Presenting GECO: An Eye-tracking Corpus of Monolingual and Bilingual Sentence Reading

Uschi Cop[1], Nicolas Dirix[1*], Denis Drieghe[2] and Wouter Duyck[1]

[1] Department of Experimental Psychology, Ghent University, Ghent, Belgium

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

[*]Corresponding author, Nicolas Dirix, Henri Dunantlaan 2, 9000 Ghent,

nicolas.dirix@ugent.be

Abstract

This paper introduces GECO, the Ghent Eye-tracking Corpus, a monolingual and bilingual corpus of eye-tracking 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 paper we describe the distributions and descriptive statistics of the most important reading time measures for the two groups of participants. This large eye-tracking corpus is perfectly suited for both exploratory purposes as well as 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 reading of long texts and narratives. The corpus is freely available at http://expsy.ugent.be/downloads/geco.

Introduction

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 to evaluate 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 that collect these large databases are often referred to as corpus studies or mega studies (e.g., Balota et al., 2007; Seidenberg & Waters, 1989). Because corpus studies gather a large amount of observations from a limited amount of participants, or vice versa, or both, 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 factorial designs used in traditional experimental settings, where 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 representativeness of processing characteristics and show only a part of possible language behavior. An advantage of the corpus approach is that effects of continuous lexical variables, such as word frequency, can be assessed over their full possible range, instead of a constricted one. Another advantage of large corpora of linguistic data is that it enables researchers to answer multiple hypotheses without the need to design a new experiment and gather new data, which is considerably time consuming, or requires expensive equipment (e.g., an eye-tracker).

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. (2007) gathered lexical decision latencies of 816 participants for 40 481 different American English words (3 400 responses on average per participants). 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 have also been used to evaluate complex computational models of word recognition (e.g. Norris & Kinoshita, 2012; Whitney, 2011), illustratingthe relevance and broad applicability of such big datasets.

Eye-tracking 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 the 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 all 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 data sets of natural text reading are necessary.

The technique of eye-tracking enables researchers to record the eye movements of participants during silent reading, with minimal instruction or interference on behalf of the researcher. Also, eye-tracking, in contrast to for example lexical decision tasks, captures language performance how it occurs in daily life, without interference of additional decision components or response mechanisms, which are inherent to lexical decision for instance. With modern day eye-tracking equipment, the position of the eye can be determined every millisecond with very high spatial accuracy, resulting in a very rich and detailed data set. 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 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. There is also an essential role for lexical access 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 eye-tracking data set including a large sample of stimuli considerably increases the richness of available eye movement data sets. Corpora of eye movements during naturalistic, contextualized reading of text will be invaluable in 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 of different processing levels (e.g. both at word and at sentence level) and the interactions among them simultaneously, as well as the specific time course of these effects. Moreover, the testing of predictions of language models in an eye-tracking corpus of natural reading could provide a test of the generalizability of parts or whole of the specific model, especially with regard to parts of the model which were inspired by findings obtained in less natural tasks.

Additionally, as already discussed for corpora of isolated word recognition, these eye-tracking 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 there is no need to continuously design new experiments or to 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 eye-tracking 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 1ms and spatial accuracy of 0.25 characters. Initially, the authors used this corpus to investigate the effect of parafoveal processing on foveal word inspection time (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 effect of syntactic and semantic constraints on fixation times (Pynte, New, & Kennedy, 2008, 2009a, 2009b), the effect of violations in reading order (Kennedy & Pynte, 2008) and the interaction between frequency and predictability (Kennedy, Pynte, Murray, & Paul, 2013) using the eye movement data of the Dundee corpus.

Other authors 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 was 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 ten percent 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 processing in sentence processing. Kuperman et al. (2012) used both the mega data of the Lexicon Project (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 the evaluation 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 seperately 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, 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 for example Boston, Hale, Kliegl, Patil, & Vasishth, 2008) and/or hypothesis testing. Some of these corpora contained monolingual reading in different languages, supporting generalizability of claims across languages. However, these existing datasets remain quite limited in their diversity of words and sentences, and have much less stimuli for instance than the large isolated word reading projects (e.g., the ELP).

In conclusion, it seems that corpora of eye movements 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 eye-tracking corpus.

Our Corpus: GECO

As the previous section shows, the building of eye-tracking 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 eye movements of participants reading an entire book (though see Radach (1996) for a corpus of 4 participants reading a selection of chapters of Gulliver’s Travels in German). This is quite surprising, as books have been read for hundreds of years in a multitude of contexts (e.g., work, studies or leisure). Our current approach allows answering several important questions. First, it would be highly interesting to examine whether findings of previous eye-tracking research using a limited set of stimuli will be preserved when put to the test in a database that contains a very large and wide range of stimuli, and not appearing in specially constructed sentences. Second, 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 (e.g. Calvo & Meseguer, 2002; Miellet, Sparrow, & Sereno, 2007; Miller, Cohen, & Wingfield, 2006). Therefore, an eye-tracking corpus of people reading a long narrative allows 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, is active.

Next, until now not a single, large eye movement database focused on, or even specified, possible differences in language knowledge between participants. All eye-tracking corpora (at least to our knowledge) implicitly assume that their participants have knowledge of only the language they are reading in. As bilingualism is most commonly defined as 'the regular use of two (or more) languages' (Grosjean, 1992), today, across most European countries, 54 % of the people are bi-or multi-linguals due to migration and the fact that foreign languages are a compulsory part of formal education (European Union & European Comission 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 non-native language (Shin & Kominski, 2007).

This is important because a plethora of evidence shows that bilingualism changes language processes and 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, there are no mega-data available for participants reading in their first language with a confirmed and assessed knowledge of another language, or of participants reading in a second language that they have acquired later in life. In short, there is no bilingual eye-tracking corpus available to researchers. In this paper we present the GECO, the Ghent Eye-tracking Corpus, which 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 L1 and the other half in L2. All participants read a total of about 5 000 sentences. The precise language history and proficiency score was gathered for all participants. This is 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 eye-tracking data are available as online supplementary materials. See Appendix A for a list of the available files and the exact contents of the files.

Exploitation of the current corpus

The data from the GECO corpus have been in two studies so far. By comparing the basic eye movement measures on sentence level between L1 and L2 reading (Cop, Drieghe, & Duyck, 2015), we provided a database of benchmark parameters of reading with attention for 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, shorter saccades, lower probability of skipping words,…). Furthermore, we observed that in L1 reading of continuous text there are no differences 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 effects of L2 learning on L1 use, 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 vs. 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 a smaller frequency effect 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 is the same for both groups. These findings are very relevant for theoretical models of monolingual and bilingual reading, and examples in themselves of the value of such data to investigate 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 model (Dijkstra & van Heuven, 2002). The authors mentioned that this model concerns the visual word recognition system and is part of a larger 'language user' system, which also includes sentence parsing and language production. They assume 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, it has not used eye movement data obtained from natural reading to inform the architecture or evaluate the system of word recognition they propose. Instead, this model has been adjusted from the BIA model (Dijkstra & Van Heuven, 1998) using the findings of 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), of 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 be able to evaluate the ecological validity of this word recognition model in a context of natural reading, but it should also be especially helpful to specify the exact nature of the interactions between the sentence context and the word recognition system. In their paper 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. Since bilingual participants read text in a unilingual context, the influence of the activation of lexical candidates (e.g., orthographic neighbors) of the non-target language could be a clear indication of a shared lexicon or non-selective 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 material is 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 is especially interesting 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 word level could be investigated and compared between these groups (e.g., orthographic (cross-lingual) neighbors, age of acquisition, homographs,…). Finally, next to word-level studies, the nature of the corpus also allows investigations of sentence-, semantic- or higher order levels of reading, which are almost non-existent for L2 reading.

We have already noted some of the differences between the current 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 of the Dundee corpus (Kennedy & Pynte, 2005) and reaction time data of the ELP (Balota et al., 2007). Our data could also be exploited by similar studies for comparing the monolingual data of the corpus to for instance the BLP (Keuleers et al., 2012), the L1 bilingual data to the DLP (Keuleers et al., 2010) and the L2 bilingual data to a potential future lexicon project in L2 (which is non-existent 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. As this model has proven to be successful in accommodating eye movement patterns of older (Rayner, Reichle, Stroud, Williams, & Pollatsek, 2006) and younger (Reichle et al., 2013) readers as well as non-alphabetic languages (Rayner, Li, & Pollatsek, 2007), we have reason to believe that it will perform well as a frame work for bilingual eye movement patterns. As 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 which is currently underexplored in eye-tracking research. Our participant group consists of both monolinguals and bilinguals, resulting in the first bilingual database of eye movements.

Method

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

Subjects

Nineteen unbalanced Dutch (L1) – English (L2) bilingual Ghent University and fourteen 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 age was 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 or master program of psychology. In the monolingual group, 6 males and 7 females participated. In the bilingual group, 2 males and 17 females participated. Participants had normal or corrected-to-normal vision. None of the participants reported to have any language and/or reading impairments.

The bilinguals started learning their L2 relatively late: The mean age of acquisition was eleven years [range: 5-14, sd = 2.46]. All participants completed a battery of language proficiency tests. This included a vocabulary test, a spelling test, a lexical decision task and a self-report language questionnaire (for results see Table 1). Vocabulary was tested with the LexTALE (Lexical Test for Advanced Learners of English, Lemhöfer & Broersma, 2012). This is an unspeeded lexical decision task, which is an indicator of language proficiency for intermediate to highly proficient language users 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 Table B.1 and B.2 in Appendix B).

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

Most important, 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 bilinguals had less relative exposure to their L1 than the monolinguals. The English (L2) proficiency is clearly lower than the Dutch (L1) proficiency (see Table 1).

Table 1

Average percentage scores [standard deviations] on the LexTALE, the Spelling test, the accuracy of the Lexical Decision task and Subjective Exposure and the score on the comprehension questions for the bilingual and monolingual group. T-values [degrees of freedom] of t-tests are presented in the last 2 columns.

	Monolinguals	Bilinguals L1	Bilinguals L2	t-value L1-L2	t-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<0.05, ** p<0.01, *** p<0.001

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 which did not have any copyright issues as all of these books were selected from the Gutenberg collection that is freely available on the Internet. We selected the novels that could be read in four hours. 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 Flesh Reading Ease for the novel was 81.3, the SMOG was 7.4, indicating that it has an above average reading ease.

[1] The D_{KL} is non-symmetric and therefore we calculated it in both directions: from distribution A to distribution B and vice versa. The possible range of $D_{KL} = [0,+\infty]$, 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.

The monolinguals read only the English version of the novel. These participants read a total of 5 031 sentences. 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), the other half read them in their second language (English). One of the bilingual participants only read the first half of the novel in English. The 10 participants reading the first part of the novel in Dutch, read 2 754 Dutch sentences and 2 449 English sentences. The 8 participants reading the first part of the novel in English, read 2 852 English sentences and 2 436 Dutch sentences. The participant that 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 version of the novel is presented in Table 2.

Table 2

Description of the Dutch and the English version of the novel 'The mysterious case at Styles.' by Agatha Christie

	Dutch			English		
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		
	M	SD	Range	M	SD	Range
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]Log10 transformed Subtlex frequencies: Subtlex-NL for Dutch words (Keuleers, Brysbaert, et al., 2010), Subtlex-UK for English words (Brysbaert & New, 2009).

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 grey background. The lines were triple spaced and 3 characters subtended 1 degree of visual angle or 30 pixels. Text appeared in paragraphs on the screen. A maximum of 145 words was 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. 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 chapter 11 to 13 were read. 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 eye tracker 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 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 the participant 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 nine-point calibration was executed. The participants were presented with three practice trials where 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 participant started reading the first chapter another nine-point calibration was carried out. After the initial calibration, re-calibration was carried out every 10 minutes. 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.

Results and discussion

We will focus on the distribution and descriptive statistics of five word-level reading time measures extracted from the GECO: 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 originating from the current word.

Fixations that were shorter than 100ms 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 were excluded in the rest of the description of the data. R (R Core Team, 2014) was used for all analyses.

Distribution of Reading Times

Figure 1 and 2 show boxplots of all reading time measures after log transformation and aggregation over subjects. 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.

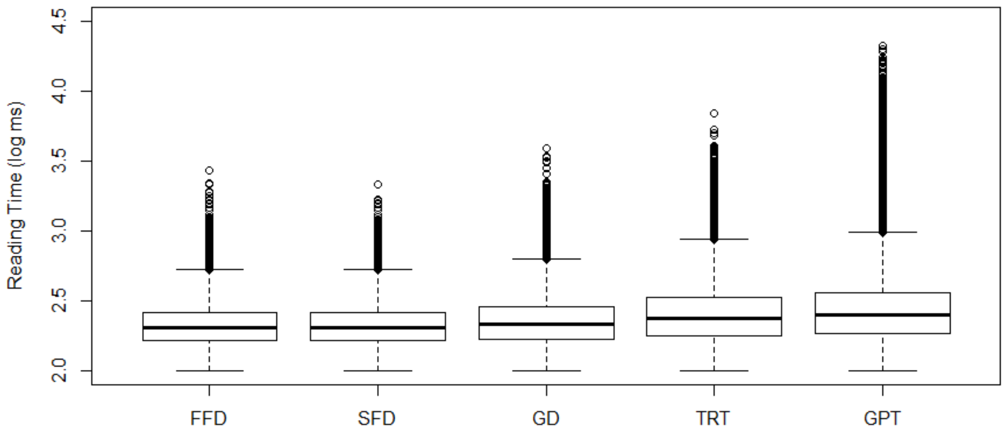


Figure 1. Boxplots of log-transformed reading time data (on the y-axis in seconds) for English monolinguals. Boxes denote the median (thick line), the lower and the upper quartile.

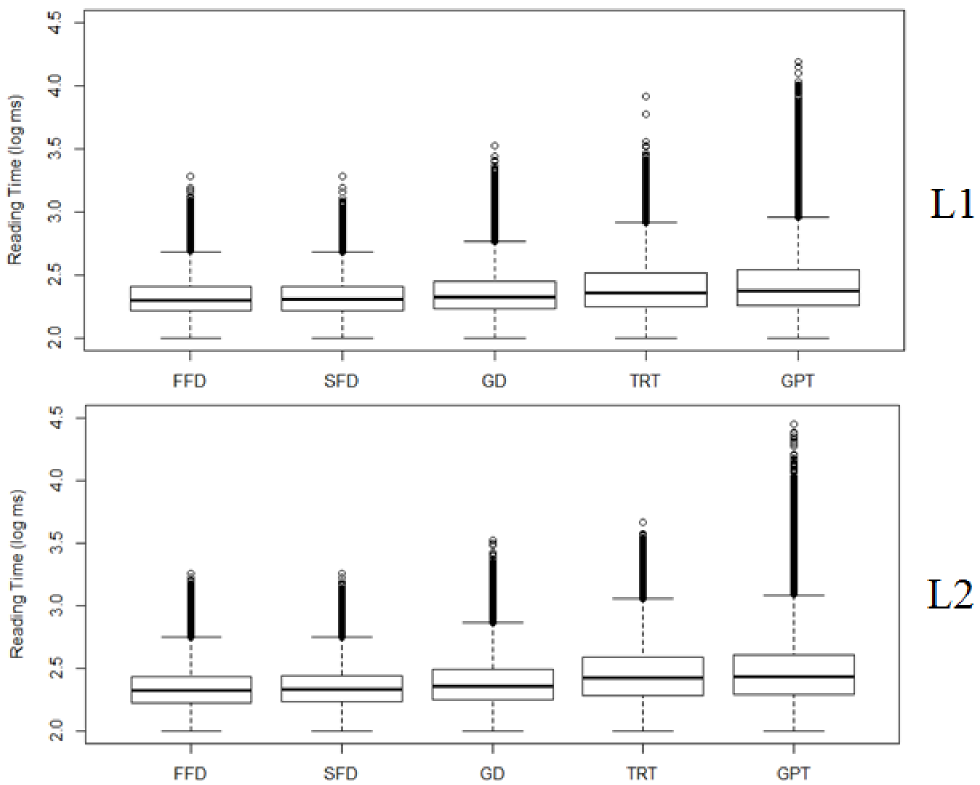
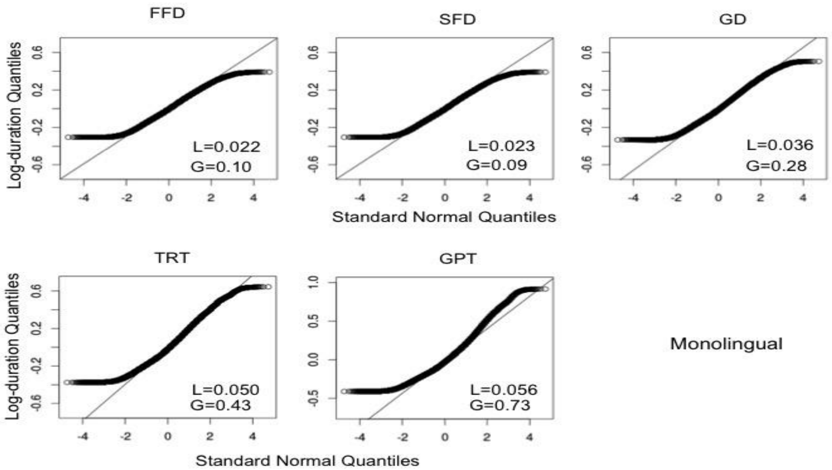
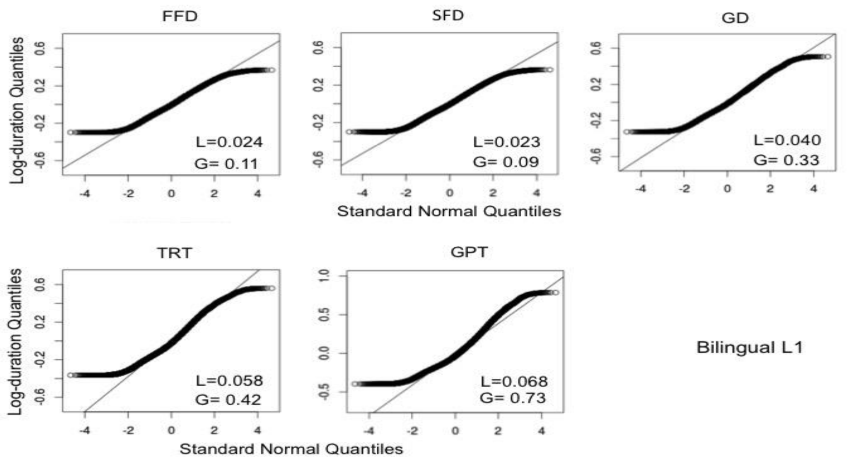


Figure 2. Boxplots of log-transformed reading time data (on the y-axis in seconds) for bilinguals in L1 (upper plot) and L2 (lower plot). Boxes denote the median (thick line), the lower and the upper quartile.

To correct for these outliers we removed all reading times that deviated more than 2.5 standard deviations from the subject mean per language. The quantile-quantile plots of the log-transformed and trimmed reading times are presented in Figure 3. The Lilliefors normality test statistic (L) is included in all panels. The p-value is smaller than 0.001 in all cases. This means that despite of trimming and log-transformation, the reading times were not normally distributed. The measures that approximated a normal distribution the most were single fixation durations and first fixation durations. 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 (skewed to the right). We can see that total reading times and go past times are more skewed than first fixation durations and gaze durations.





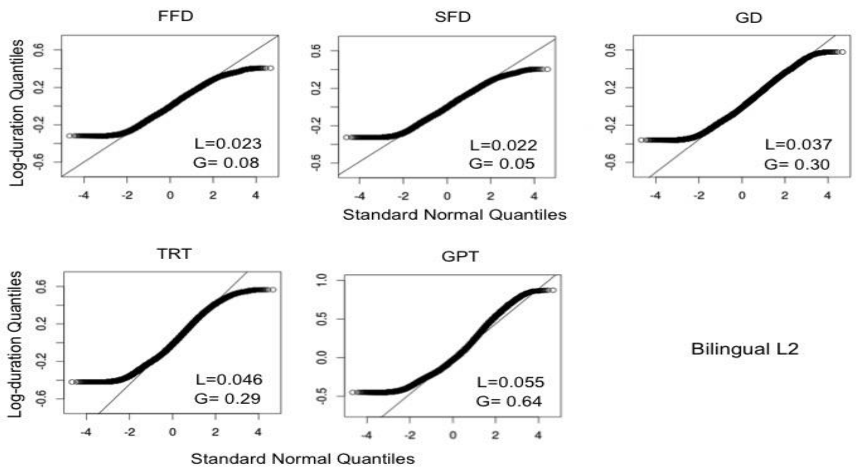


Figure 3. Quantile-quantile plots of standardized log-transformed trimmed reading time durations against a standard normal distribution. Statistic values of the Lilliefors test of normality (L) and the Pearson's moment coefficient of skewness (G) are presented on the plots. A larger value for L corresponds to larger deviation from the standard normal distribution. Positive values for G indicate a positive skewness, larger values indicate larger skewness.

We refer to Frank et al. (2013) for a similar analysis of the distribution of reading times. Their results also show that despite log-transformation, the reading times gathered by eye-tracking are often not normally distributed and are skewed to the right. This feature of our data must be taken account when choosing the preferred statistical technique for analyzing the data.

Description Reading Times

In Table 3 we present the means of first fixation duration, single fixation duration, gaze duration, total reading time and go past time for monolingual reading and L1 and L2 reading, after trimming. Standard deviations and the range 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 variance. We can see clearly that reading times are longer for L2 reading than for L1 or monolingual reading. We discussed these differences in depth in Cop et al. (2015).

Table 3

Averages (M), standard deviations (SD) and range 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	101-1060	256	117	101-852	296	194	101-978
Go Past Time	298	187	101-2140	286	168	101-1540	332	218	101-2130

Interindividual Consistency of Reading Times

As it is known that reading behavior is subject to interindividual 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 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 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 reading times (see Table 4), which illustrates the reliability of mega datasets 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.

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

Skipping probability

In addition to fxation durations, an important variable in eye movement studies of reading is the skipping probability of words. This metric represents the chance that a word does not receive a fixation in first pass. It is a marker of parafoveal processing of words and is for example influenced by word length and predictability (Brysbaert & Vitu, 1996; Rayner,

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

1998; Rayner, Slattery, Drieghe, & Liversedge, 2011). Skipping probability is also embedded in models of eye movements such as the E-Z reader model (Reichl, Tokowicz, Liu, & Perfetti, 2011).

Table 5

Averages (M), standard deviations (SD) and range of the skipping probabilities 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
Average Skipping Probability	.38	.08	.22-.52	.34	.09	.17-.47	.31	.10	.08-.52

In Table 5 the average skipping probabilities are presented for the trimmed dataset (i.e., no fixations below 100ms were included). About a third of the words are skipped while participants are reading the novel, which is similar to the proportion of skips in comparable eye-tracking research (Rayner, 1998). In Figure 4 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 to Cop et al. (2015).

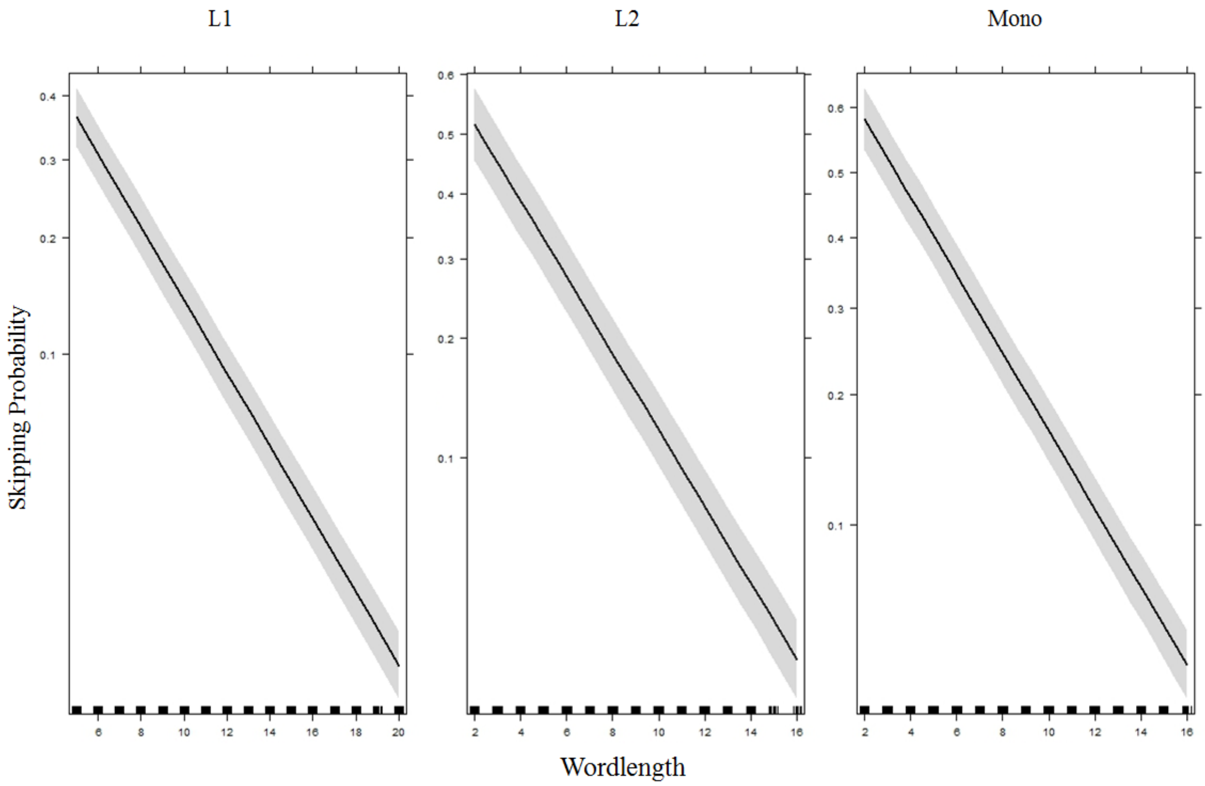


Figure 4. The effect of word length (x-axis) on skipping probabilities (y-axis) for monolinguals and bilinguals (L1 and L2).

Conclusion

In this paper, we present the first eye-tracking 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 on their language history and detailed proficiency measures were gathered. The GECO data is made freely available online for other researchers to analyze and use, provided reference to this paper and corpus is made in resulting writings. The data are perfectly suited for studies at one or multiple levels of language processing (e.g., word-level, sentence-level, 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 examing effects of L2 learning on L1 reading by comparing monolingual and bilingual L1 reading. Furthermore, the data can be useful for modelling or running virtual experiments. The novel that was used has been translated in 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 languages and to study bilingualism in different populations and language combinations.

Off course there are some limitations to the use of a natural eye-tracking corpus. First, it is much more difficult to control confounding factors compared to a more rigorously managed setting consisting of an experimentally controlled stimuli 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 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; 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 which applies to general reading of a novel in L1 and L2.

Another potential limitation of the current corpus is the difference between the mother tongues of the participants: for the monolingual group this is English whereas it is 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 differences between L1 and L2 reading or bilingual versus monolingual reading. Interesting questions are for example 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 these corpora is of a more exploratory nature. The richness in this eye-tracking data has potential in inspiring a very wide 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.

Acknowledgments

This research was funded by an aspirant scholarship of the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (FWO) and concerted research action BOF13/GOA/032 of 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*(4), 477–492.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., … Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445–459.

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*(4), 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), 1.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera 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*(4), 977–990. http://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., & Vitu, F. (1998). Word skipping: Implications for theories of eye movement control in reading. In G. Underword (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. *The Spanish Journal of Psychology, 5(*1), 66–77. doi:10.1017/S1138741600005849

Christie, A. *The Mysterious Affair at Styles*. N.p.: John Lane, 1920. *Project Gutenberg*. 1 Mar. 1997. Web. 07 Nov. 2012. <http://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*(1), 204–256. http://doi.org/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*(8), 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: Wiley.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, *109*(2), 193–210. http://doi.org/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*. http://doi.org/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). New York: Mahwah.

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*(03). http://doi.org/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. http://doi.org/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(1-2)*, 79–103. http://doi.org/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*(4), 777–813. http://doi.org/10.1037/0033-295X.112.4.777

European Union, & European Comission 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*(2), 488–496. http://doi.org/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). 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*(6), 829–834. http://doi.org/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*(4), 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*(3), 259–281. http://doi.org/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. http://doi.org/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*(3), 518–565. http://doi.org/10.1037/0033-295X.103.3.518

Grosjean, F. (1992). Another view of bilingualism. In *Cognitive processing in bilinguals* (pp. 51–62). Oxford, England: 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*(3), 662–720. http://doi.org/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*(6), 431–439. http://doi.org/10.3758/BF03208203

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, *45*(2), 153–168. http://doi.org/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*(21), 2309–2320. http://doi.org/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. *The Quarterly Journal of*

Experimental Psychology, *66*(3), 601–618. http://doi.org/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*(3), 643–650. http://doi.org/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*. http://doi.org/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*(1), 287–304. http://doi.org/10.3758/s13428-011-0118-4

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*(1), 12–35. http://doi.org/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. Ross (Ed.), *The psychology of learning and motivation* (Vol. 56, pp. 229–262).

Kuperman, V., Dambacher, M., Nuthmann, A., & Kliegl, R. (2010). The effect of word position on eye-movements in sentence and paragraph reading. *The Quarterly Journal of Experimental Psychology*, *63*(9), 1838–1857. http://doi.org/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. *The Quarterly Journal of Experimental Psychology*, *66*(3), 563–580. http://doi.org/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*(3), 802–823. http://doi.org/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*(2), 325–343. http://doi.org/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 & Cognition*, *35*(2), 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*(4), 940–967.

McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of Reading, 12*(8), 639–646. doi:10.1039/b105878a

Miellet, 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 and Review, 14*(4), 762–769. doi:10.3758/BF03196834

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). 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*(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*(22), 3786–3801. http://doi.org/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*(9), 1215–1227. http://doi.org/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*(21), 2172–2183. http://doi.org/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*(5), 544–552. http://doi.org/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*(1), 12.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Radach, R. (1996). *Blickbewegungen beim Lesen: Psychologische Aspekte der Determination von Fiationspositionen* [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*(3), 372–422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. http://doi.org/10.1080/17470210902816461

Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception & Psychophysics*, *58*(5), 734–747. http://doi.org/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*(6), 1021–1033. http://doi.org/10.1080/03640210701703824

Rayner, K., Pollatsek, A., Liversedge, S. P., & Reichle, E. D. (2009). Eye movements and non-canonical reading: Comments on. *Vision Research*, *49*(17), 2232–2236. http://doi.org/10.1016/j.visres.2008.10.013

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*(3), 448–465. http://doi.org/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 Epxerimental Psychology: Human Perception and Performance, 37(2)*, 514-528. http://dx.doi.org/10.1037/a0020990

Reichle, E. D., & Drieghe, D. (2015). Using E-Z reader to examine the consequences of fixation-location measurement error. J*ournal of Experimental Psychology: Learning, Memory and Cognition, 41*, 262-270. http://doi.org/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*(2), 110–149. http://doi.org/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*(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(7),* 993–1003. http://doi.org/10.1111/j.1469-8986.2011.01169.x

Revelle, W. (2015) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, http://CRAN.R-project.org/package=psych Version = 1.5.8.

Seidenberg, M., & Waters, G. (1989). Reading Words Aloud - a Mega Study. *Bulletin of the Psychonomic Society*, *27*(6), 489–489.

Sereno, S., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, *7*(11), 489–493. doi:10.1016/j.tics.2003.09.010

Shin, H. B., & Kominski, R. A. (2007). Language Use in the United States: 2007. U.S. Census Bureau, Washinton DC.

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*(1), 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*(3), 458–483. http://doi.org/10.1006/jmla.1998.2584

Whitney, C. (2011). Location, location, location: How it affects the neighborhood (effect). *Brain and Language*, *118*(3), 90–104. http://doi.org/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*(4), 502–529. http://doi.org/10.1016/j.jml.2009.02.001

APPENDIX A

Table A.1

Description of the file 'SubjectInformation.xlsx'. Column names are in the first column and a description of the content in that column is presented in the second column.

Column Name	Description
PP_NR	The identification number of the participant.
GROUP	Factor indicating whether the participants belonged to the unbalanced bilingual ("bilingual") or monolingual group ("monolingual")
AGE	Age of the participant in years
SEX	Sex of the participant ("f"=female, "m"=male)
AOA_ENG	Age of Acquisition of the English language or zero when monolingual
%EXP_DUTCH	Percentage of daily language exposure to Dutch
%EXP_ENG	Percentage of daily language exposure to English
LEXTALE_DUTCH	Score on the Dutch LexTALE (Lexical Test for Advanced learners of English; Lemhöfer & Broersma, 2012), NA for monolinguals
LEXTALE_ENG	Score on the English LexTALE (Lexical Test for Advanced learners of English; Lemhöfer & Broersma, 2012)
SPELLING_DUTCH	Percentage score on the Dutch spelling test (GL&SCHR; De Pessemier & Andries ,2009)
SPELLING_ENG	Percentage score on the English spelling test (WRAT4; Dell, Harrold, & Dell, 2008)
COMPR_DUTCH	Percentage score on the multiple-choice questions for the Dutch chapters of the novel
COMPR_ENG	Percentage score on the multiple-choice questions for the English chapters of the novel
LEX_DEC_ACC_DUTCH	Percentage score of accuracy on the Dutch lexical decision task on the word trails, corrected for false positives.
LEX_DEC_ACC_ENG	Percentage score of accuracy on the English lexical decision task on the word trails, corrected for false positives.

Table A.2

Description of the files 'EnglishMaterials.xlsx' and 'DutchMaterials.xlsx'. Column and sheet names are in the first column and a description of the content in that column or sheet is presented in the second column.

Sheet Name	Description
ALL	Each word presented on a separate line.
NOUNS	Each noun of the novel presented on a separate line.
SENTENCE	Each sentence of the novel presented on a separate line.
Column Name	**Description**
WORD_ID	Identification number of the word. The first number refers to the part of the novel (1,2,3 or 4), the second number refers to the trail number, and the last number refers to the word number within the trial.
SENTENCE_ID	Identification number of the sentence. The first number refers to the part of the novel (1,2,3 or 4), the second number refers to the sentence number within the part.
CHRON_ID	Chronological identification number of the current word.
WORD	The word contained in the current interest area.
PART_OF_SPEECH	The syntactic function of the current word in the sentence context.
CONTENT_WORD	Factor denoting whether the current word is a content word ("1") or a function word ("0").
WORD_LENGTH	The number of characters of the current word.
IA_AREA	The size of the current interest area around the word in pixels.
IA_TOP	The top side pixel position of the current interest area around the word.
IA_BOTTOM	The bottom side pixel position of the current interest area around the word.
IA_LEFT	The left side pixel position of the current interest area around the word.
IA_RIGHT	The right side pixel position of the current interest area around the word.
IDENTICAL_COGNATE	Factor denoting whether the current word has an identical cognate in the other language ("1') or not ("0").
CORR_LEVENSHTEIN	The corrected levenshtein distance between the current word and its translation equivalent in the other language.

SENTENCE	The sentence referred to with the current sentence-ID.
NUMBER_WORDS_SENTENCE	The number of words in the current sentence.

Table A.3

Description of the files 'L1ReadingTimedata.xlsx', 'L2ReadingTimedata.xlsx', and 'MonolingualReadingTimedata.xlsx'. Column names are in the first column and a description of the content in that column is presented in the second column.

Column Name	Description
PP_NR	The identification number of the participant.
GROUP	Factor indicating whether the participants belonged to the unbalanced bilingual ("bilingual") or monolingual group ("monolingual")
LANGUAGE_RANK	Factor indicating whether the participants read this part in their first language ("L1") or their second ('L2").
LANGUAGE	Factor indicating in which language the current part was read ("Dutch" or "English").
PART	The number of the part of the novel.
TRIAL	The number of the trial.
TRIAL_FIXATION_COUNT	The total number of fixations in the current trial.
TRIAL_TOTAL_READING_TIME	Summation of all fixation durations in the current trial.
WORD_ID_WITHIN_TRIAL	Chronological identification number of the word within the trial.
WORD_ID	Identification number of the word. The first number refers to the part of the novel (1,2,3 or 4), the second number refers to the trail number, the last number refers to the word number within the trial.
WORD	The word contained in the current interest area.
WORD_AVERAGE_FIX_PUPIL_SIZE	Average pupil size across all fixations in the current word.
WORD_FIXATION_COUNT	Total fixation falling within the current word.
WORD_FIXATION_%	Percentage of all fixations in a trial falling in the current word.
WORD_RUN_COUNT	The number of times the current word was entered and left (runs).
WORD_FIRST_RUN_START_TIM	The start time of the first run of fixations in the

E	current word.
WORD_FIRST_RUN_END_TIME	The end time of the first run of fixations in the current word.
WORD_FIRST_RUN_FIXATION_COUNT	The number of all fixations in a trial falling in the first run of the current word.
WORD_FIRST_RUN_FIXATION_%	Percentage of all fixations in a trial falling in the first run of the current word.
WORD_GAZE_DURATION	Summation of all fixation durations in the first run within the current word.
WORD_SECOND_RUN_START_TIME	The start time of the second run of fixations in the current word.
WORD_SECOND_RUN_END_TIME	The end time of the second run of fixations in the current word.
WORD_SECOND_RUN_FIXATION_COUNT	The number of all fixations in a trial falling in the second run of the current word.
WORD_SECOND_RUN_FIXATION_%	Percentage of all fixations in a trial falling in the second run of the current word.
WORD_SECOND_RUN_DWELL_TIME	Summation of all fixation durations in the second run within the current word.
WORD_THIRD_RUN_START_TIME	The start time of the third run of fixations in the current word.
WORD_THIRD_RUN_END_TIME	The end time of the third run of fixations in the current word.
WORD_THIRD_RUN_FIXATION_%	Percentage of all fixations in a trial falling in the third run of the current word.
WORD_THIRD_RUN_DWELL_TIME	Summation of all fixation durations in the third run within the current word.
WORD_FIRST_FIXATION_DURATION	The duration of the first fixation that was within the current word.
WORD_FIRST_FIXATION_INDEX	The ordinal sequence of the first fixation that was within the current word.
WORD_FIRST_FIXATION_RUN_INDEX	The number of runs of fixations have occurred when a first fixation is made to the current word. The current run is included in the tally.
WORD_FIRST_FIXATION_TIME	The start time of the first fixation to enter the current word.
WORD_FIRST_FIXATION_VISITED_WORD_COUNT	The number of different words visited before the first fixation is made into the current word.
WORD_FIRST_FIXATION_X	The horizontal coordinate position of the first fixation that was within the current word.
WORD_FIRST_FIXATION_Y	The vertical coordinate position of the first fixation

	that was within the current word.
WORD_FIRST_FIX_PROGRESSIVE	Factor indicating whether later words have been visited before the first fixation enters the current word ("0") or not ("1").
WORD_SECOND_FIXATION_DURATION	The duration of the second fixation in the current word, regardless of the run.
WORD_SECOND_FIXATION_RUN	The run index of the second fixation in the current word.
WORD_SECOND_FIXATION_TIME	The time of the second fixation in the current word, regardless of run.
WORD_SECOND_FIXATION_X	The horizontal coordinate position of the second fixation that was within the current word.
WORD_SECOND_FIXATION_Y	The vertical coordinate position of the second fixation that was within the current word.
WORD_THIRD_FIXATION_DURATION	The duration of the third fixation in the current word, regardless of the run.
WORD_THIRD_FIXATION_RUN	The run index of the third fixation in the current word.
WORD_THIRD_FIXATION_TIME	The time of the third fixation in the current word, regardless of run.
WORD_THIRD_FIXATION_X	The horizontal coordinate position of the third fixation that was within the current word.
WORD_THIRD_FIXATION_Y	The vertical coordinate position of the third fixation that was within the current word.
WORD_LAST_FIXATION_DURATION	The duration of the last fixation in the current word, regardless of the run.
WORD_LAST_FIXATION_TIME	The time of the last fixation in the current word, regardless of run.
WORD_LAST_FIXATION_X	The horizontal coordinate position of the last fixation that was within the current word.
WORD_LAST_FIXATION_Y	The vertical coordinate position of the last fixation that was within the current word.
WORD_GO_PAST_TIME	Summation of all fixation durations from when the current word is first fixated until the eyes enter a word with a higher word identification number.
WORD_SELECTIVE_GO_PAST_TIME	Summation of all fixation durations in the current word from when the current word is first fixated until the eyes enter a word with a higher word identification number.
WORD_TOTAL_READING_TIME	Summation of all fixation durations in the current word.
WORD_TOTAL_READING_TIME	Percentage of trial time spent in the current word.

_%	
WORD_SPILLOVER	The duration of the first fixation made in the next word after leaving the current word in the first pass.
WORD_SKIP	A word is considered skipped (i.e., WORD_SKIP = 1) if no fixation occurred in first-pass reading.

APPENDIX B: Results Self-report Questionnaire

Table B.1

Count of bilingual participants agreeing and not agreeing on second language skills items.

Skills	Agree	Don't Agree
Carry on normal conversation in L2	19	0
Watch television shows in L2	19	0
Listen to music in L2	19	0
Read and comprehend questions in L2	19	0
Read books or articles in L2	19	0
No problems in understanding L1 speaker	18	1
Carry on a discussion in L2	17	2
Love speaking L2	16	3
Explain difficult situation in L2	15	4
Answer difficult questions in L2	12	7
Think in L2	11	8
Speak to myself in L2	10	9
Write in L2	8	11
Make no/ almost no mistakes in L2	6	13
Dream in L2	5	14

Table B.2

Count of bilingual participants agreeing and not agreeing on second language switching items.

Switching	Agree	Don't Agree
I'm sometimes in a tip of the tongue state	16	3
I sometimes can't get the right word	14	5
I use a different language when I do not remember a word	13	6
I often use different languages intermixed	9	10
I often use different languages intermixed without noticing	5	14
I sometimes speak in a language that my dialogue partner doesn't understand	5	14

APPENDIX C: Excerpt from "The mysterious affair at Styles"

The intense interest aroused in the public by what was known at the time as "The Styles Case" has now somewhat subsided. Nevertheless, in view of the world-wide notoriety which attended it, I have been asked, both by my friend Poirot and the family themselves, to write an account of the whole story. This, we trust, will effectually silence the sensational rumours which still persist.

I will therefore briefly set down the circumstances which led to my being connected with the affair.

I had been invalided home from the Front; and, after spending some months in a rather depressing Convalescent Home, was given a month's sick leave. Having no near relations or friends, I was trying to make up my mind what to do, when I ran across John Cavendish. I had seen very little of him for some years. Indeed, I had never known him particularly well. He was a good fifteen years my senior, for one thing, though he hardly looked his forty-five years. As a boy, though, I had often stayed at Styles, his mother's place in Essex.

We had a good yarn about old times, and it ended in his inviting me down to Styles to spend my leave there.

"The mater will be delighted to see you again—after all those years," he added.

"Your mother keeps well?" I asked.

"Oh, yes. I suppose you know that she has married again?"

I am afraid I showed my surprise rather plainly. Mrs. Cavendish, who had married John's father when he was a widower with two sons, had been a handsome woman of middle-age as I remembered her. She certainly could not be a day less than seventy now. I recalled her as an energetic, autocratic personality, somewhat inclined to charitable and social notoriety, with a fondness for opening bazaars and playing the Lady Bountiful. She was a most generous woman, and possessed a considerable fortune of her own.