Eye-tracking studies in conference interpreting
Agnieszka Chmiel

Introduction
Eye-tracking is a technique of measuring the movements of the human eye. By defining the position of the eye in reference to the visual stimulus it is possible to determine where a person is looking at a given time. Rather than moving in a smooth manner, the human eye makes rapid movements called saccades and between these movements the eye remains still, or fixates on a visual stimulus (Rayner 1998). Thus, fixations are the primary object of study by means of eye-trackers. Eye-trackers typically also measure pupil dilation, i.e., changes in the size of the pupil which might reflect cognitive load or emotional response to stimuli. Modern eye-trackers emit invisible infrared light that is tracked after it reflects from the eye. Experiment participants can thus behave quite naturally although extensive head movements are discouraged and head constraining chin rests or forehead rests are sometimes used. The advantages of eye-tracking in interpreting studies include objectivity and limited invasiveness (Seeber 2015), which mean that experimental conditions are rather similar to real-life tasks and the ecological validity of the studies is sufficient. Provided that study designs are sufficiently described, such data can be reliable and easily replicable. Additionally, eye-trackers have high temporal and spatial resolution: they typically record the position of the eye and the location of the gaze from 60 to even 2000 times per second.

Tracking eye movements is interesting if we assume that these movements reflect what happens in the mind. The seminal eye-mind assumption (Just & Carpenter 1980) posits that as the eye fixates on a word, that word is being processed in the mind. Obviously, the connection between overt movements of the eye and covert cognitive processes of the mind is not straightforward because our eyes might sometimes view certain stimuli while our thoughts are engaged otherwise (Hvelplund 2014). Also, the synchronicity of the visual and cognitive focus of attention might not be exact, which led Jakobsen (2014) to compare the eye to a dog on a leash held by the mind. Although the link between the eye and the mind is indirect, eye-tracking has been recognized as a legitimate research method that enriches our knowledge of many cognitive processes. As such, it has been used in conference interpreting research for many years, with first attempts as early as in the 1980s and more recent and more frequent applications in the last ten years.
Eye-trackers provide data about the exact position of the eye (gaze coordinates) and pupil size. By combining this information with the visual stimuli and the temporal dimension of the data (how long a participant looked at a given stimulus), we can obtain information about where and how visual attention is directed. Typically, longer processing times for specific stimuli are considered to reflect difficulty of processing and the ensuing cognitive load. Having a detailed insight into the interpreters’ visual attention with a high temporal and spatial resolution might contribute to extending our knowledge about information processing in interpreting and to identifying problem triggers. Also, since interpreting involves various types of visual stimuli (the speaker, presentation slides, textual input), eye-tracking data can tell us how the interpreter’s cognitive resources are distributed to process these sources of input.

**Visual input in interpreting**

All the elements of visual processing in interpreting may be studied with eye-tracking: “[u]sing the appropriate design, eye-tracking technology might hold the potential to answer the question of why interpreters want visual input, what kind of visual input they use, and when” (Seeber 2012: 346). Although interpreting predominantly involves the auditory modality (both as regards input and output), visual processing is part and parcel of interpreting. Interpreters can gain valuable information from the visual modality, for instance, by looking at the speaker, their facial expressions and gestures, noticing audience reactions or following the written version of the speech delivered. Due to the increasing use of visual aids at conferences, we might talk about “a new profile for the conference interpreter, who has not only to interpret what s/he hears but also what s/he sees” while efficient processing of visual materials has “become indispensable for the interpreting performance” (Kumcu 2011: viii). This section will include a brief overview of visual input in four modes of interpreting: (1) simultaneous interpreting; (2) simultaneous interpreting with text; (3) sight translation; and (4) consecutive interpreting, to show the range of stimuli that can be used in research involving eye-tracking.

Kumcu (2011) provides an extensive typology of visual materials that can be involved in simultaneous interpreting. These are either image-based or text-based. Image-based visuals include the speaker (their body language, lip movements and facial expressions) and the audience, as well as pictures, graphics, videos and physical objects shown by the speaker. Text-based visual materials might be attributed to the speaker (written text of a speech, presentation slides, notes, documents the speaker refers to) and to the interpreter (conference programme, speaker’s bionote, glossaries created before simultaneous interpreting and notes taken during simultaneous interpreting—such as numbers or words scribbled to ease target language production). As exemplified by this impressive list, visual input in simultaneous interpreting involves much more than just the speaker’s face and the slides. All of these different types of visual input may have a different effect on the interpreter and may be salient at different moments.

The richness of multimodal input in simultaneous interpreting has been reflected in two theoretical models. According to Setton (1999), lip movements are part of the phonetic analysis involved in word recognition while facial expressions and gestures are part of pragmatic analysis. Seeber’s (2017) model of a cognitive resource footprint for simultaneous interpreting considers both auditory verbal and visual spatial input (e.g. speaker’s lip movements, gestures, images) that compete for resources. However, activities in different modalities tap into different resources and their interference is lower than for activities within the same modality competing for the same pool of resources. By carefully considering potential interferences between multimodal inputs in simultaneous interpreting and simultaneous interpreting with text, Seeber
concludes that the former is less complex due to fewer sources of input. It is worth mentioning that Gile’s seminal Effort model for simultaneous interpreting (Gile 2009) does not include any specific effort for visual processing. However, we might not preclude that visual processing of gestures, facial expressions or lip movements is not implicitly part of the comprehension effort.

The distinction between simultaneous interpreting and simultaneous interpreting with text is sometimes blurred because the content of slides might vary greatly: some speakers use many graphics while others provide long quotations and extensive lists of bullet points. Thus, simultaneous interpreting might become simultaneous interpreting with text even when the interpreter does not have access to the written speech—the speaker might simply read out the text provided on the slides. Typically, however, the written speech is the main visual for simultaneous interpreting with text. Gile’s Effort model (2009) for simultaneous interpreting with text includes efforts for reading, listening, memory, production and coordination, while Seeber’s model (2017) shows that visual spatial, visual verbal and auditory verbal resources have to be recruited, which results in the interference score higher than that for classic simultaneous interpreting.

Sight translation and consecutive interpreting are less researched modes of interpreting. Gile’s Effort model (2009) for sight translation includes the reading effort alongside memory, speech production and coordination. As for consecutive interpreting, the Effort model includes separate formulas for the listening and note-taking stage and only the latter includes an effort for visual processing related to note-reading alongside other efforts for remembering, production and coordination. However, one has to bear in mind that other visuals might be involved in these modes, as well. For instance, although the interpreter is busy taking notes in the note-taking stage of consecutive interpreting (see Ahrens & Orlando, Chapter 3, in this volume), he or she may glance at the speaker for additional visual input or look at the slides accompanying the speech.

**Eye-trackers**

The studies involving eye-tracking in interpreting have so far used three brands of eye-trackers from leading vendors in the field. These are EyeLink (SR Research 2020), Tobii (Tobii AB 2020) and SMI. The latter one no longer operates and its devices are neither sold nor supported. EyeLink was used in many studies reviewed below: Chmiel and Lijewska (2019), Chmiel et al. (2020), Huang (2011), Korp and Stachowiak-Szymczak (2018, 2020), Seeber and Kerzel (2011), Stachowiak (2016), Stachowiak-Szymczak and Korpal (2019). Tobii was used in the following studies reviewed below: Dragsted and Hansen (2009), Gieshoff (2018), Korp (2012), Kumcu (2011), Seeber (2012), Seubert (2019), Shreve et al. (2011), Su and Li (2019). The eye-tracker provided by the third vendor (SMI ETG 2W, a head-mounted lightweight eye-tracker) was used in one study on consecutive interpreting by Chen (2017). There are also other eye-trackers available, including such low-end solutions as EyeTribe or Gazepoint, but these have not been used so far in conference interpreting research.

Eye-trackers use near-infrared light that is reflected in the pupil and the cornea and tracked by the infrared camera. Data collection is non-invasive and such illumination is not harmful. The information obtained from the camera is then analysed to arrive at gaze estimation. As Holmqvist et al. write: “Geometrical calculations combined with a calibration procedure are finally used to map the positions of the pupil and corneal reflection to the data sample (x, y) on the stimulus” (2015: 38). Image analysis involves complex algorithms, so the raw data obtained from the eye-tracker is in fact a result of advanced computations. Good calibration is thus vital for collecting good quality eye-tracking data. During calibration, the eye-tracker
obtains data of the user’s eye, combines it with the 3D eye model and uses it to calculate gaze coordinates.

Eye-trackers offer various options for mounting cameras. For conference interpreting research, a desktop mount is most frequently used. The camera is usually placed below the screen that displays experimental stimuli. A tower mount includes a camera that tracks eye movements via a special mirror. As a result, the space in front of the study participant is free and large degrees of visual angle can be studied. To date, this type of mount has not been used in interpreting research. However, it would be applicable to studying the visual behaviour of interpreters who look also beyond the computer screen—for instance, when studying visual processing of the whole conference room. Some vendors offer such accessories as stands for mobile devices for studies that depict stimuli on a tablet or phone, or mounting brackets that help mount the eye-tracker on a laptop or monitor.

The desktop mount may be used in the head-stabilized or remote mode. In the former case, a forehead rest and a chin rest are used to stabilize the head. In the latter, a special target sticker may be placed on the participant’s head to make sure any head movements are accounted for in the data. Obviously, using a chin rest is not possible in interpreting research since it would make speech production virtually impossible. The use of a forehead rest is possible although it greatly compromises the ecological validity of the study, i.e. makes the study conditions less similar to natural interpreting conditions. Surprisingly, many researchers do not explicitly state which data collection mode (with head constraint or not) they applied in their studies.

Head-mounted eye trackers or eye-tracking glasses are also available on the market for research applications not involving a computer screen. To date, there is only one study that has used a wearable eye-tracker to collect data on conference interpreting (Chen 2017).

Eye-tracker precision is defined by the sampling rate, or the number of measurements of eye location per second. Desktop-mounted eye-trackers typically offer greater precision than wearable devices. Text-based studies require sampling rates not lower than 500–1000 Hz, while studies involving larger visual stimuli can be tracked with lower frequencies (e.g. 120 Hz).

Data recording and analysis can be done both in proprietary software and in external software that can be integrated with the eye-tracker. Currently, EyeLink offers the Experiment Builder program to design experiments and Data Viewer to analyse data. Tobii’s proprietary data recording and analysing software is Tobii Pro Lab. Depending on the eye-tracker, data collection and analysis may be integrated with third-party software and hardware, such as stimulus display software (e.g. E-Prime or PsychoPy), EEG and brain imaging software and other products, such as Translog, a key-logging system frequently used in written translation research.

Units of analysis and eye-tracking measures

The typical unit for analysing eye-tracking data is the ‘area of interest’ (AOI). The area of interest is a part of the visual area that is analysed. It might be a single word, a sentence, a slide or any other part of the visually examined area of any two-dimensional shape. Additionally, ‘interest period’ might be used as a unit of analysis. This refers to the temporal unit in which eye-tracking data is analysed. For instance, the experimenter might be interested in where visual attention is directed in simultaneous interpreting when numbers are presented on slides or someone enters the conference room (as in Seubert 2019 reviewed below).

When experimental stimuli are static (text, graphics), it is relatively easy to determine AOIs. For instance, DataViewer (data analysis program for EyeLink) does automatic segmentation of text stimuli into word-based AOIs. However, when the stimuli are dynamic (videos, non-static
Eye-tracking studies

visual space in front of the study participant—for instance, including a speaker at a rostrum and a screen with slides), AOIs have to be identified on a static image that is representative of the dynamic stimuli within a given temporal unit. For instance, Be Gaze (data analysis program for SMI) has a function of semantic gaze mapping to “map gaze data points from scene videos to reference images” (Chen 2017), while using Tobii this can be done manually or by employing the assisted mapping function.

Frequently used eye-tracking measures include number of fixations and average fixation duration, typically reflecting more effortful cognitive processing when these measures are higher in value (Holmqvist et al. 2015). These can be used in research including text stimuli or other visual stimuli (specific examples are reviewed in the subsequent sections). However, eye-tracking measures used in interpreting research are largely based on those used in reading research as an established standard. Eye-tracking measures for reading are frequently divided into early and late measures. The former typically include first fixation duration and gaze duration while the latter include go-past time and total reading time.

First fixation duration is the time initially spent by the eye on fixating an AOI, it reflects lexical access and processing difficulty when reading a word and is modulated by frequency and length (Liversedge et al. 1998). Gaze duration (also known as first-pass reading time or first-pass dwell time) is “the sum of all the fixations made in a region until the point of fixation leaves the region either to the left or to the right” (Liversedge et al. 1998: 58) and reflects early lexical processing. Go-past time (also known as regression path duration) is the sum of all fixation durations from the moment the eye first fixates the AOI to the moment the eye fixates to a region to the right of the AOI. It means that go-past time includes gaze duration and fixations to the left (i.e. refixations on previously fixated stimuli). This measure reflects integration of a viewed word into a sentence context (Liversedge et al. 1998). Finally, total reading time (also referred to as total dwell time, total gaze time or total viewing time) is the sum of all fixations on the AOI, including those during re-reading and indicates cognitive load. The same measures may sometimes be called different names and in accordance with the definitions adopted in the data analysis software of particular eye-trackers. To avoid confusion, it is always recommended to consult the eye-tracking measure definition given in the study description. Less frequently used measures might include regressions, i.e. movements back to previously read words or lines (Rayner 1998) and a regression rate understood as the “number of regressive fixations to previous AOIs divided by the total number of fixations” (Chen 2017: 143). Regressions index cognitive load in reading and their values typically increase with task difficulty, for example, when comparing reading to sight translation (Shreve et al. 2010).

Eye-trackers typically also record data on pupil dilation, which is considered to reflect local increase in cognitive effort (Jakobsen 2017). The first study employing pupillometry, or pupil diameter measurements, in interpreting studies was by Tommola and Niemi (1986), who compared cognitive load, indexed as pupil dilation, in interpreting sentences that either did or did not require syntactic restructuring. As expected, the need for restructuring induced higher cognitive load reflected through increased pupil diameter values. In another early study, Tommola and Hyönä (1990) compared pupil dilation in listening, sight translation and simultaneous interpreting. As expected, they found these tasks to be progressively demanding in terms of cognitive load. Experimental stimuli in another study by Hyönä et al. (1995) included both texts and single words with manipulated translatability (easy words with a single-word equivalent and difficult words that had to be translated by a phrase). Hyönä et al. (1995) found more pupil dilation in simultaneous interpreting as compared to listening and shadowing, in interpreting words into a B language as compared with interpreting into A and when interpreting difficult words. These results, especially regarding words as stimuli, show
that pupil dilation can be used as a sensitive index of cognitive load variations in interpreting tasks. Pupillometry has also been used in other studies (e.g. Gieshoff 2018; Seeber & Kerzel 2011, reviewed below), but it is less frequently applied than other measures. This might be due to various constraints. Pupil size is sensitive to many other factors, such as ambient light intensity, stimulants in the blood system (such as caffeine or drugs), fatigue, emotional arousal and head movements. Experiments thus have to be carefully controlled for such factors. Also, pupil size increase is not synched with a local problem leading to cognitive effort and usually happens with some delay. All these constraints make it difficult to interpret the effect of experimental stimuli on pupil size (Holmqvist et al. 2015).

**Methodological challenges**

Data quality is a common issue in eye-tracking studies. If eye movements are not tracked accurately, data might become distorted and study results skewed. Good calibration is thus of key importance. Although contact lenses and glasses should not be an obstacle, it sometimes happens that a bespectacled participant cannot be properly calibrated due to a scratched glass surface. Moreover, it might be impossible to obtain good quality data for other reasons, such as oculomotor disorders (a squint), bi-focal lenses or an idiosyncrasy (positioning of eyes). Participants are typically asked not to use heavy eye makeup and bring both their lenses and glasses to the lab just in case. If the participant’s vision is not properly corrected (the lenses are not strong enough), problems with calibration may also arise. Additionally, depending on their parameters and settings, eye-trackers might not be sensitive enough to accurately capture all relevant fixations, which is why manual inspection of data is recommended (Jensen 2008). Such inspection may also help discover drift, understood as a gradual asynchrony between recorded and true eye position progressing along data collection (Hvelplund 2014). Drift can be easily spotted on linear stimuli (text) and less easily spotted on more complex stimuli. It can be corrected manually or automatically. Because of all these potential problems, researchers should factor in approx. 20 per cent of data loss when designing their experiments (Jensen 2008; Saldanha & O’Brien 2014).

Although eye-tracking is not considered an invasive research method, some aspects of its data quality assurance might lower ecological validity, i.e. make research less similar to natural conditions. As mentioned earlier, eye-trackers can be in the form of glasses or stand-alone but requiring chin or forehead rests for good data quality. These aspects might lower the authenticity of the experimental task. Chen (2017) applied a strict criterion for assuring ecological validity of her study. As she used a head-mounted eye-tracker, she asked her participants to judge their comfort during the experiment. If participants declared below 50 per cent on the comfort scale, their data was rejected as not ecologically valid, which led to excluding 31 per cent of participants.

Stimulus design might also be problematic. When selecting complex stimuli (such as videos with a speaker, presentation slides or even whole environments), care should be taken to avoid irrelevant features that may attract visual attention due to their saliency (e.g. bright colours, large shapes). When using texts, many aspects also have to be considered for effective experimental design (Saldanha & O’Brien 2014), including font size and line spacing. Pilot studies help discover various problems that should be solved before collecting large amounts of data. One issue with texts is text length and text division. If a longer text is to be used in a study involving auditory and visual input (e.g. an 11-minute speech with approx. 1200 words as in Chmiel et al. (2020)), it has to be properly divided into excerpts to be shown on separate screens. AOIs with experimental items should be evenly distributed and preferably be placed to
Eye-tracking studies

avoid line breaks and ends of line. In a study on sight translation, Chmiel and Lijewska (2019) put experimental items on separate screens. When the participants were allowed to move to the next screen by pressing a button, there was some instances of data loss since some participants started reading the next screen ahead while still vocalizing the sight translation of the previous one. In the experiment proper, it was the experimenter who changed the screens when the sight translation of the currently visible screen finished.

Since finding professional interpreters and trainees for experimental studies in conference interpreting research is challenging (e.g. because of low numbers of trainees and high dropout rates, or due to the need for fair reimbursement for professionals), the experimental procedure should be streamlined and bug-free following a pilot study in order to minimize data loss.

Core issues

This section includes an overview of all currently available studies, known to the author, that have involved eye-tracking and conference interpreting-related tasks. Most of the studies focus on simultaneous interpreting and sight translation.

Reading patterns

Due to the dominance of visual input in sight translation, it is this hybrid mode of interpreting (see Bartłomiejczyk & Stachowiak-Szymczak, Chapter 2, in this volume) that has been most frequently researched with the use of eye-trackers. Initially, the studies compared reading patterns in sight translation and written translation. Later, scholars focused more on specific issues and problematic text elements.

Two seminal studies have compared reading in sight translation to reading in written translation. Dragsted and Hansen (2009) conducted a small-scale study and recorded more numerous fixations in written translation as compared to sight translation. They also found that reading patterns were more continuous, more linear and included fewer regressions in sight translation. This is not surprising since the production in sight translation has to be linear and quite fast. Similar conclusions come from the study by Jensen (2008), who compared reading patterns in reading for comprehension, reading for subsequent translation, sight translation and written translation. These tasks generated progressively higher values for fixation count, thus ranking sight translation as less demanding than written translation and more cognitively taxing than reading both for comprehension and for translation.

A more fine-grained comparison of sight translation, silent reading and reading aloud was conducted by Huang (2011), who used a series of eye-tracking measures. Early reading measures were similar for silent reading and sight translation and showed greater effort in reading aloud. However, late measures displayed a reverse pattern: rereading times and total viewing times were highest for sight translation as compared to silent reading and reading aloud, suggesting more need for integration at a later text processing stage. Huang (2011) also found that reading ahead (defined as viewing one sentence while sight translating a previous one) was applied to over 70 per cent of sentences. These results show that initial reading in sight translation is similar to regular silent reading while differences appear not in the comprehension stage but only in the reformulation stage.

A small-scale study by Su and Li (2019) focused on directionality in sight translation. Trainees participating in the study manifested greater processing load (higher values of task time, fixation duration and pupil dilation) when working into their foreign language. Su and Li (2019) also identified local problems generating greater cognitive load, such as low frequency
words and noun phrases requiring restructuring when sight translated between Chinese and English. In indirect confirmation of Huang’s (2011) results, they found higher cognitive load for these local problems in late and not early reading measures, suggesting again that initial reading in sight translation is similar to silent reading.

Kumcu (2011) examined reading patterns related to simultaneous interpreting. One group of interpreting trainees received a written text for a 5-minute preparation before simultaneous interpreting. Their eye movements were recorded when reading. They then performed simultaneous interpreting of that text but without the visual input, which is a rather unnatural scenario in professional practice. The other group had no preparation and received the text immediately before interpreting. Thus, they performed simultaneous interpreting with text and had their eye movements recorded then. In such a design, Kumcu (2011) directly compared the two types of reading: preparatory reading before interpreting and online reading during interpreting. He introduced incongruences between the written and the auditory text (such as added or omitted sentences, a changed order of sentences, changes in titles and proper names). Not surprisingly, he found the preparatory reading patterns to be more linear and the online reading patterns more erratic and either synchronized or non-synchronized with the auditory input. Synchronization disappeared especially due to incongruences between two inputs. Although greater cognitive load was numerically noticeable in the second group (the one performing simultaneous interpreting with text) in terms of more numerous and longer fixations, no statistically significant differences were obtained—most probably, as the author suggested—due to the low power of the experiment (only 12 participants). This study makes an important contribution towards characterizing reading types related to simultaneous interpreting.

**Multimodal processing**

The variety of visual stimuli that might be of relevance for the interpreter (see Kumcu’s typology above) makes the examination of multimodal input in interpreting a very challenging undertaking. To date, few scholars have embarked on this task.

The study that encompasses probably the widest variety of visual stimuli is that by Seubert (2019). It is an observational study, described by the author herself as exploratory, descriptive and qualitative. Seubert organized a conference simulation repeated 13 times in order to obtain eye-tracking data from 13 professional interpreters who worked in a booth under close-to-natural conditions. The participants could see the conference room, the listeners, the speaker and two screens: one showing the speaker (sometimes in close-up) and the other showing the slides that included images, data and quotations. Unsurprisingly, Seubert (2019) found that the interpreters looked at the slide immediately after it became available on the screen and alternated their visual attention between the speaker and the slides, with the majority of time spent viewing the speaker. Interestingly, when the speech became dense with numbers, facts or enumerations, there were many gaze points outside the typical areas of interest (the speaker and the slides)—as if the interpreters looked away from source-text-related stimuli to a visually-neutral area to cope with the heavy information load from the auditory channel. Looking away or closing one’s eyes was also frequent when processing quotations read out from the slides, which suggests cognitive overload. Seubert included a plethora of stimuli in her study (slides with various content, using a laser pointer by the speaker, unexpected events such as the speaker approaching one of the listeners or a person entering the conference room, speaker’s personal remarks, behaviour of the listeners, people walking past the booth) and found in general that professional interpreters were able to successfully deal with the variety of information, select the most relevant information and ignore visual input in the
Eye-tracking studies

case of cognitive overload. The study points to many interesting issues in visual processing during simultaneous interpreting. These can be further addressed in more focused experimental studies.

A more controlled experiment was conducted by Gieshoff (2018), who examined interpreter trainees (performing simultaneous interpreting) and translator trainees (performing listening) to see whether audio-visual input facilitates interpreting and reduces the cognitive load. Her participants listened to or interpreted a speech when looking at a static image of the speaker or a video of the speaker with visible lip movements. Since Gieshoff also manipulated background noise, she predicted that the cognitive load would be higher in simultaneous interpreting without video input since the interpreters would not be able to see lip movements to aid comprehension. She found an expected effect of task type since pupils dilated more during interpreting than during listening. However, pupil dilation was larger in simultaneous interpreting with audio-visual input as compared to the same task with static visual input only and remained unaffected by the level of background noise. Moreover, audio-visual input had no effect on self-reports or performance-based data. She interpreted these results by suggesting that in the case of audio-visual input, pupil dilation might reflect arousal rather than mental effort. Hence, audio-visual input in simultaneous interpreting might lead to greater arousal but not necessarily to greater mental effort or cognitive load.

Using a different paradigm, Stachowiak (2016) compared professional interpreters and trainees performing simultaneous and consecutive interpreting without notes while looking at the computer screen with a congruent or incongruent visual (a map of Poland and the emblem of Poland, respectively, when interpreting a general text about Poland). She additionally manipulated input as regards problem triggers (number, lists and narratives as the easiest type of input). As predicted, mean fixation duration was higher when processing numbers and lists (as compared to narratives) and mean fixation count was higher when dealing with an incongruent visual stimulus in both modes of interpreting. She also found a group effect, suggesting more cognitive effort on the part of trainees as compared to professionals.

Multimodal input in simultaneous interpreting with text was studied by Chmiel et al. (2020). They focused on the congruence and incongruence in multimodal input, i.e. differences between what the interpreter hears and what the interpreter sees in the written text of the speech to be interpreted simultaneously. They used proper names, numbers and control words as experimental items. Congruent items were identical in the audio input and the visually accessible text while incongruent items differed in these two input channels. According to professional standards, audio input should be prioritized and adhered to in the case of discrepancy (Pöchhacker 2004). Despite these standards, interpreters focused strongly on the visual input and interpreted items presented visually, even if those were incongruent with the auditory input. Study participants tended to look longer on items that were incongruent across the auditory and the visual modalities. In the congruent condition, there was no correlation between accuracy and skipping rate (i.e. the lack of fixation on the stimuli) or between accuracy and viewing times, suggesting no facilitatory effect of the visual input. In the incongruent condition, the less time and the shorter time the interpreters looked at the items incongruent with the auditory stimuli, the higher their accuracy was. This study shows that eye-tracking can be a useful method to broaden our knowledge about processing specific elements that are part and parcel of the interpreter’s actual work.

As can be seen from the overview above, studies examining the processing of multimodal input in interpreting involving eye-tracking are still rather scarce and employ various designs. They all show that eye-tracking can be successfully applied to shed more light on processing visual input in various modes of interpreting.
Syntactic manipulation

Eye-tracking has been used in a few studies involving syntactic manipulation, especially in sight translation. The stimuli in these studies usually include simple and more complex syntactic structures that can either be rendered in the target languages by a similar structure or have to be restructured. This restructuring is believed to trigger higher cognitive load traceable in eye movements.

Chmiel and Lijewska (2019) compared syntactic processing in sight translation performed by professional interpreters and trainees. The participants sight translated sentences with subject-relative clauses (e.g. The lady that kissed my uncle was a liar) and object-relative clauses (e.g. The lady that my uncle kissed was a liar)—the former considered to be more difficult to comprehend due to their syntax and more difficult to translate into Polish due to the need for reformulation. The study showed that professionals spent less time reading the whole sentence than trainees. The analysis of translation durations showed that professionals were generally faster than trainees and subject-relative sentences were processed more quickly than object-relative ones. The authors introduced a new measure called percentage of dwell time understood as the percentage of total translation time spent looking at the visual stimulus. The lower the percentage of dwell time, the more the participants looked away from the stimulus. It turned out that more difficult object-relative sentences generated lower percentage of dwell time, meaning that during the sight translation task the participants looked away more (probably to avoid syntactic interference when restructuring). The study shows that direct application of reading measures to the study of sight translation might be misleading. More difficult stimuli might generate shorter viewing times, which might not mean that they are easier to process but, paradoxically, that they are actually more difficult to process and need to be visually ignored to facilitate complex cross-linguistic processing.

A small-scale experimental study including a syntactic manipulation in a sight translation task was conducted by Shreve et al. (2010). The authors found expected effects of syntactic difficulty only in one experimental text used in the study and showed that performance in sight translation is prone to visual interference from the constantly available source text.

An interesting application of eye-tracking to syntactic processing in simultaneous interpreting is presented by Seeber and Kerzel (2011) who measured cognitive load when processing verb-final and verb-initial constructions. German verb-final constructions require restructuring when interpreted into English in which the verb has to precede the object. No reading measures could be used in the study since the input text was presented in an auditory modality only. Thus, Seeber and Kerzel (2011) used pupil dilation as an index of cognitive load. As expected, more difficult verb-final constructions that required asymmetrical processing generated larger pupil dilations, especially towards the end of the sentence. Similar German verb-final constructions were used by Korpal (2012) who asked interpreting trainees to sight translate either an English or a German text with the same content into Polish. The AOIs included structures with syntactic differences between English and German—German source texts generated more fixations than English texts due to the said verb-final structures. These two studies show how different eye-tracking measures applied in different interpreting modes present converging results as regards the processing of similar syntactic structures.

Numbers

Numbers are considered problem triggers in interpreting (Gile 2009: 171) due to their low redundancy and predictability, i.e. they cannot be easily inferred from context. They are also
highly informative and include, apart from the arithmetic value, such elements as the unit of measure, the order of magnitude and the context they refer to (Jones 2002). Because they tend to increase the memory load, in consecutive interpreting they are given priority in note-taking. In simultaneous interpreting, they are also frequently noted down either by the working interpreter or the booth partner. Due to their nature and frequent visual input, numbers have been the focus of several studies involving eye-tracking in interpreting.

Seeber (2012) examined how simultaneous interpreters attend to visuals when numbers are presented in the auditory input. His study participants interpreted a video presentation with numbers either gestured by the speaker (visual-spatial presentation) or shown on a screen next to the speaker (visual-verbal presentation). He analysed types of AOIs: the speaker’s head for facial expressions, the speaker’s torso for gestures and numbers on the slides. He found longer gaze durations on faces in the case of small numbers, on slides in the case of large numbers and no difference in viewing the speaker’s hands when processing small and large numbers. The study confirmed a long-held conjecture about the interpreters’ visually attending to speakers’ faces, which most probably facilitates comprehension. Interpreters were also found to actively search for information in the visual-spatial channel (slides) to support the input from the auditory channel. When such information was found, the viewing times were twice as long than when the information was not visually present (Seeber 2012). To the best of my knowledge, Seeber’s was the first study to apply eye-tracking to the multimodal processing of numbers in interpreting.

Korpal and Stachowiak-Szymczak (2018) also investigated processing numbers in simultaneous interpreting. Professional interpreters and trainees interpreted speeches accompanied with slides depicting bullet points with the most important information (including numbers and their context understood as the element the number referred to). Their AOIs included numbers and contexts separately. As predicted, they found longer mean fixation durations for numbers than for contexts and for the group of trainees as compared with professionals. The study shows that tracking eye movements and using even a single reading measure can shed more light on the cognitive load related to a specific problem trigger in interpreting. In a similar study Stachowiak-Szymczak and Korpal (2019) found no group effect in such measures as gaze time and fixation count. The only difference between professionals and trainees was found in mean fixation duration, suggesting longer fixations and greater effort expended on the part of trainees during visual processing of numbers.

The same task of simultaneous interpreting with slides was used by Korpal and Stachowiak-Szymczak (2020) in an extension of the previous study. Professional interpreters and trainees simultaneously interpreted speeches with a fast or slow delivery and looked at accompanying slides with the most important information (including numbers). The measures adopted to see how the source speech delivery rate modulates eye movements were fixation count per minute and the percentage of gaze time (i.e. the sum of all fixations on a given stimulus) devoted to reading numbers out of the gaze time for the whole slide. The effect of delivery rate was found for the fixation count for the whole slide and for AOIs including numbers. There were no group effects and no significant differences for the percentage of gaze time devoted to numbers. The authors explained these null results by claiming that the percentage of gaze time was not an adequate measure to see how eye movements are modulated by delivery rate. A faster source text did not make the participants look more at numbers than other visual elements, such as proper names and figures. Higher fixation counts in the fast delivery condition were explained by the scanning mechanism (Holmqvist et al. 2015), used by interpreters to search for supporting information to cope with overburdened working memory under temporal constraints in the fast delivery condition.
Other studies involving numbers confirm their status as problem triggers in interpreting. Chmiel et al. (2020) found that more visual interference was involved when processing numbers than control words in simultaneous interpreting with text. Texts involving numbers generate longer fixation durations than simple narratives (Stachowiak 2016) and slides with numbers can lead to such visual overload that interpreters tend to look away at visually neutral areas (Seubert 2019).

**Note-reading in consecutive interpreting**

The first study to use eye-tracking to analyse note-reading in consecutive interpreting (see Bartłomiejczyk & Stachowiak-Szymczak, Chapter 2, in this volume) is by Chen (2017), who examined professional interpreters performing Chinese-English and English-Chinese consecutive interpreting with notes. She wanted to compare note-reading (see Ahrens & Orlando, Chapter 3, in this volume) to other types of reading and to see how notes modulate cognitive load in the note-reading stage of consecutive interpreting. AOIs in her study were identified as note units (either a word, an abbreviation or a symbol). She used a number of measures and found that symbols were read faster (dwell times, first and average fixation durations were shorter) and triggered less cognitive load (smaller number of fixations and revisits) than words. When length was controlled in the comparison of full words and abbreviations, no effects emerged. Her data on average fixation duration (277 ms) are similar to those for oral reading (275 ms (Rayner 1998)), while regression rate (23 per cent) is higher than in reading (10–15 per cent (Rayner 1998)) and lower than in sight translation (26–35 per cent (Shreve et al. 2010)). All this shows that cognitive load in note-reading is higher than in silent reading and similar to sight translation since similar processes accompany reading in both modes. The language of notes turned out to have a significant effect only in B-A interpreting: notes in A language were easier to process than those in B most probably because A was the native tongue of the participants and also the target language in that particular interpreting direction. Interpreters did not have to search for translation equivalents since the notes were already in the target language. Chen’s (2017) ground-breaking study (which also included a detailed analysis of note-taking by applying an electronic pen) shows a trade-off between cognitive costs in the two stages of consecutive interpreting: interpreters decrease their cognitive load in the note-taking stage (which is speaker-paced and includes first exposure to content) and—as a result—increase their cognitive load in the note-reading stage (which is interpreter-paced and involves already known content). Chen’s study (2017) paves the way for further investigations of note-reading with the application of the eye-tracking method.

**New directions**

So far, researchers applying eye-tracking to the study of conference interpreting have mainly tapped into reading research and have typically used eye-tracking measures defined for reading. Since reading in interpreting differs by nature from regular silent reading or even reading aloud, not all measures can be successfully applied, as shown by Chmiel and Lijewska (2019), and should be reinterpreted or adjusted to the specificity of sight translation. Future studies may also include other eye-tracking measures that reflect cognitive load and have not been used so far in interpreting research, such as blink frequency and duration. Also, visuals in other modes of interpreting involve so much more than just the text, so it only seems inevitable that researchers will soon tap into other areas of study involving eye-tracking (such as video processing, interpretation of scanpaths, more advanced application of pupillometry).
Eye-tracking studies

Many studies have been exploratory in nature and already point to specific issues that could be tackled in better controlled experiments. In a truly mixed-methods approach, future studies can combine eye-tracking with self-reports and performance measures to get a more holistic and profound insight into the interpreting process. The studies on interpreting conducted so far and reviewed in this chapter confirm that eye-tracking can help shed more light into complex processes involved in this challenging task. Hopefully, as the field develops, more fitting eye-tracking measures will be identified as a golden standard in interpreting studies.

Further reading


References


