

Using eye-tracking in applied linguistics and second language research

Second Language Research

2016, Vol. 32(3) 453–467

© The Author(s) 2016

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0267658316637401

slr.sagepub.com



Kathy Conklin and Ana Pellicer-Sánchez

University of Nottingham, UK

Abstract

With eye-tracking technology the eye is thought to give researchers a window into the mind. Importantly, eye-tracking has significant advantages over traditional online processing measures: chiefly that it allows for more 'natural' processing as it does not require a secondary task, and that it provides a very rich moment-to-moment data source. In recognition of the technology's benefits, an ever increasing number of researchers in applied linguistics and second language research are beginning to use it. As eye-tracking gains traction in the field, it is important to ensure that it is established in an empirically sound fashion. To do this it is important for the field to come to an understanding about what eye-tracking is, what eye-tracking measures tell us, what it can be used for, and what different eye-tracking systems can and cannot do. Further, it is important to establish guidelines for designing sound research studies using the technology. The goal of the current review is to begin to address these issues.

Keywords

applied linguistics, eye-tracking, second language, SLA

I Introduction

Many of us use a car to get from A to B with very little understanding of the underlying mechanics of the car's operation. As technology has gotten more advanced, this has become a common phenomenon: We use things without understanding how they work. As new technologies are introduced into our research, we run the same risk of using new, seemingly innovative tools, without stopping as a field to evaluate them and without having a discussion about how they work, how to use them, and what they can realistically tell us. One of these new technologies, eye-tracking, is becoming increasingly popular in applied linguistics, second language acquisition (SLA) and testing to investigate topics

Corresponding author:

Kathy Conklin, School of English, University of Nottingham, Nottingham, NG7 2RD, UK.

Email: kathy.conklin@nottingham.ac.uk

that were traditionally examined using off-line measures (e.g. paper and pencil tests, judgment tasks, think aloud protocols, interviews, etc.). The growing interest in eye-tracking is evidenced by a special issue in *Studies in Second Language Acquisition* (Godfroid et al., 2013) and an increasing number of studies making use of eye-tracking (e.g. Bax, 2013; Godfroid et al., 2013; Pellicer-Sánchez, 2015; Siyanova-Chanturia et al., 2011; Winke, 2013; Winke et al., 2013). However, before we get too far down the road with eye-tracking, we think that it is important for the field to come to some understanding about what eye-tracking is, what eye-tracking measures tell us, and what it can be used for. We also think that it is valuable to discuss the practicalities of using eye-tracking.

II What is eye-tracking and what do eye-tracking measures tell us?

Many researchers interested in language processing make use of eye-tracking technology to monitor the eye when reading and when looking at a static scene or video while listening to auditory input. Eye-tracking is primarily used to detect and measure an eye's movements (saccades) and stops (fixations), as well as movements back in a text when reading (regressions). We will start by giving a brief overview of these terms before moving on to a discussion of what eye-tracking can tell us and the measures that are often used to report it.

Work by Rayner (1978, 1998) summarizes a lot of what is known about typical eye movement behavior. Eye movements that bring a new region of text or image into view are called saccades. In silent reading saccades take around 30 ms and are about 8 characters in length, which is reduced to approximately 6 characters in oral reading. In visual scenes, saccades cover a larger distance and average 40–50 ms. The stops between saccades are referred to as fixations and last roughly 225 ms in silent reading and 275 ms in oral reading. In visual scenes, fixations have a duration of approximately 330 ms. Movements back in a text (right to left movement) are called regressions and comprise about 10%–15% of the reading time in skilled readers. They can arise due to a number of comprehension difficulties (shorter regressions are due to overshooting a target or difficulties processing the currently fixated word; longer regressions are due to ambiguity, general text difficulty, integration, etc.). It is important to keep in mind that these are simple averages. Rayner points out that we can see saccades from 2 to 18 or more characters and fixation durations from 100–500 ms for a single reader within a single passage.

Eye-tracking has been predominately used by cognitive psychologists and psycholinguists. From its early days it 'was hailed as an opportunity for the researcher to "look into the mind" of the subject' (Rayner, 1978: 618). Such statements may lead us to believe that eye-tracking tells us more about what is going on in the mind than it realistically can. In actual fact eye-tracking, similar to other techniques that measure reaction times, has two key underlying assumptions (Pickering et al., 2004). First, the amount of time spent fixating an item reflects the cognitive effort required to process it, meaning that longer durations and more fixations indicate greater processing effort and shorter fixations and/or skipping indicate less processing effort. The second supposition is that what is being fixated is what is being considered.

There are two significant advantages of eye-tracking over other traditional techniques that measure response times or reading times. First, because eye movements are a natural part of reading and viewing, eye-tracking can be done without secondary tasks (e.g. a button press in self-paced reading) that are often subject to strategic effects. Thus, eye-tracking is often touted as allowing for natural reading or viewing. Importantly, as eye-tracking systems become more advanced, so that they no longer require participants to wear head-mounted gear or sit with chin rests, eye-tracking may become very close to natural reading and viewing on electronic devices. Second, eye-tracking provides a very rich record of reading behavior. As will be demonstrated below when discussing different measures, eye-tracking allows us to quantify what happens when a word or region is encountered. For example, we can ascertain how many times, how long and when a word is fixated during reading. We can also see where readers go back in a text when they are having difficulties.

Eye-tracking data is reported in terms of number of fixations in a region of interest (ROI, sometimes referred to as area of interest, or AOI), which are reported as counts, the duration of fixations that calculate the amount of time spent in an ROI, and the likelihood or probability of fixating an ROI. Data is often reported in terms of ‘early’ and ‘late’ measures (e.g. Altarriba et al., 1996; Staub and Rayner, 2007). Early measures (e.g. likelihood of skipping, first fixation duration, first pass reading time) tap into automatic processes and the initial stages of processing (e.g. lower-level processes like word recognition in reading). Late measures (e.g. rereading, second pass reading time, total reading time, fixation count) reflect strategic processing and include revisits and reanalysis that result from processing difficulty. Thus, they signal more effortful and/or conscious processing (e.g. lexical integration in reading). Measures like regression path duration and regression rate are sometimes said to tap into early processing and sometimes late processing, because they are indicative of a difficulty integrating a word when it is fixated, arguably an early effect, but at the same time they reflect the time overcoming this difficulty, arguably a late effect (Clifton et al., 2007). When examining regressions it is important to keep in mind that most regressions are short, simply taking the eye back to the previous word and are likely due to oculomotor errors from overshooting the intended word (Rayner et al., 2006). The far less frequent, longer regressions back in a text largely reflect comprehension failures.

In order to exemplify the measures used in eye-tracking, Figure 1 provides an example sentence and its ROI shaded in grey, with an illustrative eye movement pattern shown below it. In this example the ROI is an idiom ‘spill the beans’. ROIs can also be single words. In fact, for an idiom it would also be important to explore ‘beans’ as its own ROI to determine if the predictability of idiom completions influence word skipping. Below the sentence and eye movement pattern, the different eye-tracking measures are presented and related to the example. As the figure demonstrates, some measures are standardly reported, while other measures may only be applicable, or inapplicable, in certain contexts. For example, it is only appropriate to report first fixation duration for single words for very short ROIs. Thus, while first fixation duration is exemplified here, it would not be appropriate to report for such a long ROI.

While numerous studies have used eye-tracking during reading, it is also commonly used to monitor eye movement behavior when participants look at static images, view

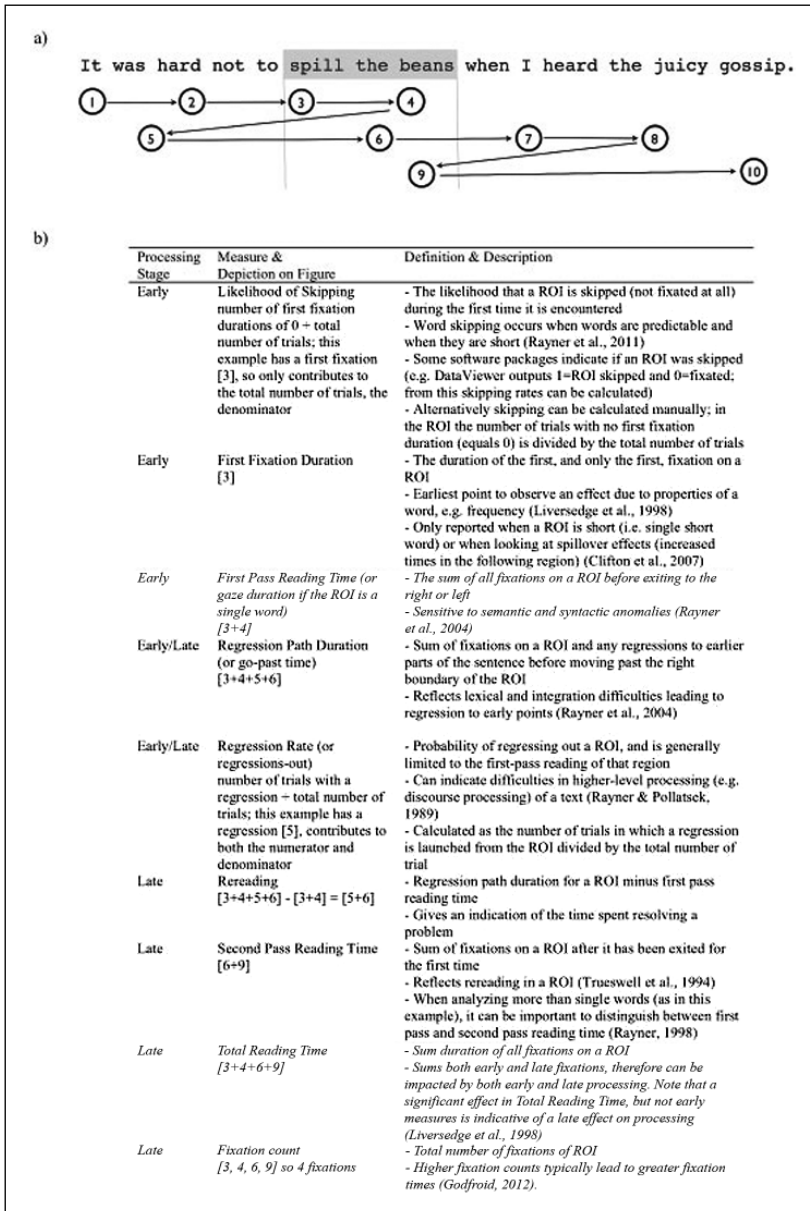


Figure 1. (a) An example sentence with a region of interest (ROI) appears in black. An illustrative eye movement pattern is depicted below the sentence, with fixations indicated by circles and the number indicating their order. (b) The eye movement pattern depicted in the sentence is related to eye-tracking measures. First, the processing stage is classified as ‘early’ or ‘late’. Second, the eye-tracking measure is given with a description of how it is calculated, as well as relating it to the fixation numbers in (a). Finally, the eye-tracking measures are defined and described. The most commonly reported eye-tracking measures are in italics, and the less commonly reported ones are highlighted in black.

visual scenes (often referred to as the visual-world paradigm), or watch a video. In such research, eye movements and fixations are time-locked to a particular linguistic variable (a word, pronoun, syntactic structure, etc.) that is presented auditorily. The different areas of the screen are defined as ROIs. Data is usually reported in terms of total fixation duration, the proportion of saccades directed to, or the proportion of time spent looking at a target ROI compared to competitor ROIs. For example, Chambers and Cooke (2009) presented English–French bilinguals with a screen containing static images of a swimming pool, a chicken (*‘la poule’* is a near homophone of pool), and two distractor objects (a strawberry and a boot). Participants listened to sentences like, *‘Marie va décrire la poule’* (*‘Marie will describe the chicken’*), and their eye movements were monitored time-locked to the noun *‘poule’*. Both the mean number of saccades and the probability of fixating the target (pool) and near homophone (chicken) competitor were calculated. ROIs can be defined for other types of visual material like videos. For example, Bisson et al. (2014) had participants watch a movie under different subtitling conditions and defined two ROIs (the subtitle region and the image region) to see how manipulating the language of the audio and the subtitles influenced looks to the subtitle region.

Many of us may have seen eye-tracking data presented as ‘heat maps’, as it is a very poignant way to present summary data. Most eye-tracking analysis packages allow researchers to easily generate a heat map of fixations, with reds and yellows indicating ‘hot zones’ of large numbers of fixations, greens and blues indicating ‘cool zones’ of few looks, and unshaded areas of no looks. While these heat maps may have a ‘wow factor’ and provide an overview of data, they do not in any way constitute an analysis of data, and therefore should be treated with extreme caution. As Bojko (2009) points out, heat maps are often used inappropriately, for the wrong reasons, and do not include the information that is necessary to interpret them. (Readers interested in the use of heat maps should refer to Bojko (2009), as it explains how heat maps are created and provides guidelines for their use.) Similarly, most eye-tracking systems allow researchers to play videos of participants’ eye movements for each stimulus superimposed over the text or image. While this can be informative when looking for unusual looking patterns or track loss, they do not provide the reading time or fixation data required to perform statistical analyses (Keating, 2014).

III What can eye-tracking be used for?

Eye-tracking can be used to examine a variety of phenomena in language processing. This section provides an overview of the linguistic phenomena that have been investigated using eye-tracking, and which we divide into three main areas: (1) auditory processing (listening); (2) visual word processing (reading); and (3) simultaneous auditory and visual processing.

I Auditory processing: Listening

Eye-tracking has been used to explore the processing of auditory stimuli. A common topic in this line of research has been the investigation of auditory word recognition and lexical activation using a visual scene or the visual-world paradigm. In this procedure

participants' eye movements are monitored when they hear auditory stimuli, usually a word or a short instruction containing a critical word, while they are presented with visual representations (e.g. objects, scenes, etc.). For example, in order to investigate whether a non-target language is activated during bilingual auditory processing, Spivey and Marian (1999) presented Russian–English bilinguals with images of a marker, a stamp – ‘marka’ in Russian and a potential phonological competitor if first-language (L1) Russian is activated while processing in second-language (L2) English – and distractor objects. They tracked participants' eye movements when they heard instructions like ‘Put the stamp below the cross.’ To determine whether knowledge of Russian influenced looking patterns, they examined the proportion of looks to the distractor objects and the potential phonological competitor (‘marka’). A higher proportion of looks to the stamp (‘marka’) would indicate that the L1 is activated in L2 listening.

This paradigm has been used to investigate a range of questions in second language research such as: whether bilinguals activate both languages (non-selective access) or only the target one (selective access) (e.g. Chambers and Cooke, 2009; Spivey and Marian, 1999), the processing of grammatical gender (e.g. Doherty, 2013; Dussias et al., 2013), and anaphor resolution (e.g. Doherty, 2013; Roberts et al., 2008). In all cases, eye movements are monitored time-locked to a particular auditory stimulus, be that a word, gender-marked article, pronoun, etc.

2 *Visual word processing: Reading*

Eye-tracking while reading has been used extensively to address a number of questions. In such studies participants are presented with phrases, sentences, stretches of discourse, or a full text (although presented across separate screens). In each case researchers define ROIs depending on their research questions. For example, to investigate language activation in visual word recognition in bilinguals, Duyck et al. (2007) presented sentences containing identical or nonidentical cognates or control words. Some of the reading time measures introduced in Figure 1 (first fixation duration, gaze duration, go-past time, and skipping rates) were compared for the three word types (ROIs in the study) to determine whether processing was speeded for cognates. Processing time for cognates was less than for control words in the early measures, which was taken as an indication that cross-linguistic lexical interactions occur early during visual word recognition.

As the discussion about Figure 1 indicated, eye-tracking while reading can be used to investigate processing beyond the level of individual words. In this case an ROI of more than one word is specified. For example, Carrol and Conklin (2015) compared reading times for idioms (‘spill the beans’) and a matched non-idiomatic control phrase (‘spill the chips’) that were embedded in sentences. Crucially, in many cases such ROIs may be further subdivided to more fully explore a range of research questions. For example, to examine the processing of idioms as a whole, as well as before and after their recognition point, Siyanova-Chanturia et al. (2011) defined three ROIs: whole idioms (‘as cold as ice’), before the recognition point (‘as cold’) and after (‘as ice’).

The eye movement record can also be used to explore questions about morphosyntactic processing (for a review of second language acquisition studies on syntactic processing using eye-tracking and their methodologies, see Frenck-Mestre, 2005; Keating and

Jegerski, 2014). For example, it has been used to investigate the processing of syntactic ambiguity (e.g. Frenck-Mestre and Pynte, 1997), sensitivity to tense incongruences (e.g. Sagarra and Ellis, 2013), grammatical gender processing (e.g. Doherty, 2013; Keating, 2009), anaphor resolution (e.g. Doherty, 2013; Doherty et al., under review; Roberts et al., 2008), and the processing of reflexive pronouns (e.g. Felser and Cunnings, 2012).

When investigating syntactic processing it is important to keep in mind, as Clifton and Staub (2011) point out, that difficulties can show up at various points in the eye-tracking record. Effects of syntax are seldom apparent at the very earliest stages of reading, and therefore first fixation time is rarely reported; however, they are occasionally found (e.g. Staub, 2007). Effects can show up as increases in first pass reading times. Sometimes effects of syntactic processes may only show up as an increase in the frequency of regressions, increased go-past times, increased second pass reading times, or as increased times in the following region (spillover effect). Because effects can be found in many measures and in subsequent regions, it is worth inspecting data for a wide range of measures and regions following the ROI.

It is important to note that when investigating questions on syntactic processing ROIs can be very short, as in the case of pronouns. On the assumption that very short words (e.g. ‘he’, ‘she’, ‘it’, ‘they’) will be seen via parafoveal view on the previous word (i.e. the ‘preview’ our eye gets while fixated on a previous word), an ROI that is larger than the critical word itself is usually defined. For example, when looking at the processing of ‘them’ used as a singular pronoun by native and non-native English speakers, Doherty et al. (under review) defined three ROIs (verb+pronoun, adverb, and final) as in the following example (1).

(1) Mr Jones was looking for the station. He saw someone on the other side of the road so he crossed over and |asked him/her/them_(verb+pronoun)| politely_(adverb)| where the station was_(final).

As example (1) shows, the critical pronoun ROI encompassed the previous word. Two subsequent ROIs were also defined to investigate any possible spillover effects. Further, as this example illustrates, sometimes a region of interest might differ in length for a critical manipulation (‘him’/‘her’ are one letter shorter than ‘them’), thus making comparisons problematic. Character length corrections (e.g. dividing a reading time measure by the number of characters) are questionable when there are only small differences in length, as in this example. In such cases it is preferable to use a residual reading time; for a discussion as well as an explanation of calculating residual reading times, see Trueswell et al. (1994).

Eye-tracking can also be used to investigate questions about language acquisition, such as learning words (e.g. Godfroid et al., 2013; Pellicer-Sánchez, 2015; Williams and Morris, 2004), and learning morphosyntax (e.g. Godfroid and Uggen, 2013; Winke, 2013). Similar to the paradigms discussed above, when looking at questions of acquisition researchers define critical ROIs and evaluate relevant eye-tracking measures. In these studies both early and late measures are generally reported for the ROIs themselves. In addition, researchers can explore the relationship between reading time for the ROIs (generally using a late measure like total reading time) and performance on a post-reading test. For example, Pellicer-Sánchez (2015) monitored participants’ eye

movements while they read a short story containing ‘new’ words (nonword ‘tloser’ meaning ‘noise’) and known, highly frequent control words (‘garden’), which were repeated several times throughout the story with each occurrence defined as an ROI. To examine whether there was a speed up for unknown words over the course of the experiment and to see if their processing became more ‘word-like’, both early (first fixation duration, gaze duration) and late measures were analysed (fixation count, total reading time). Total reading time was used to explore whether the amount of time reading contributed to different dimensions of their acquisition (form, meaning recognition and meaning recall).

Two recent studies have attempted to use eye-tracking in new and innovative ways. In order to investigate learners’ noticing of corrective feedback during online interactions, Smith (2012) asked learners to recall a story during a learner–tutor computer-assisted interaction with the instructor providing full recasts. Learners’ eye movements on the instructor’s recasts were recorded to examine the ‘noticing’ of corrective feedback. Heat maps were used as a metric of noticing in the recasts. Additionally, eye-tracking was used to investigate testees’ behavior when taking online reading tests (Bax, 2013; Bax and Weir, 2012). Learners were presented with a reading passage on the left side of the screen and questions on the right side. Reading behavior for correctly and incorrectly answered questions was compared in terms of looking patterns (first fixation duration and count and total reading time and count) to the whole text, to sections of the texts, and to words and phrases within those sections. While these studies have the potential to provide valuable insight into SLA and testing, their results should be interpreted with caution for a number of reasons: non-standard experimental design, in particular due to the lack of appropriately defined and controlled ROIs; data that is not suitable for statistical analyses (e.g. heat maps); and eye-tracking systems that are not sensitive enough to explore reading without a prohibitively large number of stimuli (an issue discussed below).

3 *Simultaneous auditory and visual processing*

Other studies have looked at the processing of visual and auditory linguistic stimuli in combined activities. For example, Bisson et al. (2015) used eye-tracking to investigate the impact of auditory and pictorial information on the incidental acquisition of foreign language vocabulary. Participants saw an image in the middle of the screen, with the corresponding foreign word below and a letter above the image. Participants’ eye movements on the three ROIs (i.e. image area, letter area, and word area) were recorded while completing a letter-search task and were explored in relation to performance on vocabulary learning tasks. As touched on above, learners’ processing of subtitles while watching and listening to films has also been explored using eye-tracking (Bisson et al., 2014; Montero Perez et al., 2015; Winke et al., 2013). In these studies, the subtitle area or parts of it are the ROI. Generally late measures like fixation count and total reading time are analysed for the ROIs and can be used to determine the relationship between reading time and language background or content familiarity on performance on post-viewing tasks. In such studies the image can also be defined as an ROI. For any analysis, the image area needs to be time-locked to the subtitle area and the audio signal. It is a challenge to

compare eye movements to various data sources, because they are different sizes, and have differences in salience and visual complexity. One way to adjust for size differences is to ‘normalize’ the data. To do this we calculate the number of pixels in each ROI and look at fixation duration or count by number of pixels. Such a comparison is still problematic and should be treated with caution. Similarly eye-tracking has been used to explore looking patterns for other types of visual materials like video-based second language listening tests and visual aids to explore their impact on performance tasks (e.g. Suvorov, 2014). When using visual stimuli, it is important to keep in mind, as with all eye-tracking studies, that ROIs need to be clearly defined and the stimuli well designed and controlled so that appropriate analyses are possible.

IV A practical ‘how to’ guide

In this final section we will look at the practicalities of doing eye-tracking research. We will discuss issues to consider when doing eye-tracking studies and briefly look at what kinds of research different eye-trackers are appropriate for.

1 Check the characteristics of the eye-tracking equipment

There are many different types of eye-trackers, all of which have different specifications, making them more or less appropriate to investigate particular language phenomena. It is essential to ensure that a system can provide the data needed to answer the research questions. For example, an eye-tracker that provides a more ‘natural’ reading experience (not head mounted, no chin rest) and is more tolerant of head movements (important when testing children), will often sacrifice precision in locating eye movements. Such systems are good enough to capture eye movements to sections of a screen (e.g. an image); however, they are often not accurate enough for ROIs of individual words or short phrases. In general, better accuracy and resolution is obtained with head supported systems and/or by using a chin rest, by doing monocular recording (instead of binocular), and by having a higher sampling rate. Most of the research done on reading uses eye-trackers with sampling at rates of 500 Hz or 1,000 Hz; however, imprecision is not usually a problem for eye-trackers that operate at 200 Hz. While systems with lower sampling frequencies can be used for reading, they will require an impractical amount of data to overcome the imprecision introduced by the sampling frequency. For a method of calculating the number of data points (number of participants and trials) needed for different sampling frequencies, see Andersson et al. (2010).

2 Familiarization with the functioning of the eye-tracker and its software

Being able to calibrate an eye-tracker well, so that accurate data is collected, is important. Failure to obtain good calibration will result in imprecise data. The usual procedure is to include a nine-point calibration at the beginning of an experiment, at additional pre-defined points in longer experiments, and when there is evidence of eye-drifting. An eye-drift correction is usually included at the beginning of each experimental trial or screen. Most eye-tracking systems come with software for experimental design and

analysis that have quite different interfaces (e.g. Experiment Builder and Data Viewer for EyeLink), which can take a considerable amount of time to become acquainted with. It is always important to do a test run of an experiment to make sure that data is being outputted properly.

3 Carefully consider the stimuli

Because they have been shown to influence fixation time, critical stimuli should be well matched for issues such as orthographic regularity, orthographic familiarity, word class, length, frequency, predictability, word frequency, word familiarity, lexical ambiguity, and syntactic structure. Studies often benefit from a control condition or stimuli, which provide a baseline. Control stimuli should be closely matched to experimental stimuli.

Critical stimuli should appear in contexts that are the same or as similar as possible, so that differential contexts do not drive effects. For example, stimuli should be in sentences that have the same number of words, are in the same syntactic frames, and are equated for predictability/bias (see Figure 2). Further, if there is any chance of spillover effects occurring, the region following the critical stimuli should be the same or carefully matched. Critical stimuli should be presented in similar positions, as reading speed generally decreases as a reader progresses through a text, and words at the end of a sentence are read more slowly, as are sentences at the end of a passage. Since studies have shown that the appearance and layout of stimuli can affect performance, it is important to ensure that the font is an appropriate and that its size is not too small to acquire accurate data (e.g. New Courier font, as all letters take the same amount of horizontal space). Additionally, eye-trackers are less accurate at measuring vertical eye movements. To make it easier to determine which line of text is being read lines should be double-spaced. Finally, when presenting longer texts that appear across multiple screens, the screens should have equivalent lengths with stimuli appearing in similar positions.

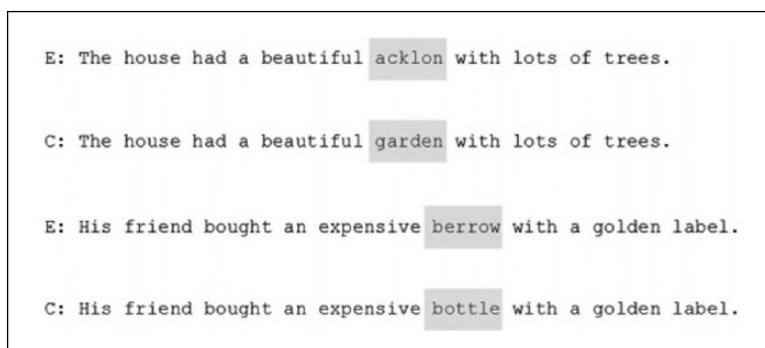


Figure 2. Example experimental (E) and control (C) sentences, with regions of interest (ROIs) shaded in grey. ROIs are controlled for length, frequency, position in the sentence, and type of syntactical structure in which they are embedded.

4 *Controlling non-linguistic visual stimuli*

There are certain features that need to be controlled for when presenting images. We tend to scan a screen with images from left to right (for languages where reading is from left to right), so it is important to counterbalance the location of objects. For example, if condition *x* always appears on the right side of the screen and condition *y* on the left side, condition *y* would likely always be fixated first, not because of the experimental manipulation, but because of its position on the screen. In addition, the images should be matched for size and salience as these have been found to influence looking patterns.

5 *Creating ROIs*

While eye-tracking software can output values for all of the reading time measures for all of the individual words, as well as larger regions, doing this could result in an unmanageable amount of data. Before conducting a study, it is important to think through what are sensible ROIs in order to answer the research questions. Many of the factors that should be considered when creating the ROIs have been mentioned above.

1. The ROIs for experimental and control conditions should be well matched, or analysis techniques that can deal with any variability in materials will need to be employed (e.g. mixed-effects modeling or multiple regression analyses).
2. If there may be spillover from a critical ROI, a successive ROI that is well matched across conditions should be established.
3. Short ROIs (e.g. a pronoun) may be skipped due to preview, so the ROI should encompass a preceding word.
4. Long ROIs may need to be further subdivided into smaller ROIs to explore factors such as skipping, which are not appropriate for longer regions.

While most analysis programs will automatically find ROIs of individual words, it is possible for researchers to ‘draw’ their own. When doing this it is important to keep in mind that ROIs should be the same across conditions (i.e. include the same amount of white space around all ROIs).

6 *Determining and reporting suitable eye-tracking measures*

As the previous discussion and Figure 1 highlighted, effects of specific manipulations are likely to be apparent in certain eye-tracking measures. To fully explore data, and to ensure that effects are detected, researchers should output data for all of the potentially relevant measures. At the same time, the selection and reporting of measures should be principled and/or theoretically justified. For example, when looking at the processing of a longer phrase like an idiom (‘spill the beans’), one ROI should encompass the entire idiom for which it would be appropriate to output and report one set of measures (first

pass reading time, total reading time and fixation count), while a second ROI should encompass only the final word of the idiom for which it would be appropriate to output and report another set of measures (likelihood of skipping, first fixation duration, first pass reading time, total reading time and regression path duration). For a detailed discussion of determining suitable eye-tracking measures for multi-word units like idioms, see Carrol and Conklin (2014).

7 *Data cleaning procedures*

Before analysing data it needs to be 'cleaned'. Trials with track loss are removed. More precisely, instances where the camera 'loses' the pupil are removed, which is often operationalized as trials having zero first pass reading times in two consecutive regions. Overly short and long times are removed. This is done in various ways. Some researchers merge short fixations (less than 50ms) to a longer one that is within 1° of it. Other researchers remove fixations that are shorter than 50–100ms and longer than 800ms. Another common procedure is to discard all data points that are 2.5 standard deviations from a participant's mean in a given condition. Whatever choices are made in the cleaning procedure, it is important to report them clearly, as well as how much data is lost due to it, and whether this loss was equivalent across conditions. After cleaning the data, it is ready to analyse.

8 *Consider the constraints of eye-tracking*

While eye-tracking has been hyped as allowing for 'natural' reading, this does not mean that we can simply present participants with 'real' material (e.g. a newspaper article, TOEFL/IELTS reading passage, etc.) and draw conclusions directly from the different reading time measures for individual words or sentences. It is important to keep in mind that when experimental materials have not been carefully controlled and designed, along the lines of what has been detailed above, the reading record will be contaminated by these factors, thereby calling into question any conclusions that are drawn from the data. If researchers would like to use 'real' material, analyses of the eye-tracking measures need to be subjected to mixed-effects modeling or multiple regression analyses. Such analyses allow researchers to account for the factors detailed above when they have not been explicitly controlled or manipulated (e.g. Carrol and Conklin, 2015; McDonald and Shillcock, 2003; Siyanova-Chanturia et al., 2011).

V **Conclusions**

Eye-tracking is becoming increasingly popular in applied linguistics and second language research. In some cases it is replacing well-established methods, which have standardized procedures for their implementation that have been validated in the field. In other cases, eye-tracking is simply being used to augment standardized methodologies. In either case, it is important for eye-tracking to be standardized in the field so that we can draw meaningful conclusions from it.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Altarriba J, Kroll J, Sholl A, and Rayner K (1996) The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory and Cognition* 24: 477–92.
- Andersson R, Nyström M, and Holmqvist K (2010) Sampling frequency and eye-tracking measures: how speed affects durations, latencies, and more. *Journal of Eye Movement Research* 3: 1–12.
- Bax S (2013) The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing* 30: 441–65.
- Bax S and Weir C (2012) Investigating learners' cognitive processes during a computer-based CAE reading test. *Cambridge ESOL: Research Notes* 47: 3–14.
- Bisson M-J, Van Heuven W, Conklin K, and Tunney R (2014) Processing of native and foreign language subtitles in films: An eye-tracking study. *Applied Psycholinguistics* 35: 399–418.
- Bisson M-J, Van Heuven W, Conklin K, and Tunney R (2015) The role of verbal and pictorial information in multi-modal incidental acquisition of foreign language vocabulary. *Quarterly Journal of Experimental Psychology* 68: 1306–26.
- Bojko A (2009) Informative or misleading? Heatmaps deconstructed. In: Jacko J (ed.) *Human-computer interaction: New trends*. Berlin/Heidelberg, Germany: Springer, pp. 30–39.
- Carrol G and Conklin K (2014) Eye-tracking multi-word units: Some methodological challenges. *Journal of Eye Movement Research* 7: 1–11.
- Carrol G and Conklin K (2015) Cross language priming extends to formulaic units: Evidence from eye-tracking suggests that this idea 'has legs'. *Bilingualism: Language and Cognition*. Epub ahead of print 20 April 2015. DOI: 10.1017/S1366728915000103.
- Chambers C and Cooke H (2009) Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35: 1029–40.
- Clifton C and Staub A (2011) Syntactic influences on eye movements in reading. In: Liversedge S, Gilchrist I, and Everling S (eds) *The Oxford handbook of eye movements*. Oxford: Oxford University Press, pp. 895–909.
- Clifton C, Staub A, and Rayner K (2007) Eye movements in reading words and sentences. In: van Gompel R (ed.) *Eye movements: A window on mind and brain*. Amsterdam, the Netherlands: Elsevier, pp. 341–72.
- Doherty A (2013) The use of gender and number cues in L2 pronoun processing. Unpublished doctoral dissertation, University of Nottingham, Nottingham, UK.
- Doherty A, Conklin K, and van Heuven W (under review) Processing of singular-them by native and non-native English speakers.
- Dussias P, Valdes Kroff J, GuzzardoTamargo R, and Gerfen C (2013) When gender and looking go hand in hand. *Studies in Second Language Acquisition* 35: 353–87.

- Duyck W, Van Assche E, Drieghe D, and Hartsuiker R (2007) Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33: 663–79.
- Felser C and Cunnings I (2012) Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics* 33: 571–603.
- Frenc-Mestre C (2005) Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second Language Research* 21: 175–98.
- Frenc-Mestre C and Pynte J (1997) Syntactic ambiguity resolution while reading in a second and native languages. *The Quarterly Journal of Experimental Psychology* 50A: 119–48.
- Godfroid A (2012) Eye tracking. In: Robinson P (ed.) *The Routledge encyclopedia of second language acquisition*. New York/London: Routledge, pp. 234–36.
- Godfroid A and Uggen M (2013) Attention to irregular verbs by beginning learners of German. *Studies in Second Language Acquisition* 35: 291–322.
- Godfroid A, Boers F, and Housen A (2013) An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye tracking. *Studies in Second Language Acquisition* 35: 483–517.
- Godfroid A, Winke P, and Gass S (eds) (2013) Special issue: Eye-movement recordings in second language research. *Studies in Second Language Acquisition* 35: 205–422.
- Keating G (2009) Sensitivity to violations of gender agreement in native and non-native Spanish: An eye-movement investigation. *Language Learning* 59: 503–35.
- Keating G (2014) Eye-tracking with text. In: Jegerski J and Van Patten B (eds) *Research methods in second language psycholinguistics*. London: Taylor and Francis, pp. 69–92.
- Keating G and Jegerski J (2014) Experimental designs in sentence processing research. *Studies in Second Language Acquisition* 37: 1–32.
- Liversedge S, Paterson K, and Pickering M (1998) Eye movements and measures of reading time. In: Underwood G (ed.) *Eye guidance in reading and scene perception*. Oxford: Elsevier, pp. 55–75.
- McDonald S and Shillcock R (2003) Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research* 43: 1735–51.
- Montero Perez M, Peters E, and Desmet P (2015) Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal* 99: 308–28.
- Pellicer-Sánchez A (2015) Incidental L2 vocabulary acquisition *from* and *while* reading: An eye-tracking study. *Studies in Second Language Acquisition*. Epub ahead of print 18 August 2015. DOI: 10.1017/S0272263115000224.
- Pickering M, Frisson S, McElree B, and Traxler M (2004) Eye movements and semantic composition. In: Carreriras M and Clifton C (eds) *On-line study of sentence comprehension: Eye tracking, ERPs and beyond*. New York: Psychology Press, pp. 33–50.
- Rayner K (1978) Eye movements in reading and information processing. *Psychological Bulletin*, 85: 618–60.
- Rayner K (1998) Eye movements in reading and information processing: 20 Years of Research. *Psychological Bulletin* 124: 372–442.
- Rayner K, Chace K, Slattery T, and Ashby J (2006) Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading* 10: 241–55.
- Rayner K and Pollatsek A (1989) *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.
- Rayner K, Slattery T, Drieghe D, and Liversedge S (2011) Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance* 37: 514–28.

- Rayner K, Warren T, Juhasz B, and Livesedge S (2004) The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30: 1290–1301.
- Roberts L, Gullberg M, and Indefrey P (2008) Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition* 30: 333–57.
- Sagarra N and Ellis N (2013) From seeing adverbs to seeing verbal morphology: Language Experience and adult acquisition of L2 tense. *Studies in Second Language Acquisition* 35: 261–90.
- Siyanova-Chanturia A, Conklin K, and Schmitt N (2011) Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research* 27: 251–72.
- Siyanova-Chanturia A, Conklin K, and van Heuven W (2011) Seeing a phrase ‘time and again’ matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition* 37: 776–84.
- Smith B (2012) Eye tracking as a measure of noticing: A study of explicit recasts in SCMC. *Language Learning and Technology* 16: 53–81.
- Spivey M and Marian V (1999) Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science* 10: 281–84.
- Staub A (2007) The parser doesn’t ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory and Cognition* 33: 550–69.
- Staub A and Rayner K (2007) Eye movements and online comprehension processes. In: Gaskell MG (ed.) *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press, pp. 327–42.
- Suvorov R (2014) The use of eye tracking in research on video-based second language (L2) listening assessment: A comparison of context videos and content videos. *Language Testing* 32: 463–83.
- Trueswell J, Tanenhaus M, and Garnsey S (1994) Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33: 285–318.
- Williams R and Morris R (2004) Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology* 16: 312–39.
- Winke P (2013) The effects of input enhancement on grammar learning and comprehension. *Studies in Second Language Acquisition* 35: 323–52.
- Winke P, Gass S, and Sydorenko T (2013) Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal* 97: 254–75.

