EYE TRACKING AND THE TRANSLATION PROCESS: REFLECTIONS ON THE ANALYSIS AND INTERPRETATION OF EYE-TRACKING DATA[1]

Kristian Tangsgaard Hvelplund

University of Copenhagen (Denmark)
bnm486@hum.ku.dk

Abstract

Eye tracking has become increasingly popular as a quantitative research method in translation research. This paper discusses some of the major methodological issues involved in the use of eye tracking in translation research. It focuses specifically on challenges in the analysis and interpretation of eye-tracking data as reflections of cognitive processes during translation. Four types of methodological issues are discussed in the paper. The first part discusses the preparatory steps that precede the actual recording of eye-tracking data. The second part examines critically the general assumptions linking eye movements to cognitive processing in the context of translation research. The third part of the paper discusses two popular eye-tracking measures often used in translation research, fixations and pupil size, while the fourth part proposes a method to evaluate the quality of eye-tracking data.

Resumen

El seguimiento ocular es un método de investigación cuantitativa de creciente popularidad en la investigación de la traducción. Este artículo aborda algunos de los aspectos metodológicos más importantes relativos el uso del seguimiento ocular en la investigación de la traducción. Se centra específicamente en el análisis y la interpretación de los datos de seguimiento ocular como reflejo de los procesos cognitivos durante la traducción. El artículo aborda cuatro tipos de aspectos metodológicos. La

1. I would like to thank Annette C. Sjørup and two anonymous reviewers for comments on earlier drafts of this article. I would like also to thank Lotte Jelsbech Knudsen for help with translating the abstract into Spanish.

MonTI Special Issue – Minding Translation (2014: 201-223). ISSN 1889-4178
http://dx.doi.org/10.6035/MonTI.2014.ne1.6

primera parte considera los pasos preparatorios previos a la grabación de datos. La segunda parte examina críticamente las hipótesis que vinculan los movimientos oculares al procesamiento cognitivo en el contexto de la investigación de la traducción. En la tercera parte se analizan dos parámetros de seguimiento ocular de uso frecuente en la investigación de la traducción (fijaciones y el tamaño pupilar), mientras que la cuarta parte propone un método para evaluar la calidad de los datos de seguimiento de los ojos.

Keywords: Translation. Eye tracking. Assumptions. Data quality. Indicators.

Palabras clave: Traducción. Seguimiento ocular. Hipótesis. Calidad de los datos. Indicadores.

Manuscript received on April 29, 2013 and accepted on September 5, 2013.

1. Introduction

Eye tracking has become a well established and increasingly popular quantitative research method in translation research to collect data about the cognitive processes involved in translation. A wide range of research questions have been explored using eye tracking, such as translation memory tools and cognitive load (O'Brien 2006), reading for translation as a particular type of reading (Jakobsen & Jensen 2008), coordination of comprehension and production processes in translation (Dragsted & Hansen 2008), directionality in translation (Pavlović & Jensen 2009; Chang 2009), reading modalities in translation (Alves *et al.* 2011), distribution of cognitive effort during translation (Hvelplund 2011), translator competence (Ehrensberger-Dow & Massey 2013), metaphor translation (Sjørup 2013), classification of translator styles (Dragsted & Carl 2013), parallel processing in translation (Balling *et al.* 2014), to name just a few.[2] The use of eye tracking as a research method raises methodological questions, and O'Brien (2009) and Alves *et al.* (2009) have dealt specifically with the methodological issues involved in the use of eye tracking in translation research. O'Brien (2009) focuses on the challenges researchers encounter when collecting eye-tracking data, and she discusses important issues such as research environment, participant selection and ethics. Alves *et al.* (2009) address the challenges involved in the use of eye tracking in combination with key logging and retrospective analysis, and raise questions on issues related to the reliability and comparability of eye-tracking data across studies.

This paper focuses specifically on the challenges involved in the analysis and interpretation of eye-tracking data as reflections of some sort of cognitive activity, and it also comments on relevant methodological aspects to consider when capturing the translation process using eye tracking. Ideally, eye-tracking data reflect the translator's object of attention with perfect spatial accuracy and perfect temporal precision. In reality, however, several issues

2. See Alves *et al.* (2012) for an overview of some of recent studies exploring the translation process using eye tracking.

complicate a straightforward interpretation of eye-tracking data as evidence of concurrent cognitive processing. For instance, since our thoughts can shift independently of eye movements, how can we rely on Just & Carpenter's (1980: 331) eye-mind and immediacy assumptions to infer something about the translator's cognitive focus during translation? With respect to process measures, can we, for instance, be certain that longer fixations actually reflect more processing intensity? Might it not be that a task attracts longer fixations because the eyes are monitoring the mechanical operation of typing? These and other questions are considered here in the context of translation research. The paper is divided into four parts. The first part discusses the suitability, advantages and disadvantages of different types of eye trackers for translation research as well as issues related to participant selection and research setting. The second part evaluates basic assumptions of eye movements as indicators of cognitive processing in the light of translation research, while the third part focuses on eye tracking measures, including fixations and pupil size, as well as their advantages and disadvantages. Finally, the fourth part proposes three methods to evaluate the quality of eye-tracking data.

2. Eye trackers, participants and research setting

Several issues can have an impact on the reliability of eye-tracking data as reflections of the cognitive processes involved during translating. O'Brien (2009: 252) points to several methodological challenges when using eye tracking to examine cognitive processes. The choice of eye tracker, the research environment and the participant's familiarity with working with an eye tracker (or lack thereof) are some of the factors which may contribute negatively to the generalisability of the recorded data. Eye-tracking data can be affected by a variety of factors not related specifically to cognitive events. For instance, eye-tracker accuracy, reflexive responses, such as pupillary responses to changes in light intensity, and a possible white-coat effect from having a translation process monitored in an unfamiliar environment are some of those factors. In order not to base findings on data which cannot easily be said to be generally representative of translation processing, careful consideration must be given to these issues that could otherwise cause problems for the interpretation of eye-tracking data as indication of cognitive processing during translation.

2.1. Types of eye trackers

An eye tracker is a device that registers and records where the eyes are looking. Most modern eye trackers use video-based technology to measure the

position of the eye, where a camera captures and records the reflection of infrared light on the eye's cornea or retina (Duchowski 2007: 54). Eye trackers are often compared on how accurately they reflect where the user is looking, and accuracy is typically measured in degrees of visual angle. The inaccuracy reported by eye-tracker manufacturers is typically between 0.5 and 1 degrees, which corresponds to roughly 0.5 to 1 centimetre, although some head-supported systems reportedly have even higher accuracy. In deciding which eye tracker is best for a given research objective, accuracy is one important factor, but the degree of invasiveness and the eye tracker's recording speed, i.e., its sampling rate, measured by the frequency of gaze sample registrations per second (Hz), are also factors to consider.

Remote eye trackers, also referred to as desktop eye trackers, are generally the preferred type in translation research (see also O'Brien 2009: 263). In remote eye trackers, cameras are integrated into a separate box which is placed in front of, or attached to, a computer monitor (e.g., SMI's RED series, Tobii's X60/120, TX300 and SR Research's EyeLink 1000) or the cameras are integrated into a dedicated monitor (e.g., Tobii's T60/120), making this type of tracker less invasive than head-mounted eye trackers and systems which require head support.[3] Accuracy is typically between 0.5 and 1 degree, and remote eye trackers have the overall advantage that the participant can freely move around and move the head without compromising the quality of the recording. In naturalistic research settings—such as most translation setups, which are intended to imitate an authentic setting—movement restriction may potentially cause the participant to become very aware of the research setting, and stress and a possible white-coat effect may arise resulting in a recording that cannot straightforwardly be assumed to contain generalizable reflections of typical translator behaviour. Since there is usually no movement restriction with remote eye trackers, this is not a serious issue, so this type constitutes a good choice for translation research.

Head-mounted systems, such as SMI's iView X HED and SR Research's EyeLink II, and eye-tracking glasses, such as the SMI Eye Tracking Glasses and the Tobii Glasses, also allow free head movement and inaccuracy is typically about 0.5 degrees. This type of eye-tracking device is more invasive than remote devices, since the participant has to wear the equipment on his/her head. Having an eye tracker strapped to the head will most likely make

3. Tobii Technology <http://www.tobii.com/> Accessed 9 March 2013.
SR Research <http://www.sr-research.com> Accessed 9 March 2013.
SensoMotoric Instruments (SMI) <http://www.smivision.com/> Accessed 9 March 2013.

the participant even more aware that s/he is being observed, and this could also influence the reliability of the eye-tracking data as reflection of cognitive processing. The head-mounted systems work at between 200–500 Hz, which is comparable to the recording frequency of most remote systems, while the glasses work at only 30 Hz. Such low frequency is not ideal for studies where high temporal resolution is needed to make detailed observations on changes in fixation duration and pupil sizes, and by this measure eye-tracking glasses are not well-suited for translation research. One advantage that head-mounted eye trackers and glasses have over remote eye trackers is that the recording area of the device is not restricted to the screen area of a computer monitor. While remote eye trackers will only capture eye movements inside the computer monitor area, head-mounted eye trackers and glasses capture eye movements outside of this area as well. For translation research, specifically, this is particularly important if the researcher is interested in the participant's use of external resources, such as printed dictionaries or other texts, or if the researcher wishes to examine how frequently the participant monitors his/her typing activities.

Eye trackers which require that the participant's head is kept still, such as SMI's iView X Hi-speed system, Arrington Research's HeadLock system and SR Research's EyeLink 1000 Head Supported, work at between 400 and 2000Hz, and inaccuracy can reportedly be as low as 0.25 degrees, corresponding to a spatial offset of around 0.25 centimetres between object and actual visual focus.[4] With this type of eye tracker, the head is stabilised using a chin rest or a bite bar, so this setup makes it the most invasive alternative of the three types discussed here. In terms of ecological validity, there is the general problem of possible stress and white-coat effects from having the head fixed to the eye tracker. Secondly, since the participant's head is fixed in a locked position, the participant will have poor if no visual contact with a keyboard, if such is used. For translation research specifically, this means that the translator will not be able to monitor his/her typing activities, which severely complicates the writing process. The translator's eye movements may thus well be related to stress from not being able to look at what key is being pressed in addition to actual problem-solving activities arising from the translation itself. Although some translators are good touch typists and do not need to frequently monitor the keyboard, a situation in which the translator is restricted by the experimental setup to look only at the monitor will result in an undesirable research

4. Arrington Research Eye Trackers <http://www.arringtonresearch.com/> Accessed 20 March 2013.

setting, which further compromises the ecological validity of the research. Overall, remote eye trackers are less intrusive than the other two types, and for that reason it constitutes a better option. However, head-mounted trackers have the advantage that they record activities which are not confined to what occurs on the computer monitor.

2.2. Participant profile

Other factors that have an impact on how well the eye tracker registers the position of the participant's eyes on the monitor have to do with the participant's eyes and the possible use of eyeglasses. Problems related to eyes and eyeglasses can to some extent be anticipated and taken into account before the eye-tracking data is collected, thus increasing the chances of a successful high quality recording. Glasses and contact lenses are normally not a problem for the quality of the recorded eye-tracking data, although the shape of the eyeglasses worn by the participant can have a negative impact. Eye trackers rely on unobstructed view between the eye tracker and the participant's eyes, and if the frame of the eyeglasses is very narrow, then the infrared signal may be obstructed by the frame and data are not captured by the tracker. Similarly, bifocal lenses can cause problems since the quality of the signal from the eye to the eye tracker can become affected by sudden variation in lens dioptre. For the same reason of obstruction of view between eye and eye tracker, even very long eyelashes and heavy mascara can make recording of good eye-tracking data problematic. To obviate these problems, it can be necessary to instruct participants not to wear heavy mascara and, if possible, to use eyeglasses that do not have a very narrow frame or bifocal lenses. The eye tracker should in any case be thoroughly calibrated, so that the researcher can identify potential problems before the recording session begins.

The problems outlined above are inherent to any research with eye tracking, but in relation specifically to translation, there is the additional problem of the participant's ability to touch type, which was discussed in relation to eye tracker type above. The participant's ability to touch type will affect how frequently s/he looks at the keyboard to monitor typing activities. The most popular type of eye-tracker system, i.e., remote eye trackers, record only the position of the eyes when the eyes are looking at the computer monitor, so if the participant is visually monitoring his/her typing activities by looking at the keyboard, then the eye tracker will not be able to record any eye movement activity and the level of *completeness* of the recording is not as high as it might have been. It might therefore be tempting to prefer only participants that are able to touch type, and ideally the selected participants would share

the same high level of touch typing ability; however, this selection criterion places severe limitations on the potential pool of candidate participants in a translation study using eye tracking, and restriction such as this is not favourable. Instead, the impact of the participant's touch-typing ability should be controlled for statistically, rather than experimentally, using inferential methods (e.g., Balling 2008, Balling & Hvelplund, in preparation).

2.3. Data collection

To compensate for the inaccuracies of the eye-tracking equipment discussed above in 2.1, large font sizes are generally preferred in translation research using eye tracking. O'Brien (2009: 261ff) suggests a font size 16 or 18, and studies typically opt for font sizes 16–20. Using a large font size is relevant only to the extent that the researcher is interested in making observations on differences at the *word* or *sentence* levels. If the researcher is interested in differences between larger items, for instance at the *text* level between source text (ST) and target text (TT), then smaller fonts could be used. For the sake of recycling, however, it might still be a good idea to use a large font size, since data could then be reused for other research objectives that rely, for instance, on mapping of eye movements to words.

Another factor that can be controlled easily is the participant's distance from the monitor. Eye-tracker manufacturers typically recommend that participants are seated at a distance of around 60-80 centimetres from the monitor (e.g., SMI's RED systems specifications and Tobii's T60 & T120 Eye Tracker User Manual).[5] Some eye tracking software (e.g., Tobii Studio and SMI's iView X software) will indicate in real-time whether the participant is sitting at a suitable distance, but not all software provides this information. In those cases, it is absolutely necessary to manually check the distance to the eye tracker before the session starts. Jensen *et al*. (2009: 325) suggested that the high discard percentage reported in this study could be related to a high distance between participant and monitor and, based on these observations, it seems that there is a close causal relationship between distance to the computer monitor and discard percentage.

5. Tobii T60 / T120 Eye Tracker User Manual < http://www.tobii.com/Global/Analysis/Downloads/User_Manuals_and_ Guides/ Tobii_T60_T120_EyeTracker_UserManual.pdf> Accessed 20 March 2013.
SMI RED Systems specifications <http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/red-red250-red-500.html> Accessed 25 October 2013.

Since eye movements and pupil size are sensitive to variation in light intensity, it is generally recommended to collect eye-tracking data in a room with a stable source of light. O'Brien (2009: 253) suggests that a consistent source of light is used and that the blinds are closed in the room where the eye-tracking data are collected, while Holmqvist *et al*. (2011: 17) suggest an entirely windowless room. The matter of pupil size and relevant precautionary measures are considered in more detail in section 4.2 below.

3. Eye movements, assumptions and translation processes

The reliability of eye-tracking data as indication of cognitive processing has not yet received much critical attention in the context of translation research. From an intuitive perspective, it makes good sense to assume that we direct our focus of attention to whatever object we are looking at. For instance, when reading a book, it is hard, if not impossible, to intentionally detach our focus of attention from that word or string of words that we are looking at. Visual exposure to letters automatically activates a processing stream that cannot be interrupted intentionally, unless looking away from those letters (Valdés *et al*. 2005: 279). For translation, it also makes good sense to suppose that the translator focuses on the words that s/he is looking at. For instance, when fixations are observed on ST words, it makes sense that the translator is engaged in ST reading and when fixations are observed on TT words, it makes sense that the translator is engaged in tasks related to the processing of the TT. In translation research using eye tracking, Just & Carpenter's (1980: 331) eye-mind and immediacy assumptions are often used as an operational basis for assuming a link between visual focus and cognitive focus. Just & Carpenter point out that "there is no appreciable lag between what is being fixated and what is being processed" and "... the interpretations at all levels of processing are not deferred; they occur as soon as possible" (1980: 331). Just & Carpenter's assumptions hold that recordings of eye movements will unveil information about the contents of conscious processing during a task, as visual focus on objects, be it letters or images, will always lead to attention being focused instantaneously to those letters or images. This is not necessarily the case, though. For instance, our thoughts may drift unintentionally during reading, and we may think of something completely unrelated to clouds when staring at the sky. Such mind wandering, or mind drifting, is a frequent and common phenomenon (Smallwood & Schooler 2006: 946), and it suggests that the eye-mind assumption perhaps only provides an approximation of the relationship between visual focus and cognitive focus. The matter of covert attention in particular has been emphasised as a possible argument against

the eye-mind assumption: "[…] it is important to distinguish between overt changes in orienting that can be observed in head and eye movements, and the purely covert orienting that may be achieved by the central mechanism alone […]" (Posner 1980: 5.)

Posner distinguishes here between behavioural changes, which can be observed, and cognitive changes, which cannot, and highlights an important limitation of the eye-mind assumption, namely that cognitive focus can shift independently of eye movement. While the eye tracker can fairly accurately identify where the eyes are looking, it cannot identify the object of thought. For translation, specifically, this means that although the translator is looking at the ST, he may well be considering possible target language (TL) equivalents of that specific ST word, and when looking at the TT, the translator may well be constructing meaning hypotheses based on ST content. In translation research, and other research disciplines as well, this problem of possible disagreement between visual focus and cognitive focus merits caution, and observations ought to be interpreted in the light of this potential weakness.

With respect to a potential weakness of the immediacy assumption, research suggests that the mind is up to 250 milliseconds ahead of the eye (cf. Holmqvist *et al.* 2011: 379). In other words, the mind focuses attentional resources to an object before it enters into visual focus. In a translation context, this could mean that the researcher cannot be certain if the translator is processing the word on which a fixation has been registered or if the translator is in fact preparing to process a successive word not yet in visual focus. With respect to this potentially asynchronous temporal relationship, Holmqvist *et al.* (2011: 379) offer a word of caution: "most eye-tracking research is conducted and interpreted as though attention and fixation were synchronous events […] they probably are not."

In terms of more technical challenges to the eye-mind assumption, there is also the issue of *drift,* which may further compromise the validity of the eye-tracking data as reflections of cognitive processing (Tobii Eye Tracking White Paper).[6] Drift is when the recorded eye position and the true eye position become gradually asynchronous as a data-collection session progresses. Drift is measured by calculating the gradual dislocation of the participant's gaze relative to screen content in degrees of visual angle. Eye tracker manufacturers typically report up to 0.3 degrees drift over time, which corresponds

6. <http://www.tobii.com/en/eye-tracking-research/global/library/white-papers/tobii-eye-tracking-white-paper/> Accessed 9 January 2013.

to roughly 0.3 centimetres. This is not likely to be a serious issue for short translation sessions but for lengthy sessions, which are not uncommon, it may well have an impact on the reliability of the eye-tracking data. For lengthy translation sessions, it may be necessary to consider dividing the session into shorter sessions with additional eye-tracking calibrations between sessions.

Despite the issues discussed above, eye tracking constitutes a very useful and powerful methodology to make observations on the cognitive processes during the translation process. While the eye-mind and immediacy assumptions have their weaknesses, they do offer a reasonable basis for assuming some sort of relationship between eye movements and translation processing. For example, while covert attention is a factor to consider, we cannot ignore the many instances during the translation process where ST words have been read for the purpose of translating them into the TL. During those instances, visual focus will have been overt manifestation of cognitive focus. Findings from reading research provide further support for assuming a link between visual and cognitive focus. Fixation duration tends to increase as an indication of increase in perceived difficulty with less frequent and less predictable words (Inhoff & Rayner 1986, Ehrlich & Rayner 1981) and with more complex and difficult genres (Rayner & Pollatsek 1989). Furthermore, since the task of translating is a cognitively demanding one, there is arguably little room for much mind wandering, and we may cautiously assume that the *majority* of eye movements during translation relate to on-going, conscious, synchronous processing of the translation task. This argument is supported by psychology research, which has found that mind wandering is more likely to occur in tasks that are simple or automatic than in attention-demanding tasks (Smallwood & Schooler 2006: 947, 956). In summary, the eye-mind and immediacy assumptions are reasonable assumptions that are not only necessary in order to be able to interpret eye movements as correlates of cognitive processing in translation but that have been successfully validated in neighbouring research disciplines.

It should be noted that in order to increase the likelihood that eye movement data reflect actual cognitive activities, other precautions can and should be taken. Having a high number of participants and filtering recordings according to eye-tracking data quality (e.g., Hvelplund 2011, Sjørup 2013) and statistical control of possibly confounding factors (e.g., Balling 2008, Balling & Hvelplund, in preparation) are some of the steps that could be taken to further secure large amounts of high quality data.

4. Eye movement measures and translation

In eye-tracking research in general, a wide range of measures have been used to make observations on eye movements during different types of tasks. These measures include fixation-based measures, pupil-based measures, saccade-based measures and transition-based measures, and they can be further classified into movement measures, position measures, numerosity measures and distance measures.[7] In most eye-tracking software, popular measures such as fixation duration, fixation count, time to first fixation, first fixation duration, etc. are readily available at the click of a button. Most eye-tracking software also offers visualisation of a recording, such as heat maps and gaze plots. Some measures, however, are not calculated automatically by the software. In those situations, manual identification is needed of the desired figures in the so-called *raw data* (see Hvelplund 2011: 112-116 for a practical example of data extraction from raw data output). For instance, access to information about pupil sizes most often requires manual identification in the raw data, and this manual work can potentially be quite time consuming.

In translation research, two indicators in particular have been popular to make observations on the cognitive processing during the translation process, namely measures of fixations and measures of pupil size. These two types of measures are considered below in some detail in relation to some of the issues involved in applying them to make inferences on the cognitive processes in translation. In addition, other quantitative measures of cognitive processing are discussed briefly and a note of caution is raised about the use of visualisation tools.

4.1. Fixations

A fixation is a type of eye movement often defined as a period of time during which the eye is relatively stable; the purpose of fixations is to bring an object of interest into visual focus (Duchowski 2007: 46). In translation process research using eye-tracking data, fixation duration and fixation count are popular measures, and they are often taken to index cognitive effort. Longer fixations and more fixations indicate more effortful processing and shorter fixations and fewer fixations indicate less effortful processing, and this more effortful processing is often linked to an increase in difficulty. While these interpretations make good sense, in particular in light of the eye-mind and

7. See Holmqvist *et al*. (2011) for a comprehensive overview of various eye movement measures.

immediacy assumptions discussed above, there are situations in which it does not seem so straightforward. As a case in point, the TT area of the monitor generally attracts longer fixations than the ST area of the monitor (e.g., Sharmin *et al*. 2008: 39, Jakobsen & Jensen 2008: 114, Pavlović & Jensen 2009: 101). But does this mean that TT reformulation, in general, is more difficult than ST reading and comprehension? Or could it be that these more and longer fixations over the TT area have to do with the eye moving more slowly across the TT in sync with the emerging TT being typed? For this specific case, it is necessary to distinguish between ST reading and TT reading as two different types of reading activity. A likely interpretation of the more and longer TT fixations is that TT fixation duration and count are functions of the mechanical time-consuming operations related to typing the TT *as well as* of the difficulty involved in TT reformulation. In other words, TT reading speed during typing is essentially defined by the typing speed with which the TT emerges on the computer screen, and TT reading is thus not necessarily more cognitively demanding than ST reading. This relationship between fixation duration and the nature of the task is supported by research in reading. Summarising previous work, Rayner (1998: 373) points out that mean fixation duration is 225 ms during silent reading and 275 ms during reading out loud. During typing, i.e., when reading and typing simultaneously, mean fixation duration is 400 ms, which is considerably longer than for 'regular' reading. Non-reading tasks such as visual search and scene perception also yield different durations, namely mean fixation durations of 275 ms and 330 ms, respectively. Since the nature of the task codetermines the duration of the fixation, it is important to interpret the fixation data in light of the *kind* of reading that the translator is performing. In the example above, the fixation durations need to be interpreted as two different reading tasks, i.e., as ST reading and as TT reading while typing, since the underlying tasks carried out during the two types of reading are fundamentally different.

In terms of comparing and replicating the findings from translation research using eye tracking, Alves *et al.* (2009: 274) and Alves *et al.* (2011: 191) point to the different filter settings used in different studies as a potentially complicating factor. A filter setting essentially defines which gaze samples should be included in a fixation, and most eye-tracking software allows the researcher to manipulate these settings. Two filter settings can typically be manipulated: a setting related to the maximum *distance* between two gaze samples (measured in millimetres or pixels) and a setting related to the minimum *duration* of the fixation (measured in milliseconds). Based on the filter setting, gaze samples will be grouped together if they are in spatial and temporal proximity

to each other according to predefined thresholds. Comparing different filter settings, Alves *et al*. (2009: 274) observe that "the data are […] inconclusive when different fixation filters are used, since […] the AOI (area of interest) having the longest fixation differs according to the filter." Alves *et al*. (ibid.) suggest that researchers should strive towards a standardisation of settings, i.e., that the duration and distance thresholds are kept the same irrespective of the study. This recommendation of standardisation is indeed very recommendable, although a uniform setting might be problematic in cases where two studies have recorded eye-tracking data at different eye-tracker speeds (i.e., *sample rates*). More specifically, the distance the eye travels between two samples is a function of the recording speed of the eye tracker. For instance, the eye travels fewer pixels, or millimetres, between two samples if the eye tracker recording has been sampled at 50 Hz (i.e., across a time span of 20 milliseconds) than if the recording has been sampled at 30 Hz (i.e., across a time span of 33 milliseconds). If a comparison is intended, sample rates and distance would have to be normalised to account for this difference. That said, standardisation of filter settings is a very recommendable proposal, and it would make comparison and replication of studies more straightforward.

4.2. Pupil size

Measures of pupil size or dilation are often taken as indicator of the working load placed on the cognitive system (e.g., Hess & Polt 1964, Holmqvist *et al*. 2011: 393). Overall, pupils that are more dilated indicate higher cognitive load, i.e., they indicate that a task is relatively more difficult, while pupils that are less dilated indicate lower cognitive load, i.e., that a task is relatively easier. In translation and interpreting research, some studies have used measures of pupil size as indicators of changes in cognitive load (including Hyönä *et al*. 1995, O'Brien 2006, Caffrey 2008, Chang 2009, Jensen *et al*. 2009, Pavlović & Jensen 2009, Hvelplund 2011).

In general, caution should be exercised when collecting and analysing pupil size data, since this type of eye movement is sensitive to not only changes in cognitive load but to many other factors. More specifically, pupils dilate and constrict as a reflexive response to changes in light intensity; in response to emotional events, such as stress, pain and fear; in response to the intake of medicine and stimulants, such as drugs and alcohol; or if the participant is ill (Holmqvist *et al*. 2011: 393ff). While it is hard to control the participant's emotional state of mind while the recording of eye-tracking data is in progress, it is less problematic to control for other factors. Obviously, prospective participants in a translation study who have consumed stimulants

or are ill should not be allowed to contribute. Maintaining the same light intensity in the room where the eye-tracking data is recorded is crucial, in particular for studies that are interested in changes in pupil sizes.

In terms of analysis of pupil size data, further caution should be exercised. Pupillary response to a stimulus occurs with some delay. Several estimates have been presented of this delay, or *pupillary latency*. For instance, during multiplication tasks, pupils react within 300 milliseconds (Ahern & Beatty 1979), and the pupillary response to light occurs after 150 to 400 milliseconds (Holmqvist *et al*. 2011: 435). For translation and interpreting, specifically, Hyönä *et al*. (1995: 605) found that pupils responded with a delay of between 300 and 500 millisecond in a study on interpreting, while Hvelplund (2011: 71, 117) estimates a pupillary delay of 120 milliseconds for ST and TT reading during translating, based on a heuristic design in which different latency values were tested. In order to reduce the risk that changes in pupil size are erroneously linked to the wrong word or object, a pupillary delay needs to be taken into consideration. This could be done either by applying a fixed pupillary delay to all recordings across all participants or by applying individual pupillary delays for each participant, assuming that not all participants' pupils respond with the same delay. The latter approach requires individual baseline measurements to be recorded for each participant before collecting the actual process data. In any situation, if this psychophysiological delay is not somehow taken into account, the researcher risks performing analyses on pupil measurements that do not reflect the actual pupil size related to a specific word or item.

4.3. Other measures

In addition to the popular position and numerosity measures discussed above, total gaze times (the sum of all fixation durations) on a word or a larger region is also a popular measure. Other measures have also been used in translation research to make observations on translation as a cognitive activity, including blink rate as an indicator of cognitive load (e.g., Chang 2009), attention shifts as indicators of cognitive management (Hvelplund 2011) and eye-key span as indicator of the translator's coordination efforts (Dragsted & Hansen 2008). While fixation data, and to some extent also pupil data, are often more easily accessible than, for instance, data from saccades, transitions, and blinks, it is very likely that translation research could benefit from exploring the possibilities of measures rarely used in translation process studies but often used in neighbouring research disciplines such as reading research. Also, measures such as first fixation duration (the duration of the fixation on a word the

first time the gaze lands on a word), second fixation duration, regression behaviour (how frequently a word is reread), have not yet been used much in translation research, and there is most probably more to be discovered about the translation process using these indicators of cognitive processing.

Visualisation of eye movement behaviour through heat maps and gaze plots is popular in translation research. Visualisation often gives a very good first impression of which part(s) of a text received most fixations during a translation. It is, for the most part, used as a supplement to the more quantifiable measures discussed above, and to generate hypotheses. However, the use of visualisation as the main source of input for hypothesis testing and exploration of research questions is not very recommendable. Visualisations cannot be subjected to statistical tests such as the data from the quantifiable measures discussed above, and hypotheses therefore cannot be verified with comparable certainty.

5. Eye-tracking data quality

As discussed above, the quality of eye-tracking data is sensitive to a variety of factors. Although precautions have been taken, eye-tracking data may still be of such poor quality that they are not realistic reflections of the translator's eye movements and pupil size. Thorough assessment of eye-tracking data quality is therefore a crucial step in the analysis process; it is, however, a step which is often neglected in process studies. Below, three methods to evaluate the quality of eye-tracking data are presented and discussed in the context of translation research.

5.1. Fixation measure

While translation process studies often do not report how eye-tracking data quality was assessed, but only that a number of recordings were discarded due to poor data quality, one relatively popular measure of eye-tracking data quality is calculations of mean fixation duration. In a study on translation directionality, Pavlović & Jensen (2009: 99) discarded recordings in which fixations were "abnormally short", namely < 200 milliseconds, noting that the mean fixation duration during silent reading is around 225 milliseconds (Rayner 1998: 373). Hvelplund (2011: 106) similarly used a mean fixation duration threshold of 200 milliseconds to discriminate acceptable data from non-acceptable data, while Sjørup (2013: 105) applied a threshold of 180 milliseconds.

Relying on the fixation duration alone as a quality measure is not entirely unproblematic. Mean fixation duration is a relatively crude measure, which ignores the potential difference in *completeness* of eye-tracking recordings. More specifically, completeness, seen as how much eye movement has been successfully recorded by the eye tracker compared to how much has not been recorded, varies between recordings as a function of various factors. For instance, as discussed above, the quality of a recording can be affected by the participant's use of optical aids, which means that the recorded eye-tracking data may only be partial reflections of the participant's eye movements. It might be that a participant's recording has a mean fixation duration of >200 milliseconds, but if this mean is calculated on the basis of just a few seconds of partial recording with just a few fixations, then it does not represent the overall quality of the eye-tracking data during a recording. In Hvelplund (2011: 260), one discarded participant had mean fixation durations of 201, 235 and 314 milliseconds in three separate recordings, which is comfortably close to Rayner's (1998: 373) 225 milliseconds mean in silent reading. However, these means were calculated from very few fixations, which represented only 1.7 per cent, 0.9 per cent and 1.4 per cent of the total recording time, respectively. During 98.3 per cent, 99.1 per cent and 98.6 per cent of the respective recordings, no fixations were detected by the equipment. Relying on mean fixation duration alone, these three recordings would have been included in the analyses and could have distorted the analyses.

5.2. Gaze time on screen

In response to the issue of completeness, Gaze Time on Screen (GTS) has been used as a measure to further gauge the quality of eye-tracking data (Hvelplund 2011: 104; Sjørup 2013: 105ff). GTS is a simple calculation of total fixation duration as a percentage of total task time [(total fixation duration / total task time) * 100]. The score provides an indication of either how much time the participant spent looking at the screen, or the quality of the eye-tracking data. A high GTS score may indicate that the participant looked at the monitor for a considerable amount of time *or* that the eye tracker captured well the eye movements of the participant. A low GTS score may indicate that the participant only looked at the monitor for a limited amount of time *or* that the eye tracker did not capture well the eye movements of the participant. In a translation setting, we can expect that the translation task requires the participant to look at the monitor for a fair amount of time in order to read the ST (and most likely also the TT or parts of it), but there will obviously be instances during which the translator looks away from the monitor, for

instance, to monitor typing activities and to consult offline dictionaries, and the GTS score will very rarely be 100 per cent. In the example above with the very low percentages, the translator did in fact manage to translate three texts consisting of a total of 419 words. The three low percentages presented above constituted only around 11 seconds of total fixation time for all three texts. It is fairly reasonable to assume that there are problems with the quality of the eye-tracking data in this translator's recordings – and not so much that the participant managed to read 38 ST words per second. Despite this measure's advantage over the fixation duration measure, it has the overall disadvantage that the percentage reflects recording quality as well as the amount of time that the participant looked at the monitor. In other words, if the translator spends a considerable amount of time looking up words in dictionaries then the percentage will be correspondingly low – irrespective of the otherwise high quality of the eye-tracking data. To sum up, the researcher cannot be certain if a GTS percentage is the product of overall poor eye-tracking data quality or if the translator looked away from the monitor for substantial periods of time.

5.3. Gaze sample to fixation percentage

In response to the drawback of the GTS measure, a third measure has been used to gauge eye-tracking data quality. The gaze sample to fixation percentage (GSF) is based on the circumstance that saccades constitute between 5 and 15 per cent of all eye movements in reading. Ideally, a recording would reflect this ratio, in which roughly 85-95 per cent of a recording's gaze samples could be categorised as belonging to fixations and roughly 5-15 per cent would belong to saccades. The GSF percentage is calculated by comparing the total number of gaze samples with the total number of gaze samples that formed part of a fixation ([number of gaze samples / number of fixation gaze samples] * 100). For instance, in a recording containing 11,000 gaze samples, of which 10,000 belong to fixations, the GSF percentage is 90.9. In Hvelplund (2011: 259), around half of the GSF percentages of the study's 81 recordings were lower than 85 per cent and a lower threshold at 75 per cent was adopted for practical reasons: "[...] the quality of eye-tracking data is prone to be affected by external factors [...] irrespective of the efforts made to minimise them [...]" (Hvelplund 2011: 105). With a 75 per cent threshold, 11 recordings (13.8 per cent) were considered to be of low quality while 70 recordings were above threshold. Unlike the GTS measure, this measure does not presuppose that the participants spend the same relative amount of time looking at the monitor, and it therefore constitutes a better alternative to measuring eye-tracking

data quality. Its main drawback, however, is that the figures are not straightforwardly accessible, and calculations have to be done on the recording's raw data, which may be potentially cumbersome and time-consuming.

Most eye tracking software provides mean fixation duration of a given recording at the press of a button. While obtaining the GSF percentage is a potentially labour-intensive process, some analysis software can be useful in determining a GTS percentage. Tobii Studio, for instance, automatically calculates how many gaze samples were correctly identified as belonging to fixations or saccades as a percentage of the total number of identification attempts made by the software. Combined, these three methods offer a robust tool to identify which recordings should be included in a study's analyses and which should not, due to data quality issues. Other methods to screen poor data from good data, for instance, one based on pupil size registrations, could be equally useful; the key point here is, however, that neglecting proper data screening may have a serious impact on the data analysis of a given study, and it could potentially distort a study's findings.

6. Summary and conclusion

The aim of this paper has been to highlight some of the relevant methodological issues when interpreting and analysing eye-tracking data from recordings of the translation process. It has also commented on relevant issues related to preparing data collection in the context of translation research.

Eye-tracking data can potentially be misrepresentative of actual cognitive processing if caution is not exercised when collecting it. In terms of the choice of eye tracker, remote eye trackers are generally better suited for translation research because they are less invasive than head-mounted and head-supported systems. With respect to interpreting eye-tracking data as manifestations of cognitive processing, the link from eye movement to cognitive processing is intuitively sound; however, there are issues that make the link less straightforward. Issues such as covert attention and mind wandering can potentially complicate interpretation, and the researcher should consider how these issues may affect the findings of a study and how the effect of these issues can be minimised. When collecting and analysing eye movement data, there is a host of confounding factors that are not necessarily linked to the translator's problem-solving activities during translation, and they could potentially distort the analyses of the recorded translations. Pupils, for instance, are sensitive to many factors, including changes in light intensity and the emotional state of the participant, and the researcher should aim at controlling for these potentially error-inducing factors when collecting the

data and also when analysing it. In addition, while measures of fixation and pupils are popular indicators of cognitive processing in translation research, other measures that are popular in other research disciplines might be useful to the study of translation processes.

Even if precautions have been taken during data collection, the quality of the eye-tracking data might still be poor and overall not reflect the translator's process. An important step in the analysis phase is, therefore, to discard low-quality data. While mean fixation duration is a reasonable indicator, other more powerful measures, such as gaze time on screen and fixation samples as a percentage of gaze samples, are recommended since they reflect the quality of the data in more detail. Careful attention to the complex interplay of factors that are inherent to translation research using eye tracking before, during and after the data collection will help increase the reliability and generalisability of the findings of an analysis.

References

AHERN, Sylvia & Beatty Jackson. (1979) "Pupillary responses during information processing vary with scholastic aptitude test scores." *Science* 205:4412, pp. 1289–1292.

ALVES, Fábio; Adriana Pagano & Igor da Silva. (2009) "A new window on translators' cognitive activity: methodological issues in the combined use of eye tracking, key logging and retrospective protocols." In: Mees, Inger M; Fábio Alves & Susanne Göpferich (eds.) 2009. *Methodology, Technology and Innovation in Translation Process Research*. Copenhagen: Samfundslitteratur, pp. 267–292.

ALVES, Fábio; Adriana Pagano & Igor da Silva. (2011) "Towards an investigation of reading modalities in/for translation: An exploratory study using eye-tracking data. In: O'Brien, Sharon (ed.) *Cognitive Explorations of Translation*. London: Continuum, pp. 175–196.

ALVES, Fábio; José Luiz Gonçalves & Karina Szpak. (2012) "Identifying instances of processing effort in translation through heat maps: An eye tracking study using multiple sources. *Proceedings of the First Workshop on Eye-tracking and Natural Language Processing, COLING December 2012*, pp. 5–20. Electronic version available at: <http://www.aclweb.org/anthology-new/W/W12/W12-49.pdf> on 9 January 2013.

BALLING, Laura Winther. (2008) "A brief introduction to regression designs and mixed-effects modelling by a recent convert." In: Göpferich, Susanne; Arnt L. Jakobsen & Inger M. Mees (eds.) *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing*. Copenhagen: Samfundslitteratur, pp. 175–192.

BALLING, Laura Winther & Kristian Tangsgaard Hvelplund. [In preparation] "Quality in quantitative translation research."

BALLING, Laura Winther; Kristian Tangsgaard Hvelplund & Annette Camilla Sjørup. (accepted, 2014) "Evidence of parallel processing during translation." *Meta* 59/2.

CAFFREY, Colm. (2008) "Using pupillometric, fixation-based and subjective measures to measure the processing effort experienced when viewing subtitled TV anime with pop-up gloss." In: Göpferich, Susanne; Arnt L. Jakobsen & Inger M. Mees (eds.) *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing*. Copenhagen: Samfundslitteratur, pp. 125–144.

CHANG, Vincent Chieh-ying. (2009) *Testing Applicability of Eye-tracking and fMRI to Translation and Interpreting Studies: An Investigation into Directionality.* PhD thesis. London: Imperial College.

DRAGSTED, Barbara & Michael Carl. (2013) "Towards a classification of translation styles based on eye-tracking and keylogging data." *Journal of Writing Research* 5:1, pp. 133–158.

DRAGSTED, Barbara & Inge Gorm Hansen. (2008) "Comprehension and production in translation." In: Göpferich, Susanne; Arnt L. Jakobsen & Inger M. Mees (eds.) *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing*. Copenhagen: Samfundslitteratur, pp. 9–30.

DUCHOWSKI, Andrew T. (2007) *Eye Tracking Methodology: Theory and Practice*. London: Springer.

EHRENSBERGER-DOW, Maureen & Gary Massey. (2013) "Indicators of translation competence: Translators' self-concepts and the translation of titles." *Journal of Writing Research* 5:1, pp. 103–131.

EHRLICH, Susan F. & Keith Rayner. (1981) "Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior* 20:6, pp. 641–655.

HESS, Eckhard H. & James M. Polt. (1964) "Pupil size in relation to mental activity in simple problem solving." *Science* 143:3611, pp. 1190–1192.

HOLMQVIST, Kenneth; Marcus Nystrom; Richard Andersson; Richard Dewhurst; Halszka Jarodzka & Joost van de Weijer (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. New York: Oxford University Press.

HVELPLUND, Kristian Tangsgaard. (2011) *Allocation of Cognitive Resources in Translation: An Eye-tracking and Key-logging Study.* PhD thesis. Copenhagen: Copenhagen Business School.

HYÖNÄ, Jukka; Jorma Tommola & Anna-Mari Alaja. (1995) "Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks." *Quarterly Journal of Experimental Psychology* 48A: 598–612.

INHOFF, Albrecht Werner & Keith Rayner. (1986) "Parafoveal word processing during eye fixations in reading: Effects of word frequency." *Perception & Psychophysics* 40:6, pp. 431–439.

JAKOBSEN, Arnt Lykke & Kristian Tangsgaard Hvelplund Jensen. (2008) "Eye movement behaviour across four different types of reading task." In: Göpferich, Susanne; Arnt L. Jakobsen & Inger M. Mees (eds.) *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing*. Copenhagen: Samfundslitteratur, pp. 103–124.

JENSEN, Kristian Tangsgaard Hvelplund; Annette Camilla Sjørup & Laura Winther Balling. (2009) "Effects of L1 syntax on L2 translation." In: Mees, Inger M.; Fábio Alves & Susanne Göpferich (eds.) *Methodology, Technology and Innovation in Translation Process Research*. Copenhagen: Samfundslitteratur, pp. 319–338.

JUST, Marcel Adam & Patricia A. Carpenter. (1980) "A theory of reading: From eye fixations to comprehension." *Psychological Review* 87:4, pp. 329–354.

O'BRIEN, Sharon. (2006) "Eye-tracking and translation memory matches." *Perspectives: Studies in Translatology* 14, pp. 185–203.

O'BRIEN, Sharon. (2009) "Eye tracking in translation-process research: methodological challenges and solutions." In: Mees, Inger M.; Fábio Alves & Susanne Göpferich (eds.) *Methodology, Technology and Innovation in Translation Process Research*. Copenhagen: Samfundslitteratur, pp. 251–266.

PAVLOVIĆ, Nataša & Kristian T. Hvelplund Jensen. (2009) "Eye tracking translation directionality." In: Pym, Anthony & Alexander Perekrestenko (eds.) 2009. *Translation Research Projects 2*. Universitat Rovira i Virgili, Tarragona, pp. 101–119. Electronic version available at: <http://www.tobii.com/Global/Analysis/Marketing/Research%20Paper/linguistics%20psychology/Eye%20tracking%20translation%20directionality.pdf>

POSNER, Michael I. (2011) "Orienting of attention." *Quarterly Journal of Experimental Psychology* 32:1, pp. 3–25.

RAYNER, Keith. (1998) "Eye movements in reading and information processing." *Psychological Bulletin* 124:3, pp. 372–422.

RAYNER, Keith & Alexander Pollatsek. (1989) *The Psychology of Reading*. Englewood Cliffs, NJ: Prentice Hall.

SHARMIN, Selina; Oleg Špakov; Kari-Jouko Räihä & Arnt L. Jakobsen. (2008) "Where on the screen do translation students look while translation, and for how long?" In: Göpferich, Susanne; Arnt L. Jakobsen & Inger M. Mees (eds.) *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing*. Copenhagen: Samfundslitteratur, pp. 31–52.

SJØRUP, Annette Camilla. (2013) *Cognitive Effort in Metaphor Translation: An Eye-tracking and Key-logging Study*. PhD thesis. Copenhagen: Copenhagen Business School.

SMALLWOOD, Jonathan & Jonathan W. Schooler. (2006) "The restless mind." *Psychological Bulletin* 132:6, pp. 946–958.

VALDÉS, Berenice; Andrés Catena & Paloma Marí-Beffa. (2005) "Automatic and controlled semantic processing: A masked prime-task effect." *Consciousness and Cognition* 14:2, pp. 278–295.

BIONOTE / NOTA BIOGRÁFICA

Kristian Tangsgaard Hvelplund is an assistant professor in the Department of English, Germanic and Romance Studies at the University of Copenhagen. He holds a PhD in translation from the Copenhagen Business School. His research interests include translation and cognition and his research has focused in particular on the cognitive processes involved in the translation process, using experimental methods such as eye tracking and key logging.

Kristian Tangsgaard Hvelplund es profesor del Departamento de Estudios Ingleses, Germánicos y Románicos de la Universidad de Copenhague. Tiene un doctorado en traducción de la Copenhagen Business School. Sus intereses de investigación giran en torno a la traducción y la cognición. En particular, su investigación se ha centrado en los procesos cognitivos del proceso de la traducción, utilizando los métodos experimentales de seguimiento ocular y registros de pulsaciones de teclado.