

Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading

Uschi Cop¹ · Nicolas Dirix¹ · Denis Drieghe² · Wouter Duyck¹

© Psychonomic Society, Inc. 2016

Abstract This article introduces GECO, the Ghent Eye-Tracking Corpus, a monolingual and bilingual corpus of the eyetracking data of participants reading a complete novel. English monolinguals and Dutch–English bilinguals read an entire novel, which was presented in paragraphs on the screen. The bilinguals read half of the novel in their first language, and the other half in their second language. In this article, we describe the distributions and descriptive statistics of the most important reading time measures for the two groups of participants. This large eyetracking corpus is perfectly suited for both exploratory purposes and more directed hypothesis testing, and it can guide the formulation of ideas and theories about naturalistic reading processes in a meaningful context. Most importantly, this corpus has the potential to evaluate the generalizability of monolingual and bilingual language theories and models to the reading of long texts and narratives. The corpus is freely available at <http://expsy.ugent.be/downloads/geco>.

Keywords Bilingualism · Reading · Eyetracking · Corpus study

Electronic supplementary material The online version of this article (doi:10.3758/s13428-016-0734-0) contains supplementary material, which is available to authorized users.

✉ Nicolas Dirix
nicolas.dirix@ugent.be

¹ Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium

² School of Psychology, University of Southampton, Southampton, UK

Over the years, linguistic data gathered in experimental settings have driven the development of ideas and theories about the cognitive processes involved in language performance. Usually, these experiments are designed to test one or more specific hypotheses and use a meticulously selected and restricted stimulus set, containing one or more, often orthogonal, experimental manipulations. More recently, with the development of larger, and more complex, computational-reading models that operate on multiple processing levels and/or cover a wide range of phenomena (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Demberg & Keller, 2008; Dilkina, McClelland, & Plaut, 2010; Friederici, 1995; Grainger & Jacobs, 1996; Harm & Seidenberg, 2004), the need for data from a larger and more naturalistic range of stimuli has become more pressing. This kind of data is necessary for evaluating the generalizability and external validity of these language models for the reading of longer texts or narratives.

The collection of large amounts of language behavior data can have an important role in the development, simulations, or confirmation of ideas and theories. The studies based on collecting these large databases are often referred to as *corpus studies* or *megastudies* (e.g., Balota et al., 2007; Seidenberg & Waters, 1989). Because corpus studies are based on a large number of observations from a limited number of participants, or vice versa, or on large numbers of both observations and participants, they usually have considerable statistical power and can detect relatively small effects. These studies are often characterized by the presentation of a large sample of a wide range of unselected stimuli, in contrast to the factorial designs used in traditional experimental settings, in which a limited set of stimuli are selected on the basis of specific characteristics. This typically constricted range usually includes very high and/or very low values and limits the stimulus set to stimuli that are rather extreme in the critical dimension, which may impede the representativeness of their processing

characteristics and show only a part of possible language behavior. An advantage of the corpus approach is that the effects of continuous lexical variables, such as word frequency, can be assessed over their full possible range, instead of over a constricted one. Another advantage of large corpora of linguistic data is that they enable researchers to answer multiple hypotheses without the need to design a new experiment and gather new data, which is considerably time-consuming and may require expensive equipment (e.g., an eyetracker).

A good example of an influential psycholinguistic corpus study in the field of visual word recognition is the English Lexicon project (ELP; Balota et al., 2007). Balota et al. gathered lexical decision latencies from 816 participants for 40,481 different American English words (3,400 responses, on average, per participant). Subsequently, this project sparked the development of similar databases for French (FLP; Ferrand et al., 2010), Dutch (DLP; Keuleers, Diependaele, & Brysbaert, 2010), and British English (BLP; Keuleers, Lacey, Rastle, & Brysbaert, 2012). These databases have been used to evaluate psycholinguistic ideas about frequency effects (e.g., Kuperman & Van Dyke, 2013), word length effects (e.g., Yap & Balota, 2009), neighborhood effects (e.g., Whitney, 2011; Yap & Balota, 2009), and the lexical decision task itself (Diependaele, Brysbaert, & Neri, 2012; Kuperman, Drieghe, Keuleers, & Brysbaert, 2012), but they have also been used to evaluate complex computational models of word recognition (e.g., Norris & Kinoshita, 2012; Whitney, 2011), illustrating the relevance and broad applicability of such big datasets.

Eyetracking corpora

Large databases of responses related to the processing of isolated word stimuli are very useful in evaluating specific hypotheses about word recognition and in simulations of models, which are mainly concerned with the process of lexical access to an isolated target word. However, when the goal is to explain how reading occurs in natural contexts, the ambition of reading models should also be to expand their generalizability beyond word-level processes, in order to cover a larger scope of potential interacting language processes. This means that they should consider how word-level processes may alter or interact with semantic or syntactic processes, for instance, when readers are processing longer text fragments. Clearly, to evaluate generalizability and the complex interactions between different representation levels, more complex datasets of natural text reading are necessary.

The technique of eyetracking enables researchers to record the eye movements of participants during silent reading, with minimal instruction or interference on behalf of the researcher. Also, eyetracking—in contrast to, for example, lexical decision tasks—captures language performance as it occurs in daily life, without interference from the additional decision

components or response mechanisms that are inherent to lexical decision, for instance. With modern-day eyetracking equipment, the position of the eye can be determined every millisecond with very high spatial accuracy, resulting in a very rich and detailed dataset. The recording of eye movements during reading has been used often to study visual word recognition in context (see Rayner, 1998, for an introduction and review of early work, and Rayner, 2009, for a more recent review). Some models of reading have focused on the influence of the characteristics of the surrounding words or sentences on reading target words (Engbert, Nuthmann, Richter, & Kliegl, 2005; Pynte & Kennedy, 2006; Reichle, Pollatsek, Fisher, & Rayner, 1998), and these models have relied heavily on experimental findings in eye movement research as a way to understand the cognitive processes of reading. One of these models, the E-Z reader model by Reichle et al. (1998), has put the modeling of eye movements central in their theorizing. Lexical access also plays an essential role in this model, based on the fact that lexical characteristics such as word frequency and word length reliably influence (the duration of) eye movements (Inhoff & Rayner, 1986; Rayner & Fischer, 1996).

Here, we propose that an eyetracking dataset including a large sample of stimuli considerably increases the richness of the available eye movement datasets. Corpora of eye movements during naturalistic, contextualized reading of text will be invaluable for informing and evaluating language models that go beyond the word level, such as the E-Z reader model. These corpora can be used to examine a large number of variables at different processing levels (e.g., at both the word and sentence levels) and the interactions among them simultaneously, as well as the specific time courses of these effects. Moreover, testing the predictions of language models in an eyetracking corpus of natural reading could provide a test of the generalizability of parts or the whole of a specific model, especially with regard to parts of the model that were inspired by findings obtained in less natural tasks.

Additionally, as we have already discussed for corpora of isolated word recognition, these eyetracking databases (a) are perfectly suited to investigate a very broad scale of phenomena—as long as certain syntactic constructions or words with certain lexical traits occur frequently enough in the corpus, they can be studied; (b) have a representative unrestricted set of stimuli, which supports generalizability; and (c) provide researchers with data, so that there is no need to continuously design new experiments or collect new data, which often requires specific, expensive equipment and is a time-intensive process, especially for sentence reading.

A first example of an existing eyetracking corpus of natural reading is the Dundee corpus (Kennedy & Pynte, 2005). Ten native French and ten native English subjects read newspaper articles (50,000 words) that were presented in paragraphs on the screen. Eye movements were recorded with a sampling rate of 1 ms and spatial accuracy of 0.25 characters. Initially,

the authors used this corpus to investigate the effect of parafoveal processing on foveal word inspection times (Kennedy & Pynte, 2005; Pynte & Kennedy, 2006; but see Reichle & Drieghe, 2015, for a criticism). Later, the same authors investigated the effect of punctuation (Pynte & Kennedy, 2007), the effects of syntactic and semantic constraints on fixation times (Pynte, New, & Kennedy, 2008, 2009a, b), the effect of violations in reading order (Kennedy & Pynte, 2008), and the interaction between frequency and predictability (Kennedy, Pynte, Murray, & Paul, 2013) using eye movement data from the Dundee corpus.

Other authors have also used this corpus to investigate specific hypotheses. Demberg and Keller (2008), for example, investigated subject/object clause asymmetry with the Dundee corpus data and were inspired by these results to build a model of syntactic processing (Demberg & Keller, 2008). The Dundee data were used to evaluate this model. Mitchell, Lapata, Demberg, and Keller (2010) used the Dundee corpus to investigate prediction in sentence reading. A nice illustration of the power of these kinds of corpora is the fact that these authors only needed 10% of the data to test their hypothesis. Both Frank and Bod (2011) and Fossum and Levy (2012) used the Dundee corpus to evaluate their language models concerned with the role of hierarchical mechanisms in sentence processing. Kuperman et al., (2012) used both the megadata of the ELP (Balota et al., 2007) and the Dundee corpus (Kennedy & Pynte, 2005) to correlate lexical decision times with natural reading data. Their results showed very low correlations between these measures, implying that these commonly used methods measure, at least to some extent, different processes. This illustrates that evaluations of language models should also use natural reading data.

There are other interesting examples of databases of eye movements in text reading. For instance, Frank, Fernandez Monsalve, Thompson, and Vigliocco (2013) gathered eye movements from 43 English monolingual subjects reading 205 sentences. Instead of presenting the sentences in paragraphs, as the Dundee corpus does, Frank et al. selected sentences from natural narrative text and presented these sentences separately on the screen. Other examples are the German Potsdam corpus (Kliegl, Nuthmann, & Engbert, 2006) and the Dutch DEMONIC database (Kuperman, Dambacher, Nuthmann, & Kliegl, 2010). In the former, 222 subjects read 144 constructed German sentences, and in the latter, 55 subjects read 224 constructed Dutch sentences. These sentences were presented in isolation and did not form a coherent story in any way. The data of these corpora have been useful for model construction (Engbert et al., 2005), evaluation (see, e.g., Boston, Hale, Kliegl, Patil, & Vasishth, 2008), and/or hypothesis testing. Some of these corpora contained monolingual reading in different languages, supporting generalizability of their claims across languages. However, these existing datasets remain quite limited in the diversity of their words and sentences, and have many fewer

stimuli than, for instance, the large, isolated-word reading projects (e.g., the ELP).

In conclusion, it seems that corpora of eye movement data have been (and still are) valuable to the field of psycholinguistics. However, two domains within this approach are yet to be explored: reading an entire novel (implying a large amount of different word stimuli) and reading in a second language. We will address these issues and their importance in the presentation of a new eyetracking corpus.

Our corpus: GECCO

As the previous section showed, the building of eyetracking corpora of natural reading can be very fruitful for the development and evaluation of monolingual models of language processing. However, whereas the act of reading isolated sentences (Kuperman et al., 2010) or short newspaper articles (Kennedy & Pynte, 2005) has been studied in experimental settings, no one has ever systematically collected and analyzed the eye movements of participants reading an entire book (though see Radach, 1996, for a corpus of four participants reading a selection of chapters from *Gulliver's Travels* in German). This is quite surprising, since books have been read for hundreds of years in a multitude of contexts (e.g., work, study, or leisure). Our present approach allows for answering several important questions. First, it would be highly interesting to examine whether the findings of previous eyetracking research using a limited set of stimuli would be preserved when put to the test in a database that contained a very large and wide range of stimuli not appearing in specially constructed sentences. Second, the reading of long texts or narratives entails additional processes (e.g., sentence integration) that typically are not present in the process of reading isolated sentences (see, e.g., Calvo & Meseguer, 2002; Mielliet, Sparrow, & Sereno, 2007; Miller, Cohen, & Wingfield, 2006). Therefore, an eyetracking corpus of people reading a long narrative would allow us to test whether the influence on reading of some well-known factors is impacted when the full range of cognitive processes that are typically at play during the reading of a novel are active.

Next, until now no single, large eye movement database has focused on, or even specified, possible differences in language knowledge between participants. All eyetracking corpora (at least to our knowledge) have implicitly assumed that their participants have knowledge of only the language they are reading in. Bilingualism is most commonly defined as “the regular use of two (or more) languages” (Grosjean, 1992), and today, across most European countries, 54% of the people are bi- or multilinguals, due to migration and the fact that foreign languages are a compulsory part of formal education (European Union & European Commission for Education and Culture, 2012). Even in developing countries such as Cameroon, more than half of the population speaks three or

more languages (Bamgbose, 1994). In the United States of America, although foreign language courses are not compulsory, about 20% of the population has some knowledge of a nonnative language (Shin & Kominski, 2007).

This is important, because a plethora of evidence shows that bilingualism changes language processes and that bilinguals need to allocate resources in a different way than monolinguals do. A major finding, for instance, is that words of both languages are activated in parallel even in unilingual contexts (for a recent review of the evidence, see Kroll, Dussias, Bogulski, & Valdes Kroff, 2012).

So far, no megadata are available for participants reading in their first language who have a confirmed and assessed knowledge of another language, or for participants reading in a second language that they have acquired later in life. In short, no bilingual eyetracking corpus is available to researchers. In this article we present the GECO, the Ghent Eye-Tracking Corpus, whose goal it is to bridge this gap, serving both the bilingual and monolingual reading research domains. We gathered eye movement data from monolingual British English participants and Dutch–English bilinguals while they read an entire novel. The bilinguals read half of the novel in their L1 and the other half in their L2. All participants read a total of about 5,000 sentences, and a precise language history and proficiency score were gathered for each of the participants. This was the first bilingual corpus study and also the first large corpus of Dutch reading of natural text (i.e., not specifically constructed for an experiment). Information on the participants and the materials of the novel, as well as the eyetracking data, are available as online [supplementary materials](#). See Appendix A for a list of the available files and their exact contents.

Exploitation of the present corpus

Data from the GECO corpus have been used in two studies so far. By comparing the basic eye movement measures on the sentence level between L1 and L2 reading (Cop, Drieghe, & Duyck, 2015), we provided a database of the benchmark parameters of reading with attention, investigating the relation between language history and changes in eye movement behavior. Here, we showed that changes in eye movement patterns from L1 to L2 closely resemble the changes observed in reading patterns from child to adult reading (e.g., longer and more fixations over time, shorter saccades, and lower probability of skipping words). Furthermore, we observed that in L1 reading of continuous text, no differences were apparent between monolinguals and bilinguals, in contrast to the disadvantages found in L1 production for bilinguals (Gollan, Montoya, Cera, & Sandoval, 2008). This finding is important for theories of bilingualism that assume that effects of L2 learning on L1 use are caused by the distributed practice across languages (e.g., the weaker links theory; Gollan et al., 2008).

The GECO was also used for a systematic analysis of the most-investigated lexical variable, word frequency, in L1 versus L2 reading (Cop, Keuleers, Drieghe, & Duyck, 2015). We showed that frequency effects are larger in L2 than in L1, and also that higher L1 (but not L2) proficiency resulted in smaller frequency effects for both languages. These analyses also showed that qualitative differences between monolingual, L1, and L2 language processing do not necessarily account for the differences in frequency effects. Indeed, our results demonstrated that for both groups, the size of the frequency effect can be explained by the target language proficiency. Moreover, the relationship between the frequency effect and L1 proficiency was the same for both groups. These findings are very relevant for theoretical models of monolingual and bilingual reading, and are examples in themselves of the value of such data for investigating specific research questions without the need to collect new data.

Avenues for future research

These two applications are only indicative of the many possible applications of the database, and many others remain—for instance, for the field of bilingualism. A prominent model of bilingual word recognition is the *bilingual interactive activation plus* (BIA+) model (Dijkstra & van Heuven, 2002). The authors mentioned that this model concerns the visual word recognition system, which is part of a larger “language user” system that also includes sentence parsing and language production. The model assumes that the linguistic (sentence) context has a direct impact on the word recognition system (Dijkstra & van Heuven, 2002), but how exactly is not specified. Because of the contained nature of their model, they did not use eye movement data obtained from natural reading to inform the architecture or evaluate the system of word recognition they proposed. Instead, the model was adjusted from the BIA model (Dijkstra & van Heuven, 1998) on the basis of findings from a multitude of experimental studies using lexical decision, progressive demasking, and identification tasks (e.g., Bijeljac-Babic, Biardeau, & Grainger, 1997; Dijkstra, Timmermans, & Schriefers, 2000; van Heuven, Dijkstra, & Grainger, 1998) with words usually not embedded in a sentence context (but see Altarriba, Kroll, Sholl, & Rayner, 1996). We believe that the large corpus of eye movements we present here will not only allow us to evaluate the ecological validity of this word recognition model in the context of natural reading, but it also be especially helpful to specify the exact nature of the interactions between the sentence context and the word recognition system. In their article presenting the BIA+ model, Dijkstra and van Heuven (2002) said,

Future studies should focus on disentangling such effects of lexical form features and language membership in sentence processing experiments. They should

examine, for instance, to which extent the language itself of preceding words in the sentence can modulate the activation of target word candidates from a non-target language. (Dijkstra & van Heuven, 2002, p. 187)

Indeed, the GECO can be exploited for such purposes. As bilingual participants read text in a unilingual context, an influence of the activation of lexical candidates (e.g., orthographic neighbors) in the nontarget language could be a clear indication of a shared lexicon or nonselective access to the lexicon (van Heuven et al., 1998). The effect of interlingual homographs (Libben & Titone, 2009) or cognates (Van Assche, Drieghe, Duyck, Welvaert, & Hartsuiker, 2011) could also be put to the test under less constrained circumstances (i.e., without specially constructed sentences). Another advantage of the dataset is that the same materials are used for monolingual and bilingual reading. A cross-language comparison between L1 and L2 for bilinguals can be made, as well as a direct comparison between L1 reading for monolinguals and bilinguals. The latter comparison would be especially interesting so as to address, for example, the weaker-links hypothesis (Gollan et al., 2008), which states that becoming a bilingual has an influence on L1 reading. Furthermore, besides our study of the word frequency effect (Cop et al., 2015), other effects at the word level could be investigated and compared between these groups (e.g., using orthographic [cross-lingual] neighbors, age of acquisition, or homographs). Finally, beyond word-level studies, the nature of the corpus would also allow investigations of the sentence, semantic, or higher-order levels of reading, which are almost nonexistent for L2 reading.

We have already noted some of the differences between the present corpus approach and other methods of studying (bilingual) reading and word recognition in psycholinguistics. In an interesting study, Kuperman et al. (2012) found little shared variance between eye movement data from the Dundee corpus (Kennedy & Pynte, 2005) and reaction time data from the ELP (Balota et al., 2007). Our data could also be exploited by similar studies for comparing monolingual data from the corpus to, for instance, the BLP (Keuleers et al., 2012), the L1 bilingual data to the DLP (Keuleers, Diependaele, et al., 2010), and the L2 bilingual data to a potential future lexicon project in the L2 (which is nonexistent, to date).

Besides the possible theoretical and empirical contributions that may be derived from the GECO, this corpus can also support advancements in computational modeling. For instance, a broader use for these data might be the evaluation and adaptation of the E-Z reader model (Reichle et al., 1998), one of the most important models of eye movements, to bilingual reading. Because this model has proven to be successful in accommodating the eye movement patterns of older

(Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006) and younger (Reichle et al., 2013) readers as well as nonalphabetic languages (Rayner, Li, & Pollatsek, 2007), we have reason to believe that it will perform well as a framework for bilingual eye movement patterns. As we discussed earlier, using GECO we found that L2 reading resembles child-like reading (Cop et al., 2015), the latter of which has been successfully simulated in the E-Z reader model by only adjusting a single parameter (i.e., the rate of lexical processing; Reichle et al., 2013). The data of GECO therefore constitute a promising avenue to extend models like E-Z reader to bilingualism.

In conclusion, we present a corpus of eye movements of participants reading an entire book, a text format that is currently underexplored in eyetracking research. Our participant group consisted of both monolinguals and bilinguals, resulting in the first bilingual database of eye movements.

Method

A more concise version of this method is presented in Cop, Keuleers, et al. (2015), who described the method as part of an investigation into frequency effects.

Participants

Nineteen unbalanced Dutch (L1)–English (L2) bilingual Ghent University and 14 English monolingual undergraduates from the University of Southampton participated either for course credit or monetary compensation. Bilingual and monolingual participants were matched on age and education level. The average ages were 21.2 years for the bilinguals (range: 18–24; $SD = 2.2$) and 21.8 years for the monolinguals (range: 18–36, $SD = 5.6$). All of the participants were enrolled in a bachelor's or master's program of psychology. In the monolingual group, six males and seven females participated. In the bilingual group, two males and 17 females participated. The participants had normal or corrected-to-normal vision, and none reported having any language and/or reading impairments.

The bilinguals started learning their L2 relatively late: The mean age of acquisition was 11 years (range: 5–14, $SD = 2.46$). All participants completed a battery of language proficiency tests, including a vocabulary test, a spelling test, a lexical decision task, and a self-report language questionnaire (for the results, see Table 1). Vocabulary was tested with the LexTALE (Lexical Test for Advanced Learners of English; Lemhöfer & Broersma, 2012). This is an unspecced lexical decision

Table 1 Average percentage scores [and standard deviations] on LexTALE, the spelling test, the accuracy of the lexical decision task, and subjective exposure, as well as scores on the comprehensionquestions for the bilingual and monolingual group; results [with degrees of freedom] from *t* tests are presented in the last two columns

	Monolinguals	Bilinguals L1	Bilinguals L2	<i>t</i> Value L1–L2	<i>t</i> Value L1–Mono
LexTALE score (%)	91.07 [8.92]	92.43 [6.34]	75.63 [12.87]	7.59 [18] ^{***}	0.49 [18]
Spelling score (%)	80.78 [7.26]	83.16 [7.80]	69.92 [8.74]	8.15 [18] ^{***}	0.99 [18]
Lexical decision accuracy (%)	77.89 [12.01]	80.19 [5.41]	56.84 [11.12]	9.93 [17] ^{***}	0.67 [17]
Subjective exposure (%)	100.00 [0]	75.00 [15.24]	25.00 [15.24]	7.10 [18] ^{***}	7.10 [18] ^{***}
Comprehension score (%)	78.27 [9.46]	79.63 [10.96]	78.95 [12.54]	0.40 [18]	0.38 [30]

^{***} $p < .001$

task, which is an indicator of language proficiency for intermediate to highly proficient language users that has been validated for English, Dutch, and German. Due to the lack of a standardized cross-lingual spelling test, we tested the English spelling with the spelling list card of the WRAT 4 (Wilkinson & Robertson, 2006) and the Dutch spelling with the GLETSCHR (De Pessemier & Andries, 2009). A classical speeded lexical decision task was also administered in Dutch and English for the bilinguals, and in English for the monolinguals. The self-report questionnaire was an adaptation of the LEAP-Q (Marian, Blumenfeld, & Kaushanskaya, 2007). This questionnaire contained questions about language-switching frequency/skill, age of L2 acquisition, frequency of L2 use, and reading/auditory comprehension/speaking skills in L1 and L2 (for a detailed summary, see Tables B.1 and B.2 in Appendix B).

Two bilinguals were classified as lower intermediate L2 language users (50%–60%), ten were classified as upper intermediate L2 language users (60%–80%), and seven were scored as advanced L2 language users (80%–100%) according to the LexTALE norms reported by Lemhöfer and Broersma (2012).

Most importantly, the Dutch (L1) proficiency of the bilinguals was matched with the English proficiency of the monolinguals for all but subjective exposure (see Table 1), indicating that both groups were equally proficient in their first language, but the bilinguals had less relative exposure to their L1 than did the monolinguals. The English (L2) proficiency was clearly lower than the Dutch (L1) proficiency (see Table 1).

Materials

The participants read the novel *The Mysterious Affair at Styles* by Agatha Christie (1920; title in Dutch: *De zaak Styles*; see Appendix C for an excerpt). This novel was selected out of a pool of books that were available in a multitude of different languages (allowing for possible

future replication in other languages) and that did not have any copyright issues, since all of these books were selected from the Gutenberg collection, which is freely available on the Internet. We selected novels that could be read in 4 h. The remaining books were examined for difficulty, as indicated by the frequency distribution of the words that the book contained. The Kullback–Leibler divergence (D_{KL} ; Cover & Thomas, 1991)¹ was used to select the novel whose word frequency distribution was the most similar to the one in natural language use, as observed in the Subtlex database (Brysbaert & New, 2009; Keuleers, Brysbaert, & New, 2010). As additional measures of the difficulty of the book, we calculated two readability scores: the Flesch Reading Ease (Kincaid, Fishburne, Rogers, & Chissom, 1975), which returns a score between 0 and 100 (closer to 100 is easier to read), and the SMOG grade (McLaughlin, 1969), which indicated how many years of education are a prerequisite for understanding the text. The Flesch Reading Ease for the novel was 81.3, and the SMOG was 7.4, indicating that it has an above-average reading ease.

The monolinguals read only the English version of the novel. These participants read a total of 5,031 sentences. The bilinguals read chapters 1–7 in one language and 8–13 in the other. The order was counterbalanced, such that half of the participants read chapters 1–7 in their mother tongue (Dutch), and the other half read those chapters in their second language (English). One of the bilingual participants only read the first half of the novel, in English. The ten participants who read the first part of the novel in Dutch read

¹ The D_{KL} is nonsymmetric, and therefore we calculated it in both directions: from distribution A to distribution B and vice versa. The possible range of D_{KL} was $[0, +\infty]$, with 0 being identical distributions. The average value of $D_{KL}(A||B)$ and $D_{KL}(B||A)$ for *The Mysterious Affair at Styles* was .598.

Table 2 Descriptive statistics of the Dutch and the English version of the novel *The Mysterious Case at Styles* by Agatha Christie

	Dutch			English		
	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range
Number of words	59,716			54,364		
Number of word types	5,575			5,012		
Number of nouns	7,987			7,639		
Number of noun types	1,777			1,742		
Number of sentences	5,190			5,300		
Number of words per sentence	11.64	8.86	[1–60]	10.64	8.20	[1–69]
Word frequency ^a	4.51	1.39	[0.30–6.24]	4.59	1.37	[0.30–6.33]
Word length	4.51	2.54	[1–22]	4.18	2.30	[1–17]

^a \log_{10} -transformed Subtlex frequencies: Subtlex-NL for Dutch words (Keuleers, Brysbaert, et al. 2010), Subtlex-UK for English words (Brysbaert & New, 2009)

2,754 Dutch sentences and 2,449 English sentences. The eight participants who read the first part of the novel in English read 2,852 English sentences and 2,436 Dutch sentences. The participant who only read the first part of the novel in English read 2,852 English sentences. In total, we collected eye movements for 59,716 Dutch words (5,575 unique types) and 54,364 English words (5,012 unique types). A summary of the characteristics of the Dutch and English versions of the novel is presented in Table 2.

Apparatus

The bilingual eye movement data were recorded with a tower-mounted EyeLink 1000 system (SR Research, Canada) with a sampling rate of 1 kHz. A chinrest was used to reduce head movements. Monolingual eye movement data were acquired with the same system that was desktop-mounted. The presentation of the material and recording of the eye movements were all implemented by Experiment Builder (SR Research Ltd.). Reading was always binocular, but eye movements were recorded from the right eye only. Text was presented in black 14-point Courier New font on a light gray background. The lines were triple spaced, and three characters subtended 1 degree of visual angle or 30 pixels. Text appeared in paragraphs on the screen. A maximum of 145 words were presented on one screen. During the presentation of the novel, the room was dimly illuminated.

Procedure

Each participant read the entire novel in four sessions of an hour and a half apiece. In the first session, every participant read chapter 1 to 4; in the second session,

chapters 5 to 7; in the third session, chapters 8 to 10; and in the fourth session, chapters 11 to 13. Every bilingual and monolingual participant completed a number of language proficiency tests. The results of these proficiency measures can be found in Table 1.

The participants were instructed to read the novel silently while the eyetracker recorded their eye movements. It was stressed that they should move their head and body as little as possible while they were reading. The participants were informed that there would be a break after each chapter and that during that break they would be presented with multiple-choice questions about the contents of the book (Comprehension scores are also reported in Table 1). This was done to ensure that participants understood what they were reading and paid attention throughout the session. The number of questions per chapter was relative to the amount of text in that chapter.

The text of the novel appeared on the screen in paragraphs. When participants finished reading the sentences on one screen, they pressed a button on the control pad to move to the next part of the novel.

Before starting the practice trials, a 9-point calibration was executed. The participants were presented with three practice trials in which the first part of another story was presented on the screen. After these trials, the participants were asked two multiple-choice questions about the content of the practice story. This part was intended to familiarize participants with the reading of text on a screen and the nature and difficulty of the questions. Before the participants started reading the first chapter, another 9-point calibration was carried out. After the initial calibration, recalibration was carried out every 10 min. Furthermore, each time participants turned to the next screen, a drift correction was included. If the error exceeded 0.5°, a recalibration was also performed.

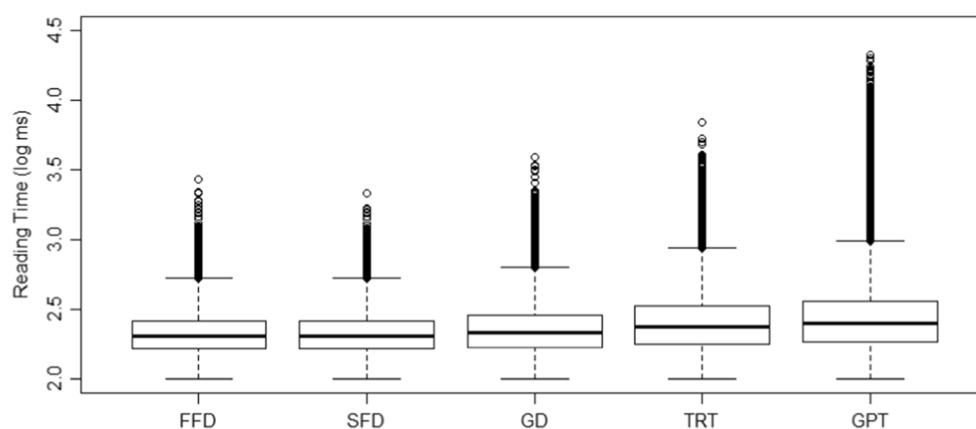


Fig. 1 Boxplots of log-transformed reading time data (on the y -axis, in seconds) for English monolinguals. Boxes denote the median (thick line) and the lower and upper quartiles

Results and discussion

We will focus on the distribution and descriptive statistics of five word-level reading time measures extracted from the GECCO: (a) *first fixation duration* (FFD), the duration of the first fixation landing on the current word; (b) *single fixation duration* (SFD), the duration of the first and only fixation on the current word; (c) *gaze duration* (GD), the sum of all fixations on the current word in the first-pass reading before the eye moves out of the word;

(d) *total reading time* (TRT), the sum of all fixation durations on the current word, including regressions; and (e) *go-past time* (GPT), the sum of all fixations prior to progressing to the right of the current word, including regressions to previous words that originated from the current word.

Fixations that were shorter than 100 ms were excluded from the analyses (but are available in the online dataset), because these are unlikely to reflect language processing (e.g., Sereno & Rayner, 2003). Words that were skipped are

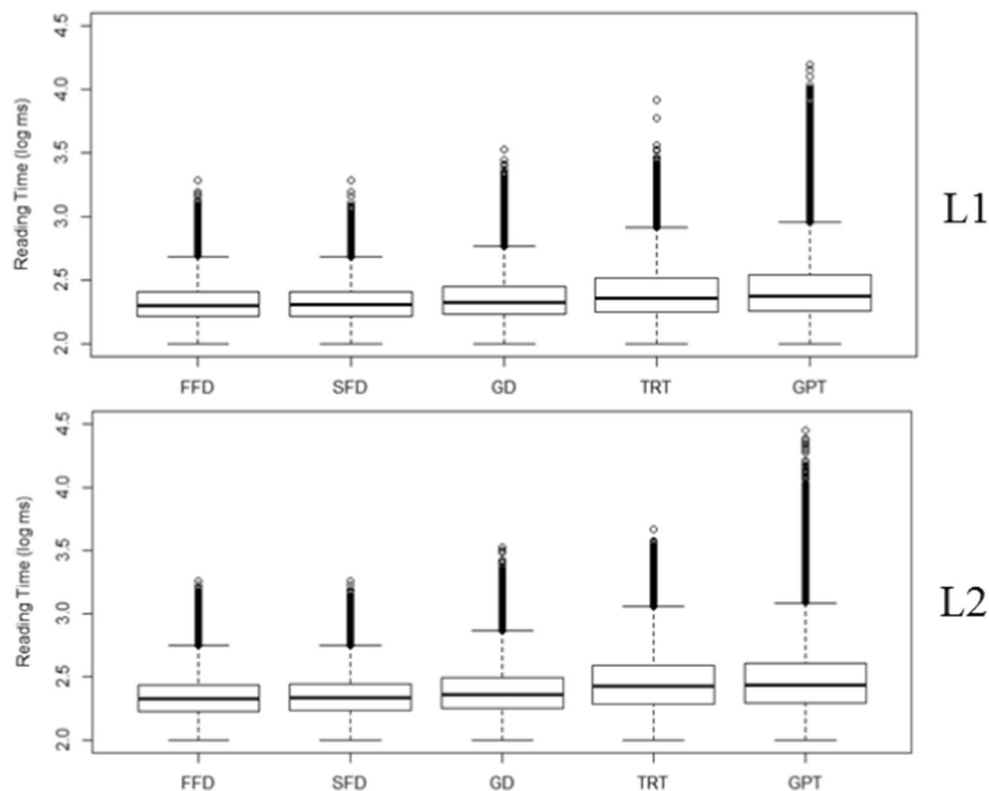


Fig. 2 Boxplots of log-transformed reading time data (on the y -axis, in seconds) for bilinguals in L1 (upper plot) and L2 (lower plot). Boxes denote the median (thick line) and the lower and upper quartiles

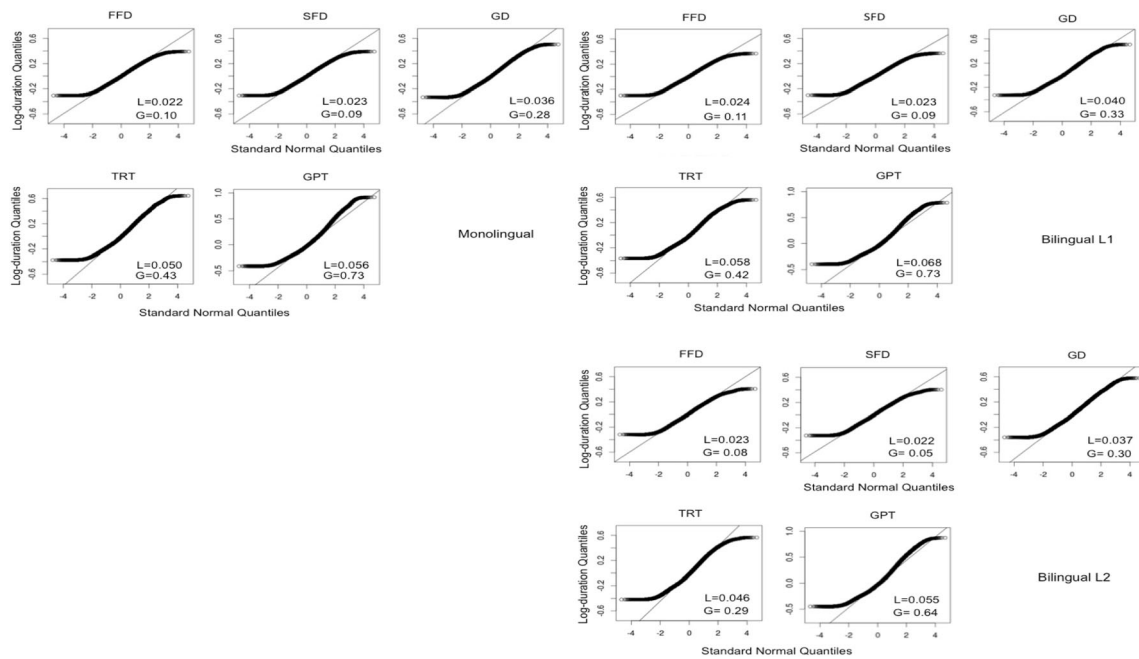


Fig. 3 Quantile–quantile plots of standardized log-transformed trimmed reading time durations against a standard normal distribution. Statistic values of the Lilliefors test of normality (L) and the Pearson’s moment coefficient of skewness (G) are presented on the plots. A larger value for L

corresponds to larger deviation from the standard normal distribution. Positive values for G indicate a positive skewness, and larger values indicate larger skewness

excluded in the rest of the description of the data. R (R Development Core Team, 2014) was used for all analyses.

Distribution of reading times

Figures 1 and 2 show boxplots of all reading time measures after log transformation and aggregation over participants. As we can see, the reading time variables are not normally distributed. Due to the exclusion criteria, they all show a minimal value of 100 ms. They also show a large number of reading time observations that are positive outliers.

To correct for these outliers, we removed all reading times that deviated more than 2.5 standard deviations from the participant mean per language. The quantile–quantile plots of the log-transformed and trimmed reading times are presented in

Fig. 3. The Lilliefors normality test statistic (L) is included in all panels. The p value is smaller than .001 in all cases. This means that, despite the trimming and log transformation, the reading times were not normally distributed. The measures that approximated a normal distribution the most were SFDs and FFDs. The Pearson’s moment coefficient of skewness (G) is also included in the panels. All G values are positive. This means that the reading times were all positively skewed (i.e., to the right). We can see that TRTs and GPTs are more skewed than FFDs and GDs.

We refer to Frank et al. (2013) for a similar analysis of the distribution of reading times. Their results also showed that despite log transformation, the reading times gathered by eyetracking are often not normally distributed and are skewed to the right. This feature of our data must be taken into account

Table 3 Averages (M), standard deviations (SD), and ranges of the reading time measures for monolingual, bilingual L1, and bilingual L2 reading

	Monolingual (English)			Bilingual L1 (Dutch)			Bilingual L2 (English)		
	M	SD	Range	M	SD	Range	M	SD	Range
First-fixation duration	214	70	101–502	209	65	101–467	222	74	101–536
Single-fixation duration	215	69	101–490	210	64	101–464	224	74	101–540
Gaze duration	232	89	101–695	226	85	101–682	250	105	101–877
Total reading time	264	127	101–1,060	256	117	101–852	296	194	101–978
Go past time	298	187	101–2,140	286	168	101–1,540	332	218	101–2,130

Table 4 Spearman–Brown split-half reliability coefficients for timed measures in the GECO database

	Monolinguals	Bilinguals L1	Bilinguals L2
First fixation duration	.649	.611	.640
Single fixation duration	.701	.701	.742
Gaze duration	.883	.844	.864
Total reading time	.907	.870	.901
Go-past time	.765	.742	.780

when choosing the preferred statistical technique for analyzing the data.

Description reading times

In Table 3, we present the means of FFD, SFD, GD, TRT, and GPT for monolingual reading and L1 and L2 reading, after trimming. Standard deviations and the ranges of values are also given. Standard deviations are larger on average for L2 reading. This means that for L2 reading, there is more variance in reading times. The larger range in language proficiency for L2 than for L1 might account for this difference in variances. We can see clearly that reading times are longer for L2 reading than for L1 or monolingual reading. We have discussed these differences in depth in Cop, Drieghe, et al. (2015).

Interindividual consistency of reading times

Because it is known that reading behavior is subject to inter-individual variance, we determined the level of consistency of reading times of the large sample of stimuli across participants. For all stimuli, we calculated the split-half correlations between two halves of the participants in every language condition, and corrected these for length by applying the Spearman–Brown formula (a procedure also applied in the DLP and BLP; Keuleers, Diependaele, et al., 2010; Keuleers et al., 2012). We used the psych package (Revelle, 2015) in R for these calculations. Even though the number of stimuli is very large, the number of readers is rather low. The results, however, show high to very high consistency of the reading times (see Table 4), which illustrates the reliability of megadatasets like GECO.² In terms of early reading measures, SFD seems to be preferable over FFD when analyzing the corpus, because the reliability scores are higher for this measure.

² These coefficients are indeed comparable to those of word-level corpora. The split-half correlation for the items from the DLP was .79 for reaction times, and for those from the BLP it was .72.

Skipping probability

In addition to fixation durations, an important variable in eye movement studies of reading is the skipping probability of words. This metric represents the chance that a word will not receive a fixation in the first pass. It is a marker of the parafoveal processing of words and is, for example, influenced by word length and predictability (Brysbaert & Vitu, 1998; Rayner, 1998; Rayner, Slattery, Drieghe, & Liversedge, 2011). Skipping probability is also embedded in models of eye movements such as the E-Z reader model (Reichle et al., 2011).

In Table 5, the average skipping probabilities are presented for the trimmed dataset (i.e., no fixations below 100 ms were included). About a third of the words are skipped while participants were reading the novel, which is similar to the proportions of skips in comparable eyetracking research (Rayner, 1998). In Fig. ZX4, we present the effect of word length on skipping probability. There is a clear decrease of word skipping with an increase of word length, which is also consistent with previous research (Drieghe, Brysbaert, Desmet, & De Baecke, 2004; Rayner et al., 2011). For a more in-depth discussion of the skipping probabilities in GECO and a further comparison between L1, L2, and monolingual reading, we refer the reader to Cop, Drieghe, et al. (2015).

Conclusion

In this article, we present the first eyetracking corpus of natural reading specifically aimed at bilingual sentence reading, the GECO, and make it available for free use in future research. Participants were selected for their language history, and detailed proficiency measures were gathered. The GECO data are freely available online for other researchers to analyze and use, provided that reference to this article and corpus is made in the resulting writings. The data are perfectly suited for studies at one or multiple levels of language processing (e.g., at the word level, sentence level, and semantic level). They allow for investigating specific research questions concerning L1 and L2 reading (e.g., differences in (cross-lingual) neighborhood effects or age-of-acquisition effects between L1 and L2), but also for examining the effects of L2 learning on L1 reading by comparing monolingual and bilingual L1 reading. Furthermore, the data can be useful for modeling or running virtual experiments. The novel that was used has been translated into more than 25 languages, including Hebrew, Finnish, and Japanese. This opens up possibilities for further data collection by other researchers to enable the comparison of natural reading across

Table 5 Averages (*M*), standard deviations (*SD*), and ranges of the skipping probabilities for monolingual, bilingual L1, and bilingual L2 reading

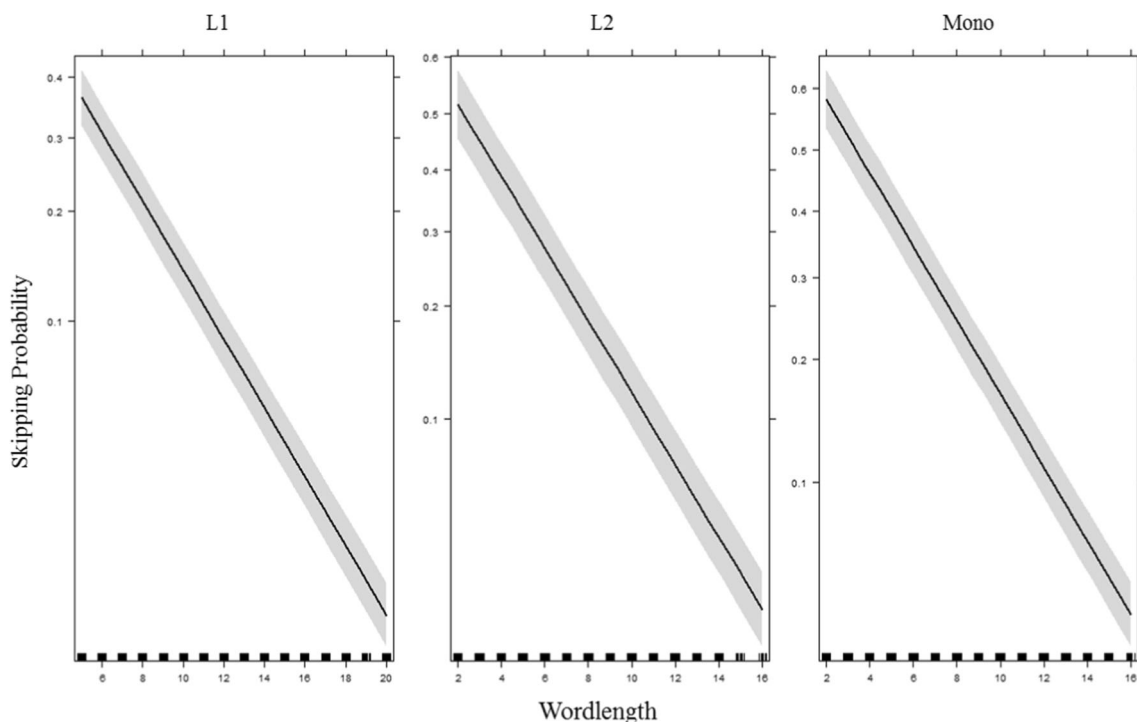
	Monolingual (English)			Bilingual L1 (Dutch)			Bilingual L2 (English)		
	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range
Average skipping probability	.38	.08	.22–.52	.34	.09	.17–.47	.31	.10	.08–.52

languages and to study bilingualism in different populations and language combinations.

Of course, there are some limitations to the use of a natural eyetracking corpus. First, it is much more difficult to control confounding factors than with a more rigorously managed setting consisting of an experimentally controlled stimulus set. However, if a suitable metric is available, the size of the dataset does allow the inclusion of possible confounding factors as covariates in the statistical model. Second, although the size of the dataset surpasses any individual experiment by far in terms of the included stimuli, it is possible that some cases or combinations of word characteristics that may be of special interest are underrepresented (e.g., extremely high- or low-frequency words, or long words that are high in frequency). For such special cases, generalization of results from these items may be compromised, due to the small number of observations. However, because the corpus contains more than 5,000 unique words for each language, it should be possible to obtain a meaningful set of results that applies to the general reading of a novel in L1 and L2.

Another potential limitation of the present corpus is the difference between the mother tongues of the participants: For the monolingual group, this was English, whereas it was Dutch for the bilinguals. This follows from the choice to keep language constant for the comparison between monolingual and L2 reading. However, a global comparison of sentence reading times, skipping probabilities, and regression probabilities yielded no significant differences between the monolinguals and the L1 of the bilinguals (Cop et al., 2015).

With this corpus, models of bilingual language processing can be evaluated, compared, and simulated using one large dataset of bilingual eye movements. This corpus can also be used to test specific hypotheses about the differences between L1 and L2 reading or between bilingual and monolingual reading. Interesting questions, for example, are whether bilinguals might use less prediction in reading than monolinguals do, or whether specific syntactic constructions are processed differently in L2 than in L1 reading. Another important contribution of this corpus is of a more exploratory nature. The richness in these eyetracking data has the potential to inspire a very wide

**Fig. 4** Effect of word length (*x*-axis) on the skipping probabilities (*y*-axis) for monolinguals and bilinguals (L1 and L2)

range of research, yielding new theoretical questions and insights about the time course of reading and specific interactions between multiple levels of a language user system.

Author note This research was funded by an aspirant scholarship from the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (FWO) and by a concerted research action, Grant No. BOF13/GOA/032, from Ghent University.

References

- Altarriba, J., Kroll, J. F., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory & Cognition, 24*, 477–492.
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445–459. doi:10.3758/BF03193014
- Bamgbose, A. (1994). Nigeria's choice. *UNESCO Courier*, pp. 22–31.
- Bijeljac-Babic, R., Biardeau, A., & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory & Cognition, 25*, 447–457.
- Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research, 2*, 1.
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977–990. doi:10.3758/BRM.41.4.977
- Brysbaert, M., & Vitu, F. (1998). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 125–147). Oxford, UK: Elsevier.
- Calvo, M. G., & Meseguer, E. (2002). Eye movements and processing stages in reading: Relative contribution of visual, lexical, and contextual factors. *Spanish Journal of Psychology, 5*, 66–77. doi:10.1017/S1138741600005849
- Christie, A. (1920). *The mysterious affair at styles*. N.p.: John Lane. Retrieved on 7 November, 2012, from Project Gutenberg, www.gutenberg.org
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256. doi:10.1037/0033-295X.108.1.204
- Cop, U., Drieghe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS ONE, 10*, e0134008. doi:10.1371/journal.pone.0134008
- Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review, 22*, 1216–1234. doi:10.3758/s13423-015-0819-2
- Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York, NY: Wiley.
- Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition, 109*, 193–210. doi:10.1016/j.cognition.2008.07.008
- De Pessemier, P., & Andries, C. (2009). *GL&SCHR*. Leuven/Apeldoorn: Garant.
- Diependaele, K., Brysbaert, M., & Neri, P. (2012). How noisy is lexical decision? *Frontiers in Psychology, 3*, 348. doi:10.3389/fpsyg.2012.00348
- Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on inter-lingual homograph recognition. *Journal of Memory and Language, 42*, 445–464.
- Dijkstra, T., & van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189–225). Mahwah, NJ: Erlbaum.
- Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*, 175–197. doi:10.1017/S1366728902003012
- Dilkina, K., McClelland, J. L., & Plaut, D. C. (2010). Are there mental lexicons? The role of semantics in lexical decision. *Brain Research, 1365*, 66–81. doi:10.1016/j.brainres.2010.09.057
- Drieghe, D., Brysbaert, M., Desmet, T., & De Baecke, C. (2004). Word skipping in reading: On the interplay of linguistic and visual factors. *European Journal of Cognitive Psychology, 16*, 79–103. doi:10.1080/09541440340000141
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*, 777–813. doi:10.1037/0033-295X.112.4.777
- European Union & European Commission for Education and Culture. (2012). *Europeans and their languages*. Retrieved from http://ec.europa.eu/public_opinion/archives
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., & Pallier, C. (2010). The French Lexicon Project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods, 42*, 488–496. doi:10.3758/BRM.42.2.488
- Fossum, V., & Levy, R. (2012). Sequential vs. hierarchical syntactic models of human incremental sentence processing. In *Proceedings of the 3rd Workshop on Cognitive Modeling and Computational Linguistics* (pp. 61–69). New York, NY: Association for Computational Linguistics. Retrieved from <http://dl.acm.org/citation.cfm?id=2390313>
- Frank, S. L., & Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science, 22*, 829–834. doi:10.1177/0956797611409589
- Frank, S. L., Fernandez Monsalve, I., Thompson, R. L., & Vigliocco, G. (2013). Reading time data for evaluating broad-coverage models of English sentence processing. *Behavior Research Methods, 45*, 1182–1190. doi:10.3758/s13428-012-0313-y
- Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and Language, 50*, 259–281. doi:10.1006/brln.1995.1048
- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58*, 787–814. doi:10.1016/j.jml.2007.07.001
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103*, 518–565. doi:10.1037/0033-295X.103.3.518
- Grosjean, F. (1992). Another view of bilingualism. In R. J. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 51–62). Oxford, UK: North-Holland.
- Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111*, 662–720. doi:10.1037/0033-295X.111.3.662

- Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*, 431–439. doi:10.3758/BF03208203
- Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, *45*, 153–168. doi:10.1016/j.visres.2004.07.037
- Kennedy, A., & Pynte, J. (2008). The consequences of violations to reading order: An eye movement analysis. *Vision Research*, *48*, 2309–2320. doi:10.1016/j.visres.2008.07.007
- Kennedy, A., Pynte, J., Murray, W. S., & Paul, S.-A. (2013). Frequency and predictability effects in the Dundee Corpus: An eye movement analysis. *Quarterly Journal of Experimental Psychology*, *66*, 601–618. doi:10.1080/17470218.2012.676054
- Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*, 643–650. doi:10.3758/BRM.42.3.643
- Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice effects in large-scale visual word recognition studies: A lexical decision study on 14,000 Dutch mono- and disyllabic words and nonwords. *Frontiers in Psychology*, *1*, 174. doi:10.3389/fpsyg.2010.00174
- Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*, 287–304. doi:10.3758/s13428-011-0118-4
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel (Research Branch Report 8-75)*. Millington, TN: Naval Technical Training, U.S. Naval Air Station.
- Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, *135*, 12–35. doi:10.1037/0096-3445.135.1.12
- Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Valdes Kroff, J. (2012). Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 56, pp. 229–262). San Diego, CA: Academic Press.
- Kuperman, V., Dambacher, M., Nuthmann, A., & Kliegl, R. (2010). The effect of word position on eye-movements in sentence and paragraph reading. *Quarterly Journal of Experimental Psychology*, *63*, 1838–1857. doi:10.1080/17470211003602412
- Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2012). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies. *Quarterly Journal of Experimental Psychology*, *66*, 563–580. doi:10.1080/17470218.2012.658820
- Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 802–823. doi:10.1037/a0030859
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343. doi:10.3758/s13428-011-0146-0
- Libben, M. R., & Titone, D. a. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 381–390. doi:10.1037/a0014875
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*, 940–967.
- McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of Reading*, *12*, 639–646. doi:10.1039/b105878a
- Miellat, S., Sparrow, L., & Sereno, S. C. (2007). Word frequency and predictability effects in reading French: An evaluation of the E-Z Reader model. *Psychonomic Bulletin & Review*, *14*, 762–769. doi:10.3758/BF03196834
- Miller, L. M., Cohen, J. A., & Wingfield, A. (2006). Contextual knowledge reduces demands on working memory during reading. *Memory & Cognition*, *34*, 1355–1367. doi:10.3758/BF03193277
- Mitchell, J., Lapata, M., Demberg, V., & Keller, F. (2010). Syntactic and semantic factors in processing difficulty: An integrated measure. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (pp. 196–206). New York, NY: Association for Computational Linguistics. Retrieved from <http://dl.acm.org/citation.cfm?id=1858702>
- Norris, D., & Kinoshita, S. (2012). Orthographic processing is universal: It's what you do with it that's different. *Behavioral and Brain Sciences*, *1*, 34–35.
- Pynte, J., & Kennedy, A. (2006). An influence over eye movements in reading exerted from beyond the level of the word: Evidence from reading English and French. *Vision Research*, *46*, 3786–3801. doi:10.1016/j.visres.2006.07.004
- Pynte, J., & Kennedy, A. (2007). The influence of punctuation and word class on distributed processing in normal reading. *Vision Research*, *47*, 1215–1227. doi:10.1016/j.visres.2006.12.006
- Pynte, J., New, B., & Kennedy, A. (2008). On-line contextual influences during reading normal text: A multiple-regression analysis. *Vision Research*, *48*, 2172–2183. doi:10.1016/j.visres.2008.02.004
- Pynte, J., New, B., & Kennedy, A. (2009a). On-line contextual influences during reading normal text: The role of nouns, verbs and adjectives. *Vision Research*, *49*, 544–552. doi:10.1016/j.visres.2008.12.016
- Pynte, J., New, B., & Kennedy, A. (2009b). On-line syntactic and semantic influences in reading revisited. *Journal of Eye Movement Research*, *3*, 12.
- R Development Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org
- Radach, R. (1996). *Blickbewegungen beim Lesen: Psychologische Aspekte der Determination von Fixationspositionen [Eye movements in reading: Psychological factors that determine fixation locations]*. Münster, Germany: Waxmann.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422. doi:10.1037/0033-2909.124.3.372
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*, 1457–1506. doi:10.1080/17470210902816461
- Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception & Psychophysics*, *58*, 734–747. doi:10.3758/BF03213106
- Rayner, K., Li, X., & Pollatsek, A. (2007). Extending the E-Z Reader model of eye movement control to chinese readers. *Cognitive Science*, *31*, 1021–1033. doi:10.1080/03640210701703824
- Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*, 448–465. doi:10.1037/0882-7974.21.3.448
- Rayner, K., Slattery, T. J., Drieghe, D., & Liversedge, S. P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 514–528. doi:10.1037/a0020990
- Reichle, E. D., & Drieghe, D. (2015). Using E-Z reader to examine the consequences of fixation-location measurement error. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 262–270. doi:10.1037/a0037090
- Reichle, E. D., Liversedge, S. P., Drieghe, D., Blythe, H. I., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2013). Using E-Z Reader to examine the concurrent development of eye-movement control

- and reading skill. *Developmental Review*, 33, 110–149. doi:10.1016/j.dr.2013.03.001
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105, 125–157. doi:10.1037/0033-295X.105.1.125
- Reichle, E. D., Tokowicz, N., Liu, Y., & Perfetti, C. A. (2011). Testing an assumption of the E-Z Reader model of eye-movement control during reading: Using event-related potentials to examine the familiarity check. *Psychophysiology*, 48, 993–1003. doi:10.1111/j.1469-8986.2011.01169.x
- Revelle, W. (2015) *psych: Procedures for personality and psychological research*. Northwestern University, Evanston, Illinois, USA. Retrieved from <http://CRAN.R-project.org/package=psych>
- Seidenberg, M., & Waters, G. (1989). Reading words aloud: A mega study. *Bulletin of the Psychonomic Society*, 27, 489–489.
- Sereno, S., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, 7, 489–493. doi:10.1016/j.tics.2003.09.010
- Shin, H. B., & Kominski, R. A. (2007). *Language use in the United States: 2007*. Washington, DC: U.S. Census Bureau.
- Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, 64, 88–107. doi:10.1016/j.jml.2010.08.006
- van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, 39, 458–483. doi:10.1006/jmla.1998.2584
- Whitney, C. (2011). Location, location, location: How it affects the neighborhood (effect). *Brain and Language*, 118, 90–104. doi:10.1016/j.bandl.2011.03.001
- Wilkinson, G. S., & Robertson, G. (2006). *Wide Range Achievement Test 4 professional manual*. Lutz, FL: Psychological Assessment Resources.
- Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, 60, 502–529. doi:10.1016/j.jml.2009.02.001