

## **Using Eye Tracking to Study the Effect of Badly Synchronized Subtitles on the Gaze Paths of Television Viewers**

Juha Lång, Jukka Mäkisalo

School of Humanities, University of Eastern Finland, FINLAND

Tersia Gowases, Sami Pietinen

School of Computing, University of Eastern Finland, FINLAND

### **ABSTRACT**

The present study had two aims: 1) to examine how breaking the synchronization of subtitles and spoken dialogue affects the attention allocation of television viewers, and 2) to assess the usability of eye-tracking methodology in studying subtitling conventions. This was achieved by analyzing the eye-tracking data of two groups of subjects who watched different versions of a subtitled television programme: in one version the subtitles were composed according to the subtitling rules used by Yle, the Finnish national public service broadcasting company, while in the other version the subtitles included thirty-one cases in which the rules were broken in various ways. The gaze data confirmed the hypothesis that disrupting the synchronization draws the gaze more to the subtitles, but the questionnaire analysis suggested that the subjects were not conscious of this effect and failed to notice most of the manipulations. Nevertheless, eye-tracking methodology has much to offer to the research of audio-visual translation.

**KEYWORDS:** eye movements, fixations, reception, subtitles.

### **1. Introduction**

Subtitling is the preferred method of translating foreign television programmes and films in many small countries, like the Nordic countries, Belgium and the Netherlands, where dubbing is not profitable because of its higher cost. The pros and cons of both methods have been widely debated (Diaz-Cintas 1999, Koolstra et al. 2002). One of the most common arguments against subtitling is that subtitles are an added visual feature that draws viewers' attention away from the main visual elements of the picture. Indeed, a study by Tosi et al. (1997:50) found that written elements on the screen, such as subtitles, draw viewers' attention even when they are in the periphery of the visual field. Furthermore, d'Ydewalle and De Buycker (2007:203) noticed that subtitles are read even when the programme's audio track is in the viewers' native language and the subtitles are in a language they do not understand.

In a more recent study, Perego et al. (2010) examined the effectiveness of subtitle processing by using multiple measures, including gaze data, word recognition, and visual scene recognition. According to subtitling conventions, parts of a noun phrase (NP) should not be divided into different lines or subtitles, and Perego's research group wanted to know how breaking this convention affects the processing of the subtitle or the picture. They concluded that subtitle processing is indeed efficient and subtitles do not have a negative effect on viewers' scene recognition. This suggests that, although subtitles draw the viewer's attention away from the main visual aspects of the screen (Tosi et al. 1997:50), they do not distract the viewer from processing the rest of the image effectively, at least when the subtitles are properly synchronized to the audio and video channels. Furthermore, Perego et al. found that dividing the parts of a noun phrase into different lines of the subtitle did not negatively affect

viewers' performances on word recognition tests (2010:257) and thus has no negative effect on the cognitive processing of NPs and the readability of subtitles.

The main function of subtitles is to convey to the viewer the message of the dialogue while at the same time allowing the viewer to follow the events on the screen. To reach this double aim of both clarity and inconspicuousness, subtitling companies have a set of subtitling rules or conventions (see Diaz-Cintas and Remael 2006 for a general summary of subtitling conventions used in Europe). These conventions seem to differ only slightly across different languages, at least in the Nordic countries. The conventions give guidelines, for example, about how the subtitle should be constructed and when the subtitle should appear and disappear. Nevertheless, there seems to be little empirical basis for these rules (Vertanen 2007:150), and as the study by Perego et al. (2010) has shown, some of these conventions might not be that important.

Although reading has always been one of the most researched topics in eye tracking studies (for a review of early eye tracking studies, see Rayner 1998), subtitles have received only a fraction of this attention. One of the few scholars to examine how subtitles are read is Géry d'Ydewalle, who has conducted several studies on the subject with his colleagues. In his early studies (d'Ydewalle et al. 1987; d'Ydewalle et al. 1991), he concluded that reading subtitles is at least partially an automated process and that subtitles are read even when they are not necessary for the understanding of the dialogue. He argued that this happens because the textual channel (written subtitles) is a more efficient channel of information than the audio channel (spoken dialogue) (1991:661-662). He has also observed that reading subtitles is comparable to reading normal text (a book or a journal article, for instance), but that there are some differences, especially when reading single-line subtitles: fixations are generally shorter and there are more regressive fixations (d'Ydewalle and De Bruycker 2007:202-203).

During the last decade, eye-tracking methodology has been introduced to translation studies. Many of the studies have concentrated on the process of translating (see Göpferich et al. 2008 for an overview). For example, Sharmin et al. (2008) examines the gaze patterns of translation students while they were translating, and O'Brien (2006; 2008) has used eye tracking in studying the use of translation memory tools. Surprisingly few of the studies, though, deal with audio-visual translation. Perego et al. (2010), mentioned above, is one of the few exceptions, and even they used eye trackers with only some of their subjects.

Caffrey (2008a; 2008b) has employed gaze tracking more extensively to study how viewers process visual non-verbal cues and additional textual information, such as text boxes containing explanations for cultural references that are presented on the screen in addition to subtitles. His results reveal that showing the extra information boxes at the same time as subtitles increases the cognitive effort needed to process either one of the textual information channels (subtitles or text boxes), and that the extra information channel causes the viewers to skip more subtitles than in normal subtitling conditions (Caffrey 2008b:141-142). This is not surprising considering the time constraints: the viewer does not have the time to read both the subtitles and the text boxes and thus time and attention are split between the two.

This paper describes a pilot study that has two aims: firstly, to examine with an eye-tracking camera how disrupting the synchronization between the audio dialogue and the subtitles affects viewers' gaze paths; and secondly, to assess the usability of eye-tracking methodology in studying subtitling conventions. Previous studies have shown that eye-tracking can be a

useful tool for text processing and reception research, because it enables researchers to tap into the subjects' unconscious reactions to test stimuli.

The primary data of the present study was collected by recording the eye movements of twenty-five Finnish subjects watching an episode of the British comedy series *Absolutely Fabulous*. Two different versions were shown: one had the subtitles used by Yleisradio (Yle), Finland's national public service broadcasting company, when the episode was aired in 2008, and the other used a manipulated version of the same subtitles. The manipulations, thirty-one in total, included structural and semantic manipulations (fifteen cases) as well as synchronization manipulations (sixteen cases). The study concentrates on the latter. Each subject saw only one version of the programme, and the analysis was conducted by comparing the gaze data of the subjects in the different subtitle versions both quantitatively and qualitatively: the statistical analysis was done on the basis of fixation counts (the number of fixations the subjects made on the subtitles), while the qualitative analysis included comparing the subjects' scan paths. Secondary data included a questionnaire in which the subjects were asked to report conscious observations about the programme and the subtitles. The main hypothesis of the study was that the manipulations cause confusion, and that this will manifest itself in the fixation data as increased fixation counts and discrepancies in scan paths.

In section 2 of this paper, an overview of the method and test setting is given, as well as a detailed account of the manipulations. Section 3 sums up the results of the study and section 4 gives a more detailed discussion of the results and a summary of the conclusions. The results confirm the initial hypothesis by showing differences in the gaze paths and fixation counts of the two groups of subjects. In general, the manipulated subtitles attract the viewers' gaze more to the subtitle area. Then again, the results from the analysis of the questionnaire did not show great subjective differences between the groups. This means that the subjects were not conscious of the disturbing effect of the manipulations.

## 2. The Test Setting

The subjects were shown a full twenty-five-minute episode of the British comedy series *Absolutely Fabulous* with one of the two different subtitle versions. The participants were recruited from the faculty staff and students of the University of Eastern Finland, and a total of 25 subjects participated in the study (9 male, 16 female, mean age = 24, standard deviation (SD) = 7). The participants were randomly assigned into two groups: approximately half of the group (henceforth control group, 12 subjects) saw the video with the same subtitles that Yle used when broadcasting the episode, while the rest (test group, 13 subjects) saw the video with manipulated subtitles. The gaze data of two participants of the test group was discarded due to bad recording quality, which means that the gaze data analysis was done on the basis of 23 participants in total.

Each of the participants watched the video clip individually. The test supervisor (one of the authors of the present paper) was the only other person in the room, and he was situated out of the participant's line of sight. After calibrating the tracker, the subjects were given written instructions in which they were asked to watch the video clip and afterwards fill in a questionnaire. The participants were told beforehand that the study involved examining how people watch television, but did not know that the aim was to study how they read the subtitles or that the subtitles may have been manipulated.

The subjects' eye movements were recorded with a Tobii X120 eye-tracker<sup>1</sup> and Tobii Studio (version 1.2) software. The Tobii X120 is a stand-alone eye-tracking unit that uses infrared cameras to track the subject's eye movements. It is relatively tolerant to head movements, which increases the reliability and accuracy of the data and eliminates the need for head rests or bite bars to keep the subject's head immobile. In our study, the video stimulus was shown in a well-lit room on a 24-inch standard definition LCD television screen, which was connected to a computer.

*Absolutely Fabulous* is a British comedy series created by Jennifer Saunders, who also plays the main role. The show follows the antics of Edina Monsoon (Saunders), a twice-divorced and middle-aged head of a successful PR agency who strolls through life smoking marijuana and drinking vodka with her best friend Patricia "Patsy" Stone (played by Joanna Lumley), in a desperate attempt to stay young. Patsy is an editor of a fashion magazine who indulges in various intoxicating substances and casual sex and is often the catalyst of bad ideas becoming reality in Edina's life. Other recurring characters include Edina's daughter Saffron (Julia Sawalha), the complete opposite of her mother and often the only voice of reason in Edina's household; Edina's mother (June Whitfield), who is always meddling with Edina's life and clearly disapproves of her lifestyle; and Edina's air-headed secretary Bubble, whose sense of fashion is even more outrageous than Edina's. The show has run a total of five seasons since its premiere in November 1992. The final episode of the fifth season originally aired in November 2003, but the story has continued with several specials, including a special twentieth anniversary episode (BBC 2012).

In the episode used in our study (season 3, episode 5, *The Fear*, first aired in 1995), Saffron has just started university and moved out, leaving Edina and Patsy living together in Edina's house. This is only one of the many changes that are happening in Edina's life, causing her great anxiety. She feels that her life has lost meaning and needs a change. Thus the episode ends with two dramatic decisions; Edina decides to sell the house and Patsy decides to move to New York.

The initial plan was to make the viewing environment as natural as possible, but this did not succeed particularly well for two reasons. Firstly, the eye tracker had to be positioned relatively close to the subject (approximately sixty centimetres from the eyes) and thus could not be hidden. Secondly, the subject has to stay relatively motionless for the eye tracker to collect data reliably. No chin rests or bite bars were used in the study, but the subjects had to sit in an office chair (as opposed to, for instance, a comfortable sofa, in which most people normally watch television) and they were instructed to remain as still as possible, which in itself can cause the situation to feel somewhat unnatural. Furthermore, it was impossible to track the subjects' eye movements covertly since in order to get accurate data the tracker has to be calibrated individually for each subject at the beginning of the recording session. It would have required extraordinary guile to do this without the subjects realizing that their eye movements were being tracked.

The manipulated subtitles were identical to the Yle subtitles except for thirty-one cases in which subtitling conventions were broken in one way or another. The cases included instances in which the structure of the subtitle was changed (for example, a compound word was split across two consecutive subtitles), and cases in which the synchronization between the subtitle and the audio dialogue was broken. Among the structural manipulations there were four cases

---

<sup>1</sup> Tobii Technology AB, Stockholm, Sweden.

that included a semantic translation error, i.e. subtitles in which a word or phrase was translated incorrectly (literally, in most cases) resulting in phrases that make little sense.<sup>2</sup> The study concentrates only on the effect of synchronization manipulations on the viewers' eye movements, but the effects of the structural manipulations are also taken into consideration in the analysis of the questionnaire answers. No subtitle included more than one type of manipulation. Yle's subtitles were chosen as the basis of the manipulations because Yle imposes strict guidelines for subtitling on its translators and its subtitles are generally thought to be of high quality (see Vertanen 2007 for a summary of the guidelines). Yle's subtitling guidelines are also generally in accordance with those used by other subtitling companies in Finland.

According to Yle's guidelines, the general rule is that the subtitle should appear when the character on screen starts to say the corresponding line and disappear when the character finishes pronouncing the corresponding line (Vertanen 2007:152). In addition to this, a full one-lined subtitle (approximately thirty-four symbols, including spaces) should stay on-screen for two to three seconds and a full two-lined subtitle for four to five seconds in order to ensure that the viewer has enough time to read the subtitle. The minimum display times of shorter subtitles are calculated accordingly, but no subtitle should ever be visible for less than one second (*ibid.*). In practice the latter rule takes priority if the dialogue is fast-paced, as in the video chosen for this study. Synchronization cases consisted of three types of manipulations that broke these rules: delayed subtitles, lengthened subtitles, and shortened subtitles. In delayed cases, the display time of the subtitle (both in cue and out cue times) was delayed 1.5, 2.0 or 2.5 seconds from the original time code. There were two cases of each type in the manipulated subtitles. In cases of lengthened subtitles, the display time was lengthened to two, three or four times the original. This meant that the subtitle stayed on the screen after the corresponding line had been said. These manipulations were done so that they did not have any effect on the time codes of the other subtitles and there was no overlapping. The final type of synchronization manipulations included instances where the subtitles were completely synchronized to the dialogue so that the on-screen time was shortened to violate the guideline for subtitle duration. There were a total of sixteen synchronization cases in the manipulated subtitles.

In addition to the gaze path data, another set of data was gathered with a questionnaire. While the gaze data gives information about the subject's involuntary and subconscious reactions to stimuli, the questionnaire was intended to ascertain the subject's conscious observations about the subtitles and the manipulations. It included questions concerning background information (gender, age, subject's own assessment of language skills, etc.) as well as questions about the video clip and its subtitles. The subjects were asked to answer the following questions (our translation from Finnish):

---

<sup>2</sup> The notion of translation error is problematic when subtitles are concerned. Because of the spatial and temporal restrictions, the message of the dialogue is often very much condensed in the subtitles, and there is no space for explanations of, for example, cultural references. This means that translators often have to detach themselves completely from the source text and invent a line that fits the context but does not necessarily have much in common with the original. Furthermore, as subtitling includes more than just translating the text, namely composing and time cueing the subtitles, sometimes problems in synchronization or line division can also be considered as translation errors.

1. Did something in the subtitles catch your attention while watching the clip? What?
2. Did you notice any translation errors in the subtitles?<sup>3</sup>
3. Did you have trouble following the show at any point?

Finally, the subjects were asked to assess the quality of the subtitles on a scale of one to five (one being the lowest).

The data was analysed both quantitatively and qualitatively. The quantitative analysis included counting the number of fixations on the subtitles and comparing the results of the two subject groups with statistical analysis tools by means of an analysis of variance (ANOVA) and  $\chi^2$ -test.<sup>4</sup> Parts of the questionnaire were also analysed statistically. Qualitative analysis involved analysing manually how gaze paths of the test group differed from the control group's gaze path in the manipulated cases. Tobii Studio allows the numerical gaze-coordinate data to be exported as video, which shows the stimulus video clip embedded with the viewer's gaze path. This allows us to see exactly where the subject is fixating on at any given moment. The temporal and spatial thresholds of a fixation were set to 100 milliseconds and 30 pixels, which are Tobii's recommendations for a mixed content stimulus.

### **3. Results: a comparative analysis of gaze data**

The results will be presented in four parts according to the three types of manipulations and the questionnaire. The fourth part will concentrate only on the relevant points of the questionnaire analysis in respect to the gaze analysis. The cases are not numbered chronologically but according to the type of the manipulation in order to make the analysis more comprehensible.

#### **3.1. Delayed Subtitles**

As mentioned above, d'Ydewalle and his colleagues have concluded in several studies (d'Ydewalle et al. 1987; d'Ydewalle et al. 1991; d'Ydewalle and De Bruycker 2007) that reading subtitles is at least partially an automated process. The present study attempted to test this theory by delaying the appearance of the subtitle in a few places during the programme.

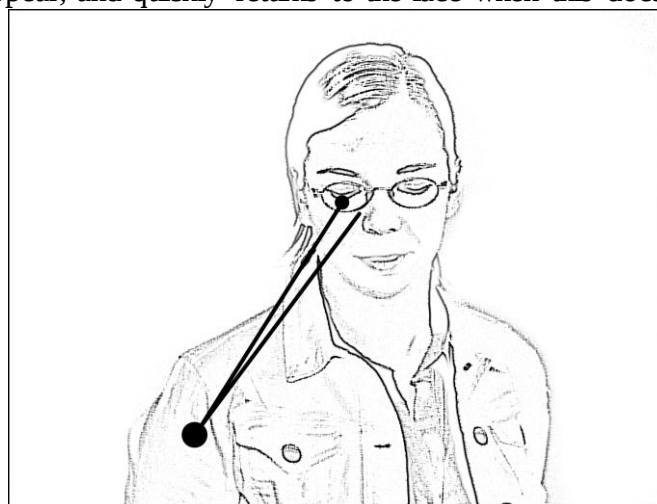
As expected, in the cases of delayed subtitles the single factor ANOVA analysis of variance did not show statistically significant differences in the number of fixations on the subtitles between the control group and the test group. The time of the delay did not seem to affect the fixation counts either. Despite this, a more detailed analysis revealed an interesting tendency in the test group's gaze paths. Many of the subjects shifted their gaze towards the subtitle area when the actor on-screen started to speak although the subtitle had not appeared there. Figure 1 illustrates the seeking behaviour, with the black line and dots representing the gaze path on the screen. At first the viewer's eyes are fixated on an interesting area of the screen, for example, on the face of the actor. When the character starts to speak, the gaze moves to the subtitle area at the bottom of the screen, but promptly returns to the main area of the screen when the subtitle is noticed to be absent.

---

<sup>3</sup> It should be noted that the term "translation error" was not specified in any way in the questionnaire, but the definition of what is a translation error was left open for the subjects' own interpretation.

<sup>4</sup> ANOVA and  $\chi^2$ -test (chi-squared test, also known as Pearson's chi-squared test) are statistical tests that are used in assessing the statistical significance of quantitative data. They both do this by comparing the mean values and variation of the data and give result as a p-value. To greatly simplify, p represents the possibility of getting a difference which is equal to or greater than the observed difference, provided that there is no dependency or difference between the datasets. If  $p < 0.05$  the difference between tested sets of data is considered to be statistically significant.

Figure 1. An illustration of a typical occurrence of the seeking behaviour: the gaze path (black line and dots) starts from the face of the on-screen figure, moves to the subtitle area where text is expected to appear, and quickly returns to the face when this does not happen.



The gaze did not necessarily always reach the place where the subtitles were supposed to start, but a distinct saccade towards that place (usually from the face area of the person on screen) was observed in many of the cases. Table 1, below, shows the number of subjects who demonstrated this tendency in each of the cases containing delayed subtitles. Note that the total number of participants is not equal in each case since gaze data was not retrieved from all of the participants in all of the cases. This has happened because the participant may have been looking away from the screen, the eyes may have been closed, or the head may have moved so that the cameras lost track of the participant's eyes.

Table 1: Number of participants in the test group who demonstrated seeking behaviour when the appearance of the subtitle was delayed.

| Manipulation Case | Delay Time (seconds) | Seeking Behaviour | No Seeking Behaviour |
|-------------------|----------------------|-------------------|----------------------|
| 1                 | 1.5                  | 0                 | 9                    |
| 2                 | 1.5                  | 8                 | 1                    |
| 3                 | 2.0                  | 2                 | 7                    |
| 4                 | 2.0                  | 3                 | 7                    |
| 5                 | 2.5                  | 3                 | 5                    |
| 6                 | 2.5                  | 6                 | 3                    |
| Total             |                      | 22                | 32                   |

It was expected that seeking behaviour would become more frequent as the delay time increased, but the data did not fully confirm this hypothesis. Instead, there were noticeable differences between cases that had the same delay time. In the cases with the shortest delay time, the difference was the clearest: in case 1, none of the nine participants sought the subtitles while in case 2 all but one demonstrated seeking behaviour. Similar inconsistency was visible in the cases with the longest delay time. Case 5 showed only three out of eight participants seeking the subtitles, and in case 6 the ratio was six out of nine. The fact that

these discrepancies can be found at both ends of the spectrum suggests that the delay time is not the most important factor in triggering seeking behaviour.

A qualitative analysis of the video clip gave a possible explanation for the discrepancies, at least in cases 1 and 2. Case 1 occurred during the first half of the clip and it was the first line of the fourth scene. The scene opens with the main character's secretary standing in the middle of an office room, trying to bounce a ping-pong ball attached to a paddle with a string. It is the first time that she appears in the episode, and she is dressed in a very original way in white tights and fluffy underwear. It is possible that the sight caught the viewers by surprise and thus drew their attention so that they did not even notice that the subtitle was delayed. Then again, case 2 is located in the middle of a heated argument between the main character and her daughter. The dialogue is quite fast-paced in the scene, and the viewer has to be attentive to each subtitle change in order to have enough time to read every subtitle. Then at one point the viewer sees with peripheral vision that the subtitle which s/he has read disappears. The viewer assumes that this means that, since the dialogue is still continuing, the next subtitle is going to appear promptly and is then caught by surprise when this does not happen.

Unfortunately cases 5 and 6 cannot be explained in a similar way. They both take place in similar contexts, in the middle of a dialogue, with similar close-up shots and no obvious attention grabbers in the scene. The dialogue is not very fast-paced in either of the cases. Cases 3 and 4 are also quite similar to cases 5 and 6, and they also yield the most consistent results. In fact, if the delay time is ignored, and the cases are examined solely from the point of view of contextual issues, cases 3, 4 and 5 give consistent results, while case 6 creates a discrepancy. The issue is clearly not simple, because when talking about subtitles the context includes multiple channels (audio verbal and non-verbal as well as the graphic channel and the surrounding subtitles), and on the basis of this analysis it is impossible to pin-point any single aspect. Nevertheless, cases 1 and 2 suggest that viewers adapt to the pace of the subtitles, and abruptly changing the pace causes confusion.

### *3.2. Lengthened Subtitles*

The manipulated subtitles included five cases where the length of the on-screen time of the subtitle was increased to two, three or four times that of the original. In order to make the fixation counts between the groups comparable, the time span of the analysis in the control group was the same as in the test group, which means that it exceeded the display time of the subtitles. Table 2, below, sums up the results.

Table 2: Mean fixation counts to subtitles in cases 7-11, in which the on-screen time of the subtitles was lengthened by the factor of 2, 3 or 4.

| Manipulation case  | 7     | 8     | 9     | 10    | 11    | All   |
|--------------------|-------|-------|-------|-------|-------|-------|
| Lengthening factor | 2     | 2     | 2     | 3     | 4     |       |
| Test group         | 4.80  | 5.10  | 5.14  | 3.80  | 7.20  | 5.21  |
| Control group      | 3.82  | 3.10  | 5.89  | 2.75  | 5.22  | 4.04  |
| p=                 | 0.273 | 0.027 | 0.634 | 0.082 | 0.077 | 0.015 |

The hypothesis was that lingering subtitles draw the viewers' attention again and again, which would result in increased fixation counts in the test group. The ANOVA analysis of variance of the fixation data confirmed the hypothesis. The test group made an average of 5.21 fixations on the subtitles in all five cases combined, while the control group's mean value was

4.04 fixations ( $p = 0.015$ ,  $F = 6.15$ ). Three of the cases had the same lengthening factor (the manipulated subtitles were visible twice the time of the original), and when these cases were investigated as a group, the difference between the test group and the control group was not as clear: the mean values were 5.0 fixations in the test group and 4.2 fixations in the control group ( $p = 0.207$ ,  $F = 1.63$ ). In case 10, in which the manipulated subtitle was on-screen three times the duration of the original, the mean values were 3.8 and 2.8 fixations ( $p = 0.082$ ,  $F = 3.35$ ), respectively, while in case 11 (four times longer duration in the manipulated subtitles) the test group's mean value was 7.2 fixations and the control group's was 5.2 fixations ( $p = 0.077$ ,  $F = 3.53$ ).

Although the data fails to show statistical significance, it should be noted that the mean values show a clear tendency to increase as the subtitles lengthen. This is not echoed by the statistical analysis, possibly because the data is rather limited: the manipulated subtitles included only five cases of lengthened subtitles, and in three of these the lengthening factor was two. Thus there was only one of each case in which the subtitle was lengthened to three or four times of the duration of the original. Furthermore, there are notable differences between cases 6-8 (see Table 2, above), which all have the same lengthening factor and are structurally quite similar, too. In case 8, the test group made noticeably more fixations to the subtitles than the control group (mean values 5.10 and 3.10 fixations, respectively,  $p = 0.027$ ,  $F = 5.83$ ), while in cases 7 and 9 there is no statistically significant difference.

### *3.3. Shortened Subtitles*

In cases where the duration of the subtitle was shortened to be completely in synchronization with the audio dialogue (Table 3, below), the differences in the fixation counts between the two groups were noticeable. The mean values for the test group and the control group were 4.23 fixations and 6.85 fixations, respectively, and the difference was statistically very significant ( $p < 0.01$ ). The average word length in these cases was five words, which suggests that the subjects in the test group did not have time to fixate on every word, which is customary in normal reading of Finnish (Hyönä 1996). A reasonable conclusion is that the test group's subjects did not have enough time to read the subtitle completely. This is confirmed by a gaze path analysis which shows that often in such cases the subtitle disappeared when the subject was in the middle of reading them. The fixation counts of each case also show that the test group's mean fixation count is less than the number of words in the corresponding subtitle (see Tables 3 and 4, below).

Table 3: Cases 12-16: Subtitles, spoken lines, and on-screen durations in the original and manipulated versions of the subtitles. The manipulated duration is in full synchronization with the dialogue.

| Manipulation case | Subtitle                                      | Spoken line                                                                       | Orig. duration (sec) | Manip. duration (sec) |
|-------------------|-----------------------------------------------|-----------------------------------------------------------------------------------|----------------------|-----------------------|
| 12                | Hitto. Perseen suti.<br>- Huomenta, Eddy.     | Damn, damn, bugger, bugger, bollocks.<br>- Good morning, Eddy.                    | 3.48                 | 2.17                  |
| 13                | Aina saisi olla somistamassa uudelleen.       | We'd have to keep replacing it.                                                   | 3.04                 | 1.13                  |
| 14                | Toimituksissa tutistaan, kun minä olen irti.  | I can tell you, some of them are shaking in their shoes knowing I'm on the loose. | 4.32                 | 2.71                  |
| 15                | Upeen itseni. Mitä sinä?                      | Me looking fabulous. What do you see?                                             | 3.04                 | 1.42                  |
| 16                | Miten pitkäksi aikaa?<br>- Ikuisiksi ajoiksi. | For how long?<br>- Forever.                                                       | 3.44                 | 1.67                  |

Some differences between the cases became obvious when they were examined individually. Table 4, below, shows the mean fixation counts for these cases in both groups. Single factor ANOVA analysis of variance shows that in each of the cases 13-15 the difference is very significant ( $p < 0.01$ ), while in case 12 the difference is of borderline significance ( $p = 0.056$ ,  $F = 4.30$ ). In each of these cases, the subjects in the test group made fewer fixations on the subtitles than the number of words contained in them. However, in case 16 the difference is highly insignificant ( $p = 0.485$ ,  $F = 0.51$ ), and this calls for a more careful inspection of the case.

Table 4: Mean fixation counts to subtitles in cases 12-16.

| Manipulation case | 12    | 13    | 14    | 15    | 16   | All   |
|-------------------|-------|-------|-------|-------|------|-------|
| Test group        | 5.75  | 3.44  | 5.36  | 2.91  | 3.89 | 4.23  |
| Control group     | 9.33  | 7.56  | 8.56  | 4.90  | 4.40 | 6.85  |
| p=                | 0.056 | <0,01 | <0.01 | <0.01 | 0.49 | <0.01 |

In case 16, the subtitle consisted of five words in two lines. The duration of the subtitle was shortened from 3.46 seconds to 1.67 seconds in the manipulation, which means that the test group had less than half of the time of the control group to read the subtitle. Despite this, the control group did not make noticeably more fixations on the subtitles: the mean values show that on average not every word of the subtitle was fixated on in either one of the groups. According to Rayner (1998:375), short functional words are often skipped in normal reading which means that they are processed without fixating on them. Case 16, though, contains only content words, and quite long ones at that (mean word length = 6.8 characters,  $SD = 1.6$ ), so Rayner's observation does not explain why some of the words in the case were skipped.

Case 12, in which the difference between the groups also failed to reach significance, is similar to case 16: it also consists of five content words divided into two lines, but here the time difference was not as great (2.17 seconds in the manipulated version versus 3.50 seconds in the original). It should be noted, though, that in case 12 the subjects of both groups made

more fixations on the subtitles than the number of words contained in them, and that the difference between the fixation counts is noticeable. Nevertheless, the statistical analysis is based on a rather small number of subjects and the results should be verified with a larger data set before drawing firm conclusions.

### 3.4. Questionnaire

Since the gaze data analysis showed clear differences in the gaze paths between the two groups, it was assumed that this would show up in the results of the questionnaire as well. To some degree this was the case, but the analysis also produced some surprises. Table 5, below, sums up the results of the questionnaire analysis.

Table 5: Subjects' answers to key questions of the questionnaire (percentages in brackets).

|               | No. of participants who noticed something odd in the subtitles | No. of participants who noticed any translation errors | No. of participants who had difficulties in following the programme at any point |
|---------------|----------------------------------------------------------------|--------------------------------------------------------|----------------------------------------------------------------------------------|
| Test group    | 8 / 11 (73 %)                                                  | 5 / 11 (45 %)                                          | 4 / 11 (36 %)                                                                    |
| Control group | 6 / 12 (50 %)                                                  | 0 / 12 (0 %)                                           | 4 / 12 (33 %)                                                                    |

In the test group, 8 out of 11 participants noticed something odd in the subtitles, and four of them could specify at least one translation error. All of these mentioned one specific error, namely a case in which the word “stoned” (‘in a state of drug-induced euphoria’, OED 2012) was mistranslated as “kivi” (‘a stone’). In addition to this, one of the subjects mentioned another error, a mistranslation of the term “platform shoes” as “lavakengät” (lava, ‘a platform, or stage’; kengät, ‘shoes’). Furthermore, one participant said that he noticed a translation error, but did not specify the case. This means that two out of the four translation errors in the test group’s stimulus were not identified by any of the group’s subjects.

In the control group, 6 out of 12 participants noticed something odd in the subtitles, but none could specify a translation error or other type of disturbance. The number was surprisingly high since the control group’s subtitles were composed according to the guidelines of Yle and did not contain any translation errors (apart from one mistyped name, which was noticed by the authors of the present study only in the analysis). In both groups, four participants wrote that they had problems in following the programme at some point. Because the same number of participants in both groups answered positively to the last question, it is reasonable to assume that the confusion might not have been caused by the subtitles alone but that instead there was something in the programme itself that made it difficult to follow.

The last question of the questionnaire asked the participants to evaluate the quality of the subtitles on a scale of 1 (the lowest) to 5 (the highest). The mean values of the evaluations were almost the same between the test group and the control group, the test group giving a slightly higher score (3.41, SD = 0.63) than the control group (3.23, SD = 0.58). This is very surprising since it means that the 31 cases of various subtitling errors did not contribute negatively to the subjects’ opinion about the quality of the subtitles, as would be reasonable to expect. Instead this echoes the results of some of the other questions: each group had roughly the same amount of people who had difficulties in following the programme and who noticed something odd in the subtitles. One reason why the subtitles that were shortened in duration did not result in more confused viewers in the test group is that there were only five such cases in a 25-minute episode and they were scattered evenly across the scenes. Probably

missing just one sentence in a conversation was not enough to make the viewer lose track of the point of the dialogue.

#### **4. Discussion**

The scan path data of cases in which the appearance of the subtitle was delayed showed that some viewers had a tendency to search the subtitles with their gaze when a line was spoken on screen, even though the subtitle did not appear. This can be seen as evidence in support of the idea first introduced by d'Ydewalle et al. (1987:320) that reading subtitles is at least partially an automated process. Then again, the data showed great differences between similar cases, which means that the issue is not as simple as that. In fact, there were notable differences between similar cases in all three types of synchronization manipulations. Clearly the context of the subtitles (dialogue, background sound effects and music, visual elements as well as surrounding subtitles) affects reading behaviour. For example, a qualitative gaze path analysis of cases 1 and 2 showed that when the dialogue was fast-paced, viewers adapted to this pace and thus almost all of the test groups' subjects demonstrated the seeking behaviour when one of the subtitles was delayed. Then again when the delayed subtitle was in the beginning of the scene, and there was an unusual visual element on the screen, none of the test group's subjects sought the subtitle before it appeared. It is evident that the effect of context should be studied more carefully in future research projects.

One aspect of the context that possibly affected the results of the present study, but was not taken into consideration in the planning of the test setting, was the subjects' language skills. All of the subjects were either students or staff members of the University of Eastern Finland, so it is reasonable to assume that they had better-than-average knowledge of English, the stimulus material's language. This means that they did not have to rely completely on the subtitles to follow the dialogue, which may have had unpredictable effects on the gaze paths. Future research should attempt to eliminate this problem either by using material that uses a less-familiar language in the dialogue or assessing the subjects' language skills with some sort of test.

The fixation counts in the cases of lengthened subtitle display time showed that lingering subtitles draw viewers' attention again and again as long as they are visible. One of the most important attributes of a good subtitle is that it is easy to understand, and this means that the font and colour of the text has to be chosen so that the subtitle stands out and is as clear as possible. Thus the subtitle is an unnatural element of the screen: it does not belong to the setting of the programme and in a way it is between the viewer and the setting. It is likely that this unnatural element draws the viewer's attention simply because it stands out and the effect is even stronger when the usual cue for the subtitle, spoken language from the audio channel, is not there. This attention-drawing aspect of subtitles could also explain the observation by d'Ydewalle et al. (1987:320) and d'Ydewalle and De Bruycker (2007:203) that people who understand perfectly the spoken language of the programme still often fixate on the subtitles when they appear, even when they do not understand the language of the subtitles.

Although the effect of the synchronization problems on the gaze behaviour was clearly visible, the analysis of the answers to the questionnaire did not show similar results. The manipulations did not have a significant effect on the answers, as both groups had the same amount of people who had trouble following the programme at some point. Since the questionnaire did not ask the subjects to elaborate on the reasons for the difficulties in understanding, one can only speculate about it, but again this brings the question of context to mind. The programme was rather old (first aired in 1995) and the humour is partly culture-

specific. Most of the participants were just over twenty years old, and they might have been simply too young to remember the cultural elements to which the show referred. Another reason why the control group noticed issues but could not identify them was that the questions themselves led the participants to think that there must have been some errors. Whatever the reason, the design of the questionnaire could have been improved by using more detailed questions. Furthermore, the length of the program probably caused the subjects to forget many of the issues they noticed, which resulted in vague answers (for example, “I noticed an error but can’t remember what it was”).

The questionnaire also showed that most of the manipulations went unnoticed even though they did affect the gaze paths. This means, that the disturbance in these cases was qualitatively or quantitatively not strong enough to cause conscious reactions and only affected unconscious reactions. It raises a question for future research: Is it possible to find a ‘border’ when the disturbing effect of de-synchronization (or another type of subtitling error) reaches viewers’ consciousness more often than not?

It is also curious that the manipulations, thirty-one cases in all, did not negatively affect subjects’ assessment of the quality of the subtitles. One possible reason for this is that the subjects simply did not pay conscious attention to the subtitles, and thus gave them an average grade just to be safe (3 was the most common grade given in both groups on the scale of 1 to 5). Then again, both groups found roughly the same amount of ‘odd things’ in the subtitles, and this correlates with the question about quality. Since the original subtitles were in accordance with the subtitling norms of Yle, this suggests that television viewers think about the quality of subtitles from a very different perspective than that of subtitling professionals, and thus the current study did not address the right sort of issues. Whatever the right answer is, it is clear that the notion of subtitling quality deserves a more detailed study, as it seems possible that it consists of something entirely different from what has been thought before.

One of the limitations of the study was that the universality of the findings can, and should, be questioned. The demographic range of the subjects was not very extensive as a vast majority of them were young university students. They probably had better than average language skills in English, the language of the programme’s dialogue, and young people in general are likely to read faster than the elderly or young children. This reveals one of the interesting directions that future research could take, echoing the proposal by Perego et al. (2010:266): the differences in subtitle reading and television viewing behaviour between different demographic groups. This might identify possible special needs or requirements of the groups and enable translators to modify subtitles according to the programmes’ assumed target audiences.

The test setting was designed so that it would enable the investigation of as many aspects of subtitles as possible and this meant that the manipulations included many different types of subtitling errors. This proved to be both a drawback and an advantage. On the one hand, it meant that the number of cases for each type of error was small, in many cases too small for reaching reliable conclusions and statistical significance. On the other hand, the data gave many pointers for further research. It showed that even a rather rough analysis of viewer eye movements can identify some of the problems that bad subtitle design causes, and thus gives hope that a more refined eye movement study can elaborate and clarify the phenomenon even further.

## Acknowledgements

The authors would like to thank Dr Roman Bednarik as well as the two anonymous reviewers for their suggestions and valuable comments. Preparation for this article was financially supported by grants from the C V Åkerlund Foundation and Emil Aaltonen Foundation, and Langnet, the Finnish doctoral programme in language studies.

Main Author's Address:

Juha Lång

juha.lang(a)uef.fi

Philosophical Faculty, School of Humanities

University of Eastern Finland

P.O. Box 111

FI-80101 Joensuu, Finland

## Research Material

BBC (1995) *Absolutely Fabulous*, season 3, episode 5, 'Fear'. Original air date May 4, 1995.  
Re-run aired in Finland December 28, 2008.

## References

- BBC (2012) BBC One on-line programme guide: *Absolutely Fabulous*.  
[<http://www.bbc.co.uk/programmes/b00jm3ms>] (accessed on 18 January 2012).
- Caffrey, Colm (2008a) 'Viewer perception of visual nonverbal cues in subtitled TV Anime', *European Journal of English Studies* 12(2): 163-178.
- Caffrey, Colm (2008b) 'Using pupillometric, fixation-based and subjective measures to measure the processing effort experienced when viewing subtitled TV anime with pop-up gloss', in Susanne Göpferich, Arnt Lykke Jakobsen and Inger M. Mees (eds.) *Looking at Eyes. Eye-Tracking Studies of Reading and Translation Processing*. Copenhagen: Samfundslitteratur, 125-144.
- d'Ydewalle, Géry and Wim De Bruycker (2007) 'Eye Movements of Children and Adults While Reading Television Subtitles', *European Psychologist* 12(3): 196-205.
- d'Ydewalle, Géry, Caroline Praet, Karl Verfaillie and Johan Van Rensberger (1991) 'Watching Subtitled Television: Automated Reading Behavior', *Communication Research* 18(5): 650-666.
- d'Ydewalle, Géry, Johan Van Rensbergen and Joris Pollet (1987) 'Reading A Message When The Same Message Is Available Auditorily In Another Language: The Case of Subtitling', in J. Kevin O'Regan and Ariane Lévy-Schoen (eds.) *Eye Movements: From Physiology to Cognition*, Amsterdam: North Holland, 313-321.
- Diaz-Cintas, Jorge (1999) 'Dubbing or Subtitling: The Eternal Dilemma', *Perspectives: Studies in Translatology* 7(1): 31-40.
- Diaz-Cintas, Jorge and Aline Remael (2006) *Audiovisual translation: subtitling*, Manchester: St. Jerome.
- Göpferich, Susanne, Arnt Lykke Jakobsen and Inger M. Mees (eds.) (2008) *Looking at eyes: eye-tracking studies of reading and translation processing*, Copenhagen: Samfundslitteratur.
- Hyönä, Jukka (1996) 'Silmänliikkeet lukemisessa [Eye movements in reading]' in Revonsuo, Antti, Heikki Lang and Olli Aaltonen (eds.) *Mieli ja aivot: Kognitiivinen neurotiede*

- [*Mind and brain: Cognitive neuroscience*], Turku: Centre for Cognitive Neuroscience, University of Turku, 283-292.
- Koolstra, Cees M., Allerd L. Peeters and Herman Spinhof (2002) 'The Pros and Cons of Dubbing and Subtitling' *European Journal of Communication* 17(3): 325-354.
- O'Brien, Sharon (2006) 'Eye Tracking and Translation Memory Matches', *Perspectives - Studies In Translatology*, 14(3): 185-205.
- O'Brien, Sharon (2008) 'Processing fuzzy matches in Translation Memory tools: an eye-tracking analysis', in Susanne Göpferich, Arnt Lykke Jakobsen and Inger M. Mees (eds.) *Looking at Eyes. Eye-Tracking Studies of Reading and Translation Processing*. Copenhagen: Samfundslitteratur, 79-102.
- OED (2012) Oxford English Dictionary. Available online at [www.oed.com] (accessed on 18 January 2012)
- Perego, Elisa, Fabio Del Missier, Marco Porta and Mauro Mosconi (2010) 'The Cognitive Effectiveness of Subtitle Processing', *Media Psychology* 13(3): 243-272.
- Rayner, Keith (1998) 'Eye Movements in Reading and Information Processing: 20 Years of Research', *Psychological Bulletin* 124(3): 372-422.
- Sharmin, Selina, Oleg Špakov, Kari-Jouko Räihä and Arnt Lykke Jakobsen (2008) 'Where on the screen do translation students look while translating, and for how long?', in Susanne Göpferich, Arnt Lykke Jakobsen and Inger M. Mees (eds.) *Looking at Eyes. Eye-Tracking Studies of Reading and Translation Processing*. Copenhagen: Samfundslitteratur, 31-51.
- Tosi, Virgilio, Luciano Mecacci and Elio Pasquali (1997) 'Scanning Eye Movements Made When Viewing Film: Preliminary Observations', *Intern. J. Neuroscience* 92(1-2): 47-52.
- Vertanen, Esko (2007) 'Ruututeksti tiedon ja tunteen tulkkina [Subtitles as conveyors of knowledge and emotion]' in Oittinen, Riitta and Tiina Tuominen (eds.) *Olellaisen äärellä - Johdatus audiovisuaaliseen kääntämiseen [Introduction to audiovisual translation]*, Tampere: Tampere University Press, 149-170.