# Effect of age of first exposure on L2 contextual lexical semantic learning: an ERP investigation

Shuang Xu[1] (iD), Hailing Wang[2], Shouxin Li[2] and Guang Ouyang[1]

[1]Faculty of Education, University of Hong Kong, Hong Kong, China and [2]School of Psychology, Shandong Normal University, Jinan, China
**Corresponding authors:** Shouxin Li and Guang Ouyang; Emails: shouxinli@sdnu.edu.cn; ouyangg@hku.hk

**Abstract**

Age of first exposure (AoFE) is an important factor that influences the quality of L2 acquisition. This study aims to investigate the AoFE effect on the contextual learning of L2 novel words at the neural level, as measured by the N400 component from event-related potentials (ERPs). Eighty-eight participants were recruited for the experiment of L2 pseudo-word learning, which includes a learning session and a testing session. The participants' EEG data were recorded from the testing session, and the N400 effect was derived from target words that were either congruous or incongruous with the context. The linear mixed model and multiple regression model revealed a positive AoFE effect on the N400 effect in discourses that were designed for testing retrieval of episodic and semantic memory even after accounting for the variance contributed by several confounding factors. In addition to AoFE, the effects of total L2 exposure, L2 proficiency and personality on the L2 novel word learning performance indicated by the N400 effect were also confirmed in the statistical results.

## 1. Introduction

### 1.1. AoFE effect in different linguistic domains

Under the trend of globalization, the importance of L2 acquisition has been recognized in social and educational domains, which has led to increasing attention in scientific research, especially in the psycho-linguistic area (Larsen-Freeman & Long, 2014). Driven by its importance, there is increasing interest in identifying the main factors that contribute to the learning outcome of L2 acquisition (Butler & Hakuta, 2006; Hurtado & Estrada, 2010). For each individual, the outcome of L2 acquisition must be determined by many factors – two primary factors are the amount of total exposure (AoTE) to L2 and the age of first exposure (AoFE) to L2 (Saville-Troike &

Barto, 2017). AoTE simply refers to the total amount of time a learner is exposed to or engaged in L2 learning, which should certainly be correlated with the learners' learning outcome. AoFE in L2 learning, also termed age of acquisition in some studies (e.g., Ojima et al., 2011), refers to the age at which L2 begins to be learned in a systematic, sustained and usually intensive way, such as in school teaching (Kovelman et al., 2008). The AoFE effect thus refers to how AoFE influences L2 learners' learning performance in a way that is not accounted for by total L2 exposure (i.e., AoTE). This issue is more commonly discussed with a linkage to the concept of the critical period effect in L2 acquisition, which states that there exists a critical period in the course of life development after which new L2 learners would not be able to catch up with early learners (Hartshorne et al., 2018). Until now, the critical period effect in L2 acquisition is still a controversial but attention-grabbing topic (Lucio, 2020). One of the major complications in the critical period effect is that the effects are likely to differ across different linguistic domains, such as phonology/phonetics, syntax and semantics (Singleton & Ryan, 2004). Some previous behavioral and neurophysiological research works have provided evidence for the existence of critical periods of acquiring phonetics and grammar or syntax patterns of L2 (Dollmann et al., 2020; Hartshorne et al., 2018; Kotz, 2009; Marinova-Todd et al., 2003), whereas there is also counter-evidence against the critical period hypothesis (CPH). For example, Flege (2019) raised an alternative explanation of the critical period effect as the differential quality and quantity of L2 input to learners at different ages. Compared to other aspects of linguistic domains, there is little research on the effect of AoFE on L2 semantics. Slabakova (2006) reviewed and discussed the previous findings on L2 semantic processing assessed with neuroimaging and electrophysiology methods and concluded that the AoFE effect on semantics is mainly manifested in the syntax and phrasal semantics. Liu et al. (2017) found that within the semantic modules, compared with late bilingual learners, early bilingual learners are endowed with advantages that are manifested in significantly more robust functional connectivity in the neural system.

## 1.2. AoFE effect in lexical semantic network

Lexical semantics, one of the most essential parts of semantics, mainly concerns the meaning of words. Kasparian and Steinhauer (2016) stated that AoFE may not be a crucial factor affecting the acquisition of lexical semantics based on the neural evidence that L2 learners' ERP responses to semantically incorrect stimuli were not modulated by AoFE but by language proficiency. However, Kotz and Elston-Güttler (2007) reviewed a series of relevant studies and claimed that the significant neurocognitive differences in bilingual's lexical-semantic processing are modulated by AoFE, which supports the existence of the AoFE effect in the domain of lexical semantics. Newman et al. (2012) also found that, compared to native speakers, late L2 learners showed weakened and deferred neurocognitive responses to lexical-semantic violations in reading. In sum, the research findings on the effect of AoFE on L2 lexical semantics are still mixed, which is still inadequate for making solid conclusions.

The semantics behind lexicons are 'concepts of words', and words are conceptually connected. As such, the essential process underlying the acquisition of novel concepts is to integrate the novel concepts into the existing semantic network. As

Steyvers and Tenenbaum (2005) proposed, based on the expanding process of semantic network development, the earlier the semantic nodes are integrated into the semantic network, the richer connections these semantic nodes can form in the semantic network and the more robust and abundant the semantic network would be in adulthood. If the semantic network is better formed and connected. It would be easier for a new and unfamiliar semantic node to connect with the existing semantic nodes in the established semantic network. Thus, an early formed semantic network would be more efficient in hosting of the process of semantic acquisition than a late-formed one.

In terms of connection formation, the relational associations can be either theme-based relations or category-based relations. Thematic relation refers to the linkage between concepts that co-occur in the same situations (Lin & Murphy, 2001), or concepts that are clearly attached to certain roles, events or situations (e.g., Goldwater et al., 2011; Jones & Love, 2007; Markman & Stilwell, 2001). Moreover, in thematic relations, there usually exist certain semantic associations among concepts beyond the mere co-occurrence of the words (Yee et al., 2009). For example, 'cheese' and 'bird' are usually not in a thematic relation, whereas 'spider' and 'web' are in a thematic relation. Another relation in the semantic network is the taxonomic relation, which is based on the classification of concepts into categories with similar semantic features (Rogers et al., 2004). For example, 'sparrow' is a subcategory of 'bird' since sparrow and bird share similar features, so they form a taxonomic relation. Taxonomic relations are multilevel, and they form a huge hierarchical structure in the semantic network (Nelson, 1988). Taken together, the acquisition of L2 words can be seen as the process of building up thematic and taxonomic relations, which should be largely determined by the ability (particularly neural plasticity) of the neural network system to form associated neural substrates. Based on the assumption that neural plasticity is a function of age (Arcos-Burgos et al., 2019; Burke & Barnes, 2006), we expect that AoFE would have a significant effect on the L2 learning outcome that is not accounted for by AoTE.

## 1.3. AoFE effect in different memory systems

From the perspective of the memory system, the establishment of taxonomic and thematic relations in the semantic network reflects the process of novel concepts entering the semantic memory (Mirman et al., 2017). Semantic memory refers to the memory of established facts and general knowledge about the world (Kumar, 2021; Yee et al., 2014), as well as the newly established representations (thematic or taxonomic) that can be detached from the original contexts and generalized to new ones (Baddeley et al., 2009). In contrast to semantic memory, episodic memory is another category of long-term memory (Tulving, 1972), which mainly involves the recurrence of original contexts (Squire & Zola, 1998). Although a new word is defined as a new piece of information in the semantic system, the specific episodic memory in which the new word was learnt may also contribute to (or facilitate) the absorption of the new word into the semantic system. For example, when an individual recalls a specific episodic memory that contains learning a new word, the new word is rehearsed. Therefore, language learning implicates both encoding and retrieval of episodic memory. It was found that early bilingual experiences may shape bilingual episodic memory (Schroeder & Marian, 2014). Previous research on bilingualism has

shown that bilinguals who learned L2 at a very early age did perform better on memory tasks probing not only semantic memory but also episodic memory as compared to monolinguals (Kormi-Nouri et al., 2003, 2008; Schroeder & Marian, 2012). Based on these rationales, both the processes of episodic and semantic memory formation should be examined when studying the association between L2 learning and AoFE.

### 1.4. Other confounding factors of the AoFE effect

For the purpose of examining how lexical-semantic learning may be influenced by AoFE, other confounding factors, such as the AoTE, language proficiency, working memory capacity, personality, socioeconomic status (SES) and gender, should be considered. First, the AoTE is a straightforward factor that will undoubtedly confound the effect of AoFE (Ojima et al., 2011). Second, language proficiency is one of the key indicators of the strength of links among different semantic codes in the semantic network (Kotz & Elston-Güttler, 2004) and should be a predictor of the performance of new word learning under contextual learning. It is thus necessary to include it as an independent variable in the present study. Third, working memory capacity refers to the ability to process and store information for instantaneous use, emphasizing the capacity of both short-term memory storage and processing (Cowan, 2012). Working memory capacity has been proven to be associated with many cognitive abilities in L2 learning domains (Juffs & Harrington, 2011; Wen et al., 2019). Compared with L1, the working memory capacity measured in L2 is more closely related to L2 learning and reading-related performance (Alptekin & Erçetin, 2010; Harrington & Sawyer, 1992). Fourth, personality (e.g., introversion or extroversion) is another important factor that may influence L2 learning in addition to other factors (Kezwer, 1987), and extroverted character facilitates language education compared with introverted character. Fifth, SES is also one of the important environmental factors influencing our brain development and language acquisition (Brito & Noble, 2014; Raizada & Kishiyama, 2010; Thomas et al., 2013), and it is a construct of multiple dimensions that include not only measures of social attributes such as hierarchies, reputation and power but also economic conditions (Bateman, 2014; Hackman & Farah, 2009). Therefore, the assessment of participants' SES needs multiple indicators, including income, education and occupation (Oakes & Andrade, 2017). Finally, in the language learning domain, the influence of gender on language acquisition has always been one of the major interests of neurolinguists (Zoghi et al., 2013). It has been found that compared with males, females performed better in episodic memory in language tasks (Kormi-Nouri et al., 2003).

### 1.5. Contextual learning paradigm

In order to examine the process of new word learning through building up different relations with the existing semantic network, it is necessary to design a learning paradigm that is based on multi-contextual learning. The contextual paradigm is previously commonly used in the research on L2 acquisition, especially in L2 word learning (Elgort et al., 2015; Elgort & Warren, 2014; Ferreira & Ellis, 2016; Nassaji, 2003). The formation and strengthening of new word-related memory greatly benefit from various episodic memories from diverse contexts. It has been established that

context diversity enhances both the encoding and retrieval of newly acquired concepts (Frances et al., 2020; Johns et al., 2016; Pagán & Nation, 2019). A multi-contextual learning paradigm would also benefit the participants in inferring the meaning of novel words (Bolger et al., 2008).

### 1.6. *The present study*

Based on the rationales above, we hypothesized that AoFE could influence the acquisition of lexical semantics (L2 word learning) in a way that is not accounted for by AoTE and other confounding factors, which would be manifested as better retrieval of episodic memory and more efficient establishment of new semantic representation in the semantic memory (especially the establishment of thematic relations). AoTE is a very straightforward factor that will influence the outcomes of L2 acquisition because the longer the learners are exposed to English, the better their English learning outcome will be. Participants can have the same AoTE but different AoFEs. There are some previous studies that aimed to study similar questions, but limitations remain. For instance, in some studies, the experimental materials adopted were based on the participants' native language, which is not suitable for examining genuine learning effects for L2 (Jalali-Moghadam & Kormi-Nouri, 2017; Kormi-Nouri et al., 2003). In the present study, all of the experimental materials are written in English (participants' L2). Many previous studies compared two populations at the extreme spectrum, that is, bilinguals with both languages at the native level and monolinguals who never use L2 (Kaushanskaya & Marian, 2009; Kormi-Nouri et al., 2003). In contrast, the population of bilinguals whose AoFE to L2 varies to different degrees was rarely explored. To address this limitation, the present study recruited participants with a relatively large range of AoFE.

In addition to addressing the limitations of the previous study as raised above, the major aim of this research is to examine the influence of AoFE on L2 new word learning, particularly on the process of L2 novel concepts being integrated into the conceptual semantic network in the semantic memory through taxonomic and thematic relations. In designing our study, three novel elements were involved. The first one is, as mentioned, that we recruited participants with a relatively broad range of AoFE, which would allow a reliable statistical estimation of the AoFE effect on learning outcomes. Second, by employing a multi-contextual learning paradigm combined with new-theme testing materials, we aim to differentiate the effects of new word learning in terms of whether it is from episodic memory or semantic memory. Third, we will directly measure the neural indicator of the learning outcome to serve as evidence of the learning effect directly from the neural level. In measuring performance related to learning L2 novel concepts, previous research has been primarily based on behavioral data such as accuracies and reaction times (RTs) (Tamminen & Gaskell, 2013), which cannot reveal the underlying subtle neural effects of learning. Therefore, we will apply ERP (event-related potentials) technology, particularly the extraction of the N400 effect that has been discovered by Kutas and Hillyard (1980) under the 'semantic violation' paradigm and has been widely used to indicate neural activity of semantic processing to examine the learning performance in the present study (Benau et al., 2011; Lau et al., 2008; Mathalon et al., 2002). The magnitude of the N400 effect is usually calculated by subtracting the ERP elicited by the keyword in a congruous condition from the ERP elicited by the

keyword in an incongruous condition. A larger N400 effect suggests that participants' brain respond more strongly to the incongruency, which is only possible when the word is better learned. In the current research, it is hypothesized that in the testing session, learners with earlier AoFE to L2 would exhibit larger N400 amplitudes in all three types of discourses and sentences than late learners, which means that earlier AoFE can have advantages in acquiring L2 unknown concepts both in episodic and semantic memory. In addition to the AoFE effect, we also hypothesized that the results could show positive effects evoked by other confounding factors, such as AoTE to L2 and L2 proficiency.

## 2. Materials and methods

### 2.1. Participants

Eighty-eight adults (18–25 years old, mean = 20.7, SD = 1.8, 46 females) participated in this research. Eighty-seven valid data were retained after removing the data of one participant because of excessively low accuracy (did not reach the chance level accuracy of the task: 50%). All of the participants were right-handed, and their vision and hearing were normal or corrected-to-normal. They were all native Chinese speakers and had English as their L2. They had no psychological or physical disorders. They were all undergraduate or postgraduate students at Shandong Normal University, and they were not majoring in language or psychology. This research has received Ethical approval from the Human Research Ethics Committee at the University of Hong Kong.

### 2.2. Collection of demographic and background information related to L2 learning

#### 2.2.1. Assessment of AoFE to English and AoTE to English

Since the highlight of this study is to examine the AoFE effect on the acquisition of L2 novel concepts, measuring the participants' AoFE is one of the key parts of this study. In the present study, it is important to note that we aim to collect information about the time from which the L2 learners began receiving systematic, intensive and continuous exposure to the new language rather than shallow and occasional smattering (Kovelman et al., 2008). To determine participants' AoFE to L2, some previous studies recruited participants from bilingual schools, and the AoFE can be easily defined as the time of entering the bilingual school. Bilingual schools refer to the schools in which 50% of teaching activities are conducted in L1 and 50% in L2 (Kovelman et al., 2008). In the present study, the participants did not attend a bilingual school as defined above but rather learned English in a predominantly monolingual setting with the major aim of learning English oriented to preparation for the entrance examination but not for daily use purposes. This is a typical formal education setting in China at the time when the current group of participants grew up (but it is much different now). Therefore, the AoFEs of the participants in the present study were difficult to evaluate precisely, but the advantage here is that they formed a relatively wide range primarily due to different situations in different regions in China. In this case, the questionnaire in this study was deliberately designed by the authors to evaluate their AoFE. Systematic exposure to English mainly refers to the English education received in school teaching and training in qualified training centers. However, in our participants, there were many complicated factors in their

experience of systematic exposure to English. The complications were rooted in the quality of their English education, the continuity of the exposure and the degree of their engagement. Considering all the complication issues in assessing AoFE, we attempted to improve the genuineness of AoFE by fully explaining the meaning of AoFE to our participants before asking them to report their AoFE. In our explanation, we have given some exemplary cases that may be mistaken as AoFE. For example, if someone started to learn English at a very early age but only in a very informal or superficial way or was interrupted for some years before they formally picked up the learning again, they would identify their AoFE as the later time point of formal learning but not the first point of informal and superficial exposure. After obtaining the participants' report of their AoFE to L2, we verified their answers by asking them what English program they participated in, what textbook they used, how many hours they learned English each week, and how the teachers' quality and their learning quality (achieved grades) were during the time when they first learned English. Moreover, since the measurement of AoTE is based on the calculation of time that participants spend in their English-related activities, in their recalling process, they were required to recall the English activities that they participated in at a certain stage of learning, which inevitably involved the starting age of English learning. As such, AoFE and AoTE that participants reported mutually verified each other. Although there may still exist some ambiguity, we think this detailed explanation would result in a more precise assessment of the time of first exposure to English.

In the present study, AoTE was assessed in a very detailed way based on a deliberately designed English exposure questionnaire according to the English teaching situation in mainland China around 20 years ago (participants in this study grew up at that time). The purpose of this questionnaire is to more precisely calculate participants' total English learning time from childhood to adulthood. In this questionnaire, the participant was required to evaluate the amount of time devoted to English learning-related activities in different phases of their formal education. The six phases are the pre-school (e.g., kindergarten) period, primary school period, junior high school period, senior high school period, undergraduate period and master period. Each learning phase is separated into several grades (years) (first grade, second grade, … etc.), and participants reported their English learning time related to certain English activities within a certain grade. Within each grade (year), participants' English learning time can be recalled separately on the basis of the school day (from Monday to Friday), weekend (Saturday and Sunday) and vacation days. The participants filled in the information based on their memory recollection and the filled information was further checked and verified by the experimenter to avoid vague description and obvious mistakes. AoTE was then calculated as the sum over all phases for each participant.

### 2.2.2. Assessment of participants' English proficiency

The measurement of participants' English proficiency can be subjective and objective (Hulstijn, 2012), both of which were used in the current study. The subjective measurement is based on participants' self-assessment of their English proficiency in listening, speaking, reading and speaking (Gullifer et al., 2021). The objective measurement is normally based on tests or test-style questionnaires, for example, a pragmatic and grammatical questionnaire or a vocabulary size test. The Vocabulary

Size Test by Nation and Beglar (2007) was used in this study to assess English lexical proficiency and lexicon knowledge of the participants. The pragmatic and grammatical questionnaire was used in the current study to examine the participants' English pragmatic and grammatical abilities and awareness. This questionnaire is adapted from the test designed by Bardovi-Harlig & Dörnyei (1998). All separate scores were converted to a percentage value. Considering that L2 proficiency measurement based on objective performance provides more useful information than subjective ratings (Brysbaert, 2013), the proficiency score was calculated as 0.1 × subjective assessment score + 0.9 × objective assessment score. The 90% portion of the objective assessment score was composed of 30% for the assessment of vocabulary knowledge and 60% for the assessment of pragmatic and grammatical knowledge.

### 2.2.3. Testing of participants' English working memory capacity

In this study, a standard and commonly used reading operation task was used to examine the participants' working memory capacity in L2. It is commonly used to specifically measure learners' ability to memorize and retain L2 information during the flow of processing L2 information in the field of language processing research (Daneman & Carpenter, 1980; Daneman & Merikle, 1996). Specifically, participants are required to read several groups of English sentences that do not contain any difficult English words. Each group includes a fixed number of sentences, gradually increasing from 2 to 6. For each sentence, they have to judge the semantic correctness and remember the last word on the basis of fully understanding the sentence. After the presentation of each group of sentences, participants are required to recall and write down the last word of each sentence as many as they can. This reading span task can also be administered to require participants to orally report the words in the final position of sentences. Here in our study, in order to have a better documentation process and quantitative analysis, we chose to require participants to write down the final words of sentences. In this working memory capacity test, participants are simultaneously assigned two cognitive tasks: processing the sentences' meaning and memorizing the last words of each sentence. The number of words that have been correctly recalled and written down by the participants is used to represent their performance, and the correctness of their judgment on this task is used to show the validity of their performance. The English materials used in the reading operation task were compiled by the authors, and 94% of the participants in this study reported that they could understand the meaning of all the sentences. The average accuracy for all participants in the semantic correctness judgment task reached 89%, which also showed that the difficulty of English sentences is low, and the semantic judgment of these sentences is easy.

### 2.2.4. Testing of participants' personality related to English learning

In this research, we adopted the personality type test that was designed by Qiufang (2003) to study the effect of personality on the learning lexical semantics of unknown L2 concepts. The test includes 15 questions that examine the personality type of English learners in terms of the degree of extroversion in the context of L2 learning. The participants' scores measured from this test were recorded. Higher value means more extroverted.

### 2.2.5. Assessment of participants' SES information

In this study, the participants' SES was measured from the following three aspects: educational attainment, place of birth and living and household income (Braveman et al., 2005; Krieger et al., 1997; Oakes & Andrade, 2017). For educational attainment, the participants' own educational levels and their parents' and relatives' education level were all considered. Their own and their parents' education level were evaluated on a scale ranging from 0 to 5 (0 represents elementary school while 5 represents graduate school [doctor]). The education level of their relatives was evaluated by the proportion of relatives who obtained at least a bachelor's degree, which was evaluated on a scale ranging from 0 to 4. Since these participants were all students at Shandong Normal University, they shared the same current residence (Jinan, Shandong, China) but differed in their growth environment; thus, the places where they were born and raised were recorded. The places of birth were divided into the urban, county, town and rural areas according to the regional development and were scored as integers from 0 to 3, respectively. In addition, the cities in which they were born and raised were also ranked from first-tier cities, new first-tier cities, to fifth-tier cities according to the 'Ranking of Cities' Business Attractiveness in China 2021″' issued by the YiMagazine in the *New First-tier Cities Summit* and were evaluated on a scale ranging from 0 to 5. Finally, the household income of the participants was classified into seven levels. In the calculation of participants' SES scores, education, place of birth and living and household income are equally weighted. Before summing up the final SES score, all separate scores were converted to a percentage value. The final SES score was finally calculated as ((participant's education level + parent's education level + relative's education level)/3 + (region score + city score)/2 + household income)/3.

### 2.3. Behavioral and EEG experiment

A concise description of the experiment with sufficient information for the current study is provided in the following. The experiment is divided into a learning session and a testing session. Both tasks in the two sessions were prepared and implemented by Psychtoolbox-3 in MATLAB. Before the formal part of the learning and testing sessions began, there were practice sessions separately for the learning session and testing session. The practice session contained six learning discourses and questions, which were embedded with two novel words, as well as ten testing discourses for the two novel words. Before the beginning of the learning and testing sessions, participants were carefully instructed and ensured that they understood the experimental flow.

### 2.3.1. Learning session

*Learning materials.* In the learning session, 60 contextual English discourses were presented, with every three discourses forming a set (thus 20 sets in total). Within a set, three discourses describe three different themes in which a novel word may appear. In this way, 20 novel words were learned from the 60 contextual English discourses. Each novel word was placed in the final position of the three discourses (see reasons below). All of these novel words were selected from the experimental materials used in the work of Deacon et al. (2004). The novel words are nonderivational pseudowords and conform to the English orthographic and pronunciation

rules but cannot be derived from any existing words. The definitions of these novel words are created according to people's existing world knowledge, but with a slight distortion. They are not peculiar concepts and can be understood by participants. The definitions contain information about the categories to which the novel concepts belong and one to two extra features or functions that seem atypical to the categories, for example, 'the chopsticks that can heat themselves'. The length of these novel words ranges from four to six letters, and they were presented in bold in the learning discourses. All of these pseudowords are nouns, and they play a concrete role in the discourses.

In the learning session, participants are not told the definitions directly. Instead, the learning discourses provided restrictive information for participants to infer the meaning of the novel words. In order to ensure that the discourses are highly restrictive for participants to infer the definition of the novel words, there were 15 participants who did not participate in the main study completing the inference task. The major features or categories of novel words are marked by several bullet points in the definition of each novel word. Participants were required to write down the meaning of novel words or the bullet points for the definition of the novel word as much as possible in light of their comprehension of the novel word. The responses that they wrote down were examined by the authors, and then we calculated the number of bullet points contained in their responses. Ten pilot participants correctly inferred more than 80% of the bullet points describing the meanings of the novel words. The five pilot participants who left produced the following scores: 78.87%, 77.46%, 76.06%, 76.06% and 42.25%. Although the results showed that one-third of the participants achieved less than 80% accuracy, the majority of them had satisfactory performance, indicating that the materials prepared were not too difficult yet were sufficient to differentiate participants. In addition to the contextual constraints, to verify the contextual variety of learning discourses, there were 10 additional participants judging whether the three discourses for the same novel word describe diverse contexts or not. Seven participants marked all learning discourses for the same novel word as describing multiple contexts. The length of the discourses is balanced.

An example of a novel word and its discourses is provided here:

The novel word: speth.

Learning discourses:

*John is a construction worker, and he has to eat outdoors in winter. To keep his fingers warm, he turns on the heating button on the speth.*
*John's child is five years old, and he no longer needs to use the spoon when eating. As a Chinese, his child has learned to pick up food using the speth.*
*John is making the dessert. According to the recipe, firstly, he should add the flour to the eggs and mix them well using the speth.*

Designed meaning of 'speth': *chopsticks that can heat themselves.*

All learning discourses are divided into four blocks and are presented one by one, with each block including 15 discourses for five novel words. The presentation and learning of each block were followed by 15 two-choice questions for consolidating the participants' learning outcomes. The options for the two-choice question are composed of two phrases. Only one of them is thematically related to the novel words. The

other option is completely unrelated. The purpose of these choice questions is to ensure that the participants have actually learned the discourses in the learning session. The correct options basically reflect the key contexts or themes in the learning discourses and inevitably involve some keywords from the learning discourses. There are three choice questions corresponding to the same novel word; therefore, for each block of 15 learning discourses, there are 15 choice questions. To evaluate the thematic relatedness between the learning discourses and the correct option (thematically related phrase) used in the choice question, another 20 participants were asked to fill out the 7-point Likert scale. An example of a choice question is shown below:

> *Which one of the following two options, A and B, is thematically related to the novel word* **speth** *you just learned?*
> *A: speth – Chinese way of picking up food.*
> *B: speth – open the curtain after getting up.*
>
> *Correct answer: A.*

*Learning procedure.* The flow of the learning session is depicted in Figure 1. Participants completed the self-paced reading task in a computer lab. The viewing distance between the screen and the participants is approximately 60 cm. The texts are displayed in white color on a black screen on a 27-inch (diagonal) Dell monitor (resolution: 1024 × 768 pixels; refresh rate: 100 Hz). All learning materials are in Times New Roman font, with a size of 42. The height of each letter of the words is about 0.55 cm–0.65 cm on the screen, and the width of each letter of the words is about 0.3 cm–0.4 cm on the screen. Before the start of the experimental program, participants were asked whether they could clearly see the materials on the screen, and affirmative answers were obtained. First, in the center of the screen, a fixation cross is presented for 800–1200 ms. After the fixation cross, for each novel word, three discourses are displayed one after another and finally appear together on the screen. Then, the participants' comprehension of discourses was evaluated by self-rating on a 5-point scale (from 'cannot understand' to 'completely understand'). Participants gave a high rating on their understanding of these learning discourses (M = 3.87, SD = 0.53), which indicated a good comprehension of the learning discourses. After the rating, a blank screen appears and lasts for 2000–2800 ms before the onset of the next trial. The presentation order of three discourses for the same novel word within a set was randomized across all participants. The presentation order of the discourses for five novel words within the same block is randomized. The presentation order of the four blocks was also randomized.

After the presentation of 5 sets of discourses, the participants were asked to answer 15 two-choice questions one by one in a randomized order. Each question remained on the screen until participants pressed the left or right key to answer. The program does not allow three consecutive answers to appear on the same side. For each choice question, participants' incorrect answers would again lead to the presentation of the corresponding discourse. Then, a retest of the same choice question is performed by the participants, and this cycle ends when the answer is correct. The next question appears after the offset of the duration (equally distributed between 2000 ms and 2800 ms). These questions were easy for the participants: most participants could give the right answer when they encountered the choice questions for the first round (the average correct rate was 95.84%, SD = 4.18%).
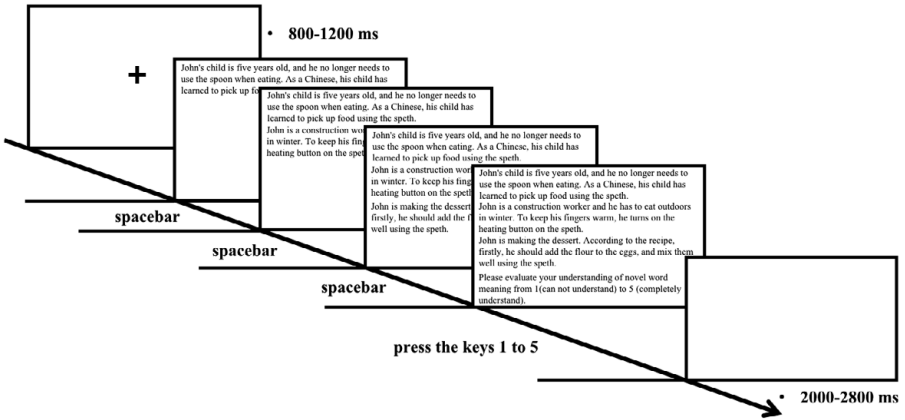
A. Schematic overview of the learning session (learning discourses or sentences).
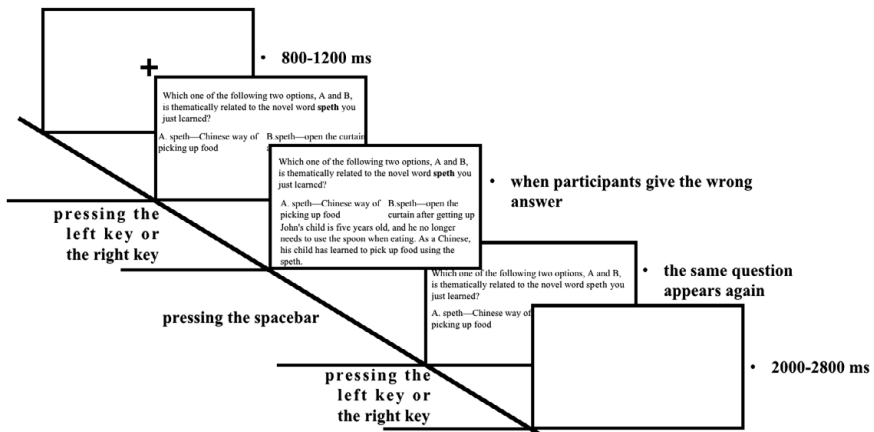


B. Schematic overview of the learning session (choice questions).

**Figure 1.** Schematic overview of the learning session.

### 2.3.2. Testing session

*Testing materials.* After the learning session, the participants wore the EEG cap and performed the testing session. The testing session includes 60 recurrence type discourses, 20 new-theme type discourses, and 20 category-feature type sentences, all of which are written in English. For each type of discourse or sentence, half of them are made congruous by making the contexts described by the discourses or sentences fit the novel words, while the other half are made incongruous by substituting the original novel words with unmatched novel words selected from the list of other novel words in the learning session. All novel words are placed at the final position of testing discourses or sentences to create incongruency or congruency relationships between target words and preceding contexts. Novel words in the testing discourses or sentences are underlined. The pairing of novel words and discourses or sentences is pseudorandom.

The recurrence type discourses refer to the same discourses that have been learned in the learning session, and they can be used to examine whether learners have mainly

absorbed the novel words into their episodic memory because the processing of recurrence type discourses mainly involved the recollection and recognition of original contexts. It is possible that participants can make the right judgment on the congruency in the recurrence type discourses even though they do not understand the target word because their judgment was based on familiarity with the original contexts that they saw. An example of a recurrence type discourse is shown below:

> *John is a construction worker, and he has to eat outdoors in winter. To keep his fingers warm, he turns on the heating button on the jommer.*
>
> *Answer: false.*

The new-theme type discourses provide new themes that were not encountered by participants in the learning session. The introduction of the new-theme type discourses is to examine whether the participants have truly understood the novel words and are able to make sense of them under new thematic relations independent of the contexts. The length of new-theme type discourses is designed to be similar to that of recurrence type discourses. To ensure that the new-theme type discourses describe different contexts from those in the learning discourses, another 15 participants were recruited to evaluate these discourses by reading them, and on average, 83% of new-theme discourses were reported as describing different themes from the recurrence type discourses. It should be noted that although the three original contexts (for each novel word) provided in the learning session may have built some thematic linkages between novel words and semantic networks, the new-theme discourses are still valid in testing the establishment and extension of thematic relations. The difference in the themes used in learning and testing sessions makes sure that they do not interfere, and the effects generated by the new themes would mark a qualitative shift from episodic memory to semantic memory. An example of a new-theme type discourse is shown below:

> *John was eating lunch. The food dropped and dirtied John's new T-shirt when he tried to pick up food using the over-heated speth.*
>
> *Answer: true.*

Category-feature sentences describe the categories and features of novel words. The aim of this condition is to investigate the establishment of taxonomic relations in the semantic network by asking participants to differentiate the category features of novel words. An example of a category-feature sentence is shown below:

> *The chopsticks that can heat themselves are speth.*
>
> *Answer: true.*

It should be noted that in the new-theme type discourses and category-feature type sentences, it is inevitable to have a few words repeating the learning discourses. The judgment of recurrence type is assumed to mainly rely on episodic memory even though it cannot completely rule out the involvement of semantic memory (which should be much less, if it exists, than the other two conditions). And the judgment of new-theme type and category-feature type should mainly employ semantic memory.
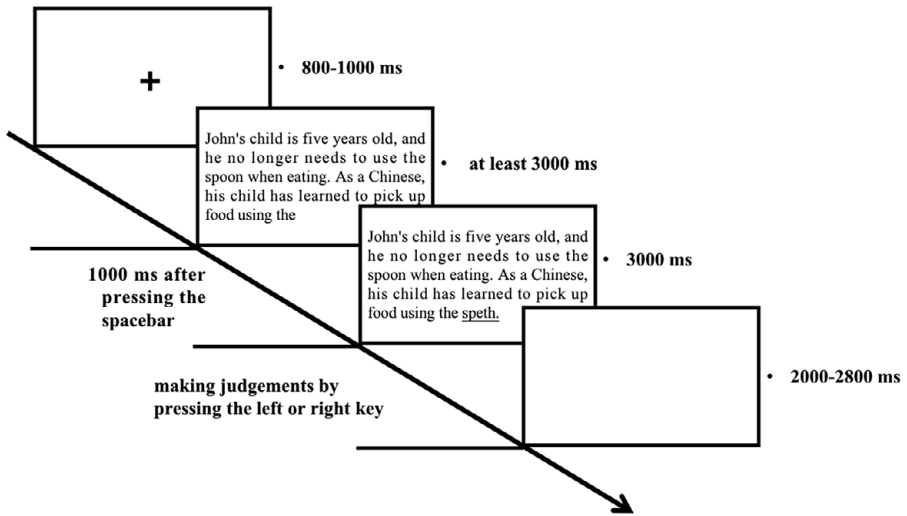
**Figure 2.** Schematic overview of the testing session.

*Testing procedure.* The presentation procedures of the testing session are depicted in Figure 2. First, in the center of the screen, a fixation cross is presented for 800–1200 ms. Discourses or sentences appear on the screen one after another. Participants can read the entire discourse or sentence except for the newly learned word in the final position. After at least 3000 ms, participants can press the spacebar, after which the novel word will appear in the final position of the discourse or sentence (1000 ms after the spacebar press). Here, participants were notified before the testing session that they should focus on reading and understanding the discourses or sentences before they pressed the spacebar because they were given only 3000 ms to judge the congruency between the novel word and the discourse by pressing the left (congruous) or right key (incongruous). Participants were required to make their judgments as quickly and as accurately as possible. After the judgments are made, a blank screen will appear and stay for 2000–2800 ms before the onset of the next trial. Assignment of the left and right keys to congruous and incongruous conditions is counterbalanced. Discourses originally designed for the same novel words or including the same novel words are separated by other discourses.

In the testing session, all discourses or sentences are systematically separated into three types and are presented alternatively. The presentation order of the 100 discourses is as follows: 20 recurrence discourses, 20 new-theme discourses, 20 recurrence discourses, 20 category-feature sentences and 20 recurrence discourses. The category-feature type was placed at the relatively back position to avoid participants' exposure to the word meaning implied by the category-feature type sentences before processing the recurrence type and new-theme type discourses. The main reason we administered 60 trials of recurrence discourses and divided them into three sections is to demonstrate that the insignificant effects from the other two conditions (new theme and category-feature), if found, are not caused by fatigue. During the whole testing session, participants were instructed to keep still and try to avoid blinking during the judgment process. Participants were required to take a short break after every 10 trials.

### 2.3.3. *EEG data collection and preprocessing*

Participants wore an elastic electrode cap with standard 64 Ag/AgCl electrodes (Brain Products, Germany) according to the international 10–20 system for recording their Electroencephalogram signals during the testing session. The signals were sampled at 1000 Hz, and the physical reference electrode was FCz. The ground electrode was placed between FP1 and FP2. Throughout the EEG collection, the electrode impedance was maintained below 5 kΩ. We used Letswave 7 (André Mouraux, Brussels, Belgium) and EEGLAB (Delorme & Makeig, 2004) to preprocess and analyze the EEG data. The data were first resampled to 250 Hz and referenced to the cross-electrode average. Then, a band-pass filter in the range of 0.05–30 Hz was used. Afterward, the SPA toolbox (Ouyang et al., 2022) was employed on all participants to remove EEG artifacts based on a threshold parameter of 30. SPA removes artifacts based on decomposed principal components with abnormal amplitudes instead of discarding the whole trial. Therefore, after the SPA, no trial is completely deleted from the dataset.

It has to be noted that our current preprocessing does not employ the commonly used procedure with independent component analysis (ICA). The reasons are as follows. In the beginning, we applied the ICA method to remove the ocular artifacts in combination with both automatic artifact removal packages and visual inspections of artifacts. Following several documented publications (Chaumon et al., 2015; Dimigen, 2020; Winkler et al., 2015) and the official EEGLAB tutorial website (https://eeglab.org/tutorials/06_RejectArtifacts/RunICA.html), we learned that the ICA method' s performance would be compromised if the low cutoff of band-pass filtering is lower than 1 Hz. Our observation of the ICA application results was in line with this statement. We found that lowering the low cutoff increasingly retains ocular artifacts. When the low cutoff is low (e.g., 0.05 Hz), it is difficult to unambiguously identify ocular artifacts from the ICs for both automatic tools and visual inspection. It appeared that the ocular artifacts shattered across many ICs, including the low-energy ones. We could have applied 1 Hz cutoff filtering. However, our N400 effects appear to dominate the low-frequency part, which is also in line with the literature of N400 studies (Chen et al., 2014; Ding et al., 2017; Zhang et al., 2019). This left us a dilemma 1) having better ICA decomposition but sacrifice our N400 component (as said in the official EEGlab tutorial and documented literature) or 2) having compromised ICA decomposition by applying low cutoff but better retaining our N400 component. To address this dilemma, eventually, we resorted to methods that simply removed the large-variance PCs.

The last step is to obtain the average ERPs locked to the presentation time of the novel word at the end of each discourse. The ERP data used for analysis were obtained by segmenting EEG trials from 200 ms before the onset of the novel word to 1000 ms after the stimulus. The baseline correction was based on $[-200\text{ ms}, 0\text{ ms}]$. The average ERPs for congruous and incongruous conditions were separately obtained, and the difference wave between them was calculated to obtain the N400 component. The mean amplitude of the N400 component was calculated from the time window of $[300\text{ ms}, 500\text{ ms}]$ after stimulus onset. Except for the outliers that have been removed (see below), all trials under the same conditions are involved in the averaging process.

## 3. Statistical analysis

This research primarily focuses on examining the effect of AoFE on L2 learning outcomes, particularly on semantic learning of new L2 words, as indicated by the

N400 effect. To prove the existence of the learning effect at the neural level, the ERP amplitudes elicited by congruous and incongruous discourses from specific time windows and electrodes (reported in the Results section) were separately calculated, and their difference serves as the neural indicator of learning outcome. For statistical testing of the significance of the N400 effect, we first applied repeated measures ANOVA across all conditions and then paired *t* tests in separate conditions. The *T* value, p value, degree of freedom, SEM (standard error of the mean) and effect size measures were reported. Multiple comparison corrections based on the Bonferroni correction were conducted wherever needed (Shaffer, 1995). The average accuracy and RT separately for all discourses combined, different discourse types and different blocks of the recurrence type were reported. The outliers (beyond mean ± 3 SD) from the behavioral variables (0.25% excluded) and neural variables (0.82% excluded) were excluded before calculating the descriptive statistics.

Afterward, to investigate the AoFE effect on the N400 component, several other major factors that may also significantly affect L2 learning and thus confound the effect of AoFE were identified according to the literature. This leads to the following set of independent variables: 1) AoFE, 2) AoTE, 3) English proficiency, 4) English working memory capacity, 5) gender, 6) personality and 7) SES. The dependent variable is the N400 effect, which was calculated as the difference between the congruous and incongruous conditions, and was further separated into four versions: 1) N400 for all discourses combined, 2) N400 for recurrence type discourses, 3) N400 for new-theme type discourses and 4) N400 for category-feature type sentences. The outliers (beyond mean ± 3 SD) from the participant samples were excluded, resulting in an exclusion rate of 0.29%. We first conducted Pearson correlations among variables to provide a coarse-grained overview of their internal relationships. Before conducting the multiple linear regression (MLR) analyses on the separate conditions, we first applied a linear mixed model (LMM) analysis (using R) to examine the independent variable's effects across all the conditions in the same model. This model thus includes all discourse types (three types) and all the above-mentioned independent variables as the whole set of independent variables. The independent variables (except for gender, that is, binary) were standardized as Z scores before feeding to the LMM fitting. After the LMM analysis, we conducted three separate MLR analyses (conducted in SPSS 28.0.1.0) for the three discourses and sentences separately. It should be noted here that in the current study, the research scope does not include the interaction effects among various independent variables on the N400 components (dependent variables). Therefore, no specific hypotheses about the interactions were formulated and tested.

We decided not to remove the inaccurate trials in our ERP analysis. In this research, the effect of AoFE was mainly manifested in the neural data rather than the behavioral data. We hypothesized that the semantic processing of the novel words occurs in the brain, regardless of whether participants gave the wrong or right answer. In this sense, the relevant neural effect of semantic processing at single trials was treated as a continuous variable. Below a threshold at this continuous variable, the participant may generate a wrong answer. However, we have to include all the trials to have all the neural effects (no matter to what degree) retained in our statistical analysis to reveal the neural effects. We also analyzed the data based on correctly responded trials for comparison. The differences between the two versions of analysis will be discussed in the Discussion section.

## 4. Results

### 4.1. Descriptive statistics of the independent variables

In this study, there are, in total, seven independent variables that are examined in their contribution to the outcomes of lexical-semantic learning of L2/English unknown concepts measured by N400. The results of their descriptive statistics are summarized in Table 1.

### 4.2. Behavioral and EEG data measurements

#### 4.2.1. Behavioral data measurements

We separately report the results of the descriptive statistics of behavioral data for different discourse types and different blocks of the recurrence type in Tables 2 and 3. The validity of the behavioral data is supported by the relatively high average accuracy (substantially higher than 50%) and the cross-individual variability, as well as the relatively short average RT (ranging from 1 to 1.3 seconds). This also means that the judgment tasks in the testing session are not so difficult that participants did not have to take many seconds to perform them. Therefore, the ERP method is an appropriate neural measurement method since it is usually used to investigate stimulus-elicited neural processes that occur within 1–2 seconds.

The accuracy of the recurrence type discourses was largely in line with the pilot results of learning discourses. Although the range is from 0.43 to 1.00, the majority of them showed satisfactory performance. This can be demonstrated by the histogram as follows (Figure 3), which shows that most participants achieved an accuracy exceeding 70% in the judgment of recurrence type discourses. Only very few participants completed an accuracy below 60%. These behavioral results implied that most participants achieved an acceptable performance in the recurrence type of testing session. We believe that the wide distribution of the accuracy is appropriate since it can differentiate participants' learning outcomes. If the learning discourses are too easy for participants to learn, there would be little individual differences in the N400 effect elicited by their judgment of testing discourses.

**Table 1.** Descriptive statistics for the independent variables related to English learning

|  | *N* | Min. | Max. | M | SD |
|---|---|---|---|---|---|
| AoFE | 87 | 3.00 | 14.00 | 8.72 | 2.82 |
| English proficiency | 87 | 20.05 | 76.32 | 50.86 | 11.86 |
| AoTE | 87 | 1953.87 | 10791.22 | 5250.74 | 1933.71 |
| English working memory capacity | 87 | 19.00 | 57.00 | 39.96 | 8.29 |
| Gender | 87 | .00 | 1.00 | .51 | .50 |
| Personality | 87 | 1.00 | 14.00 | 7.95 | 3.22 |
| SES | 87 | 10.00 | 91.11 | 43.90 | 17.37 |

*Note*: *N*: sample size. M: mean. SD: standard deviation. AoFE: age of first exposure to English. English proficiency: score obtained from the weighted sum of many tests and questionnaires measuring English proficiency. AoTE: the amount of total English exposure measured as the sum of participants' English learning time at the unit of the minute. English working memory capacity: score of participants' working memory capacity to store and process English materials. Personality score: measurement of outgoingness. The higher the personality score is, the more outgoing the participant is. SES: score of socioeconomic status.

**Table 2.** Descriptive statistics for accuracy

| Accuracy | N | Max | Min | Mean | SD | CV |
|---|---|---|---|---|---|---|
| Overall | 87 | .98 | .52 | .79 | .13 | .16 |
| Recurrence type | 87 | 1.00 | .43 | .80 | .13 | .17 |
| New–theme type | 87 | 1.00 | .35 | .76 | .15 | .20 |
| Category–feature type | 87 | 1.00 | .45 | .81 | .14 | .18 |
| Recurrence type block 1 | 87 | 1.00 | .40 | .77 | .14 | .19 |
| Recurrence type block 2 | 87 | 1.00 | .30 | .80 | .17 | .21 |
| Recurrence type block 3 | 87 | 1.00 | .50 | .82 | .13 | .16 |

*Note*: N refers to the sample size used in this descriptive analysis. SD refers to standard deviation. CV refers to the coefficient of variation.

**Table 3.** Descriptive statistics for reaction time

| Reaction time | N | Max | Min | Mean (s) | SD | CV |
|---|---|---|---|---|---|---|
| Overall | 87 | 1.77 | .59 | 1.18 | .26 | .22 |
| Recurrence type | 87 | 1.73 | .58 | 1.16 | .26 | .22 |
| New–theme type | 86 | 1.84 | .63 | 1.25 | .29 | .23 |
| Category–feature type | 87 | 1.75 | .56 | 1.16 | .28 | .24 |
| Recurrence type block 1 | 87 | 1.98 | .60 | 1.30 | .34 | .26 |
| Recurrence type block 2 | 87 | 1.73 | .52 | 1.15 | .28 | .24 |
| Recurrence type block 3 | 87 | 1.71 | .52 | 1.04 | .25 | .24 |

*Note*: N refers to the sample size used in this descriptive analysis. SD refers to standard deviation. CV refers to the coefficient of variation.
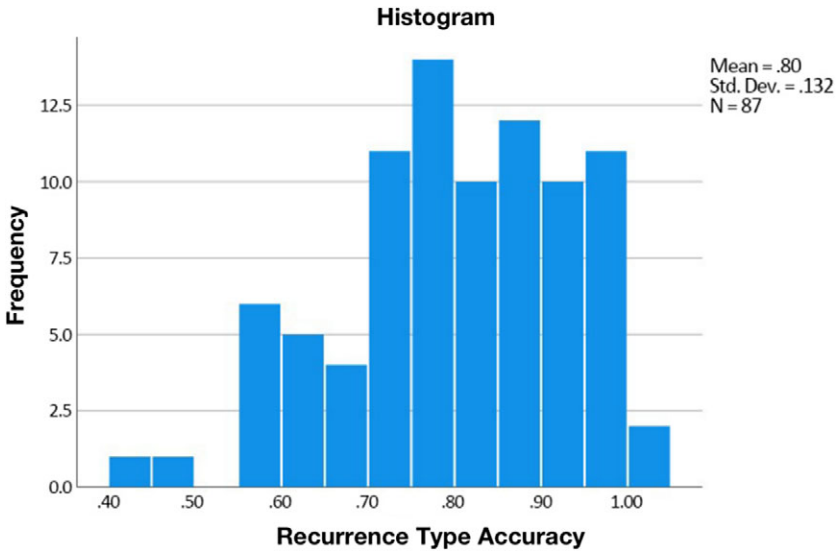


**Figure 3.** Distribution of accuracy in the recurrence type discourses.

### 4.2.2. EEG data measurements

The ERPs obtained from the novel words (target words) in the final position of a sentence are expected to show different amplitudes depending on whether the target word is congruous with the sentence or not. The effect caused by congruency is called the N400 effect. N400 is an ERP component with an amplitude reversely proportional to the degree of semantic violation between the target word and its contexts. Therefore, participants' mastery of novel words can be reflected in the indexes of N400. The N400 effect is calculated as the ERP difference between the averaged ERP waveforms from congruous and incongruous conditions (specifically, incongruous minus congruous). In this study, the time window from 300 ms to 500 ms after the onset of the novel word was selected to measure the N400 effect. The region of interest includes the central and parietal areas, which are represented by electrodes Cz, CPz and Pz since this area is conventionally adopted to observe N400 (Hajra et al., 2018; Mathalon et al., 2002). The analysis was separately done on different conditions, that is, all discourses combined, the recurrence type discourses, the new-theme type discourses and the category-feature type sentences.

We first conducted a repeated measures ANOVA with discourse type (recurrence type block 1, new-theme type, recurrence type block 2, category-feature type and recurrence type block 3) and congruence (congruous vs. incongruous) as two within-subject factors and found a significant interaction effect between the two factors ($F = 2.899$, $p = 0.022$, Partial Eta Squared = 0.03). Besides, the main effects of both discourse type and congruence were significant (type: $F_{(4,332)} = 3.008$, $p = 0.018$, Partial Eta squared = 0.04; congruence: $F_{(1,83)} = 33.649$, $p < 0.001$, Partial Eta Squared = 0.29). After the ANOVA test, paired *t* tests were conducted between congruous and incongruous on all discourses combined, the three types and three blocks of recurrence type, separately. For multiple *t* tests, we adopted the Bonferroni correction to adjust the significance level α. For the *t*-test results for all discourses combined (the entire dataset), we retain the original standard significance level (α = 0.05) because testing on the whole dataset of the entire study can be treated as a single test. Then, when analyzing (testing) the three types (recurrence type; new-theme type; category-feature type), they are treated as three tests. Therefore, based on the Bonferroni correction, the significant level α is adjusted by dividing the original value by three; thus, the new α to be compared with is 0.017. Since the recurrence type is further divided into three blocks, the new α value for the further divided subsets becomes 0.006, i.e., p values less than 0.006 (rather than 0.05) will be treated as significant results. As shown in Table 4, for all discourses combined, there is a significant N400 effect. For separate types, the recurrent discourses elicited a statistically significant N400 effect; new-theme type discourses also elicited a statistically significant N400 effect; category-feature type sentences did not elicit a statistically significant N400 effect. The recurrence discourses in blocks 1 and 3 elicited significant N400 effects.

Figure 4 shows the descriptive results of ERP waveforms and topography of the N400 effect for all discourses combined. Figure 5 shows the topographies of averaged ERP differences between congruous and incongruous conditions within a time window of [300 ms, 500 ms] for all discourses combined and different types of discourses or sentences.

**Table 4.** The N400 effect of the target word from all discourses combined, three types and three blocks of recurrence type at the average electrode

| Averaged electrode | *t* value | *p* value | Adjusted α | df | SEM |
|---|---|---|---|---|---|
| Overall | −5.83*** | <.00 | 0.05 | 85 | .07 |
| Recurrence type | −5.81*** | <.00 | 0.017 | 85 | .08 |
| New–theme type | −3.33** | .00 | 0.017 | 86 | .15 |
| Category–feature type | −0.60 | .55 | 0.017 | 86 | .15 |
| Recurrence type block 1 | −4.55*** | <.00 | 0.01 | 85 | .15 |
| Recurrence type block 2 | −2.63 | .01 | 0.01 | 84 | .12 |
| Recurrence type block 3 | −2.87* | .01 | 0.01 | 86 | .12 |

*Note*: df refers to the degree of freedom. Regarding the dfs for different types of discourses or sentences, the reason is that we separately examined the number of outliers in the data sample for different types (but based on the same criterion) and removed the different numbers of outliers in different types. Therefore, the actual numbers of participants that were used in the *t*-test analyses for different types of discourses or sentences were different. SEM refers to the standard error of the mean.
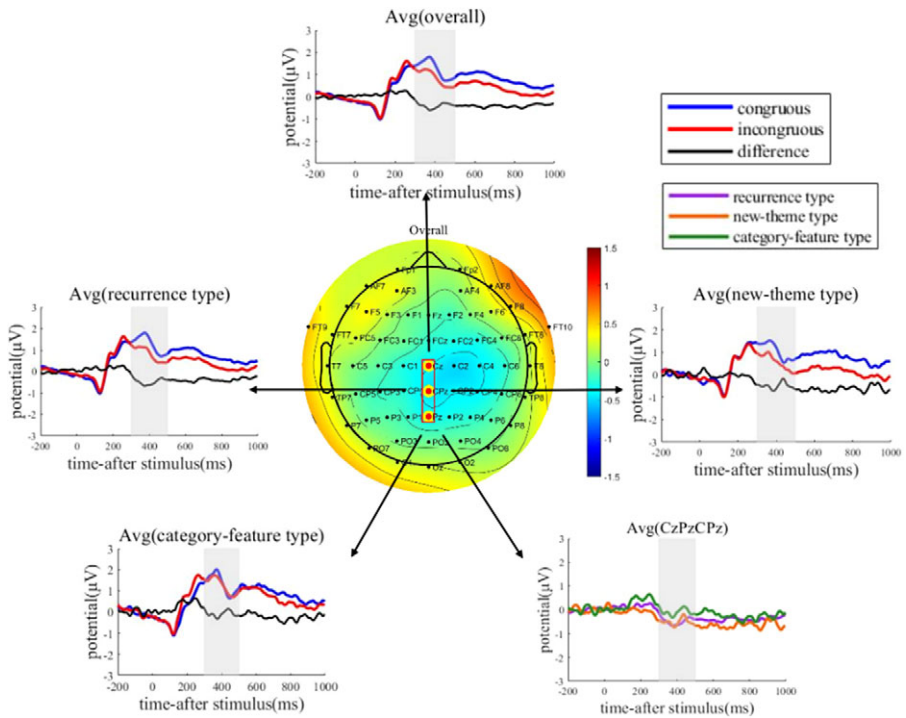


**Figure 4.** Time courses of ERP and topography of N400 effect. The ERPs are averaged from all types of discourses or sentences. The difference waves from three types of discourses or sentences are also presented together in a single waveform. The topography at the center is the ERP difference from all types of discourses combined between congruous and incongruous conditions averaged from 300 to 500 ms.

## 4.3. Correlation and regression analyses between all variables

The Pearson correlation analysis was conducted to examine the correlation among all variables to provide an overview of their internal relationships among them. The results of the correlation matrix are shown in Table 5. The correlation table is for the

## Topographies



**Figure 5.** Topographies of N400 effect obtained by subtracting ERP waveforms of congruous conditions from ERP waveforms of incongruous conditions for the overall data (all types combined) and different types of discourses or sentences.

purpose of providing information for readers' evaluation of the internal relationships but not for making statistical conclusions about the AoFE effect on lexical semantic learning. As such, we did not specifically carry out multiple comparison adjustments. Statistical significance analysis was carried out in the LMM and MLR models below.

After the Pearson correlation analysis, we conducted the LME analysis to evaluate the degree to which different factors contribute to the variability of the magnitude of the N400 effect across participants. The results showed that participants' AoFE, AoTE, English proficiency and personality had significant effects on the variability of the N400 effect across participants (Table 6).

After the LME analysis, we conducted three separate MLR analyses on the three types of discourses or sentences. The results are shown in Tables, and. Multiple comparison corrections based on the Bonferroni correction were conducted on the results obtained from multiple regression models (Shaffer, 1995). The significant level α was reduced to one-third of 0.05, that is, 0.017, because the test was conducted three times on each of the three sub-datasets. The positive t value of AoFE means that the N400 effect value is higher when AoFE is larger (i.e., starting L2 learning later). This is because the N400 effect here was calculated as

**Table 5.** Pearson correlations among 11 variables

| | N400 overall | N400 type1 | N400type2 | N400type3 | AoFE | AoTE | Proficiency | WMC | Personality | Gender | SES |
|---|---|---|---|---|---|---|---|---|---|---|---|
| N400 overall | 1.00 | .75** | .61** | .46** | .59** | −.58** | −.55** | −.31** | −.22* | .26* | −.21 |
| N400 type1 | .75** | 1.00 | .16 | −.01 | .46** | −.50** | −.36** | −.28** | −.16 | .33** | −.28** |
| N400 type2 | .61** | .16 | 1.00 | .15 | .47** | −.52** | −.50** | −.16 | .11 | .30** | −.11 |
| N400 type3 | .46** | −.01 | .15 | 1.00 | .04 | −.10 | −.27* | −.11 | −.36** | −.22* | .06 |
| AoFE | .59** | .46** | .47** | .04 | 1.00 | −.42** | −.43** | −.15 | .05 | .17 | −.33** |
| AoTE | −.58** | −.50** | −.52** | −.10 | −.42** | 1.00 | .53** | .26* | .07 | −.21 | .20 |
| Profici–ency | −.55** | −.36** | −.50** | −.27* | −.43** | .53** | 1.00 | .42** | −.05 | −.09 | .29** |
| WMC | −.31** | −.28** | −.16 | −.11 | −.15 | .26* | .42** | 1.00 | .01 | −.31** | .29** |
| Personality | −.22* | −.16 | .11 | −.36** | .05 | .07 | −.05 | .01 | 1.00 | .08 | −.04 |
| Gender | .26* | .33** | .30** | −.22* | .17 | −.21 | −.09 | −.31** | .08 | 1.00 | −.16 |
| SES | −.21 | −.28** | −.11 | .06 | −.33** | .20 | .29** | .29** | −.04 | −.16 | 1.00 |

*represents that correlation is significant at the 0.05 level (2-tailed).
**represents the correlation is significant at the 0.01 level (2-tailed).
*Note*: N400 overall refers to the N400 effect of all discourses combined. N400 type 1 refers to the N400 effect from the recurrence type discourses. N400 type 2 refers to the N400 effect from the new-theme type discourses. N400 type 3 refers to the N400 effect from the category-feature sentences. AoFE: age of first exposure to English. AoTE: the amount of total English exposure. WMC: English working memory capacity.

**Table 6.** LME results

| Random effects: | | | |
|---|---|---|---|
| Groups | Name | Variance | Std. Dev |
| Participant ID | (Intercept) | 0.000 | 0.000 |
| Residual | | 1.11 | 1.05 |

Number of obs: 260, groups: Participant ID, 87

Fixed effects:

| | Estimate | Std. Error | df | t value | Pr(>|t|) |
|---|---|---|---|---|---|
| (Intercept) | −0.75 | 0.18 | 251 | −4.06 | 6.47e−05 *** |
| Discourse type | 0.15 | 0.08 | 251 | 1.93 | 0.06 |
| AoFE | 0.18 | 0.08 | 251 | 2.21 | 0.03* |
| AoTE | −0.18 | 0.08 | 251 | −2.15 | 0.03* |
| Proficiency | −0.29 | 0.09 | 251 | −3.33 | <0.01*** |
| WMC | <−0.01 | 0.08 | 251 | −0.02 | 0.98 |
| Personality | −0.15 | 0.07 | 251 | −2.29 | 0.02* |
| Gender | 0.13 | 0.14 | 251 | 0.93 | 0.35 |
| SES | 0.09 | 0.07 | 251 | 1.26 | 0.21 |

'incongruent ERP - congruent ERP', and therefore, a higher N400 effect value (closer to zero) refers to a lower magnitude of the N400 effect. This relationship applies to the results in Tables, and. Regarding the collinearity problems among the 11 variables, the value of variance inflation factor (VIF) is provided, which shows that there are no collinearity problems among those variables in this study since the values of VIF are relatively small (<5).

The summary of the MLR model results for the N400 averaged from recurrence type discourses is shown in Table 7, which shows that AoFE and AoTE significantly contribute to the variance of the N400 effect averaged from recurrence type discourses. Specifically, lower AoFE and higher AoTE elicited larger N400 effects.

The summary of MLR model results for the N400 averaged from new-theme type discourses is shown in Table 8, which shows that AoFE, AoTE and English

**Table 7.** MLR results for recurrence type discourses

| Model | $R$ | $R^2$ | Adjusted $R^2$ | SE of the estimate | $R^2$ change | $F$ change | Sig. $F$ change | Durbin–Watson |
|---|---|---|---|---|---|---|---|---|
| | .64 | .41 | .36 | .60 | .41 | 7.91 | <.001 | 1.67 |

| | Unstandardized coefficients | | Standardized coefficients | | | Collinearity | |
|---|---|---|---|---|---|---|---|
| | B | Std. error | Beta | $t$ | Sig. | Tolerance | VIF |
| (Constant) | −.62 | .09 | | −6.76 | <.00 | | |
| AoFE | .20 | .08 | .26 | 2.54 | .01 | .71 | 1.42 |
| AoTE | −.21 | .08 | −.28 | −2.63 | .01 | .65 | 1.54 |
| Proficiency | −.03 | .09 | −.04 | −.37 | .72 | .57 | 1.77 |
| WMC | −.05 | .08 | −.06 | −.61 | .55 | .72 | 1.39 |
| Personality | −.13 | .07 | −.18 | −2.01 | .05 | .97 | 1.03 |
| Gender | .31 | .14 | .21 | 2.23 | .03 | .85 | 1.18 |
| SES | −.06 | .07 | −.08 | −.81 | .42 | .83 | 1.21 |

**Table 8.** MLR results for new-theme type discourses

| Model | $R$ | $R^2$ | Adjusted $R^2$ | SE of the estimate | $R^2$ change | $F$ change | Sig. $F$ change | Durbin–Watson |
|---|---|---|---|---|---|---|---|---|
| | .67 | .45 | .40 | 1.08 | .45 | 9.23 | <.001 | 1.81 |

| | Unstandardized coefficients | | Standardized coefficients | | | Collinearity | |
|---|---|---|---|---|---|---|---|
| | B | Std. error | Beta | $t$ | Sig. | Tolerance | VIF |
| (Constant) | −.78 | .17 | | −4.66 | <.00 | | |
| AoFE | .34 | .14 | .25 | 2.47 | .016 | .71 | 1.42 |
| AoTE | −.38 | .15 | −.27 | −2.65 | .01 | .65 | 1.54 |
| Proficiency | −.42 | .16 | −.30 | −2.70 | .01 | .57 | 1.77 |
| WMC | .13 | .14 | .10 | .97 | .33 | .72 | 1.39 |
| Personality | .12 | .12 | .08 | .98 | .33 | .97 | 1.03 |
| Gender | .58 | .25 | .21 | 2.32 | .02 | .85 | 1.18 |
| SES | .18 | .13 | .13 | 1.37 | .17 | .83 | 1.21 |

proficiency significantly contribute to the variance of the N400 effect averaged from new-theme type discourses. Specifically, lower AoFE, higher AoTE and higher English proficiency elicited larger N400 effects.

The summary of MLR model results for the N400 averaged from category-feature type sentences is shown in Table 9, which shows that the N400 effect did not reach significance in the category-feature type even though English proficiency and personality significantly contribute to the variance of the N400 effect averaged from category-feature type sentences. Specifically, higher English proficiency and extroverted personality elicited larger N400 effects.

To explicitly show the direction of the relationship between the N400 effect and various individual difference measures, the scatterplot and the best-fit line are provided in Figure 6.

**Table 9.** MLR results for category-feature type sentences

| Model | $R$ | $R^2$ | Adjusted $R^2$ | SE of the estimate | $R^2$ change | $F$ change | Sig. $F$ change | Durbin–Watson |
|---|---|---|---|---|---|---|---|---|
| | .52 | .27 | .21 | 1.11 | .27 | 4.21 | <.001 | 1.90 |

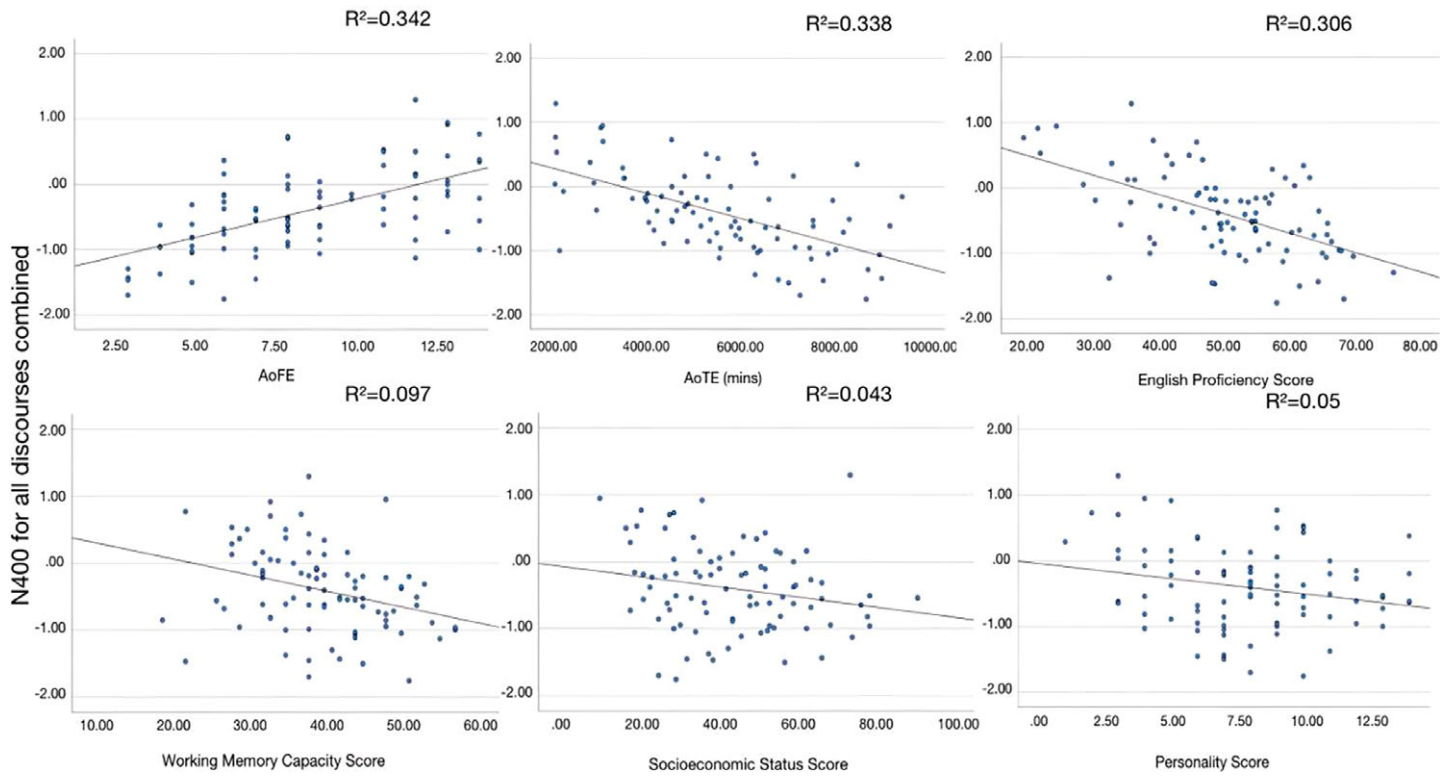| | Unstandardized coefficients | | Standardized coefficients | | | Collinearity | |
|---|---|---|---|---|---|---|---|
| | B | Std. error | Beta | $t$ | Sig. | Tolerance | VIF |
| (Constant) | .08 | .17 | | −.48 | .63 | | |
| AoFE | .01 | .15 | .01 | .07 | .94 | .65 | 1.54 |
| AoTE | .08 | .15 | .07 | .55 | .58 | .64 | 1.56 |
| Proficiency | −.41 | .16 | −.33 | −2.58 | .01 | .56 | 1.78 |
| WMC | −.10 | .14 | −.08 | −.71 | .48 | .71 | 1.41 |
| Personality | −.45 | .12 | −.36 | −3.66 | <.00 | .97 | 1.03 |
| Gender | −.52 | .26 | −.21 | −1.99 | .05 | .85 | 1.18 |
| SES | .16 | .13 | .13 | 1.20 | .23 | .82 | 1.22 |

**Figure 6.** Scatterplot and best-fit line showing the relationship between the N400 effect and various individual factors.

## 5. Discussion

According to the LME results obtained from all discourses combined (see Table 6), it can be seen that lower AoFE, longer AoTE, higher L2 proficiency and higher extroversion elicited larger N400 effects in individuals. The key finding for the current study is that AoFE is significantly associated with the neural N400 effect (earlier AoFE, larger N400 effect), even after controlling for various major confounding factors, including AoTE, language proficiency and so forth. The N400 effect serves as a neural marker of L2 lexical semantics acquisition measured by a novel-word learning task designed in this study. In the following, we will discuss the major results found.

### 5.1. AoFE effect manifested in episodic and semantic memory of newly acquired words

The MLR models show that the N400 amplitude effects elicited by the target words at the end of both recurrence type discourses and new-theme type discourses are significantly correlated with the participants' AoFE in a way that early AoFE leads to a stronger N400 effect. Based on the assumption that N400 reflects the performance of the lexical semantic learning, these results imply that earlier AoFE to English leads to an enhanced process of integrating unknown L2 concepts into both episodic and semantic memory.

The AoFE effect on episodic memory formation has been supported by the findings of previous relevant studies. Luk et al. (2011) found that the early bilingual learning experience of young adults enhances cognitive control, which was claimed to accelerate the retrieval process of learners' episodic memory (Barredo et al., 2015; Wagner, 2002). Similarly, Schroeder and Marian (2012) asked participants to perform several tasks on episodic memory and found that the bilingual learners who started their language learning at a relatively early age performed better in the recall tasks. Their research concluded that the higher performance of bilingual learners with earlier L2 is rooted in their superior executive functioning (Schroeder & Marian, 2012). In contrast to late bilingual learners, early bilingual learners have the advantage of higher efficiency in searching and locating relevant clues during episodic memory retrieval. From the perspective of the cognitive reserve hypothesis, being exposed to more than a single language from an early age can, to a large extent, alleviate the decline of cognitive ability caused by the ageing of the brain. Thus, bilingual learners' brain is suggested to be more endurable to degeneration of episodic memory during ageing (Angel et al., 2022; Perquin et al., 2013).

The AoFE effect on semantic memory formation can also be linked to previous models. There were very few models specifically focusing on the AoFE effect in L2, especially in L2 lexical-semantic processing. It has been suggested that AoFE effects in L1 and L2 are based on similar underlying mechanisms or processes (Hernandez & Li, 2007). Here, we propose a general account of the AoFE effect from the computational perspectives (Li et al., 2004; Seidenberg & Zevin, 2006; Smith et al., 2000): Through semantic connections, the meaning of novel words can be merged into the individual lexical-semantic network during novel word learning. According to the growing network model (Steyvers & Tenenbaum, 2005), the age at which a person starts to learn a new language influences the quality of their lexical-semantic network. The quality of the L2 lexical-semantic network determines learners' ability to infer novel word meanings through contextual learning, which leads to different learning outcomes. L2 learners with late AoFE have comparably less adequately formed L2

lexical semantic networks, which leads to less effective contextual learning and suboptimal establishment of associations between the novel and existing semantic nodes. On the contrary, learners with early AoFE have a well-developed L2 lexical-semantic network that supports effective contextual learning. In addition to the growing network model, the theory of neural plasticity also supports the effect of AoFE on semantic memory formation. Ellis and Lambon Ralph (2000) put forward the connectionist model, stating that early learners enjoy greater neural plasticity for the formation of a semantic network while starting late reduces plasticity. Such an effect of age-dependent plasticity brings great advantages to early learners in lexical semantic acquisition. Many previous neurocognitive studies that employed the N400 component as the neural marker to study novel word semantic learning (Borovsky et al., 2012; Liu & van Hell, 2020) did not study how this ERP component is affected by AoFE. Our study was designed to examine this specific question.

We now discuss the differential effects of AoFE on new-theme discourses and category-feature sentences. Novel concepts are acquired based on their various relations with existing concepts, thematically or categorically. Previous studies have not yet studied how the AoFE effect differs across thematic and taxonomic relations. One relevant line of study is that the strength of semantic relations has been shown to be influenced by age (Cicirelli, 1976; Kogan, 1974; Smiley & Brown, 1979). To be specific, the preference for thematic relations declines from childhood to adulthood but reverses after adulthood and regains its dominance for elderly people. In contrast, adults are more sensitive to taxonomic relations as compared to children and the elderly. These findings may explain why L2 learners with larger AoFE perform worse in connecting L2 novel concepts with existing nodes through thematic relations. To elaborate, when learners begin to learn English at a late age, their preference for thematic relations is lowered, and they tend to employ more taxonomic relations. This leads to further decreased intensity in the cultivation of the thematic relations in their L2 semantic network. Consequently, their understanding of the thematic relations is not as good as that of early learners, and they performed worse in developing and extending new thematic relations when acquiring unknown concepts in L2. From another perspective, the preference for learning via thematic relations can be consolidated in people at a young age and persists into adulthood, thus accounting for the result that lower AoFE leads to larger N400 effects for words seen in the thematic condition.

## 5.2. Independent effect of AoFE from AoTE

One of the key novel findings in the present study is that AoFE shows an independent effect that is not accounted for by AoTE. The pattern is consistent in both recurrence type and new-theme type discourses. If the learning process is built on a simple model in which the learning outcome is linearly correlated with the amount of time input to the learning, then the starting age would not exhibit a significant effect in the MLR model analysis. In other words, the learning outcome, which is quantified by the N400 effect, would be fully explained by AoTE. This result demonstrates that the contribution of the total amount of English exposure (AoTE) does not fully explain the variance in the learning outcomes of the participants. Instead, AoFE plays a unique role. Although it may be trivial to state that learning outcome depends on the starting age of learning because the brain's functions naturally decline during ageing,

this study focuses on young adult populations and provides statistical evidence that AoFE indeed plays an important role in L2 learning outcomes.

### 5.3. Proficiency effect

Aside from the AoFE and AoTE effects, it can also be found from the MLR results that the N400 effects that were elicited by the final words both in the new-theme discourses and in the category-feature sentences showed a statistically significant increase in the level of English proficiency. This pattern was not found in the recurrence type discourses. This result confirmed the unique role played by English proficiency in L2 learning. Here, one may argue that English proficiency should be a direct outcome of L2 learning and a direct indicator of the level of L2 achieved. Therefore, this factor should explain the majority of the new word learning as measured by the N400 effect, leaving little variance to be explained by AoFE and AoTE because AoFE and AoTE have 'injected' their effects into English proficiency – a variable that is expected to be the direct outcome of language learning and the direct determinant of N400 effects. To this point, our explanation is that the proficiency level assessed by the behavioral tests and questionnaires may not precisely capture the ability of new L2 word learning based on a multi-contextual learning scheme. In addition, different individuals might achieve an L2 proficiency level that is not accounted for by external factors (e.g., AoFE, AoTE) but by internal factors (e.g., language learning abilities determined genetically, learning habits and methods, etc.). Consequently, the factor of language proficiency contributes a unique amount of effect to the word learning performance.

### 5.4. Personality type effect

Another interesting finding from the study is that L2 learners' personality type measured by extroversion also has significant effects on new word learning performance. Specifically, the N400 effect (all discourse types combined) is stronger in individuals with higher personality scores (which means more extroverted). The personality effect is mainly located in the category-feature type sentences. This result indicates that extroverted character enhances L2 learners' learning of novel concepts, whereas introverted character impedes this process. This is largely in line with the intuition that extroversion is associated with more intense language communication, which is a defining feature of it. This relationship has been supported by findings of previous studies on the correlation between extroversion and L2 learning. In the aspect of the breadth of vocabulary knowledge, Golaghaei and Sadighi (2011) found that extroverted English learners performed better than introverted English learners. Ara (2015) also found that success in word learning can be largely attributed to the cultivation of positive personality skills. Overall, extensive previous research supported that extroversion positively contributes to L2 learning performance (Ghapanchi et al., 2011; Kayaoğlu, 2013; Pourfeiz, 2015; Wakamoto, 2000).

## 6. Limitation and future direction

The current study examined the association between AoFE and L2 lexical semantic learning performance and showed a significance between them modulated by the type of memory. The association cannot be fully accounted for by AoTE and other

major confounding factors, meaning that the starting age does play a unique role. The results strongly suggest a causal effect of AoFE on L2 learning. However, the nature of the current study could not demonstrate the causality relationship, which will be left for future studies (e.g., longitudinal research). As for the issue regarding critical points, our results only showed the existence of the association but did not identify the exact time point (if it exists). Future work needs to be done with a larger sample size and a more specific experimental design to address the issue of the existence of a critical point in AoFE on lexical semantics.

The testing session was arranged immediately after the learning session, which could be another limitation of the current study. This design aims to investigate whether the integration process of L2 novel concepts occurs immediately after learning, which may be argued to be implicated with the contribution of episodic memory. The consolidation effect on novel word learning over a longer time scale is not under the scope of the current research but will be an interesting topic for future studies. In the future, we can arrange the semantic testing of the L2 unknown concepts at least one day after the learning session. Alternatively, the semantic testing can be repetitively set on Day 1, Day 2 and one week after the learning session. In this way, we can investigate the dynamical courses in terms of how the semantic integration process of L2 unknown concepts lasts and how the role of thematic and taxonomic relations developed or diverged in the semantic network during this process.

It was not precisely controlled in how much time passed between the learning and testing sessions and during the testing and learning sessions. However, when the factors of session duration and inter-session interval were added as an independent variable into the multiple regression models, the results showed that the time spent on the learning and testing sessions, as well as the interval between the learning and testing sessions, did not have a significant effect on the N400 effect.

According to the current research results, there is no significant AoFE or AoTE effect on the N400 amplitudes elicited by the category-feature type sentences. Here, we discussed the possibility that led to the insignificant AoFE and AoTE effects on the N400 amplitudes under the category-feature condition. First, as mentioned before, it may be due to the insufficient trials that resulted in low statistical power. Second, we do agree that the difficulty level of learning discourses could be a compromised design. It is possible that more intense and quality learning will eventually lead to better acquisition of the categorical concepts. Our current work presented a scenario in which the establishment of taxonomic relations appears to be more complicated than the other ones. Third, the design of the two-choice option in the learning session is imperfect in ensuring high-quality and genuine word learning for the participant, as some participants would possibly use the content repetition to conduct the choice question. At the same time, participants may simply choose the alternative response option after failing in one attempt, which also raises concerns about the genuineness of the learning processes. While the two-option design is imperfect, participants cannot correctly choose the correct options, and the N400 effects would not exist in the testing session had the participants not learned the semantic meaning of the words from the learning discourses. In other words, if the two-choice question were presented before the discourses, their correct rate would be 50%. However, the participants had a very high correct rate in the two-choice questions (95.84%), which means most of them had acquired the meaning of the novel words before the first round of two-choice questions. Therefore, contextual learning still served as the foundation of their learning. We agree

that the two-choice question served as a confirmatory procedure, which might further enhance their memory and understanding of the novel words. However, the learning effects should still be primarily from contextual learning (if they did not learn, the memory could not be enhanced). Therefore, we think there was a sufficient amount of participants who managed to acquire the word meanings from learning discourses. The current design would, to a certain degree, be able to reflect individual differences in learning qualities. This limitation does not invalidate the ERP session because the N400 effect can only be elicited if the participant truly links the target word with the sentence context. Fourth, the acquisition of category features of novel concepts may tap into a more complex level of the learning process in the cognitive system, which requires more cognitive resources. In any case, the learning process should be improved in future studies. Improvement of learning quality may change the results regarding the category-feature sentences: the N400 effect in category-feature sentences may become significant if the design is improved.

In the testing session, it would be a good procedure to include a recognition test or a recall test to behaviorally examine if the participants can recall the meaning of the novel word or not. Not including this procedure is one of the limitations of our research. Our assumption was that the participants largely understood the meaning of the novel words, as reflected by the high correct rate in the two-choice question and by the reports from the 15 participants in the pilot study phase. The significant N400 effects can, to a certain extent, though not totally, prove that the acquisition of novel concepts occurred. The current results can still compare the degree of learning effects between the thematic and taxonomical relations (in our study, the former appeared to be stronger).

The last limitation that needs to be discussed regards the involvement of trials in our analysis. In our formal statistical analysis, as shown in the result section, we included all trials regardless of response correctness based on the assumption that relevant neural activity of language processing exists in both correct and incorrect trials. Therefore, to boost the statistical power, given the limited number of trials and low signal-to-noise ratio in neural activity data in general, we included all trials. We also analyzed the data with only correctly responded trials included and indeed found that the statistical results were not fully consistent with the results based on all trials (with some significant effects becoming insignificant). The primary reason we speculated for the inconsistency is the inadequacy of trial numbers, which led to insufficient power for the statistical results. This is, however, not directly testable in the current study due to the limited number of trials, which is noted as a limitation in the current study for the reader's information.

## 7. Conclusion

This research demonstrated the advantages of the early L2 learning experience in L2 new concept learning in the domain of lexical semantics. Early AoFE learners show stronger neural indicators of new concept learning performance, which implies that the effect of AoFE has an influence on the neural system in semantic processing. Most importantly, the AoFE effect is shown to be significant even after accounting for other major confounding factors, including AoTE, L2 proficiency, personality, and so forth. The AoFE effect in L2 is shown to be manifested through both the formation processes of episodic and thematic memory.

# References

Alptekin, C., & Erçetin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading*, 33(2), 206–219.

Angel, L., Guerrerro-Sastoque, L., Bernardo, M., Vanneste, S., Isingrini, M., Bouazzaoui, B., Kachouri, H., Fay, S., & Taconnat, L. (2022). Metamemory mediates the protective effect of cognitive reserve on episodic memory during aging. *Acta Psychologica*, 228, 103627.

Ara, J. (2015). *The effect of personality traits on adult learners' vocabulary learning* (Doctoral dissertation, BRAC University).

Arcos-Burgos, M., Lopera, F., Sepulveda-Falla, D., & Mastronardi, C. (2019). Neural plasticity during aging. *Neural Plasticity*, 2019, 1–3.

Baddeley, A. D., Eysenck, M. W., & Anderson, M. C. (2009). *Memory*. Psychology Press.

Bardovi-Harlig, K., & Dörnyei, Z. (1998). Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. *Tesol Quarterly*, 32(2), 233–259.

Barredo, J., Öztekin, I., & Badre, D. (2015). Ventral fronto-temporal pathway supporting cognitive control of episodic memory retrieval. *Cerebral Cortex*, 25(4), 1004–1019.

Bateman, L. B. (2014). Socioeconomic status, measurement. In *The Wiley Blackwell encyclopedia of health, illness, behavior, and society* (pp. 2227–2232). Wiley Online Library.

Benau, E. M., Morris, J., & Couperus, J. W. (2011). Semantic processing in children and adults: incongruity and the N400. *Journal of Psycholinguistic Research*, 40(3), 225–239.

Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. A. (2008). Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes*, 45(2), 122–159.

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development*, 8(3), 278–302.

Braveman, P. A., Cubbin, C., Egerter, S., Chideya, S., Marchi, K. S., Metzler, M., & Posner, S. (2005). Socioeconomic status in health research: One size does not fit all. *JAMA*, 294(22), 2879–2888.

Brito, N. H., & Noble, K. G. (2014). Socioeconomic status and structural brain development. *Frontiers in Neuroscience*, 8, 276.

Brysbaert, M. (2013). Lextale_FR a fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica*, 53(1), 23–37.

Burke, S. N., & Barnes, C. A. (2006). Neural plasticity in the ageing brain. *Nature Reviews Neuroscience*, 7(1), 30–40.

Butler, Y. G., & Hakuta, K. (2006). Bilingualism and second language acquisition. In *The handbook of bilingualism* (pp. 114–144). Wiley Online Library.

Chaumon, M., Bishop, D. V., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, 250, 47–63.

Chen, S., Wang, L., & Yang, Y. (2014). Acquiring concepts and features of novel words by two types of learning: Direct mapping and inference. *Neuropsychologia*, 56, 204–218.

Cicirelli, V. G. (1976). Categorization behavior in aging subjects. *Journal of Gerontology*, 31(6), 676–680.

Cowan, N. (2012). *Working memory capacity*. Psychology Press.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3(4), 422–433.

Deacon, D., Dynowska, A., Ritter, W., & Grose-Fifer, J. (2004). Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. *Psychophysiology*, 41(1), 60–74.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134, 9–21.

Dimigen, O. (2020). Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. *NeuroImage*, 207, 116117.

Ding, J., Chen, S., Wang, L., & Yang, Y. (2017). Thematic and taxonomic relations of novel words learned from action and perceptual features. *Journal of Neurolinguistics*, 41, 70–84.

Dollmann, J., Kogan, I., & Weißmann, M. (2020). Speaking accent-free in L2 beyond the critical period: The compensatory role of individual abilities and opportunity structures. *Applied Linguistics*, 41(5), 787–809.

Elgort, I., Perfetti, C. A., Rickles, B., & Stafura, J. Z. (2015). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and event-related brain potential (ERP) indicators of learning. *Language, Cognition and Neuroscience*, 30(5), 506–528.

Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, 64(2), 365–414.

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.

Ferreira, R. A., & Ellis, A. W. (2016). Effects of contextual diversity on semantic decision and reading aloud: evidence from a word learning study in English as a second language/Efectos de la diversidad contextual en la decisión semántica y la lectura en voz alta: evidencia de un estudio de aprendizaje de palabras en inglés como segundo idioma. *Estudios de Psicología*, 37(1), 162–182.

Flege, J. E. (2019). A non-critical period for second-language learning. In *A sound approach to language matters: In honor of Ocke-Schwen Bohn*. Aarhus University.

Frances, C., Martin, C. D., & Duñabeitia, J. A. (2020). The effects of contextual diversity on incidental vocabulary learning in the native and a foreign language. *Scientific Reports*, 10(1), 1–11.

Ghapanchi, Z., Khajavy, G. H., & Asadpour, S. F. (2011). L2 motivation and personality as predictors of the Second language proficiency: Role of the Big Five traits and L2 motivational self System. *Canadian Social Science*, 7(6), 148–155.

Golaghaei, N., & Sadighi, F. (2011). Extroversion/introversion and breadth of vocabulary knowledge. *Director*, 1(3), 69.

Goldwater, M. B., Markman, A. B., & Stilwell, C. H. (2011). The empirical case for role-governed categories. *Cognition*, 118(3), 359–376.

Gullifer, J. W., Kousaie, S., Gilbert, A. C., Grant, A., Giroud, N., Coulter, K., Klein, D., Baum, S., Phillips, N., & Titone, D. (2021). Bilingual language experience as a multidimensional spectrum: Associations with objective and subjective language proficiency. *Applied Psycholinguistics*, 42(2), 245–278.

Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, 13(2), 65–73.

Hajra, S. G., Liu, C. C., Song, X., Fickling, S. D., Cheung, T. P., & D'Arcy, R. C. (2018). Accessing knowledge of the 'here and now': A new technique for capturing electromagnetic markers of orientation processing. *Journal of Neural Engineering*, 16(1), 016008.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, 14(1), 25–38.

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, 177, 263–277.

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin*, 133(4), 638.

Hulstijn, J. H. (2012). The construct of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism: Language and Cognition*, 15(2), 422–433.

Hurtado, L. M., & Estrada, C. (2010). Factors influencing the second language acquisition of Spanish vibrants. *The Modern Language Journal*, 94(1), 74–86.

Jalali-Moghadam, N., & Kormi-Nouri, R. (2017). Bilingualism and reading difficulties: An exploration in episodic and semantic memory. *Journal of Cognitive Psychology*, 29(5), 570–582.

Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review*, 23(4), 1214–1220.

Jones, M., & Love, B. C. (2007). Beyond common features: The role of roles in determining similarity. *Cognitive Psychology*, 55(3), 196–231.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching*, 44(2), 137–166.

Kasparian, K., & Steinhauer, K. (2016). Confusing similar words: ERP correlates of lexical-semantic processing in first language attrition and late second language acquisition. *Neuropsychologia*, 93, 200–217.

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, 16(4), 705–710.

Kayaoğlu, M. N. (2013). Impact of extroversion and introversion on language-learning behaviors. *Social Behavior and Personality: An International Journal*, 41(5), 819–825.

Kezwer, P. (1987). The extroverted vs. the introverted personality and second language learning. *TESL Canada Journal*, 5, 45–58.

Kogan, N. (1974). Categorizing and conceptualizing styles in younger and older adults. *Human Development*, 17(3), 218–230.

Kormi-Nouri, R., Moniri, S., & Nilsson, L. G. (2003). Episodic and semantic memory in bilingual and monolingual children. *Scandinavian Journal of Psychology*, 44(1), 47–54.

Kormi-Nouri, R. E. Z. A., Shojaei, R. S., Moniri, S., Gholami, A. R., Moradi, A. R., Akbari-Zardkhaneh, S. A. E. E. D., & Nilsson, L. G. (2008). The effect of childhood bilingualism on episodic and semantic memory tasks. *Scandinavian Journal of Psychology*, 49(2), 93–109.

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, 109(2–3), 68–74.

Kotz, S. A., & Elston-Güttler, K. (2004). The role of proficiency on processing categorical and associative information in the L2 as revealed by reaction times and event-related brain potentials. *Journal of Neurolinguistics*, 17(2–3), 215–235.

Kotz, S. A., & Elston-Güttler, K. E. (2007). Bilingual semantic memory revisited: ERP and fMRI evidence. In *The neural basis of semantic memory* (pp. 105–132). Cambridge University Press.

Kovelman, I., Baker, S. A., & Petitto, L. A. (2008). Age of first bilingual language exposure as a new window into bilingual reading development. *Bilingualism: Language and Cognition*, 11(2), 203–223.

Krieger, N., Williams, D. R., & Moss, N. E. (1997). Measuring social class in US public health research: Concepts, methodologies, and guidelines. *Annual Review of Public Health*, 18(1), 341–378.

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, 28(1), 40–80.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207(4427), 203–205.

Larsen-Freeman, D., & Long, M. H. (2014). *An introduction to second language acquisition research*. Routledge.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920–933.

Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*, 17(8–9), 1345–1362.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General*, 130(1), 3.

Liu, X., Tu, L., Wang, J., Jiang, B., Gao, W., Pan, X., Li, M., Zhong, M., Zhu, Z., Niu, M., Li, Y., Zhao, L., Chen, X., Liu, C., Lu, Z., & Huang, R. (2017). Onset age of L2 acquisition influences language network in early and late Cantonese-Mandarin bilinguals. *Brain and Language*, 174, 16–28.

Liu, Y., & van Hell, J. G. (2020). Learning novel word meanings: An ERP study on lexical consolidation in monolingual, inexperienced foreign language learners. *Language Learning*, 70(S2), 45–74.

Lucio, R. (2020). Critical Period Controversies for Second Language Acquisition: Implications for Language Teaching.

Luk, G., De Sa, E. R. I. C., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, 14(4), 588–595.

Marinova-Todd, S. H., Mayo, M., & Lecumberri, M. (2003). Know your grammar: what the knowledge of syntax and morphology in an L2 reveals about the critical period for Second/Foreign language acquisition. In *Age and the acquisition of English as a foreign language* (pp. 59–76). Clevedon.

Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental & Theoretical Artificial Intelligence*, 13(4), 329–358.

Mathalon, D. H., Faustman, W. O., & Ford, J. M. (2002). N400 and automatic semantic processing abnormalities in patients with schizophrenia. *Archives of General Psychiatry*, 59(7), 641–648.

Mirman, D., Landrigan, J. F., & Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin*, 143(5), 499.

Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *Tesol Quarterly*, 37(4), 645–670.

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9–13.

Nelson, K. (1988). Where do taxonomic categories come from? *Human Development*, 31(1), 3–10.

Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. T. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of English. *Journal of Cognitive Neuroscience*, 24(5), 1205–1223.

Oakes, J. M., & Andrade, K. E. (2017). The measurement of socioeconomic status. *Methods in Social Epidemiology*, 2, 23–42.

Ojima, S., Matsuba-Kurita, H., Nakamura, N., Hoshino, T., & Hagiwara, H. (2011). Age and amount of exposure to a foreign language during childhood: Behavioral and ERP data on the semantic comprehension of spoken English by Japanese children. *Neuroscience Research*, 70(2), 197–205.

Ouyang, G., Dien, J., & Lorenz, R. (2022). Handling EEG artifacts and searching individually optimal experimental parameter in real time: A system development and demonstration. *Journal of Neural Engineering*, 19(1), 016016.

Pagán, A., & Nation, K. (2019). Learning words via reading: Contextual diversity, spacing, and retrieval effects in adults. *Cognitive Science*, 43(1), e12705.

Perquin, M., Vaillant, M., Schuller, A. M., Pastore, J., Dartigues, J. F., Lair, M. L., Diederich, N., & Group, M. V. (2013). Lifelong exposure to multilingualism: new evidence to support cognitive reserve hypothesis. *PloS One*, 8(4), e62030.

Pourfeiz, J. (2015). Exploring the relationship between global personality traits and attitudes toward foreign language learning. *Procedia-Social and Behavioral Sciences*, 186, 467–473.

Qiufang, W. (2003). *The road to success in English learning.* Shanghai Foreign Language Education Press.

Raizada, R. D., & Kishiyama, M. M. (2010). Effects of socioeconomic status on brain development, and how cognitive neuroscience may contribute to leveling the playing field. *Frontiers in Human Neuroscience*, 4, 1075

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, 111(1), 205.

Saville-Troike, M., & Barto, K. (2017). *Introducing second language acquisition.* Cambridge University Press.

Schroeder, S. R., & Marian, V. (2012). A bilingual advantage for episodic memory in older adults. *Journal of Cognitive Psychology*, 24(5), 591–601.

Schroeder, S. R., & Marian, V. (2014). Bilingual episodic memory: How speaking two languages influences remembering. In *Foundations of bilingual memory* (pp. 111–132). Springer.

Seidenberg, M. S., & Zevin, J. D. (2006). Connectionist models in developmental cognitive neuroscience: Critical periods and the paradox of success. In *Processes of change in brain and cognitive development: Attention and performance* (Vol. XXI, pp. 315–347). Oxford University Press.

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology*, 46(1), 561–584.

Singleton, D. M., & Ryan, L. (2004). *Language acquisition: The age factor* (Vol. 9). Multilingual Matters.

Slabakova, R. (2006). Is there a critical period for semantics? *Second Language Research*, 22(3), 302–338.

Smiley, S. S., & Brown, A. L. (1979). Conceptual preference for thematic or taxonomic relations: A nonmonotonic age trend from preschool to old age. *Journal of Experimental Child Psychology*, 28(2), 249–257.

Smith, M., Cottrell, G., & Anderson, K. (2000). The early word catches the weights. *Advances in Neural Information Processing Systems*, 13, 52–58.

Squire, L. R., & Zola, S. M. (1998). Episodic memory, semantic memory, and amnesia. *Hippocampus*, 8(3), 205–211.

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.

Tamminen, J., & Gaskell, M. G. (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *Quarterly Journal of Experimental Psychology*, 66(5), 1001–1025.

Thomas, M. S., Forrester, N. A., & Ronald, A. (2013). Modeling socioeconomic status effects on language development. *Developmental Psychology*, 49(12), 2325.

Tulving, E. (1972). *Episodic and semantic memory*. Academic Press.

Wagner, A. D. (2002). Cognitive control and episodic memory. *Neuropsychology of Memory*, 3, 174–192.

Wakamoto, N. (2000). Language learning strategy and personality variables: Focusing on extroversion and introversion. *IRAL: International Review of Applied Linguistics in Language Teaching*, 38(1), 71.

Wen, Z. (Edward), & Li, S. (2019). Working Memory in L2 Learning and Processing. In J. W. Schwieter & A. Benati (Eds.), *The Cambridge Handbook of Language Learning* (pp. 365–389). Cambridge: Cambridge University Press.

Winkler, I., Debener, S., Müller, K. R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. In *2015 37th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 4101–4105). IEEE.

Yee, E., Chrysikou, E. G., & Thompson-Schill, S. L. (2014). 17 Semantic Memory. *The Oxford Handbook of Cognitive Neuroscience, Volume 2: The Cutting Edges*, 2, 353.

Yee, E., Overton, E., & Thompson-Schill, S. L. (2009). Looking for meaning: Eye movements are sensitive to overlapping semantic features, not association. *Psychonomic Bulletin & Review*, 16(5), 869–874.

Zhang, M., Ding, J., Li, X., & Yang, Y. (2019). The impact of variety of episodic contexts on the integration of novel words into semantic network. *Language, Cognition and Neuroscience*, 34(2), 214–238.

Zoghi, M., Kazemi, S. A., & Kalani, A. (2013). The effect of gender on language learning. *Journal of Novel Applied Sciences*, 2(4), 1124–1128.