Journal of Audiovisual Translation

Modelling Gaze Behaviour in Subtitle Processing: The Effect of Structural and Lexical Properties

Juha Lång ✉
University of Eastern Finland

Hana Vrzakova ✉
University of Eastern Finland & Kuopio University Hospital

Lauri Mehtätalo ✉
Natural Resources Institute Finland

Abstract

One of the main rules of subtitling states that subtitles should be formatted and timed so that viewers have enough time to read and understand the text but also to follow the picture. In this paper we examine the factors that influence the time viewers spend looking at subtitles. We concentrate on the lexical and structural properties of subtitles. The participant group ($N = 14$) watched a television documentary with Russian narration and Finnish subtitles (the participants' native language), while their eye movements were tracked. Using a linear mixed-effects model, we identified significant effects of subtitle duration and character count on the time participants spent looking at the subtitles. The model also revealed significant inter-individual differences, despite the fact that the participant group was seemingly homogeneous. The findings underline the complexity of subtitled audiovisual material as a stimulus of cognitive processing. We provide a starting point for more comprehensive modelling of the factors involved in gaze behaviour when watching subtitled content.

Key words: audiovisual translation, eye-tracking, linear mixed-effect models, reception, subtitles.

Citation: Lång, J., Vrzakova, H. & L. Mehtätalo (2021). Modeling Gaze Behavior in Subtitle Processing: The Effect of Structural and Lexical Properties *Journal of Audiovisual Translation, 4*(1), 71–95.
https://doi.org/10.47476/jat.v4i1.2021.104

Editor(s): J.K. Kruger

Received: October 19, 2019

Accepted: January 5, 2021

Published: April 27, 2021

✉ jlang@uef.fi, https://orcid.org/0000-0002-8333-5775
✉ hanav@uef.fi, https://orcid.org/0000-0002-5624-8588
✉ lauri.mehtatalo@uef.fi, https://orcid.org/0000-0002-8128-0598

1. Introduction

Subtitled video content bombards our cognitive system with multiple feeds of information, both visual and aural (Gottlieb, 1998). This makes it, as a stimulus for our cognitive system to process, very different from typical narrative text. The factors involved in processing standard texts are fairly well known (see Rayner, 1998; Rayner & Pollatsek, 2006 for a review), but with audiovisual translation (AVT), of which subtitling is perhaps the most commonly used form, the picture is less clear. Compared to static text, subtitles are time-constrained, without a predictable duration of display, and are rendered on dynamic audiovisual stimuli. Due to the aforementioned aspects, the utilization of eye movement methodology in subtitle reception studies has not been as straightforward as in the case of typical reading studies (Kruger et al., 2015). The rapid advances in technology have helped to overcome some of these challenges and enabled a fast growth in the number of eye-tracking studies related to AVT in the last decade.

In recent years, the landscape of consuming audiovisual content has been in turmoil with the rise of internet streaming and video-on-demand services. Consequently, viewing habits have changed from traditional television and cinema screens to a wide variety of devices. Content has become more easily available and more easily consumed. This has possibly been a main factor in changes in the field of AVT as well, as subtitling has gained popularity among consumers as the preferred method of translation, even in countries traditionally classified as dubbing or voice-over countries (Perego et al., 2016).

Previous studies on the reception of subtitled audiovisual content have provided mixed results on the distracting effect that subtitles can have. On one hand, Lavaur and Bairstow (2011), for example, concluded that subtitles can have a negative impact on general comprehension, while Lee et al. (2013) showed that subtitles can cause extra strain to cognition. On the other hand, Perego et al. (2010) as well as Perego et al. (2015) found no negative effect on general content comprehension, scene recognition, dialogue recall, or general enjoyment of the film. In fact, subtitles have been shown to boost the recall of dialogue, especially in cases where the viewer can follow both the audio feed and the subtitles (Hinkin et al., 2014; Lång, 2016). Then again, Perego et al. (2018) showed that with videos that are characterised by a complex visual or narrative style, subtitles can increase the cognitive load in the overall processing of the stimulus.

If the issue is examined from a practical point of view, it can be assumed that in order to effectively process both the text of the subtitles and the image, the viewer must have enough time to divide their attention between the two. This is one of the main points of what have traditionally been considered good subtitles (Ivarsson & Carroll, 1998, p. 64). In this paper we investigate which textual and structural properties of subtitles have an effect on the time viewers spend looking at them. Although reception studies in AVT have become more frequent in the last decade, the field still lacks a comprehensive baseline regarding the factors involved in typical gaze behaviour when watching video material with interlingual subtitles. We build on prior research on reading of narrative texts and reception studies of subtitled audiovisual material, and by using statistical modelling we attempt to narrow this knowledge gap in order to gain better insight into the factors that have an impact on gaze behaviour when viewing subtitled audiovisual material.

2. Background

2.1 Previous Reception Studies on the Effect of Structural Properties of Subtitles

One of the pioneers in combining eye tracking and AVT is Gery d'Ydewalle. In one of his most extensive studies on the eye movement behaviour of people watching subtitled videos, d'Ydewalle and De Bruycker (2007) compared the gaze behaviour of children and adults when watching video fragments that had either native language subtitles and foreign language dialogue (standard condition) or vice versa (reversed condition). Although the study used more or less typical subtitled video material as the stimulus (in the standard condition), the only structural variable that was included in the analysis was the number of lines. The results showed that even when subtitles were in an unknown language, the viewers processed them to some degree. The participants' eye movement patterns resembled typical eye movements in reading more with two-line subtitles than with one-liners: with two-line subtitles, fewer subtitles were skipped (not fixated at all), fewer regressive fixations were made, and overall fixation count was higher (also resulting in a higher word fixation probability). The overall time viewers looked at two-line subtitles (proportional to subtitle duration) was significantly higher than with one-line subtitles (45% vs. 37%, respectively). The authors speculated that one possible reason for this is that short subtitles include redundant information that can be deduced from other information channels (i.e., the picture or the audio track, even when it is in a foreign language).

A similar increase in the relative dwell time on subtitles with a higher number of presented lines was reported by Szarkowska and Gerber-Morón (2018a). Normally, interlingual subtitles are presented in a maximum of two lines, but three lines of text are sometimes used in fan-made subtitles and subtitles for the deaf and hard-of-hearing (SDH). Szarkowska and Gerber-Morón (2018a) compared the reception of three-line and two-line (intralingual) subtitles in a study that included eye movement data and various self-evaluated cognitive measures. The participant group included normal hearing as well as hard-of-hearing and deaf participants. The number of lines did not negatively influence general comprehension in any of the participant groups, but cognitive load increased with the number of lines. With regard to enjoyment, deaf participants differed from those with normal hearing: normal hearing participants enjoyed two-line subtitles significantly more, while no significant effect was found with deaf participants. The increasing tendency in time spent on subtitles with the increase of lines was uniform across the participant groups.

Instead of structural properties of the subtitles, Jensema, Danturthi et al. (2000) concentrated on their temporal aspects. More precisely, they examined the effect of different subtitle presentation speeds (ranging from 100 to 180 words per minute) to the time deaf participants spent looking at subtitles, and found a small positive correlation between the presentation speed and the number of gaze points in the subtitle area. With a presentation speed of 100 words per minute, the participants spent 82% of the subtitle duration looking at the subtitles, and the proportional time increased to 86% with the presentation speed of 180 words per minute.

With a similar goal in mind, Szarkowska et al. (2011), Szarkowska et al. (2016), and Szarkowska and Gerber-Morón (2018b) have looked into the effect of different text editing styles and presentation speeds on the gaze behaviour of deaf, hard-of-hearing, and normal hearing people, with both intralingual and interlingual subtitles and various types of video clips.

Szarkowska et al. (2011) examined the effect of different subtitling styles and speeds on the gaze behaviour of deaf (nine participants), hard-of-hearing (21 participants), and normal hearing (10 participants) people. The stimulus material consisted of an animated film that was dubbed and subtitled in Polish, and the different subtitling conditions were verbatim, standard, and edited, with the presentation speeds of 13, 10, and 7 characters per second, respectively. Text editing had a significant effect on general comprehension: the scores were much better with verbatim subtitles (presented at a speed of 13 cps) than with edited subtitles in all participant groups. The analysis showed distinct differences between normal hearing participants and hearing-impaired participants in the total time they spent looking at the subtitle area. With normal hearing participants there were no statistically significant differences between the subtitling conditions. As the authors point out, one of the main limitations of the study is the rather small group size, especially for the deaf and normal hearing participants. It should also be noted that the standard presentation speed of Polish SDH is 12 cps, making the slowest speed used in the study much slower than the industry standard. Stemming from the very slow presentation speed, the text was also greatly edited and simplified, which may explain the lower comprehension scores.

Szarkowska et al. (2016) expanded on the issue with larger participant groups and with the inclusion of interlingual subtitles. The stimulus material included video clips with both intralingual (Polish-Polish) and interlingual (English-Polish) subtitles. The participants included a total of 144 Polish adults, 60 of whom had normal hearing. The different presentation speeds were 12 cps (in which the text was edited slightly) and 15 cps (in which the text was near verbatim rendering of the spoken audio). The analysis verified the previous results of Szarkowska et al. (2011) in that the gaze behaviour of normal-hearing and hearing-impaired differed noticeably in many respects. The relative dwell time on subtitles increased with the presentation speeds in all groups, but the normal hearing participants spent noticeably less time on subtitles than the hearing-impaired. Normal hearing participants also made fewer and shorter fixations and skipped subtitles (made zero fixations on them) more often. This is an important finding as it means that researchers should be careful when generalizing the results of subtitle reception studies across the two groups (i.e., deaf and hard-of-hearing vs. normal hearing people). Comprehension scores differed between the participant groups (with the normal hearing achieving the best results), but no difference was observed between the presentation speeds in any of the participant groups, contrary to the findings of Szarkowska et al. (2011).

In a more recent study, Szarkowska and Gerber-Morón (2018b) examined the effect of even faster subtitle presentation speeds in two experiments. Again, the analysed metrics were general comprehension and gaze behaviour, with the addition of self-reported measures of cognitive load (namely frustration, effort, and difficulty related to following the subtitles). The participant group (the same group in both experiments, 74 individuals in total) included native English, Polish, and Spanish speakers. The material included clips from motion pictures and television shows, which had in Experiment 1 Hungarian dialogue and English, Polish, or Spanish subtitles, according to the native language of the respective participant, and in Experiment 2 English audio with the subtitle languages similar to Experiment 1. The critical difference between the two experiments was in the knowledge of the language of the stimuli: none of the participants could understand Hungarian but all were proficient in English. The different subtitle presentation speeds were 12, 16, and 20 cps in Experiment 1, while in Experiment 2 only the two extremes (12 cps and 20 cps) were used.

The researchers found that the participants could cope well even with the fast subtitling speeds. In both experiments the proportional time spent on subtitles increased with the presentation speed, and this effect was uniform in all three participant groups. Comprehension was not affected in either experiment by the subtitling condition. Cognitive measures gave mixed results, as in Experiment 1 the slowest subtitles evoked the lowest scores on effort and difficulty, but no effect on frustration was observed. Contrary to this, in Experiment 2 no effect of speed was found on effort and difficulty, but frustration was significantly lower with faster subtitles. Furthermore, presentation speed did not have a negative effect on scene or subtitle recognition, suggesting that the participant could follow both the image and the subtitles effectively, even with the fastest subtitle conditions. In fact, slower subtitles were generally enjoyed less, especially in the case of intralingual subtitles.

Szarkowska and Gerber-Morón (2018b) concluded that the results confirm the subtitle effectiveness hypothesis first introduced by Perego et al. (2010), which states that subtitles can be processed effectively and that they do not hinder the processing of other visual elements of the video. Despite this, the external validity of this hypothesis can be questioned, since Perego et al. (2018) showed that on video stimuli that have complex narrative or visual structure subtitles can further increase the difficulty of processing.

2.2 Factors in Reading

Although Jensema, El Sharkawy et al. (2000, p. 275) stated that adding captions to a video resulted in “the viewing process becoming primarily a reading process,” some differences in typical gaze behaviour between reading conventional texts and reading subtitles have been reported (d’Ydewalle & De Bruycker, 2007). The multimodal context, in which subtitles are presented, is one probable cause of these differences, as previous research has identified noticeable differences in the gaze behaviour of different individuals when reading typical linear texts, but also within individuals between different tasks, text formats, and contexts (Radach et al., 2008).

There have been several different approaches to the modelling of eye movement control in reading (see Rayner, 2009; Snell et al., 2018, for an overview). Although the models have some fundamental differences (the main one being whether processing words is seen as a parallel or a serial process), they all share the core concept that processing difficulty is related to the lexical and semantic properties of words and the relationships between words. The most often used variables for the quantification of processing difficulty are word length, frequency, and predictability.

The empirical basis for using these three factors as measures of processing difficulty is strong. To briefly sum up the main findings of decades of eye tracking research on reading, words that are short, more frequent in general language use, or easier to predict from the textual context, attract shorter fixations (Kliegl et al., 2004; Rayner, 1998). Although every word is typically fixated approximately once, short words are also skipped more often than long words, and long words in turn attract refixations more often (Brysbaert & Vitu, 1998). High word frequency and predictability have also been associated with increased probability of skipping the word (Rayner & Raney, 1996; Rayner & Well, 1996). It has been speculated that the skipped words, especially when they are short high-frequency words, are actually processed while the reader is fixated on the previous word, which results in the inflation of that fixation (Kliegl & Engbert, 2005).

Most eye tracking studies on reading have used English as the source language. This raises the question of how well the results compare to other languages. Liversedge et al. (2016) compared reading behaviour across three languages: English, Finnish, and Chinese. They found that there were surprisingly strong similarities in the eye movements of native speakers of the three languages when variation caused by differences in linguistic characteristics and writing systems was accounted for. This means that languages that use the same writing system and use similar semantic structures should be comparable in terms of gaze behaviour.

3. Methods

3.1 Experiment Task, Stimulus Material, and Test Procedure

The experiment that is reported here was conducted in a soundproof studio. The participants received written instructions that explained the procedure. After this, the eye tracking equipment was calibrated with a 3-point calibration method and the calibration was validated. The participants then continued with the experiment task, which consisted of watching a short documentary film and comprehension testing with a written questionnaire about the contents of the documentary. The questionnaire had a total of 28 open-ended questions, 12 of which concentrated on the contents of the subtitles (see Lång, 2016, for further details about the questionnaire).

The topic of the documentary was Fridtjof Nansen, the Norwegian explorer and Nobel Peace Prize laureate. It was narrated in Russian and subtitled in Finnish. The soundtrack included only the narration and subtle background music, that is, no sound effects or spoken dialogue. The total duration of the video was approximately 7 minutes. The subtitles were composed according to Finnish subtitling standards and presented in white text without a black background, aligned to the left of the screen.

3.2 Participants and Apparatus

A total of 20 participants took part in the experiment and they received a meal coupon as compensation for their time. The data of six participants was omitted from the analysis due to its low quality, more specifically, miscalibration of the equipment, errors in fixation filtering (abnormal long or short fixations), or high proportion of missing data. Consequently, the analysis reported here is based on the data of 14 participants (10 females and 4 males, mean age = 24 years, *SD* = 3.088). All participants had normal or corrected-to-normal vision. The participants' reading skills were not tested, but as they were all university students (with majors including linguistics, media, theology, and literature) and of fairly similar age, noticeable differences in reading skills were not assumed. All participants were native Finnish speakers and had little to no knowledge of the Russian language (by their own assessment). We included the demographic descriptors, namely participants' age, gender, and whether they had any knowledge of the source language (as a binary variable), in the model development. In addition to these, we also used participants' comprehension scores, that is, the percentage of correct answers to questions about the contents of the subtitles, in the modelling process.

Participants were equipped with over-ear headphones to reduce outside noise and to provide clear audio. Video stimulus was projected on a 22-inch LCD monitor connected to a computer. SMI Eye Tracking Glasses 2.0 (ETG 2.0), with a binocular sampling rate of 30 Hz, were used to record participants' eye movements and the scene view using SMI Experiment Center 3.6. The relatively low sampling rate was considered to be an acceptable compromise for spatial accuracy and portability: As the eye tracker is worn as glasses, head or body movement does not affect the calibration of the tracker and there was no need for head-stabilizing aids (e.g., chin rests or bite bars). Fixations were detected using the SMI ETG Event Detection algorithm, which is a velocity-based algorithm (SensoMotoric Instruments, 2015, p. 366). The statistical analysis was conducted in R (version 3.4.4, R Core Team, 2018) using the packages lme4 (Bates et al., 2015), nlme (Pinheiro et al., 2018), and lmerTest (Kuznetsova et al., 2017).

3.3 Structural and Temporal Characteristics of Subtitles

Each subtitle was characterised by the following measures: number of lines, words, and characters (including spaces) in the subtitle, the time the subtitle was visible (duration of the subtitle), and average word length in the subtitle. We also included the presentation speed of the subtitle (duration of the subtitle divided by number of characters, including spaces) because this metric is commonly used in the subtitling industry to ensure the readers have enough time to read a subtitle. The subtitles were timed so that they followed the pace of the video (narration and style) as closely as possible, which meant that the presentation speed was not constant across subtitles.

Previous research on reading has shown that word frequency and predictability are good predictors of processing difficulty and speed (Engbert et al., 2005; Rayner, 1998, 2009; Reichle et al., 2003). Of these two we chose to include only word frequency in our analysis. This measure was calculated from the Finnish Sub-corpus of the Newspaper and Periodical Corpus of the National Library of Finland, Kielipankki Version (National Library of Finland, 2011), which consists of Finnish journals and periodicals published in the 1990s and 2000s (a total of 149.38 million words). Because the analysis was grouped by subtitle, the word frequencies had to be averaged for each subtitle. The frequencies, expressed as occurrences per a million words, had an extremely wide range and high standard deviation, which signals high variability. This fact could undermine the analysis since extremely high values could skew the means and eliminate the effect of low frequency words. Thus, to minimize the effect of extreme values, we transformed the word frequencies using the common logarithm before averaging them for each subtitle.

Table 1 outlines the structural and lexical properties of the subtitles used in the analysis. The stimulus video included a total of 89 subtitles, but approximately one minute from the beginning of the stimulus video (12 subtitles) was excluded from the analysis as an adjustment period. Thus, the analysis was executed with a set of 77 subtitles, which covered 79.28% of the run-time of the stimulus. Out of this subset, 9 were single-lined subtitles, and 68 subtitles consisted of two lines of text. The video included short periods where there was no narration and thus also no subtitles visible. These were obviously not included in the analysis.

Table 1

Overview of Lexical and Structural Properties of the Subtitles

Property	Mean	Median	Min	Max	*SD*
duration of subtitles (seconds)	3.96	3.58	1.75	7.84	1.32
number of words	5.55	6	2	9	1.52
number of characters (including spaces and punctuation)	45.71	46	16	66	10.68
average word length (in characters)	7.51	7.14	5	12	1.49
presentation speed (characters per second)	12.20	12.29	6.47	20.95	3.26
mean frequency of words in a subtitle	1802.36	871.14	8.30	8406.07	2034.50
mean log frequency of words in a subtitle	1.83	1.80	0.43	3.23	0.58

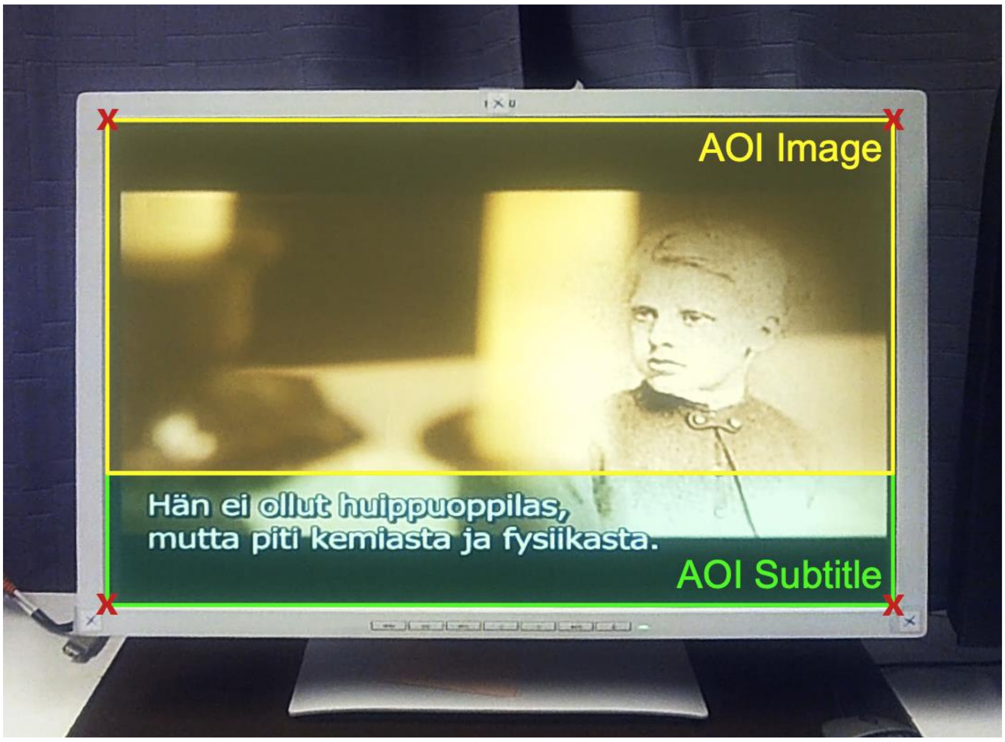
3.4 Spatial and Temporal Mapping of Eye Gaze to Subtitles

The SMI ETG 2.0 eye tracker produces two streams of data: a video with egocentric scene view and gaze coordinates with related properties of recorded eye movements. The gaze coordinates are mapped onto the scene view video, which means that if analysis is directed to specific areas of interest (AOIs), these must be defined separately on each participant's scene view video. In order to estimate dynamic AOIs and to map participants' gaze data to these areas (taking into account their head movement), we created a three-step process that included: 1) extracting the coordinates of the stimulus screen and relevant AOIs inside the screen area on each participants' scene view video; 2) annotating the timestamps of the subtitles in the scene view video; and 3) mapping the fixations into relevant AOIs using the annotated timestamps and coordinates, and calculating the eye-tracking metrics. Motion tracking and spatial annotations of the stimulus screen in the scene videos were performed in Kinovea (version 0.8.24, Charmant & contributors, 2014). Temporal annotations of subtitles in the video were done in ELAN version 4.9.2 (Brugman & Russel, 2004). All data streams were joined and analysed with custom-made scripts using Python 2.7 with Pandas (McKinney, 2010), Numpy (van der Walt et al., 2011) and Shapely (Gillies et al., n.d.).

First, the area of the screen in the scene video was annotated frame-by-frame with Kinovea (Charmant & contributors, 2014) using key reference points around the stimulus screen, as illustrated in Figure 1. As a result, each frame of the scene view video was defined by a bounding box specifying the position of the stimulus screen in each participants' gaze data. Next, the bounding coordinates were processed to define the relevant AOIs, with respect to stimulus scaling and the rotation of the scene. The Subtitle AOI consisted of the bottom 25% of the screen, and Image AOI covered the rest of the stimulus screen (75%) (see Figure 1).

Figure 1

A Screenshot of a Participant Scene View Video, With Overlaid Key Reference Points and Areas of Interest


AOI Image
Hän ei ollut huippuoppilas,
mutta piti kemiasta ja fysiikasta.
AOI Subtitle
X
X
X
X

Note. Key reference points used in detecting the stimulus screen location are marked with a red letter X. The green and yellow areas represent the Subtitle AOI and Image AOI, respectively.

Although each participant watched the same video, the length of the overall scene view video differed between participants. The reason for this is that before starting the playback of the stimulus video, calibration of the eye tracker was validated. Therefore, the subtitle time stamps were annotated into each participant's scene video manually with ELAN, with 1ms precision. The third step of data processing consisted of combining the gaze event data, the subtitle time codes (defined by the annotations from ELAN), and the AOIs (defined by the coordinates from Kinovea). The resulting data frame allowed us a direct comparison of each participant's gaze data and to calculate the desired metrics for statistical analysis. Finally, participants' visual attention in each subtitle was calculated as total dwell time, where a single dwell represents the sum duration of all events, that is, fixations, saccades, and blinks, that occur between the first and last fixation in the subtitle area. We opted for this rather inclusive definition of dwell in order to minimise the effect of noise in the data. During the initial examination of the fixation data, we observed that the event detection filter produced an abnormally large proportion of short fixations (fixation durations less than 100 ms). One possible explanation could be that the SMI ETG Event Detection algorithm struggled to identify separate fixations in the cases when participants were moving their heads while fixating at a stationary stimulus.

3.5 Linear Mixed-Effect Model of Visual Attention to Subtitles

We analysed participants' total dwell durations in the subtitle area using a linear mixed-effects model and built the statistical model bottom-up with the library lme4 (Bates et al., 2015). We chose linear mixed-effects modelling because of the two-level grouped data design (subtitles crossed with participants). Table 2 summarises all independent variables that were systematically employed in the model development.

We began building the model by including only the random effects and residual error and tested iteratively which of the possible independent variables as a fixed effect improved the model fit the most. Model fitness was evaluated graphically by plotting the random effects against the possible independent variables, and by comparing the log-likelihood and Akaike Information Criterion (AIC, Akaike, 1973) scores of the model vs. the restricted model. The significance of the variables included in the final model was tested using t-test with Satterthwaite's method (with library lmerTest, Kuznetsova et al., 2017) and the restricted maximum likelihood method was used for estimating the parameters of the final model.

The data violates the prerequisite of homoscedasticity of residual variance, but this was modelled by giving the residuals a weight as a function of fitted values (Pinheiro & Bates, 2000, p. 208). The value for δ (see Equation 2 below) was first estimated with R library nlme (Pinheiro et al., 2018) by building a model with subtitles as random intercept and character count, subtitle duration, participants, and the interaction of subtitle duration and participants as fixed effects, and using the argument varPower. The estimated δ was thereafter used as a fixed value to determine the weighting in the final model fitted using lme4 (Mehtätalo & Lappi, 2020).

Table 2

Independent Variables Included in the Modelling Process

Name	Description	Variable type
Subtitle ID (*)	Identifier for the subtitle	categorical
Participant ID (*)	Identifier for the participant	categorical
Subtitle properties:		
Subtitle duration (*)	Duration of the subtitle	numerical, continuous
Character count (*)	Number of characters in the subtitle, including spaces	numerical, discrete
Word count	Number of words in the subtitle	numerical, discrete
Line count	Number of lines in a subtitle	categorical, binary
Subtitle speed	Presentation speed as characters per second	numerical, continuous
Mean word length	Character count divided by word count	numerical, continuous
Word frequency	Mean frequency (per one million words) of all the words in the subtitle, calculated from the Finnish newspaper corpus (National Library of Finland, 2011) with common logarithm transformation	numerical, continuous
Participant-related variables:		
Age	Participant’s age	numerical, discrete
Gender	Participant’s gender	categorical, binary
Language skill	Participant’s Russian language proficiency	categorical, binary
Comprehension score	The percentage of correct answers to questions about the contents of the subtitles	numerical, continuous

Note. The variables included in the final model are marked with an asterisk (*).

The final model is defined in *Equation 1*, where y_{ij} denotes the total dwell time in subtitle i for participant j. β_1 and β_2 are the fixed effect coefficients for the subtitle duration and the number of characters, respectively, while $b_i^{(1)}$ is the random intercept for subtitle i, and $b_j^{(2)}$ the random slope of subtitle duration for participant j, independent of $b_i^{(1)}$. The residuals ε_{ij}, independent of both $b_i^{(1)}$and $b_j^{(2)}$, were weighted as a function of fitted values $\hat{y}_{ij}$ (*Equation 2*).

$$y_{ij} = \beta_0 + \beta_1 x_{ij}^{(1)} + \beta_2 x_{ij}^{(2)} + b_i^{(1)} + b_j^{(2)} + \varepsilon_{ij} \qquad (1)$$
$$i = 1 \dots 77, j = 1 \dots 14$$

$$var(\varepsilon_{ij}) = \sigma^2 |\hat{y}_{ij}|^{2*\delta} \qquad (2)$$

Interestingly, none of the participant-related variables (age, gender, language familiarity, or comprehension score) improved the model fit. Similarly, neither did the rest of the structural characteristics of the subtitle, such as the number of words or lines, presentation speed, mean word length, or word frequency. With respect to the number of words, this finding was expected, because the number of words and characters was strongly correlated (Pearson's *r* = 0.785, *p* < .001). Out of those two measures, the number of characters was included in the model because it improved the model fit more. Similarly, presentation speed was omitted from the model because, although the log-likelihood and AIC scores suggested that it could improve the model fit, it correlated with subtitle duration (Pearson's *r* = −0.628, *p* < .001) and character count (Pearson's *r* = 0.279, *p* = 0.014).

4. Results

Table 3 shows the estimates for the variances of the random effects and fixed effects in the final model. The model shows that both the temporal and the structural length of the subtitle have significant effects on the total dwell time. On average the dwell time increased 286.13 ms for each second the subtitle was visible, *SE* = 57.70, *t*(30.93) = 4.96, *p* < 0.001, and by 16.37 ms for each character in the subtitle, *SE* = 4.51, *t*(67.67) = 3.63, *p* < 0.001.

Table 3

Final Model Estimates

Fixed effects

	Estimate	Std.Error	*df*	*t*	*p*
β_0 (Intercept)	-328.761	188.598	62.558	-1.743	0.0862
β_1 Subtitle duration	286.132	57.697	30.929	4.959	< 0.001
β_2 Number of characters	16.366	4.511	67.672	3.628	< 0.001

Random effects

Grouping factor	*SD*	95% CI
Subtitle (random intercept)	313.148	(247.092, 381.745)
Participant (random intercept)	227.065	(85.999, 409.707)
Subtitle duration per participant (random slope)	160.467	(103.847, 242.371)
Correlation between random intercept and slope	-0.952	(-1.000, -0.637)

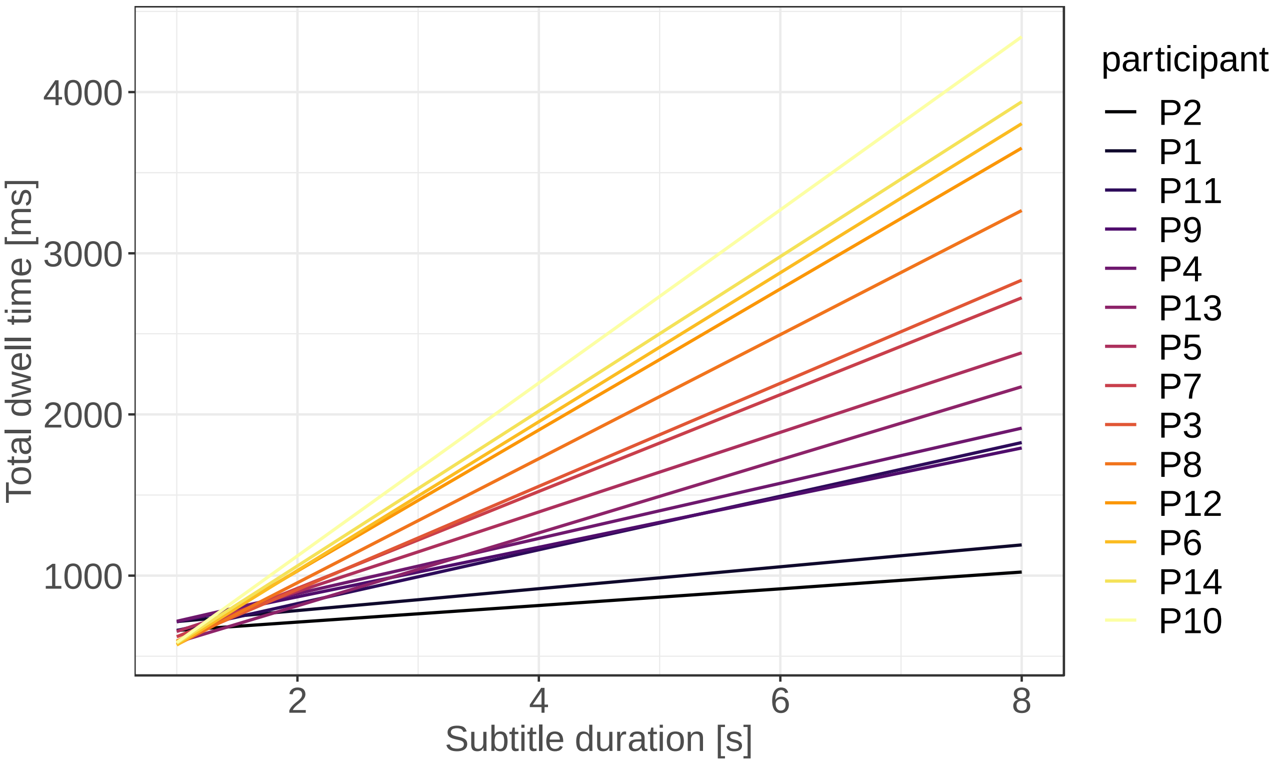
Residual variance parameters	*SD*	95% CI
σ	5.783	(5.536; 6.051)
δ	0.65	-

To get a more meaningful interpretation of the model, we can use it to make predictions of average gaze behaviour. The typical subtitle in our data consisted of 46 characters (6 words) and was visible for 3.58 seconds. In this case, the model predicts that an average viewer will spend $-328.76 + 46 * 16.37 + 3.58 * 286.13 = 1448.605$ milliseconds looking at the subtitle area. That is approximately 40% of the time the subtitle is visible. If the subtitle is one second longer (and the character count stays the same), the total dwell time on subtitles increases to 1734.735 milliseconds (which is 48% of the subtitle duration). Likewise, if the subtitle included one average length word (8 characters) less, the total dwell time decreases by 8 * 16.37 = 130.96 to 1317.645 milliseconds (37% of subtitle duration).

The standard deviation of subtitles in the random part (313.15 ms) expresses the unexplained variation between the groups (i.e., variation in total dwell times between different subtitles that is not explained by the subtitle characteristics described in Table 2). The number is quite high compared to the estimates of the fixed effects, which suggests that the total dwell time is affected by factors that were not considered in the modelling process. These will be discussed in more detail in Section 5.

Figure 2

Model estimates of the random effect of the subtitle duration for each participant (all other model variables remain constant). The figure shows that the differences between participants in the total dwell time become more pronounced with longer subtitles[1]



Likewise, the random part shows that there are noticeable differences between participants, as the standard deviations for both the random intercept and the random slope of subtitle duration between participants are quite high (227.07 ms and 160.47 ms, respectively). Figure 2 illustrates the model estimates of total dwell durations for different participants as the duration of the subtitle increases and other model variables remain constant. The minimum and maximum slopes are noticeably different (model coefficients min = 51.78 ms for P2; max = 536.74 for P10). The figure also shows that participants are spread normally between the extremes, suggesting that the extremes are

[1] Figure 2 was created using the R library sjPlot (Lüdecke, 2018).

not simple outliers or anomalies, but that a similar distribution could be found even with larger participant groups.

In relation to the typical subtitle in our data, the difference between minimum and maximum coefficients becomes even more pronounced. The model predicts that participant P2 looked at an average subtitle the total of 873.57 ms, which is only 24% of the subtitle duration. At the other extreme, the model predicts that participant P10 looked at the same subtitle for 2050.46 ms, 57% of the subtitle duration.

5. Discussion

Watching subtitled audiovisual content is a dynamic and complex process. Previous studies on subtitle reception have usually concentrated either on intralingual subtitles (Jensema, Danturthi, et al., 2000; Jensema, El Sharkawy, et al., 2000; Szarkowska & Gerber-Morón, 2018a; Szarkowska et al., 2011), or on exploring some specific phenomenon, such as text editing and presentation speed (Gerber-Morón et al., 2018; Rajendran et al., 2013; Szarkowska & Gerber-Morón, 2018b; Szarkowska et al., 2011) or the number and position of lines (Caffrey, 2008; Szarkowska & Gerber-Morón, 2018a). Our goal was to describe, with statistical modelling, the typical gaze behaviour of viewers looking at interlingually subtitled video material, and to identify the lexical and structural properties of subtitles that have an impact on gaze behaviour.

5.1 Significant Effect of Subtitle's Temporal and Structural Characteristics

Our model shows that both temporal and structural attributes of the subtitles have an effect on how long viewers look at subtitles. The data was grouped into units of analysis according to subtitles, which means the duration of the subtitle provided the time restraints for the dwell durations. Thus, it is perfectly logical that we found a strong positive correlation with the total dwell time and the duration of the subtitle. Along with this, the model was able to distinguish the effect of character count on the time viewers spend looking at the subtitles.

Character count naturally correlates heavily with word count and word length, which means that in our model it can be seen to represent the lexical structure of the subtitle. Out of the oculomotor movements (categorised by SMI event detection algorithm as fixations, saccades, and blinks) that constitute a dwell, fixations are typically the longest, which means that the increase in total dwell time in our data is most likely caused by an increase in fixation duration or fixation count (including re-fixations or regressive fixations). With a typical linear text each word is fixated approximately once, while long and semantically complex words can draw re-fixations (Rayner, 1998). Indeed, word length is considered to be one of the “Big Three” factors that have most influence on how fast a word is processed (Clifton Jr. et al., 2016, p. 5). In our data, higher character counts represent more words or longer words, which means that the model is in agreement with previous research.

Nevertheless, it is interesting that the number of words or the average word length in a subtitle (separately or together) had less predictive power in our model than the total character count. One probable reason for this lies in methodological issues which we will discuss in section 5.3. It is possible that the effect of word count and word length was masked by the fact that we examined dwell durations at the level of the subtitle, and, consequently, averaging the word length, for instance, reduces the predictive power of the variable.

The ratio of the character count to the duration of the subtitle is commonly used to estimate whether viewers have enough time to read the subtitles. In this study we called this ratio presentation speed, and it was one of the possible independent variables in the model-building process. It was not included in the final model because the two elements of the ratio provided a better fit separately. Nevertheless, the model estimates show that presentation speed is a valid metric for assessing the proportional time viewers spend looking at the subtitles: by increasing the subtitle duration or decreasing the character count the presentation speed decreases, and the model predicts that the proportional time spent looking at the subtitles also decreases. Similar results have been reported previously in studies involving both intralingual (Jensema, Danturthi, et al., 2000) and interlingual subtitling (Szarkowska & Gerber-Morón, 2018b). In contrast, (Szarkowska et al., 2011) discovered no effect of presentation speed with normal hearing participants watching a video with intralingual subtitles.

The predicted proportional dwell times (40% for the median subtitle) are somewhat consistent with previous findings on how attention is divided between subtitles and image when watching foreign language videos with interlingual subtitles. Szarkowska et al. (2011) reported that normal hearing participants spent approximately 45% of the subtitle duration looking at the subtitles. In contrast, in the studies by d'Ydewalle and De Bruycker (2007) and Szarkowska et al. (2016) the data showed lower percentages (d'Ydewalle & De Bruycker, 2007: 31% one-liners and 37% for two-liners; Szarkowska et al., 2016: 34%). The model reveals a possible cause for these discrepancies, as it predicts that the proportional dwell time is linked to the ratio of subtitle duration to character count. It should be noted that the definition of dwell is not identical in the cited studies: Szarkowska et al. (2011) define dwell as the sum of the fixation durations in the subtitle area, while d'Ydewalle and De Bruycker (2007) and Szarkowska et al. (2016) use the sum of both fixation and saccade durations. In the present study the definition of dwell included all events that occur between the start of the first and end of the last consecutive fixation in the relevant area of interest.

5.2 Individual Differences

The results also revealed noticeable differences between the participants (see Figure 2). Previous studies on reading have shown that factors such as age (Rayner et al., 2006), reading skill (Ashby et al., 2005), or even cultural background (Chua et al., 2005; Rayner et al., 2007) can have a significant impact on eye movement behaviour. In respect of video material, people tend to look first at the centre of the screen, but disparities in the gaze behaviour between individuals grow larger as the scene continues (Brasel & Gips, 2008; Dorr et al., 2010). The content of the video has an impact on this: scenes that have been designed to draw gaze to certain parts of the screen usually have less dispersion than natural scenes, which suggests that it is possible to guide the viewers' attention at least to a degree (Goldstein et al., 2007).

Reflecting on these previous studies, it is not surprising that inter-individual differences were observed in the present study. What is interesting, though, is the scale of the differences: with the average subtitle the proportional time spent looking at the subtitles varied by a factor of two. It was unexpected to discover such large differences between participants who formed a demographically homogeneous group: they were all Finnish university students in a narrow age range, and only three of the participants declared having introductory-level Russian skills (the rest reported having no Russian skills). The differences cannot be explained by comprehension scores; including comprehension scores in the model did not improve the model fit, which means that more time spent looking at the subtitles did not result in higher comprehension.

This finding has considerable methodological relevance. Despite the vast technological advances, eye tracking studies are still relatively labour-intensive to conduct, in both the data gathering and analysis stages. This is especially true when dealing with multimodal stimuli, such as subtitled videos. This has led, at least in the past, to studies usually including a fairly small number of participants. The fewer participants there are, the stronger the influence of a single participant on the results, which means that with small participant groups one idiosyncratic participant can skew the results and cause false interpretations. Of course, when conducting statistical analysis, the data is usually scanned for outliers, which are then controlled for in one way or another. The analysis presented in this study has shown that the participants were spread evenly across the extremes, which means that the perceived variance is not a matter of outliers, but natural inter-individual variation. In order to reach reliable and repeatable results, this must be taken into account when conducting eye tracking studies of audiovisual material.

5.3 Non-Significant Effects and Limitations of the Study

The variables that were included in the final model provide interesting information about the factors that affect gaze behaviour when watching subtitles audiovisual content, but equally interesting are the variables that did not contribute to the model fit.

Contrary to previous eye tracking studies on watching subtitled audiovisual material (d'Ydewalle & De Bruycker, 2007; Szarkowska & Gerber-Morón, 2018a), our data manifested no significant effect of the number of lines on the total dwell time. Quite possibly the reason for this is that the data was skewed heavily in favour of the two-line subtitles (only 9 out of the 77 analysed subtitles had one line of text), and the data for one-liners was simply too scarce to reveal statistically significant differences. An alternative explanation is that the fixed effects included in the model accounted for the differences, which means that the differences identified in previous studies actually stem from other structural properties. In other words, lower proportional dwell time with one-lined subtitles versus two-line subtitles could be explained by the tendency of one-liners to consist of fewer and/or shorter words. Unfortunately, we cannot confirm the validity of this explanation with our data.

Another variable that unexpectedly did not have a significant effect in our model is word frequency. Word frequency is one of the most commonly used metrics of semantic complexity in models of normal reading (Clifton Jr et al., 2016; Engbert et al., 2005; Reichle et al., 2003), which is why it was expected to have predictive power in our model as well. One possible reason for why this effect was not found in our model is the lack of accuracy: the word frequencies were averaged per subtitle, which may have masked the effect of frequency altogether.

An alternate explanation may be a methodological issue that is common to all eye tracking studies that deal with subtitles (originally pointed out by Kruger & Steyn, 2014, p. 109). The problem is that it is difficult to make a reliable distinction between looking at the subtitles and reading them, especially with the procedure and the equipment used in the present study. With our method we could not map fixations to single words, which is one of the main reasons why we decided to take a dwell-based approach in our analysis instead. As stated before, fixation durations most likely contribute the most to the total dwell time. As the link between word frequency and fixation durations in normal reading is well documented, and as there is no reason to assume that processing text in subtitles differs from normal text in this regard, the logical conclusion is that the dwells in the subtitle area include fixations that are not connected with text processing.

This is not an unreasonable conclusion. Subtitles are, after all, usually placed so that they overlap with the image area, as was also the case in the present study (see Figure 1). Although the main visual elements in the stimulus video are placed in the central parts of the screen, that is above the subtitle area, the bottom of the screen (the subtitle area) may have included some visually interesting elements. These elements would be a likely trigger for fixations that cause the “noise” in the data, that is fixations which are unrelated to the processing of the text. Consequently, the noise could mask effects that would be identified if the analysis were able to isolate the gaze data that are relevant to processing textual information.

In order to distinguish actual reading behaviour from other cognitive processes, first we would need more accurate eye tracking equipment. In this study we used SMI Eye Tracking Glasses 2.0, which are somewhat lacking in temporal accuracy (although Andersson et al., 2010, have argued that low temporal resolution can be compensated for by a larger sample size). Second (and perhaps more important) requirement is a reliable procedure for distinguishing fixations that are involved in actual reading processes from the “noise”. One possibly viable tool for this would be the Reading Index for Dynamic Texts (RIDT), proposed by Kruger and Steyn (2014). RIDT takes into account the ratio of fixations to words in a subtitle, and the ratio of forward saccade length to word length, to calculate a comparable score that represents the eye movements that are likely connected to text processing. The prerequisite for using RIDT is that the direction of saccades can be extracted from the data, which would not have been easily achieved in our data processing procedure.

In addition to between-subjects variance, the model also showed considerable variance between subtitles. Since the model attempted to control the effect of lexical and structural properties of the subtitles, and individual differences, it can only be concluded that the remaining variance is caused by factors that were not considered in the modelling process. One viable metric of lexical complexity that was left out is word predictability. The reason for this omission was already explicated in Section 3, and in order to effectively include the variable in the modelling process, the data would have to be mapped with word-level precision.

Another possible cause of the between-subtitles variation is the visual context. The contents of the image have been proven to have an impact on eye movements both when looking at print advertisements with text (Rayner et al., 2008) and when watching movie clips (Brasel & Gips, 2008; Dorr et al., 2010; Goldstein et al., 2007). Subtitles have also been proven to increase the cognitive load in films that have complex visual or narrative structures (Perego et al., 2018). The stimulus video used in the current study was characterised by a fairly slow-paced visual style. It consisted mostly of still photographs that were slowly panned across and zoomed in on, and there were only a few short acted scenes. Nevertheless, each subtitle was presented in a visual context that differed from other subtitles, even if the change was ever so slight in the case of consecutive subtitles. Some scenes may have included elements that were especially interesting (for example, people's faces, or pictures containing interesting details), which could have caused participants to spend more time on the image than on the text, while if the image contained few interesting elements, the viewer may have spent more time on the text. In addition, as discussed previously, the subtitle area overlapped partly with the image, and the visual elements in the subtitle area may have drawn fixations that are unrelated to text processing.

The result highlights the importance of the visual context, an element which is often completely neglected in reception studies of audiovisual content. Parametrizing the visual elements should have a high priority if further attempts to model eye movement behaviour when watching subtitled videos are made. One possible approach to this is the concept of visual saliency (Elazary & Itti, 2008).

Furthermore, there is also the concern for the external validity of the results. The study was conducted in Finland with native Finnish speakers. In Finland, subtitling is the main method for translating foreign audiovisual material, which suggests that Finnish people become accustomed to watching subtitled television material at a very young age. Nevertheless, subtitling has become more and more common even in countries that have traditionally preferred other forms of audiovisual translation as consumer viewing habits have changed from traditional television to more versatile sources, with various internet services spearheading the change. Although there is some evidence of universal tendencies in reading (Liversedge et al., 2016) and of the reception of subtitled video content (Perego et al., 2016), too little of this issue is known to make strong claims about how the findings presented here can be generalised across different languages or viewing habits.

6. Conclusions

Watching subtitled television content is cognitively a very complex process, involving multiple channels of information that need to be processed either in parallel or quickly in sequence. There are multiple variables that have an effect on the difficulty or ease of processing the information, and this is reflected in the gaze behaviour.

In this experiment we identified factors that have an impact on the time viewers spend looking at the subtitles when watching a subtitled video. The duration of the subtitle and the number of characters had a significant effect, which means that the viewer's gaze behaviour is affected by both temporal and structural properties of the subtitle. Our model also revealed noticeable differences between participants, even though the participant group was homogeneous in terms of age and there was no reason to assume noticeable differences in reading skills.

The model presented in this study can act as a starting point for a more comprehensive and accurate model of gaze behaviour of viewers of subtitled audiovisual material. An important aspect that was ignored in the present model is the visual contexts of the subtitles, and future studies should attempt to rectify this shortcoming. It would also be beneficial to attempt to duplicate our results with different age groups, language pairs, and types of stimulus video to test the universality of the results.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N., Petrov, & F. Caski (Eds.), *Proceedings of the second international symposium on information theory* (pp. 267–281). Akademiai Kiado.

Andersson, R., Nyström, M., & Holmqvist, K. (2010). Sampling frequency and eye-tracking measures: How speed affects durations, latencies, and more. *Journal of Eye Movement Research*, *3*(3), 1–12. https://doi.org/10.16910/jemr.3.3.6

Ashby, J., Rayner, K., & Clifton Jr, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology Section A*, *58*(6), 1065–1086. https://doi.org/10.1080/02724980443000476

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using {lme4}. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Brasel, S. A., & Gips, J. (2008). Points of view: Where do we look when we watch TV? *Perception*, *37*(12), 1890–1894. https://doi.org/10.1068/p6253

Brugman, H., & Russel, A. (2004). Annotating multi-media/multi-modal resources with ELAN. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, & R. Silva (Eds.), *Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation* (pp. 2065–2068). ELRA - European Language Resources Association.

Brysbaert, M., & Vitu, F. (1998). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 125–147). Elsevier. https://doi.org/10.1016/B978-008043361-5/50007-9

Caffrey, C. (2008). Using pupillometric, fixation-based and subjective measures to measure the processing effort experienced when viewing subtitled TV anime with pop-up gloss. In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), *Looking at eyes: Eye-tracking studies of reading and translation processing* (Copenhagen studies in language, Vol. 36, pp. 125–144). Samfundslitteratur Press.

Charmant, J., & contributors. (2014). Kinovea (Version 0.8.24) [Computer software]. http://www.kinovea.org

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(35), 12629–12633. https://doi.org/10.1073/pnas.0506162102

Clifton Jr, C., Ferreira, F., Henderson, J. M., Inhoff, A. W., Liversedge, S. P., Reichle, E. D., & Schotter, E. R. (2016). Eye movements in reading and information processing: Keith Rayner's 40-year legacy. *Journal of Memory and Language*, *86*, 1–19. https://doi.org/10.1016/j.jml.2015.07.004

d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist*, *12*(3), 196–205. https://doi.org/10.1027/1016-9040.12.3.196

Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, *10*(28), 1–17. https://doi.org/10.1167/10.10.28

ELAN. (2015). (Version 4.9.2) [Computer software]. Max Planck Institute for Psycholinguistics, The Language Archive. https://archive.mpi.nl/tla/elan

Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision*, *8*(3), 1–15. https://doi.org/10.1167/8.3.3

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777–813. https://doi.org/10.1037/0033-295X.112.4.777

Gerber-Morón, O., Szarkowska, A., & Woll, B. (2018). The impact of text segmentation on subtitle reading. *Journal of Eye Movement Research*, *11*(4). https://doi.org/10.16910/11.4.2

Gillies, S., Adair, A., Blakey, A., Freeland, A., Kadouri, A., Bierbaum, A., Broere, B., Couwenberg, B., Root, B., Gervais, B., Hards, B., Wood, B., Hawkins, C., Prior, C., Quest, C., Pradal, C., Esposti, D., Collins, D., Baumgold, D., ... Ware, Z. (n.d.). Shapely: Manipulation and analysis of geometric objects. GitHub. https://github.com/Toblerity/Shapely

Goldstein, R. B., Woods, R. L., & Peli, E. (2007). Where people look when watching movies: Do all viewers look at the same place? *Computers in Biology and Medicine*, *37*(7), 957–964. https://doi.org/10.1016/j.compbiomed.2006.08.018

Gottlieb, H. (1998). Subtitling. In M. Baker (Ed.), *Routledge encyclopedia of translation studies* (pp. 244–348). Routledge.

Hinkin, M. P., Harris, R. J., & Miranda, A. T. (2014). Verbal redundancy aids memory for filmed entertainment dialogue. *The Journal of Psychology*, *148*(2), 161–176. https://doi.org/10.1080/00223980.2013.767774

Ivarsson, J., & Carroll, M. (1998). *Subtitling*. TransEdit.

Jensema, C. J., Danturthi, R. S., & Burch, R. (2000). Time spent viewing captions on television programs. *American Annals of the Deaf, 145*(5), 464–468. https://doi.org/10.1353/aad.2012.0144

Jensema, C. J., El Sharkawy, S., Danturthi, R. S., Burch, R., & Hsu, D. (2000). Eye movement patterns of captioned television viewers. *American Annals of the Deaf*, *145*(3), 275–285. https://doi.org/10.1353/aad.2012.0093

Kliegl, R., & Engbert, R. (2005). Fixation durations before word skipping in reading. *Psychonomic Bulletin & Review*, *12*(1), 132–138. https://doi.org/10.3758/BF03196358

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, *16*(1–2), 262–284. https://doi.org/10.1080/09541440340000213

Kruger, J.-L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly*, *49*(1), 105–120. https://doi.org/10.1002/rrq.59

Kruger, J.-L., Szarkowska, A., & Krejtz, I. (2015). Subtitles on the moving image: An overview of eye tracking studies. *Refractory : a journal of entertainment media*, *25*, 1-14.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, Articles*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lång, J. (2016). Subtitles vs. narration: The acquisition of information from visual-verbal and audio-verbal channels when watching a television documentary. In S. Hansen-Schirra & S. Grucza (Eds.), *Eyetracking and applied linguistics* (pp. 59–82). Language Science Press. https://doi.org/10.17169/langsci.b108.235

Lavaur, J.-M., & Bairstow, D. (2011). Languages on the screen: Is film comprehension related to the viewers' fluency level and to the language in the subtitles? *International Journal of Psychology*, *46*(6), 455–462. https://doi.org/10.1080/00207594.2011.565343

Lee, M., Roskos, B., & Ewoldsen, D. R. (2013). The impact of subtitles on comprehension of narrative film. *Media Psychology*, *16*(4), 412–440. https://doi.org/10.1080/15213269.2013.826119

Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition*, *147*, 1–20. https://doi.org/10.1016/j.cognition.2015.10.013

Lüdecke, D. (2018). sjPlot: Data visualization for statistics in social science (Version 2.6.2). Zenodo. https://doi.org/10.5281/zenodo.1308157

Mehtätalo, L., & Lappi, J. 2020. *Biometry for forestry and environmental data: With examples in R.* CRC Press.

McKinney, W. (2010). Data structures for statistical computing in python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference: Vol. 445* (pp. 51–56). https://doi.org/10.25080/Majora-92bf1922-00a

National Library of Finland. (2011). *The Finnish sub-corpus of the newspaper and periodical corpus of the National Library of Finland, Kielipankki Version* [text corpus]. http://urn.fi/urn:nbn:fi:lb-2016050302

Perego, E., Del Missier, F., & Bottiroli, S. (2015). Dubbing versus subtitling in young and older adults: Cognitive and evaluative aspects. *Perspectives: Studies in Translation Theory and Practice, 23*(1), 1–21. https://doi.org/10.1080/0907676X.2014.912343

Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology*, *13*(3), 243–272. https://doi.org/10.1080/15213269.2010.502873

Perego, E., Del Missier, F., & Stragà, M. (2018). Dubbing vs. subtitling: Complexity matters. *Target*, *30*(1), 137–157. https://doi.org/10.1075/target.16083.per

Perego, E., Laskowska, M., Matamala, A., Remael, A., Robert, I. S., Szarkowska, A., Vilaró, A., & Bottiroli, S. (2016). Is subtitling equally effective everywhere? A first cross-national study on the reception of interlingually subtitled messages. *Across Languages and Cultures*, *17*(2), 205–229. https://doi.org/10.1556/084.2016.17.2.4

Pinheiro, J., & Bates, D. (2000). *Mixed-effects models in S and S-PLUS*. Springer.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2018). nlme: Linear and nonlinear mixed effects models. The Comprehensive R Archive Network. https://cran.r-project.org/package=nlme

R Core Team. (2018). R: A language and environment for statistical computing (Version 3.4.4) [Computer software]. The Comprehensive R Archive Network. https://www.R-project.org/

Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological Research*, *72*(6), 675–688. https://doi.org/10.1007/s00426-008-0173-3

Rajendran, D. J., Duchowski, A. T., Orero, P., Martínez, J., & Romero-Fresco, P. (2013). Effects of text chunking on subtitling: A quantitative and qualitative examination. *Perspectives: Studies in Translation Theory and Practice*, *21*(1), 5–21. https://doi.org/10.1080/0907676X.2012.722651

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K. (2009). Eye movements in reading: Models and data. *Journal of Eye Movement Research*, *2*(5). https://doi.org/10.16910/jemr.2.5.2

Rayner, K., Li, X., Williams, C. C., Cave, K. R., & Well, A. D. (2007). Eye movements during information processing tasks: Individual differences and cultural effects. *Vision Research*, *47*(21), 2714–2726. https://doi.org/10.1016/j.visres.2007.05.007

Rayner, K., Miller, B., & Rotello, C. M. (2008). Eye movements when looking at print advertisements: The goal of the viewer matters. *Applied Cognitive Psychology*, *22*(5), 697–707. https://doi.org/10.1002/acp.1389

Rayner, K., & Pollatsek, A. (2006). Eye-movement control in reading. In M. Traxler & M. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 613–657). Elsevier.

Rayner, K., & Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review*, *3*(2), 245–248. https://doi.org/10.3758/BF03212426

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*(3), 448–465. https://doi.org/10.1037/0882-7974.21.3.448

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*(4), 504–509. https://doi.org/10.3758/BF03214555

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*(4), 445–476. https://doi.org/10.1017/S0140525X03000104

SensoMotoric Instruments. (2015). *Begaze manual* (version 3.5). SensoMotoric Instruments.

Snell, J., van Leipsig, S., Grainger, J., & Meeter, M. (2018). OB1-reader: A model of word recognition and eye movements in text reading. *Psychological Review*, *125*(6), 969–984. https://doi.org/10.1037/rev0000119

Szarkowska, A., & Gerber-Morón, O. (2018a). Two or three lines: A mixed-methods study on subtitle processing and preferences. *Perspectives: Studies in Translation Theory and Practice*, 1–21. https://doi.org/10.1080/0907676X.2018.1520267

Szarkowska, A., & Gerber-Morón, O. (2018b). Viewers can keep up with fast subtitles: Evidence from eye movements. *PloS One*, *13*(6), 1–30. https://doi.org/10.1371/journal.pone.0199331

Szarkowska, A., Krejtz, I., Klyszejko, Z., & Wieczorek, A. (2011). Verbatim, standard, or edited? Reading patterns of different captioning styles among deaf, hard of hearing, and hearing viewers. *American Annals of the Deaf*, *156*(4), 363–378. https://doi.org/10.1353/aad.2011.0039

Szarkowska, A., Krejtz, I., Pilipczuk, O., Dutka, Ł., & Kruger, J.-L. (2016). The effects of text editing and subtitle presentation rate on the comprehension and reading patterns of interlingual and intralingual subtitles among deaf, hard of hearing and hearing viewers. *Across Languages and Cultures*, *17*(2), 183–204. https://doi.org/10.1556/084.2016.17.2.3

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science Engineering*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37