



**UNIVERSITY
OF TURKU**

This is a self-archived – parallel-published version of an original article. This version may differ from the original in pagination and typographic details. When using please cite the original.

AUTHOR	Tuominen Tiina
TITLE	Exploring the gaps in linguistic accessibility of media: The potential of automated subtitling as a solution
YEAR	2023
VERSION	Publisher's PDF
CITATION	Koponen, M., Vitikainen K., Sulubacak, U. and Tiedemann J. (2023) 'Exploring the gaps in linguistic accessibility of media: The potential of automated subtitling as a solution', <i>JoSTrans : the Journal of Specialised Translation</i> , 39, pp. 77-98. Available at: https://jostrans.org/issue39/art_tuominen.php
LICENSE	CC BY

Exploring the gaps in linguistic accessibility of media: The potential of automated subtitling as a solution

Tiina Tuominen, University of Turku

Maarit Koponen, University of Eastern Finland

Kaisa Vitikainen, University of Helsinki and Yle

Umut Sulubacak

Jörg Tiedemann, University of Helsinki

ABSTRACT

Linguistic accessibility presents a challenge for public broadcasters. While demand for multilingual content to support accessibility is growing, limited resources do not allow translating to a wide range of languages. One possibility to increase linguistic accessibility in media would be to provide automated translations into languages that cannot be served otherwise. However, implementing automation with the objective of supporting linguistic accessibility requires careful, proactive investigation together with the prospective audience. This article explores automated interlingual subtitling from the audience's perspective, based on focus group discussions and an online survey conducted in association with the Finnish public broadcaster Yle. We investigated English-speaking viewers' reactions to automated English subtitles in Finnish-language news and current affairs video clips. Our analysis indicates that while viewers are able to understand the gist of a programme with automated subtitles, shortcomings in the quality of automated speech recognition, translation and timing of subtitles result in significant cognitive load that limits the usability of the subtitles. Participants expressed clear interest in automated subtitles for breaking news and other important local content, as well as to facilitate access to local culture and entertainment. However, quality improvements are required before automated subtitles can be deployed.

KEYWORDS

Linguistic accessibility, public broadcasting, audiovisual translation, subtitling, machine translation, reception research, focus group, questionnaire, cognitive load.

1. Introduction

Public broadcasters face a great responsibility in providing services to a wide and diverse audience. One of the cornerstones of public services is accessibility. While accessibility is often seen as services to people with disabilities (e.g. visual or hearing impairments), it also includes linguistic accessibility, which means content needs to be provided in multiple languages (Hirvonen and Kinnunen 2021). However, linguistic accessibility is limited by necessity, as public broadcasters do not have the resources to make all their content multilingual. One possible way to approach linguistic accessibility in media could be to use technology like automatic speech recognition (ASR) and machine translation (MT) to provide automatic interlingual subtitles for audiovisual content.

This paper examines automatic interlingual subtitling as a potential solution for linguistic accessibility in the context of the Finnish public broadcaster Yle. At the end of 2020, there were over 432,800 people living in Finland who did not speak one of the national languages (Finnish, Swedish or Sami)

as their native language (Statistics Finland n.d.), a significant number in a country with a population of just over 5.5 million. While some in this population also speak Finnish, most of the Finnish media remains inaccessible for many. Yle is required by law to provide services in Finnish, Swedish, Romani, Sami and the Finnish sign language, “and in the languages of other language groups in the country, where applicable” (Laki Yleisradio Oy:stä; translation by authors). Yle provides news content in the required languages, as well as English and Russian, on a regular basis. Early in the COVID-19 pandemic, Yle also provided some COVID-19 related news in Arabic, Somali, Kurdish and Persian/Dari for a limited time. However, the scope of news in languages other than Finnish and Swedish is limited. A proactive approach to identifying new technologies that improve linguistic accessibility could also benefit Yle. The availability of content in languages other than Finnish and Swedish could be drastically increased if automatic translations of sufficiently high quality could be provided.

The research discussed in this paper was part of the EU-funded research project MeMAD (Methods for Managing Audiovisual Data: Combining Automatic Efficiency with Human Accuracy, grant no. 780069). One part of the project, carried out mainly as collaboration between Yle and University of Helsinki, focused on automatic interlingual subtitling as a means to offer broader access to audiences who do not (sufficiently) understand the language of the original content. For this purpose, the project developed a pipeline for automatic interlingual subtitling (see Laaksonen *et al.* 2021). Automatically generated subtitles were evaluated in post-editing experiments with professional subtitle translators (see Koponen *et al.* 2020a, 2020b), and in audience reception studies consisting of focus group discussions and an online survey. Preliminary observations from the reception studies have been reported in Koponen *et al.* (2020c) and Braun *et al.* (2021). This article expands on those observations and investigates whether viewers think automated subtitles could support linguistic accessibility. We also examine the key considerations arising from the viewers’ perspectives and their potential implications for future development and implementation of automated interlingual subtitling, especially in a public broadcasting context.

This paper is structured as follows. Section 2 presents an overview of related work on automatic interlingual subtitling and audience reception. Section 3 describes the reception study design and methodology as well as materials and tools used in the study. In section 4, we present the analysis of our findings, followed by conclusions in section 5.

2. Background: automatic interlingual subtitling and audience reception

Automatic subtitle translation has been explored since the 1990s (e.g. Popowich *et al.* 2000; Piperidis *et al.* 2004; Volk *et al.* 2010), but has received increased research interest in recent years following developments

in neural MT and automatic speech translation. Much of early work focused on machine translating intralingual subtitles (or closed captions) created by humans, although some initiatives like the MUSA project combined MT with ASR (Piperidis *et al.* 2004). While text-based MT of intralingual subtitles has also been examined in more recent work (e.g. Bywood *et al.* 2017; Matusov *et al.* 2019; Koponen *et al.* 2020a, 2020b), research has been increasingly turning towards automatic speech translation and subtitling (e.g. Di Gangi *et al.* 2019; Karakanta *et al.* 2020). Some have suggested that subtitles are particularly suited for MT because they generally consist of short sentences and relatively simple structures (Popowich *et al.* 2000; Volk *et al.* 2010). Volk *et al.* (2010) present a successful example of implementing a system for machine translating subtitles in the workflows of a subtitling company. However, the closely related language pairs (Swedish into Norwegian and Danish) made the situation particularly favourable (Volk *et al.* 2010: 56; see also Bywood *et al.* 2017: 499-500 for observations regarding different language pairs).

Conversely, some subtitling features may be challenging for automation. Subtitles are intended as a written representation of speech, and therefore often contain idiomatic and colloquial expressions, grammatical irregularities, hesitations, interruptions and ellipsis which have been found problematic for MT (see Popowich *et al.* 2000; Burchardt *et al.* 2016; Bywood *et al.* 2017). Subtitles also involve different genres covering a wide range of domains, and data from one genre may not be directly applicable to another (Burchardt *et al.* 2016). For example, automatic subtitle translation may struggle more with unscripted broadcasts compared to scripted dialogue (Bywood *et al.* 2017) or content involving creative use of language, such as comedies (Volk *et al.* 2010; Matusov *et al.* 2019). Features like unscripted speech also pose difficulties for ASR when used to create intralingual subtitles (e.g. Vitikainen and Koponen 2021), which can be another source for accuracy errors in the final translated subtitles.

Difficulties are also caused by technical restrictions regarding the number of characters displayed on screen and the display speed. For example, the Finnish guidelines specify that subtitles should contain one or two lines, and the subtitle speed should be no greater than 12-14 characters per second (Käännöstekstitysten laatusuosituksset 2020). This generally requires condensing the speech, which verbatim transcriptions generated by ASR cannot provide (Karakanta *et al.* 2020: 210). Segmenting subtitles along syntactic and semantic boundaries to support readability is also challenging to automate (Matusov *et al.* 2019: 85). Based on their comparison of different approaches, Karakanta *et al.* (2020) propose that end-to-end systems, which generate interlingual subtitles directly from the speech rather than through an intermediate ASR transcript, have promise for producing subtitles that meet the guidelines.

Further complications are added by the audiovisual context. Subtitling is increasingly conceptualised as a multimodal activity, where meaning is

constructed in an interplay between dialogue and other visual and auditive modes (Pérez-González 2014: 185). In this view, subtitling is an interpretive process, for which all modes are necessary meaning-making resources. Therefore, presenting the verbal dialogue as the sole source text in subtitling can be criticised as reductive. Although research has started to explore multimodal MT solutions that also incorporate video information (see Sulubacak *et al.* 2020 for an overview), accounting for the multimodal context as a whole remains a challenge.

Prior work has examined the usability of MT as a tool for subtitle translators (e.g. Volk *et al.* 2010; Bywood *et al.* 2017; Matusov *et al.* 2019; Koponen *et al.* 2020a, 2020b), but the experience of potential viewers and reception of fully automatic subtitle translation appears relatively unexplored. Reception of subtitles, in general, is a growing research area, where a body of studies is accumulating and “providing significant and stimulating insights into the world of the receivers, their preferences and needs” (Di Giovanni and Gambier 2018: x). Studies have observed, for example, that the experience and cognitive load of viewing subtitles is affected by synchrony between the audio and subtitles (Lång *et al.* 2013), the presentation speed of subtitles (Szarkowska and Gerber-Morón 2018), and overall complexity of the content (Perego *et al.* 2018). Most research on reception has involved human-created subtitles, and so far, few studies have addressed fully automatic interlingual subtitling.

Armstrong *et al.* (2007) and Flanagan (2009) examined the reception of intralingual movie subtitles translated from English into German using an example-based MT system. The experiment reported by Armstrong *et al.* (2007) involved rating MT subtitles for intelligibility and accuracy, interviews, and an online survey comparing the acceptability of unedited MT and post-edited movie subtitles. The initial evaluations and interviews suggested that machine-translated subtitles could offer some benefit for viewers who did not understand the source language, but would require post-editing to be usable in any commercial setting (Armstrong *et al.* 2007: 174-176). In the final survey, most of the 12 respondents deemed raw MT unacceptable for any purpose, although findings reported by Armstrong *et al.* (2007) vary depending on the movie clip and MT version. The highest acceptance rate was five positive answers for the scenario of watching a pirated DVD, whereas only one participant considered MT acceptable for watching a purchased DVD (Armstrong *et al.* 2007: 177-179). Building on this work, Flanagan (2009) investigated the reception of machine-translated subtitles in a survey addressing intelligibility and acceptability. Flanagan (2009: 247) found that MT subtitles “were deemed intelligible and acceptable to a certain degree,” but the majority of the respondents (60-80% depending on the MT version) were not willing to use such subtitles to watch a movie in a language they did not understand. Respondents were more accepting if they understood the soundtrack language (Flanagan 2009: 250).

Matamala *et al.* (2015) conducted an experiment where Spanish students with varying levels of English skills viewed short English news clips without any subtitles, with intralingual subtitles generated by ASR, and interlingual subtitles machine translated into Spanish. Based on a comprehension questionnaire, participants with low English proficiency benefited somewhat from the MT subtitles, although overall their comprehension levels remained low, while in the group with intermediate proficiency, MT subtitles seemed to affect comprehension negatively (Matamala *et al.* 2015: 15-16). Hu *et al.* (2020) used eye tracking and questionnaires to examine viewers' cognitive processing, comprehension and attitude towards English-to-Chinese MT subtitles for a Massive Open Online Course. Raw MT subtitles generally received lower comprehension scores and involved more cognitive processing compared to post-edited and human-translated subtitles, and attitudes towards raw MT were more negative than towards post-edited subtitles (Hu *et al.* 2020: 532-533). Unexpectedly, the respondents' attitude was most negative towards the subtitles translated by a human, but Hu *et al.* (2020: 533) note that this may have been affected by the translator not being a professional.

The present article adds to research on the reception of automatic interlingual subtitling of news and current affairs content by investigating the audience's experience of comprehension, cognitive load and acceptability. Through an analysis of focus group and survey data, the article contributes to the understanding of what factors may motivate viewers to use automated subtitles, and what obstacles may limit their use.

3. Study design

The reception study consisted of three phases. First, we conducted a focus group discussion to gauge initial views and reactions of a small group of viewers. Then, we designed a questionnaire based on experiences from the focus group to collect more quantitative data from a larger number of respondents. Finally, we conducted another focus group to develop a deeper understanding of the topics emerging from the questionnaire. All three stages explored the following questions:

1. How comprehensible are video clips with raw machine-translated subtitles that have been created with MeMAD tools?
2. How acceptable are raw machine-translated and fully automated subtitles to viewers? What are some key factors that determine subtitle quality and acceptability from viewers' point of view?
3. How much cognitive load do raw machine-translated subtitles cause for viewers in comparison to professionally produced subtitles?
4. What use contexts do viewers envision for raw machine-translated subtitles (Braun *et al.* 2021: 78)?

The following subsections describe first the materials and tools used for the experiments and then the methodology and set-up for the focus groups and questionnaire.

3.1. Video material and automatic subtitle generation

Both the focus group discussions and the questionnaire were structured around short video clips of Finnish-language news and current affairs programmes that were automatically subtitled in English. The first focus group was shown one five-minute clip from the beginning of a current affairs programme. In the questionnaire, the respondents were shown two three-minute clips: a news broadcast segment and a shortened version of the current affairs clip from the first focus group. The second focus group was shown the same two clips as were used in the questionnaire, although with slightly different subtitles (see below).

An in-depth discussion of the interlingual subtitle generation pipeline¹ is not within the scope of this article, rather, we provide a brief top-level view of the approach. First, we used the ASR system by the project partner Lingsoft combined with automatic timecoding and segmentation by the project partner Limecraft to automatically generate intralingual Finnish subtitles, which were then pre-processed to reconstruct sentences, standardise punctuation and casing, and fix some ASR features (e.g. abbreviations) which caused recurring translation problems. For MT into English, we used a subtitle translation model based on the transformer implementation of Marian² and trained on all available Finnish-to-English data from OPUS³ (excluding a small development set sampled from OpenSubtitles). Finally, machine-translated sentences were post-processed and fit back into the original timed segments. For details, see Laaksonen *et al.* (2021).

For the first focus group and the questionnaire, subtitles were generated without any human input. It was apparent from the quality of the subtitles that they were not professionally created. The subtitles were frequently out of synchrony with the spoken dialogue, and their segmentation did not follow sentence structures, making them challenging to follow and creating a sense of rush, even though the duration and display speed of the subtitles were largely within the norms for professional subtitles. In addition, some of the language was awkward and confusing, and there were occasional mistranslations or other distortions of meaning. Nevertheless, the subtitles provided the gist of the clips, and it was possible to follow the narrative with their help.

For the second focus group, the ASR output of the current affairs clip was edited manually before segmentation and translation to correct misidentified punctuation, sentence boundaries and speaker turns, incorrect number formatting and proper names. Some of the quality shortcomings mentioned above stemmed from these errors. Therefore, we wished to test whether such minor adjustments would improve the MT quality and viewer

reception effectively. Because resources for comprehensive manual correction of the ASR output were deemed unlikely in a real-world scenario, the “pre-editing” was limited to mechanical issues which could be fixed with relatively little effort compared to fully editing the text, and could potentially be automated with further development of ASR, named entity recognition and other processing steps. Correcting sentence boundaries and speaker changes, in particular, was deemed necessary because some of the confusion in the first version arose from the ASR incorrectly combining speech turns, which the MT naturally treated as a whole, leading to confusing statements.

3.2. Focus groups

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. Allowing the participants to discuss the topic with each other can produce richer data than interviews, where participants simply answer questions (Wilkinson 2006: 50-52). Due to the relatively early stage of automatic interlingual subtitling technology and scarcity of previous audience research, we selected focus groups as a useful exploratory method. This approach can reveal previously undiscovered opinions and perspectives that can be used to develop further, more specific research questions.

Due to the COVID-19 pandemic, focus groups were conducted online (Google Meet). The main selection criterion for focus group participants was their linguistic background. They were expected to be native or near-native speakers of English with little or no understanding of Finnish, which was the source language of the video clips. To include a variety of perspectives, we aimed to recruit participants representing a broad selection of backgrounds in terms of age, gender, education and occupation. The first focus group in June 2020 was attended by seven participants, four men and three women, ranging in age from 24 to 44. Three participants were native English speakers and four near-native speakers of English. The second focus group in October 2020 was attended by seven participants, four men and three women. The participants were 23 to 65 years old, and six of them were native English speakers, while one was a near-native speaker.

The first focus group session started with warm-up questions about how often the participants watch subtitled programming. Then, they watched the video clip, after which the moderator asked open-ended questions, first about the participants’ general reactions and first impressions, then more specifically about comprehension and cognitive load, including the pace of the subtitles and the amount of mental effort the participants needed to invest in reading the subtitles. Then, they were asked for their general opinion on automated subtitles, and whether they would be willing to watch something with such subtitles or prefer to wait a few days for human-translated subtitles. Finally, participants were invited to add any other

questions or comments. The structure of the second focus group was similar, but the participants were shown two clips with shorter discussion about each topic. In addition, the second group was asked to compare their impressions about the two clips.

3.3. Online questionnaire

The questionnaire was open for four weeks in October-November 2020, and received 74 responses. The questionnaire contained a combination of multiple choice, five-point Likert scale, and open questions to gauge viewers' comprehension, cognitive load, and attitudes towards the clips and towards automatic subtitles in general.⁴ The four open questions were voluntary, but received a good number of responses (53 responses to question 29, 50 to question 30, 29 to question 31, and 20 to question 32). The questionnaire link was distributed through social media (e.g. project participants' Twitter and Facebook accounts, Yle's English language news service Facebook account), and other online networks (e.g. the mailing list of the Finnish branch of the Erasmus Student Network).

Most respondents were highly educated and between 30 and 59 years old (see Tables 1 and 2). As such, this group was not fully representative of the general population, but they represent a meaningful variety of backgrounds and plausible potential subtitle users.

Highest level of education completed	Percent of respondents
Secondary education	4%
Vocational qualification	3%
Some studies at a university or other institute of higher education, but no degree	15%
Undergraduate degree	22%
Postgraduate degree	34%
Doctoral degree or higher	23%

Table 1. Educational background of questionnaire respondents

Age	Percent of respondents
18-29	16%
30-44	38%
45-59	32%
Over 59	14%

Table 2. Age of questionnaire respondents

The Finnish comprehension skills of a significant majority of the respondents were low or nonexistent, indicating that they would be a realistic target audience for the subtitles. The respondents' self-evaluations of their Finnish skills are presented in Table 3.

Finnish comprehension	Percent of respondents
Not at all	58%
Some words and phrases in simple Finnish (CEFR A1) ⁵	15%
Main points in simple Finnish (CEFR A2)	10%
Majority of simple Finnish (CEFR B)	4%
Complex Finnish (CEFR C1)	7%
No difficulty understanding any Finnish (CEFR C2)	7%

Table 3. Finnish comprehension skills of questionnaire respondents

Correspondingly, most respondents stated that they had never lived in Finland (see Table 4). The questionnaire respondents therefore differ from the focus groups, because all focus group participants lived in Finland at the time of the study. Thus, the questionnaire data provided a broader look at reception by including views from people with little connection to Finland and little contextual knowledge of Finnish media.

Living in Finland	Percent of respondents
Never	68%
Currently live in Finland and have lived there for 1-5 years	5%
Currently live in Finland and have lived there for 6-10 years	5%
Currently live in Finland and have lived there for more than 10 years	18%
Have previously lived in Finland	4%

Table 4. Length of time lived in Finland by questionnaire respondents

4. Analysis

In the following analysis, we first discuss responses to questions regarding comprehension and appreciation of the automatic interlingual subtitles in both the focus groups and the questionnaire. Then, we explore how the cognitive load caused by reading these subtitles was addressed by the study participants. Finally, we discuss the acceptability of the subtitles. On the one hand, we describe potential use contexts suggested by the study participants and reasons why they might use automatic subtitles. On the other hand, we examine potential obstacles that may limit the usefulness of automatic subtitles.

4.1. Comprehension and appreciation

The responses in both the focus groups and the questionnaire indicated that viewers were able to understand the video clips reasonably well with the automatic subtitles. When asked to explain the contents of the clips, focus group participants gave fairly accurate descriptions, even though they evidently felt the viewing experience was confusing and required some effort. In the questionnaire, respondents answered six multiple-choice comprehension questions (questions 9-11 and 18-20). The correct answers to these questions could not be inferred from nonverbal information on screen, so the participants had to use information in the subtitles as a basis for their responses. A substantial majority of responses to three of them were correct (76%, 95% and 95%, respectively), and in a fourth question, more respondents (45%) answered correctly than incorrectly (34%), while 22% chose "I don't know." In the other two questions a majority of the responses were either incorrect or "I don't know." The respondents' self-assessment of their comprehension was positive (questions 8 and 17): on a scale of 1 to 5, the mean scores were 3.6 (median 4) for the current

affairs video, and 4.2 (median 4) for the news video. Similarly, respondents who viewed automatic subtitles in the study by Hu *et al.* (2020: 529) answered most comprehension questions correctly, although they rated ease of understanding and clarity of raw MT lower than human or post-edited translations. In contrast, Matamala *et al.* (2015) found that automatic interlingual subtitles did not help participants with low or intermediate level of source language proficiency to answer most comprehension questions.

However, quality deficits in the subtitles affected the viewers' appreciation of the clips. In both focus groups, the predominant reaction to the subtitles was negative. For example, in the first group, one participant called them "ropey," while in the second group, a participant compared the quality to the comically incorrect English of the film character Borat. In both groups, the main topics of criticism included obvious mistranslations, unclear and unidiomatic target language, problems in timecoding and segmentation, and a fast pace that was difficult to follow. The contrast between reasonably good comprehension and negative reactions to quality was expressed by one participant in the second focus group in the following way: "I felt like it was like translated by Google Translate. With like five different languages. But then, in the end, you still get the main picture of what's happening." Pre-editing the ASR output prior to translation seemed to have limited effect. In the second focus group, comments regarding the latter, pre-edited clip also reflected confusion and difficulty following the narrative. However, some comments were cautiously more positive, and participants claimed they were able to tell that some adjustments had been made to improve the text.

The questionnaire respondents were asked three Likert-scale questions concerning their appreciation of the subtitles for each clip. The current affairs video (questions 12-14) received the mean scores of 2.7 (median 3) for pleasantness, 3.6 (median 4) for usefulness and 3.0 (median 3) for accuracy. The news video (questions 21-23) was rated more positively, with mean scores of 3.5 (median 4) for pleasantness, 3.8 (median 4) for usefulness, and 3.6 (median 4) for accuracy. In other words, the experience of watching particularly the current affairs clip was not very enjoyable, but the respondents rated the subtitles as reasonably useful and accurate. Responses to the open questions reinforce the sense that the viewing experience was uncomfortable. Some responses stated that automated subtitles could be useful, but a larger number of comments expressed misgivings about the quality and the technology, and many suggested that subtitles always need human involvement, such as post-editing. The following improvement suggestion demonstrates some of these negative feelings:

The pacing could be more consistent. Some groupings of words were easily read in the amount of time they appeared on the screen, but some groupings were gone

before I could finish reading them. The result was that I felt nervous, reading everything a little too quickly for solid comprehension.

4.2. Cognitive load

One recurrent theme arising from the focus group and questionnaire data is that viewers assess the cognitive load caused by automated subtitles as higher than that of professionally made subtitles. Focus group participants frequently mentioned that viewing the video clips was demanding, because it required concentration, and because it was challenging to divide attention between the subtitles and the rest of the programme. One participant in the second focus group described this process as follows:

[...] it's the news, and there's lots of graphics going on, but I'm trying to read the charts, and look at subtitles, and then. You know, I look up to see what they're talking about, and then by the time I look at the subtitles again, it's gone, probably, past a couple of sentences [...] I'm sort of struggling to keep up with the subtitles, let alone decipher some of the grammar, to what they were saying, and then, see what's happening on the screen.

Even if the gist of the clips was understandable, the demanding nature of the subtitles meant that participants often expressed a need to focus all their attention on the clip. As one participant in the first focus group stated: "It requires your full attention. I don't think this is that kind of a programme you could sit and just half-watch with one eye, while you are scrolling on your phone." This experience may be partly due to the dense information load of the video clips, since the complexity of the structure, information content, language and narrative also affect cognitive processing of subtitles and enjoyment of the viewing experience (see Perego *et al.* 2018). However, the comments about deciphering grammar, for example, indicate that subtitle quality was a contributing factor. In addition, it is likely that issues such as poor segmentation and rushed pace, which respondents frequently mentioned, had a detrimental effect on cognitive load.

Responses in both focus groups suggested two distinct reasons why automatic subtitles can cause high cognitive load. First, the quality deficiencies mean that processing the subtitles is not possible with just a glance. The subtitles therefore claim most of the viewers' attention and distract them away from the rest of the programme. Second, errors and other awkward and even unintentionally funny features can distract viewers' attention altogether and shift their focus to the error instead of the programme. Hu *et al.* (2020) also observed distracting effects in their eye-tracking data, where gaze times on subtitles were longest on raw MT subtitles.

Similar views on cognitive load were apparent in the questionnaire. Cognitive load was addressed in three questions: two that were posed separately for each clip and one to gauge the respondents' general sense of cognitive load after watching both clips. When asked how well the viewers

were able to read the subtitles all the way through (questions 15 and 24), the mean score was 2.6 (median 3) for both clips, which suggests that the respondents coped with the subtitles equally or almost as well as usual. When asked about mental effort (questions 16 and 25), the mean scores of 2.0 (median 2) for the current affairs clip and 2.5 (median 2) for the news clip indicate that more mental effort was needed than usual. The final question (26) which compared this viewing experience to a general experience of watching subtitled content received a mean score of 2.4 (median 2), suggesting that most respondents found the viewing experience laborious. Viewing machine-translated subtitles was also found by Hu *et al.* (2020) to involve more mental effort than human-translated or post-edited ones.

4.3. Potential use contexts, motivating factors and obstacles to use

Although the study participants' assessments of the automatic interlingual subtitles were not very positive, many did indicate they would be motivated to use them in specific circumstances. On the one hand, many participants suggested automatic subtitles would be useful for urgent information that would not be otherwise available in the target language, such as breaking news or local information. In this case, the need to receive such information quickly could motivate users to accept imperfect quality as better than nothing. On the other hand, several participants expressed concern over the reliability of the information conveyed via automation. Therefore, another proposed scenario involved less important content, where the motivation for using automatic subtitles would come from gaining access to local culture and society. This could mean light entertainment such as celebrity interviews, music programmes, morning shows or other "everyday silly things," as one participant in the second focus group put it. With these lower-stakes programmes, accuracy would be a smaller concern. In the first focus group, one participant mentioned that such linguistic support "gives you a lot more access, a lot more, ability [...] to actually take part in the cultural events. There are these things you'd like to talk about with Finnish colleagues and friends the next day." This comment demonstrates the broader value of linguistic accessibility as a means of inclusion in society, and explains why those who do not fully understand the local language may be motivated to accept automatic subtitles. However, the data suggest that more study participants consider news and other informative or time-sensitive content to be the primary type of programming that could benefit from automatic subtitles. These programmes were the most frequent answer in the focus groups as well as in the questionnaire's open question asking for potential uses for automatic subtitles.

Although many study participants mentioned potential uses for automatic interlingual subtitles, many also pointed out that the quality was not yet sufficient. In addition, many focus group participants simply stated that they would prefer subtitles that have been created or at least post-edited

or checked by humans. Most questionnaire respondents also indicated a preference for human-made subtitles (question 28), as seen in Figure 1.

If you could choose either automatic subtitles that were available immediately or human-made subtitles that were available a couple of days later, which one would you choose?

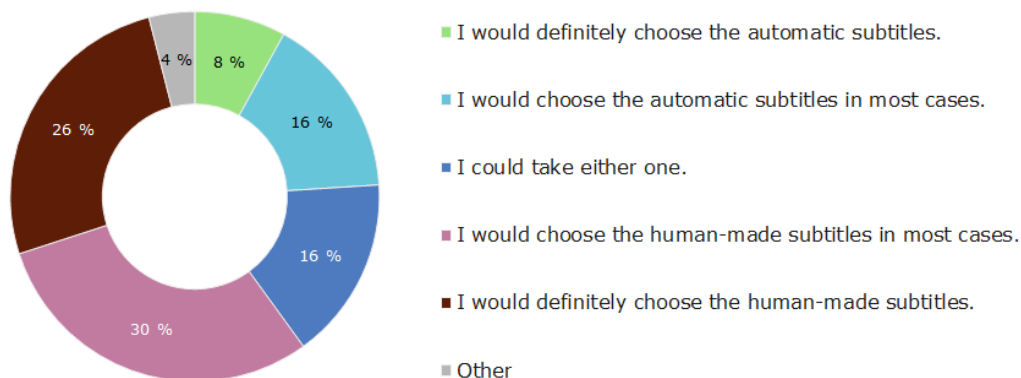


Figure 1. Stated preference for automated or human-made subtitles

However, the responses also reflect some interest in automatic subtitles, including 24% of respondents who prefer automatic subtitles immediately over human-made ones later. This openness to automatic subtitles was more obvious when respondents were presented with different scenarios for choosing automated subtitles (question 27). The most permissive option (“Any time automatic subtitles are available for a foreign-language video that interests me”) was the most popular choice with 43% of responses, and only 9% chose “never” (see Figure 2). This result seems somewhat more promising than the findings of Armstrong *et al.* (2007) and Flanagan (2009), since most of their respondents did not find automatic interlingual subtitles acceptable. The difference may reflect both technological developments and different audience expectations regarding movie subtitles, which were used in the older studies. In addition, it is possible that the attitudes of the participants in this study were more permissive because they were directed to think of online contexts, such as websites and social media (see question 27). The threshold for accepting automatic subtitles for television broadcasts, for example, may be higher.

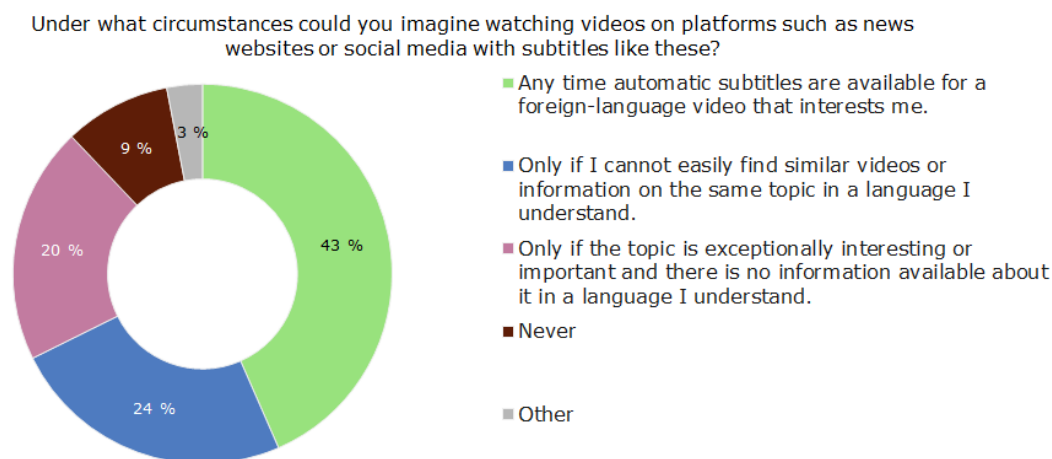


Figure 2. Potential use scenarios for automated subtitles

In all, insufficient quality of the automatic interlingual subtitles and the resulting mistrust in their contents were seen as a significant obstacle for use. Problems mentioned by respondents included fluency and consistency of the target language, interference, spelling, punctuation, accuracy of the translation, segmentation of the subtitles, pacing and reading speed, visual presentation of the subtitles, timecoding and synchronisation, and issues related to condensation and omissions in the text. While many mentioned that having any English subtitles available would be a welcome addition to Finnish public broadcasts, many also expressed a lack of trust in the subtitles at their current quality level. The first focus group also raised the point that programmes with fully automated subtitles should be accompanied by a disclaimer which would inform the audience's expectations concerning subtitle quality and accuracy.

The cognitive load caused by the subtitles emerged as another obstacle for their use. Research on subtitle synchronisation has found that breaking the synchrony (delayed, extended or shortened display time) increases cognitive load of subtitle processing even if viewers are not consciously aware of the problems (Lång *et al.* 2013). In our study, participants appeared very aware of this, commenting on synchrony issues and the ensuing mental effort frequently in both the focus groups and the questionnaire. It should be noted that their observations are based on short clips: trying to follow a long programme would certainly be even more mentally taxing. Improving subtitle timing as well as other factors affecting cognitive load is therefore an important direction for future research.

5. Discussion and conclusions

Focus groups along with the questionnaire proved to be a useful exploratory method. The findings provided a relatively uniform picture of audience views, suggesting that a saturation point was reached where comments start repeating themselves. Some slight differences were observed in that

respondents who indicated they rarely watch foreign-language content subtitled into English seemed somewhat more reluctant to use automatic subtitles. The majority of these respondents (11 out of 15) stated they would use automated subtitles only if no information were available on the topic and/or the topic was exceptionally interesting or important, and nearly all (13 out of 15) preferred human-made subtitles later rather than immediately available automatic subtitles. In comparison, approximately half of respondents who indicated they watch subtitled content often or occasionally were willing to use automatic subtitles any time they were available, and preference for automatic subtitles immediately or human-made subtitles later varied in these groups. No other meaningful variation emerged based on the respondents' background. Although the number of respondents is relatively low and views expressed in the evaluations are not statistically representative, they come across as a fairly realistic reflection of audience views, particularly among highly educated audience segments, which were most heavily represented in the data. Moving forward, it would be useful to complement these subjective views with more objective methods of observation, such as eyetracking, to gain a deeper understanding of the viewing process, and particularly the cognitive load associated with viewing automatic subtitles.

The viewer reactions were somewhat ambivalent, even contradictory, and did not suggest an easy answer to how well automatic interlingual subtitles could facilitate linguistic accessibility. While viewers expressed some conditional acceptance, additional work is needed to reach genuinely usable levels of quality. For any scenario where the intention is to provide subtitles in fully automated form, further work should prioritise technical solutions that affect the cognitive load caused by subtitles, such as reducing the amount of text, improving segmentation and synchronisation, and smoothing out target language readability issues. Ensuring the accuracy of the translation is of course vital, but even a factually accurate translation would not be fully acceptable if the cognitive load remains high. However, reaching suitably high quality for all use purposes may not be feasible through automation alone, and the need for post-editing needs to be considered carefully.

Even with the shortcomings in quality, the findings demonstrate a clear interest in and need for automated subtitling solutions. Technical improvements leading to better quality could make a plausible case for introducing automated subtitling for some public broadcasting content and situations. Introducing automated subtitling and finding the optimal use contexts is, however, a complicated task. News and current affairs content seems like an obvious place to start, especially for an audience like English-speakers in Finland. However, the nature of news reports presents challenges for automatic subtitling. The fast-paced, continuous speech that is characteristic of news content can cause excessive cognitive load if the subtitles contain too much text or are not segmented for readability. Issues related to cognitive load such as condensation and segmentation therefore

emerge again as crucial challenges for further development. Additionally, accuracy is even more vital in news and current affairs than in other genres; errors in reporting local emergencies, for example, could have a direct effect on the viewers' lives.

An important factor in using automated subtitles is motivation: viewers can accept even imperfect quality if it gives them access to something they genuinely want to or need to access. The converse is also true: content can become more interesting, and viewing motivation can increase, if the subtitles are comfortable to follow. Consequently, better quality may enhance viewers' engagement with content, and working to improve subtitle quality is in the broadcaster's interest, even if lesser quality might suffice for some contexts. Another option to explore could be to simply provide automatic subtitles for all content and allow viewers to choose what they are most motivated to watch. However, such a decision involves risks which would need to be carefully weighed by the broadcaster. Shortcomings in quality would become particularly visible if even the most demanding content was subtitled, which may affect the viewers' perception of, and trust in, both the content and the broadcaster. Transparency and managing audience expectations are also important concerns, and automated subtitles should be clearly labelled as such. If possible, an introductory explanation of the automation and its current stage of development could also be offered to interested audience members.

One question at the start of our study was whether familiarity with tools like Google Translate and the imperfect output they produce would lead people to find the level of automatic interlingual subtitling acceptable. Based on our findings, human involvement is appreciated and preferred to full automation, even if it means waiting for the translation. Automatic interlingual subtitling was deemed helpful by viewers under some circumstances, but it does not seem ready for comprehensive implementation. Nevertheless, it holds promise for populations that are being poorly served by current media provision. Automatically generated subtitles cannot replace professional subtitlers, and caution is needed, so that automation experiments do not create an atmosphere where striving for optimum quality would be deemed unnecessary. Constant development towards better quality is an important goal, particularly in the context of public broadcasting. Still, automatically generated subtitles would be a welcome solution in circumstances where professional subtitling is impossible. By providing access to news and other content that is important for linguistic minorities, or to light entertainment that could facilitate better inclusion in local society, automated subtitling solutions can contribute towards the crucial goals of linguistic accessibility and inclusion.

Acknowledgements

This work is part of the MeMAD project (2018-2021), funded by the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement No 780069).

References

- **Armstrong, Stephen et al.** (2007). "Leading by example: Automatic translation of subtitles via EBMT." *Perspectives* 14(3), 163–84. <https://doi.org/10.1080/09076760708669036>.
- **Braun, Sabine et al.** (2021). *D6.9 Evaluation report, final version*. MeMAD Project. <https://memad.eu/wp-content/uploads/D6.9-Evaluation-Report-final-version-version-1.0-merged.pdf>.
- **Burchardt, Aljoscha et al.** (2016). "Machine translation quality in an audiovisual context." *Target* 28(2), 206–21. <https://doi.org/10.1075/target.28.2.03bur>.
- **Bywood, Lindsay, Panayota Georgakopoulou and Thierry Etchegoyhen** (2017). "Embracing the threat: Machine translation as a solution for subtitling." *Perspectives* 25(3), 492–508. <https://doi.org/10.1080/0907676X.2017.1291695>.
- **Di Gangi, Mattia A., Matteo Negri and Marco Turchi** (2019). "Adapting transformer to end-to-end spoken language translation." *Proceedings of the Annual Conference of the International Speech Communication Association INTERSPEECH 2019*, 1133–37. <https://doi.org/10.21437/Interspeech.2019-3045>.
- **Di Giovanni, Elena and Yves Gambier** (2018). "Introduction." Elena Di Giovanni and Yves Gambier (eds) (2018). *Reception Studies and Audiovisual Translation*. Amsterdam: John Benjamins, vii-xii.
- **Flanagan, Marian** (2009). *Recycling Texts: Human Evaluation of Example-Based Machine Translation Subtitles for DVD*. PhD thesis. Dublin City University. <http://doras.dcu.ie/14842/1/MarianFlanaganPhD.pdf>.
- **Hirvonen, Maija and Tuija Kinnunen** (2021). "Accessibility and linguistic rights." Kaisa Koskinen and Nike K. Pokorn (eds) (2021). *The Routledge Handbook of Translation and Ethics*. Abingdon: Routledge, 470–83. <https://doi.org/10.4324/9781003127970-35>.
- **Hu, Ke, Sharon O'Brien and Dorothy Kenny** (2020). "A reception study of machine translated subtitles for MOOCs." *Perspectives* 28(4), 521–38. <https://doi.org/10.1080/0907676X.2019.1595069>.
- **Karakanta, Alina, Matteo Negri and Marco Turchi** (2020). "Is 42 the answer to everything in subtitling-oriented speech translation?" *Proceedings of the 17th International Conference on Spoken Language Translation (IWSLT2020)*, 209–19. <https://doi.org/10.18653/v1/P17>.
- **Koponen, Maarit et al.** (2020a). "MT for subtitling: User evaluation of post-editing productivity." André Martins et al. (eds) (2020). *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation (EAMT 2020)*, 115–24. <https://eamt2020.inesc-id.pt/>.

- **Koponen, Maarit et al.** (2020b). "MT for subtitling: Investigating professional translators' user experience and feedback." John E. Ortega et al. (eds) (2020). *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas 1st Workshop on Post-Editing in Modern-Day Translation*, 79–92. Association for Machine Translation in the Americas.
- **Koponen, Maarit et al.** (2020c). "User perspectives on developing technology-assisted access services in public broadcasting." *Bridge: Trends and Traditions in Translation and Interpreting Studies* 1(2), 47–67. <https://www.bridge.ff.ukf.sk/index.php/bridge/article/view/8>.
- **Käännöstekstityksen laatusuositukset** [Quality recommendations for translated subtitles] (2020). Av-kääntäjät. <https://www.av-kaantajat.fi/gallery/laatusuositukset%20taitettu%20ei%20allek.pdf>.
- **Laaksonen, Jorma, Umut Sulubacak and Jörg Tiedemann** (2021). *D4.3 Tools and models for multimodal, multilingual and discourse-aware machine translation*. MeMAD Project. <https://doi.org/10.5281/zenodo.4630083>.
- **Laki Yleisradio Oy:stä** [Law regarding Yleisradio Oy (Yle)], <https://finlex.fi/fi/laki/ajantasa/1993/19931380> (consulted 21.12.2021).
- **Lång, Juha et al.** (2013). "Using eye tracking to study the effect of badly synchronized subtitles on the gaze paths of television viewers." *New Voices in Translation Studies* 10(1), 72–86.
- **Matamala, Anna et al.** (2015). "The reception of intralingual and interlingual automatic subtitling: An exploratory study within the HB4ALL project." João Esteves-Ferreira et al. (eds) (2015). *Translating and the Computer* 37, 12–17.
- **Matusov, Evgeny, Patrick Wilken and Yota Georgakopoulou** (2019). "Customizing neural machine translation for subtitling." *Proceedings of the Fourth Conference on Machine Translation (WMT), Volume 1: Research Papers*. Florence: Association for Computational Linguistics, 82–93.
- **Perego, Elisa, Fabio Del Missier and Marta Stragà** (2018). "Dubbing vs subtitling: Complexity matters." *Target* 30(1), 137–57.
- **Pérez-González, Luis** (2014). *Audiovisual Translation: Theories, Methods and Issues*. London: Routledge.
- **Piperidis, Stelios et al.** (2004). "Multimodal multilingual resources in the subtitling process." *Proceedings of Fourth International Conference on Language Resources and Evaluation (LREC 2004)*. Lisbon: European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2004/pdf/680.pdf>
- **Popowich, Fred et al.** (2000). "Machine translation of closed captions." *Machine Translation* 15(4), 311–41. <https://doi.org/10.1023/A:1012244918183>.
- **Sulubacak, Umut et al.** (2020). "Multimodal machine translation through visuals and speech." *Machine Translation* 34(2–3), 97–147. <https://doi.org/10.1007/s10590-020-09250-0>.
- **Szarkowska, Agnieszka and Olivia Gerber-Morón** (2018). "Viewers can keep up with fast subtitles: Evidence from eye movements." *PLoS One* 13(6), e0199331. <https://doi.org/10.1371/journal.pone.0199331>.

- **Statistics Finland.** "Vieraskieliset [Foreign-language speakers]." <https://www.stat.fi/tup/maahanmuutto/maahanmuuttajat-vaestossa/vieraskieliset.html> (consulted 21.12.2021).
- **Vitikainen, Kaisa and Maarit Koponen** (2021). "Automation in the intralingual subtitling process: Exploring productivity and user experience." *Journal of Audiovisual Translation* 4(3), 44–65. <https://doi.org/10.47476/jat.v4i3.2021.197>.
- **Volk, Martin et al.** (2010). "Machine translation of TV subtitles for large scale production." *Second Joint EM+/CNGL Workshop*, 53–62.
- **Wilkinson, Sue** (2006). "Analysing interaction in focus groups." Paul Drew, Geoffrey Raymond and Darin Weinberg (eds) (2006). *Talk and Interaction in Social Research Methods*. London: SAGE, 50–62.

Data Availability Statement

The questionnaire used in this study is available on Zenodo: <https://doi.org/10.5281/zenodo.7244340>.

Biographies

Tiina Tuominen is an acting Professor of English with a specialisation in translation at the University of Turku, Finland. She has previously worked as a developer of subtitling and translation for the Finnish public broadcaster Yle, as a Lecturer at the University of Glasgow, and in various roles at the University of Tampere, Finland. Her research focuses on subtitling, particularly reception, user-centered translation, translators' workplace studies, and multimodality. She has worked as a translator and subtitler for several years.

ORCID: 0000-0002-0665-6970

E-mail: tiina.k.tuominen@utu.fi



Maarit Koponen currently works as Professor of Translation Studies at the University of Eastern Finland. She has previously worked as a post-doctoral researcher at the University of Helsinki and as a lecturer at the University of Turku. She obtained her PhD in Language Technology at the University of Helsinki. Her research focuses on the use of machine translation and other translation technologies, machine translation post-editing and quality evaluation. She has also worked as a professional translator for several years.

ORCID: 0000-0002-6123-5386

E-mail: maarit.koponen@uef.fi



Kaisa Vitikainen works as a live subtitler at the Finnish Broadcasting Company Yle. They are also a doctoral student of Language Studies at the University of Helsinki. Their research focuses on the application of automatic speech recognition in intralingual subtitling.

ORCID: 0000-0003-2067-3969

E-mail: kaisa.vitikainen@yle.fi

Umut Sulubacak is a former doctoral student in the Language Technology research group at the University of Helsinki. His research focused on the development of machine translation systems optimised for the translation of media subtitles. Previously, he has also published on computational morphology and syntax, and contributed to the creation of several treebanks.



Jörg Tiedemann is Professor of Language Technology at the University of Helsinki. He received his PhD in computational linguistics for work on bitext alignment and machine translation from Uppsala University before moving to the University of Groningen for 5 years of post-doctoral research on question answering and information extraction. His main research interests are connected with machine translation, massively multilingual data sets and data-driven natural language processing and he currently runs an ERC-funded project on representation learning and natural language understanding.

ORCID: 0000-0003-3065-7989

E-mail: jorg.tiedemann@helsinki.fi

Notes

¹ The pipeline is available on Github <https://github.com/MeMAD-project/subtitle-translation>.

² For more information on Marian, see <https://marian-nmt.github.io/>

³ The OPUS corpora are available at <https://opus.nlpl.eu/>

⁴ The questionnaire is available on Zenodo: <https://doi.org/10.5281/zenodo.7244340>.

⁵ Definitions of proficiency levels according to the Common European Framework of Reference for Languages (CEFR): <https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale>.