Introduction

The object of study in research on audiovisual translation (AVT), the audiovisual text, is a complex polysemiotic text where information is presented simultaneously in a number of semiotic codes (i.e. verbal and non-verbal, visual and auditory codes). The processing of this text involves a variety of cognitive processes. It is therefore understandable that research in this field is very much concerned with the reception of the audiovisual text by various audiences with different needs.

The primary (but not exclusive) goal of all subfields within AVT is to provide or improve access to audiovisual texts to audiences who are either excluded from one or more of the auditory or visual codes, or who only have partial access to these codes. These audiences and their needs are by no means always easy to define. Let us take the example of audio description (AD). Blind audiences only have access to the soundtrack containing the auditory codes, and have to be given access to those elements in the image relevant to their understanding and enjoyment of the text. However, in considering what is relevant in the visual code, and how to present this to the audience, the audio describer has to consider the fact that the audience is made up of individuals with a range of needs. Some have been blind from birth (congenital blindness), whereas others may have lost their sight later in life, significantly impacting on their frames of reference. Congenital blindness, for example, means that an individual would have no memory of sight or colours, whereas, depending on the age of the onset of blindness, other blind individuals may have some memories to draw on in interpreting a scene. Furthermore, significant proportions of the user groups of AD are not totally blind, and have varying degrees of sight.

In the case of subtitling, the user groups who rely on or can benefit from the subtitles range from Deaf viewers (again spanning viewers who lost their hearing at different ages and who have varying levels of reading skills), to hard of hearing viewers (with varying degrees of hearing loss), to viewers who do not understand the language of the original text (in varying degrees), to viewers who have none of these barriers, but who could benefit from same-language subtitles to improve literacy or comprehension. Dubbing is the only form of AVT that really only benefits one group, namely those who cannot understand the language in the original text.
When we are therefore interested in the processing of audiovisual texts, we have to understand how the texts create meaning, and how they are processed by different audiences. The focus of this chapter will be on the processing of the visual codes, in other words both verbal (e.g. text on screen) and non-verbal (e.g. camera angles, editing, objects, etc.) visual elements in these texts, with a particular focus on how an investigation of eye movements can yield valuable information that is relevant to most fields within AVT. For example, in subtitling, the viewer has to divide his or her attention between the subtitles and the rest of the screen, and the meaning is created through a combination of information from the subtitles, the soundtrack and the image (or only the subtitles and the image in the case of Deaf viewers). The main emphasis will be on eye movements during the processing of subtitles, but some reference will also be made to the use of eye tracking for AD.

Eye tracking, or the study of eye movements, is a well-developed research methodology in reading research, psycholinguistic research, translation process research, and also in the field of AVT. In the case of AVT, eye-tracking research became established in the context of subtitle processing even before it had gained prominence within translation studies in the late 1990s.

**Eye tracking**

According to Findlay (1985: 101–102), ‘[v]ision is the primary human sense modality. The quantity of information passed along the optic nerve is far greater than that in any other pathway of special sense, and a large percentage of the posterior part of the cerebral cortex is primarily concerned with the analysis of this message’. Given the importance of this modality, it is understandable that researchers have been interested in gaining insights into the human cognitive system, by looking at how people attend to scenes and texts, for more than a century. In the past half a century, eye tracking has become an important experimental method in psycholinguistic research in the study of reading (cf. Keating 2014: 69, Rayner 1998), and it is particularly in this research that we find some of the most useful information and methodologies for understanding the processing of subtitles.

Eye tracking, at a basic level, concerns the study of eye movements. In particular, the eye movements that can be detected by an eye tracker provide us with a window on the internal systems of the mind (cf. Marchant et al. 2009). But applying eye tracking to dynamic scenes such as film is not without problems. On the one hand, the objects being fixated are constantly moving (cf. Papenmeier and Huff 2010). On the other hand, any text that appears in film (e.g. subtitles), is on screen for a limited period of time, forcing the reader to adopt reading strategies that differ slightly from those in the reading of static text where the reader is much more in control of the pace of reading (cf. Kruger and Steyn 2013).

Information is taken in by the eye in a process that allows light in through the pupil and projects it (upside down) onto the retina (the back of the eyeball). Importantly, we only have full visual acuity in a very small area in our field of vision (the fovea, which comprises less than 2 degrees of the visual field), which means that we constantly have to change the position of our eyes to focus on (or foveate) specific elements of the visual scene (for more details, see Holmqvist et al. 2011: 21–24, Liversedge et al. 2013). Eye movements consist of short periods of relative inactivity (fixations), followed by rapid movements (saccades) to the next point of fixation. During fixations, visuospatial processing occurs—the eyes collect and take in information. During saccades, however, the eyes do not take in any information, although cognitive processing continues in the brain (see Irwin 2004, Holmqvist et al. 2011).
The typical duration of a fixation is between 200 and 300 milliseconds (ms), and that of a saccade between 30 and 80ms (Holmqvist et al. 2011: 23), although there is much more variance than these averages suggest. Rayner (2009), for example, reports differences in fixation duration during different visual activities: the mean fixation duration during reading ranges from 225–250ms during silent reading, to 275–325ms during reading aloud, 260–330ms when looking at a scene (scene perception), and 180–275ms when looking for something specific in a scene (visual search). According to Holmqvist et al. (2011: 382), the difference in fixation duration during different tasks indicates ‘functional links between what is fixated and cognitive processing of that item—the longer the fixation, the ‘deeper’ the processing’. More specifically, the more complicated the task, the longer the fixations become. Likewise, the number of fixations in a particular area tends to increase as the task becomes more complicated (cf. Holmqvist et al. 2011). But even these qualities of fixations have to be approached with caution as different activities require different visual behaviour. Reading, for example, elicits specific patterns of fixations and saccades in a particular order that differ substantially from the more random patterns associated with scene perception and visual search. It is therefore important not to generalize across activities.

By studying the eye movements (including number of fixations and duration of fixations) of viewers watching any form of audiovisual media, researchers can therefore gain an understanding of, or at least glimpses into, the workings of the mind of the audience. However, there has been growing evidence that these two processes (fixating and processing cognitively) are not fully synchronous—often our visual attention precedes the position of our eyes by a fraction of a second, meaning that the eye does not provide a direct window onto our cognitive processes (cf. Irwin 2004). To overcome the fact that knowing where and how viewers look does not always provide direct evidence of their cognitive processing, eye tracking is increasingly used in combination with other data such as comprehension measures, physiological measures like heart rate, retrospective protocols, and even electroencephalography (EEG), which involves the measurement of electrical activity in the brain by means of contacts on the scalp (see, for example, Kruger and Doherty 2016). Together these data are providing us with a picture of human cognitive processing, also in the field of AVT research.

Eye tracking remains a reliable source of information on what viewers look at, and more importantly, how they look at it (how long, in what patterns, with what pupil dilation—all providing pieces to the puzzle of cognitive processing and cognitive load). This chapter will therefore attempt to provide a basic introduction to the principles that underlie this type of research. In particular, the aims will be to give an overview of important trends in studies using eye tracking in the context of AVT, and also to identify the areas that will benefit most from the use of this technology. In striving to reach these aims, the chapter will also discuss the limitations of eye tracking and identify typical pitfalls in the use of eye tracking in terms of scientific rigour related to experimental design, sampling, and statistical analysis. It is just as important to know when to use eye tracking as it is to know when not to use it.

**Overview of the use of eye tracking in audiovisual translation research**

In the field of AVT, eye tracking gained prominence with the substantial body of work generated in Belgium by Gery d’Ydewalle and colleagues from the 1980s (see, for example, d’Ydewalle et al. 1985, d’Ydewalle et al. 1987, d’Ydewalle and Van Rensbergen 1989, d’Ydewalle et al. 1991, d’Ydewalle and Gielen 1992, d’Ydewalle and Van de Poel 1999, d’Ydewalle and De Bruycker 2007). Since 2010 a number of new studies have

A number of studies also used eye tracking to investigate the processing of audiovisual texts by sighted viewers in order to inform the field of AD (e.g. Igareda and Maiche 2009, Kruger 2012, Orero and Vilaró 2012, Krejtz et al. 2012, Di Giovanni 2014).

**The early studies: attention distribution**

The bulk of the studies involving eye tracking and AVT in the 1980s and 1990s were conducted in Belgium by Gery d’Ydewalle and colleagues. In the first studies, the main interest was on attention allocation. Findings included that participants fixate only one or two words per subtitle (d’Ydewalle et al. 1985), that participants look at the subtitles for about 30 per cent of the presentation time (d’Ydewalle et al. 1987), and that genre has an impact on the amount of time spent looking at subtitles with children spending less time looking at subtitles when watching action film than animation (d’Ydewalle and Van Rensbergen 1989). In the latter study, the assumption that reading behaviour is elicited automatically when text appears on screen is introduced, an assumption that is tested and confirmed in later studies. For example, in d’Ydewalle et al. (1991) the authors found that participants still read redundant subtitles (L1 subtitles on a film with an L1 soundtrack) 20 per cent of the time.

Slightly later, Jensema and colleagues conducted a rather small-scale research project on the eye movement strategies used by six Deaf viewers when watching captioned television (Jensema et al. 2000b), as well as a slightly larger study on the time spent by twenty-three Deaf viewers viewing captions at different speeds (Jensema et al. 2000a). Strangely, neither of these studies took any cognisance of the work done in Belgium by d’Ydewalle and colleagues, and both studies are severely limited in terms of the number of participants and the duration of the clips studied. However, since the studies engage with the reception of the users of subtitles it is perhaps useful to note the conclusions. In the first study (2000b), the findings include that people who view a particular video segment have similar eye movements, that adding captions turns the viewing process into more of a reading process (i.e. little time is spent looking at the image), and that higher captioning speed results in an increase in the time reading the captions. In the second study, they find that viewers looked at the captions 84 per cent of the time, at the picture 14 per cent of the time, and outside of the picture 2 per cent of the time. The very high percentage of time spent on the captions in these studies could very well be ascribed to the experimental design where Deaf viewers would have to try to process extremely short video clips without any other contextual or co-textual information.

As the technology and software for analyzing eye movement data improves, the amount of data obtained from eye tracking and the complexity of the research questions have increased. In more recent studies, the attention of researchers has shifted to matters related to the content of subtitles and away from a mere descriptive investigation of viewing behaviour in the
presence of subtitles. Furthermore, most studies in the past decade in this area have turned to the use of inferential statistics and away from mere descriptive statistics to make sense of the eye-tracking data.

**Subtitles, translation and language**

The 2007 study by d’Ydewalle and De Bruycker takes eye-tracking research in AVT to the next level. Again confirming the automatic processing of subtitles, the authors find that ‘even when the task does not require linguistic processing of text [e.g. in reversed subtitling where the L1 audio is subtitled into a foreign language], looking at words is almost obligatory’ (2007: 203). However, reversed subtitles were skipped more often than standard subtitles (21 per cent vs. 4 per cent), were fixated less (0.59 fixations per word vs. 0.91 fixations per word), and elicited less gaze time (26 per cent vs. 41 per cent of presentation time), indicating significant differences in processing. Furthermore, two-line subtitles elicited more regular reading than one-line subtitles, were skipped less often, and elicited more fixations and fewer regressions (when the viewer makes a saccade back to an earlier location against the direction of reading).

Bisson *et al.* (2014) conducted a similar study, looking at the difference in the processing of standard subtitles (from a foreign language into the first language of the viewer), reversed subtitles (from the first language into a foreign language), and intralingual, or same-language, subtitles. As in d’Ydewalle and De Bruycker’s (2007) study, they found that participants read the subtitles regardless of the subtitle condition, but that the participants displayed more regular reading behaviour in the standard condition when they did not understand the language of the soundtrack and had to rely on the subtitles. Bisson *et al.* (2014) found no statistically significant differences between the standard and intralingual conditions in terms of fixation duration, number of fixations, number of skipped subtitles, or number of consecutive fixations in the subtitle area, although the intralingual condition did elicit more attention to the image than the standard condition.

It is perhaps surprising that not many eye-tracking studies have focused on more micro-level processing of the linguistic features of subtitles. Notable exceptions are two studies in Perego’s collection on eye tracking in AVT, namely the study by Ghia (2012) who investigated the impact of translation strategies on subtitle reading, and that by Moran (2012) on the effect of linguistic variation on the reception of subtitles. Ghia (2012) looks at the reading behaviour of Italian intermediate learners of English when watching a clip subtitled from English into Italian using either a literal or a non-literal translation strategy. In this study, the main focus was on attention dynamics when viewers look at subtitled film—a particularly complex issue due to the constantly changing image and the resultant changes in the importance of different sources of information. Ghia makes the obvious but long overdue link between eye-tracking studies of reading in the context of static texts, and the reading of subtitles. She calls attention to psycholinguistic studies on reading that have shown that ‘saccades and fixations are affected by a series of both linguistic and contextual factors’, including ‘linguistic variables such as word class, length, and morphological complexity’, and ‘semantic-pragmatic features of texts’ (Ghia 2012: 158). In order to investigate the impact of translation strategies on subtitle reading she considers regressions, and deflections (i.e. shifts between the subtitle and the image vertically). In terms of regressions, she did not find any significant difference between the two translation strategies, although linguistic features such as word length and frequency did result in increased regressions as is the case in static reading. However, she did find significantly more deflections in the presence of
non-literal translation. In other words, non-literal language caused the readers to switch their visual attention more often between the subtitle text and the image, indicating a degree of confusion. On this basis she concludes that subtitle reading is indeed affected by translation and translation-specific features (2012: 177).

Moran’s study (2012) focuses on the effects of word frequency and lexical cohesion on the cognitive effort required by the viewer to process the text as a whole. By showing that variables of low frequency (i.e. words that the typical reader would encounter less frequently) and low cohesion within the subtitle (in spite of co-occurring with shorter subtitle length) result in higher processing effort, she therefore finds a contradiction of the subtitling norm of condensing the dialogue and reducing the characters during subtitling. Cohesion in particular is an important consideration due to the fact that during subtitle reading, unlike during static reading, the viewer cannot refer back to earlier parts of the text to establish cohesion. Her measures included mean fixation duration as an indication of cognitive effort (finding longer mean fixation durations on subtitles in the presence of low-frequency words as well as low-coherence subtitles). Although this study was limited in terms of the number of participants, it explains the limitations of word frequency and cohesion thoroughly, and identifies global trends that can be investigated in more detail as methodologies are refined to allow for the investigation of individual words or linguistic features in a more robust manner across large numbers of subtitles without having to revert to manual inspection.

**Audience and language**

Winke *et al.* (2013) turn their attention to the impact of different relations between the language of the film and that of the audience on subtitle viewing, particularly in the context of language learning. Their study takes eye-tracking research in the context of static reading as the starting point for a discussion of the reading of different languages, building on the small number of eye-tracking studies that have investigated foreign language processing (*cf.* Winke *et al.* 2013: 257–258). Their study in particular responds to d’Ydewalle and De Bruycker’s (2007) study and that of Bisson *et al.* (2014), with the difference that they are less interested in the difference between standard (FL audio with L1 subtitles), reversed (L1 audio with FL subtitles) and intralingual (L1 audio with L1 subtitles) subtitling, and more in the difference between the processing of subtitles in different L2 languages by studying the eye movements of learners of these languages. In addition to finding a high percentage of time spent reading subtitles (68 per cent), they find significant differences between the time learners of Arabic spent reading subtitles and that spent by learners of Russian or Spanish. Learners of Chinese spent less time on the subtitles when the content was familiar. On the whole, learners of Russian and Spanish and learners of Arabic and Chinese displayed different reading behaviour, indicating the influence of the distance between the L1 and the L2 on reading, with the time spent processing subtitles increasing as the distance between L1 and L2 increases. Limitations of the study acknowledged by the authors include the fact that the time it takes to acquire languages further from the L1 (e.g. English students learning Chinese and Arabic in this study), inevitably means that the levels of language proficiency were rather different. Also, the different alphabets and characters make it harder to make comparisons between the languages. Unfortunately, the study does not report on fixation durations, as that may have yielded more useful information, and furthermore, the fact that saccades are not considered in spite of the authors identifying the fact that saccadic movements differ between the reading of these languages, also leaves a number of unanswered questions.
Subtitle presentation rate

The speed at which subtitles are presented (normally measured in words per minute or wpm, or in characters per second or cps) has been the subject of a few studies, not least because this has significant implications for the amount of reduction that has to be done to allow sufficient display time. Although Jensen et al. (2000b) found only a marginal difference in the amount of time allocated by Deaf viewers to the subtitle area when subtitles were presented at a rate of 100wpm (82 per cent) and when subtitles were presented at 180wpm (86 per cent) this remains an area that warrants more attention.

In their study, Szarkowska et al. (2011) found that Deaf viewers spent just under 70 per cent of the time reading verbatim subtitles, which drops to around 60 per cent for standard subtitles and 50 per cent for edited subtitles (a much higher range in low to high presentation speed than in the earlier study). They found a similar range in the hard of hearing group from around 60 per cent in the verbatim condition to under 50 per cent in the edited condition, but very little difference between the three conditions for the hearing audience. In terms of the dwell time on the subtitles, they find that the edited subtitles were relatively the easiest to comprehend for all groups with the least dwell time, but that the deaf group had significantly longer dwell time on the verbatim subtitles than hearing participants (but no longer than hard of hearing participants). For all three groups the verbatim subtitles elicited significantly more fixations. Interestingly, they found that Deaf viewers displayed different reading behaviour with more fixations and fewer deflections between subtitle and image than the other groups.

In a later study (Szarkowska et al. 2016), they again investigated the difference between verbatim and edited subtitles, but this time they found no significant difference within groups between verbatim and edited subtitles, although the difference between groups on verbatim and edited subtitles remained. In this study they only offered subtitles at two presentation rates, namely 12cps and 15cps (or around 140wpm and 180wpm respectively), and found no significant differences within any of the three groups (deaf, hard of hearing and hearing) between the two rates in dwell time.

In spite of the apparent inconsistencies in the findings, which should signal that more studies are required to establish the impact of presentation rate more comprehensively, broadcasters sometimes abuse preliminary findings—indicating that verbatim subtitles do not pose any additional cognitive effort and can be processed effectively. What we have to bear in mind is that film is a polysemiotic text, and even if viewers can read the subtitles at a higher rate, this does not mean that they have sufficient time also to explore the primary visual information in the image.

Romero-Fresco (2015) specifically refers to the notion of viewing speed, as opposed to presentation speed or reading speed, which also takes into account the time for processing the other semiotic signs in the audiovisual material. He based his findings on an extensive eye-tracking experiment and found that at a speed of 120wpm, viewers spent approximately 60 per cent of the viewing time on the image and 40 per cent on the subtitle. At 150wpm viewers divided their attention equally between image and subtitles. When the speed was increased to 180wpm, viewers started spending less time on the image (between 30 per cent and 40 per cent), and more on the subtitles (between 60 per cent and 70 per cent). At 200wpm, viewers only spent 20 per cent of their viewing time looking at the image.

Subtitle rules and cognitive processing

Other than presentation rate, the rules, conventions and guidelines for subtitling can also be interrogated by using eye tracking to study viewers’ eye movements. One such study was
Eye tracking in AVT research

done by Perego et al. (2010), who investigated, among other things, the impact of line segmentation on eye movements and comprehension. They compared the fixation count, total fixation time and mean fixation duration as well as the number of shifts between the subtitle area and the image when participants viewed well-segmented subtitles to when they viewed ill-segmented subtitles. Although they found no differences between the two conditions in terms of fixation count, total fixation duration and shifts, they did find that participants’ fixations were statistically significantly longer in the ill-segmented condition. Although the authors do not consider this difference to be particularly important due to the small magnitude of the difference (around 12ms), the impact of poor line segmentation does therefore induce significantly higher cognitive load. A limitation of the study is that a very small number of subtitles was investigated (28 in total), and that the line segmentations were manipulated only within subtitles (i.e. the subtitles still largely contained sense units, and did not split sense units between successive subtitles). The study primarily investigated the cognitive effectiveness of subtitle processing by testing not only comprehension, but also the processing of visuals in the presence of subtitles, and concludes that the cognitive processing of subtitles is effective.

Investigating different types of line segmentation or text chunking produced in live subtitling created through respeaking (i.e. using speech recognition software), Rajendran et al. (2013) compared no segmentation (where words keep appearing until the subtitle area is full before the next subtitle starts the same process), chunking by phrase (phrases appearing one by one on a single line), word-by-word presentation, and chunking by sentence. Looking at gaze points (using more lenient algorithms than conventionally used to define fixations) as well as ‘saccadic crossovers’ or revisits, they found that the word-by-word condition elicited more gaze points and also resulted in more vertical shifts to the image than the other conditions. The limitations of this study include the very short video clips used as well as the fact that the subtitles were displayed without sound to hearing participants.

Krejtz et al. (2013) investigated another subtitling rule, namely that subtitles should not remain on screen over a shot change, as it will induce re-reading of the subtitle which will be perceived as having changed with the image. Although they find that most viewers do not re-read the subtitles after the shot change, they did establish that about 3 per cent of viewers did re-read the subtitles, and about 30 per cent of viewers returned their gaze to the beginning of the subtitle after the shot change even if they soon realized that it was the same text and then resumed their reading. They also found that shifts from the subtitle to the image were significantly higher after a shot change than before, which seems to suggest that the reading process was interrupted, even if it did not result in re-reading. Another indication of the increased cognitive cost associated with subtitles that stay on screen over shot changes, is the fact that the first fixation duration after a shot change for those who returned to the beginning of the subtitle was significantly longer.

Breaking the rules

But of course not all subtitles adhere to the norms and conventions. The rise of fansubbing which employs many experimental techniques and routinely break with conventions, as well as other examples of films that test the boundaries of conventions has resulted in attention to such deviations from scholars using eye tracking. A handful of eye-tracking studies investigate viewing behaviour in the presence of non-standard subtitles. Caffrey (2012), for example, draws on the dependent variables used by d’Ydewalle and De Bruycker (2007), namely percentage skipped subtitles (i.e. subtitles that were not looked at), percentage gaze
time in subtitles (i.e. the time viewers looked at the subtitles as a percentage of the time the subtitle was displayed), mean fixation duration, and word fixation probability (the probability that any individual word would be fixated at least once, calculated by dividing the number of fixations in the subtitle by the number of words in the subtitle), and adds median pupil size for pupillometric data. Pupil size has been established as a measure of cognitive effort, with larger pupil size indicating more effort. His study focuses on the impact of pop-up glosses used in the fansubbing of Japanese anime subtitled into English. His findings include that the presence of a pop-up gloss resulted in a higher number of skipped subtitles, and a lower percentage gaze time in the subtitle area, lower mean fixation duration, and lower word fixation probability. Similar to d’Ydewalle and De Bruycker’s findings in 2007, Caffrey finds that one-line subtitles are significantly more likely to be skipped in the presence of pop-up glosses. His use of pupillometric data yields findings that suggest higher cognitive effort when both two-line subtitles and pop-up glosses are on screen, although this measure is notoriously unreliable in reading as well as in texts where the luminosity varies since the pupil may respond to the intensity of the light by contracting, which is not related to cognitive effort.

Fox (2016) investigates the processing of integrated titles, a form of subtitling where the titles are placed on the screen in positions that would require less effort from the viewer to process, such as close to where their eyes would naturally be such as the current speaker, and also integrating the titles more aesthetically with the image. In her study she addresses the questions whether individual placement and design of (sub)titles have an impact on reading time, time on the image, saccade length between titles and focal points, and overall viewing experience. Her study also uses eye tracking to identify focal points in a film in order to locate the optimal position for title placement, thereby adding another process element to the use of eye tracking. She finds statistically significantly shorter reading time for the integrated titles when compared to the reading time for traditional subtitles, and also statistically significantly more time for exploring the image in the presence of integrated titles. The study provides empirical support for an innovative form of integrated titles where the titles are placed in such a way that they complement the image, and minimize the effort required of the viewer.

Eye tracking and audio description

Eye tracking can also yield valuable information in the field of AD. The main challenge in AD is to understand how sighted viewers make sense of what they see—what they look at on screen in what sequence and in what manner, and how that relates to their interpretation of the film, to allow for better AD. Vilaró et al. (2012), for example, investigate the impact of sound on visual attention distribution, finding that the manipulation of the soundtrack influences where people look, and therefore also presents an important source of information to be taken into account in AD.

Likewise, Kruger (2012) combines the use of eye tracking with an analysis of viewer constructions of the narrative to provide information that could be beneficial in determining what should be included in AD. Using dwell time, fixation count and glance counts to specific areas of interest, and comprehension scores, he found a positive correlation between higher comprehension scores and glances to specific areas that were central to an understanding of the narrative, but were not very visually salient. In particular, this study points to the importance of top-down cognitive processes where viewers direct their eyes at particular areas on the screen.
to get narrative information, as opposed to bottom-up cognitive processes where the eyes are drawn to areas on the screen because of factors external to the viewer. Orero and Vilaró (2012), in a similar vein, use eye tracking to investigate whether viewers attend to minor details, and whether these details should therefore be audio described. They find that all participants look at the same regions in the same order, but their results concerning the importance of minor details are inconclusive. Taken together with the study by Kruger (2012), the study could, however, be seen as adding evidence that minor details only become important to AD when they are important to the central narrative of the film. Again, like the previous two studies, Di Giovanni (2014) uses eye tracking to identify visual priorities of sighted individuals, but then goes further to use these to inform the audio description. When comparing the visual-priority-derived AD to conventional AD, she finds significant benefits.

In the context of AD, eye-tracking research to date tended to use predominantly qualitative data obtained from heat maps although some of the studies mentioned also use quantitative data. The complexity of the visual information presented in film, however, makes eye-tracking data extremely difficult to interpret if the goal is to obtain information on the cognitive processing of the image. It is very tempting to use heat maps (visualizing where the majority of fixations were made by using shades of red and orange and yellow, providing very convincing red blots on the screen where most viewers attended) or focus maps (that do the same except that they blur those areas of the screen that did not receive much attention and bring into focus areas that received more attention). However, these maps can be misleading as they tend to be an accumulation of eye-tracking data across a number of frames visualized on the last frame. This is confirmed by the fact that studies comparing eye-tracking data to comprehension data often produce inconclusive results, signalling that current conventions in AD to focus on the description of visually salient elements may need to be revisited.

What further complicates the use of eye-tracking data for AD is the fact that there are many features in film that elicit visual attention automatically (i.e. bottom-up visual attention that is not under the cognitive control of the viewer), such as contrast, movement, faces (specifically eyes and mouths in close-up shots), and text on screen. In other words, viewers will be looking at these features involuntarily, making it very hard to distinguish between automatic viewing behaviour and viewing behaviour that is under the cognitive control of the viewer (or top-down attention). Nevertheless, eye tracking can give many important insights into the processing of audiovisual texts, insights that will doubtlessly be sought increasingly with this methodology in years to come.

**Eye-tracking measures used in audiovisual translation research**

As should be evident from the discussion above, different eye-tracking studies utilize slightly different measures that will be summarized in this section. As this methodology becomes more commonplace in the field of AVT research, it is important to start using standardized methodologies and terminology. Most eye-tracking measures relate to fixations, although there are a few measures that also include saccades. Since qualitative measures obtained through the use of heat maps and focus maps really only provide the starting point for more rigorous quantitative work, they will not be discussed here. In the next paragraphs, I will discuss some of the most common eye-tracking measures that have been used in studying the processing of subtitles beyond the mere distribution of attention between the subtitle area and the image area.
Jan-Louis Kruger

The valuable contributions made by d’Ydewalle and colleagues provided the first attempts at investigating the way in which subtitles are processed by viewers. The 2007 study discussed above identified some of the most important measurements in eye-tracking research in this field that remain the mainstay of this methodology for investigating eye movement in the presence of subtitles. In particular, d’Ydewalle and De Bruycker identify the following variables:

- percentage of skipped subtitles (i.e. the percentage of subtitles that did not receive any fixations or were not looked at by a viewer);
- latency time (i.e. the time between the appearance of the subtitle and the first fixation in the subtitle);
- percentage time in subtitles (also called percentage dwell time including fixations and saccades, i.e. the total time spent in the subtitle as a percentage of the time the subtitle was displayed—also called percentage dwell time in other studies);
- word fixation probability (i.e. the probability that a word would be fixated at least once);
- saccade amplitude in degrees and characters (also called saccade length, i.e. the distance between two consecutive fixations);
- percentage regressive eye movements (i.e. the percentage of times the saccades between consecutive fixations went against the direction of reading namely from right to left);
- number of back and forth shifts (also called revisits to the subtitle area, and referring to the number of times the participant shifted their gaze back and forth between subtitle and image).

D’Ydewalle and De Bruycker point out that they could not compute a measure of gaze for the simple reason that ‘eye-movement data could not be mapped onto the individual words, but only onto the subtitle area as a whole’ (2007: 199). This drawback means that it is not possible to calculate reading statistics when investigating subtitles because the subtitle text is part of the image rather than text that can be chunked automatically into individual words. In the study of static reading, by contrast, software has been developed to chunk words automatically into individual areas of interest and then to calculate various reading statistics. The fact that this software cannot be used for subtitles, means that eye-tracking research on subtitle processing has traditionally been a particularly time-consuming task, or could not really move beyond the study of the subtitle area as opposed to the words in the subtitle.

In her doctoral thesis, Specker (2008) investigates subtitle reading by looking at fixation counts, mean fixation duration and successive fixations before inspecting the fixation plots for individual subtitles. This methodology yields very rich data, taking into account the multimodal nature of the text, but still does not give a useful method for investigating the reading of a large number of subtitles as it relies on the manual inspection of scan fixation plots or scan paths. Ghia (2012) looks at regressions in the subtitle area and deflections to the subtitle area to study subtitle reading, whereas Moran (2012) focuses primarily on mean fixation duration, again not really relating the eye movements to the reading process.

Bisson et al. (2014) investigate the processing of subtitles in different conditions by looking at the following eye-tracking measures in the subtitle area:

- total fixation duration (i.e. the sum of the duration of all the fixations in the subtitle area);
- number of fixations (i.e. the sum of all the individual fixations in the subtitle area);
• average fixation duration (i.e. the total fixation duration divided by the number of fixations);
• number of skipped subtitles (as explained above);
• proportion of consecutive fixations (i.e. the percentage of fixations that were preceded by another fixation in the subtitle area).

The last measure in particular is intended to be a measure of reading behaviour based on the assumption that a higher proportion of consecutive fixations would indicate a higher degree of (continuous) reading, whereas the word fixation probability of d’Ydewalle and De Bruycker (2007) rather measures the probability that all the words in the subtitle were fixated. Although both these measures provide a useful quantification of reading behaviour, they lack nuance for investigating the actual reading of specific subtitles.

Winke et al. (2013) use a variation on the percentage time in subtitles (used by d’Ydewalle and De Bruycker 2007 and d’Ydewalle et al. 1991) or percentage dwell time (used by Kruger and Steyn 2014), namely the total fixation duration as percentage of the time the subtitle was displayed. Although they say that this measure is used in order to compare their results with that of d’Ydewalle and colleagues rather than the percentage of number of fixations on the areas of interest used in other studies, they do not, like the three studies mentioned above, include saccade duration, somewhat undermining the comparison. They also look at mean fixation duration as many of the other studies discussed here.

This brings us to the 2014 study by Kruger and Steyn which introduces a reading index for dynamic texts (RIDT). This measure makes it possible to determine the extent to which a subtitle was read by presenting an index of the reading that takes place in individual subtitles based on the following:

• Number of unique fixations per standard word. Unique fixations exclude those fixations that are so close to previous fixations that they could not be considered to add to the processing but are rather refixations, as well as excluding fixations following regressions where the eyes moved back to a word that has already been processed previously. Standard words use a standardized word count calculated from the number of characters in the subtitle divided by the average number of characters per word across the video as a whole.
• Average forward saccade length as a function of the standard word length—excluding saccades between refixations and also saccades that are so long that words between successive fixations were skipped. This index provides a way to obtain large-scale data on the reading of subtitles over extended text using eye-tracking measures. It must be noted, however, that all eye tracking yields a large volume of data.

Depending on the research question, the most useful eye-tracking measures for research in AVT can therefore be summarized as follows:

• Mean fixation duration: This measure provides a useful index of the processing effort, but should be used only to compare similar activities (e.g. reading of different types of subtitles and not reading to scene perception due to the fact that the eyes behave very differently when reading, which is a structured activity, than when exploring a visual scene).
• Dwell time: This provides a measure of the total time viewers spent looking at a particular area of interest, including both fixations and saccades, and is therefore a useful measure of attention allocation.
Number of fixations per word: This is particularly useful when an attempt is made to measure the extent to which subtitles were processed, although a more nuanced measure would take into account refixations and regressions.

Average forward saccade length: Like the number of fixations per word, this measure makes it possible to determine whether viewers performed regular reading or whether they skipped words.

Glance count: This measure gives an indication of the number of times viewers shifted their eyes between different areas on the screen (e.g. between the image and the subtitles).

Number of skipped subtitles: This is a fairly rough indication of how many subtitles were not even noticed by viewers, but should be considered together with other measures.

RIDT (reading index for dynamic texts): This measure provides a useful quantification of the extent to which subtitles were read.

Using these measures, researchers can investigate a wide range of issues such as:

- the impact of linguistic and translation features on the processing of subtitles;
- the impact of user-specific qualities such as language proficiency, educational level, degree of hearing, etc. on subtitle processing;
- the impact of subtitle-specific qualities such as presentation speed, layout or line division, font size and style, position, and other subtitling guidelines such as the position of subtitles in relation to visual boundaries such as shot changes on the processing of subtitles.

### Future trajectories

In the next decade the attention in AVT is bound to shift from research questions concerned with the impact of subtitles on comprehension, cognitive load, and attention distribution (as well as questions on the attention distribution of film viewers for AD), to research questions that are more aligned with research in other fields such as psycholinguistics and cognitive science. By now it is becoming evident that the basic questions have been answered, and the full potential of eye-tracking methodologies should be harnessed to interrogate various aspects associated with subtitling and AD conventions and guidelines, and with the utilization of subtitles and AD in education, therapy and other fields.

In the case of AD, subtitling has already yielded a wealth of information on the way sighted audiences view audiovisual texts, and where they get the most relevant visual information that can be utilized in creating AD that better produces an equivalent experience to blind and partially sighted audiences. With early indications that both subtitling and AD can be beneficial in focussing the attention of viewers with specific cognitive disorders, eye tracking in this domain will certainly gain momentum. Likewise, although many studies have shown that subtitles can be beneficial in terms of literacy, language proficiency, comprehension, and therefore in various educational contexts, future research will need to engage with a deeper understanding on which qualities of language and of the language of subtitles contribute to these benefits. Eye tracking will remain an important tool in this regard. However, due to the fact that eye tracking only offers an indirect indication of cognitive processing, methodologies employing eye tracking will have to be combined with methodologies such as EEG that could yield valuable online information on cognitive processing, particularly when combined with eye tracking.
As technology improves and becomes more user friendly and more affordable, eye-tracking studies are set to increase in the coming decades. In particular, with more sophisticated algorithms and more powerful computers, it is already becoming possible to perform much more large-scale eye-tracking studies with bigger numbers of participants resulting in more robust findings. However, there is also a need for researchers to use standardized methodologies and terminology and to report their experimental design in sufficient detail to allow for the replication of studies that will move the field closer to scientifically responsible and robust findings.

**Summary**

Using eye tracking in AVT research provides a powerful methodology for understanding how audiences process the audiovisual text. Particularly in subtitling, eye tracking can yield very detailed information on whether or how thoroughly viewers read the subtitles, which words or segments they looked at more, and how they switched their attention between the subtitles and the image. For this reason, eye tracking has become an indispensable tool in experimental research on AVT.

As an objective instrument for measuring visual attention distribution when viewers process the various visual signs contained in an audiovisual text, eye tracking can yield valuable information on how text and image are processed. It can help researchers determine how text-specific qualities such as subtitle style, presentation speed and line division impact on the effective processing of subtitles. It can also help us to understand how viewer-specific qualities such as language proficiency and hearing impact on subtitle processing. Also in the case of AD, eye tracking holds tremendous potential for helping us understand how sighted viewers process different visual codes, which could be used in decisions on which parts of the visual code are more important for AD.

In this chapter an overview is given of the most prominent studies in the field of AVT that have used eye tracking to look at attention distribution, language and translation of subtitles, the language of the audience, the presentation speed and other rules and conventions in subtitling, non-standard subtitles that break with conventions, and also AD. From these studies it becomes clear that eye-tracking research is becoming much more rigorous and provides a scientific tool that has become almost indispensable to AVT researchers. The chapter concludes with a summary of the most important eye-tracking measurements that are used in AVT research before sketching a few future trajectories in this approach.

**Further reading**

Holmqvist, K., M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka and J. van de Weijer (2011) *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford: Oxford University Press | This is a definitive book on eye-tracking methodology with detailed discussions not only of the more technical aspects of eye tracking, but also of experimental design, equipment and data analysis.


Jan-Louis Kruger

the development of the usage of eye tracking in the context of empirical research on subtitling and captioning, including foreign and same-language subtitling and captioning. It forms a critique of the eye-tracking measures and methodologies in this field of research.


Related topics

5 Voice-over: practice, research and future prospects
6 Subtitling for deaf and hard of hearing audiences: moving forward
8 Audio description: evolving recommendations for usable, effective and enjoyable practices
14 Psycholinguistics and perception in audiovisual translation
17 Multimodality and audiovisual translation: cohesion in accessible films
23 Audiovisual translation and audience reception
31 Accessible filmmaking: translation and accessibility from production
32 Technologization of audiovisual translation

References

Eye tracking in AVT research


