

Two or three lines: a mixed-methods study on subtitle processing and preferences

Agnieszka Szarkowska^{a*} and Olivia Gerber-Morón^b

*^aCentre for Translation Studies, University College London, London, United Kingdom;
and Institute of Applied Linguistics, University of Warsaw, Warsaw, Poland;*

*^bDepartment of Translation and Interpreting & East Asian Studies, Faculty of
Translation and Interpreting, Universitat Autònoma de Barcelona, Barcelona, Spain*

*a.szarkowska@uw.edu.pl, Institute of Applied Linguistics, University of Warsaw,
Dobra 55, 00-312 Warsaw, Poland

Two or three lines: a mixed-methods study on subtitle processing and preferences

The typically recommended maximum number of lines in a subtitle is two. Yet, three-line subtitles are often used in intralingual English-to-English subtitling on television programmes with high information density and fast speech rates. To the best of our knowledge, no prior empirical work has contrasted the processing of three-line with two-line subtitles. In this study, we showed participants one video with two-line subtitles and one with three-line subtitles. We measured the impact of the number of lines on subtitle processing using eye tracking as well as comprehension, cognitive load, enjoyment and preferences. We conducted two experiments with different types of viewers: hearing native speakers of English, Polish and Spanish as well as British hard of hearing and deaf viewers. Three-line subtitles induced higher cognitive load than two-line subtitles. The number of lines did not affect comprehension. Viewers generally preferred two-line over three-line subtitles. The results provide empirical evidence on the processing of two- and three-line subtitles and can be used to inform current subtitling practices.

Keywords: subtitling; audiovisual translation; cognitive load; number of lines; eye tracking; hearing loss; preferences

Introduction

We all know what subtitles look like: relatively short lines of text usually displayed at the bottom of the screen. But how many lines exactly can a subtitle have? The answer to this question largely depends on the type of the subtitles: ‘classic’ interlingual cinema subtitles usually take up no more than two lines, but three or even four lines are not uncommon in intralingual English-to-English subtitles for the deaf and hard of hearing on British or American television (Díaz Cintas & Remael, 2007; Robson, 2004).

In this paper, we look at subtitling in general, without differentiating between interlingual subtitling, intralingual subtitling for the deaf and hard of hearing (SDH) or interlingual SDH (Neves, 2009; Szarkowska, 2013). We take this broad perspective

because the differences between various types of subtitling and their respective target audiences are now becoming less clear-cut and their boundaries overlapping. For example, in the absence of SDH in cinemas in the UK, British deaf and hard of hearing people often use standard interlingual subtitles with no sound information; Polish deaf and hard of hearing viewers benefit from interlingual subtitles to foreign films released on DVD/VOD as there is no separate interlingual SDH version targeted for them specifically; and finally, hearing people across the world use intralingual SDH for various reasons, such as improving their proficiency in another language, watching television in noisy environments, etc. (Vanderplank, 2016).

Despite the omnipresence of subtitles (Dwyer, 2017), little is known about the impact of a higher number of lines on subtitle processing, enjoyment, comprehension or cognitive load in different types of viewers. From previous eye-tracking studies we know that viewers spend relatively more time looking at two-line subtitles compared to one-line subtitles (d'Ydewalle, Praet, Verfaillie, & Van Rensbergen, 1991; Praet, Verfaillie, De Graef, Van Rensbergen, & D'Ydewalle, 1990), possibly because two-line subtitles stay on screen longer, thus attracting viewers' attention, or because of their more informative content.

But how exactly does the number of lines affect subtitle reading? Are three-line subtitles easier or more difficult to process than two-liners? Should the number of lines depend on the genre of the audiovisual material? And finally, what do the viewers prefer? To address these questions, we conducted a mixed methods study, combining eye tracking with self-reported cognitive load, enjoyment, performance measures and semi-structured interviews. In this study, we use intralingual English subtitles as it is in this type of subtitling that three-liners are most common.

The number of lines in subtitling

Scholars and subtitle guidelines generally recommend a maximum of two lines for a subtitle, although three lines are acceptable on condition that they do not obscure the image (Baker, Lambourne, & Rowston, 1984; BBC, 2017; Díaz Cintas & Remael, 2007; Ivarsson & Carroll, 1998; Ofcom, 2017). The US-based Described and Captioned Media Program (2017) allows three and even four lines of text if “a one- or two-line caption¹ would interfere with pre-existing graphics” (p. 9).

Three-line subtitles are used on English-speaking television for different reasons, for example in sentences which are too long to fit into a single two-line subtitle. Three-line subtitles can be found in some TV programmes and documentaries with high information density and fast speech rates. The reason for using three rather than two lines in a subtitle may be attributed to avoiding two-line subtitles being displayed very quickly one after the other (Bogucki, 2009). The use of three lines also depends on the type of audiovisual content: in fast quiz shows, for instance, it may be important to show the entire question in one subtitle (BBC, 2017). Given the well-known preference for verbatim subtitles by deaf and hard of hearing viewers (Neves, 2008; Romero-Fresco, 2009; Szarkowska, Krejtz, Kłyszejko, & Wieczorek, 2011) – who are the main target group of intralingual English subtitles – three-liners may also reflect an attempt to include as much text from the dialogues as possible (Kalantzi, 2008).

Studying subtitling with eye tracking

Drawing largely on the eye-mind hypothesis, according to which “there is no appreciable lag between what is fixated and what is processed” (Just & Carpenter, 1980, p. 331), eye tracking has been used as an online measure of visual attention allocation in

subtitle processing, allowing researchers “unprecedented access into the eyes, and arguably the minds, of the viewers” (Doherty & Kruger, 2018). In recent years, eye tracking has been used to study different aspects of subtitling: line breaks and segmentation (Gerber-Morón & Szarkowska, 2018; Gerber-Morón, Szarkowska, & Woll, 2018; Perego, Del Missier, Porta, & Mosconi, 2010; Rajendran, Duchowski, Orero, Martínez, & Romero-Fresco, 2013), text editing and reading speed (d'Ydewalle, Rensbergen, & Pollet, 1987; Jensema, Danturthi, & Burch, 2000; Koolstra, Van Der Voort, & d'Ydewalle, 1999; Szarkowska & Gerber-Morón, 2018b; Szarkowska, Krejtz, Pilipczuk, Dutka, & Kruger, 2016), shot changes (Krejtz, Szarkowska, & Krejtz, 2013), the impact of sound and type of subtitling on subtitle processing (Bisson, Van Heuven, Conklin, & Tunney, 2012), cognitive processing (Kruger & Doherty, 2016; Kruger, Hefer, & Matthew, 2014), and types of viewers (d'Ydewalle & De Bruycker, 2007; De Bruycker & d'Ydewalle, 2003; Matamala, Perego, & Bottiroli, 2017; Perego, Del Missier, & Stragà, 2018; Perego et al., 2016). The increasing body of research has shown that the processing of subtitles not only depends on the type of subtitles, but is also conditioned by the characteristics of the audience, including their linguistic abilities or hearing status (d'Ydewalle & De Bruycker, 2007; De Bruycker & d'Ydewalle, 2003; De Linde & Kay, 1999; Krejtz, Szarkowska, & Łogińska, 2016; Łuczak, 2017; Muñoz, 2017; Szarkowska et al., 2016; Winke, Gass, & Sydorenko, 2013).

Although to the best of our knowledge no previous studies compared the reading of three-line and two-line subtitles, we can draw some conclusions from the studies on one- and two-line subtitles. Previous research (d'Ydewalle et al., 1991; De Bruycker & d'Ydewalle, 2003; Praet et al., 1990) showed that people spend relatively more time reading two-line than one-line subtitles in standard interlingual subtitling (as opposed to reversed subtitling²). A possible reason for this suggested by Praet et al. (1990) was that

two-line subtitles are more syntactically and semantically complex than one-liners. Another explanation was that unlike two-liners, one-line subtitles do not add much relevant information to what viewers can already infer from the screen. The finding that viewers spend proportionally more time reading subtitles with two lines than those with one line prompted some authors to argue that a more regular reading pattern can be expected with two-line subtitles (d'Ydewalle & De Bruycker, 2007). Along these lines, we may wonder whether a more regular reading pattern occurs in the case of three-line subtitles compared to two-liners. Are subtitles containing more text easier to read and less cognitively demanding? Or are they more difficult given the complex nature of audiovisual processing?

Cognitive load in subtitling

When watching subtitled content, viewers need to process several sources of information simultaneously: moving images, sounds, and on-screen text (Kruger, Szarkowska, & Krejtz, 2015). This can be a taxing task in itself and it should not be made more difficult by the poor quality of subtitles or any subtitle features potentially hampering the viewing process. One such feature may possibly be an excessive number of lines.

Cognitive load is “a multidimensional construct that represents the load that performing a particular task imposes on the cognitive system” of a person (Paas & Merriënboer, 1994, p. 353). According to cognitive load theory (Chandler & Sweller, 1991; Paas, Renkl, & Sweller, 2003, 2004, 2016; Plass, Moreno, & Brünken, 2010; Sweller, 2011), limitations in working memory may lower a person’s processing capacity, causing them to experience high cognitive load, overload or frustration (Sweller, Ayres, & Kalyuga, 2011). Originating from the fields of instructional design and multimedia learning (Moreno & Mayer, 2002; Paas et al., 2016; Plass et al., 2010),

cognitive load has also been studied in subtitle processing (Kruger & Doherty, 2016; Kruger, Doherty, Fox, & de Lissa, 2017; Kruger, Hefer, & Matthew, 2013; Kruger et al., 2014). Early studies on subtitling and cognitive load investigated whether the very presence of subtitles increases viewers' cognitive load (Kruger et al., 2013). It has been found that although subtitles do change the distribution of viewers' visual attention, they do not result in increased cognitive load (Kruger et al., 2013; Perego et al., 2010).

The amount of cognitive load depends on the one hand on the nature of the task and on the other hand on the characteristics of the individual person (Paas, Renkl, et al., 2003). Among the features of the task which can impact on cognitive load are task difficulty, topic, pace, etc. Individual characteristics of people include age, reading skills, hearing status, expertise, or prior experience.

Cognitive load can be measured using mental effort and performance (Brunken, Plass, & Leutner, 2003; Kruger et al., 2013; Paas & Van Merriënboer, 1993). Mental effort is the amount of cognitive processing invested in a task (Paas, Tuovinen, Tabbers, & Van Gerven, 2003; Paas & Van Merriënboer, 1993). It is often measured using self-reports or physiological techniques, such as electroencephalography (EEG) or eye tracking. Performance – understood as “the effectiveness in accomplishing a particular task” (Paas & Van Merriënboer, 1993, p. 738) – is usually measured as a test score, for instance a comprehension score. If the difficulty of the task increases, more effort may need to be allocated to carry out the task and its execution may result in poorer performance.

Current study

The main goal of this study is to compare the cognitive processing of two- and three-line subtitles and to examine viewers' preferences with regard to the number of lines in subtitles. With this goal in mind, we presented participants with subtitled videos with

two- and three-line subtitles, and measured their comprehension, self-reported cognitive load, enjoyment, eye movements, and preferences (see the Method section).

We wanted to know whether three-line subtitles are more cognitively demanding than two-liners. If so, we would expect a higher cognitive load as well as lower comprehension and enjoyment reported by the participants. In terms of eye tracking, higher cognitive load would result in longer mean fixation duration and time spent in the subtitle area (Holmqvist et al., 2011). We predicted that people would spend more time reading three-line subtitles as they are longer and contain more words (see Table 1 and 2). If three-line subtitles are more cognitively demanding, we would expect people to spend *proportionally* more time on reading them compared to two-line subtitles (Koolstra et al., 1999; Praet et al., 1990). Finally, we analysed the number of revisits to the subtitle area, also referred to as ‘returns’ and ‘rechecks’ (Holmqvist et al., 2011). On the one hand, revisits have been found to show the semantic informativeness of an area (Holmqvist et al., 2011), but on the other hand they may reflect processing difficulty associated with returning to an area to recheck it (Conklin & Pellicer-Sánchez, 2016). In the context of subtitled videos, revisits show how people divide their visual attention between viewing the scene and reading the subtitles; a higher number of revisits may indicate less fluent and thus less efficient reading.

Another research question we asked was whether subtitle processing depends on the participants’ linguistic background. We conducted two experiments with different groups of participants: Experiment 1 with hearing native speakers of English, Spanish and Polish, and Experiment 2 with hearing, hard of hearing and deaf British viewers. We hypothesised that hearing viewers accustomed to subtitling, or deaf and hard of hearing who rely on subtitles in their everyday life, would experience lower cognitive load than those whose previous exposure to subtitling was limited.

Thanks to using a combination of different methods, such as eye tracking, questionnaires, and semi-structured interviews, we have been able to analyse the impact of three and two lines on the processing of subtitled videos, modulated by the viewers' linguistic background. To the best of our knowledge, no prior work has contrasted the processing of three-line with two-line subtitles. The empirical examination of this issue is important to confirm the validity of current subtitling practices, particularly in the UK.

Method

Two experiments were conducted. The experimental design was mixed factorial. We used the same methodology and materials in both experiments. The main independent variable was experimental condition (two vs. three lines). The main within-subject factor was language (English, Polish, Spanish) in Experiment 1 and hearing loss (hearing, hard of hearing, deaf) in Experiment 2.

All data related to the study, including the subtitle files, experimental protocol with all the questions and raw data, are openly available in the RepOD data repository hosted by the University of Warsaw (Szarkowska & Gerber-Morón, 2018a).

Participants in Experiment 1

The total of 74 people took part, 46 females and 28 males. Their mean age was $M=26.96$ ($SD=6.66$). They were native speakers of English ($N=27$), Polish ($N=21$) and Spanish ($N=26$). Due to data quality issues, for eye-tracking analyses we used data from 24 English, 19 Polish and 24 Spanish participants.

We recruited participants based on the traditionally conceived audiovisual translation landscape, assuming for instance that Spanish participants are the least familiar with subtitling. However, when asked about their preferred type of audiovisual

translation, most participants chose subtitling (24 English, 11 Polish and 22 Spanish), whereas others stated they prefer watching films in the original version (1 English, 10 Polish and 3 Spanish); only 1 Spanish participant declared to prefer dubbing. We believe this may reflect recent changes in viewers' preferences and a wider availability of different AVT types (Perego et al., 2016; Szarkowska & Laskowska, 2015) as well as the fact that they were living in the UK at the time the study was conducted. In spite of these changes and the preferences expressed by the participants, we still believe that it is true that Spanish participants had been exposed to dubbing most compared to the other groups, which – despite their current preferences – may have possibly affected the results.

Because the subtitles used in this study were in English, we also asked Polish and Spanish participants to self-report their proficiency in reading English³. All participants declared their proficiency to be at least B1. 3 Polish participants reported to be at the C1 level and 18 at C2 level, whereas 1 Spanish person declared to be at B1 level, 4 at B2 level, 5 at C1 and 16 people at C2 level. Overall, our sample was quite proficient in English, which is not surprising given that they were living in the UK at the time the study was conducted.

Participants in Experiment 2

The total of 46 participants from the UK took part in Experiment 2 (27 were hearing, 10 hard of hearing, and 9 deaf). Their mean age was 34.57 ($SD=13.29$). Due to data quality issues, for eye-tracking analyses we used data from 24 hearing, 9 hard of hearing, and 5 deaf people. Hearing participants in Experiment 2 were the same as English participants in Experiment 1.

Prior to the study, participants were asked to complete a demographic questionnaire, where they stated the degree and onset of their hearing loss as well as if

they considered themselves to be deaf or hard of hearing. 7 participants were born deaf or hard of hearing, 4 lost hearing under the age of 8, 2 lost hearing aged 9-17 and 6 lost hearing between the ages of 18 and 40. Ten participants were profoundly deaf, 6 were severely deaf and 4 had moderate hearing loss. All deaf and hard of hearing people were regular subtitle users.

Stimuli

Two short videos were used from the Netflix show *Chelsea* by Chelsea Handler (2016, see Table 1). Both clips featured the “talking head” of the presenter.

Table 1. Characteristics of clips used in the study

	Clip 1	Clip 2
Duration	1 min 10 sec	1 min 24 sec
Number of words	188	199
Number of subtitles in two-line condition	22	23
Number of subtitles in three-line condition	16	18
Total duration of subtitles in two-line condition	56800 ms	74560 ms
Total duration of subtitles in three-line condition	59240 ms	74800 ms
Percentage of total subtitle display time as a function of total clip duration in two-line subtitles	81%	88%
Percentage of total subtitle display time as a function of total clip duration in three-line subtitles	84%	89%

The clips were subtitled intralingually, from English to English, in two versions: two and three lines. The text in both versions was identical and it was a verbatim version of the dialogue (see Table 2). No sound descriptions were included. The maximum number of characters per line was 40 and the subtitle speed was 17 characters per second for all clips and conditions.

Table 2. Comparison of text in two- and three-line subtitles

No.	Two-line subtitles	Three-line subtitles
1	I've invited my doctor of ten years here	I've invited my doctor of ten years here and instructed him to bring all of my medical records.

2	and instructed him to bring all of my medical records.	Hopefully, by exposing my very own medical history, I will inspire our future leaders to do the same.
3	Hopefully, by exposing my very own medical history,	It's the old: "I'll show you mine if you show me yours."
4	I will inspire our future leaders to do the same.	
5	It's the old "I'll show you mine if you show me yours."	

Procedure

The study took place at University College London and received full ethical approval of the UCL Research Ethics Committee. Before taking part in the experiment, participants received information sheets and signed the informed consent forms.

Participants were tested individually in an eye-tracking lab. They were shown two clips with the sound on: one with two-line subtitles and the other with three-line subtitles. The order of clips and conditions was counterbalanced. After each clip, participants answered five true-false comprehension questions and self-reported their cognitive load and enjoyment. After watching both clips, they were asked which subtitles they prefer (two lines, three lines or no preference). Finally, they were invited to discuss their views on the number of lines in subtitles in a short semi-structured interview, where we asked them to elaborate on their choices and elicited their views on subtitling quality.

Apparatus

Participants' eye movements were recorded with SMI RED mobile eye tracker with the sampling rate of 250 Hz. The experiment was created and conducted using SMI Experiment Centre and data was analysed with SMI BeGaze and IBM SPSS Statistics 24. Participants whose tracking ratio was lower than 80% were excluded from eye-

tracking analyses.

Design and dependent variables

The study design was 3 (group in Experiment 1: English, Polish and Spanish; in Experiment 2: hearing, hard of hearing and deaf) x 2 (condition: two vs. three lines). For each dependent variable, we conducted a 3 x 2 mixed ANOVA with language/hearing loss as an independent factor and the number of lines as an independent within-subjects variable. The dependent variables were: comprehension score, three indicators of cognitive load, enjoyment, preferences and four eye-tracking measures.

Comprehension. The comprehension score was calculated as a percentage of correct answers to 5 true/false questions after each clip. The questions were related to the content of the subtitles (not the visuals), for instance: *Chelsea suggests that Trump's diet consists of xenophobia and KFC* (true/false).

Cognitive load. We used three indicators of cognitive load: difficulty (*Was it difficult for you to read the subtitles in this clip?*), effort (*Did you have to put a lot of effort into reading subtitles in this clip?*) and frustration (*Did you feel annoyed when reading the subtitles in this clip?*). Participants were asked to self-report their load after each clip using a 1-7 scale (1 = *very low*, 7 = *very high*).

Enjoyment. Similarly to cognitive load, enjoyment was a self-report measured on 1-7 scale: *I enjoyed watching the film with these subtitles. Please assess it on the scale from 1 to 7, where 1 means "not at all" and 7 means "very much"*.

Preferences. After watching the two clips, viewers were asked a question on their preferences regarding the number of lines: *Which subtitles do you prefer: two-line subtitles, three-line subtitles, I don't mind.*

Eye-tracking measures. We used four eye-tracking measures (see Table 3). Areas of interest (AOIs) were drawn on each subtitle in each clip. The results reported here come from these AOIs and are averaged per participant and per clip.

Table 3. Description of eye-tracking variables

Eye-tracking measure	Description
Absolute reading time	Dwell time, i.e. the sum of durations of all fixations and saccades, starting with the first fixation on the subtitle AOI. Reported in milliseconds (ms).
Proportional reading time	The percentage of time that a participant spent in the AOI as a function of subtitle display time ⁴ , where 100% is the total subtitle display time. Reported in percentages.
Mean fixation duration	The duration of a fixation in a subtitle AOI, averaged per clip and per participant. Reported in ms.
Revisits	The number of glances a participant made to the subtitle AOI after visiting the subtitle for the first time. Reported as count.

Experiment 1

Results

Comprehension

As a first step, to assess the participants' performance after watching videos with two- and three-line subtitles, we conducted a 3 x 2 mixed ANOVA with language (English, Polish and Spanish) as an independent factor and the number of lines as an independent within-subjects variable. The number of lines had no effect on comprehension,

$F(1,71)=1.178$, $p=.281$, $\eta_p^2=.016$ (Table 4). Polish and Spanish participants had lower comprehension scores for the three lines compared to two-line subtitles, whereas

English people achieved slightly higher scores in the three-line condition, but the interaction was not significant and the differences were small.

Table 4. Comprehension scores by number of lines and language in Experiment 1

Language	2 lines		3 lines	
	<i>M (SD)</i>		<i>M (SD)</i>	
English	89.63 (16.04)		91.11 (10.12)	
Polish	95.24 (10.77)		90.48 (13.59)	
Spanish	84.62 (22.13)		80.00 (21.16)	
Total	89.46 (17.58)		87.03 (16.36)	

Comprehension differed depending on the participants' linguistic background, $F(2,71) = 4.473, p=.015, \eta_p^2=.112$. Spanish participants achieved a significantly lower score than the Polish, $p=.02, 95\% \text{ CI} [-19.82, -1.28]$ and a marginally lower than the English, $p=.077, 95\% \text{ CI} [-16.74, .62]$. The highest comprehension score was found in Polish participants.

Cognitive load

Next, to test the effect of lines on the cognitive load experienced by participants, we conducted a 3 x 2 mixed ANOVA with language (English, Polish and Spanish) as an independent factor and the number of lines (2 x 3) as an independent within-subjects variable. The dependent variables were three indicators of cognitive load (difficulty, effort, frustration). We found a significant main effect of the number of lines in all three indicators of cognitive load (see Table 5). As predicted, difficulty, effort and frustration were reported by all the participants to be higher in the three-line condition, although the effect size was not large.

Table 5. Mean cognitive load indicators in Experiment 1

	Number of lines		<i>df</i>	<i>F</i>	<i>p</i>	η_p^2
	2	3				
Difficulty			1	24,750	.000*	.258
English	2.37 (1.39)	3.63 (1.64)				

Polish	1.81 (1.12)	3.05 (1.68)				
Spanish	1.96 (1.18)	2.96 (1.9)				
Effort			1	16,371	.000*	.187
English	2.78 (1.57)	3.41 (1.84)				
Polish	1.67 (1.19)	2.9 (1.67)				
Spanish	2.04 (1.18)	2.85 (1.91)				
Frustration			1	33,959	.000*	.324
English	2.37 (1.27)	3.00 (1.79)				
Polish	1.43 (.81)	3.14 (1.85)				
Spanish	1.54 (.76)	2.81 (1.87)				

There were no between-subject differences, which means that the load experienced by participants was similar regardless of their mother tongue.

Enjoyment

In line with our expectations, an analogous ANOVA confirmed that the number of lines had a significant effect on enjoyment, $F(1,71)=5.039$, $p=.028$, $\eta_p^2=.066$. Enjoyment was higher in the two-line condition compared to the three lines, as shown in Table 6.

Table 6. Descriptive statistics for enjoyment by the number of lines and language

Language	2 lines	3 lines
	<i>M (SD)</i>	<i>M (SD)</i>
English	4.48 (1.47)	4.15 (1.46)
Polish	5.38 (1.53)	4.24 (1.64)
Spanish	4.96 (1.70)	4.77 (1.68)
Total	4.91 (1.59)	4.39 (1.59)

There were no differences between the three groups, $F(2,71)=1.719$, $p=.187$, $\eta_p^2=.046$.

This means that all groups of participants declared to have enjoyed the three lines less than two lines.

Preferences

To explore people's views on the maximum number of lines in subtitles, we calculated a chi-square test comparing the preferences for two and three lines, which turned out to be

significant, $\chi(2)=28.405$, $p<.000$. Most participants preferred two-line subtitles over three-liners (see Table 7).

Table 7. Descriptive statistics for preferences on number of lines in Experiment 1

Language	2 lines	3 lines	<i>I don't mind</i>
English	51.9%	11.1%	37.0%
Polish	81%	0%	19%
Spanish	57.7%	30.8%	11.5%
Total	62.2%	4.9%	23%

A significant interaction was found between participants, $\chi(4) = 13.485$, $p = .009$. This means that while all groups of participants preferred two-lines over three in subtitles, over one third of English participants showed no preference, and one third of Spanish participants – perhaps somewhat surprisingly – declared they preferred three lines (see the Interviews section).

Eye-tracking measures

To examine people's reading patterns while watching two- and three-line subtitles, we measured the time they spent on reading subtitles in the two conditions. We found a main effect of lines on both absolute reading time and proportional reading time (see Table 8). In line with our expectations, people spent more time – in absolute and proportional terms – when reading subtitles with three lines. The number of lines, however, did not affect revisits to the subtitles area.

Having measured mean fixation duration, we found a significant interaction between the number of lines and participant group, $F(2,64) = 4.912$, $p = .010$, $\eta_p^2 = .133$. We decomposed this interaction with simple effects using Bonferroni correction, and analysed each group separately. There was an effect of the number of lines among English participants, $F(1,23) = 9.017$, $p = .006$, $\eta_p^2 = .282$. Their mean fixation duration was shorter in the two-line condition ($M = 202\text{ms}$) compared to three-line condition

($M=225\text{ms}$), 95% CI [7.17, 38.95]. However, there was no effect of the number of lines on mean fixation duration among Polish participants, $F(1,18)=.695$, $p=.415$, $\eta_p^2=.037$, or Spanish participants, $F(1,23)=.147$, $p=.705$, $\eta_p^2=.006$, for whom mean fixation duration did not differ across the two conditions.

Table 8. Differences between two- and three-line subtitles in eye-tracking measures in Experiment 1

	Number of lines		<i>df</i>	<i>F</i>	<i>p</i>	η_p^2
	2	3				
Absolute reading time			1,64	84.374	.000*	.569
English	1528	2199				
Polish	1438	1937				
Spanish	1739	2393				
Proportional reading time			1,64	9.408	.003*	.128
English	46	50				
Polish	42	45				
Spanish	52	55				
Mean fixation duration			1,64	3.394	.070	.050
English	202	225				
Polish	201	197				
Spanish	218	220				
Revisits			1,64	2.497	.119	.038
English	.74	.86				
Polish	.83	.89				
Spanish	.66	.84				

Note: Absolute reading time and mean fixation duration as reported in milliseconds. Proportional reading time is reported as a percentage. Revisits are reported as count.

When it comes to differences between the groups, we found a tendency in absolute reading time, $F(2,64)=2,443$, $p=.095$, $\eta_p^2=.071$, and proportional reading time, $F(2,64)=2,616$, $p=.081$, $\eta_p^2=.076$. Polish participants tended to spend less time in the subtitle area compared to Spanish, but the difference did not reach statistical significance.

Correlations

To explore the parallels between participants' cognitive load, comprehension and enjoyment while watching the clips with two- and three-line subtitles, we correlated the

self-reported results with eye-tracking data. Using Spearman's rank correlation, we found that in the two-line condition, there was a significant positive correlation between comprehension and enjoyment, $r_s(74)=.315, p=.006$. At the same time, enjoyment correlated negatively with difficulty, $r_s(74)=-.434, p<.000$, effort, $r_s(74)=-.455, p<.000$, and frustration, $r_s(74)=-.566, p<.000$. This means that people's enjoyment increased together with higher comprehension and decreased with higher cognitive load. There were no significant correlations with any of the eye tracking variables.

In the three-line condition, comprehension was negatively correlated with the absolute time spent in the subtitle area, $r_s(67)=-.342, p=.005$, and with the proportional time, $r_s(67)=-.301, p=.013$ as well as fixation count, $r_s(67)=-.300, p=.014$, showing that the longer the people gazed at three-line subtitles, the lower their comprehension score was. This was also confirmed by comprehension score correlating negatively with mean fixation duration, $r_s(67)=-.262, p=.032$, which is an indication of higher processing effort. Higher cognitive effort was negatively related with enjoyment: participants reported enjoying the clips with three-line subtitles less when they found them more difficult, $r_s(67)=-.626, p<.000$, effortful, $r_s(74)=-.552, p<.000$, and frustrating, $r_s(74)=-.601, p<.000$. Overall, the more cognitive effort the participants experienced, the less enjoyment they reported and the lower their comprehension was.

Experiment 2

Results

Comprehension

Similarly to Experiment 1, we started with measuring viewers' comprehension scores (Table 9). Again, the number of lines had no significant effect on comprehension,

$F(1,43)=.778, p=.38, \eta_p^2=.018$. Hard of hearing and deaf participants had slightly higher scores in the two-line condition, but this interaction was not significant.

Table 9. Comprehension scores by number of lines and hearing status in Experiment 2

Hearing status	2 lines	3 lines
	<i>M (SD)</i>	<i>M (SD)</i>
Hearing	89.63 (16.04)	91.11 (10.12)
Hard of hearing	88.00 (10.32)	86.00 (13.49)
Deaf	86.67 (14.14)	80.00 (20.00)
Total	88.70 (14.39)	87.83 (13.64)

Although there was a tendency for hearing people to have higher comprehension than hard of hearing and deaf, the differences between subjects were not significant,

$F(1,2)=1.384, p=.26, \eta_p^2=.06$, possibly due to the small number of participants and small numerical differences between the groups and conditions.

Cognitive load

Next, we proceeded to check whether the number of lines affected people's cognitive load (see Table 10). All cognitive load indicators were higher in the three-line condition apart from difficulty in the deaf group, where there was no difference between the number of lines. There was a main effect of number of lines on effort and a tendency in difficulty. Mean effort was lower in two-line subtitles ($M=2.9, SD=.27$) than in three-line subtitles ($M=3.69, SD=.3$). No significant differences were found in frustration, though it was consistently higher in three lines than in two lines.

Table 10. Mean cognitive load indicators in Experiment 2

	Number of lines		<i>df</i>	<i>F</i>	<i>p</i>	η_p^2
	2	3				
Difficulty			1,43	3.305	.076	.071
Hearing	2.37 (1.39)	3.63 (1.64)				
Hard of hearing	2.80 (1.22)	3.50 (1.84)				
Deaf	3.22 (1.85)	3.22 (1.85)				
Effort			1,43	5.881	.020*	.120
Hearing	2.78 (1.57)	3.41 (1.84)				

Hard of hearing	2.70 (1.63)	3.80 (1.54)				
Deaf	3.22 (1.92)	3.89 (1.96)				
Frustration			1,43	2.642	.111	.058
Hearing	2.37 (1.27)	3.00 (1.79)				
Hard of hearing	1.90 (1.19)	2.50 (1.65)				
Deaf	3.11 (1.96)	3.33 (1.58)				

We found no significant differences in cognitive load between groups, although numerical results show a tendency for deaf participants to declare higher effort and frustration than other groups, but not in the case of difficulty.

Enjoyment

Unlike in Experiment 1, this time there was no significant effect of the number of lines on enjoyment, $F(1,43)=.086$, $p=.771$, $\eta_p^2=.002$ (see Table 11).

Table 11. Descriptive statistics for enjoyment by number of lines in Exp. 2

Hearing status	2 lines	3 lines
	<i>M (SD)</i>	<i>M (SD)</i>
Hearing	4.48 (1.47)	4.15 (1.46)
Hard of hearing	4.90 (1.66)	4.70 (2.00)
Deaf	5.22 (1.56)	5.44 (1.42)
Total	4.72 (1.53)	4.52 (1.62)

There was an almost significant difference between groups, $F(2,43)=2,865$, $p=.068$, $\eta_p^2=.118$. Deaf people differed from the hearing, $p=.074$, 95% CI [-2.11, .07], in that they enjoyed the three-line subtitles a bit more than the two lines. There were no differences between the hearing and hard of hearing, or the hard of hearing and deaf. In general, participants who were deaf or hard of hearing enjoyed both clips more than the hearing.

Preferences

Similarly to Experiment 1 and despite the above results on enjoyment, there was a marked preference for two-line subtitles, particularly among the deaf and the hard of hearing participants, who unanimously opted for two-liners (see Table 12). We

calculated a chi-square comparing the preferences for two and three lines, and the difference in preferences was significant, $\chi(2)=32.130, p<.000$.

Table 12. Descriptive statistics for preferences on number of lines in Experiment 2

Hearing status	2 lines	3 lines	<i>I don't mind</i>
Hearing	51.9%	11.1%	37%
Hard of hearing	100%	0%	0%
Deaf	100%	0%	0%
Total			

There was a significant interaction between the preference and participant group, $\chi(4), p=.013$. While all hard of hearing and deaf participants preferred two lines over three, about one in three hearing participants expressed no preference.

Eye-tracking measures

Similarly to Experiment 1, we found a main effect of the number of lines on absolute and proportional reading time (see Table 13). More time was spent on reading three- than two-liners. The effect size was not large, but it was particularly discernible in the case of deaf participants, who spent 48% of the subtitle display time looking at the subtitle area in the two lines condition and 57% in three lines. No interactions were significant.

In contrast to Experiment 1, there was no effect of the number of lines on mean fixation duration. Mean fixation duration was higher for hearing and hard of hearing people in the two-line condition, but lower for the deaf. This interaction did not reach statistical significance, probably owing to the small sample size.

Unlike in Experiment 1, we found a main effect on the number of revisits, which was significantly higher in the three-line condition compared to the two-lines. This was particularly discernible among deaf participants. Higher number of revisits observed may be related to the higher subtitle display times in the three-line condition and to the

way deaf people were reading subtitles (possibly comparing the text with lip movements).

Table 13. Differences between two- and three-line subtitles in eye-tracking measures in Experiment 2

	Number of lines		<i>df</i>	<i>F</i>	<i>p</i>	η_p^2
	2	3				
Absolute reading time			1,35	46.849	.000*	.572
Hearing	1528	2199				
Hard of hearing	1706	2209				
Deaf	1559	2523				
Proportional reading time			1,35	10.148	.003*	.225
Hearing	46	50				
Hard of hearing	50	52				
Deaf	48	57				
Mean fixation duration			1,35	.962	.334	.027
Hearing	202	225				
Hard of hearing	187	194				
Deaf	210	201				
Revisits			1,35	9.311	.004*	.210
Hearing	.74	.86				
Hard of hearing	.65	.81				
Deaf	.41	1.00				

Note: Absolute reading time and mean fixation duration as reported in milliseconds. Proportional reading time is reported as a percentage. Revisits are reported as count.

There were no statistically significant differences between groups in any of the eye-tracking measures, possibly because of the small sample.

Correlations

Similarly to Experiment 1, in the two-line condition, enjoyment was negatively related to difficulty, $r_s(46)=-.369$, $p=.012$, effort, $r_s(46)=-.397$, $p=.006$, and frustration, $r_s(46)=-.460$, $p=.001$, as well as to mean fixation duration, $r_s(38)=-.330$, $p=.043$. This shows that the higher the cognitive effort reported by participants, the lower their enjoyment. In the three-line condition, enjoyment was also negatively correlated to difficulty, $r_s(46)=-.460$, $p=.001$, effort, $r_s(46)=-.361$, $p=.014$, and frustration, $r_s(46)=-.361$, $p=.014$. There were no significant correlations with any of the eye-tracking variables.

Semi-structured interviews

After watching the two clips, participants were invited to express their views on different aspects of subtitling in a short semi-structured interview. One of the main findings which transpired from the interviews was that many participants believed that the number of lines in subtitling should depend on the genre: whereas they would prefer to have two lines in fiction films, they would accept three lines in dialogue-heavy non-fiction programmes, such as news or chat shows, as the one used in our study. This may explain why a number of people, particularly the Spanish, chose three lines as their preferred option. A deaf participant said that three-line subtitles in fast-paced programmes “*help to make out what the person is saying*”. Someone else noted that with three-liners, the continuity of text is less cut. One Polish participant said “*I was kind of surprised because I always thought that three lines is too much, and here actually it wasn't that bad*”, but they still preferred two-liners.

Overall, however, the vast majority of the participants declared a clear preference for two-line subtitles. An English participant explained that “*two liners are better, because if there are three liners, you focus too much on the subtitles and you don't focus too much on what's going on and it's just too much information.*” One Spanish participant found the information in three-line subtitles “*too concentrated*” and another stated that “*three lines require more attention to the subtitle*”. Some participants argued that although three-liners provide more information at once, which can be useful when there are more characters speaking simultaneously, two-liners allow for more involvement and flow, require less eye movement and do not obscure the images.

Discussion

By conducting this study, we investigated the impact of the number of two- and three-line subtitles on the cognitive processing, comprehension, enjoyment and preferences of

viewers with different linguistic backgrounds.

One of the most important findings of this study is that three-line subtitles resulted in higher cognitive load than two-line subtitles, as evidenced by self-reports and the longer time spent in the subtitle area. Cognitive effort was declared to be higher in the case of three-line subtitles by all study participants, regardless of their linguistic background. These findings may be attributed to an increase of the type of cognitive load known as ‘extraneous’ (Sweller, Van Merriënboer, & Paas, 1998), related to the way information is presented. Three-line presentation may impose more demands on participants’ working memory (Baddeley, 2007), thus leaving them with fewer cognitive resources necessary not only to read three lines of text on screen, but also to attend to the soundtrack and on-screen action. We acknowledge that the effect size and numerical differences we found were not large, however they were already discernible with short videos, and we believe they could be compounded in longer videos. Longer exposure to three-line subtitles could cause more strain on viewers’ cognitive processing, in line with the results reported by Ackerman and Kanfer (2009), who found that cognitive fatigue increases with task length.

Another significant finding of this study is that the vast majority of viewers were quite clear in their preferences: two lines were favoured over three lines. Interestingly, about one third of English hearing participants declared no preference, which on the one hand may be attributed to the fact that the programme was presented in their mother tongue and, on the other hand, to their familiarity with three-line subtitles on British television. That habit may affect preference was also confirmed by a Polish participant who said in the interview: *“I don’t know, maybe I just got used to those two-line subtitles, that’s why I prefer them”*. We acknowledge that viewers’ preferences may also be dependent on other factors, such as film genre or type of subtitling (intralingual

vs. interlingual). Preferences may also be related to the type of screen used for watching, for instance Gerber-Morón, Soler-Vilageliu, and Castella (2019) found that viewers preferred shorter two-line subtitles for larger screens and longer one-line subtitles for smartphones.

The preferences declared by the participants in our study were in line with enjoyment, which was generally higher in two-line subtitles. Enjoyment is often defined as “a pleasurable response to media use” (Tamborini, Bowman, Eden, Grizzard, & Organ, 2010, p. 758) and we acknowledge it may be difficult to isolate the impact of subtitles from the impact of the show itself. Viewers may declare higher enjoyment simply because they liked the content of the show rather than the nature of the subtitles. Similarly to the results from the self-reported cognitive load, the effect size in enjoyment scores in this study was not large, so our results would need to be corroborated by other studies, using different clips and languages.

Despite the differences in the self-reported cognitive load, enjoyment and preferences, comprehension was not found to be affected by the number of lines. This is in line with some previous studies, which also found that the manipulation of various subtitle parameters did not impact on comprehension, as was the case in the study by Perego et al. (2010) on subtitle segmentation, and two studies on subtitle speeds by Szarkowska et al. (2011) and Szarkowska and Gerber-Morón (2018b). The lack of differences in comprehension scores between the conditions may be attributed to viewers' previous exposure to different types of subtitling – both high-quality subtitles fully conforming to the standards and poor-quality subtitles not necessarily compliant with all subtitling rules. Viewers may have learnt to process different subtitles efficiently, regardless of their quality.

Our eye-tracking results confirm previous findings related to the processing of one- and two-line subtitles, where a higher number of lines resulted in proportionally more time spent in the subtitle area (d'Ydewalle et al., 1991; Praet et al., 1990). Previous studies showed, for instance, that American hearing subjects reading intralingual English to English subtitles spent 21.87% time on two-line subtitles and only 16.37% on one-liners (d'Ydewalle et al., 1991). Similarly, d'Ydewalle and De Bruycker (2007) reported that in standard interlingual subtitling, adult viewers spent proportionally more time reading two-line subtitles (37%) than one-line subtitles (31%). Interestingly, the values reported in these studies are substantially lower compared to our study, in which participants looked at the subtitles for about half of their display time. These differences may be attributed to a fast pace of contemporary productions and to the speed of subtitles (for more on subtitle speed see Romero-Fresco (2015) and Szarkowska and Gerber-Morón (2018b)). In our study, participants spent slightly more time in the three-line compared to the two-line condition, which means that they had relatively less time to follow the on-screen action. It does not mean, however, that their reading pattern was more regular with three lines in comparison with two lines (cf. d'Ydewalle and De Bruycker (2007)). This is because while reading, people were gazing a lot at the image too: their reading of subtitles was not linear and smooth, particularly in the case of hard of hearing and deaf participants, who moved their eyes more between the image and the subtitle (i.e. made more revisits) in the three-line condition. The higher number of revisits from the image to the subtitle in the three-line condition may also stem from the fact that these subtitles were displayed on screen longer than two-liners. It is also possible that when reading three-line subtitles, viewers found it more difficult to return to the correct position after having looked at the image or when moving eyes between the lines, which resulted in more return sweeps (Rayner,

1998). Because they had to make corrective eye movements, the time they spent in the subtitle was longer, as was also argued by d'Ydewalle et al. (1991) in their interpretation on the differences in reading time between one- and two-line subtitles.

We also found considerable differences between participants depending on their linguistic background. Spanish participants achieved the lowest comprehension scores and had the longest mean fixation duration, indicating higher processing effort (Holmqvist et al., 2011). In contrast, Polish participants had the highest comprehension scores and the lowest mean fixation duration. This result may be attributed to the fact that Spanish participants grew up in a dubbing country. On the other hand, however, the country of provenance cannot be unequivocally equated with higher or lower familiarity with subtitling, as people's preferences and habits are changing (Matamala et al., 2017; Szarkowska & Laskowska, 2015) due to technological developments and new solutions mushrooming on the audiovisual translation market. It is also possible that the lower results for Spanish people are simply coming from the characteristics of this particular sample. On their part, English hearing participants had longer mean fixation duration in the three-line condition compared to the two-line conditions. This may be interpreted as reflecting more effortful processing (Holmqvist et al., 2011), possibly due to the simultaneous presentation of dialogue and subtitles with identical words in the participants' mother tongue, resulting in verbal redundancy (Kalyuga, Chandler, & Sweller, 1999; Moreno & Mayer, 2002).

Subtitle processing was also affected by hearing loss. Firstly, deaf participants spent significantly more time reading the subtitles than other groups. Such results tend to be ascribed to deaf people's lower reading skills (Mayberry, del Giudice, & Lieberman, 2011; Musselman, 2000; Trezek, Wang, & Paul, 2010). Secondly, the number of revisits to the subtitle in the three-line condition was much higher for hard of

hearing and deaf participants compared to the hearing. That hard of hearing people made more revisits to the subtitle than the hearing was an observation previously made by Szarkowska et al. (2011). It is also in line with the results reported by De Linde and Kay (1999), who found that for the clip where the speaker was visible on screen, deaf people made more regressions (than the hearing) “caused by constant deflections to the screen image in order to check lip movements” (p. 70). This may reflect a reading pattern typical of people with hearing loss, who tend to compare the subtitles with the speaker’s lip movements much more than the hearing. Thirdly, hard of hearing and deaf participants generally declared to enjoy the subtitled clips more than hearing English people, which may be explained by their inherent need for subtitling and their previous extensive exposure to this type of audiovisual translation. Finally, although deaf participants reported higher enjoyment in the three- rather than two-line condition, they unanimously preferred two lines.

Conclusions

This study has shown that three-line subtitles are more cognitively demanding than two-line subtitles and that people prefer subtitles with two rather than three lines. Examining viewers’ cognitive load and enjoyment is important in order to design good quality subtitles as “a successful film viewing experience is at the same time cognitively effective and pleasant” (Perego, 2016). Effective subtitle design should therefore limit the extraneous load on viewers’ working memory to optimise subtitle processing. Our findings – once replicated and confirmed by other studies – may constitute a step towards a change in subtitling guidelines and practices.

Acknowledgements

The research reported here has been supported by a grant from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 702606, from "La Caixa" Foundation (E-08-2014-1306365) and from Transmedia Catalonia Research Group (2017SGR113).

References

- Ackerman, P. L., & Kanfer, R. (2009). Test length and cognitive fatigue: an empirical examination of effects on performance and test-taker reactions. *Journal of Experimental Psychology: Applied*, 15(2), 163-181. doi:10.1037/a0015719
- Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.
- Baker, R. G., Lambourne, A., & Rowston, G. (1984). *Handbook for television subtitlers*. Winchester: Independent Broadcasting Authority.
- BBC. (2017). BBC subtitle guidelines. Retrieved from BBC Subtitle Guidelines website: <http://bbc.github.io/subtitle-guidelines/>
- Bisson, M.-J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2012). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, 35(02), 399-418. doi:10.1017/s0142716412000434
- Bogucki, Ł. (2009). Amateur subtitling on the internet. In J. Díaz Cintas & G. Anderman (Eds.), *Audiovisual Translation: Language Transfer on Screen* (pp. 49-57). London: Palgrave Macmillan UK.
- Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, 38(1), 53-61.
- Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8(4), 293-332. doi:10.1207/s1532690xci0804_2
- Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research*, 32(3), 453-467. doi:10.1177/0267658316637401
- d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist*, 12(3), 196-205. doi:10.1027/1016-9040.12.3.196
- d'Ydewalle, G., Praet, C., Verfaillie, K., & Van Rensbergen, J. (1991). Watching subtitled television: automatic reading behavior. *Communication Research*, 18(5), 650-666.
- d'Ydewalle, G., Rensbergen, J. v., & Pollet, J. (1987). Reading a message when the same message is available auditorily in another language: the case of subtitling. In J. K. O'Regan & A. Levy-Schoen (Eds.), *Eye movements: from physiology to cognition* (pp. 313-321). Amsterdam/New York: Elsevier.
- De Bruycker, W., & d'Ydewalle, G. (2003). Reading native and foreign language television subtitles in children and adults. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research* (pp. 671-684). Amsterdam: North-Holland.
- De Linde, Z., & Kay, N. (1999). *The semiotics of subtitling*. Manchester: St. Jerome.

- Díaz Cintas, J., & Remael, A. (2007). *Audiovisual translation: subtitling*. Manchester: Manchester : St. Jerome.
- Doherty, S., & Kruger, J.-L. (2018). The development of eye tracking in empirical research on subtitling and captioning. In J. Sita, T. Dwyer, S. Redmond, & C. Perkins (Eds.), *Seeing into Screens* (pp. 46-64). London: Bloomsbury.
- Dwyer, T. (2017). *Speaking in subtitles: revaluating screen translation*. Edinburgh: Edinburgh University Press.
- Gerber-Morón, O., Soler-Vilageliu, O., & Castella, J. (2019). The effects of screen size on subtitle layout preferences and comprehension across devices. *Hermeneus*, 21.
- Gerber-Morón, O., & Szarkowska, A. (2018). Line breaks in subtitling: an eye tracking study on viewer preferences. *Journal of Eye Movement Research*, 11(3), 1-22.
- Gerber-Morón, O., Szarkowska, A., & Woll, B. (2018). The impact of text segmentation on subtitle reading. *Journal of Eye Movement Research*, 11(4), 1-18. doi:10.16910/jemr.11.4.2
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: a comprehensive guide to methods and measures*. Oxford: Oxford University Press.
- Ivarsson, J., & Carroll, M. (1998). *Subtitling*. Simrishamn: TransEdit HB.
- Jensema, C., Danturthi, R. S., & Burch, R. (2000). Time Spent Viewing Captions on Television Programs. *American Annals of the Deaf*, 145(5), 464-468.
- Just, M., & Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4), 329-354.
- Kalantzi, D. (2008). *Subtitling for the deaf and hard of hearing: A corpus-based methodology for the analysis of subtitles with a focus on segmentation and deletion*. (PhD), University of Manchester, Manchester.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied cognitive psychology*, 13(4), 351-371.
- Koolstra, C. M., Van Der Voort, T. H. A., & d'Ydewalle, G. (1999). Lengthening the presentation time of subtitles on television: effects on children's reading time and recognition. *Communications*, 24(4), 407-422. doi:10.1515/comm.1999.24.4.407
- Krejtz, I., Szarkowska, A., & Krejtz, K. (2013). The effects of shot changes on eye movements in subtitling. *Journal of Eye Movement Research*, 6(5:3), 1-12. doi:10.16910/jemr.6.5.3
- Krejtz, I., Szarkowska, A., & Łogińska, M. (2016). Reading function and content words in subtitled videos. *Journal Of Deaf Studies And Deaf Education*, 21(2), 222-232. doi:10.1093/deafed/env061
- Kruger, J.-L., & Doherty, S. (2016). Measuring cognitive load in the presence of educational video: Towards a multimodal methodology. *Australasian Journal of Educational Technology*, 32(6), 19-31.
- Kruger, J.-L., Doherty, S., Fox, W., & de Lissa, P. (2017). Multimodal measurement of cognitive load during subtitle processing: Same-language subtitles for foreign language viewers. In I. Lacruz & R. Jääskeläinen (Eds.), *New Directions in Cognitive and Empirical Translation Process Research* (pp. 267-294). London: John Benjamins.
- Kruger, J.-L., Hefer, E., & Matthew, G. (2013). *Measuring the impact of subtitles on cognitive load: eye tracking and dynamic audiovisual texts*. Paper presented at

- the Proceedings of the 2013 Conference on Eye Tracking South Africa, Cape Town, South Africa.
- Kruger, J.-L., Hefer, E., & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. *Journal of Eye Movement Research*, 7(5), 1-15.
- Kruger, J.-L., Szarkowska, A., & Krejtz, I. (2015). Subtitles on the moving image: an overview of eye tracking studies. *Refractory*, 25.
- Łuczak, K. (2017). *The effects of the language of the soundtrack on film comprehension, cognitive load and subtitle reading patterns. An eye-tracking study.* (MA), University of Warsaw, Warsaw.
- Matamala, A., Perego, E., & Bottiroli, S. (2017). Dubbing versus subtitling yet again? *Babel*, 63(3), 423-441. doi:10.1075/babel.63.3.07mat
- Mayberry, R. I., del Giudice, A. A., & Lieberman, A. M. (2011). Reading achievement in relation to phonological coding and awareness in deaf readers: a meta-analysis. *Journal Of Deaf Studies And Deaf Education*, 16(2), 164-188. doi:10.1093/deafed/enq049
- Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: when reading helps listening. *Journal of Educational Psychology*, 94(1), 156-163. doi:10.1037/0022-0663.94.1.156
- Muñoz, C. (2017). The role of age and proficiency in subtitle reading. An eye-tracking study. *System*, 67, 77-86. doi:10.1016/j.system.2017.04.015
- Musselman, C. (2000). How do children who can't hear learn to read an alphabetic script? A review of the literature on reading and deafness. *Journal Of Deaf Studies And Deaf Education*, 5(1), 9-31.
- Neves, J. (2008). Ten fallacies about subtitling for the d/Deaf and the hard of hearing. *Journal of Specialised Translation*, 10, 128-143.
- Neves, J. (2009). Interlingual subtitling for deaf and hard of hearing. In J. Díaz Cintas & G. Anderman (Eds.), *Audiovisual translation. Language transfer on screen* (pp. 151-169). Basingstoke: Palgrave Macmillan.
- Ofcom. (2017). Ofcom's Code on Television Access Services. 1-25.
- Paas, F., & Merriënboer, J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review*, 6(4), 351-371. doi:10.1007/BF02213420
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist*, 38(1), 1-4. doi:10.1207/s15326985ep3801_1
- Paas, F., Renkl, A., & Sweller, J. (2004). *Advances in cognitive load theory, methodology and instructional design.* London: Kluwer Academic Publishers.
- Paas, F., Renkl, A., & Sweller, J. (2016). *Cognitive Load Theory: A Special Issue of Educational Psychologist:* Routledge.
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38(1), 63-71. doi:10.1207/s15326985ep3801_8
- Paas, F., & Van Merriënboer, J. (1993). The efficiency of instructional conditions: an approach to combine mental effort and performance measures. *Human Factors*, 35(4), 737-743. doi:10.1177/001872089303500412
- Perego, E. (2016). Gains and losses of watching audio described films for sighted viewers. *Target*, 28(3), 424-444. doi:10.1075/target.28.3.04per

- Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology, 13*(3), 243-272. doi:10.1080/15213269.2010.502873
- Perego, E., Del Missier, F., & Stragà, M. (2018). Dubbing vs. subtitling. *Target, 30*(1), 137-157. doi:10.1075/target.16083.per
- Perego, E., Laskowska, M., Matamala, A., Remael, A., Robert, I. S., Szarkowska, 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. doi:10.1556/084.2016.17.2.4
- Plass, J. L., Moreno, R., & Brünken, R. (2010). *Cognitive load theory*. Cambridge: Cambridge University Press.
- Praet, C., Verfaillie, K., De Graef, P., Van Rensbergen, J., & D'Ydewalle, G. (1990). A one line text is not half a two line text. In R. Groner, G. D'Ydewalle, & R. Parham (Eds.), *From Eye to Mind: Information Acquisition in Perception, Search and Reading* (pp. 205-213). Amsterdam: Elsevier.
- Program, D. a. C. M. (2017). *DCMP Captioning Key*. Retrieved from <http://www.captioningkey.org/>
- 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, 21*(1), 5-21. doi: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. doi:10.1037/0033-2909.124.3.372
- Robson, G. D. (2004). *The Closed Captioning Handbook*. Burlington: Focal Press.
- Romero-Fresco, P. (2009). More haste less speed: Edited versus verbatim respoken subtitles. *Vigo International Journal of Applied Linguistics, 6*, 109-133.
- Romero-Fresco, P. (2015). Final thoughts: Viewing speed in subtitling. In P. Romero - Fresco (Ed.), *The Reception of Subtitles for the Deaf and Hard of Hearing in Europe* (pp. 335-341). Bern: Peter Lang.
- Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation, 55*, 37-76. doi:10.1016/B978-0-12-387691-1.00002-8
- Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. New York: Springer.
- Sweller, J., Van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296. doi:10.1023/a:1022193728205
- Szarkowska, A. (2013). Towards interlingual subtitling for the deaf and the hard of hearing. *Perspectives, 21*(1), 68-81. doi:10.1080/0907676X.2012.722650
- Szarkowska, A., & Gerber-Morón, O. (2018a). *SURE Project Dataset*. Retrieved from: <https://repod.pon.edu.pl/dataset/sure-project>
- Szarkowska, A., & Gerber-Morón, O. (2018b). Viewers can keep up with fast subtitles: Evidence from eye movements. *Plos One, 13*(6). doi:10.1371/journal.pone.0199331
- Szarkowska, A., Krejtz, I., Kłyszczko, Z., & Wiczorek, 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.
- 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.
doi:10.1556/084.2016.17.2.3

- Szarkowska, A., & Laskowska, M. (2015). Poland – a voice-over country no more? A report on an online survey on subtitling preferences among Polish hearing and hearing-impaired viewers. In Ł. Bogucki & M. Deckert (Eds.), *Accessing audiovisual translation* (pp. 179–197). Bern: Peter Lang.
- Tamborini, R., Bowman, N. D., Eden, A., Grizzard, M., & Organ, A. (2010). Defining media enjoyment as the satisfaction of intrinsic needs. *Journal of communication*, 60(4), 758-777. doi:10.1111/j.1460-2466.2010.01513.x
- Trezek, B. J., Wang, Y., & Paul, P. V. (2010). *Reading and deafness: theory, research, and practice*. Clifton Park, N.Y.: Delmar Cengage Learning.
- Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching: subtitles for the deaf and hard-of-hearing as tools for language learning*. London: Palgrave Macmillan.
- Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: an eye-tracking study. *The Modern language journal*, 97(1), 254-275.

¹ In this paper we follow the European convention and use the term ‘subtitling’ as opposed to ‘captioning’ to denote both intra- and interlingual transfer. We retain the term ‘captions’ in citations only.

² Reversed subtitles contain text in a foreign language accompanying a programme in the soundtrack in viewers’ mother tongue, for instance English subtitles to a Polish programme for Polish viewers.

³ We are well aware of the problematic nature of using self-reports to assess one’s own level of proficiency in a foreign language. Yet, owing to the time limitations and the nature of the study, it was not possible for us to test participants’ language proficiency directly. Participants were given a sheet (Szarkowska & Gerber-Morón, 2018b) describing the skills required at each proficiency level using the Common European Framework of Reference for Languages. They made their assessment based on the sheet as well as their own experience with the European system which they were familiar with from their education in schools and at university.

⁴ For example, if a subtitle lasted for 3 seconds and the participant spent 2.5 seconds in that subtitle, the percentage dwell time was $2500/3000 \text{ ms} = 83\%$, i.e. while the subtitle was displayed for 3 seconds, the participant was looking at that subtitle for 83% of the time.