The Routledge Handbook of Research Methods in Applied Linguistics

Jim McKinley, Heath Rose

Eye tracking as a data collection method

Publication details
Ana Pellicer-Sánchez, Kathy Conklin
Published online on: 02 Jan 2020

How to cite: Ana Pellicer-Sánchez, Kathy Conklin. 02 Jan 2020, Eye tracking as a data collection method from: The Routledge Handbook of Research Methods in Applied Linguistics Routledge
Accessed on: 26 Oct 2023

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: https://www.routledgehandbooks.com/legal-notices/terms

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Introduction

Understanding the cognitive processes that underpin the acquisition, representation, production and comprehension of language is a central concern in applied linguistics. Because cognitive processes are not available to direct observation, research has traditionally relied on participants’ verbal accounts, using techniques such as stimulated recall (see Sanchez & Grimshaw, this volume) and think aloud protocols (see Zhang & Zhang, this volume). Other research has observed and/or measured participants’ behaviours and inferred something about cognition based on their performance (Ellis, 2015). Eye tracking is just such a technique; it provides a rich moment-to-moment measure of eye fixations which can be linked to ‘real-time’ processing. Since its inception, eye tracking has been said to allow us to “look into the mind of the subject” (Rayner, 1978, p. 618). In recent years, due to its many advantages, there has been an unprecedented increase in the number of studies in applied linguistics using eye-tracking technology.

This chapter provides a brief introduction to eye tracking and the methodological considerations to keep in mind when using it. The discussion touches on some of the areas and topics that have been explored using the technology in applied linguistics research and we consider how it can be used to investigate different types of verbal and nonverbal stimuli.

Overview of eye tracking

When we are presented with visual information – anything from written words to paintings or quickly changing images in videos – our eyes stop to process information at one location and then move to another where other information is available. Eye-tracking technology tells us about the eyes’ movements (saccades) and where people’s eyes land (fixate) when looking at visual stimuli. We can use this information to calculate how many times the eyes land in a position or region (fixation count), how long each fixation lasts (fixation duration) and determine whether the eyes go back to a previous point (regressions), as well as measuring saccade duration and length. Fixations, saccades and regressions generally occur ‘automatically’, without our conscious awareness (Rayner et al., 2012). Thus, tracking eye movements provides a window into a largely unconscious behaviour.
Eye tracking as a data collection method

The technology has a number of important benefits for researchers (Conklin, Pellicer-Sánchez, & Carrol, 2018; Scherr, Agauas, & Ashby, 2016). The temporal precision of eye tracking provides a record of behaviour from the first moment a visual stimulus (text, image, scene, video, etc.) is perceived until the stimulus is removed or a participant stops looking at it. For example, we can see when the eyes land on a word, as well as when and if they go back to re-read it at some point. It provides a ‘direct’ measure of processing effort during a task, rather than at the output of a decision, recall or production task, that are often subject to strategic effects. Although it is generally used in a laboratory setting, eye tracking allows readers and viewers to engage with visual stimuli much as they normally would (when they are presented on a computer screen). This means that participants can read and re-read at their own pace and look where they want at scenes, images and videos without the need to impose an additional task.

Eye tracking is premised on the assumption that there is a tight relationship between eye movements and cognitive processing. This is sometimes referred to as the eye-mind assumption (Just & Carpenter, 1980), which relies on two underlying beliefs (Pickering, Frisson, McElree, & Traxler, 2004; for a longer discussion and some caveats see Conklin et al., 2018). First is the supposition that what is being fixated on is what the processing system is working to decode and understand. Second, the amount of time spent fixating an item or region reflects the cognitive effort required to process it. Thus, by tracking people’s eye movements, we can make inferences about what people are attending to and how much cognitive effort they expend in doing so. The difficulty and complexity of what the eyes are looking at is believed to influence fixations and saccades, such that when the input is more difficult, fixation durations and regressions increase and saccade size decreases (Castelhano & Rayner, 2008).

This means that in reading, more difficult texts elicit more and longer fixations and regressions, while saccades get shorter. When looking at scenes or images that are more crowded, cluttered or dense, fixations also get longer and saccades get shorter. Fixations and saccades also vary based on the task. For example, in reading, fixations tend to be longer and saccades shorter in oral than silent reading, likely so that the eyes do not get too far ahead of what we are saying. Fixations and saccades in image/scene perception tend to be longer than in either type of reading because the eyes take in information from a wider area. Further, reading for comprehension or looking at a scene for memory induces more and longer fixations relative to reading for gist or a visual search task. Because of these differences, it makes it challenging to directly compare eye-movement patterns across different tasks and sources of visual input. For the remainder of this section, we will focus on the general properties of eye movements in reading and image/scene perception, which will be informative when designing tasks that make use of both sources of visual input.

In reading, during fixations readers need to ‘recognize’ words and ‘integrate’ them into a larger sentence or context (Clifton, Staub, & Rayner, 2007). The distinction between word recognition and integration roughly maps onto the classification of ‘early’ (recognition) and ‘late’ (integration) eye-tracking measures. More specifically, early measures are characterized as reflecting highly automatic word recognition and lexical access processes, while later measures tend to reflect more conscious, controlled, strategic processes (Staub & Rayner, 2007). Some common eye-tracking measures are demonstrated and described in Figure 31.1 and classified as ‘early’ or ‘late’. In broad terms, this distinction maps onto ‘first pass’ measures (location and duration of fixations during the first reading of a piece of text) and ‘total’ measures (location and duration of all fixations, including any re-reading that is required). Notably, not every element of reading can be classified neatly as recognition or integration, nor are they entirely independent of each other. For example, a predictable word is easier to recognize and identify, but the predictability of a word relies on comprehending a context greater than this single word.
The researcher was studying effects of enhancement on learning.

(a)

Figure 31.1 Illustrative eye-movement pattern (a) and related eye-tracking measures (b)

(b)

<table>
<thead>
<tr>
<th>Processing Stage and Measures</th>
<th>Definition and Depiction on Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Measures</td>
<td>Number of first fixation durations of $0 \div$ total number of trials. There is a first fixation [3], which contributes to the total number of trials, the denominator.</td>
</tr>
<tr>
<td>Likelihood of Skipping</td>
<td>The duration of the first, and only the first, fixation on a return on investment. [3]</td>
</tr>
<tr>
<td>First Fixation Duration</td>
<td>The sum of all fixations on a ROI before exiting to the right or left. [3 + 4]</td>
</tr>
<tr>
<td>First Pass Reading Time</td>
<td>Sum of fixations on a ROI and any regressions to earlier parts of the sentence before moving past the right boundary of the ROI. [3 + 4 + 5 + 6]</td>
</tr>
<tr>
<td>(or gaze duration if the ROI is a single word)</td>
<td>Number of trials with a regression $\div$ total number of trials. This example has a regression [5], which contributes to both the numerator and denominator.</td>
</tr>
<tr>
<td>Early/Late Measures</td>
<td>Regression path duration for a ROI minus first pass reading time. [3 + 4 + 5 + 6] $-$ [3 + 4] = [5 + 6]</td>
</tr>
<tr>
<td>Regression Path Duration</td>
<td>Sum of fixations on a ROI after it has been exited for the first time. [6 + 9]</td>
</tr>
<tr>
<td>(or go-past time)</td>
<td>Total duration of all fixations on a ROI. [3 + 4 + 6 + 9]</td>
</tr>
<tr>
<td>Regression Rate</td>
<td>Total number of fixations on ROI [3, 4, 6, 9] which makes four fixations</td>
</tr>
<tr>
<td>(or regressions-out)</td>
<td></td>
</tr>
</tbody>
</table>

*Note:* (a) An example sentence with a region of interest shaded in grey. 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. The processing stage is classified as ‘early’ or ‘late’ and 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).
As Figure 31.1 demonstrates, eye tracking allows us to collect and analyze data for a wide range of different measures that can be related to language processing more generally. Other measures are also available, and depending on the research question some of these might be relevant (for a description see Conklin & Pellicer-Sánchez, 2016; Pickering et al., 2004). While we do not need to decide on the appropriate measures when designing our study, it is important to be aware of the measures and consider which ones will allow us to best address our research questions. Something else to consider when designing a study involving reading are the various variables that can impact saccades and fixations (Clifton et al., 2007; Conklin et al., 2018): word class, length, frequency, familiarity, age of acquisition, number of meanings/polysemy, plausibility, contextual constraint, sentence length and syntactic complexity. Many of these may need to be controlled or accounted for in an analysis.

In addition to presenting written text, eye-tracking technology can be used to present static images and visual scenes (often referred to as the visual-world paradigm), as well as visual media like films and television programs. Eye tracking in this context is often used to explore auditory processing (listening), as well as in some cases speaking. In such tasks, various types of visual stimuli can be presented and they generally have a relationship to the linguistic input. For example, researchers have explored eye-movement patterns in various situations: giving spoken directions based on a visual display (e.g. Michel, Gilabert, & Révézès, 2013), audio storybooks with written text and pictures (e.g. Pellicer-Sánchez et al., 2018), and movies or TV shows with subtitles (e.g. Bisson, van Heuven, Conklin, & Tunney, 2014).

In listening studies, eye movements and fixations to visual input are generally time-locked to a particular linguistic variable (a word, pronoun, syntactic structure, etc.) that is presented auditorily. A key assumption in such tasks is that listeners tend to look at what they think the linguistic input refers to. Thus in Figure 31.2, when listeners hear ‘the exhausted maid’ they look at the maid depicted on the screen and when they hear ‘the broom’ they look at the broom. The different areas/images on the screen are defined as ‘regions of interest’ – which in Figure 31.2 would be the maid, broom and dustpan – and we would be interested in eye-movement patterns to these. Unlike in reading, where there is widespread agreement about what eye-tracking measures to look at and where effects are likely to show up, the situation for eye-movement patterns to visual data is much more complex (Conklin et al., 2018; see Altmann, 2011, for an overview of various ways of examining eye movements to visual input). A fairly standard practice is to look at the percentage of fixations to the regions of interest. In Figure 31.2 this would mean the percentage of fixations to the broom upon hearing ‘broom’, relative to the maid and the dustpan. However, some researchers measure the onset of a saccade towards a region of interest, or they calculate the percentage of trials in which a saccade to the regions of interest was launched within a specified time interval. As with reading, the eye-mind assumption holds. Thus, what is what is being looked at is what is being processed and number of fixations or saccades and fixation time reflect processing effort.

In relation to the example in Figure 31.2, we would consider the onset of the word ‘broom’ to be time point 0, and then at successive time points or intervals we can calculate the percentage of fixations to the various regions of interest. Thus, we could calculate the percentage of fixations to the broom at 50 ms intervals after the start of the word. We would expect to see an effect emerge at around 200 ms, because it is thought that it takes roughly 200 ms to plan and launch an eye movement. This means that there should begin to be more fixations to the broom than anything else on the screen about 200 ms after hearing the word ‘broom’. Importantly, for each item we need to establish the precise onset of this critical word in the audio file.

Notably, in such research, it is not just the eye-tracking measures that are debated. There are also questions about how precise the spatial regions of analysis should be. In Figure 31.2,
should the regions of interest stop at the outer boundary of the maid, broom and dustpan? Alternatively, is it acceptable to ‘draw’ a rectangle around the object to demarcate the region of interest, which includes areas of the screen that do not actually contain the image, simply because this is easier to do using the software available with eye-tracking systems? Defining a region of interest that extends beyond the exact boundary of the object will change the absolute number of fixations that ‘count’ as landing on the object; however, the overall pattern across conditions will essentially be unchanged (Altmann, 2011). Thus, it seems acceptable to define slightly larger areas as regions of interest that extend beyond the boundaries of the objects and entities of interest.

When doing research using images, scenes or video material it is important to keep in mind a number of factors that influence eye movements so that materials are well matched (for a fuller discussion of these factors see Conklin et al., 2018). Viewers primarily fixate on informative areas of a scene, as well as on salient aspects, which are generally defined in terms of features like contrast, colour, intensity, brightness, spatial frequency, and so on. Viewing is influenced by the real-world knowledge. For example, in a scene containing the sky and a road, when participants hear the word ‘car’, fixations are largely constrained to the road, which is where a car is likely to be found. Finally, viewers tend to fixate near the centre of an object, and when looking at scenes they look at people (or characters) and concentrate their fixations on the face.

In sum, eye tracking has many benefits. Maybe at the forefront of these for applied linguists is the possibility to use tasks that allow for natural reading and viewing. That being said, we cannot
simply present authentic materials and draw conclusions directly from eye-movement data. As the previous discussion highlights, many factors influence eye movements (e.g. word frequency, visual salience of images). These either need to be controlled or accounted for in our analysis.

Methodological considerations

As discussed in the previous section, eye tracking provides information about what parts of the input and stimuli participants are processing and the cognitive effort involved. Crucially, the capacity of eye tracking to tell us anything about cognitive processes depends on having a good experimental design. This section provides an overview of some of the main methodological issues that need to be considered in the design of high-quality eye tracking studies.

An initial consideration is ensuring that our research design conforms to the technical features and specifications of the eye-tracking system we are using. For example, if we want to make claims about the processing of smaller areas of the stimuli (e.g. specific words in a text), we need to have an eye-tracking system that records eye-movement data of high-enough quality (i.e. high precision, accuracy, sampling rate). If we have an eye-tracker that records data of lower quality, the interpretations that we can draw from eye-movement data to smaller regions of interest will be seriously compromised. The negative effect that certain technical specifications have on data quality can sometimes be compensated with more data (Andersson, Nyström & Holmqvist, 2010). Further, the magnitude of this negative effect might be different for saccades and fixations (Raney, Campbell & Bovee, 2014). Thus, before deciding on specific aims and an experimental design, the technical specifications of the system we are using and their effect on data quality should be considered (see Conklin et al., 2018, for a detailed discussion of the features of different systems and how they might impact research questions and designs).

Eye tracking allows us to examine the cognitive effort involved in processing different types of stimuli, being that written/oral verbal stimuli, or non-verbal stimuli. For example, we might be interested in finding out whether the processing of words that have specific characteristics requires more/less/equal cognitive effort. Since processing effort is always relative, we will need to have a matched comparison set. Only through a proper comparison would we be able to interpret eye movements and attribute eye-movement patterns to greater or less processing effort. Designing experiments that can make this comparison valid is one of the most important methodological considerations in eye tracking research. In what follows we discuss some of the key considerations when creating experiments using the same type of stimuli (e.g. text only containing written words) and across stimulus types (e.g. audio book with pictures, written words and audio recording). These considerations should be taken into account when designing our own experimental stimuli or when modifying previously designed materials.

Examining eye movements within the same input type

Written verbal input

Investigating eye movements during reading allows us to explore the processing of texts of different lengths, i.e. a word, a phrase or longer extracts. Most eye tracking research on reading has focused on examining processing patterns at the word level. For example, as applied linguists we might be interested in the processing of legal documents and we might hypothesize
that more difficult words such as ‘covenant’ in Example 1 (next) require more cognitive effort and make the reading more arduous.

**Example 1.** Experimental sentences used to compare the reading of sentences with lower frequency words (a) and the reading of matched sentences with higher frequency words (b). The underlining appears for illustration purposes and it would not be visible in the experimental stimuli. The duration of all fixations (total fixation duration) made within the target items appears below the target words in brackets.

(a) This Agreement constitutes the entire agreement between the parties. None of the terms and covenants in the agreement can be further modified. 
(540 ms)
(b) This Agreement constitutes the entire agreement between the parties. None of the terms and contracts in the agreement can be further modified. 
(235 ms)

Any claims about the cognitive effort that a word such as ‘covenant’ requires can be made only through a proper comparison. Finding that the word ‘covenant’ in Example 1 (a) is read at 540 ms does not mean anything by itself. We would need to compare the processing of ‘covenant’ to the processing of a matched control. A viable comparison in this context would control for as many factors as possible that are known to affect the processing of text, including word length, frequency, age of acquisition, predictability, number of meanings, contextual richness and position in the text. The processing of the word ‘covenant’ in Example 1 (a) could be compared to the reading of the word ‘contract’ in Example 1 (b). In this example both target items are matched for lexical features (i.e. length, word class) and are inserted in the same sentence contexts and in the same position within the sentence, which means we can rule out the possibility of those factors exerting an influence on reading times. Thus, we will be able to claim that ‘covenant’ entails more processing effort than its matched control (see Clifton et al., 2007, for a discussion of factors affecting eye movements in reading). The position of critical stimuli in a passage should also be controlled for. As a general rule, we tend to avoid placing critical words at the beginning and end of a sentence, as research has shown that these positions might attract longer and more fixations (see Conklin et al., 2018, for a detailed discussion of desired properties of textual stimuli). When target words are embedded in longer texts, the font size should be big enough and interline spacing should at least be double to ensure that fixations can indeed be attributed to specific words. A larger font size and increased spacing make it easier to attribute a fixation to a particular word and therefore attribute processing effort to it.

Eye tracking can also be used to examine the processing of items beyond the single word (e.g. idioms such as ‘spill the beans’ and collocations such as ‘heavy rain’). In this case, it is important to control not only for the features of the individual constituent words discussed earlier, but also for the features of the whole multiword combination, including characteristics such as the frequency of the multiword unit, association strength and mutual information (MI) score (see Carrol & Conklin, 2014, for a detailed discussion on methodological considerations when examining eye movements to multiword units).

When examining the processing of written stimuli, we may be interested in eye-movement patterns to the whole text, as opposed to specific words or phrases. For instance, we might be interested in examining the effect of previous background knowledge on the processing and comprehension of texts. We want to ask participants to read two texts: one on a topic that is
familiar and one that is not. For this comparison to be appropriate, the two experimental texts need to be matched for factors such as overall length, syntactical and lexical complexity. Otherwise, it will not be possible to attribute eye-movement patterns to differences in familiarity. In addition, we will need to ensure that the text layout (i.e. font style, font size and interline spacing) is comparable across the texts. Equally, this is critical if we would like to compare different sources of written input. For example, in the context of translation we might be interested in how translators process a source text and an automatically generated translation that are presented concurrently on the computer screen (see Figure 31.3). If one of the textual areas has a smaller font size, as in Figure 31.3, a fixation will capture different amounts of information. If we find that the number and duration of fixations is the same in the source and target text, we cannot be sure whether this means that the two require equivalent levels of processing. Alternatively, we could conclude that the target text requires more processing effort: one fixation allows the readers to see more, but they still need as many fixations as when they see less. Thus, similar eye-movement patterns might be hiding important processing differences.

Figure 31.3 shows hypothetical stimuli representing two textual interest areas that contain different amounts of information.

![Figure 31.3](image-url)
We tend to think of text areas as something static. However, verbal stimuli can also be dynamic. Recent applications of eye tracking to the study of writing have examined how writers process the text they are producing at different time points. This involves examining eye movements to a text that is continuously changing. In this case, the main methodological challenge is to match eye movements to the specific parts of the produced text as it develops. Integrating eye movements to the developing text can be done by specific software that synchronizes eye-movement and keystroke logging data (e.g. Wengelin et al., 2009).

Non-verbal input

Eye movements to non-verbal stimuli (which are often presented in conjunction with auditory stimuli) can also be explored to make inferences about the cognitive effort they entail. A long tradition of research in psychology and psycholinguistics has shown the connection between the linguistic stimuli we hear and eye movements to referents in a visual display (e.g. Altmann, 2011; Cooper, 1974). In this context fixations to visual stimuli are an indication of what we think the linguistic stimuli are referring to. For example, if participants see a screen with a picture of a mouse and some cheese, when they hear ‘the cheese’ they will look at the cheese depicted on the screen.

Fixations to images and videos are affected by the perceptual features of the stimulus. An obvious consideration is the size of the stimuli. Images of different sizes attract different number of fixations and fixations of different durations. More salient visual input also attracts more attention. Properties such as contrast, colour and brightness determine how salient visual input is. Images or scenes that are denser also attract longer fixations. Research has shown that particular sections of the non-verbal display receive different amounts of attention, with most fixations being close to objects and entities (Henderson, 2003). A well-controlled design involving non-verbal stimuli will therefore account for as many of these features as possible. For example, we might be interested in the use of pictures as modes of meaning presentation in the learning of new words of different levels of difficulty. We would need a set of pictures that illustrate the meanings of our target words. Using pictures like the ones depicted in Example A in Figure 31.4 would allow us to make claims about any differences in eye movements being due to the difficulty of the words. On the contrary, processing differences observed between the two pictures in Example B will clearly be affected by their distinctive perceptual features.

Eye tracking research on the processing of non-verbal stimuli has also shown that speakers of languages that are read from left to right tend to scan a screen with images from left to right. Therefore, the position of the target images needs to be counterbalanced, ensuring that target images appear in different locations on the screen across experimental lists.

Controlling for the features that affect the processing of dynamic, non-verbal stimuli is more challenging, as most of the time we use authentic videos that cannot be modified. In this case an important consideration will be to ensure that the regions of interest that we select within those videos have comparable characteristics. For example, when comparing the processing of different video frames, it is important to ensure that they are presented for the same length of time. Otherwise, a part of the visual stimuli that stays on the screen for longer will attract more fixations. Often we are interested in the processing of video frames when specific words appear in the auditory stimulus. In this case we should try to select target items that have similar characteristics and that appear in similar contexts.
Eye tracking as a data collection method

Examining eye movements across input modes

The previous section discussed some of the main methodological considerations when looking at eye movements to text or to images/videos. Thus, experimental comparisons were being made to the same type of stimulus (written word to written word; image to image). However, many of the questions that applied linguists seek to answer involve the comparison of eye movements to different types of stimuli. For example, we might be interested in the processing of text and pictures in a storybook (see the example in Figure 31.5), or in how people process the subtitle and the image areas in movies. This section focuses on the main methodological issues that arise from these types of comparisons.

As we have seen, the processing of verbal and non-verbal stimuli is affected by different factors. Patterns of eye movements to text and image/video areas will be by nature different. The duration of fixations and length of saccades are shorter when processing text than when

---

**Example A:**

![Example A](image)

**Example B:**

![Example B](image)

*Figure 31.4  Non-verbal stimuli with standardized (Example A) vs. non-standardized (Example B) perceptual features*

*Note: Example A shows two images for the words ‘ant’ and ‘bed’ from MultiPic (Duñabeitia, et al., 2018) and normed for name agreement and visual complexity. In Example B the normed image of the ‘ant’ is compared to a non-standardized image of ‘bed’.*
looking at images and scenes (Rayner, 2009). Thus, because total fixation duration on the image area in Figure 31.5 is 2,450 ms and in the text area is 3,467 ms, we cannot conclude that the processing of the text requires more cognitive effort, because we are simply comparing things that are not alike. Even when differences in the size of the text and image areas are accounted for (normalising the fixation durations by size of the areas), the direct comparison of how a text and an image areas are processed would still be irrelevant.

In these cases, the interest is not in the examination of number of fixations and fixation durations but on the proportion of overall time that participants spend fixating each of the areas of interest and the saccadic movements between them. This gives us an indication of how participants allocate their attention to the different sources of input. Even when looking at proportion of time spent processing the different areas, it is important to note that eye-movement patterns will still be affected by the distinctive nature of the stimuli we are examining. The comparison of fixations to a text and to an image by itself do not tell us much, unless we have another experimental manipulation. For example, the study by Pellicer-Sánchez et al. (2018) examined fixations to image areas relative to fixations to the written text when auditory input was present or absent (i.e. reading vs. reading-while-listening).

When examining the processing of verbal and non-verbal static stimuli, the relative position of the stimuli needs to be controlled for to make sure that none of the areas are particularly salient (Conklin et al., 2018). Previous research has shown that spatial contiguity in multimedia learning leads to a better integration of words and pictures (e.g. Johnson & Mayer, 2012). Thus, as in the study by Pellicer-Sánchez et al. (2018), the relative position of text and pictures should be counterbalanced so that all text and image areas appear to the right and left of the screen across experimental lists.

Most of the methodological considerations explored in this chapter assume that we will have the opportunity to create our own stimuli or to manipulate the experimental stimuli. However, very often our designs will involve the use of authentic materials and controlling for all the potentially confounding factors might not be a realistic endeavour. We could instead consider those factors in the selection of regions of interest or collect ‘norming’ data to allow us to account for the differences. For example, when looking at the processing of pictures and
text in an authentic book, we could norm the pictures for saliency and complexity and select only those that receive a similar rating. In addition, we can exclude from the analysis any pages in the book that have a significantly different feature or layout. It is important to note the materials we use should be sufficient so that even after the appropriate exclusions are made, we end up with an appropriate amount of data. Crucially, statistical analysis (i.e. mixed-effects modelling) that allow us to account for a variety of factors (e.g. contrast, colour, brightness) could also be used to allow us to examine the effect of those factors that could not be controlled for in the experimental design (see Conklin et al., 2018, for a detailed discussion of the analysis of eye-movement data).

Conclusion

In recent years eye tracking has become a valuable tool and an important source of data in applied linguistics research. This chapter has illustrated some of the main uses that the technology has been put to, as well as the main methodological concerns that are important to consider in order to obtain good-quality eye-movement data. If we do not consider the factors affecting eye movements, either in the design or analysis of data, this will restrict the interpretations we can draw from eye-movement data and the claims we can make about the cognitive effort involved in the processing of different types of stimuli. There is no doubt that the uses of eye tracking and its application within our field will continue to flourish in innovative, and sometimes unpredictable, ways, further contributing to the advancement of eye-tracking-based applied linguistics research. Regardless of the specificities of the topics and questions addressed, we will always be dealing with the processing of either verbal or non-verbal (static or dynamic) stimuli, albeit combined in different ways and presented in different contexts. The same methodological considerations that have guided eye tracking research in other fields and that have been discussed in this chapter will therefore still apply. The present chapter should serve as a good foundation not only for the interpretation of current research findings but for further developing research applications in the field.

References


