the focus of the SAVAS project (Del Pozo *et al.* 2014). EU-BRIDGE[2] addressed various use cases, including ASR and MT for television broadcast captioning (The EU-BRIDGE Consortium 2015). The SUMAT project[3] developed MT for subtitling in multiple language pairs. The usability of MT and its effect on subtitle translators' processes was evaluated through both process data and the translators' subjective evaluations (Bywood *et al.* 2017, 496–497). MT and post-editing is also being investigated in the recently launched GoURMET project[4] as a way for creating multilingual news media.

While AD itself has received great research interest, use of technology in AD appears less common. The HBB4ALL project[5] explored various access services for television and video-on-demand, covering intralingual and interlingual subtitling with the help of ASR and MT (Matamala *et al.* 2015), audio access services like clean audio, AD and spoken subtitles, as well as sign language interpretation. Experiments on automating part of AD production have also been made. Szarkowska and Jankowska (2011) report on a study on using automatic text-to-speech

---

[1] http://sifnos.ilsp.gr/musa/.
[2] https://www.eu-bridge.eu/.
[3] http://www.fp7-sumat-project.eu/.
[4] https://gourmet-project.eu/.
[5] https://pagines.uab.cat/hbb4all/.

to replace the human voice in AD delivery. The ALST Linguistic and Sensorial Accessibility project has investigated MT and post-editing for multilingual voice-overs and AD of documentaries (Ortiz-Boix and Matamala 2016; Matamala and Ortiz-Boix 2016). The CineAd system developed by Campos *et al.* (2020) creates audio descriptions by analyzing the video script and the subtitles.

Verbal descriptions are also needed as metadata for archival and production purposes. Metadata are data about data that enable the identification of a resource (Pomerantz 2015, 73–74). One type of metadata is content description (CD): a more or less detailed description of a video's visuals that helps to find places, people, objects and actions appearing in the content. CD and AD share similarities, but their purposes and users are different. While AD is a service for partially sighted people to fully grasp audiovisual communication, CD serves to locate material in "audiovisual big data" (e.g. television archives) for the benefit of the media industry. The project described in this article aims to develop automatic methods for both CD and AD. This means machine-learning methods that convert visual information nested in video footage into machine-interpretable representations, and subsequently, converting these representations into verbal text. Automatic CD is developed in computer vision and natural language processing (see Aafaq *et al.* 2019). The interest in automatizing metadata production is in line with goals of the international broadcasting community (see EBU 2019).

3. The MeMAD project

The work discussed in this paper has been carried out within the European research project MeMAD (Methods for Managing Audiovisual Data: Combining Automatic Efficiency with Human Accuracy, grant nr 780069)[6]. The aim of this project has been to develop language-based manual, automatic and semi-automatic models and approaches to advance the accessibility of audiovisual content within Creative Industries, particularly television broadcasting and on-demand media. The project is a joint effort by researchers, broadcasting and audiovisual media organizations as well as companies developing language technology and digital production tools. MeMAD researches and develops solutions for speech-to-text applications, intralingual and interlingual subtitling, and for producing verbal-textual descriptions and structured metadata representations of visual and auditory multimedia content in multiple languages and for varied contexts and audiences. By combining manual and automatic methods, the project aims to combine the

---

[6] https://memad.eu/.

accuracy and rich understanding of humans with the efficiency of automatic methods.

One of the project partners is Yleisradio Oy (Yle), the Finnish public broadcasting corporation. Yle's programming and accessibility needs are the focus of this article. Public broadcasting is particularly relevant for media accessibility because of its broad service mission and legal status. Finnish law specifies the duties of Yle to include providing "versatile and comprehensive television and radio programming" to all, to foster "tolerance, equal treatment, equality, and cultural diversity and provide programming for minority and special groups" (Act on Yleisradio Oy 1993, c 3, s. 7). In order to fulfil these obligations, Yle must maintain accessibility as a key consideration. However, it is also worth noting that Yle's access services do not currently serve all audience groups equally. For example, Finnish legislation requires subtitling for the deaf and hard of hearing to be provided comprehensively on Yle programming, but no similar legal obligation exists for AD. Yle's AD offering is rather limited, less than 1% of total program hours in 2020[7]. On the other hand, audio subtitling is provided for broadcasts with interlingual subtitles. The situation highlights the limitations of a public broadcaster's ability to serve diverse populations and presents a rationale for initiatives like MeMAD to offer technological solutions that could facilitate the provision of higher volumes of access services. Since the work is conducted in Finland, accessibility solutions involve the country's official languages, Finnish and Swedish. This context introduces a particular challenge, as most existing technological solutions are primarily developed for more dominant languages such as English, and quality may be lower in other languages. Consequently, questions of quality arise in our use cases and illustrate the situation of less-resourced languages in technological development.

MeMAD follows a user-centered ethos (see User-Centered Design Basics), where the usability of tools is iteratively tested at different stages of the project with various user groups, and users' views are taken into account in further development decisions. The objective is to develop tools that are as usable as possible. In the early stages of the project, in 2018, discussions were held with selected media professionals to scope early reactions to the technologies proposed by the project. During the fall of 2019, a round of use case validations was conducted both within Yle and more broadly with an international group of external experts. At Yle, four workshops were formed with professionals representing different departments and areas of expertise to discuss use cases relevant to the participants. Ultimately, specific use cases were

---

[7] In comparison, BBC provides AD for 20% of its programs on different channels (BBC 2020).

targeted for evaluation based on the priorities identified, the maturity of different technologies, and expert review by the external collaborators' group.

In the next section, we describe four use cases which involve using

- ASR and MT to generate fully automated interlingual subtitles for viewers (Section 4.1),

- MT (and ASR) as support for interlingual subtitle translators (Section 4.2),

- ASR as support for intralingual subtitlers (Section 4.3), and

- automatic video description technologies to create metadata for audiovisual, nonlinguistic content as a support for video content management (Section 4.4).

Each of these cases addresses some aspect of accessibility. Interlingual subtitling and other multilingual content offer broader access across language and culture barriers to audiences who do not (sufficiently) understand the language of the original content. Intralingual subtitling supports users with hearing impairments and simultaneously benefits other users, like language learners and anyone who cannot use the auditory content (e.g. due to a noisy environment). Finally, video description technologies and automatic metadata can help production and archival staff with easier access to relevant content (e.g. by facilitating search for content in media archives) and ultimately provide audiences more efficient access to audiovisual materials relevant to their interests.

4. Exploring users' perspectives on technological access solutions

This section describes work involving technological and semi-automatic solutions for access services in four use cases and represents a snapshot of the situation in fall 2020. A detailed description of each user study is not within the scope of this article. Experiments in 2019 have been reported in Hirvonen *et al.* (2020). At the time of writing, analysis of the last experiments in 2020–2021 is ongoing, and will be reported in the final evaluation deliverable to be published in 2021. Because the technological solutions are at different stages of maturity, the individual studies have attempted to answer different questions related to usability, depending on priorities at the particular stage of development, and have involved different methods including keylogging and screen recording, questionnaires, as well as individual and focus group interviews. The purpose of this article is to explore themes raised by the participants in their own words through a qualitative approach. Therefore, we focus on

information collected in the individual and focus group interviews. The following subsections summarize key findings of the interview analyses in each case.

a. Viewers and machine-translated subtitles

Interlingual subtitles advance linguistic accessibility by making audiovisual content available to audiences across language barriers, and MT can increase the amount of translated content. MeMAD has been developing a fully automated subtitling solution where speech is transcribed through ASR, timecoded and segmented automatically, and then machine translated. Project partners are conducting audience research to explore the usefulness and usability of these automated subtitles. At the time of writing this article, two focus group sessions have been conducted, and a questionnaire study as well as two further focus group sessions are being organized in late 2020.

The focus group is a research method where a small number of participants discusses the topic of research, led by a moderator who steers the conversation with questions and comments. The participants are allowed to discuss the topic with each other, which can produce richer data than interviews where participants simply answer questions (Wilkinson 2006, 50–52). In the first focus groups, the participants were shown an approximately five-minute clip from the beginning of a current affairs program. One of the groups consisted of six Finnish-speakers who saw a Swedish-language video clip subtitled in Finnish. The other group consisted of seven English-speakers living in Finland who saw a Finnish-language clip subtitled in English. After viewing the clip, the groups were invited to answer the moderator's questions and discuss their opinions on the subtitles. The questions revolved around four broad themes: 1) comprehension of the clips, 2) cognitive load caused by the subtitles, 3) appreciation of and reactions towards the clips, and 4) usefulness and potential uses of automated subtitling. The discussions lasted approximately an hour, and they were conducted online via Google Meet.

The two language groups are in different positions in terms of linguistic accessibility. Whereas Finnish-speakers are served well by both Yle and commercial providers in Finland, both with local content and professionally translated international programming, English-speakers have difficulties accessing local content if they do not understand Finnish. Therefore, to Finnish-speakers, automated subtitles, which are still not close to reaching human quality, may look like a reduction in service, but for English-speakers they can be a welcome relief, a way to access local content that was previously inaccessible to them. Some potential use contexts do exist for the native population in niche content that may interest specific audience segments but not be popular enough to be

either translated into Finnish or available in widely understood languages such as English. In other words, automated subtitles can provide migrant populations access to their immediate surroundings and daily topics, and native populations access to remote cultures and marginal topics. As access to matters of daily concern has more urgency, the migrant population's interest seems a more significant access issue than the native population's access to less critical matters.

In this experiment of automated linguistic accessibility, trust emerged as a significant consideration. The participants' willingness and ability to trust the automated subtitles as a source of accurate information was mentioned in the discussions on several occasions. In this case, trust is closely related to quality, because automated subtitles have noticeable quality issues, which may make them appear less reliable than professional, high-quality subtitles. The question of quality was prominent in the focus group discussions, and both groups pointed out quality problems that eroded the credibility of the subtitles. This was also reflected in discussions concerning the usefulness of the subtitles. In the Finnish group, one participant suggested that automated subtitles could provide instant access to breaking international news, but another observed that poor subtitle quality may make the information unreliable. Similarly, in the English-speaking group, a participant reflected on the possibility of using automated subtitles to learn Finnish, but another participant argued that automated subtitles may not reliably reflect the meaning of the spoken Finnish and would be a problematic learning tool. Thus, while both groups acknowledged that automation can help audiences gain wider access, they also highlighted the importance of trust in making the access meaningful. If the user cannot trust the translations, the access the subtitles provide is limited to low-stakes contexts where potential misunderstandings do not carry significant consequences. Nevertheless, automated provision of linguistic access can benefit both local and migrant populations, as it opens up more options for participating in society and culture and for following one's interests.

b. Subtitle translators and post-editing machine-translated subtitles

While fully automatic subtitle translation could be envisioned in some use contexts, quality remains a concern. Therefore, the project has explored semi-automatic use of MT where a subtitle translator checks and corrects the automatic translations. The rationale for such post-editing workflows, in general, is that they increase productivity (e.g. Cadwell *et al.* 2018). For accessibility, this potentially means that more content could be translated, and that translated content could be made available sooner after broadcast of the original content. However, this

potential relies on sufficient quality of tools and outputs to support the translators' work.

Yle's current subtitling workflow involves translating "from scratch", meaning that the translators translate directly from the audio and timecode subtitles manually[8]. To investigate whether machine-translated and pre-segmented subtitles could offer benefits in the workflow, the project has carried out two rounds of tests (2019 and 2020) with twelve[9] subtitle translators working in-house at Yle or as freelancers. The participants subtitled approximately three-minute video clips (a total of six clips in each round) by post-editing machine-translated subtitles and tested different MT system outputs. The first round involved also process data collection with keylogging, and in the second round, task times were measured. Feedback was collected with questionnaires and semi-structured interviews. The tests have mainly involved machine translating human-created intralingual subtitles, although the second round included also speech-to-text with ASR and MT. The content involved unscripted election debates and semi-scripted cultural or lifestyle programming translated from Finnish into English and Swedish, and from English and Swedish into Finnish.

Findings of the experiments were mixed. Based on process data collected with keylogging in the first experiment round, post-editing made the translators' work slightly faster on average, but differences between language pairs and translators were considerable. The translators' assessments of the experience indicated that while they did not find post-editing particularly difficult (although all noted that it was unfamiliar to them in the first round), they tended to characterize it as somewhat annoying and limiting. The quality of the speech-to-text translations in the second round was deemed too poor to be useful: due to the frequency of errors, the translators stated that too much effort was required to correct these outputs and found the task very unpleasant and frustrating. However, most translators saw promise for post-editing machine-translated intralingual subtitles. In particular, they mentioned situations where quick translations of topical affairs are needed as a potential scenario.

The translators conceptualize MT and post-editing as a potential tool for accessibility: the practice could promote access to more content, sooner. However, also in this case trust was a recurring theme. Sometimes errors and "oddities" encountered in the output made it difficult to fully trust the MT. In contrast, trust was also deemed dangerous. Seemingly fluent and reasonable translations could in fact

---

[8] Subtitling templates, which are common in AVT, are not used by Yle.
[9] Two translators were unavailable for the second round and participated only in the first.

hide considerable errors with regard to meaning, which the translators worried might remain unnoticed in a hurry. In contrast, one translator noted that when the MT could be trusted, post-editing was indeed quite enjoyable and helped feed creativity.

c. Subtitlers and post-editing automatic intralingual subtitles

ASR is already widely used to produce intralingual subtitles, whether without any human intervention or through respeaking or manual post-editing. However, in many smaller languages fully automated subtitling through ASR is not feasible. MeMAD project partners have developed ASR technology specifically for Finnish and Swedish (particularly the variety of Swedish spoken in Finland). In these languages, ASR is not common for subtitling, as suitable tools with sufficient quality have not been readily available, although "offline" respeaking, i.e. respeaking short segments and pausing the video while correcting the subtitles, is used by Yle. The project has explored semi-automatic approaches where the subtitles are produced by ASR, segmented automatically, and then checked and corrected by subtitlers. Similarly to the previous case, the rationale is that automation would increase subtitlers' productivity, thereby allowing for more material to be subtitled with the same resources.

To test whether ASR could improve productivity, two rounds of evaluations (2019 and 2020) have been conducted with four intralingual subtitlers[10], along with a proof-of-concept period where subtitlers used the technology in production. All participants were Yle's in-house subtitlers. Their current workflows involve making subtitles from scratch, either manually or with the help of offline respeaking. All participants had at least some experience in offline respeaking, though only one had used it extensively. That same participant had also previously tested post-editing ASR subtitles. In the experiments, the participants subtitled approximately three-minute video clips (a total of seven clips in the first round and three clips in the second round). The experiments were carried out on Finnish content produced by Yle, and two different genres: unscripted election debates (both rounds) and semi-scripted lifestyle or cultural programming (first round only). Like in the MT post-editing case, questionnaires and semi-structured interviews were used to collect feedback. Keylogging was used in the first round to compare subtitling from scratch, subtitling with respeaking, and post-editing ASR subtitles. Task times were measured in the second round.

---

[10] One participant was unavailable for the second round of testing and participated only in the first.

While only one out of four participants in the first round clearly improved productivity when post-editing subtitles, three out of four found the experience very positive and potentially helpful for their work. Based on this feedback, the three subtitlers used ASR and automatic timecoding as part of their normal workflow for a proof-of-concept period during the summer of 2020 and reported their observations and impressions. A second round of productivity tests to examine the effect of practice was conducted in the fall of 2020.

Quality of the ASR output emerged as a key topic in the interviews. In the first round, the participants discussed ASR errors and commented on problems with the automatic segmentation, which caused them some frustration. Opinions differed somewhat, with some participants assessing the overall quality more positively than others. They also noted that quality varied between the clips and genres; for example, quality was deemed lower in clips with more colloquial language. One of the participants also noted that small errors in the ASR output were sometimes very difficult to detect. The subtitlers felt that the technology might limit their creativity, which could affect the quality of the subtitles. In the interviews after the second round of testing, the subtitlers commented positively on the quality of the ASR developed by the project, and all three participants were eager to use ASR as part of their daily work.

d. Production and archival staff using automatic description of visual content

In this subsection, we describe research on the production and use of CD in the work of archival and production staff. The first rounds of user interviews and experiments focused on charting the perspectives and needs of the different groups of media professionals.

To understand media professionals' perspective on the purpose and use of content descriptions, we first conducted interviews in 2019 with six professionals in charge of CD at Yle: three production coordinators who write content descriptions, and three archive editors who edit them. The interviews highlight the purpose of CD as enabling the effective reuse of content. The descriptions entail different levels of detail depending on their reuse value: programs with complete reuse rights should be described in the most comprehensive manner. The level of detail also needs to be balanced: descriptions that are either too detailed or too sparse make effective searches difficult. These distinctions suggest that different types of output may also be needed from technological solutions. As a first step, machine-generated CD might be most useful for content with low reuse value. The interviewees also discussed quality in terms of the relevance of information, noting that the human user needs

background knowledge and understanding of the context to determine what is relevant.

Experiments and interviews in 2019 explored how production and archival staff use different types of automatic metadata. In one test, three video editors completed an editing task; in the other, journalists and archivists (five people in total) used the metadata for searching and browsing tasks. Automatic CD was not included in this test due to the immaturity of the technology. Instead, the metadata comprised automatically generated transcripts of the dialogues in the original language (French and Finnish) and in translation into English, with tags referring to names of people, organizations and places that appear in the spoken content identified through named-entity recognition. Additionally, automatic face recognition to identify people could be used in the editing case, whereas the search case was provided with manually generated CD. In interviews following these tasks, the participants were asked about their experience with the access provided and about the task and tools.

The participants were content with the rapidity and ease of finding material as long as the search was successful. However, a recurring topic raised by them was that the software did not work properly, or they were unfamiliar with it. It was deemed important to ensure that the users understand the "logic" behind the technology before they can benefit from its functionalities. Especially the possibility to search content via translated speech was considered as helpful although the risk of errors in MT was also mentioned. A further positive aspect was the quantity of data which makes the searching more versatile and gives more results and therefore possibilities. On the negative side is the risk of the machine producing too much and irrelevant information. With regard to quality, the participants focused mainly on the software and mentioned to some extent the quality of the automatic transcripts. Despite the absence of automatic CD, the professionals mentioned this type of description could be beneficial in both work settings. Improved user experiments including machine-generated CD are being conducted at the time of writing the article in 2020–2021.

5. Lessons from the users: trust, quality and limits of technology

The cases discussed above each address a specific user group, and a different perspective to accessibility, usability and technology. In the first case, we discussed end users of audiovisual content: viewers who need interlingual subtitles to understand the program they are watching. For them, access is facilitated by high and consistent quality of the subtitles. They need to be able to trust that the subtitles accurately reflect the contents of the program. The nature of the technology used to

produce the subtitles is less important to them. The users' comments in this case are similar to findings in other studies exploring the use of MT for information purposes: quality and trust are recurring themes. Trust in translation is of course a wider issue. Although literature on the topic has mainly focused on interpersonal or sometimes organizational trust, Moorkens and Rocchi (2021, 328–330) suggest that the increasing use of technology in the translation industry means that trust in technology should be combined with these more traditional, interpersonal conceptualizations of trust. With regard to quality, other projects have also observed that users may accept varying levels of quality in different contexts and for different purposes. For example, lower quality content may be preferred if the alternative is no content at all (see Nurminen and Koponen 2020; Szarkowska *et al.* 2015), which was a sentiment echoed in the English-language group in this case.

In the second case, we see translators as users of technology while engaged in producing an access service. While the subtitles created by the translators using the automated "raw material" ultimately serve the needs of the end users that were at the center of the first case, the subtitle translators have needs and interests of their own. The translators see (semi-)automation as facilitating audience access to more information and providing access more quickly, if they can act as effective intermediaries in this process. In order to do that, the translators need sufficiently high and consistent quality of automated output to allow them to produce a professional end result. The technology itself also needs to be reliable and suitable for easy adoption into the subtitling process. The professional intermediaries generally emphasize quality and trust in order to use automated output effectively. Bywood *et al.* (2017) similarly discuss the importance of quality and attention to user perspectives when integrating MT into subtitling workflows. More generally, quality, (mis)trust and perceptions of creativity have been found to be key factors affecting translators' acceptance of MT as a tool (see Cadwell *et al.* 2018). The translators also appeared particularly bothered by the effect of technology on their creativity, although more research would be needed to study the questions of trust and creativity.

The third case also focused on professional intermediaries, this time in intralingual subtitling. The findings are more optimistic than in the first two cases. The ASR technology is more mature than MT, and the experiment was successful enough to progress to the proof-of-concept phase where subtitlers started using ASR in their daily work. Although the intralingual subtitlers also discussed quality and errors, trust did not emerge as an equally significant theme as for the translators. One reason may be that ASR reproduces the spoken source text in a more mechanical fashion than interlingual MT, which by nature involves a more

interpretive dimension. For intralingual subtitling, the interpretive aspect is more prevalent in editing and condensing speech into subtitles. In addition, accuracy is easier to assess, because ASR only transcribes the spoken words, so intralingual subtitlers can check the ASR against the spoken language more easily than in the interlingual scenario. Therefore, less trust is needed as both source and target content is immediately available to the users.

In the fourth case, the participants were again content producers, but they are not providing direct access services to end users. Rather, their role is either to manage audiovisual content as "raw material" to be reused by other media professionals or to reuse that material themselves in journalistic work. The experiment focused on the audiovisual elements of content, for which the conversion of audiovisual data into verbal text data is essential. Overall, from the perspective of production and archival staff, an important driver of accessibility in this case is the reliability and relevance of the technology and software, which have to adhere to the logic of their work processes in order to be usable. Although the study did not test automatic CD in this round, the relevance of having textual access to the visual content was brought up as a key feature of metadata.

While ASR has reached a maturity level sufficient for integration into intralingual subtitling workflows, and MT was ready for user tests involving interlingual subtitling (although not yet for actual production), technology for automatic description of the visual content remains at a more preliminary stage. Usefulness of technological support for audio description remains an open question. As became evident in analyses of automatic video descriptions, this technology is not easily compatible with audio description (see Braun and Starr 2019). Semi-automatic solutions for supporting audio description could, however, offer benefits for both end users and professionals and this direction would therefore merit further research (see also Remael *et al.* 2016). It would also be interesting to explore whether and how the concepts of quality, trust and creativity emerge in the users' perspectives. In contrast to the other cases, the interpretive aspect again comes in a slightly different stage of the process when creating verbal descriptions of visual information. Intersemiotic translation requires interpretation and fundamental decisions of what is relevant to be verbalized, whereas in intralingual and interlingual translation these decisions already exist to a larger extent in the verbal source text.

6. Discussion and concluding remarks

The example cases discussed in this article demonstrate how many meanings accessibility has for different user groups at various stages of

the production process. For the end user, accessibility means access to the audiovisual product, whether by crossing a linguistic boundary or by overcoming limitations of auditory or visual perception. For translators and subtitlers, facilitating access for the audience is the purpose of their work, but they also benefit from efficient access to reliable material with the help of automation. For audiovisual content describers and users, automation provides intermediary metadata which allows them to process the audiovisual content more efficiently and accurately. For the professionals, access comes from well-functioning interfaces that facilitate their work processes and from information about the content (metadata) that is relevant to them. Additionally, the development of technologies is at different stages for different access services and different professional groups. Whereas intralingual subtitlers can already benefit from speech-to-text tools in their work processes, machine translation still needs further development to be genuinely useful, and tools for intersemiotic translation are at an even earlier stage of development.

Our use cases suggest that technological innovations can increase access by lowering barriers of language and perceptual limitation more efficiently than human action alone, and by increasing the productivity of professionals dealing with audiovisual content. Automatic and semi-automatic solutions can be helpful in tackling the vast amounts of media content that would otherwise remain inaccessible, as suggested by Georgakopoulou (2012). However, that increased access is by no means a given. New tools and approaches need to be accepted by their users and tailored to users' interests in order to make a genuine difference. The MeMAD cases demonstrate how crucial the input from different user groups is in identifying potential weaknesses of the solutions, envisioning use contexts and revealing users' approaches and attitudes when encountering new technologies and negotiating access.

In most cases, trust emerges as a question that determines the usability of these tools. The quality of the automated output can be too low to be trusted, the tools themselves may need further development in order to win the users' trust, and getting used to new work processes may take time. Trust may also be eroded by seemingly high quality, if surface-level fluency risks making errors harder to detect. On the other hand, inability to rely on the quality may lead to an even more insidious type of mistrust, as it may affect users' attitudes towards everything the system produces. This may erode their trust in not only the technology, but also the information provided to them, which in turn can damage trust in the provider of that information (cf. Nurminen and Koponen 2020). Although perspectives of trust vary in the different cases, the question in all is fundamentally the same: Can users trust that automation produces reliable, accurate and useful results? Therefore, the

process of introducing new technologies and making them genuinely usable is a delicate one, and transparency about the process is important in building trust. We can also see that when trust ceases to be a constant concern, as was the case with the intralingual subtitlers, technological advances may be adopted rather quickly. As suggested by Moorkens and Rocchi (2021, 330), further research could also benefit from applying models which have been used in other contexts to investigate trust in technology.

To understand quality and usability in accessibility, it is important to conduct design and development processes with the participation of diverse user groups that represent different perspectives and positions, as Greco and Jankowska (2020, 8) recommend. The MeMAD cases constitute one step in that direction, and offer examples of user-centered development. However, further research is needed to understand more comprehensively the use of technology and automation in media access services. One promising direction for future work would be the interplay of accessibility and usability. While the two concepts should not be conflated (Suojanen *et al.* 2015, 56–58), they are intertwined: in order to achieve genuine accessibility, solutions and services must be usable, and usability must include considerations of accessibility to achieve its objectives, i.e. the efficient and enjoyable use of products and solutions. Furthermore, further research on applying usability to the context of communication would be valuable. For example, UCT (Suojanen *et al.* 2015) and UCD could provide even more useful tools for exploring the user perspective in media accessibility.

It is also crucial to continue to investigate the interaction of humans and technology in the context of access services. MeMAD's studies reveal that any production phase involving creativity or decision-making is difficult to hand over to technology, and the more challenging and open-ended the decision-making is, the more difficult the process is to automate. Yet, automation provides the kind of efficiency that is recognized by professionals and end users alike as important in ensuring access when volumes of audiovisual content keep increasing beyond the ability of humans to process on their own. Therefore, research should aim to find the appropriate balance that allows humans to rely on and trust technological solutions, and gain the most benefit from them.

References:

Aafaq, Nayyer, Mian, Ajmal, Liu, Wei, Gilani, Syed Zulqarnain, and Shah, Mubarak. 2019. Video description: A survey of methods, datasets and evaluation metrics. In: *ACM Computing Surveys*. 52(6): article no. 115.

Act on Yleisradio Oy. 1993. https://finlex.fi/en/laki/kaannokset/1993/en19931380.pdf. Accessed on: 15 September 2020.

BBC. 2020. Access services at the BBC. https://www.bbc.com/aboutthebbc/whatwedo/publicservices/accessservices. Accessed on: 8 January 2021.

Braun, Sabine, and Starr, Kim. 2019. Finding the right words: Investigating machine-generated video description quality using a corpus-based approach. In: *Journal of Audiovisual Translation*. 2(2): pp. 11–35.

Bywood, Lindsay, Georgakopoulou, Panayota, and Etchegoyhen, Thierry. 2017. Embracing the threat: Machine translation as a solution for subtitling. In: *Perspectives: Studies in Translatology*. 25(3): pp. 492–508. https://doi.org/10.1080/0907676X.2017.1291695. Accessed on: 15 December 2020.

Cadwell, Patrick, O'Brien, Sharon, and Teixeira, Carlos S.C. 2018. Resistance and accommodation: Factors for the (non-) adoption of machine translation among professional translators. In: *Perspectives: Studies in Translatology*. 26 (3): pp. 301–321. https://doi.org/10.1080/0907676X.2017.1337210. Accessed on: 15 December 2020.

Campos, Virginia P., de Araújo, Tiago M. U., de Souza Filho, Guido L., and Gonçalves, Luiz M. G. 2020. CineAD: A system for automated audio description script generation for the visually impaired. In: *Universal Access in the Information Society*. 19(1): pp. 99–111. https://doi.org/10.1007/s10209-018-0634-4. Accessed on: 15 December 2020.

Del Pozo, Arantza, Aliprandi, Carlo, Álvarez, Aitor, Mendes, Carlos, Neto, Joao P., Paulo, Sérgio, Piccinini, Nicola, and Raffaelli, Matteo. 2014. SAVAS: Collecting, annotating and sharing audiovisual language resources for automatic subtitling. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*. pp. 432–436.

Di Giovanni, Elena. 2018. Audio description and reception-centred research. In: Di Giovanni, Elena, and Gambier, Yves (eds.). *Reception Studies and Audiovisual Translation*. Amsterdam: John Benjamins. pp. 225–250.

EBU (= The European Broadcasting Union). 2019. AI and automatic metadata extraction. https://tech.ebu.ch/groups/ame#goals. Accessed on: 18 September 2020.

Georgakopoulou, Panayota. 2012. Challenges for the audiovisual industry in the digital age: The ever-changing needs of subtitle production. In: *The Journal of Specialised Translation* 17: pp. 78–103. http://www.jostrans.org/issue17/art_georgakopoulou.pdf. Accessed on: 15 December 2020.

Greco, Gian Maria. 2016. On accessibility as a human right, with an application to media accessibility. In: Matamala, Anna, and Orero, Pilar (eds.). *Researching Audio Description: New Approaches*. London: Palgrave Macmillan. pp. 11–33. https://doi.org/10.1057/978-1-137-56917-2. Accessed on: 20 December 2020.

Greco, Gian Maria. 2018. The nature of accessibility studies. In: *Journal of Audiovisual Translation* 1(1): pp. 205–232. http://jatjournal.org/index.php/jat/article/view/51. Accessed on: 15 December 2020.

Greco, Gian Maria, and Jankowska, Anna. 2019. Framing media accessibility quality. In: *Journal of Audiovisual Translation* 2(2): pp. 1–10. http://www.jatjournal.org/index.php/jat/article/view/114/28. Accessed on: 20 December 2020.

Hirvonen, Maija, and Kinnunen, Tuija. 2021. Accessibility and linguistic rights. In: Koskinen, Kaisa, and Pokorn, Nike K. (eds.). *The Routledge Handbook of Translation and Ethics*. Abingdon/Oxon: Routledge. pp. 470–483.

Hirvonen, Maija, Koponen, Maarit, Saarikoski, Lauri, Sulubacak, Umut, Vitikainen, Kaisa, and Van Rijsselbergen, Dieter. 2020. D6.6 Evaluation report, intermediate version. Technical report, MeMAD Project. https://memad.eu/deliverables/. Accessed on: 15 December 2020.

Jankowska, Anna. 2020. Audiovisual media accessibility. In: Angelone, Erik, Ehrensberger-Dow, Maureen, and Massey, Gary (eds.). *The Bloomsbury Companion to Translation Industry Studies*. London: Bloomsbury Academic. pp. 231–259.

Matamala, Anna, Oliver, Andreu, Álvarez, Aitor, and Azpeitia, Andoni. 2015. The reception of intralingual and interlingual automatic subtitling:

An exploratory study within the HB4ALL project. In: *Proceedings of Translating and the Computer* 37: pp. 12–17.

Matamala, Anna, and Ortiz-Boix, Carla. 2016. Accessibility and multilingualism: An exploratory study on the machine translation of audio descriptions. In: *Trans* 20: pp. 11–24. https://doi.org/10.24310/TRANS.2016.v0i20.2059. Accessed on: 10 December 2020.

Melero, Maite, Oliver, Antoni, and Badia, Toni. 2006. Automatic multilingual subtitling in the ETITLE project. In: *Proceedings of Translating and the Computer* 28: pp. 1–18.

Moorkens, Joss, and Rocchi, Marta. 2021. Ethics in the translation industry. In: Koskinen, Kaisa, and Pokorn, Nike K. (eds.). *The Routledge Handbook of Translation and Ethics*. Abingdon/Oxon: Routledge. pp. 320–337.

Nurminen, Mary, and Koponen, Maarit. 2020. Machine translation and fair access to information. In: *Translation Spaces* 9(1): pp. 150–69. https://doi.org/10.1075/ts.00025.nur. Accessed on: 15 December 2020.

Ortiz-boix, Carla, and Matamala, Anna. 2016. Post-editing wildlife documentary films: A new possible scenario? In: *Journal of Specialised Translation*. 26: pp. 187–210. http://www.jostrans.org/issue26/art_ortiz.pdf. Accessed on: 10 December 2020.

Piperidis, Stelios, Demiros, Iason, Prokopidis, Prokopis, Vanroose, Peter, Hoethker, Anja, Daelemans, Walter, Sklavounou, Elsa, Konstantinou, Manos, and Karavidas, Yannis. 2004. Multimodal multilingual resources in the subtitling process. In: *Fourth International Conference on Language Resources and Evaluation (LREC 2004)*: (no pagination).

Pomerantz, Jeffrey. 2015. *Metadata*. (The MIT Press Essential Knowledge Series). Cambridge, MA: The MIT Press.

Remael, Aline, Reviers, Nina, and Vandekerckhove, Reinhild. 2016. From translation studies and audiovisual translation to media accessibility. In: *Target* 28(2): pp. 248–260. https://doi.org/10.1075/ target.28.2.06rem. Accessed on: 10 December 2020.

Suojanen, Tytti, Koskinen, Kaisa, and Tuominen, Tiina. 2015. *User-Centered Translation.* London: Routledge.

Szarkowska, Agnieszka, and Jankowska, Anna. 2011. Text-to-speech audio description: towards wider availability of AD. In: *Journal of Specialised Translation* 15: pp. 142–162.

Szarkowska, Agnieszka, Pietrulewicz, Joanna, and Jankowska, Anna. 2015. Long questionnaire in Poland. In: Romero-Fresco, Pablo (ed.). *The Reception of Subtitles for the Deaf and Hard of Hearing in Europe: UK, Spain, Italy, Poland, Denmark, France and Germany*. Bern: Lang. pp. 45–74.

The EU-BRIDGE Consortium. 2015. *EU-BRIDGE Technology Catalogue*, 2nd edition, January 2015. http://project.eu-bridge.eu/downloads/fseu bridge tech20150311screen.pdf. Accessed on: 18 September 2020.

User-Centered Design Basics. https://www.usability.gov/what-and-why/user-centered-design.html. Accessed on: 15 September 2020.

Wilkinson, Sue. 2006. Analysing interaction in focus groups. In: Drew, Paul, Raymond, Geoffrey, and Weinberg, Darin (eds.). *Talk and Interaction in Social Research Methods*. London/Thousand Oaks/New Delhi: SAGE Publications. pp. 50–62.