The first- and second-language age of acquisition effect in first- and second-language book reading

Nicolas Dirix[1*] and Wouter Duyck[1]

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

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

nicolas.dirix@ugent.be

Abstract

The age of acquisition (AoA) effect in first/monolingual language processing has received much attention in psycholinguistic research. However, AoA effects in second language processing were only investigated rarely. In the current study, we investigated first (L1) and second language (L2) AoA effects in a combined eye tracking and mega study approach. We analyzed data of a corpus of eye movements to assess the time course of AoA effects on bilingual reading. We found an effect of L2 AoA in both early and late measures of L2 reading: fixation times were faster for words that were learned earlier in L2. This suggests that the L2 AoA effect has an influence throughout the entire L2 reading process, analogous to the L1 AoA effect. However, we are also the first to find an early effect of L1 AoA on L2 processing: if the L1 translation of the L2 word was learned earlier, the L2 word was also read faster. We discuss the implications of these findings for two important hypotheses that offer an explanation for the AoA effect: the mapping and semantic hypothesis. We propose that the current results suggest an integration between these accounts.

Keywords

Age of acquisition

Bilingualism

Eye tracking

Visual word recognition

Corpus study

Introduction

Through our lifetime, we continuously encounter and learn new words. The age of acquisition (AoA) of words has been identified as an important factor in language processing. A well-established finding, at least in first-language (L1) processing, is that words with an earlier AoA are processed faster than words with a late AoA. This effect has a long history of replications in a multitude of experiments, including different paradigms and techniques.

L1 age of acquisition

In the very first study that revealed an influence of word-level AoA, Carroll and White (1973) found that pictures were named faster when their name was learned at an earlier age. This AoA effect in picture naming has been replicated with different sets of stimuli and in different languages (Belke, Brysbaert, Meyer, & Ghyselinck, 2005; Morrison, Ellis, & Quinlan, 1992; Pérez, 2007) and was also found in word naming studies (Brysbaert, Lange, & Van Wijnendaele, 2000; Gerhand & Barry, 1999b; Morrison & Ellis, 1995).

AoA also influences word recognition: in lexical decision, reaction times (RTs) are faster for earlier acquired words (e.g., Brysbaert, Lange, et al., 2000; Bonin, Chalard, Méot, & Fayol, 2001; Butler & Hains, 1979; Gerhand & Barry, 1999a; Wilson, Cuetos, Davies, & Burani, 2013). Interestingly, in several of these studies (Bonin et al., 2001; Gerhand & Barry, 1999a; Wilson et al., 2013) an interaction was found between AoA and word frequency, with larger the AoA effects for low frequency words.

In two investigations of the English Lexicon Project (ELP; Balota et al., 2007), which consists of lexical decision data for 40 481 English words, the role of word-level AoA was investigated in combination with a large set of other linguistic variables (for example word frequency, length, … ; Cortese & Khanna, 2007; Cortese & Schock, 2012). Both studies found an AoA effect, with shorter RTs for earlier learned words. The above interaction between word frequency and AoA also showed up in Cortese and Schock (2012).

Finally, a few studies investigated the AoA effect by means of eye tracking. In this paradigm, the eye movements of participants are recorded while they read pieces of natural text or sentences, without performing an artificial task like lexical decision. In two eye tracking studies, Juhasz and Rayner (2003, 2006) investigated AoA effects in sentence reading. In the 2003 study, AoA and other predictors were included as continuous variables, whereas in the 2006 study an orthogonal design was applied (early vs late AoA). In both studies, early and late timed measures were analyzed, and both yielded significant AoA effects (i.e., shorter fixations for early AoA words). In the 2006 study, an AoA effect was found for all eye tracking measures, whereas the 2003 study only found the AoA effect in early measures (single fixation duration and gaze duration). Juhasz and Rayner argue that the orthogonal design with extreme AoA values was more sensitive to detect AoA effects in late word processing. These L1 AoA effects in eye tracking were recently replicated in a corpus study by Dirix and Duyck (in press), in which eye movement data of monolinguals reading an entire novel was investigated. L1 AoA effects on 7158 nouns were found in all timed measures (single, first fixation and gaze duration and total reading time), as well as an interaction between AoA and word frequency in total reading times (cf. the lexical decision studies discussed above). Finally, Juhasz, Gullick, and Shesler (2011) investigated the AoA effect with ambiguous words that had an early and late learned meaning (e.g., *straw*, *volume*). The sentence context disambiguated the meaning of the target word, and target words received shorter fixations (both in early and late measures) when the early learned meaning of the ambiguous word was relevant.

In sum, the AoA effect seems to be quite robust in the literature on monolingual/L1 language processing. Faster processing of earlier learned words has been found in a large variety of paradigms and in different modalities (see Johnston and Barry (2006) or Juhasz (2005) for reviews). Recent monolingual/L1 mega studies of lexical decision (e.g., Cortese &

Schock, 2012) and eye movements (Dirix & Duyck, in press) validated the pioneer findings of smaller scale experiments.

Second language age of acquisition

Although the monolingual/L1 domain now approaches 45 years of AoA research, it has only been 15 years since word-level AoA has been investigated in the field of bilingualism, and studies are very rare. This is remarkable, because there is much more interindividual variability in the age at which words are learned for a second language (L2), so that the variable is possibly of greater relevance than for L1 processing. The majority of the words that we learn in L2 will also be known already in our L1, which creates an interesting situation: L2 words have an L2 AoA (the age at which the word was learned in L2), but also an L1 AoA (the age at which the L1 translation of the L2 word was learned). These L1 and L2 AoAs do not necessarily correspond: words that were learned early in L1 can be learned late in L2 and vice versa. Two main questions were addressed in the few L2 AoA studies that have been carried out. First, researchers investigated whether a word-level AoA effect may indeed be found in L2 processing. Second, it was investigated what mainly drives this AoA effect: the order at which the words were learned in the L1 or L2?

Izura and Ellis (2002) first addressed these questions. In their Experiment 1 (picture naming) and 2 (lexical decision), they found shorter RTs for earlier acquired words in L1 and L2, thus confirming the existence of a L2 AoA effect. To further assess whether it was the L1 or L2 AoA of the words that caused the AoA effect in L2, Izura and Ellis orthogonally manipulated the L1 and L2 AoA of their stimuli in Experiment 4 (lexical decision). Results showed only within-language AoA effects: in L1, RTs were faster for words learned early in L1, irrespective of when the words were learned in L2. Similarly, L2 reading was only influenced by order of acquisition in L2, not L1. The AoA seems to only have an impact within each language. Izura and Ellis (2004) later replicated these findings in both translation judgments and lexical decision. To date, these are the only two visual word recognition studies that investigated both the roles of L1 and L2 AoA in a full orthogonal design. For production, similar within-language AoA effects were also obtained in a bilingual picture naming task (Hirsh, Morrison, Gaset, & Carnicer ,2003).

In a spin-off of AoA research, the order of acquisition (OoA) effect of newly acquired stimuli is investigated. These “laboratory studies of AoA” allow researchers to study the impact of learning new stimuli at different points in time, while characteristics such as frequency can be controlled. Typically, a part of the stimuli set is introduced at the beginning of the study phase (“early acquired”); another part is presented at a later time (“late acquired”). This generally results in processing advantages for earlier learned items. For example, participants were faster to categorize “early” learned abstract checkerboard stimuli than a “later” learned set (Stewart & Ellis, 2008). In studies that involved linguistic material, similar results were obtained. Izura et al. (2011) found that early learned novel words for existing objects were processed faster in a series of behavioral tasks up to 35 days after the learning phase. Joseph, Wonnacott, Forbes, and Nation, (2014) found OoA effects on eye movements: total reading times decreased for novel words between the training and testing phase both for early and late learned items, but this effect was significantly larger for the early trained set.

These OoA studies support the robustness of acquisition effects, as OoA effects emerge even with a minimal delay between the presentation of the early and late stimuli set. Second, Izura et al. (2011) claim that these effects mirror real-life AoA effects, as the advantage for the early learned set can persist for weeks after training. Finally, studies involving linguistic materials could be interpreted as learning vocabulary of a novel language, mapping new lexical forms onto existing semantics, analogous to real life L2 learning.

To summarize, in the previous parts we have shown that L1 AoA is a well-established effect in psycholinguistic research. For L2 processing, some rare studies have confirmed L2 AoA effects, independent of L1 AoA, but the number of studies and stimuli is limited. Also, only isolated L2 word reading was investigated, and AoA eye tracking research for L2 sentence reading is completely lacking, until the present study. Our study will shed light on the specific time-course of AoA effects. Further, we will also argue that this approach may clarify the etiology of the (L1) AoA effect, about which two hypotheses exist.

The origin of the age of acquisition effect

The first hypothesis about the mechanism behind the AoA effect is the *semantic hypothesis*. According to this hypothesis, AoA effects originate from the organization of the semantic representational network of words (Brysbaert, Van Wijnendaele, & De Deyne, 2000; Steyvers & Tenenbaum, 2005). As we learn new words or concepts, they are linked to semantic representations we already know. Early learned words take up a more central place in the semantic network, so that they are more easily accessible than later learned words. In a study of semantic networks, Steyvers and Tenenbaum (2005) indeed found that most nodes in the network have few connections, but they are joined through a few nodes with many connections, so-called “hubs” (cf. the early learned words).

There are a few sources of empirical evidence for a semantic locus of the AoA effect. First, earlier learned words were categorized faster in semantic categorization tasks (Brysbaert, Van Wijnendaele et al, 2000; Menenti & Burani, 2007), In a more complex design, Ghyselinck, Custers, and Brysbaert (2004) presented names of living and non-living stimuli, which were either printed in upper- or lowercase. Participants were instructed to judge the letter case of targets words by responding verbally, using the labels “living” and “non-living”, so that responses were either congruent or incongruent with the semantic category of the words. Ghyselinck et al. found a larger congruency effect for early than late

AoA words. The authors concluded that the meaning of early AoA words is activated faster than that of late AoA words, and proposed that semantics indeed have an important role in the AoA effect. Second, the magnitude of the AoA effect seems to increase with a higher need of semantic activation: it is smallest in word naming tasks, larger in lexical decision tasks and largest in object naming (Barry, Johnston, & Wood, 2006).

The second hypothesis explaining AoA effects is the *mapping* or *connectionist hypothesis* (Ellis & Lambon Ralph, 2000; Lambon Ralph & Ehsan, 2006; Monaghan & Ellis, 2010). In a connectionist modeling study, Ellis and Lambon Ralph (2000) conducted a series of simulations involving OoA. They trained networks on learning patterns; one set of patterns was introduced immediately ("early"), one set after a number of training cycles ("late"). Importantly, in analogy with human language learning, the training on the "early" set was continued together with learning the "late" set. They found that in a variety of circumstances the early set was always learned better than the late set. Ellis and Lambon Ralph relate this to AoA effects by the principle of mapping of word forms on meaning representations. Information that enters the network first has the advantage of network plasticity: it can have a large influence on the connections between input (word form) and output (meaning) representations, making them more easily accessible. With new information entering the network, it becomes more settled or entrenched. This allows for a progressive smaller influence on the connection weights and thus a disadvantage for later learned items. The mapping hypothesis does not specify a particular linguistic level at which these effects take place.

A first line of evidence for the mapping hypothesis comes from the OoA investigations. In particular the study of Stewart and Ellis (2008) shows that even when learning random patterns, without semantics, an OoA effect emerges. Additional evidence comes from the L2 AoA literature. In their 2002 study, Izura and Ellis applied the following reasoning: if words in two languages share semantic representations (e.g., Kroll & Stewart, 1994; Van Hell & De Groot, 1998b), AoA effects of the (shared) semantic representations should transfer from L1 to L2. So, the semantic hypothesis predicts that the AoA effect in L2 should correspond to the age at which the L1 translations of the words were learned. However, the evidence on L2 AoA effects only shows within-language AoA effects, without an influence of L1 AoA on L2 processing (e.g., Hirsh et al., 2003; Izura & Ellis, 2002, 2004). Hence, Izura and Ellis situate the etiology of the AoA effect at the lexical level: when a second language is learned, new mappings (and connections) between in- and output have to be formed. These connections will be subject to the same mechanisms that apply to an L1 AoA effect: an advantage for the early items (because they profit from the network plasticity)

To summarize, whereas one etiological hypothesis about the AoA effect refers to the organization of the semantic network, the other hypothesis assumes a representational plasticity principle. At this point, evidence for both of these hypotheses creates a lack of consensus, although recent studies mostly support the latter. The approach to involve L2 AoA to determine the mechanism behind AoA effect therefore seems very interesting.

The present study

We investigated this matter further by conducting the first bilingual eye tracking sentence reading study of the (L2) AoA effect. The goal of the current study was twofold. First, we wanted to extend the L2 isolated word reading studies (e.g. Izura and Ellis, 2002) by applying eye tracking during natural reading, providing a better insight in the time course of the AoA effect in L2 reading. We were also interested to see whether L2 reading is indeed only influenced by L2 AoA, and not by L1 AoA, which is very informative for the etiology of the AoA effect.

We investigated L1 and L2 AoA effects in fixation time data of the Ghent Eye tracking COrpus (GECO; Cop, Dirix, Drieghe, & Duyck, in press), which contains eye movement data of bilinguals reading an entire novel in L1 and L2. Before analysis, lacking L2 AoA ratings for our stimuli needed to be collected. For L1, vast databases with AoA ratings are freely available (e.g., for Dutch, Brysbaert; Stevens, De Deyne, Voorspoels, & Storms, 2014; for English, Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). AoA ratings are usually collected by asking participants at which age they think they learned words. Although this might seem to be a rather poor indication of AoA, such ratings score well on validity (Brysbaert, in press; De Moor, Ghyselinck, & Brysbaert, 2000). In L2 AoA studies, ratings are typically only gathered for the (few tens of) stimuli presented in the experiment(s).

Experiment 1

In this experiment, we present the L2 AoA ratings for nearly 5 000 words, including the target stimuli from Experiment 2, using the method of Brysbaert et al. (2014). Participants had to indicate at which age they thought they learned a list of words. The ratings are available in the Supplementary Material. These ratings are freely available[1] and may be used in future L2 AoA studies that use similar late Dutch-English bilinguals.

Method

Participants and Materials. 126 undergraduates of Ghent University took part in this experiment (100 female, M_{age} = 18.94 [2.60]). They received course credit for their participation. All participants were unbalanced Dutch – English bilinguals, who received formal English education from age 13 on.

One part of the stimuli consisted of the 1 742 English nouns of the Ghent Eye tracking Corpus (GECO; Cop, Dirix, Drieghe, et al., in press), which was analyzed in Experiment 2. In addition, we selected 3 158 words from the English vocabulary test “wordORnot” of Ghent University (CRR, 2014).[2] To obtain a diverse sample of words, the 61 850 words from this test were divided into ten bins according to word frequency, and then subdivided again in ten

[1] An URL will be provided on acceptance of the paper

[2] The authors would like to thank Emmanuel Keuleers for providing the stimuli list.

bins according to word length. From each of the 100 resulting bins, 31 to 32 words were randomly selected. Combined with the GECO nouns, this resulted in 4 900 words. These were divided into six lists which were all matched on average word frequency and word length: two consisting of the GECO nouns and four of the remaining words.

Procedure. Each participant rated one of the word lists in an excel sheet. They were asked to indicate for each word at which age they believed they learned it, in analogy with L1 AoA rating studies (e.g., Brysbaert et al., 2014; Kuperman et al., 2012). The specific instructions can be found in the Appendix. Participants were also encouraged to complete the list in good conscience and not to fill in random numbers, as their data could not be used if it correlated poorly with the average ratings. All participants needed maximum one hour to complete their list.

Results and Discussion

The data of four participants correlated less than .60 with the average AoA ratings. It can be expected that AoA ratings for L2 words show a smaller inter-individual consistency than for L1 words, as people may start to learn the language at a different age, but with such low correlations we cannot exclude the possibility of random responses. Therefore, we excluded these ratings from the dataset (cf. Ghyselinck, De Moor, & Brysbaert, 2000). The final dataset thus included 20 – 21 ratings for each word list.

The average correlation between the ratings was .76 (*sd* = .05), which is indeed somewhat lower than reported in L1 AoA studies (around .90). However, as mentioned before, this does make sense as the L2 learning onset differs more between participants that for L1 learning. Furthermore, as formal English education only starts at age 13 in Flanders, vocabulary acquisition before that age depends largely on which words participants encounter in their daily life.

In Figure 1, the distribution of the L2 AoA ratings is presented. This resembled a normal distribution, as was the case in the large scale L1 AoA ratings (e.g. Brysbaert et al., 2014)

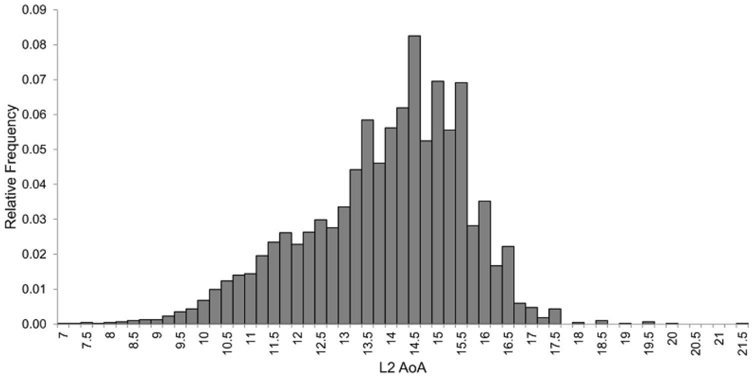

Figure 1. Frequency distribution of the L2 AoA ratings of Experiment 1.

We also visualized the relation between our L2 ratings and (a) their Dutch L1 translation AoA ratings (Figure 2; L1 AoA ratings from Brysbaert et al., 2014) and (b) their word frequency (Figure 3; SUBTLEX-UK frequencies from van Heuven, Mandera, Keuleers, & Brysbaert, 2013), as these were often reported as (highly) correlated in previous research.

L2 AoA was moderately correlated with L1 AoA (r = .52), which shows that word learning order roughly corresponds across languages, although some later learned L1 words may be earlier learned L2 words, and vice versa.

The correlation between L2 AoA and word frequency was somewhat higher (r = -.66). This further confirms the established relationship between these two lexical variables, and shows that also in L2 learning, high frequency words are learned earlier.

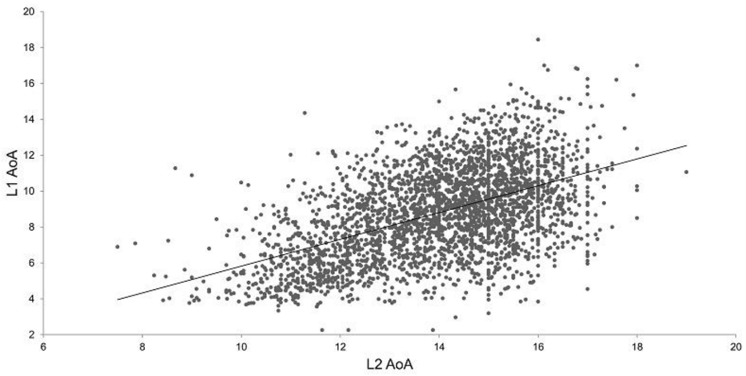

Figure 2. Scatter plot of the L2 AoA ratings of the words and the L1 AoA of their translation.

In conclusion, L2 ratings seem to show more inter-individual variability than L1 AoA ratings, but they show a lot of resemblance in terms of their characteristics: their distribution is similar and their relation with other lexical variables is in line with what could be expected. As such, they may be considered valid measures of the age at which our participants learned the L2 words for our L2 eye tracking analyses.

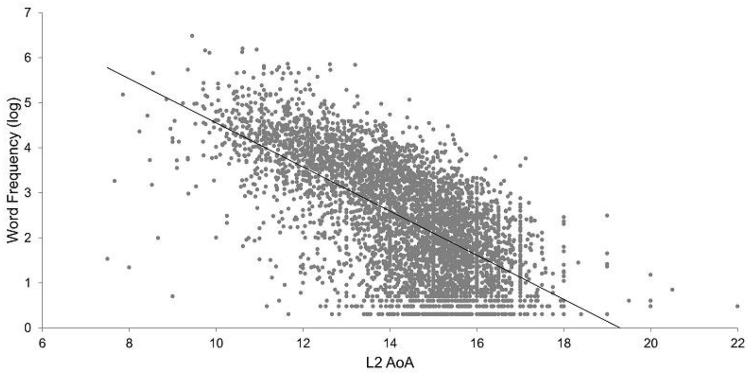

Figure 3. Scatter plot of the L2 AoA ratings of the words and their log word frequency.

Experiment 2

We investigated the L1 and L2 AoA effect in various reading measures of GECO (Cop, Dirix, Drieghe, et al., in press). This corpus consists of eye movement data of participants reading an entire book in their L1 (Dutch) and L2 (English). It has previously been successfully applied in investigations on sentence-level bilingual reading (Cop, Drieghe,

& Duyck, 2015), the word frequency effect (Cop, Keuleers, Drieghe, & Duyck, 2015), the cognate effect (Cop, Dirix, Van Assche, Drieghe, & Duyck, in press) and cross-lingual orthographic neighborhood effects (Dirix, Cop, Drieghe, & Duyck, in press).

This study is the first investigation of the L2 AoA effect during reading of meaningful, longer passages of natural text, using eye tracking. While the studies of for example Izura and Ellis (2002, 2004) were of critical importance to the field and introduced L2 AoA as a concept, this eye tracking study offers additional value and insight to AoA research. First, eye tracking allows “natural reading”, as the participants only have the instruction to read the presented text, without further influence of task demands, or decision or response processes. Second, a detailed analyses of the time course of (AoA) effects can be made, as different timed measures represent different underlying processes (Rayner, 1998, 2009). So-called early measures (i.e., single fixation duration, first fixation duration and gaze duration) reflect lexical access, or selecting the correct representation of a word in the memory and accessing it. Late measures, such as total reading time, reflect higher order processes such as the semantic activation of the word, word verification and sentence comprehension. In analogy with the monolingual AoA studies by Juhasz and Rayner (2003, 2006) and Dirix and Duyck (in press), we investigated first fixation durations (the duration of the first fixation on a word), single fixation durations (the duration of the first fixation on words that were only fixated once), gaze durations (the duration of all fixations on a word before the eyes move to the right of the word) and total reading times (the summed duration of all fixations on a word, including refixations after regressions). By analyzing these measures, we can investigate whether AoA effects in L2 reading on both early and late processes of word recognition are similar to those in L1 reading, or whether the time course of L1 and L2 AoA differs across languages.

Furthermore, this is the first mega study of L2 AoA. By including a large amount of stimuli and data points, the variability in included word characteristics allows to assess their independent and simultaneous effects as continuous predictors in the analyses. As most previous studies orthogonally manipulated AoA, this is also beneficial to our insights in the AoA effect and the reliability of earlier results. Indeed, Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004) argue that there are several disadvantages of applying a factorial design to study lexical processing. For example, there can be a decrease in statistical power and reliability when categorizing a continuous variable, or there are potentially contaminating factors that confound the factors of the design, which makes it difficult to estimate the influence of a categorical variable in a small set of stimuli.

Our results are also crucial for the discussion on the etiology of the AoA effect, as the semantic and mapping hypotheses make different predictions for the L2 AoA effect (cf. Izura & Ellis, 2002). The semantic hypothesis predicts that the L1 AoA of the translation of L2 words should influence word processing in L2, given that L2 words are mapped onto existing L1/semantics (Kroll & Stewart, 1994). In this case, L1 AoA should have an effect both on L1 and L2 reading; L2 AoA should have no effect at all. This predicts an effect of L1 AoA on all fixation durations in L1 reading, similar to previous monolingual research; L2 reading should also be influenced by L1 AoA. In contrast, the mapping hypothesis predicts that L2 reading should only be influenced by the age at which the words were learned in L2. In this case, the AoA effects should only operate within languages: L1 AoA should only influence L1 processing; and L2 AoA should influence L2 processing. As the mapping hypothesis does not specify a single level at which AoA effects can occur, we would expect within-language AoA effects on both early and late measures.

We also considered the possibility that L1 or L2 AoA could interact with other word characteristics, such as word frequency (Dirix & Duyck, in press; Gerhand & Barry, 1999a;

Wilson et al., 2013). We covered this by including interactions between the predictors in our primary statistical models. Next to word frequency, we also included word length, cross-lingual orthographic overlap and rank of occurrence as word characteristics. Orthographic overlap was included because Van Assche, Drieghe, Duyck, Welvaert and Hartsuiker (2011) showed that reading times are shorter as cross-lingual orthographic overlap between translation equivalents increases. This was operationalized by calculating the Corrected Levenshtein Distance[3] between each noun and its translation (Schepens, Dijkstra, & Grootjen, 2012). Rank of occurrence was included to control for repetition effects as some of the words are repeated multiple times throughout the novel. Finally, L1 and L2 proficiency were included as participant characteristics.

Method

Participants and Materials. The following criteria were used to select which nouns of the bilingual part of GECO (Cop, Dirix, Drieghe, et al., in press) were included in the current study: (a) only nouns for which an AoA rating of the word and its translation were available (in Brysbaert et al. (2014) for Dutch; in the new collected ratings for English, see Supplementary Material); (b) as we wanted reliable AoA estimates, we only selected words that at least 75% of the raters knew (similar to Izura & Ellis, 2002); (c) identical cognates or interlingual homographs were excluded. This resulted in 1069 unique Dutch nouns and 966 unique English nouns. See Table 1 for the characteristics of these nouns. The participants of GECO were 19 unbalanced Dutch-English bilinguals (17 females, $M_{age} = 21.2$, $SD_{age} = 2.2$). They all read the entire novel “The mysterious affair at Styles” by Agatha Christie (1920; in Dutch: “De zaak Styles). The L1 (Dutch) and L2 (English) proficiency of the participants was rated with the Dutch and English version of the LexTALE (Lemhöfer & Broersma, 2012). For

[3] The formula used for calculating the Corrected Levenshtein Distance: $Orthographic\ Overlap = 1 - \frac{Distance}{Length}$

Dutch, their average LexTALE score was 92.43 (SD = 6.34, range = [73.75 – 100]). For English, their average LexTALE score was 75.63 (SD = 12.87, range = [51.25 – 98.75]).

Table 1

Descriptive Statistics for the GECO nouns analyzed in the current study, averaged over stimuli per language (standard deviations between parentheses).

	Word Frequency[a]	Word Length	L1 AoA[b]	L2 AoA[c]	Rank of Occurrence[d]	CLD[e]
Dutch	1.51 (1.01)	6.49 (2.35)	6.83 (1.93)	12.35 (1.46)	18.53 (33.59)	0.33 (0.26)
English	1.63 (0.95)	5.91 (2.12)	7.04 (2.11)	12.33 (1.55)	12.32 (16.88)	0.37 (0.30)

[a]Log10 Subtlex frequencies per million words: SUBTLEX-NL for Dutch words (Keuleers, Brysbaert, & New, 2010), SUBTLEX-UK for English words (van Heuven et al., 2013); [b]For “Dutch”, this means the L1 AoA of the Dutch words, for “English” the L1 AoA of the Dutch translation of the English words (Brysbaert et al., 2014); [c]For “Dutch”, this means the L2 AoA of the English translation of the Dutch words, for English the L2 AoA of the English words (from the Experiment 1 ratings, see Supplementary Material).[d]The average amount of repetitions of each word throughout the novel. [e]The average amount of orthographic overlap, expressed by the Corrected Levenshtein Distance.

Procedure. All participants read the entire novel while their eye movements were recorded, spread over 4 separate sessions. Half of the novel was read in Dutch (L1), the other half in English (L2); the order was counterbalanced. Multiple-choice questions were presented after each chapter to ensure participants were paying adequate attention and reading for comprehension. We refer to Cop, Dirix, Drieghe, et al. (in press) and Cop, Keuleers, et al. (2015) for further details on the procedure.

Analyses of Eye Movements. The timed measures were fitted in a linear mixed model using the lme4 (version 1.1-11) and the lmerTest (version 2.0-30) packages in R (version 3.2.4, R Core Team, 2016). The initial models contained the fixed factors L1 AoA (continuous), L2 AoA (continuous), Language (Dutch or English), Word Frequency (continuous) per million, Word Length (continuous), L1 Proficiency (continuous), L2

Proficiency (continuous), the Rank of Occurrence (continuous) and Orthographic Overlap (continuous). Word frequency was log transformed with base 10 to normalize its distribution. All continuous variables were centered to reduce collinearity between main effects and interactions. Next to the fixed factors, a random intercept per subject and per word was included in all initial models (Baayen, Davidson, & Bates, 2008). This was done to ensure that differences between subjects concerning genetic, developmental or social factors were modeled on one hand, and because our stimuli set does not contain all possible nouns in a language on the other.

A separate analysis was carried out for each dependent variable (i.e., the timed measures). First, the timed measure was log transformed with base 10 to normalize its distribution (see Lo & Andrews, 2015 for an alternative approach). Second, a full model including the two random clusters and all interactions (up to 3-way) was fitted. By backward fitting of the fixed effects, forward fitting of the random effects and again backward fitting of the fixed effects, the optimal model was discovered (Barr, Levy, Scheepers, & Tily, 2013). An omnibus analysis was conducted for each dependent variable, so that we did not split up the analysis for each language in the first stage of the analysis. As a consequence, we considered effects that did not interact with the factor language as equally large for L1 and L2 reading. If there however was a significant interaction with language, we also conducted language-separate analyses to interpret this interaction. Finally, the Variance Inflation Factor (VIF) was calculated for each model to estimate the influence of multicollinearity on the regression coefficients. A VIF larger than 5 indicates moderate influence, larger than 10 is considered to be problematic (Fox & Weisberg, 2010).

Results

Single Fixation Duration. Only nouns that received exactly one fixation were selected for this analysis (54.60% of the data). Furthermore, single fixation durations that differed more than 2.5 standard deviations from the subject means were considered as outliers and excluded (2.35%). The outcome of the final model for single fixation durations is presented in Table 2. The maximum VIF for this model was 3.789.

Effects of L1 AoA. Across languages, there was no significant main effect of L1 AoA. The three-way interaction between L1 AoA, language and word length was significant (β = 0.0008, se = 0.0004, t = 2.133, p < .05).

Effects of L2 AoA. The main effect of L2 AoA was not significant. There was however a significant interaction between L2 AoA and Language (β = 0.0043, se = 0.0015, t = 2.863, p < .01; see Figure 4) and between L2 AoA and word frequency (β = -0.0018, se = 0.0007, t = -2.577, p < .05). There was a facilitating effect of L2 AoA on nouns with a log word frequency per million of 1.482 or less (χ^2 = 3.85, df = 1, p < 0.05). Finally, there was a significant interaction between L2 AoA and L1 proficiency (β = -0.0002, se < 0.0001, t = -2.431, p < .01) and a marginally significant one between L2 AoA and orthographic overlap (β = 0.0035, se < 0.0019, t = 1.822, p < .1). Post hoc contrasts showed that the effect of L2 AoA was significant when the L1 proficiency score of the participants was lower than 89.20 (χ^2 = 3.86, df = 1, p < 0.05; see Figure 11). No significant effects were found in the contrasts for the L2 AoA and orthographic overlap interaction.

L1 Reading. *Effects of L1 AoA.* Separate analyses for each language showed that the interaction between L1 AoA and word length was not significant for L1 reading (β = -0.0004, se = 0.0003, t = -1.475, p > .1).

Effects of L2 AoA. There was no main effect of L2 AoA on L1 reading (β = 0.0014, se = 0.0008, t = 1.433, p > .1).

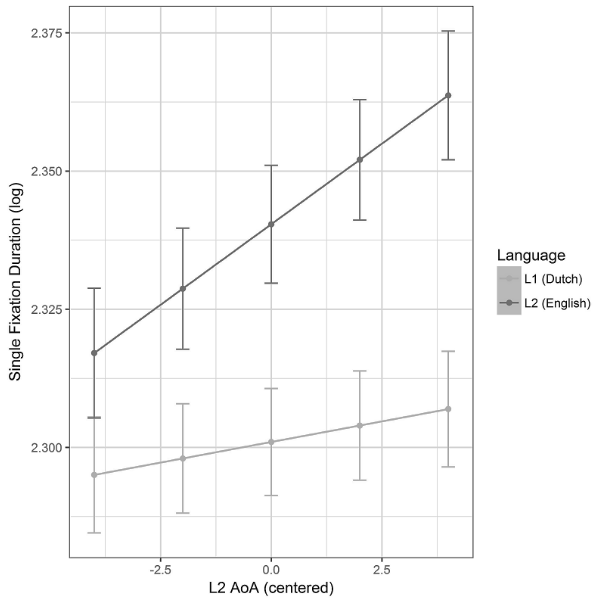

Figure 4. The interaction between Language (lines) and L2 AoA (x-axis) for single fixation durations (y-axis). Error bars represent standard errors.

L2 Reading. *Effects of L2 AoA.* Separate analyses for each language showed that there was facilitation for earlier L2 AoA in L2 reading ($\beta = 0.0057$, se = 0.0013, $t = 4.454$, $p < .001$; see Figure 4).

Effects of L1 AoA. The L2 analysis showed that the interaction between L1 AoA and word length was significant for L2 reading ($\beta = 0.0006$, se = 0.0003, $t = 2.028$, $p < .05$). Post-hoc contrasts showed that there was a marginally facilitating effect of an earlier L1 AoA, but only for words that contained 10 or more letters ($\chi^2 = 2.76$, df = 1, $p < 0.1$).

Table 2

Estimates, standard errors, t-values and p-values for the fixed and random effects of the final general linear mixed effect model for single fixation duration for bilingual reading.

	Estimate	SE	t-value	p-value	
Fixed Effects					
Intercept	2.3018	0.0097	237.792	<.001	***

L1 Age of Acquisition	0.0011	0.0008	1.409	.159	
L2 Age of Acquisition	0.0015	0.0010	1.483	.138	
Language	0.0394	0.0041	9.714	<.001	***
Word Frequency	-0.0042	0.0021	-1.998	.050	.
Word Length	0.0027	0.0009	2.959	.005	**
L1 Proficiency	-0.0019	0.0015	-1.229	.235	
L2 Proficiency	0.0005	0.0008	0.642	.529	
Rank of Occurrence	-0.0001	<0.0001	-1.470	.142	
Orthographic Overlap	-0.0021	0.0029	-0.736	.462	
L1 AoA * Language	-0.0013	0.0010	-1.281	.200	
L1 AoA * Word Length	-0.0003	0.0003	-1.083	.279	
L2 AoA * Language	0.0043	0.0015	2.863	.004	**
L2 AoA * Word Frequency	-0.0018	0.0007	-2.577	.010	*
L2 AoA * L1 Proficiency	-0.0002	0.0001	-2.431	.015	*
L2 AoA * Orthographic Overlap	0.0035	0.0019	1.822	.069	.
Language * Word Frequency	-0.0050	0.0026	-1.928	.054	.
Language * Word Length	0.0011	0.0009	1.140	.254	
Word Frequency * Word Length	-0.0020	0.0006	-3.231	.001	**
L1 AoA * Language * Word Length	0.0008	0.0004	2.133	.033	*
Language * Word Frequency * Word Length	0.0023	0.0009	2.580	.010	**

	Variance	SD
Random Effects		
Word		
(Intercept)	0.0003	0.0167
Subject		
(Intercept)	0.0017	0.0417
Language	0.0002	0.0155
Word Frequency	<0.0001	0.0056
Word Length	<0.0001	0.0027
Word Frequency * Word Length	<0.0001	0.0008

p<0.1 . p<0.05 * p<0.01 ** p<0.001***

First Fixation Duration. First fixation durations that differed more than 2.5 standard deviations from the subject means were considered as outliers and excluded from the dataset (2.29%). In comparison to single fixation duration, this measure included all first fixations on the target nouns, irrespective of later refixations. The maximum VIF was 4.174. The final model is presented in Table 3.

Effects of L1 AoA. There was a general significant main effect of L1 AoA ($\beta = 0.0017$, se = 0.0007, $t = 2.397$, $p < .05$): first fixation durations were shorter for nouns with an earlier L1 AoA. The two-way interaction between L1 AoA and language was marginally significant ($\beta = -0.0016$, se = 0.0009, $t = -1.713$, $p < .1$). Furthermore, the three-way interaction between L1 AoA, language and word length was significant ($\beta = 0.0007$, se = 0.0003, $t = 1.995$, $p < .05$; see Figure 5).

Effects of L2 AoA. The main effect of L2 AoA across languages was not significant, but the interaction with language was marginally significant ($\beta = 0.0025$, se = 0.0013, $t = 1.889$, $p < .1$; see Figure 6). The interaction between L2 AoA and word frequency was significant ($\beta = -0.0013$, se = 0.0006, $t = -2.081$, $p < .05$): a facilitatory effect of L2 AoA was present for nouns with a word frequency of 1.280 or less (χ^2=3.86, df=1, $p < 0.05$). Finally, the interaction between L2 AoA and L1 proficiency was significant ($\beta = -0.0001$, se = 0.0001, $t = -2.104$, $p < .05$). Post hoc contrasts showed that the effect of L2 AoA was significant when the L1 proficiency score of the participants was lower than 87.25 ($\chi^2 = 3.85$, df = 1, $p < .05$; see Figure 11).

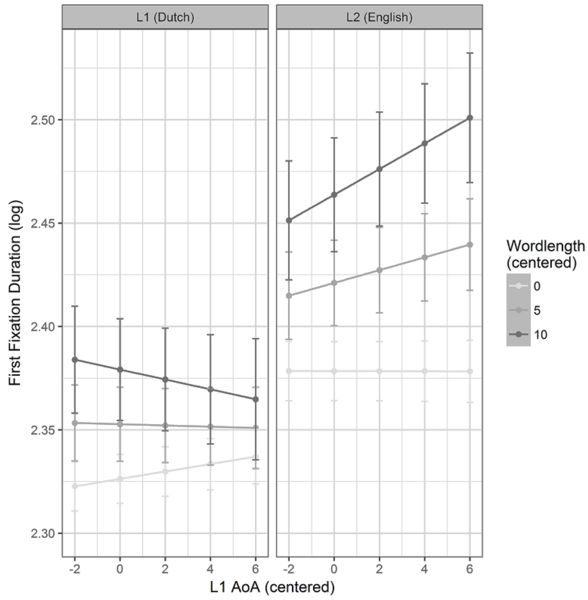

Figure 5. The interaction between L1 AoA (x-axis) and word length (lines) in each language (panels) for first fixation durations (y-axis). Error bars represent the standard error.

L1 Reading. *Effects of L1 AoA.* In the L1 analysis, there was a facilitatory effect of L1 AoA on L1 nouns (β = 0.0016, se = 0.0007, t = 2.489, p < .05). The interaction between L1 AoA and word length was not significant for L1 reading (β = -0.0003, se = 0.0002, t = -1.310, p > .1).

Effects of L2 AoA. We found no main effect of L2 AoA on L1 nouns (β = 0.0013, se = 0.0009, t = 1.486, p > .1).

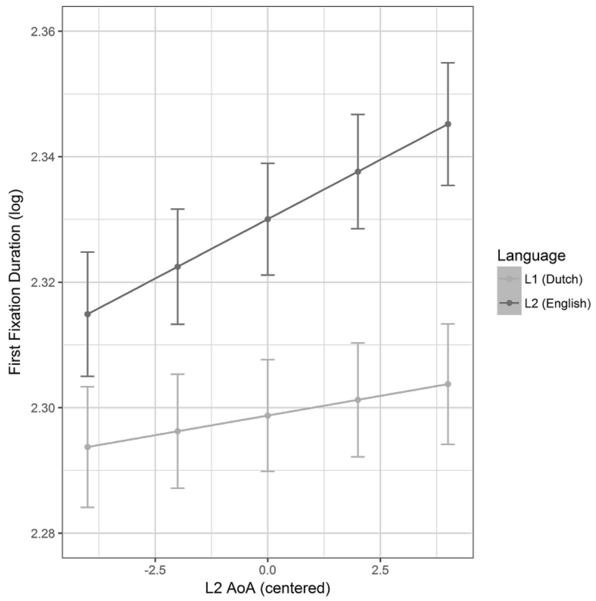

Figure 6. The interaction between Language (lines) and L2 AoA (x-axis) for first fixation durations (y-axis). Error bars represent standard errors.

L2 Reading. *Effects of L2 AoA.* The L2 analysis showed that there was a facilitating effect of L2 AoA for L2 nouns (β = 0.0036, se = 0.0011, t = 3.194, p < .01; see Figure 6).

Effects of L1 AoA. There was no main effect of L1 AoA on L2 nouns (β = <-0.0001, se = 0.0006, t = -0.009, p > .05). However, the interaction between L1 AoA and word length was significant in L2 (β = 0.0005, se = 0.0010, t = 1.964, p < .05; see Figure 5): there was a facilitatory effect for nouns of 12 or more letters when the L1 AoA was earlier (χ^2 = 3.88, df = 1, p < 0.05).

Table 3

Estimates, standard errors, t-values and p-values for the fixed and random effects of the final general linear mixed effect model for first fixation duration for bilingual reading.

	Estimate	SE	t-value	p-value	
Fixed Effects					
Intercept	2.2990	0.0089	258.241	<.001	***
L1 Age of Acquisition	0.0017	0.0007	2.397	.017	*

L2 Age of Acquisition	0.0013	0.0009	1.383	.167	
Language	0.0313	0.0017	18.767	<.001	***
Word Frequency	-0.0036	0.0017	-2.149	.034	*
Word Length	0.0014	0.0007	1.954	.055	.
L1 Proficiency	-0.0023	0.0020	-1.196	.249	
L2 Proficiency	0.0009	0.0010	0.926	.368	
Rank of Occurrence	<0.0001	<0.0001	-0.501	.616	
Orthographic Overlap	-0.0032	0.0026	-1.234	.217	
L1 AoA * Language	-0.0016	0.0009	-1.713	.087	.
L1 AoA * Word Length	-0.0002	0.0002	-1.016	.310	
L2 AoA * Language	0.0025	0.0013	1.889	.059	.
L2 AoA * Word Frequency	-0.0013	0.0006	-2.081	.038	*
L2 AoA * L1 Proficiency	-0.0001	0.0001	-2.104	.036	*
Language * Word Frequency	-0.0033	0.0023	-1.466	.143	
Language * Word Length	-0.0001	0.0008	-0.115	.909	
Word Frequency * Word Length	-0.0006	0.0005	-1.202	.229	
L1 AoA * Language * Word Length	0.0007	0.0003	1.995	.046	*
Language * Word Frequency * Word Length	0.0013	0.0007	1.712	.087	.

	Variance	SD
Random Effects		
Word		
(Intercept)	0.0003	0.0166
Subject		
(Intercept)	0.0015	0.0384
Word Frequency	<0.0001	0.0034
Word Length	<0.0001	0.0021

p<0.1 . p<0.05 * p<0.01 ** p<0.001***

Gaze Duration. Gaze durations that differed more than 2.5 standard deviations from the subject means were considered as outliers and excluded from the dataset (2.67%). The final model had a maximum VIF of 3.894; it is presented in Table 4.

Effects of L1 AoA. Across languages, there was a marginally significant main effect of L1 AoA ($\beta = 0.0018$, se = 0.0009, $t = 1.939$, $p < .1$): gaze durations were shorter for an earlier L1 AoA. The three-way interaction between L1 AoA, language and word length was again significant ($\beta = 0.0010$, se = 0.0004, $t = 2.452$, $p < .05$; see Figure 7).

Effects of L2 AoA. There was no general significant main effect of L2 AoA. There was a significant interaction between L2 AoA and Language (β = 0.0041, se = 0.0017, t = 2.335, p < .01; see Figure 8). L2 AoA also interacted significantly with orthographic overlap (β = 0.0047, se = 0.0022, t = 2.135, p < .05). Post-hoc contrasts showed that the L2 AoA effect was larger for words with a CLD of 0.51 or higher (χ^2 = 3.87, df = 1, p < 0.05). The interaction between L2 AoA and word frequency was marginally significant (β = -0.0014, se = 0.0008, t = -1.776, p < .1). Contrast revealed that there was a facilitatory effect of an earlier L2, which was only significant for nouns with a log word frequency up to 1.265 (χ^2 = 3.85, df = 1, p < 0.05).

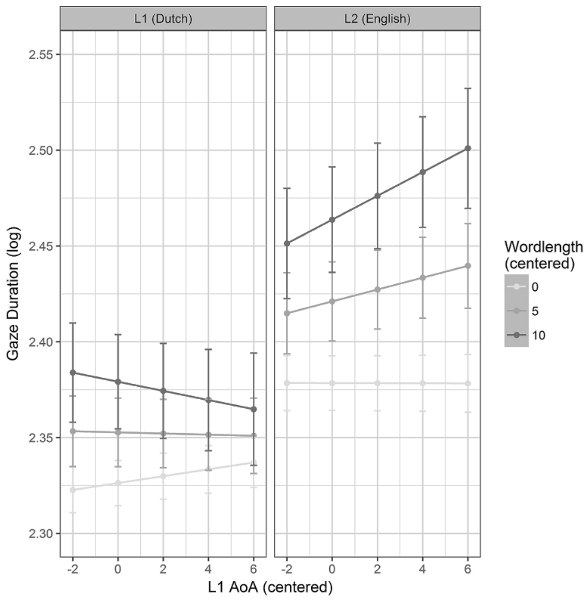

Figure 7. The interaction between L1 AoA (x-axis) and word length (lines) in each language (panels) for gaze durations (y-axis). Error bars represent standard errors.

L1 Reading. *Effects of L1 AoA.* Separate analyses for each language showed that there was a facilitatory effect of L1 AoA on L1 reading (β = 0.0017, se = 0.0008, t = 2.005, p < .05). The interaction between L1 AoA and word length was marginally significant in L1 (β = -0.0005, se = 0.0003, t = -1.694, p < .1). Post-hoc contrasts revealed no significant effects.

Effects of L2 AoA. There was no main effect of L2 AoA on L1 reading (β = 0.0017, se = 0.0011, t = 1.502, p > .1).

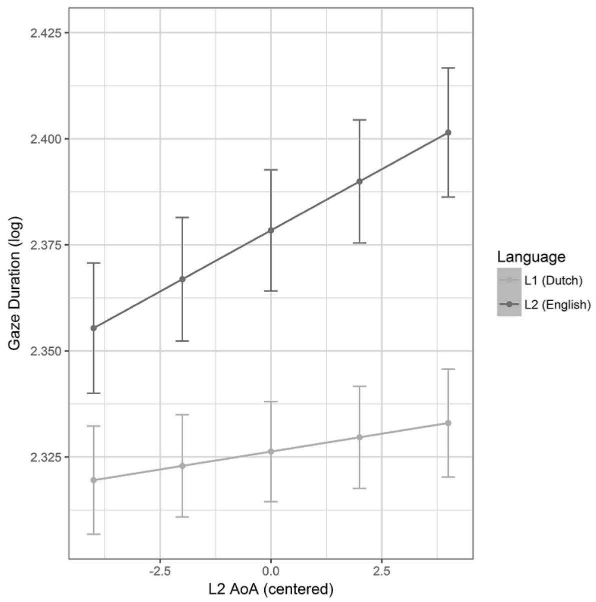

Figure 8. The interaction between Language (lines) and L2 AoA (x-axis) for gaze durations (y-axis). Error bars represent standard errors.

L2 Reading. *Effects of L2 AoA.* The L2 analysis showed that the facilitatory effect of an earlier L2 AoA was significant for L2 reading (β = 0.0058, se = 0.0015, t = 3.781, p < .001; see Figure 8).

Effects of L1 AoA. The interaction between L1 AoA and word length was significant for L2 reading (β = 0.0007, se = 0.0003, t = 2.086, p < .05; see Figure 7). Post-hoc contrasts in L2 revealed that the facilitatory effect of an earlier L1 AoA was only significant for nouns with 14 characters or more (χ^2 = 3.87, df = 1, p < 0.05).

Table 4

Estimates, standard errors, t-values and p-values for the fixed and random effects of the final general linear mixed effect model for gaze duration for bilingual reading.

	Estimate	SE	t-value	p-value	
Fixed Effects					
Intercept	2.3269	0.0118	197.070	<.001	***
L1 Age of Acquisition	0.0018	0.0009	1.939	.053	.
L2 Age of Acquisition	0.0017	0.0012	1.416	.157	
Language	0.0522	0.0060	8.716	<.001	***
Word Frequency	-0.0065	0.0024	-2.744	.008	**
Word Length	0.0053	0.0015	3.623	<.001	**
L1 Proficiency	-0.0005	0.0012	-0.442	.663	
L2 Proficiency	-0.0003	0.0006	-0.532	.601	
Rank of Occurrence	0.0000	0.0000	-1.179	.238	
Orthographic Overlap	-0.0033	0.0033	-0.998	.319	
L1 AoA * Language	-0.0018	0.0012	-1.500	.134	
L1 AoA * Word Length	-0.0004	0.0003	-1.338	.181	
L2 AoA * Language	0.0041	0.0017	2.335	.020	*
L2 AoA * Word Frequency	-0.0014	0.0008	-1.776	.076	.
L2 AoA * Orthographic overlap	0.0047	0.0022	2.135	.033	*
Language * Word Frequency	-0.0046	0.0030	-1.544	.123	
Language * Word Length	0.0032	0.0011	3.083	.002	**
Word Frequency * Word Length	-0.0025	0.0007	-3.609	<.001	***
L1 AoA * Language * Word Length	0.0010	0.0004	2.452	.014	*
Language * Word Frequency * Word Length	0.0030	0.0010	3.090	.002	**

	Variance	SD
Random Effects		
Word		
(Intercept)	0.0006	0.0253
Subject		
(Intercept)	0.0026	0.0508
Language	0.0006	0.0240
Word Frequency	<0.0001	0.0062
Word Length	<0.0001	0.0055
Word Frequency * Word Length	<0.0001	0.0010

p<0.1 . p<0.05 * p<0.01 ** p<0.001***

Total Reading Times. Total reading times that differed more than 2.5 standard deviations from the subject means were considered as outliers and removed from the dataset (2.89%). The final model for total reading times is presented in Table 5. The maximum VIF for this model was 3.894.

Effects of L1 AoA. The main effect of L1 AoA was significant across languages (β = 0.0029, se = 0.0011, $t = 2.608$, $p < .01$). Total reading times were shorter for nouns with an earlier L1 AoA. The interaction between L1 AoA and language was significant (β = -0.0034, se = 0.0015, $t = -2.329$, $p < .05$; see Figure 9). The three-way interaction between L1 AoA, language and word length was marginally significant (β = 0.0009, se = 0.0005, $t = 1.770$, $p < .1$).

Effects of L2 AoA. The main effect of L2 AoA did not reach significance, but there was a significant interaction between L2 AoA and language (β = 0.0087, se = 0.0021, $t = 4.158$, $p < .001$; see Figure 10). The interaction between L2 AoA and L1 proficiency was again significant (β = -0.0003, se = 0.0001, $t = -2.874$, $p < .01$). Post hoc contrasts showed that the effect of L2 AoA was significant when the L1 proficiency score of the participants was lower than 85.60 ($\chi^2 = 3.85$, df = 1, $p < 0.05$; see Figure 11).

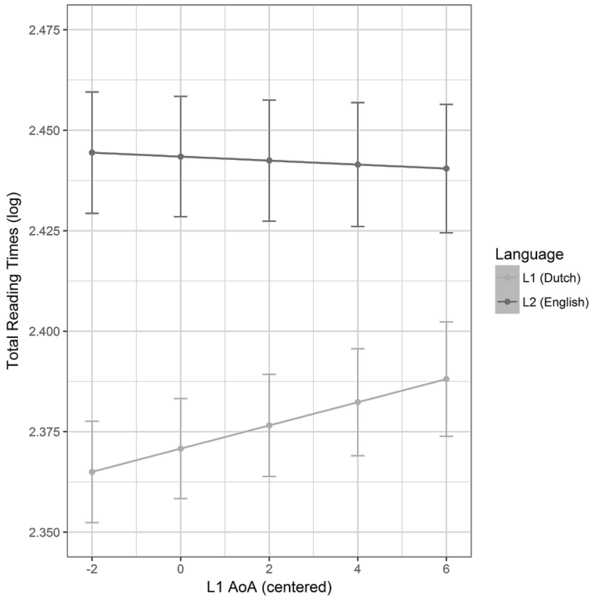

Figure 9. The interaction between Language (lines) and L1 AoA (x-axis) for total reading times (y-axis). Error bars represent standard errors.

L1 Reading. *Effects of L1 AoA.* The separate L1 analysis showed that the facilitatory effect of L1 AoA was significant for L1 reading (β = 0.0027, se = 0.0010, t = 2.608, p < .01; see Figure 9). Furthermore, the interaction between L1 AoA and word length was not significant for L1 nouns (β = -0.0005, se = 0.0004, t = -1.320, p > .1).

Effects of L2 AoA. The effect of L2 AoA was not significant for L1 reading (β = 0.0015, se = 0.0013, t = 1.150, p > .1).

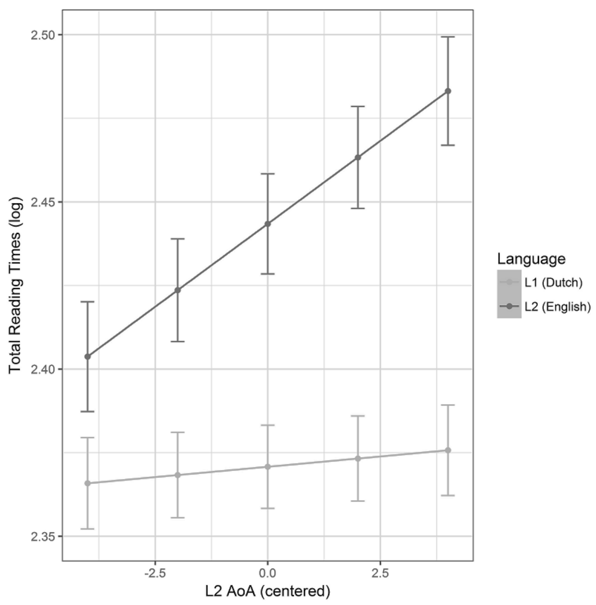

Figure 10. The interaction between Language (lines) and L2 AoA (x-axis) for total reading times (y-axis). Error bars represent standard errors.

L2 Reading. *Effects of L2 AoA.* In the separate L2 analysis it was revealed that the facilitatory effect of an earlier L2 AoA was significant for L2 reading (β = 0.0095, se = 0.0017, t = 5.494, p < .001; see Figure 10).

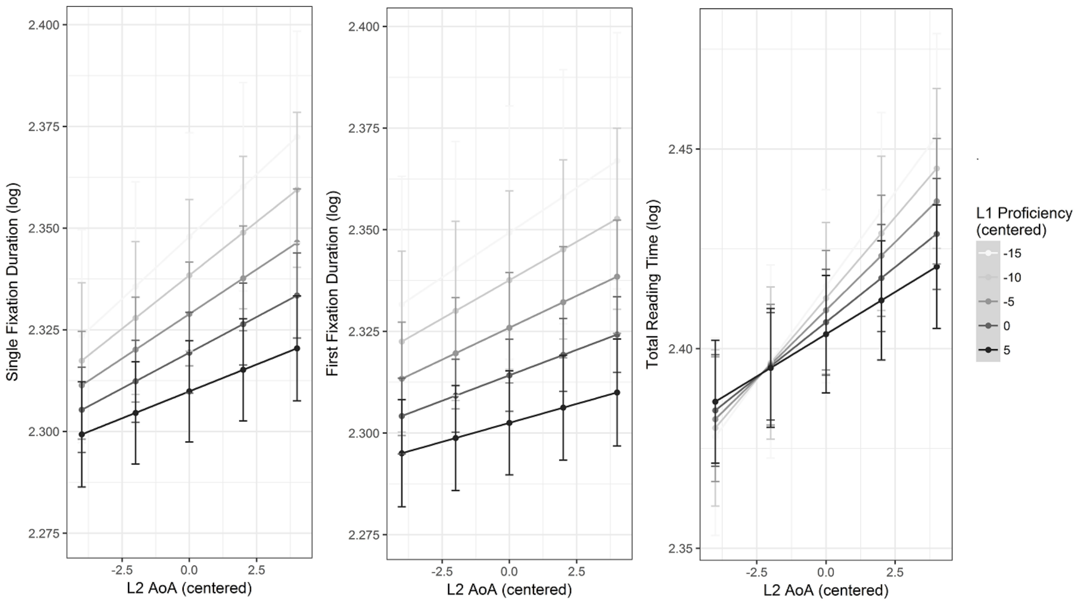

Figure 11. The interaction between L2 AoA (x-axis) and L1 proficiency (lines) for single fixation duration (left panel) , first fixation duration (middle panel) and total reading time (right panel). Error bars represent standard errors.

Effects of L1 AoA. The facilitatory effect of L1 AoA was not significant in L2 (β = -0.0006, se = 0.0010, t = -0.587, p > .1). Furthermore, the interaction between L1 AoA and word length was also not significant for L2 reading (β = 0.0005, se = 0.0004, t = 1.470, p > .1).

Table 5

Estimates, standard errors, t-values and p-values for the fixed and random effects of the final general linear mixed effect model for total reading time for bilingual reading.

	Estimate	SE	t-value	p-value	
Fixed Effects					
Intercept	2.3726	0.0125	190.535	<.001	***
L1 Age of Acquisition	0.0029	0.0011	2.608	.009	**
L2 Age of Acquisition	0.0012	0.0014	0.905	.366	
Language	0.0727	0.0073	9.959	<.001	***
Word Frequency	-0.0066	0.0031	-2.122	.038	*
Word Length	0.0070	0.0015	4.524	<.001	***

L1 Proficiency	-0.0006	0.0013	-0.451	.658	
L2 Proficiency	-0.0002	0.0006	-0.337	.741	
Rank of Occurrence	-0.0001	0.0000	-2.695	.007	**
Orthographic Overlap	-0.0009	0.0040	-0.226	.821	
L1 AoA * Language	-0.0034	0.0015	-2.329	.020	*
L1 AoA * Word Length	-0.0004	0.0004	-1.045	.296	
L2 AoA * Language	0.0087	0.0021	4.158	<.001	***
L2 AoA * L1 Proficiency	-0.0003	0.0001	-2.874	.004	**
Language * Word Frequency	-0.0083	0.0035	-2.336	.020	*
Language * Word Length	0.0043	0.0013	3.430	.001	***
Word Frequency * Word Length	-0.0050	0.0007	-6.688	<.001	***
L1 AoA * Language * Word Length	0.0009	0.0005	1.770	.077	.
Language * Word Frequency * Word Length	0.0042	0.0011	3.716	<.001	***

	Variance	SD
Random Effects		
Word		
(Intercept)	0.0009	0.0305
Subject		
(Intercept)	0.0029	0.0534
Language	0.0008	0.0292
Word Frequency	0.0001	0.0092
Word Length	<0.0001	0.0055

p<0.1 . p<0.05 * p<0.01 ** p<0.001***

Discussion

L1 Reading. In all timed measures except for single fixation durations, there was a significant main effect of L1 AoA on L1 reading: words with an earlier L1 AoA received shorter fixations than words with a later L1 AoA. These effects are largely consistent with previous monolingual AoA research in general (e.g., Brysbaert, Lange, et al., 2000; Carroll & White, 1973; Cortese & Khanna, 2007; Gerhand & Barry, 1999b; Morrison et al., 1992) and the few AoA eye tracking studies in particular (Dirix & Duyck, in press; Juhasz & Rayner, 2003, 2006). Even in the natural reading of long texts, AoA consistently and reliably influences word recognition throughout early and late stages of processing.

Furthermore, across reading languages, we encountered an interaction between L2 AoA and word frequency on two timed measures. In monolingual AoA investigations, an interaction between AoA and word frequency has been reported on a few occasions (e.g., Bonin et al., 2001; Dirix & Duyck, in press; Gerhand & Barry, 1999a; Wilson et al., 2013), with a stronger facilitatory AoA effect for low frequent words. The L2 AoA by word frequency interaction followed the same pattern.

To summarize, the AoA effects in L1 reading are partially consistent with previous research (i.e., the facilitating L1 AoA effect on L1 reading), but we also discovered some minor L2 AoA influences on L1 reading.

L2 Reading. For all of the timed measures that we analyzed, there was a significant interaction between L2 AoA and language: a facilitatory effect of L2 AoA was only present in L2 reading. These results are in line with previous L2 AoA studies (Hirsh et al., 2003; Izura & Ellis, 2002, 2004), which also found a facilitatory L2 AoA effect on L2 isolated word processing.

Furthermore, for single and first fixation duration, the interaction between word frequency and L2 AoA was significant: the AoA effect was again larger for low frequent words. This is consistent with the findings of monolingual studies (Dirix & Duyck, in press; Gerhand & Barry 1999a;Wilson et al. 2013). Wilson et al. argue that this interaction can be explained through the processing speed of high vs low frequency words: orthographic familiarity is higher for high frequency words, so that they are more easily and rapidly accessible. For low frequency words, lower familiarity and processing speed leaves more room for an additional influence of other word characteristics, such as a faster access for early AoA words.

An interaction between L2 AoA and L1 proficiency was present in single fixation duration, first fixation duration and total reading time. The L2 AoA effect was less pronounced when the L1 proficiency of the participants was higher. A similar interaction has been found in the word frequency study by Cop, Keuleers, et al. (2015) using the same database and participant characteristics, between L1 proficiency and L2 word frequency. They argued that the L1 proficiency measure probably entails more than L1 exposure, possibly a general language skill or aptitude. In analogy with their proficiency – word frequency interaction, it is indeed not unreasonable to assume that more language proficient participants not only show reduced frequency effects, but also reduced AoA effects.

There was also a significant interaction between L2 AoA and orthographic overlap in gaze durations: the AoA effect was larger for words with a high amount of orthographic overlap with their translational equivalent. As words with high orthographic overlap are accessed more easily (e.g., Van Assche et al., 2011), it could be that they receive an additional boost when they have already resided for a long time in the representational network. Alternatively, cognate-like words may have larger semantic overlap across languages (Van Hell & De Groot, 1998a), so that they should yield a larger AoA effect if part of that effect originates from semantics.

Surprisingly, whereas L1 AoA did not have an influence on L2 processing in previous research (e.g., Hirsh et al., 2003; Izura & Ellis, 2002), we are the first to find an influence of L1 AoA on L2 reading: for all timed measures except single fixation duration, there was a (marginally) significant main effect of L1 AoA on L2 reading: words with an earlier L1 AoA received shorter fixations than words with a later L1 AoA. After further inspection, it seemed that the L1 AoA effect especially arose for L2 words that take longer to process (i.e. longer words of at least nine to twelve letters): there was a facilitatory effect on single/first fixation and gaze duration when the translation of these words were learned early in L1. This is plausible given that the L1 AoA effect on L2 reading is assumed to originate from shared semantics across languages, which takes time to activate during reading, especially for longer words.

In conclusion, an earlier learning age of L2 words facilitates L2 reading. In addition, L1 AoA also seems to play a role in several measures of L2 natural reading.

General Discussion

The age at which we learn words influences their processing speed (e.g., Brysbaert & Ghyselinck, 2006; Gerhand & Barry, 1999a; Juhasz & Rayner, 2006; Morrison et al., 1992). This mechanism also applies to L2 (Hirsh et al., 2003; Izura & Ellis, 2002, 2004), although earlier findings are limited to isolated L2 word processing. In the current study, we analyzed eye movement data of a corpus of bilingual natural reading (GECO; Cop, Dirix, Drieghe, et al., in press). Our first goal was to investigate L1 and L2 AoA effects in L1 and L2 reading using eye tracking, in order to provide a detailed analysis of the time course of AoA effects. Of particular interest was the effect of L2 (and potentially L1) AoA on L2 reading. Furthermore, we wanted to test the predictions of the semantic and mapping hypotheses, in order to clarify the origin of the AoA effect.

Consistent with previous monolingual research (Dirix & Duyck, in press; Juhasz & Rayner, 2006), we found an L1 AoA effect for L1 reading on both early (first fixation and gaze duration) and late measures (total reading times). It seems that AoA has an influence throughout the entire reading process, making it easier for earlier learned words to access the representations of words in lexical memory on the one hand, and to activate their meaning and integrate them into sentences on the other hand.

The effects of L2 AoA on L2 processing were consistent with the previous research on isolated word reading: fixation times were shorter for words that were learned earlier in L2. The current study however was the first investigation providing evidence from eye movements, showing that the L2 AoA effect affects the entire time course of L2 word recognition (in analogy with L1 AoA and L1 reading): L2 words that are learned earlier yield benefits for eye tracking measures that reflect initial lexical access, as well as for measures that reflect semantic access and integration. This is consistent with the notion that the origin of the AoA effect may situate itself at different representational levels.

Interestingly, we are also the first to find a cross-lingual AoA influence on L2 reading: in the early reading stages (single/first fixation and gaze duration), longer L2 words were processed faster when their L1 translation was learned early. This is consistent with a semantic etiology of the AOA effect: if one assumes that L2 translational equivalents are mapped onto the existing semantic representations that also serve L1 (Duyck & Brysbaert, 2004; Kroll & Stewart, 1994; Van Hell & De Groot, 1998b), L2 processing should indeed be influenced by L1 AoA, because that measure reflects when the semantic representation that the L2 word is mapped onto, was created (Izura & Ellis, 2002). As noted, this effect interacted with word length: processing is slower for longer words (especially in L2), so it could be that only for these words sufficient time surpasses for this semantic activation to occur. Only then the L1 AoA influence, which originates from the semantic organization of the word network, may influence L2 word recognition.

This cross-lingual AoA effect contrasts with earlier investigations of L2 AoA, who only reported L2 AoA effects on L2 processing (Hirsh et al., 2003; Izura & Ellis, 2002, 2004). There are several reasons to explain this discrepancy. First, there is a potential influence of task characteristics. Whereas participants simply have to read the presented text in natural reading, in other paradigms there can be influences of decision components or answer strategies. Kuperman, Drieghe, Keuleers, and Brysbaert (2013) indeed show that shared variance between lexical decision RTs and eye movement measures may be surprisingly low. This could mean that these two tasks, although they both involve visual word recognition, partially tap into different processes. Second, because we included a large amount of stimuli and the AoA variables as continuous, as opposed to factorial designs, our approach might be more sensitive to discover the subtle effects of L1 AoA on L2 reading. Third, this approach also allowed us to include complex interactions. The small L1 AoA influence on L2 reading was found in the interaction with word length and language. Finally, note that the lexical decision tasks of Izura and Ellis (2002, 2004), who only found within-language AoA effects, likely involve semantics, in order to determine whether the letter string corresponds to an existing meaning. In the present study, the eye tracking measures that reflect later stages of word recognition (e.g. total reading times) also only showed within-language AoA effects, similar to Izura and Ellis, and in contrast with the early eye tracking measures that reflect initial lexical access. In the Supplementary Materials, we present data from a lexical decision task with the target words of the current study, in which we replicate the null cross-lingual AoA effect of Izura and Ellis.

Finally, we found that L2 AoA has an influence on processing of very low-frequent L1 words. A possible explanation may lie in the higher activation threshold of low-frequency words (McClelland & Rumelhart, 1981). Because we know that lexical access in bilinguals is non-selective (Dijkstra & Van Heuven, 2002; Van Assche, Duyck, Hartsuiker, & Diependaele, 2009), and following the rationale of the semantic hypothesis (i.e., organization of words in a semantic network; (Brysbaert, Van Wijnendaele, et al., 2000; Steyvers & Tenenbaum, 2005), it could be that exposure to L2 words that are learned early affects the position of the underlying semantic representation in the network sufficiently if that representation was very weak to start with (very low-frequent L1 words). Note that this L2 AoA effect was stronger for participants with low L1 proficiency scores. It could be that especially for low-proficient participants, who are supposed to have weaker representations of low frequency words (e.g., Cop, Keuleers, et al., 2015), the L2 AoA order indeed influences these low frequency representations even more strongly, as argued before.

Hypotheses of the AoA effect

We outlined two important hypotheses explaining the AoA effect. The semantic hypothesis situates the origin of AoA effects in the accessibility of semantic representations. It predicts that L2 reading should be affected by L1 AoA, if one assumes that L1 and L2 translational equivalents share semantic representations (Kroll & Stewart, 1994; Van Hell & De Groot, 1998b). In contrast, the mapping hypothesis postulates that it is only the order in which information enters a network determines AoA effects, as there is a plasticity (and processing) advantage for early entered information. As new input-output mappings (corresponding to word form and meaning) have to be installed when learning a new language, within-language AoA effects should emerge: the L2 AoA effect should be in accordance with the age at which the words were learned in L2.

In our study, we found results that support both these hypotheses. First, and most clearly, there is a within-language effect of (L2) AoA on L2 processing: on all timed measures, L2 reading was faster when the words were learned earlier in L2. This finding supports the mapping hypothesis, as the L2 learning order is determining here. The mapping hypothesis does not specify a particular linguistic level at which AoA effects could arise, and we indeed found that L2 AoA influenced both measures that reflect lexical access (single/first fixation and gaze duration) and access to the meaning or verification of the words (total reading time). However, there was also a limited but reliable effect of L1 AoA on L2 reading in our data, which supports the semantic hypothesis. For longer L2 words (only), processing is speeded if their L1 translational equivalent is early acquired. To sum up, it seems that AoA effects in late language processing are language-exclusive, but cross-lingual L1 AoA effects show op for longer L2 words that take longer to process.

In the L2 AoA or OoA literature, the semantic and mapping hypotheses are often portrayed as opposites, with specific predictions, that usually result in support of the mapping hypothesis. In an attempt to reconcile this with the current findings, we suggest an integration between the mapping and semantic hypotheses. In AoA/OoA research, there seems to be a general principle of “first learned, faster processed”. The mapping hypothesis provides an excellent and parsimonious explanation for this finding. However, we have to keep in mind that we are studying language. It is not unreasonable that words in different languages, but with the same meaning, share semantic representations. These representations are more easily accessed when learned earlier, whether it is through the L1 or L2. It can indeed be the case that early learned words can alter a network’s weights in its advantage more than late learned words, but at the same time it may also be that the semantic representation of the early learned word takes up a more central place in the network. Both of these AoA mechanisms then may influence the processing speed of words independently and simultaneously.

This brings us to two additional related topics: the organization of the (bilingual) lexicon and the critical acquisition period. From the interpretation of our results, we can conclude that AoA heavily influences the organization of the lexicon: the age at which you learn a word has a large impact on the position it will take up in the lexicon, and how easily accessible it will be. In the specific case of bilinguals, our results also point towards a shared lexicon for the two languages (see Dijkstra & van Heuven, 2002), as direct influences of L1 word characteristics on their L2 counterpart seem to take place. Furthermore, as AoA effects also emerge in late-learned L2, Izura & Ellis (2002) argued that the AoA effect is probably not due to some kind of critical period of ‘easy’ language acquisition (see Marinova-Todd, Marshall, & Snow, 2000) that would only apply to the period of L1 learning. There is indeed evidence that the onset of learning a new language has a later limited influence on word recognition processes (if controlling for proficiency; e.g., Cardimona, Smith & Roberts, 2015; Foote, 2014; Montrul & Foote, 2014). It seems that there however is a “relative” critical period in language acquisition: irrespective of the language or the age at which you start learning it, the order in which you learn the words will have an impact on their processing, with an advantage for what was acquired first.

Conclusion

In this eye tracking mega study of bilingual reading, we confirmed that L2 AoA also influences L2 natural reading. The AoA effect is however not only determined by the age at which the word was learned in L2, but also to a lesser extent by the age at which its translational equivalent was learned in L1. As the semantic and mapping hypotheses are not mutually exclusive, we propose an integration between these two to account for these results.

Acknowledgements

This research was supported by a grant of concerted research action BOF13/GOA/032 awarded by the special research fund of Ghent University.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. http://doi.org/10.1016/j.jml.2007.12.005

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283–316. http://doi.org/10.1037/0096-3445.133.2.283

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*(3), 445–459. http://doi.org/10.3758/BF03193014

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. http://doi.org/10.1016/j.jml.2012.11.001

Barry, C., Johnston, R. A., & Wood, R. F. (2006). Effects of age of acquisition, age, and repetition priming on object naming. *Visual Cognition*, *13*(7–8), 911–927. http://doi.org/10.1080/13506280544000101

Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: evidence for a lexical-semantic competition hypothesis. *Cognition*, *96*, B45-54. http://doi.org/10.1016/j.cognition.2004.11.006

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2001). Age-of acquisition and word frequency in the lexical decision task: Further evidence from the French language. *Cahiers de Psychologie Cognitive - Current Psychology of Cognition*, *20*(6), 401–443.

Brysbaert, M. (n.d.). Age of acquisition ratings score better on criterion validity than frequency trajectory or ratings "corrected" for frequency. *Quarterly Journal of Experimental Psychology*.

Brysbaert, M., & Ghyselinck, M. (2006). The effect of age of acquisition: Partly frequency related , partly frequency independent. *Visual Cognition*, *13*, 992–1011. http://doi.org/10.1080/13506280544000165

Brysbaert, M., Lange, M., & Van Wijnendaele, I. (2000). The effects of age-of-acquisition and frequency-of-occurrence in visual word recognition : Further evidence from the Dutch language. *European Journal of Cognitive Psychology*, *12*(1), 65–86.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta Psychologica*, *150*, 80–84. http://doi.org/10.1016/j.actpsy.2014.04.010

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 441–

458. doi:10.1037/xhp0000159

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, *104*, 215–226.

Cardimona, K., Smith, P., & Roberts, L. S. (2015). Lexical Organization in Second Language Acquisition: Does the Critical Period Matter? *TESOL Journal, 7*(3), 540–565. doi:10.1002/tesj.219

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, *25*(1), 85–95. http://doi.org/10.1080/14640747308400325

Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (n.d.). Presenting GECO : An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods*. http://doi.org/10.3758/s13428-016-0734-0

Cop, U., Dirix, N., Van Assche, E., Drieghe, D., & Duyck, W. (n.d.). Bilinguals reading a novel: An eye movement study of cognate facilitation in L1 and L2 natural reading. *Bilingualism: Language and Cognition*. http://doi.org/10.1017/S1366728916000213

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. http://doi.org/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. http://doi.org/10.3758/s13423-015-0819-2

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *The Quarterly Journal of Experimental Psychology*, *60*(8), 1072–1082. http://doi.org/10.1080/17470210701315467

Cortese, M. J., & Schock, J. (2012). Imageability and age of acquisition effects in disyllabic word recognition. *The Quarterly Journal of Experimental Psychology*, *66*(5), 946–972. http://doi.org/10.1080/17470218.2012.722660

CRR. (2014). wordORnot. Retrieved September 1, 2015, from http://crr.ugent.be/archives/1533

De Moor, W., Ghyselinck, M., & Brysbaert, M. (2000). A validation study of the age-of-acquisition norms collected by Ghyselinck, De Moor, & Brysbaert. *Psychologica Belgica*, *40*(2), 99–114.

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*(3), 175–197. http://doi.org/10.1017/S1366728902273010

Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: insights from connectionist networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1103–1123. http://doi.org/10.1037/0278-7393.26.5.I103

Foote, R. (2014). Age of Acquisition and Sensitivity to Gender in Spanish Word Recognition. *Language Acquisition, 21*(4), 365–385. doi:10.1080/10489223.2014.892948

Gerhand, S., & Barry, C. (1999a). Age of acquisition, word frequency, and the role of phonology in the lexical decision task. *Memory & Cognition*, *27*(4), 592–602. http://doi.org/10.3758/BF03211553

Gerhand, S., & Barry, C. (1999b). Age-of-acquisition and frequency effects in speeded word naming. *Cognition*, *73*(2), B27–B36. http://doi.org/10.1016/S0010-0277(99)00052-9

Ghyselinck, M., Custers, R., & Brysbaert, M. (2004). The Effect of Age of Acquisition in Visual Word Processing: Further Evidence for the Semantic Hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 550–554.

http://doi.org/10.1037/0278-7393.30.2.550

Hirsh, K. W., Morrison, C. M., Gaset, S., & Carnicer, E. (2003). Age of acquisition and speech production in L2. *Bilingualism: Language and Cognition*, *6*, 117–128. http://doi.org/10.1017/S136672890300107X

Izura, C., & Ellis, A. W. (2002). Age of acquisition effects in word recognition and production in first and second languages. *Psicológica*, *23*, 245–281.

Izura, C., & Ellis, A. W. (2004). Age of acquisition effects in translation judgement tasks. *Journal of Memory and Language, 50*, 165–181. Retrieved from http://ac.els-cdn.com/S0749596X03001165/1-s2.0-S0749596X03001165-main.pdf?_tid=19b3bb46-5789-11e5-9bbe-00000aab0f26&acdnat=1441868385_cc6f17a2dd58a1c4b65b15075c5a815b

Izura, C., Pérez, M. a., Agallou, E., Wright, V. C., Marín, J., Stadthagen-González, H., & Ellis, A. W. (2011). Age/order of acquisition effects and the cumulative learning of foreign words: A word training study. *Journal of Memory and Language*, *64*, 32–58. http://doi.org/10.1016/j.jml.2010.09.002

Johnston, R. a., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition*, *13*, 789–845. http://doi.org/10.1080/13506280544000066

Joseph, H. S. S. L., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition*, *133*(1), 238–48. http://doi.org/10.1016/j.cognition.2014.06.015

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. *Psychological Bulletin*, *131*(5), 684–712. http://doi.org/10.1037/0033-2909.131.5.684

Juhasz, B. J., Gullick, M. M., & Shesler, L. W. (2011). The Effects of Age-of-Acquisition on Ambiguity Resolution : Evidence from Eye Movements. *Journal of Eye Movement*

Research, *4*(1), 1–14.

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *29*(6), 1312–1318. http://doi.org/10.1037/0278-7393.29.6.1312

Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, *13*(7–8), 846–863. http://doi.org/10.1080/13506280544000075

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. B*ehavior Research Methods, 42*(3), 627–633. doi:10.3758/brm.42.3.627

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), 643–50. http://doi.org/10.3758/BRM.42.3.643

Kroll, J. F., & Stewart, E. (1994). Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections Between Bilingual Memory Representations. *Journal of Memory and Language*, *33*(2), 149–174. http://doi.org/10.1006/jmla.1994.1008

Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). 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*, 563–580. http://doi.org//10.1080/17470218.2012.658820

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990. http://doi.org/10.3758/s13428-012-0210-4

Lambon Ralph, M. a., & Ehsan, S. (2006). Age of acquisition effects depend on the mapping between representations and the frequency of occurrence: Empirical and computational evidence. *Visual Cognition*, *13*(7–8), 928–948. http://doi.org/10.1080/13506280544000110

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: a quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–43. http://doi.org/10.3758/s13428-011-0146-0

Marinova-Todd, S. H., Marshall, D. B., & Snow, C. E. (2000). Three Misconceptions about Age and L2 Learning. *TESOL Quarterly*, *34*, 9–34. http://doi.org/10.2307/3588095

McClelland, J. L., & Rumelhart, D. E. (1981). An Interactive Activation Model of Context Effects in Letter Perception: Part I. An Account of Basic Findings. *Psychological Review*, *88*, 375–407. http://doi.org/10.1016/B978-1-4832-1446-7.50048-0

Menenti, L., & Burani, C. (2007). What causes the effect of age of acquisition in lexical processing? *The Quarterly Journal of Experimental Psychology*, *60*(5), 652–660. http://doi.org/10.1080/17470210601100126

Monaghan, P., & Ellis, A. W. (2010). Modeling reading development: Cumulative, incremental learning in a computational model of word naming. *Journal of Memory and Language*, *63*(4), 506–525. http://doi.org/10.1016/j.jml.2010.08.003

Montrul, S., & Foote, R. (2014). Age of acquisition interactions in bilingual lexical access: A study of the weaker language of L2 learners and heritage speakers. *International Journal of Bilingualism, 18*(3), 274–303. doi:10.1177/1367006912443431

Morrison, C. M., & Ellis, A. W. (1995). Roles of Word Frequency and Age of Acquisition in Word Naming and Lexical Decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 116–133.

Morrison, C. M., Ellis, a W., & Quinlan, P. T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition*, *20*(6), 705–714. http://doi.org/10.3758/BF03202720

Pérez, M. A. (2007). Age of acquisition persists as the main factor in picture naming when cumulative word frequency and frequency trajectory are controlled. *The Quarterly Journal of Experimental Psychology*, *60*(1), 32–42. http://doi.org/10.1080/17470210600577423

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. http://doi.org/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*(8), 1457–1506. http://doi.org/10.1080/17470210902816461

Stewart, N., & Ellis, A. W. (2008). Order of acquisition in learning perceptual categories: a laboratory analogue of the age-of-acquisition effect? *Psychonomic Bulletin & Review*, *15*(1), 70–4. http://doi.org/10.3758/PBR.15.1.70

Steyvers, M., & Tenenbaum, J. B. (2005). The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science*, *29*(1), 41–78. http://doi.org/10.1207/s15516709cog2901_3

Van Hell, J. G., & De Groot, A. M. B. (1998). Disentangling context availability and concreteness in lexical decision and word translation. *The Quarterly Journal of Experimental Psychology Section A*, *51*(1), 41–63. http://doi.org/10.1080/713755752

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2013). SUBTLEX-UK: a new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. http://doi.org/10.1080/17470218.2013.850521

Wilson, M. A., Cuetos, F., Davies, R., & Burani, C. (2013). Revisiting age-of-acquisition effects in Spanish visual word recognition: the role of item imageability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1842–1859. http://doi.org/10.1037/a0033090

APPENDIX: L2 AoA rating instructions

“Welcome to this experiment. You will have to rate words on the age at which you've learned them. By this we mean the age at which you completely understood the word when someone used it, even if you didn't use it yourself. Please only use round numbers in the list (you'll get an error message if you don't). Some examples:

If you think you learned the word "love" at age 8, fill in 8.

If you think you learned the word "neuroscientist" at age 18, fill in 18.

You have to fill in the number in the column "age learned". If you don't know the word, put an 'x' in the column "word unknown". When your input is correct (either a number or an x), the column next to your input will turn from red to green. Please make sure to fill in every row. Try to use the whole age range from when you started to learn English words up to now. Try to make an estimation as good as possible, but don't think too long about a word.”