Language and Cognition (2023), 15: 2, 355–376
doi:10.1017/langcog.2022.40

ARTICLE

The age of acquisition effect in processing second language words and its relationship with the age of acquisition of the first language

Jue Wang[1], Lijuan Liang[2] and Baoguo Chen[1]*

[1]Beijing Key Laboratory of Applied Experimental Psychology, Faculty of Psychology, Beijing Normal University, Beijing 100875, China; [2]Bilingual Cognition and Development Lab, Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou 510420, China
*Corresponding author. Email: chenbg@bnu.edu.cn

(Received 24 May 2022; Revised 18 December 2022; Accepted 18 December 2022)

Abstract

The present study investigated the age of acquisition (AoA) effect in processing second language (L2) words and how it is related to the AoA of the corresponding first language (L1) words. We adopted a lexical decision task in three experiments. The filler words were orthographically illegal in Experiment 1 to elicit more word form processing, while Experiment 2 used legal fillers to shift the bias toward semantic processing. In Experiment 3, we used a larger amount of stimuli containing more longer words with legal fillers. Our results showed that L2 AoA has a weak effect at the orthographical processing level and a stable effect at the semantic processing level. The L1 AoA modulates the L2 AoA effect at the semantic processing level, which is more likely to appear in long words. These results suggest that it is important to take bilingual representation and activation into consideration to explain the L2 AoA effect.

Keywords: age of acquisition; second language; first language; word processing; word length

1. Introduction

Age of acquisition (AoA) refers to the age at which people learn a word, which is a robust variable affecting word recognition (Rochford & Williams, 1962). Generally, early acquired words are processed more quickly and accurately than late acquired words, which is known as the AoA effect. This effect reveals the impact of early learning experience on adult word processing (Arnon et al., 2017; Chang & Lee, 2020). The AoA effect has been widely investigated in first language (L1) and has been found in different populations, languages, and tasks (see review papers by Brysbaert & Ellis, 2016; Ellis, 2012; Hernandez & Li, 2007; Johnston & Barry, 2006; Juhasz, 2005).

Recently, the AoA effect in the second language (L2) has gradually attracted the attention of researchers (Dirix & Duyck, 2017; Wang & Chen, 2020). Given that L2 word learning generally involves associating the new word forms with the pre-existing semantic representations created in L1, L2 words may have two AoAs. One is the L2 AoA and the other is the AoA of their L1 translation equivalents (i.e., L1 AoA; Dirix & Duyck, 2017; Izura & Ellis, 2002, 2004; Wang & Chen, 2020). The two AoAs may be synchronous, that is, words are acquired early both in L1 and L2 (e.g., words related to animals) or acquired late both in L1 and L2 (e.g., abstract proper nouns). On the other hand, the two AoAs may be asynchronous, that is, words are acquired early in L1 but late in L2 (e.g., words related to fairy tales) or acquired early in L2 but late in L1 (e.g., words related to traveling). The asynchronization allows the separation of the two AoAs, so that the investigation of the L2 AoA effect can be independent of L1 AoA. Moreover, considering that L2 word processing differs from L1 word in many aspects (Kroll et al., 2015; Kroll & Stewart, 1994), the AoA effect in L2 may also be different from that in L1. As far as we know, studies on the L2 AoA effect are very rare (Dirix & Duyck, 2017; Hirsh et al., 2003; Izura & Ellis, 2002, 2004; Xue et al., 2017), and one important question remains unclear, that is, what is the relationship between the L2 AoA effect and L1 AoA? We will elaborate on this question below.

1.1. Previous studies on the L2 AoA effect and its relationship with L1 AoA

To the best of our knowledge, Izura and Ellis (2002) were the first researchers to propose the concept of L2 AoA. They reported that early acquired L2 words were processed faster and more accurately than late acquired L2 words in picture naming and lexical decision tasks. Then, different tasks (e.g., semantic relatedness judgment and text reading tasks) were used in several studies and replicated the results that L2 AoA does affect L2 word recognition (Dirix & Duyck, 2017; Hirsh et al., 2003; Izura & Ellis, 2004; Xue et al., 2017). Furthermore, Volkovyskaya et al. (2017) extended the L2 AoA research from psycholinguistics to memory, showing the role of L2 AoA in picture and word free recall tasks.

For the L2 AoA effect observed in those previous studies mentioned above, what is its relationship with the L1 AoA (i.e., the AoA of its L1 translation equivalents)? To answer this question, Izura and Ellis (2002, Experiment 3) manipulated the consistency of L1 AoA and L2 AoA. They selected Spanish (L1)–English (L2) translation word pairs as materials. For half of the word pairs, L1 words were learned early and L2 words were learned late. For the other half of the word pairs, L1 words were learned late and L2 words were learned early. The results showed that L2 lexical decisions were only affected by L2 AoA regardless of the age at which the corresponding L1 words were acquired. However, since partial factor design could not examine the interaction between the two AoAs, it may not be an appropriate method to fully testify whether L2 AoA effects are truly unrelated to L1 AoA. In a further study, Izura and Ellis (2004) adopted a full factorial design. They manipulated L2 AoA and L1 AoA simultaneously, using Spanish (L1)–English (L2) translation word pairs as materials. Participants were asked to decide whether the Spanish (L1) and English (L2) words had the same meaning or not. Results showed significant main effects of L1 AoA and L2 AoA, but their interaction was not significant. Dirix and Duyck (2017) extended L2 AoA research from isolated words to text reading. Unbalanced Dutch–English bilinguals were asked to read the novel "The Mysterious Affair at Styles," which was presented half in Dutch (L1) and half in English (L2). They analyzed the eye movement measures of 1,069 Dutch nouns and 966 English nouns,

https://doi.org/10.1017/langcog.2022.40 Published online by Cambridge University Press

including single/first fixation duration, gaze duration, and total reading time. The results showed within-language AoA effects, that is, L1 AoA influenced L1 reading and L2 AoA influenced L2 reading. In addition, they found cross-language AoA effects in long (at least 9 to 12 letters) L2 words in the single/first fixation and gaze duration measures, showing that longer L2 words were processed faster when their L1 translation equivalents were learned early.

Generally speaking, Izura and Ellis (2002, 2004) found that the L2 AoA effect is independent of L1 AoA. Dirix and Duyck (2017) did not report the relationship between the L2 AoA effect and L1 AoA, but a cross-language AoA effect was observed such that L1 AoA influenced L2 long words processing. These studies differed in many aspects, such as experimental designs and tasks. Specifically, Izura and Ellis (2002, 2004) adopted factorial designs combined with isolated word paradigms, using picture naming, lexical decision (i.e., fillers were legal nonwords), and translation judgment tasks. Dirix and Duyck (2017) adopted a mega study combined with a text reading task, using eye-tracking technology. In sum, it still remains unclear about the exact relationship between the L2 AoA effect and L1 AoA. Our study aimed to further explore this issue.

1.2. Mechanisms underlying the AoA effect

There are mainly two important theories to explain the AoA effect, the Semantic Hypothesis and the Arbitrary Mapping Hypothesis. The Semantic Hypothesis holds that the AoA effect primarily originates from the semantic system (Brysbaert et al., 2000; van Loon-Vervoorn, 1989). Late acquired words are learned by establishing a link with the semantic representation of early acquired words. So early acquired words locate at the hub of the semantic network system, which makes them easily accessible. According to the Revised Hierarchical Model, for unskilled bilinguals, their two languages share semantic representations and the semantic access of L2 words is mediated by L1 translation equivalents (Kroll & Stewart, 1994). So, if we combine the Semantic Hypothesis and the Revised Hierarchical Model together, AoA effects that appear in L2 could probably be attributed to the AoA of the corresponding L1 words in the task of semantic processing by unskilled bilinguals.

Unlike the Semantic Hypothesis, the Arbitrary Mapping Hypothesis proposes that AoA effects could appear at multiple processing levels rather than a single level. The AoA effect depends on the nature of the mappings between input (e.g., orthography) and output (e.g., semantics) representations that are formed during the development of the lexical network. Words that are learned early have a large influence on the connection weights of the lexical network. With the loss of network plasticity, late acquired words need to be adapted to the lexical network constructed by early acquired words (Ellis & Lambon Ralph, 2000; Monaghan & Ellis, 2002). When learning L2 words, learners establish new mappings among orthographical, phonological, and semantic representations, and bring about new AoA effects (i.e., L2 AoA effect), which is independent of L1 AoA. Izura and Ellis (2002, 2004) found that the L2 AoA effect was independent of the AoA of L1 translation equivalents, supporting the Arbitrary Mapping Hypothesis but not the Semantic Hypothesis. However, Dirix and Duyck (2017) found that L2 AoA has an influence not only on initial lexical access, but also on later semantic access. In addition, at the initial lexical access stage, longer L2 words were processed faster when their L1 translation equivalents were learned early. They proposed that these results suggest a necessity to combine the Semantic Hypothesis and the Arbitrary Mapping Hypothesis.

Here, it should be noted that the Semantic Hypothesis and the Arbitrary Mapping Hypothesis are mainly used to explain the L1 AoA effect. Given that L2 words differ from L1 words in many aspects, such as accessibility and conceptual representation (Kroll et al., 2015; Kroll & Stewart, 1994), the mechanisms of L2 and L1 AoA effects may not be exactly the same. In order to explain the L2 AoA effect, the processing characteristics of L2 words may need to be considered. Taking Chinese–English bilinguals, for example, the two languages have different word forms, but share semantic representations (Ding et al., 2003). So, the word form processing of the two languages may be independent of each other. This way, if AoA influences word form processing, the L2 AoA effect might not be modulated by L1 AoA. However, this may not be the case at the semantic processing level, according to the Revised Hierarchical Model. This model proposes that if bilinguals acquire L2 quite late and their L2 proficiency level is lower than that of their native language, the semantic access of L2 words is usually mediated by the L1 words (Kroll & Stewart, 1994), which leads to the influence of L1 AoA on the L2 AoA effect.

In sum, empirical studies on the L2 AoA effect and its relationship with L1 AoA are rare, and controversies still exist among the existing studies. Moreover, current theories about the AoA effect are mainly based on L1 word processing, and have not taken the characteristics of L2 lexical representation and access into consideration. Therefore, in the present study, we conducted three experiments to further investigate the relationship between the L2 AoA effect and L1 AoA, and the origins of the L2 AoA effect.

In Experiments 1 and 2, we manipulated L2 AoA and L1 AoA simultaneously. To avoid the interference of different tasks, both experiments used lexical decision task. The filler words in Experiment 1 were orthographically illegal (created with random consonant strings) and the fillers in Experiment 2 were orthographically legal (created by changing one of the letters in real words). Previous studies have found that different types of filler words can elicit different processing strategies (Armstrong & Plaut, 2016; Evans et al., 2012). If filler words are orthographically illegal nonwords (e.g., random consonant strings, like "jqmssp"), lexical judgment will be mainly based on word form information and involve less semantic information. If nonword fillers are more word-like (e.g., fillers created by changing one letter in real English words, like "trith" for "truth"), the magnitude of semantic activation will increase significantly, because it would be wrong to rely solely on word form information to make lexical decision (Menenti & Burani, 2007). In other words, the type of filler words determines the amount of semantic activation in a lexical decision task: more semantic activation for orthographically legal fillers, and more word form processing/less semantic activation for orthographically illegal fillers.

Experiments 1 and 2 used a factorial design. Factorial designs are frequently used to investigate the variables of interest under strict control of irrelevant variables. However, this design, as a small-scale design, has great restrictions on materials, which might affect the generalizability of the results (Brysbaert & Ellis, 2016). To address this issue, we conducted a mega study in Experiment 3, using a large number of materials. Mega studies, based on the regression analysis, allow us to investigate the independent and simultaneous effects of multiple variables. Using continuous variables instead of dichotomous variables improves the validity of statistical analyses (Balota et al., 2004; Cortese et al., 2018; González-Nosti et al., 2013).

https://doi.org/10.1017/langcog.2022.40 Published online by Cambridge University Press

According to the Arbitrary Mapping Hypothesis, the L2 AoA effect would not be influenced by L1 AoA at either word form level or semantic processing level. According to the Semantic Hypothesis and the nature of L2 lexical semantic access for unskilled bilinguals, the L2 AoA effect would be influenced by L1 AoA in the task biased toward semantic processing.

2. Experiment 1

In Experiment 1, we manipulated L2 AoA and L1 AoA to explore the AoA effect at the word form level. We adopted a lexical decision task with orthographically illegal filler (i.e., consonant strings).

2.1. Method

2.1.1. Participants

Forty unbalanced Chinese (L1)–English (L2) bilinguals participated in Experiment 1 for monetary compensation. Data from one participant were deleted because the accuracy rate (ACC) was below 85%, so statistical analyses were conducted using the final set of 39 participants (30 females). The participants were from several universities in Beijing and were born in China with no background of immigration or overseas education. All participants had begun learning English after childhood in a classroom setting. The time (years) they spent on learning English is presented in Table 1. Their Chinese and English proficiency levels were assessed by six-point self-assessment ratings, and their English proficiency was also assessed using the Oxford Placement Test (OPT; Table 1). The participants were right-handed with normal or corrected-to-normal vision. They signed a consent form before the experiment. They were not told about the purpose of the experiment. This study was approved by the Ethics Committee of the Faculty of Psychology, Beijing Normal University.

2.1.2. Design and materials

The experiment was a 2 (L2 AoA: early and late) × 2 (L1 AoA: early and late) within-subjects design.

Seventy-six English words were selected from a recently developed database of the AoA ratings of English words and their Chinese translation equivalents (Wang & Chen, 2020). In this database, L2 AoA was rated by asking college students whose native language is Chinese to write the age (in years) at which they thought they had

Table 1. The background information of the participants in Experiment 1

	Mean	*SD*	Range
Age (in years)	21.69	1.88	18–25
Years of L2 learning	12.81	2.57	9–19
L1 proficiency	5.24	0.42	4.90–5.68
L2 proficiency	3.41	0.70	3.17–3.60
OPT	39.09	4.61	27–49

Note: Years of L2 learning represents the total time (in years) participants spent on learning English. L1/L2 proficiency, measured on a six-point scale, is an average of the four self-reported scores in listening, speaking, reading, and writing. OPT is an English proficiency test, which consists of 25 multiple-choice questions and a cloze test, with a total score of 50 (Allan, 2004).

learned each English word in either spoken or written form (Brysbaert et al., 2014a). The participants were given the following instruction: "If you think you learned 'arm' at 7, please fill in 7." L1 AoA was estimated using a 1–7 scale, where "1" meant the word was acquired at 1–2 years old, "2" meant it was acquired at 3–4 years old, and "7" meant the word was acquired at 13 years old (Gilhooly & Gilhooly, 1979). The instruction presented for L1 AoA rating was: "If you think you learned 糖果 (candy) when you were 3 years old, please fill in 2.". Among the 76 English words adopted in the present study, 19 words were learned early in both L2 and L1, 19 were learned early in L2 but late in L1, 19 were learned late in L2 but early in L1, and 19 were learned late in both L2 and L1.

In Experiment 1, we manipulated L2 AoA and L1 AoA simultaneously (i.e., early L2 AoA and early L1 AoA, early L2 AoA and late L1 AoA, late L2 AoA and early L1 AoA, late L2 AoA and late L1 AoA), and repeated-measures ANOVA was used to test whether L2 AoA and L1 AoA effects were significantly different across these four conditions. When L2 AoA was manipulated, there was a significant difference between L2 early acquired (ranged from 11.09 to 13.05) and L2 late acquired condition (ranged from 14.00 to 16.05) ($F(1,72) = 476.83$, $p < 0.001$), and neither the main effect of L1 AoA nor the interaction between the two AoAs was significant. Similarly, when L1 AoA was manipulated, there was a significant difference between L1 early acquired (ranged from 2.50 to 3.83) and L1 late acquired condition (ranged from 4.78 to 6.89) ($F(1,72) = 418.33$, $p < 0.001$), and neither the main effect of L2 AoA nor their interaction was significant. See Table 2 for examples.

The word frequency of the target words was calculated based on the SUBTLEX-UK (van Heuven et al., 2014), and concreteness (a five-point scale, 1 low to 5 high) was cited from Brysbaert et al. (2014a). Repeated-measures ANOVA was adopted to test whether word frequency, concreteness, and word length were matched across these four conditions. The results showed neither the main effects of L2 AoA and L1 AoA, nor their interaction was significant ($Fs < 1.64$, $ps > 0.10$), indicating that the four conditions were matched on these three variables. The means and standard deviations for the characteristics of the English words are shown in Table 3.

2.1.3. *Procedure*

Participants were tested individually in a quiet room. Stimuli were presented on a computer using E-prime software version 2. The words were displayed in Times New Roman, font size 32 in white color on a black background. The distance between the participant and the screen was about 60 cm. The experiment began with the presentation of a fixation cross at the center of the screen for 500 ms. After the fixation, a blank screen was shown for 500 ms, and then the target word appeared on the screen until a response was registered or after 2,000 ms elapsed with no response.

Table 2. Example words used in Experiment 1

L2 AoA	L1 AoA	Target word	Chinese translation equivalent
Early	Early	Season	季节 (*jijie*)
Early	Late	Picnic	野餐 (*yecan*)
Late	Early	Pillow	枕头 (*zhentou*)
Late	Late	Helmet	头盔 (*toukui*)

Note: All Chinese translation equivalents are two-character words. Italics represents the pronunciation (i.e., *pinyin*) of Chinese.

https://doi.org/10.1017/langcog.2022.40 Published online by Cambridge University Press

Table 3. Descriptive statistics for target words used in Experiment 1 (standard deviation)

L2 AoA	L1 AoA	L2 AoA	L1 AoA	L2 Frequency	L2 Concreteness	L2 Length
Early	Early	12.09 (0.48)	3.31 (0.37)	42.81 (40.81)	4.26 (0.68)	6.11 (2.73)
Early	Late	12.30 (0.44)	5.29 (0.41)	36.15 (44.43)	4.45 (0.66)	6.26 (1.91)
Late	Early	14.49 (0.44)	3.41 (0.34)	22.14 (26.28)	4.50 (0.80)	5.68 (1.20)
Late	Late	14.65 (0.61)	5.36 (0.46)	34.42 (53.68)	4.29 (0.62)	5.95 (1.58)

Note: L2 AoA was rated by asking participants to write the age (in years) at which they thought they learned the words (Brysbaert et al., 2014a). L1 AoA, rated on a seven-point scale, was cited from Wang and Chen (2020) and the present study.[1] L2 Concreteness, rated on a five-point scale (1 low to 5 high), was cited from Brysbaert et al. (2014b). The L2 word frequency (per million) was cited from the SUBTLEX-UK (van Heuven et al., 2014).

Participants were asked to judge whether the English word was a real word or a nonword as quickly and accurately as possible. Half of the participants were asked to press the F key for real word and J key for nonword. The other half were asked to press the J key for real word and F key for nonword. After an interval of 1,000 ms, the next trial began. The order of stimulus presentation was randomized for each participant. Before the formal experiment, 10 practice trials were conducted to help the participants get familiar with the procedure.

2.1.4. Data analysis

Response time (RT) and ACC were analyzed in R using linear mixed-effects models (LMEMs) from the lme 4 package (Baayen et al., 2008; Bates et al., 2014). We tried to include the maximal random effects structure (Brown, 2021; Karimi & Diaz, 2020). First, we considered the full model with random intercepts for both participants and stimuli, as well as by-participants random slopes of the main effects and their interaction. If this model failed to converge, it was simplified by first removing the correlation between random intercepts and random slopes, and then removing the by-participants random intercept. Since no models converged, we further simplified the random effects structures by removing the random slopes of the main effects and their interaction. The random intercepts for both participants and stimuli were retained. Note that it might not be reasonable to include the by-stimuli random slopes in our experimental design, because each word has only one value for L2 AoA and L1 AoA. RT analyses were carried out on inverse RTs (inverse RTs = −1000/RTs) since they may produce a distribution that better matches the analysis assumptions (Brysbaert & Stevens, 2018; Catling & Elsherif, 2020). The ACCs were analyzed using logistic models. The materials, data, and scripts for Experiment 1 and subsequent experiments are available via the Open Science Framework (https://osf.io/fr9k7/).

2.2. Results

RTs faster than 300 ms and those beyond three standard deviations of the average mean, as well as incorrect responses, were excluded from the analyses (6%). Mean RTs and ACCs for the items are presented in Table 4. The parameter estimates from the linear models are presented in Table 5 with significant estimates shown in bold.

RT analysis showed that neither the main effects of L2 AoA and L1 AoA, nor their interaction was significant. Although there were no significant effects, it was still needed to determine whether there was no true difference or whether the above

https://doi.org/10.1017/langcog.2022.40 Published online by Cambridge University Press

Table 4. Mean response times (RTs, ms) and accuracy rates (ACCs, %) (standard deviations)

L2 AoA	L1 AoA	RT	ACC
Early	Early	610 (130)	96 (0.19)
Early	Late	604 (126)	98 (0.14)
Late	Early	620 (140)	93 (0.27)
Late	Late	616 (130)	98 (0.13)

Table 5. Linear mixed-effects models (LMEMs) estimates of fixed effects for response times (RTs) and accuracy rates (ACCs)

	RT			ACC		
	b	*SE*	*t*	*b*	*SE*	*z*
Intercept	**−1.71**	**0.03**	**−51.87**	**4.30**	**0.48**	**8.88**
L2 AoA	0.03	0.04	0.77	−1.12	0.57	−1.95^{+}
L1 AoA	−0.01	0.04	−0.33	0.49	0.63	0.79
L2 AoA × L1 AoA	0.003	0.06	0.05	1.34	0.89	1.52

Note: $+p < 0.10$.

analysis method could not detect it. Therefore, we used the "Bayes factor (BF)" to further examine these two possibilities (Cui et al., 2021). It was calculated in the R using the "Bayes Factor" package (Morey et al., 2015). BFs smaller than 1 favor that there is no true difference; otherwise, the alternative hypothesis is supported. The default prior width of $r = \sqrt{2}/2$ was used from the package. For RT, there was no difference between early L2 AoA and late L2 AoA conditions (BF = 0.20), between early L1 AoA and late L1 AoA (BF = 0.03). No interaction between L2 AoA and L1 AoA was found (BF = 0.04). BF regression analysis was consistent with the LMEMs analysis.

For the ACC, there was a marginally significant effect of L2 AoA, such that early acquired L2 words were identified more accurately than late acquired L2 words. The main effect of L1 AoA and the interaction of L2 AoA and L1 AoA were not significant.

2.3. Discussion

Experiment 1 found a marginally significant effect of L2 AoA on ACCs. This weak L2 AoA effect was probably related to the simple lexical decision task, in which filler words were consonant strings. Participants were able to rely more on word from information to make quick judgments, leading to a smaller L2 AoA effect.

To further study whether the L2 AoA effect would appear in a task involving more semantic processing, we changed the fillers in the lexical decision task from orthographically illegal ones (Experiment 1) to legal ones (Experiment 2). If the L2 AoA effect would be observed in Experiment 2, it may indicate that AoA plays a role at the semantic processing level.

Furthermore, a comparison of AoA effects in Experiments 1 and 2 would allow us to further study the relationship between the magnitude of the AoA effect and the need for semantic activation. If a larger AoA effect is found in Experiment 2, it means that semantic processing plays an important role in the AoA effect indeed.

3. Experiment 2

In Experiment 2, we investigated the L2 AoA effect at the semantic processing level, using a lexical decision task. The target words in Experiment 2 were identical to Experiment 1. The only difference between Experiments 1 and 2 was the filler words. In Experiment 2, fillers were orthographically legal, created by changing one of the letters in real words.

3.1. Method

3.1.1. Participants

Forty unbalanced Chinese (L1)–English (L2) bilinguals, none of whom had participated in the previous experiment or any rating assessments, participated in Experiment 2. Data from[1] participants were removed due to low accuracy (below 80%[2]), and the analyses were conducted on the dataset of 39 participants (31 females). The background of the participants in Experiment 2 was the same as that in Experiment 1. They were late unbalanced Chinese–English bilinguals (see Table 6).

3.1.2. Design and materials

The design and target words were identical to Experiment 1. Seventy-six filler words were orthographically legal and pronounceable nonwords, which were created by changing one of the letters in real English words (e.g., “trith” for “truth”) using Wuggy software (http://crr.ugent.be/programs-data/wuggy; Keuleers & Brysbaert, 2010). Since those filler words were more word-like, participants might rely more on semantic information rather than just word form information to make lexical decisions (Menenti & Burani, 2007). The word length of the nonwords was matched across the four conditions.

Table 6. The background information of the participants in Experiment 2

	Mean	SD	Range
Age (in years)	21.54	2.19	18–25
Years of L2 learning	12.76	2.66	6–19
L1 proficiency	5.57	0.21	5.23–5.77
L2 proficiency	3.89	0.29	3.60–4.00
OPT	38.73	3.58	29–46

Note: Years of L2 learning represents the total time (in years) participants spent on learning English. L1/L2 proficiency, measured on a six-point Likert scale, is averaged by four self-reported scores in listening, speaking, reading, and writing. OPT is an English proficiency test with a total score of 50 (Allan, 2004).

[1]Eighteen participants were recruited in the present study to rate the L1 AoA of 102 Chinese two-character words, with instructions and procedures similar to those of Wang and Chen (2020). We found the 102 words' AoA ratings showed a normal distribution (skewness is -0.42). AoA significantly correlated with word frequency ($r = -0.30$) and the number of strokes ($r = 0.20$). Besides, we compared our L1 AoA ratings with that of Xu et al. (2020), which collected AoA ratings for 19,716 simplified Chinese words. For the 101 words that overlapped, the correlation between the two AoA ratings was 0.83. These results indicated that the L1 AoA data collected in the present study were valid. The AoA data are available in the Supplementary Material.

[2]The filler words in Experiment 1 were orthographically illegal, but those in Experiment 2 were orthographically legal, so the task of Experiment 2 was more difficult than Experiment 1. Therefore, the deletion standard of accuracy rate was set at 85% for Experiment 1 and 80% for Experiment 2.

https://doi.org/10.1017/langcog.2022.40 Published online by Cambridge University Press

3.1.3. Procedure

The procedure was identical to Experiment 1.

3.2. Results

Data analyses method was the same as Experiment 1. The outlier trimming method mentioned in Experiment 1 was used, and this led to 11% of data being excluded. The model was established in the same way as Experiment 1, and we tried to keep the random effects maximal. Inverse RTs were used. Mean RTs and ACCs for items are presented in Table 7, and the parameter estimates from the linear models are presented in Table 8 with significant estimates shown in bold.

RT analysis showed the main effect of L2 AoA was significant; early acquired L2 words were identified faster than late acquired L2 words. There was no significant main effect of L1 AoA, indicating that L2 word processing was not affected by the AoA of Chinese translation equivalents. The interaction between L2 AoA and L1 AoA was not significant, indicating that the L2 AoA effect was not influenced by L1 AoA.

For the ACC, the main effect of L2 AoA was significant; early acquired L2 words were identified more accurately than late acquired L2 words. The main effect of L1 AoA and the interaction between L2 AoA and L1 AoA were not significant.

We further analyzed the combined results of Experiments 1 and 2 to check whether there was an increase in the AoA effect with the increase of semantic activation.[3] Before that, we compared the OPT scores of the two experiments and found that there was no significant difference in the English proficiency of the participants ($p > 0.10$), suggesting the results of the two experiments were

Table 7. Mean response times (RTs, ms) and accuracy rates (ACCs, %) (standard deviations)

L2 AoA	L1 AoA	RT	ACC
Early	Early	806 (257)	93 (0.25)
Early	Late	789 (246)	97 (0.18)
Late	Early	877 (269)	83 (0.38)
Late	Late	857 (262)	88 (0.33)

Table 8. Linear mixed-effects models (LMEMs) estimates of fixed effects for response times (RTs) and accuracy rates (ACCs)

	RT			ACC		
	b	*SE*	*t*	*b*	*SE*	*z*
Intercept	**−1.34**	**0.04**	**−32.41**	**3.40**	**0.40**	**8.59**
L2 AoA	**0.11**	**0.05**	**2.19***	**−1.35**	**0.51**	**−2.66****
L1 AoA	−0.03	0.05	−0.70	0.61	0.55	1.11
L2 AoA × L1 AoA	0.01	0.07	0.14	−0.02	0.74	0.03

Note: $^{**}p < 0.01$; $^{*}p < 0.05$.

[3]Because Experiments 1 and 2 may be underpowered, the comparison between the two experiments can only serve as an auxiliary illustration of the role of semantic processing depth and should not be treated as strong evidence.

https://doi.org/10.1017/langcog.2022.40 Published online by Cambridge University Press

comparable. The design was a 2 (Experiment: Experiment 1 and Experiment 2) × 2 (L2 AoA: early and late) × 2 (L1 AoA: early and late) mixed design.

In the RT analysis, we found faster RTs for Experiment 1 compared with Experiment 2, indicating that the processing depth may be shallower and less semantic processing was involved in Experiment 1. More importantly, an interaction between experiment and L2 AoA was found ($b = 0.08$, $SE = 0.03$, $t = 3.08$), such that the L2 AoA effect was larger in Experiment 2 than that in Experiment 1. For ACCs, the L2 AoA effect was similar between the two experiments ($b = -0.27$, $SE = 0.33$, $z = -0.83$), which might be because all the ACCs were high.

3.3. Discussion

The AoA effect of L2 was found, indicating that the age at which L2 words were acquired influences L2 word processing. This finding replicated previous reports of L2 AoA effects (Dirix & Duyck, 2017; Hirsh et al., 2003; Izura & Ellis, 2002, 2004; Xue et al., 2017). L1 AoA effects were not found in L2 word processing, which is consistent with Izura and Ellis (2002; Experiments 3 and 4). There was no interaction between L2 AoA and L1 AoA, indicating that the L2 AoA effect may be independent of L1 AoA.

The combined results of the two experiments showed shorter RTs and higher ACCs for Experiment 1 compared with Experiment 2. These differences could be attributed to the manipulation of fillers, indicating that this manipulation is a valid operation to change participants' processing strategy. Furthermore, the L2 AoA effect observed in RT in Experiment 2 was larger than that observed in Experiment 1, suggesting that semantic processing modulates the magnitude of the AoA effect, such that it would increase with a greater need for semantic activation.

Taken together, we obtained the L2 AoA effect at both the shallow word form level (i.e., marginally significant) and deep semantic processing level. However, given the limitation of the materials used in the orthogonal factorial design, the word length of the stimulus adopted in Experiments 1 and 2 was mainly short, with approximately an average of 5.95 letters. Therefore, a mega study with a larger number of stimuli containing more long words was adopted in Experiment 3 to investigate the relationship between the L2 AoA effect and L1 AoA comprehensively.

4. Experiment 3

In Experiment 3, a mega study and a lexical decision task with orthographically legal fillers were adopted, in which we mainly focused on whether word length influences the relationship between the L2 AoA effect and L1 AoA.

4.1. Method

4.1.1. Participants

Thirty-two late unbalanced Chinese (L1)–English (L2) bilinguals (24 females) participated in Experiment 3, none of whom had participated in the previous experiments or any rating assessments (see Table 9). Participants in Experiment 3 had the same background as in Experiment 1.

https://doi.org/10.1017/langcog.2022.40 Published online by Cambridge University Press

Table 9. The background information of the participants in Experiment 3

	Mean	*SD*	Range
Age (in years)	22.26	2.29	18–30
Years of L2 learning	13.84	2.93	9–20
L1 proficiency	5.43	0.23	5.21–5.75
L2 proficiency	3.89	0.30	3.41–4.29
OPT	38.36	3.99	30–49

4.1.2. Materials

The materials were 1,136 English words, which were selected from the database of Wang and Chen (2020). The L2 AoA of these English words (M = 13.54 years, SD = 1.31) was rated by asking participants to write the age (in years) at which they thought they learned the words (Brysbaert et al., 2014a). Word frequency (M = 52.39/million, SD = 110.02) was cited from the SUBTLEX-UK (van Heuven et al., 2014). Concreteness (M = 3.80, SD = 1.03), rated on a five-point scale (1 low to 5 high), was cited from Brysbaert et al. (2014b). To explore the potential effect of word length on the relationship between the L2 AoA effect and L1 AoA, long words were included as many as possible (45% of words had at least seven letters). The average word length was 6.51 letters (SD = 2.24). The AoA of corresponding L1 words (M = 4.39 years, SD = 0.98) was cited from Wang and Chen (2020) and the present study, using a seven-point scale.

The 1,136 fillers used in the lexical decision task were orthographically legal and pronounceable nonwords, created by changing one of the letters in real English words (e.g., “trith” for “truth”).

4.1.3. Procedure

The procedure was similar to Experiment 1. Participants were asked to make lexical decision on 2,272 words and nonwords (including 1,136 target words and 1,136 fillers). These words were presented pseudorandomly and were divided into two lists. To avoid the sequence effect, the order of the two lists was counterbalanced among the participants. There was a rest about every 15 min, and participants were free to decide how much time they needed to rest. The whole experiment lasted about 130 min.

4.1.4. Data analysis

Data were analyzed using LMEMs. The models contained many fixed factors, including L2 AoA, L1 AoA, word length, the interaction of L2 AoA × L1 AoA × word length, word frequency, familiarity, concreteness, OPT, all of which were continuous. Models were established in the same way as Experiment 1. We started with the full model, and then built the model without the correlation between random intercepts and random slopes as well as the model with slopes only. Since no models converged, we further simplified random-effects structure. Because we focused on the relationship among L2 AoA, L1 AoA, and word length, we tried to retain the random slopes for these three variables (Karimi & Diaz, 2020). Prior to analysis, all continuous variables were centered to reduce collinearity. RTs were inversely transformed as in Experiment 1. The degree of multicollinearity was measured using the Variance Inflation Factor (VIF), which should not exceed 5 (Fox & Weisberg, 2010).

Table 10. Mean response times (RTs, ms) and accuracy rates (ACCs, %) of the data, which were split using median L2 AoA and L1 AoA (standard deviation)

	AoA	RT	ACC
L2	Early	716 (198)	95 (0.23)
	Late	810 (230)	87 (0.33)
L1	Early	743 (216)	92 (0.27)
	Late	793 (227)	90 (0.31)

4.2. Results

The same outlier rejection was used as in Experiment 1, and 11% of data was excluded. We split the data using median L2 AoA (early 9.55–13.67 and late 13.68–17.00) and L1 AoA (early 1.33–4.43 and late 4.50–6.89) to show the average effects of L2 AoA and L1 AoA. The mean RTs and ACCs are presented in Table 10. The results of the linear models are presented in Table 11 with significant estimates shown in bold.

For RTs, the maximum VIF was 3.40. We found no main effect of L2 AoA. The L1 AoA effect was marginally significant, such that L2 words were processed marginally faster, when their L1 translation equivalents were acquired early. The effects of word length, word frequency, familiarity, and concreteness were significant. Words tended to be processed faster when they were shorter, more frequent, more familiar, and evoked concrete experiences. The effect of OPT was not significant, indicating that the RT was not influenced by L2 proficiency level. More importantly, the three-way interaction among L2 AoA, L1 AoA, and length was significant (see Fig. 1A). The interaction between L2 AoA and L1 AoA was not observed for short words (shorter than nine letters; $b = 0.01$, $SE = 0.004$, $t = 1.49$). For longer words (at least 10 letters), the interaction between the two AoAs was significant ($b = -0.06$, $SE = 0.02$, $t = -3.46$); the L2 AoA effect was significant when L1 AoA was early ($b = 0.01$, $SE = 0.003$, $t = 2.19$) and this effect disappeared when L1 AoA was late ($b = 0.001$, $SE = 0.002$, $t = 0.46$).

Table 11. Linear mixed-effects models (LMEMs) estimates of fixed effects for response times (RTs) and accuracy rates (ACCs) in Experiment 3

	RT			ACC		
	b	*SE*	*t*	*b*	*SE*	*z*
Intercept	**−1.39**	**0.02**	**−61.63*****	**3.15**	**0.10**	**32.90*****
L2 AoA	0.01	0.01	1.34	−0.01	0.08	−0.12
L1 AoA	0.01	0.01	1.91^{+}	−0.08	0.06	−1.29
Length	**0.05**	**0.002**	**20.42*****	0.003	0.03	0.10
Frequency	**0**	**0**	**−3.73*****	**0.002**	**0.001**	**2.49***
Familiarity	**−0.09**	**0.01**	**−8.82*****	**1.13**	**0.11**	**10.73*****
Concreteness	**0.01**	**0.01**	**2.82****	**−0.11**	**0.06**	**−1.96***
OPT	−0.004	0.01	−0.69	0.01	0.02	0.43
L2 AoA × L1 AoA	−0.002	0.003	−0.56	0.03	0.04	0.89
L2 AoA × length	−0.003	0.002	−1.71^{+}	−0.004	0.02	−0.20
L1 AoA × length	−0.01	0.003	−1.82^{+}	0.04	0.03	1.38
L2 AoA × L1 AoA × length	**−0.01**	**0.002**	**−4.04*****	**0.05**	**0.02**	**2.88****

Note: $^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

https://doi.org/10.1017/langcog.2022.40 Published online by Cambridge University Press

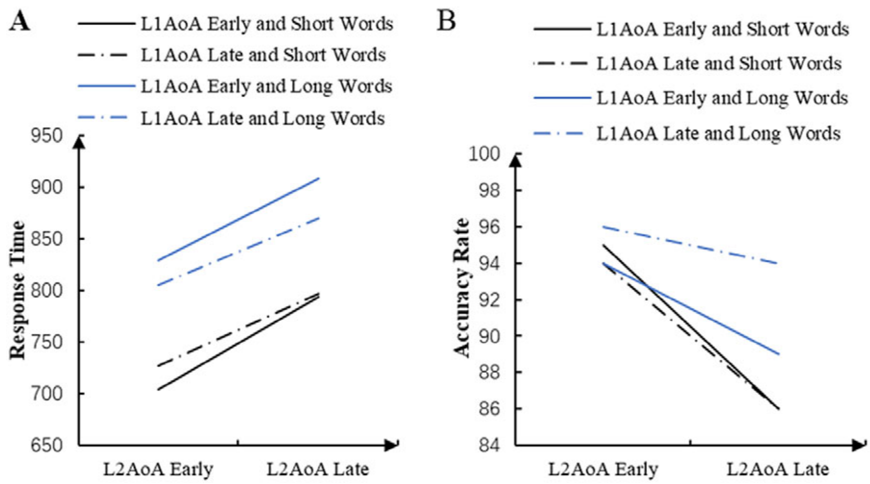

Fig. 1. The interactions among L2 AoA, L1 AoA, and length in Response Time (Panel A) and Accuracy Rate analyses (Panel B).

For ACCs, the maximum VIF was 4.30. The main effects of L2 AoA and L1 AoA were not significant. The effects of word frequency, familiarity, and concreteness were significant, that is, the ACC was higher when the words were more frequent, more familiar, and evoked concrete experiences. The effects of word length and OPT were not significant, indicating that the ACC was not influenced by word length and L2 proficiency level. More importantly, the three-way interaction among L2 AoA, L1 AoA, and length was significant (see Fig. 1B). Specifically, the interaction between L2 AoA and L1 AoA was not significant when word length was shorter (shorter than 10 letters; $b = -0.01$, $SE = 0.03$, $z = -0.21$). For longer words (at least 10 letters), the interaction between the two AoAs was significant ($b = 0.87$, $SE = 0.24$, $z = 3.63$); the L2 AoA effect was significant when L1 AoA was early ($b = -1.21$, $SE = 0.42$, $z = -2.91$) and this effect disappeared when L1 AoA was late ($b = 0.32$, $SE = 0.20$, $z = 1.57$).

4.3. Discussion

In Experiment 3, we explored the existence of the L2 AoA effect and its relationship with L1 AoA using a mega study. By comparing the results with Experiment 2, we aimed to find out whether word length modulates the effect of L1 AoA on L2 AoA.

The results showed a three-way interaction among L2 AoA, L1 AoA, and word length: for shorter words, the L2 AoA effect was significant and this effect was not influenced by L1 AoA; for longer words, the influence of L1 AoA on the L2 AoA effect arised only when words were learned early in L1. This three-way interaction, to our knowledge, has not been reported so far. We will discuss it in more detail in Section 5.

5. General discussion

The present study aimed to explore the L2 AoA effect and its relationship with L1 AoA. Three experiments were conducted at the word form and semantic processing levels. In order to detect the word form processing level, Experiment 1 adopted a lexical decision task in which filler words were consonant strings. We found that early acquired L2 words were processed marginally more accurately than late acquired L2 words. In Experiment 2, filler words in the lexical decision task were created by changing one of the letters in real English words, to explore the semantic processing level. The results showed that early acquired L2 words were processed more accurately and faster than late acquired L2 words. The L1 AoA effect and the interaction of L1 AoA and L2 AoA were not found in either experiment. The combined analysis of Experiments 1 and 2 showed that the L2 AoA effect observed in RTs was larger in Experiment 2 than that in Experiment 1, indicating that the magnitude of the L2 AoA effect increased with a greater need for semantic activation. Furthermore, in Experiment 3, which used a larger number of stimuli, we observed the L2 AoA effect again. Importantly, an interaction between L2 AoA and L1 AoA was found for longer words, that is, there was an L2 AoA effect when L1 AoA was early, but the L2 AoA effect disappeared when L1 AoA was late.

5.1. The L2 AoA effect and its relationship with L1 AoA

The L2 AoA effect was not influenced by L1 AoA in either Experiment 1 or Experiment 2. These results are in line with previous investigations on L2 AoA, which used factorial design and did not find the influence of L1 AoA on the L2 AoA effect either (Izura & Ellis, 2002, 2004). In Experiment 3, we observed for the first time the interaction of L2 AoA and L1 AoA with the mediation of word length. Specifically, for long L2 words, when their L1 equivalents were acquired early, earlier learned L2 words had processing advantages. However, for short L2 words, the L2 AoA effect was independent of the age at which their L1 equivalents were acquired.

We explained the three-way interaction among L2 AoA, L1 AoA, and word length by taking the characteristics of bilingual lexical representation and access into consideration. According to the Revised Hierarchical Model, bilinguals' two languages are separated at the lexical level and shared at the conceptual level (Kroll & Stewart, 1994). Skilled bilinguals could activate L2 semantic representations directly, while unskilled bilinguals may activate L2 semantic representations via the L2–L1 translation equivalents' link. Although the participants in the present study were unskilled bilinguals, they had started to learn English at school since 7–9 years old, and they had been learning English for a total of 12 years on average in a classroom setting. Therefore, when processing short and familiar words, participants performed similarly to highly proficient bilinguals, who might have direct access to the word's meaning. Perhaps this is why the L2 AoA effect was observed in short familiar L2 words. However, for long L2 words, due to larger processing difficulty, participants may rely more on their L1 translation equivalents to activate semantic representations. If this is the case, there should be no L2 AoA effect. But in fact, there was still an L2 AoA effect when L1 translation equivalents were acquired early. One possible reason is that bilinguals could directly link L2 word forms to conceptual representations for L1 early acquired long words, showing L2 AoA effects. However, for L1 late acquired long words, the conceptual representations of L2 words may be accessed via the L2–L1 link (i.e., through L1 translation equivalents), hence no L2 AoA effect showed up.

Dirix and Duyck (2017) found that L2 AoA effects appeared at early lexical access and late semantic access processing levels in L2 reading. The cross-lingual AoA effect, reflected by the reading times of longer L2 words being affected by the AoA of their L1 translational equivalents, only appeared in measures reflecting

https://doi.org/10.1017/langcog.2022.40 Published online by Cambridge University Press

early lexical access (i.e., single/first fixation and gaze duration), but not in late semantic access phase (i.e., total reading time). Here, if the total reading time reflects semantic access, it means cross-lingual effect would not be found at the semantic processing level. However, text reading involves complex processes such as word recognition, sentence comprehension, and text integration. So, what does the index of total reading time truly reflect? Whether it merely reflects semantic access requires further investigation.

In sum, the present study found a weak L2 AoA effect at the word form processing level, which was independent of L1 AoA. At the semantic processing level, for longer words, the L2 AoA effect was influenced by L1 AoA. That is, early acquired L2 words had processing advantage when their L1 translation equivalents were also acquired early.

5.2. Theoretical implications for understanding the L2 AoA effect and its relationship with L1 AoA

There are two predominant theories, Semantic Hypothesis and Arbitrary Mapping Hypothesis, seeking to explain the mechanisms of AoA effects. The Semantic Hypothesis proposes that the AoA effect primarily originates from the semantic processing system. The semantic activation of early acquired words is faster than late acquired words (Brysbaert et al., 2000; van Loon-Vervoorn, 1989). The Arbitrary Mapping Hypothesis holds that the AoA effect depends on the mapping relationship between input (e.g., orthography) and output (e.g., semantics) representations established in the lexical network. Early acquired words have a processing advantage because of the plasticity of the network. With the loss of network plasticity, it is difficult for late acquired words to change the connection weights between input and output representations (Ellis & Lambon Ralph, 2000; Monaghan & Ellis, 2002).

According to the Semantic Hypothesis as well as the viewpoint of the Revised Hierarchical Model (see the "Introduction" section for detailed information; Kroll & Stewart, 1994), the L2 AoA effect should be modulated by L1 AoA in the semantic processing task. Our results showed that for L2 long words, the L2 AoA effect emerged only under the condition of early L1 AoA. This finding supports the Semantic Hypothesis. However, we observed a weak L2 AoA effect in the task biased toward orthographical processing and an independent L2 AoA effect for L2 short words, which could not be explained by the Semantic Hypothesis.

According to the Arbitrary Mapping Hypothesis, L2 AoA effects should occur in tasks biased toward orthographical and semantic processing, and L2 AoA effects should not be affected by L1 AoA. Our results showed a weak L2 AoA effect in the task biased toward orthographical processing, and an independent L2 AoA effect for short words, supporting the Arbitrary Mapping Hypothesis. However, for L2 long words, we found that the L2 AoA effect was modulated by L1 AoA, which could not be explained by the Arbitrary Mapping Hypothesis. To sum up, our findings that the L2 AoA effect was independent of L1 AoA in some cases, but dependent on L1 AoA in other cases, could not be explained fully by a single theory.

We do not believe that there are totally different mechanisms for L1 and L2 AoA effects. For both L1 and L2, the AoA effect can be determined by the order in which words enter the lexical network. However, the L2 AoA effect may be somewhat different from the L1 AoA effect, since L2 interacts with the L1 knowledge of bilinguals, which could influence L2 word recognition processes. So, we tried to combine the Arbitrary Mapping Hypothesis and the Revised Hierarchical Model to explain our results.

First, the Arbitrary Mapping Hypothesis predicts that the new learned order of L2 words determines the L2 AoA effect, similar to the L1 AoA effect. Therefore, the L2 AoA effect exists in the task biased toward orthographical processing, and it is not affected by L1 AoA. Second, the Arbitrary Mapping Hypothesis holds that early acquired L2 words have a processing advantage at the semantic level because of the early establishment of form-meaning mappings. This could account for the significant L2 AoA effect in the task biased toward semantic processing in Experiment 2. However, the Arbitrary Mapping Hypothesis could not explain the results of Experiment 3, that is, the L2 AoA effect was not affected by L1 AoA in short L2 words but was affected by L1 AoA in long L2 words. To better understand the relationship between the L2 AoA effect and L1 AoA at the semantic level, we suggest to take the viewpoints of bilingual lexical representation and accessibility into consideration.

According to the Revised Hierarchical Model, bilingual words share conceptual representations (Kroll & Stewart, 1994). When processing L2 words, skilled bilinguals can activate semantic representations directly through the link of L2 words and concepts, while unskilled bilinguals need to rely on the link of L1 words and concepts. The bilinguals in the present study were unskilled bilinguals, whose L2 proficiency level was lower than that of their L1, so L2 words may be parasitic on L1 words. This provides the premise that L1 AoA may influence the L2 AoA effect.

In addition to the proficiency level of L2 mentioned above, other factors may also influence the semantic access of L2 words, for example, L2 word familiarity. For example, proficient bilinguals are more likely to activate the semantic representations of familiar L2 words directly. Taken together, our results could be accounted for like this: When L2 words are short and familiar, unskilled Chinese–English bilinguals (i.e., participants in the present study) could process these words directly through the link of L2 word forms and conceptual representations, similar to highly proficient bilinguals, so there would be no influence of L1 AoA on the L2 AoA effect. However, for long L2 words, due to their processing difficulty, participants may rely more on their L1 translation equivalents to activate semantic representations, so there would be the L1 AoA effect only. But in fact, we still found an L2 AoA effect when L1 translation equivalents are acquired early. The possible reason for this might be that bilinguals could build a direct link between L2 word forms and their conceptual representations if their L1 translation equivalents are learned earlier. However, for late acquired L1 words, the conceptual representations of L2 words rely on the mediation of L1 translation equivalents, leading to the absence of the L2 AoA effect.

Based on the above reasoning, we hypothesize that in L2 word processing, L1 AoA may play a role only when L2 semantic access requires the help of L1 translation equivalents or L2 word activates the corresponding L1 translation words. This occurs when L2 words are long or difficult to process. Izura and Ellis (2002, 2004) found that the L2 AoA effect is independent of L1 AoA, which may be because their participants were highly proficient and/or short experimental words were easy to process, so the meaning of L2 words could be directly accessed without the mediation of L1, and hence there would be no influence from L1 AoA.

In sum, combining the Arbitrary Mapping Hypothesis with the characteristics of L2 lexical representation and accessibility can better explain the L2 AoA effect and its relationship with L1 AoA. Nevertheless, our theoretical explanations still need further examination, since there have been few studies on the L2 AoA effect and more evidence are needed.

5.3. Guidelines for future studies and limitations of the present study

As we know, compared with L1 words, the representation and accessibility of L2 words are different (Kroll et al., 2015; Kroll & Stewart, 1994). The Revised Hierarchical Model proposes that the two languages are represented separately at the lexical level but share conceptual representation. Therefore, the L2 AoA effect may be independent of L1 AoA at word form level, especially for Chinese–English bilinguals, whose two languages have different orthographies. Here, it is important to note that the findings for Chinese–English bilinguals may not apply to bilinguals whose two languages belong to the same language family, for example, Spanish–English bilinguals. According to the Bilingual Interactive Activation Model (BIA), for two languages with similar orthographies, word forms are represented in a common lexicon (Dijkstra & van Heuven, 1998). On this occasion, the influence of L1 AoA on the L2 AoA effect may occur at the word form level. So, the influence of similarity (or language distance) between the two languages on the L2 AoA effect and its relationship with L1 AoA can be further studied in the future.

According to the Revised Hierarchical Model, L2 proficiency affects the mechanism of L2 semantic access. Skilled bilinguals could directly access semantic representations, while unskilled bilinguals activate semantic representations with the help of L1 translation equivalents. It is possible that highly proficient participants directly access L2 words, so the L2 AoA effect is independent of L1 AoA at the semantic processing level. However, for unskilled bilinguals, the L2 AoA effect may be influenced by L1 AoA to a greater extent. The L2 AoA effect and its relationship with L1 AoA in bilinguals at different proficiency levels can be further explored in subsequent studies.

Moreover, our study found that the relationship between the L2 AoA effect and L1 AoA is influenced by lexical characteristics. For example, for short L2 words, bilinguals can directly access the semantic representations, so no influence of L1 AoA on the L2 AoA effect is found. We can further investigate how L2 lexical characteristics, such as word length, frequency, concreteness, and other factors, affect the L2 AoA effect and its relationship with L1 AoA in future studies.

The present study has some limitations. First, in Experiment 1, the L2 AoA effect is marginally significant, so it is difficult to determine the role of L2 AoA at the word form level. However, Juhasz and Sheridan (2020) found that the AoA effect emerged as early as 158 ms, using a survival analysis. Adorni et al. (2013) found a role of AoA by using a typical orthographic detection task. Combined with these studies, we believe that the AoA effect may emerge at the word form processing level, and survival analysis can be used to further investigate the time course of the L2 AoA effect at this level.

Second, we manipulated the type of filler words in the lexical decision task to investigate the L2 AoA effect at the word form and semantic processing levels in Experiments 1 and 2. Adopting the same task can avoid interferences caused by different tasks. However, semantic and orthographic processing are not completely mutually exclusive. They are a relative state and we try to use a comparative (or subtractive) way to illustrate that the AoA effect is smaller when orthographic processing is dominant, compared with a dominant semantic processing condition. In future research, the event-related potentials (ERPs) technique can be used to further explore L2 AoA effects at different processing levels. This technique has high temporal resolution (milliseconds) and ERPs components could reflect specific cognitive processes. For example, N400 is a typical component reflecting semantic processing (Kutas & Hillyard, 1980).

Third, Kuperman et al. (2013) and Dirix et al. (2019) found low correspondence between reaction times and reading times of various paradigms. This raises the question whether these tasks tap into the same underlying cognitive processes, and also points out the importance of converging evidence from multiple sources. Furthermore, Balota et al. (2004) summarize some important potential pitfalls in applying small-scaled experiments that orthogonally manipulate variables and arbitrarily categorize variables in "high" and "low" categories, and point toward the advantage of "big data" studies which allow to study variables continuously. Our study used both a factorial design and a mega study and obtained different results, indicating the importance of applying multiple experimental paradigms to collect evidence from multiple sources.

6. Conclusion

In the present study, we combined a factorial design and a mega material corpus to address the relationship between the L2 AoA effect and L1 AoA. The results indicate that L2 AoA has a weak effect at the orthographical processing level, but a stable effect at the semantic processing level. The AoA of L1 translation equivalents modulates the L2 AoA effect at the semantic processing level, which is more likely to appear in long words. Theoretically, our results suggest that it is important to take both the Arbitrary Mapping Hypothesis and bilingual representation and activation characteristics into consideration in order to explain the L2 AoA effect.

Data availability statement. The materials, data, and analysis scripts are available through the Open Science Framework (https://osf.io/fr9k7/). None of the experiments were preregistered.

Acknowledgments. We thank Nicolas Dirix and one anonymous reviewer for their helpful comments.

Funding. This work was supported by funding of National Natural Science Foundation of China (31970976) for B.C.

Competing interests. There are no competing interests to declare.

References

Adorni, R., Manfredi, M., & Proverbio, A. M. (2013). Since when or how often? Dissociating the roles of age of acquisition (AoA) and lexical frequency in early visual word processing. *Brain and Language*, 124(1), 132–141. https://doi.org/10.1016/j.bandl.2012.11.005

Allan, L. (2004). *The Oxford placement test*. Oxford University Press.

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience*, 31(7), 940–966. https://doi.org/10.1080/23273798.2016.1171366

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, 92, 265–280. https://doi.org/10.1016/j.jml.2016.07.004

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. https://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. https://doi.org/10.1037/0096-3445.133.2.283

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R Package Version*, *1*(7), 1–23. https://CRAN.R-project.org/package=lme4

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, 4(1), 1–19. https://doi.org/10.1177/2515245920960351

Brysbaert, M., & Ellis, A. W. (2016). Aphasia and age of acquisition: Are early-learned words more resilient? *Aphasiology*, 30(11), 1240–1263. https://doi.org/10.1080/02687038.2015.1106439

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1(1), 1–20. https://doi.org/10.5334/joc.10

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

Brysbaert, M., Van Wijnendaele, I., & De Deyne, S. (2000). Age-of-acquisition effects in semantic processing tasks. *Acta Psychologica*, 104(2), 215–226. https://doi.org/10.1016/S0001-6918(00)00021-4

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014b). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46, 904–911. https://doi.org/10.3758/s13428-013-0403-5

Catling, J. C., & Elsherif, M. M. (2020). The hunt for the age of acquisition effect: It's in the links! *Acta Psychologica*, 209(1), 103138. https://doi.org/10.1016/j.actpsy.2020.103138

Chang, Y. N., & Lee, C. Y. (2020). Age of acquisition effects on traditional Chinese character naming and lexical decision. *Psychonomic Bulletin & Review*, 27(2), 1317–1324. https://doi.org/10.3758/s13423-020-01787-8

Cortese, M. J., Yates, M., Schock, J., & Vilks, L. (2018). Examining word processing via a megastudy of conditional reading aloud. *Quarterly Journal of Experimental Psychology*, 71(11), 2295–2313. https://doi.org/10.1177/1747021817741269

Cui, L., Wang, J., Zhang, Y. L., Cong, F. J., Zhang, W. X., & Hyönä, J. (2021). Compound word frequency modifies the effect of character frequency in reading Chinese. *Quarterly Journal of Experimental Psychology*, 74(4), 610–633. https://doi.org/10.1177/1747021820973661

Dijkstra, A., & 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). Lawrence Erlbaum Associates.

Ding, G. S., Perry, C., Peng, D. L., Ma, L., Li, D. J., Xu, S. Y., Luo, Q., Xu, D., & Yang, J. (2003). Neural mechanisms underlying semantic and orthographic processing in Chinese–English bilinguals. *Neuroreport*, 14(12), 1557–1562. https://doi.org/10.1097/01.wnr.0000087906.78892.8e

Dirix, N., Beken, H. V., Bruyne, E. D., Brysbaert, M., & Duyck, W. (2019). Reading text when studying in a second language: An eye-tracking study. *Reading Research Quarterly*, 55(3), 371–397. https://doi.org/10.1002/rrq.277

Dirix, N., & Duyck, W. (2017). The first- and second-language age of acquisition effect in first- and second-language book reading. *Journal of Memory and Language*, 97, 103–120. https://doi.org/10.1016/j.jml.2017.07.012

Ellis, A. W. (2012). The acquisition, retention, and loss of vocabulary in aphasia, dementia, and other neuropsychological conditions. In M. Faust (Ed.), *The handbook of the neuropsychology of language* (pp. 637–660). Blackwell Publishing Ltd.

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. https://doi.org/10.1037/0278-7393.26.5.1103

Evans, G. A., Ralph, M. A. L., & Woollams, A. M. (2012). What's in a word? A parametric study of semantic influences on visual word recognition. *Psychonomic Bulletin & Review*, 19(2), 325–331. https://doi.org/10.3758/s13423-011-0213-7

Fox, J., & Weisberg, S. (2010). *An R companion to applied regression*. SAGE.

Gilhooly, K. J., & Gilhooly, M. L. (1979). Age-of-acquisition effects in lexical and episodic memory tasks. *Memory & Cognition*, 7(3), 214–223. https://doi.org/10.3758/BF03197541

González-Nosti, M., Barbón, A., Rodríguez-Ferreiro, J., & Cuetos, F. (2013). Effects of the psycholinguistic variables on the lexical decision task in Spanish: A study with 2,765 words. *Behavior Research Methods*, 46 (2), 517–525. https://doi.org/10.3758/s13428-013-0383-5

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin*, 133(4), 638–650. https://doi.org/10.1037/0033-2909.133.4.638

Hirsh, K. W., Morrison, C. M., Gaset, S., & Carnicer, E. (2003). Age of acquisition and speech production in L2. *Bilingualism: Language and Cognition*, 6(2), 117–128. https://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(2), 245–281.

Izura, C., & Ellis, A. W. (2004). Age of acquisition effects in translation judgement tasks. *Journal of Memory and Language*, 50(2), 165–181. https://doi.org/10.1016/j.jml.2003.09.004

Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition*, 13(7), 789–845. https://doi.org/10.1080/13506280544000066

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

Juhasz, B. J., & Sheridan, H. (2020). The time course of age-of-acquisition effects on eye movements during reading: Evidence from survival analyses. *Memory & Cognition*, 48(5), 83–95. https://doi.org/10.3758/s13421-019-00963-z

Karimi, H., & Diaz, M. (2020). When phonological neighborhood density both facilitates and impedes: Age of acquisition and name agreement interact with phonological neighborhood during word production. *Memory and Cognition*, 48, 1061–1072. https://doi.org/10.3758/s13421-020-01042-4

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

Kroll, J. F., Dussias, P. E., Bice, K., & Perrotti, L. (2015). Bilingualism, mind, and brain. *Annual Review of Linguistics*, 1(1), 377–394. https://doi.org/10.1146/annurev-linguist-030514-124937

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory & Language*, 33 (2), 149–174. https://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. *Quarterly Journal of Experimental Psychology*, 66(3), 563–580. https://doi.org/10.1080/17470218.2012.658820

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203–205. https://doi.org/10.1126/science.7350657

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

Monaghan, J., & Ellis, A. W. (2002). What exactly interacts with Spelling—Sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory & Cognition*, 28(1), 183–206. https://doi.org/10.1037/0278-7393.28.1.183

Morey, R. D., Rouder, J. N., & Jamil, T. (2015). *BayesFactor: Computation of Bayes factors for common designs* (R package version 0.9.12-2). https://cran.r-project.org/package=BayesFactor.

Rochford, G., & Williams, M. (1962). Studies in the development and breakdown of the use of names. *Journal of Neurology Neurosurgery & Psychiatry*, 28(5), 407–413. https://doi.org/10.1136/jnnp.28.5.407

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, 67(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

van Loon-Vervoorn, W. A. (1989). *Eigenschappen van basiswoorden*. Swets & Zeitlinger.

Volkovyskaya, E., Raman, I., & Baluch, B. (2017). Age of acquisition (AoA) effect in monolingual Russian and bilingual Russian (L1)-English (L2) speakers in a free recall task. *Writing Systems Research*, 9(2), 148–163. https://doi.org/10.1080/17586801.2017.1405136

Wang, J., & Chen, B. G. (2020). A database of Chinese-English bilingual speakers: Ratings of the age of acquisition and familiarity. *Frontiers in Psychology*, 11, 554785. https://doi.org/10.3389/fpsyg.2020.554785
Xu, X., Li, J., & Guo, S. (2020). Age of acquisition ratings for 19,716 simplified Chinese words. *Behavior Research Methods*, 53(2), 558–573. https://doi.org/10.3758/s13428-020-01455-8
Xue, J., Liu, T., Marmolejo-Ramos, F., & Pei, X. (2017). Age of acquisition effects on word processing for Chinese native learners' English: ERP evidence for the Arbitrary Mapping Hypothesis. *Frontiers in Psychology*, 8, 1–14. https://doi.org/10.3389/fpsyg.2017.00818

Cite this article: Wang, J., Liang, L. & Chen, B. (2023). The age of acquisition effect in processing second language words and its relationship with the age of acquisition of the first language *Language and Cognition* 15: 355–376. https://doi.org/10.1017/langcog.2022.40