Machine-Translated Subtitles and Education: The Future of More Accessible Learning? – Insights from an Eye-Tracking Study

Frederike Schierl[1][0000-0001-8271-731X]

[1] Tampere University, Kalevantie 4 33100 Tampere, Finland
frederike.schierl@tuni.fi

Abstract. Globalization has flattened the world. This does not apply only to the global economy, but also increasingly to education. In line with this, educational material available online and in video format has become an integral part of today's education. Despite the availability, support with the aid of audiovisual translation is sometimes needed to make the content more accessible, which especially applies to content in a language other than the student might be proficient in. In these cases, subtitling has become an important aid to support this need. However, as of today, the material to be translated exceeds the capacities of human subtitlers. Machine translation can contribute to a cost- and time-efficient alternative. Nevertheless, since research on the topic is still scarce, the question of how the consumer perceives such subtitles in an educational context, in which the correct transmission of information is crucial, remains. When it comes to machine translation, the benefit of information gain and learning effect is only one aspect, sufficient quality is another issue. This study presents the results of an eye-tracking study contrasting the reception of machine-translated and human-translated subtitles in German, Finnish, and English. The overall results show that the translation mode itself is not decisive, but rather, gaze behaviour depends on the general language competence of the viewer to understand the video's original language, therefore the dependence on the subtitling to understand the content. These quantitative results give insight into the potential that machine translation has in the context of broad accessibility.

Keywords: Subtitling, Machine Translation, Education, Accessibility, Reception.

1 Introduction

In the world of today, focusing on visual, video material on the Internet has become a competitive, sometimes crucial, source of information. It also can support the understanding of a subject or can give further information on a topic, especially in an educational context [1-4]. This applies in particular to a globalized education world. But despite the sheer availability of the information or the video, the content might not always be accessible in the user's language or the user's language level might not be good enough. To overcome the language barrier, machine-translated subtitles (MTST) can

Proceedings of New Trends in Translation and Technology, pages 156–169
3–6 July, 2024, Varna, Bulgaria. ©NeTTT 2024
https://doi.org/10.26615/issn.2815-4711.2024_013

help and are already applied on video platforms such as YouTube (already as of 2008). However, the quality of the provided subtitling can still be an issue, and therefore human-translated subtitles (HTST) can still be preferred.

For the purpose of investigating how a consumer deals with a certain product, reception studies have gained more and more in importance over the last decade, also due to the recognition that there is a need to put the consumer in the spotlight [5] and especially more studies have been conducted in the context of subtitling reception due to the popularity of this easy-to-access tool of audiovisual translation and the need for more accessibility [6-12]. Further, methods such as eye-tracking have also become more and more popular to gain insight into (potential) cognitive load while consuming subtitles [13, 8, 3]. Also, according to Kruger and Doherty [14], the combination of offline methods (such as questionnaires) and online methods (such as eye-tracking) have been considered fruitful and a valid tool for triangulation.

Nevertheless, when it comes to the investigation of the reception of subtitles in an educational context, other factors have to be considered, such as the cognitive load [15,16], redundancy effect [17], and attention split [18]. Therefore, it is crucial to shed light on the reception of subtitles, in particular when the language of instruction is not English (today's *lingua franca*), and to further investigate to what extent subtitled audiovisual material can foster understanding, increase or remove cognitive load, and influences attention allocation in general with particular regard to the differences in machine-translated and human-translated subtitles.

2 Research Question and Methodological Framework

This paper is part of a PhD project and investigates the reception of machine-translated and human-translated subtitles (MTST vs. HTST) with regard to the 3R concept [19]: response, reaction, repercussion, i.e. how does the reception of machine-translated and human-translated subtitles differ in terms of their visual reception, their cognitive treatment and also general attitudes and expectations towards (machine-translated) subtitles. In this part of the research, I focused on the aspects of response, whether and to what extent the machine-translated condition influences the general watching behaviour and experience (i.e. whether and to what extent there is a shift from the presentation/speaker to the subtitling area), and reaction, i.e. the cognitive load.
Therefore, the research hypotheses are as follows:

- H1 (for response): Informants in the conditions of MTST will pay more attention to the subtitles than to the image (compared to HTST) since they concentrate more on the written text when the subtitling quality is suspected to be low.
- H2 (for reaction): Informants in the conditions of MTST will report a lack of comprehension, which will be reflected in lower rating scores compared to HTST. The cognitive load will also be higher.

Due to space constraints, results of comprehension and rating of the subtitles are not presented in this paper.

Early results of a reception study conducted in the ongoing project [20] indicate that the reception does not differ significantly in the information retrieval, i.e. the general information gain, but human-translated subtitles are perceived as better on a personal scale, i.e. when being evaluated according to the User Experience Questionnaire [21] and NASA Task Load Index [22], the general user experience is perceived as more pleasant with human-translated subtitles, especially when it comes to aspects such as speed (display and standing time of the subtitles), which makes the reading more burdensome.

3 Experiment Setup and Procedure

3.1 Video Material and Subtitling

The video corpus consisted of two video recordings by two researchers from Tampere University who, upon request, volunteered to record a presentation in video format in which each presented a topic of their teaching or research interest in their respective native language, i.e. Finnish and German. The presentation topics were *Regional and Culture Studies in the German-speaking area* and *Professional Listening in on-site Interpreting*. The presentations were recorded on the video platform *Panopto*. This platform also supports the creation of subtitles through automatic speech recognition (ASR). The intralingual subtitles based on the ASR were then exported as SRT file for further processing. For the machine-translated condition, the subtitles were slightly post-edited, i.e. only misrecognitions and punctuation errors were corrected. The machine translation was performed in the open-source tool *SubtitleEdit*, which has a built-in machine translation function, which uses *GoogleTranslate* as tool. The subtitles therefore were translated with a neural network (*Transformer* model), which is also due to the starting point of the project and production of the videos (2022) where this technology was state-of-the-art (i.e. generative AI had not yet reached popularity).

For the human-translated condition, the task was assigned to a professional subtitler with over 25 years of experience in both creating subtitles from scratch and with templates. In this case, the subtitler was not provided with templates.

Finnish was the only subtitle language for the German presentation whereas a German and English set of subtitles was used for the Finnish presentation. It was also considered to use an English set of subtitles for the German presentation, but this consideration was rejected since it was considered not beneficial for the outcome of the study, also since the main location of the research is Finland. Therefore, it was expected that more highly proficient Finnish speakers would volunteer for the experiment.

3.2 Participants

Participants for the experiment were recruited between October and December 2023 via the Intranet of Tampere University (i.e. a post informed about the research and potentially interested people could contact me via e-mail). Overall, 20 participants volunteered to take part in in the experiment, and as a result, 19 datasets could be included

in the analysis: 7 male and 12 female [age groups: 20-29: 7x, 30-39: 6x, 40-49: 2x, 50-59: 3x, 60-69: 1x). It also should be considered that all participants had a background of higher education, which meant that the test group was comprised of relatively highly educated participants.

Further, participants were asked about the language(s) they considered themselves most proficient in ('native' or 'near-native' level; the question 'What is your native language?' was avoided on purpose) and participants responded as follows: English (16x), Finnish (13x), and German (4x), Chinese, Danish, Dutch, Indonesian, Italian, Polish, Swedish (1x respectively). Regarding their level of Finnish and German respectively (since those languages were the audio languages of the presentations), the following distribution was obtained (Table 1).

Table 1. Language competence distribution

Level/ Language	No knowledge	Beginner	Intermediate	Advanced	Native Speaker
Finnish	2	3	2	0	12
German	5	7	1	3	3

3.3 Methods

To investigate the differences in reception of machine-translated and human-translated subtitles, the participant's gaze was tracked through a Tobii T60 eye-tracking device (remote eye-tracking via a 17" monitor working at 60 Hz sampling rate) while watching the presentation videos on full screen mode on the video platform YouTube (videos were not listed and the platform was used for watching only to facilitate the experimental procedure). The analysis of the data was performed via the Tobii Studio Version 3.8.4 software in which areas of interest (subtitle area, presentation area, and speaker area) were created to perform the analysis of the gaze behaviour. The analyzed aspects were first fixation duration, mean fixation duration, total fixation duration, total fixation count, mean visit duration, total visit duration, and total visit count. For the purpose of the paper and due to limitations in space, only total fixation duration, fixation count, total visit duration, and visit count are presented. For the analysis, independent variables were the condition (machine-translated or human-translated), language competence in the respective language of the video (Finnish or German; high, medium, or low), and the level of dependence on subtitles (entirely, somewhat, almost not, not at all; of note: the categories of 'somewhat' and 'a little' were merged into one group).

Further, questionnaires were used to gain insight into general information retrieval as well as the subjective perception of the subtitling. To do this, the questionnaires contained on the one hand, 10 questions on the general video content and, on the other hand, evaluation items to be evaluated on a scale from 1 (not at all/very poor) to 7 (very much/perfect). These were taken and adapted from the User Experience Questionnaire [21] and NASA Task Load Index [22].

3.4 Experimental Procedure

The experiment took place at Tampere University in a designated space for user experience, called UsabilityLab. Participants who agreed to take part in the experiment were provided with an information sheet on the experiment, a privacy notice, and it was obligatory to sign an informed consent before beginning the experiment. Instructions and explanations for the experiment were also given orally. The participant sat in front of the screen at an appropriate distance (approx. 60 cm), so that the tracking device would recognize the eyes well. Before the recording, a 5-point calibration was performed. The participants were not informed beforehand whether they would watch the presentations with machine-translated or human-translated subtitles. They were also instructed to watch each presentation only once without pausing and rewinding in-between; note-taking was also not permitted.

After these preparatory steps, the recording started and continued until the video presentation ended. Then, the participants were asked to fill in the questionnaires, which lasted approx. 10-15 minutes, during which the gaze was not tracked. The procedure was repeated for the second video presentation. After the experiment, participants were invited to see the tracking results and also to comment on how they felt while watching the subtitled video presentations. The whole experiment duration was approx. one hour per participant.

4 Results

This section presents the results of the data for the language pairs German – Finnish, Finnish – German and Finnish – English. The same analyses were performed on different datasets. In this analysis, total fixation duration, fixation count, total visit duration, and visit count were the predicted (dependent) variables, whereas the condition (machine-translated or human-translated), language competence, and the level of dependence on subtitles were treated as predictor variables.

4.1 German – Finnish

Overall, 20 participants volunteered to take part in the experiment, resulting in 19 datasets that could be used (1 dataset had to be excluded due to the poor data quality, i.e. sufficient gaze tracking was not successful). Of the 19 participants included, 10 watched the presentation in the human-translated and 9 participants in the machine-translated condition. As for the conscious attention allocation (‘Did you read the subtitles consciously all the time?’), 13 participants confirmed that they watched/read them consciously all the time (6 in HTST/7 in MTST), 2 indicated having watched them ‘sometimes’ (1 in MTSTS/1 in HTST), 3 said ‘no’ (all HTST) and 1 participant stated ‘when it was difficult to understand’ (MTST).

When it came to the dependence on the subtitles, ranging from ‘entirely’ dependent to ‘not at all’, 10 participants stated that they were entirely dependent on the subtitling to understand the content (5 in HTST/5 in MTST), 3 were ‘somewhat’ dependent on

them (1 in HTST/2 in MTST), one participant was 'a little' dependent (MTST), one was 'almost not' dependent, and 4 were 'not at all' dependent (3 in HTST/1 in MTST).

Total Fixation Duration. The results for the total fixation duration in the subtitling area did not yield significant effects. Although subtitles in the machine-translated condition were fixated shorter and there was a positive (or longer) effect when participants had a low command of the presentation language, and especially when they were entirely dependent on the subtitles, the results were not statistically significant; an interaction between language competence and dependence on subtitles was not significant.

Analyzing the gaze behaviour at the presentation area, the results showed a significant main effect for total fixation duration and dependence on subtitles (β = -111.89; SE [50.94]; t = -2.2; $p < 0.05$). There was also a marginally significant main effect for total fixation and when being somewhat dependent on the subtitles (β = -105.98; SE [50.14]; t = -2.11; $p < 0.06$). These findings therefore confirm that participants being entirely dependent on the subtitles would fixate on the presentation area less overall. However, it might still be interesting that participants in the machine-translated condition fixated on this area longer, even if the result was not significant. Taking interactions into account, a significant interaction between the machine-translated condition and non-dependence on subtitles was visible (β = 136.49; SE [29.43]; t = 4.64; $p < 0.001$).

For the gaze behaviour of the speaker area, results showed a marginally significant main effect of total fixation duration and machine-translated condition (β = -32.09; SE [16.12]; t = -1.99; $p < 0.07$). Therefore, the participants' attention in this condition seemed to affect the attention allocation, but still taken all factors together (also presentation and subtitling area), the importance of the result is questionable since main effects treated as independent do not account for other influential factors. The results also showed a significant interaction of the machine-translated condition and non-dependence on the subtitles (β = -144.76; SE [42.11]; t = -3.44; $p < 0.007$), which could point towards the behaviour that participants looked more at the presentation than at the speaker, and a significant interaction between the machine-translated condition and low language competence (β = -103.68; SE [36.71]; t = -2.82; $p < 0.02$). The latter finding could then again point towards a tendency to shift the attention also more to other parts of the presentation although the overall subtitling condition might have been influential in that regard that it was available to the participants rather than posing a difference towards human-translated subtitles.

Fixation Count. The analysis of the fixation count showed a significant main effect of fixation count in the subtitle area and the dependence on subtitles when being entirely dependent on them (β = 515.59; SE [213.90]; t = 2.410; $p < 0.03$). Further, no significant interactions between fixation count, condition, language competence, and dependence on subtitles could be found.

Analyzing the gaze behaviour in the presentation area, results showed a significant main effect of fixation count and entire dependence on subtitles (β = -421.86; SE [128.87]; t = -3.27; $p < 0.006$) as well as limited ('somewhat') dependence on subtitles

(β = -317.77; SE [126.86]; t = -2.51; p < 0.03). There was no direct interaction between the different effects.

The results for the speaker showed a significant main effect of fixation count and machine-translated condition (β = -70.08; SE [30.31]; t = -2.31; p < 0.04). However, as pointed out above, this might have been not entirely due to the machine translation, but maybe the general dependence on the subtitles, but interactions could not be shown either.

Total Visit Duration. The total visit duration in the subtitling area showed a barely significant main effect for low language competence in German (β = 74.92; SE [38.52]; t = 1.95; p < 0.08). Total visit duration in the machine-translated condition was also higher, but not statistically significant. The conditions 'machine translation' and 'low language competence' showed a significant interaction effect (β = 167.16; SE [73.06]; t = 2.29; p < 0.05). Overall, the results were not surprising, but it is of particular interest here that the machine-translated condition is suspected to have an influence. As for the watching behaviour in the presentation area, results showed a significant main effect of the total visit duration and the dependence on subtitles (entirely dependent: β = -141.73; SE [49.88]; t = -2.84; p < 0.01; somewhat dependent: β = -107.64; SE [49.10]; t = -2.19; p < 0.05), but there was no significant difference in terms of the subtitling mode. However, the reported findings on dependence on subtitles were in line with expectations, therefore not surprising, and it confirmed that being dependent on the subtitling draws considerably away the attention. Looking into the interactions, a finding showed an interaction between the machine-translated condition and non-dependence on subtitles (β = 154.79; SE [38.70]; t = 4.00; p < 0.003). Therefore, those factors influenced the outcome (total visit duration) indeed or both factors depended on one another.

Investigating the speaker area, there was a barely significant effect of attention allocation and the machine-translated condition (β = -37.83; SE [17.95]; t = -2.11; p < 0.06), which means that participants in this condition fixated the speaker less or the visit duration was shorter, however, this might have been again due to other factors. The analysis for interactions showed indeed that there was an influence of the condition (MT) and low language competence (β = -103.34; SE [44.58]; t = -2.32; p < 0.05) and the machine-translated condition and non-dependence on subtitles (β= -149.92; SE [51.13]; t = -2.93; p < 0.02). Those results may seem contradictory at the first glance, but may also point towards a gaze behaviour for different reasons: in the former case, the attention might have been drawn to other areas of the screen (the subtitling area for example) and in the latter case, the machine-translated condition may also have drawn the attention to other areas of the screen (for instance the presentation area) since the subtitles were not needed.

Visit Count. Lastly, the total visit count, i.e. how often the participant visited the respective area of interest, was of interest as well. Here again, the machine-translated condition had fewer visit counts. Surprisingly, also in contrast to the aspect of fixation and fixation durations, there were no significant main effects. It should be perhaps also

noted that the aspect of low language competence is inferior in visit counts, which could be an indicator that those participants spent overall more time in the area and left it less often to look at other parts of the presentation.

Further, an interaction between machine-translated condition and low language competence could be found (β = -102.67; SE [41.48]; t = -2.48; p < 0.04), which could confirm that the visit count can again depend on more than one factor and that machine translation could have played a role as well.

The results for the analysis for the presentation area showed that the visit count in the machine translation condition was lower, but the result was not significant. However, taking this small finding, visiting this area less, was in line with the previous findings, which would indicate that the gaze might travel less between the different areas. Further, there were no interactions which affected the visit count.

Finally, looking into the reading behaviour for the speaker, results showed less visit counts in the speaker area in the machine-translated condition as well, but the result was not statistically significant. An interaction between the machine-translated condition and low language competence (β = -78.53; SE [36.00]; t = -2.18; p < 0.06) was found, but the result was barely significant and probably less surprising.

4.2 Finnish - German

Twelve participants were watching a Finnish video presentation with a German set of subtitles. Of those 12 datasets, 11 could be included in the analysis (1 had to be excluded due to low data quality). Five of the participants were presented with the machine-translated condition and 6 with the human-translated condition. Of those 11 participants, 3 stated having watched the subtitles 'all the time' (1 in HTST/2 in MTST), 4 stated 'sometimes' (3 in HTST/1 in MTST), 4 said that they did not watch them (2 in HTST/2 in MTST). Further, 2 participants stated being 'entirely' dependent on the subtitling to understand the content (1 in HTST/1 in MTST), 1 was 'somewhat' dependent (MTST), 1 participant stated having been 'a little' dependent (HTST) and 7 were 'not at all' dependent (4 in HTST/3 in MTST).

Total Fixation Duration. Investigating the total fixation duration in the subtitle area, results showed no significant differences for the condition, i.e. the subtitle area in the machine-translated condition was not fixated longer, but there was a significant main effect of total fixation duration and low and medium language competence (low: β = 119.79; SE [21.36]; t = 5.61; p < 0.001; medium: β = 89.34; SE [26.33]; t = 3.39; p < 0.01) and also a significant main effect on non-dependence on subtitles (β = -94.15; SE [18.62]; t = -5.06; p < 0.002) although those effects were not surprising. There were also no significant interactions for the machine translation condition.

The analysis for the presentation area did not show significant results for main effects and no interactions. Therefore, there was no effect on total fixation duration in the presentation area.

The speaker area also did not show any significant main effects and no interactions. Overall, also these analyses did not show an indication that the subtitling in the

machine-translated condition would draw more attention, but this might also be due to the fact that most participants had a high competence in Finnish, i.e. were native or near-native speaker. There were also no significant interactions.

Fixation Count. For the fixation count in the subtitling area, results showed a significant main effect for the language competences low and medium (low: $\beta = 706.19$; SE [166.87]; $t = 4.23$; $p < 0.005$; medium: $\beta = 823.39$; SE [205.74]; $t = 4.00$; $p < 0.007$), which was interesting, but not surprising since it was expected that participants with a low or medium command would fixate the subtitle area more often. However, there was no significant effect in terms of the condition, i.e. participants fixated the area less in the machine-translated condition, but the result was not statistically significant. Further, there were no interactions between other coefficients.

For the presentation area, no significant main effects and interactions between coefficients were revealed in the analysis.

There were no main significant effects for the speaker area although it might be of note that participants in the machine-translated condition would fixate the speaker more (more fixations), although this might be related to their language competence. The results did not yield significant interactions either.

Total Visit Duration. Regarding the total visit duration in the subtitling area, results showed a significant main effect for low language competence ($\beta = 162.48$; SE [57.45]; $t = 2.83$; $p < 0.03$) and a marginally significant main effect for medium language competence ($\beta = 152.16$; SE [70.83]; $t = 2.15$; $p < 0.08$), but the condition did not affect the results in a significant manner. There were also no significant interactions which could be reported.

There were no significant main effects for the analysis in the presentation area although results also showed a positive effect for the machine-translated condition. There were also no interaction coefficients, but it might be worth mentioning that there was a negative interacting effect of the machine-translated condition and low language competence (although not significant), which could point towards a tendency that those participants were more drawn away from the presentation area.

There were no significant main effects in the speaker area as well as no interactions although, again, one could see a negative interactive effect of low language competence and the machine-translated condition.

Visit Count. Lastly, it was interesting to investigate the visit count in the different areas, but the results showed no significant differences in main effects across the different coefficients. Participants with the machine-translated mode of presentation showed less visit counts, but those were not statistically significant. However, this could indicate that their gaze either travels less in this area and the participants looked at this area less or, the other way around, their gaze stayed in the area for a longer time. There were also no significant interactions.

As for the presentation area, there were no significant main effects and no significant interactions.

For the speaker area, participants in the machine-translated condition showed a significant main effect (β = 13.65; SE [3.82]; t = 3.57; p < 0.01), which seemed surprising on the first glance, but since this only described the main effect, there might have been other factors influencing the result. However, interactions were not significant.

4.3 Finnish – English

Eight participants were presented with a Finnish video presentation and an English set of subtitles of which all datasets could be included in the analysis. Of those 8 participants, 4 saw the presentation in the machine-translated condition and 4 in the human-translated condition. Five participants stated having read the subtitles all the time (3 in HTST/2 in MTST), 1 stated 'sometimes' (HTST), 1 indicated 'no' (MTST). Also, one participant did not answer this question.

Further, 3 participants indicated having been 'entirely dependent' on the subtitling to understand the content (1 in HTST/2 in MTST), 1 was 'somewhat' dependent (HTST), and 4 indicated not having been dependent on the subtitling at all (2 in HTST/2 in MTST).

Total Fixation Duration. There was a significant main effect for total fixation duration and low Finnish language competence (β = 170.69; SE [59.50]; t = 2.87; p < 0.05), but there was no significant main effect for the subtitling condition. Further, there was an interaction between the machine-translated condition and low language competence, but this was not statistically significant.

For the presentation area, results showed a significant main effect for low language and medium language competence (low: β = -108.08; SE [11.00]; t = -9.83; p < 0.0006; medium: β = -83.96; SE [16.80]; t = -5.00; p < 0.007); interactions were not significant.

The analysis for the speaker area showed a significant main effect for low language competence (β = -91.75; SE [27.31]; t = -3.36; p < 0.03). Which was again in line with what could have been expected, i.e. participants with a low command of Finnish paid more attention to the subtitling area than to other parts of the screen.

Fixation Count. As for the fixation count, results showed significantly more fixations in the subtitling area when the language competence was low (β = 701; SE [208.5]; t = 3.36; p < 0.03), which was not surprising since one can suspect a higher dependence on the subtitling in this case. Interactions were not statistically significant. Maybe these results could then be again influenced by the proficient Finnish speakers who were presented with the machine-translated condition, which would balance out the result.

For the analysis in the presentation area, results showed a significant main effect for low and medium language competence (low: β = -513.70; SE [16.48]; t = -31.17; p < 6.31e-06; medium: β = -349.90; SE [25.17]; t = -13.90; p < 0.0001), which were in line with the finding presented above, i.e. when being strongly dependent on the subtitling, the attention shifted away from other parts of the screen, which then could have

reflected in less fixation counts. Still, there was no significant effect of the machine translation condition alone and interactions were not significant.

The analysis for the speaker area also showed a significant main effect of fixation count and low language competence (β = -131.30; SE [46.06]; t = -2.85; p < 0.05), but no effect for the condition. Interactions between the machine-translated condition and low language competence were also not significant.

Total Visit Duration. There was again a significant main effect for the total visit duration in the subtitling area for low and medium Finnish language competence (low: β = 257.50; SE [30.71]; t = 8.39; p < 0.001, medium: β = 172.60; SE [46.91]; t = 3.68; p < 0.02). The main effect for the machine-translated condition was not significant. Overall, interactions were also not significant.

For the presentation area, results showed a significant main effect for low and medium language competence (low: β = -157.38; SE [10.19]; t = -15.45; p < 0.0001; medium: β = -111.36; SE [15.56]; t = -7.16; p < 0.002); interactions with a machine translation coefficient were not significant.

The analysis for the total visit duration in the speaker area also showed a significant main effect for low and medium language competence (low: β = -105.34; SE [32.33]; t = -3.26; p < 0.03; medium: β = -107.71; SE [49.39]; t = -2.18; p < 0.09), although the effect for medium language competence was only marginally significant. Interactions between coefficients were not significantly different.

Visit Count. The analysis for the total visit count only showed a significant main effect for medium language competence (β = 103.90; SE [25.44]; t = 4.09; p < 0.02), which still could be an indicator that the gaze travels more with medium language competence, which then would reflect in more visit counts. Interactions were not significant.

For the presentation area, a significant main effect for low language competence (β = -32.80; SE [10.32]; t = -3.18; p < 0.03) was visible, however this result was not surprising or was in line with previous findings. But interactions were again not significant.

The results revealed a significant main effect for the total visit count in the speaker area for low language competence (β = -37.50; SE [11.66]; t = -3.22; p < 0.03), which again confirmed the previous findings in that manner that the gaze travels less when being dependent on the subtitling, therefore, one could see less visit counts. Still, there were no significant interactions. Therefore, the condition in which the subtitling was watched did not have a significant influence.

5 Conclusion

The main purpose of the eye-tracking experiment presented in this article was to investigate the reading or gaze behaviour of participants when presented with audiovisual material which was subtitled in different languages, regardless of whether they were

fluent in at least one of the languages of the video material. For example, fluent Finnish speakers were presented with foreign language subtitles to create equal conditions for both video presentations. The main research questions, or hypotheses, consisted of investigating to what extent participants would receive the presented subtitling, i.e. it was assumed that participants would pay more attention to the subtitling area when the subtitling is machine-translated, resulting in longer fixations and potentially more distraction. Further, this would also lead to a higher cognitive load.

Contrary to this assumption, the results could not confirm the hypotheses in either language pair, i.e. machine-translated subtitles do not seem to have a significantly distracting effect, especially with regard to the subtitling area, neither in the German presentation with Finnish subtitles nor the Finnish presentation with German or English subtitles. Occasionally, significant main effects could be found when it came to attention allocation to the speaker area and when it came to interactions with other coefficients, such as low language competence. On the other hand, results were very often significant (across the attention allocation for the different areas of interest) when it came to language competence. That is, the less competent a person is in the respective language of the video, the more the participant would pay attention to the subtitling area.

6 Limitations

There were some limitations to the study, first and foremost potential drawbacks in the study design: especially with regard to the area of interest, it has to be acknowledged that the subtitle area was designed as one big block at the lower part of the screen and the length of the subtitles in the different versions (human-translated and machine-translated) were not taken into account. This potential design lack was also due to compensating for potential inaccuracies in the calibration. That is, the areas were designed purposely bigger to retrieve more (inaccurate) hits. For instance, if the accuracy would cause that the gaze was recognized slightly above, below, a little to left or right from the actual subtitle. Further, this design (procedure) should also guarantee a certain test and analysis equality since the visual differences for the subtitling in the machine-translated and human-translated version are striking (e.g. the distribution over the whole length of the lower part of the screen in the machine-translated version vs. centredness in the human-translated version).

It should also be noted that the experiment most likely does not reflect an authentic reading or watching behaviour. First, participants were forced to see the presentation with foreign language subtitles although they also even might not know the language of the subtitles. Second, in this study, participants were not allowed to skip back and forth to check again on the information they received, which could eventually have an influence on their overall understanding, but also the gaze behaviour in general. For further research, it might be beneficial to investigate the reading, understanding, and gaze in a more natural setting, where skipping back and forth is allowed. In the experiment, I also did not analyze whether the familiarity of subtitles had an influence on the overall watching behaviour since the demographic data analysis showed that the

general experiment population was indeed quite familiar with subtitling, probably also due to the fact that people in Finland are exposed to subtitles on an everyday basis, so, after all, it was considered as an important factor.

Acknowledgments. The research is supported by the Faculty of Information, Technology and Communication at Tampere University. Thanks to the video presenters who provided the video material for the research and also thanks to the subtitler who created the subtitles for the research.

Disclosure of Interests. The author has no competing interest.

References

1. Gernsbacher, M.A.: "Video Captions Benefit Everyone." *Policy Insights from the Behavioral and Brain Sciences* 2 (1): 195–202 (2015). https://doi.org/10.1177/2372732215602130.
2. Negi, S., Mitra, R.: "Native Language Subtitling of Educational Videos: A Multimodal Analysis with Eye Tracking, EEG and Self-reports." *British Journal of Educational Technology* 53 (6), pp. 1793–1816 (2022). https://doi.org/10.1111/bjet.13214.
3. Van Hoecke, S.: *Subtitles for Access to Education: The Impact of Subtitles, Subtitle Language and Linguistic Proficiency on Cognitive Load, Comprehension, Reading and Processing in Different Styles of Asynchronous, Online University Lectures*. Vol. ix. Antwerp: University of Antwerp & Macquarie University (2023).
4. Zhang, D.: "Interactive Multimedia-Based E-Learning: A Study of Effectiveness." *American Journal of Distance Education* 19 (3), pp. 149–62 (2005). https://doi.org/10.1207/s15389286ajde1903_3.
5. Suojanen, T., Koskinen K., Tuominen, T.: *User-Centered Translation*. London: Routledge (2015).
6. Chan, W.S., Kruger J.-K., Doherty, S.: "Comparing the Impact of Automatically Generated and Corrected Subtitles on Cognitive Load and Learning in a First- and Second-Language Educational Context." *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 18, pp. 237-272 (2019). https://doi.org/10.52034/lanstts.v18i0.506.
7. Chan, W.S., Kruger J.-L., Doherty S.: "An Investigation of Subtitles as Learning Support in University Education." *Journal of Specialised Translation*, no. 38, pp. 155–79 (2022).
8. Orrego-Carmona, D.: *The Reception of (Non) Professional Subtitling*. PhD Thesis, Universitat Rovira i Virgili (2015).
9. Tuominen, T.: *The Art of Accidental Reading and Incidental Listening. An Empirical Study on the Viewing of Subtitled Films*. Tampere University Press (2013).
10. Tuominen, T.: "Multi-Method Research: Reception in Context." In: Di Giovanni, E., Gambier, Y. (eds.): *Reception Studies and Audiovisual Translation*, pp. 69–90. Amsterdam: John Benjamins Publishing Company (2018). https://doi.org/10.1075/btl.141.05tuo.
11. Tuominen, T., Koponen, M., Vitikainen, K., Sulubacak, U., Tiedemann, J.: "Exploring the Gaps in Linguistic Accessibility of Media: The Potential of Automated Subtitling as a Solution." *Journal of Specialised Translation*, no. 39, pp. 77–89 (2023).
12. Van Hoecke, S., Schrijver, I., Robert. I.: "Methodological Preparation of a Within-Subject Audiovisual Cognition, Reception and Perception Study." *Journal of Audiovisual Translation* 5 (1), pp. 94-128 (2022). https://doi.org/10.47476/jat.v5i1.2022.163.

13. Hu, K., O'Brien, S., Kenny, D.: "A Reception Study of Machine Translated Subtitles for MOOCs." *Perspectives* 28 (4), pp. 521–38 (2020). https://doi.org/10.1080/0907676X.2019.1595069.
14. Kruger, J.-L., Doherty, S.: "Triangulation of Online and Offline Measures of Processing and Reception in AVT." In: Di Giovanni, E., Gambier, Y. (eds.): *Reception Studies and Audiovisual Translation*, pp. 91–110. Amsterdam: John Benjamins Publishing Company (2018). https://doi.org/10.1075/btl.141.06kru.
15. Paas, F., Tuovinen, J.E., Tabbers, H., and Van Gerven, P.: "Cognitive Load Measurement as a Means to Advance Cognitive Load Theory." *Educational Psychologist* 38 (1), pp. 63–71 (2003). https://doi.org/10.1207/S15326985EP3801_8.
16. Sweller, J., Ayres, P. Kalyuga, S.: *Cognitive Load Theory*. New York, NY: Springer New York (2011). https://doi.org/10.1007/978-1-4419-8126-4.
17. Kalyuga, S.: "Instructional Benefits of Spoken Words: A Review of Cognitive Load Factors." *Educational Research Review* 7 (2), pp. 145–59 (2012).
18. Ayres, P., Sweller. J.: "The Split-Attention Principle in Multimedia Learning." In: Mayer, R. (ed.): *The Cambridge Handbook of Multimedia Learning*. 1st ed., pp. 135–46. Cambridge University Press (2005). https://doi.org/10.1017/CBO9780511816819.009.
19. Gambier, Y.: "Challenges in Research on Audiovisual Translation." In: Pym A., Perekrestenko, A. (eds.): *Translation Research Projects 2*, pp. 17–25. Tarragona: Intercultural Studies Group (2009).
20. Schierl, F. "Reception of Machine-Translated and Human-Translated Subtitles – A Case Study." In: Yamada, M., do Carmo, F. (eds.): *Proceedings of Machine Translation Summit XIX, Vol. 2: Users Track*, pp. 42–53. Macau SAR, China: Asia-Pacific Association for Machine Translation (2023). https://aclanthology.org/2023.mtsummit-users.4.
21. Laugwitz, B., Held, T., Schrepp, M.: 2008. "Construction and Evaluation of a User Experience Questionnaire." In: *Symposium of the Austrian HCI and Usability Engineering Group*, pp. 63–76. Springer (2008).
22. NASA. NASA TLX: Task Load Index (2006).