CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Processing of novel L2 compounds across repeated exposures during reading: A growth curve analysis

Wei Yi (ID)

School of Chinese as a Second Language, Peking University, Beijing, China
Email: weiyisla@pku.edu.cn

Abstract

Using eye tracking, this study examined L2 learners' real-time processing of novel compounds across repeated exposures during reading. Sixty-one L2 speakers of Chinese read 12 stories over two days. Unbeknown to them, 12 novel compounds were embedded, each occurring six times. Growth curve analyses showed that semantic transparency, working memory capacity, and morphological awareness had no impact on fixation durations for the novel compounds. However, participants with a larger L2 vocabulary size processed novel opaque compounds significantly faster than those with a smaller L2 vocabulary size. For both transparent and opaque compounds, first fixation durations did not change across exposures, yet similar curvilinear decreasing patterns were found in gaze duration and total reading time, with the rates of decrease moderated by L2 vocabulary size and working memory capacity, respectively. Taken together, such findings provide converging evidence supporting the incidental nature of vocabulary learning through natural reading.

Keywords: incidental vocabulary learning; reading; compounds; eye tracking; individual differences

Incidental vocabulary learning, especially through reading, is an important source to promote the growth of L2 vocabulary (Rott, 1999; Schmitt, 2008). Current literature on this topic has mainly focused on the quantity of vocabulary knowledge that can be gained from reading, leaving it largely unknown how L2 learners process novel words in real time under incidental conditions. Vocabulary learning is an incremental process (Fukkink et al., 2001; Schmitt, 1998) that involves complex interactions between lexical characteristics and individual learner differences (Peters, 2020). Nevertheless, little research has been done to examine L2 learners' processing of unfamiliar lexical items across repeated exposures, as well as how lexical characteristics and individual learner differences might contribute to such processes. Additionally, most studies on incidental vocabulary learning have focused on English, with little research being carried out in other languages, such as Chinese. To address these gaps, the present study adopted the eye-tracking

CrossMark

technique to explore L2 Chinese learners' processing of novel compounds, which varied in semantic transparency and were encountered repeatedly during natural reading. Meanwhile, it also considered the influences of three main types of individual differences, namely L2 vocabulary size, working memory capacity, and morphological awareness.

Novel L2 word processing during incidental vocabulary learning through reading

Vocabulary learning takes place either intentionally or incidentally. In the field of second language acquisition (SLA), the notions of intentional and incidental vocabulary learning are generally distinguished in the operational sense. Under incidental learning conditions, L2 learners are engaged with meaning-focused activities, whereas intentional learning activities mostly focus on linking L2 form to meaning (Webb, 2020). Additionally, under incidental learning conditions, L2 learners are not forewarned of the existence of subsequent vocabulary tests, whereas they are instructed that vocabulary tests will follow the learning session under intentional learning conditions (Uchihara et al., 2019).

In the past few decades, SLA has witnessed a steadily growing interest in incidental vocabulary learning, especially through extensive reading (Schmitt, 2008; Waring & Nation, 2004; Webb, 2020). Incidental vocabulary learning through reading is conceptualized as a process in which L2 learners' primary focus is on the meaning of texts, with vocabulary knowledge acquired as a by-product of the reading experience (Hulstijn, 2003). Overall, most research efforts on this topic have taken a product-oriented approach, investigating research questions such as the amount of vocabulary knowledge that can be gained from extensive reading and the number of repeated exposures needed to achieve measurable vocabulary gains (for meta-analyses, see Swanborn & de Glopper, 1999; Uchihara et al., 2019). By contrast, only a few studies (Elgort et al., 2018; Godfroid et al., 2018; Joseph, Wonnacott, Forbes & Nation, 2014; Mohamed, 2018; Pellicer-Sánchez, 2016) have been carried out to examine how L2 learners process unknown lexical items during reading. Unlike intentional vocabulary learning, in which L2 learners' attention was drawn to unknown lexical items, online processing of novel L2 words under incidental learning conditions, such as reading, might be quite different and highly variable due to the following reasons. First, L2 learners' top priority is to understand the textual meaning. As a result, they may not pay as much attention to the form, meaning, and form-meaning association of novel words (Schmitt, 2008) as they do under intentional learning conditions. Second, under incidental learning conditions, L2 learners have no access to word definitions or explanations. Instead, they have to infer the meaning of novel words based on sublexical, lexical, and contextual cues. Third, just as lexical characteristics and individual learner differences have been found to moderate incidental vocabulary gains (e.g., Elgort & Warren, 2014; Uchihara et al., 2019), such factors may also influence how novel L2 words are processed in real time by L2 learners, especially in terms of the amount of attention paid to unfamiliar lexical items and the processing strategies they may apply (Webb, 2020). Lastly, vocabulary learning is incremental (Fukkink et al., 2001; Schmitt,

1998), meaning that repeated exposures are generally required before a new word can be fully mastered. Compared with intentional vocabulary learning, incidental vocabulary learning, such as through reading, may demand more encounters with novel lexical items. Moreover, given the incidental nature of the input, online processing of unknown L2 words may follow intriguing patterns across repeated exposures (e.g., Godfroid et al., 2018; Mohamed, 2018; Pellicer-Sánchez, 2016).

The role of repeated exposures

In recent years, a handful of studies on incidental L2 vocabulary learning through reading have adopted the eye-tracking technique to investigate whether online processing of novel words is associated with vocabulary gains (e.g., Elgort et al., 2018; Godfroid, Boers & Housen, 2013; Godfroid et al., 2018; Mohamed, 2018; Pellicer-Sánchez, 2016). According to the E-Z Reader model (Reichle et al., 1998; Reichle et al., 2006)—one of the most influential theoretical frameworks of eye movement control during reading—lexical processing consists of three sequential stages, namely familiarity check of orthographic forms, access to lexical (phonological/semantic) information, and integration of word meaning with sentential/discourse contexts. Such cognitive processes can be captured by different eye-tracking measures. Specifically, first fixation duration (i.e., the duration of the first fixation on a word) is assumed to indicate familiarity check of lexical items (Juhasz & Rayner, 2003), whereas gaze duration (i.e., the sum of the duration of all fixations made on a word before the eyes move away from it) and total reading time (i.e., the sum of all fixation durations made on a word) are believed to reflect lexical access and post-lexical semantic integration at the discourse level, respectively. Following such assumptions, SLA researchers have used different combinations of eye-tracking measures and found that various eye-tracking measures are associated with different aspects of word knowledge acquired incidentally from reading. Mohamed (2018) had advanced English L2 speakers read pseudowords embedded in a graded reader. The results showed that first fixation duration was a significant predictor of L2 participants' form recognition of pseudowords, whereas gaze duration was associated with meaning recall of such lexical items. Positive relationships between total reading time and incidental vocabulary gains have also been reported in the current literature. Mohamed (2018) found that total reading time spent on pseudowords predicted L2 learners' performance on all vocabulary knowledge measures, including form recognition, meaning recall, and meaning recognition. Following experimental designs similar to that in Mohamed (2018), Pellicer-Sánchez (2016) and Godfroid et al. (2018) found that longer total reading times on target lexical items led to better meaning recall performance of L2 learners. Additionally, Godfroid and colleagues (Godfroid et al., 2018) reported that total reading time independently contributed to L2 learners' recognition of word meaning.

Vocabulary learning takes place incrementally (Fukkink et al., 2001; Schmitt, 1998). Therefore, to reveal a complete picture of L2 learners' processing of novel words during reading, one non-negligible aspect is to examine how unfamiliar words are processed in real time across repeated exposures. Joseph et al. (2014) investigated incidental vocabulary learning by embedding English nonwords into individual sentences. L1 speakers of English were recruited to read the sentences

for meaning over a five-day period. Following the exposure phase, a surprise meaning recognition test and a reading test were administered, in which participants read the novel nonwords in semantically neutral sentences. Statistical analysis revealed that reading times for novel nonwords decreased over exposures, and early presented targets were processed for longer time than late-presented ones. Such findings were replicated by later L2 studies (Elgort et al., 2018; Mohamed, 2018; Pellicer-Sánchez, 2016). Mohamed (2018) and Pellicer-Sánchez (2016) reported that fixation durations on target pseudowords embedded in texts decreased across repeated exposures, with longer reading times spent on earlier encounters. Focusing on the first eight encounters, Elgort and colleagues (Elgort et al., 2018) found that L2 learners' reading times on target low-frequency words decreased steadily across exposures. These findings suggest that online processing of novel L2 words goes through certain decreasing patterns over repeated exposures. According to Mirman (2014), cognitive processing over time is generally nonlinear. Nevertheless, the above-mentioned research all adopted linear models when running statistical analyses, leaving it unclear whether the decreasing patterns of processing of novel words during reading are nonlinear. As far as the author knows, Godfroid et al. (2018) is the only study that has addressed this issue. Using growth curve analyses, Godfroid and colleagues concluded that processing of novel words over exposures followed nonlinear, S-shaped curves—that is, reading times decreased sharply after the first few encounters and were followed by a plateau, and then a further, more gradual decrease.

Compounding and semantic transparency

Words vary with respect to visual, phonological, orthographic, semantic, and syntactic characteristics, which can make vocabulary learning more or less difficult (Schmitt, 2019). Laufer (1990) listed a number of intralexical factors that may moderate the difficulty of L2 vocabulary learning, including pronounceability, length, part of speech, inflectional/derivational complexity, abstractness, specificity, idiomaticity, register restriction, and number of meanings. More recently, Peters (2020) added additional factors to this list, such as cognateness and synforms (i.e., words with similar sound/spelling/meaning, originally put forward by Laufer, 1988). Words can be monomorphemic or multimorphemic. Specifically, morphologically complex words are created through three different ways, namely inflection (e.g., *walk-walks-walked*), derivation (e.g., *happy-happiness-happily-unhappy*), and compounding (e.g., *snow-man-snowman*). Laufer and Peters focused on English and did not mention compounds as well as lexical properties central to the acquisition and processing of such lexical items, probably because compounding is not the dominant principle for word formation in English. However, in many other languages (e.g., Chinese, German, and Finnish), compounding is quite productive. Compounds are different from other types of words in that they consist of multiple free morphemes, with the semantic relationship between the morphemes and the compound varying in the degree of transparency (Libben, 1998). For semantically transparent compounds (e.g., *snowball*), the meaning of the word is the combination of the meanings of its constituent morphemes. By contrast, for

semantically opaque compounds (e.g., *honeymoon*), the meaning of the word cannot be derived from the constituent morphemes.

Research has shown that morphologically complex words are represented and processed differently from monomorphemic words (Zhou & Marslen-Wilson, 2000). Libben (1998) proposed a seminal model of representation and processing of semantically transparent and opaque compounds, in which he outlined three distinct levels of structure. At the stimulus level, morphological parsing is assumed to take place from left to right recursively for novel compounds, thus isolating each constituent morpheme before checking their lexical and orthographic status. At the lexical level, lexical representations of familiar compounds, as well as connections between their constituent morphemes, are activated. Lastly, at the conceptual level, the notion of semantic transparency is represented, with the semantic componentiality between constituent morphemes and the whole compound being recognized. Many studies (e.g., Juhasz, 2007; Pollatsek & Hyönä, 2005) have found that decomposition occurs for both transparent and opaque compounds. Following Libben's model, if compounds are comprehended by decomposing and combining the constituents' meanings, then one would expect that opaque compounds will be more difficult to process than transparent compounds. Such semantic transparency effects have been observed by a plethora of eye-tracking studies, focusing on the processing of familiar compounds by L1 speakers during reading. Underwood et al. (1990) embedded English transparent and opaque compounds in sentential contexts that primed either the constituent morphemes or the whole compound. They found that transparent compounds were read faster than opaque compounds in gaze duration, regardless of context type. Similarly, Juhasz (2007) found that L1 speakers of English read semantically opaque compounds embedded in sentences for significantly longer time compared with transparent compounds, in terms of gaze duration and go-past duration. Frisson et al. (2008) found no significant difference between the processing of unspaced opaque and transparent compounds. However, when a space was inserted between constituent morphemes (i.e., the compounds were presented as two separate words), a semantic transparency effect was observed in gaze duration. Brusnighan and Folk (2012) investigated the role of semantic transparency in cognitive processing of novel compounds. In their study, L1 speakers of English were instructed to read individual sentences with novel compounds embedded. Results showed that novel opaque compounds in informative contexts did not receive significantly longer gaze durations than novel transparent compounds. However, skilled L1 readers did spend significantly less time rereading novel transparent compounds than novel opaque compounds in informative contexts, which indicates that both contextual information and semantic transparency played a role in the processing of unfamiliar compounds. As far as the author knows, not a single study has examined how novel transparent and opaque compounds are processed by L2 learners under incidental learning conditions, such as reading.

Compound word learning in Chinese as an L2

Chinese relies heavily on compounding to construct words. More than 70% of Chinese words are compounds, among which 73.6% are disyllabic (Institute of Language Teaching and Research, 1986). Given that a morpheme in Chinese

generally corresponds to a syllable in spoken form and to a character in written form (Zhou et al., 1999), one would not be surprised to find out that most Chinese compounds consist of two free morphemes (i.e., two characters in written form). Unlike English compounds (e.g., *blackboard*), compounds in Chinese do not have primary stress on the first constituent morpheme. Moreover, orthographic rules such as the use of hyphens (e.g., *son-in-law*) or space (e.g., *prime minister*) also do not apply to Chinese compounds (Dronjic, 2011). As a result, compounds in Chinese may be less salient to readers than English compounds in the visual or auditory modality. As already mentioned, compounds can be semantically transparent or opaque depending on the degree to which the meaning of the compound can be predicted from its constituent morphemes. According to an analysis by J. Li (2011), 28% of bi-morphemic compounds in Chinese are fully transparent (e.g., 果-园/*fruit-garden: fruit garden*), whereas 72% of bi-morphemic Chinese compounds are more or less opaque (e.g., 水-母/*water-mother: jellyfish*). Taken together, these statistics clearly demonstrate the central role of compounds in the learning of Chinese as an L2, as well as the importance of semantic transparency as a factor affecting the acquisition of Chinese compounds, which has been supported by many studies. Gan (2008) found that intermediate-level Chinese L2 learners' inferencing of transparent compounds was significantly better than that for opaque compounds, when contextual support was not provided. Similarly, Chen (2021) reported that L2 Chinese learners' recall of transparent compounds was better than that for opaque compounds, after explicitly learning such items without any contextual support. Hong et al. (2017) had Korean learners of Chinese read compounds in sentential contexts. Their results showed that L2 learners' lexical inferencing performance was much worse for opaque compounds compared with that for transparent compounds, especially when sentential contexts were uninformative. To date, little research has been carried out on incidental learning of compounds through reading in non-English languages. However, based on the Chinese studies reviewed above, one might expect to replicate the semantic transparency effects widely reported in English eye-tracking studies (e.g., Underwood et al., 1990; Juhasz, 2007; Frisson et al., 2008; Brusnighan & Folk, 2012), when examining L2 learners' processing of Chinese compounds under reading-for-meaning conditions.

Individual learner differences

Language learners vary enormously in social, cognitive, and affective aspects (Ellis, 2004). In a meta-analysis, Uchihara et al. (2019) found that learner variables, including age and L2 vocabulary knowledge, significantly moderate the effect of repeated exposures on incidental vocabulary learning through different input modes. When narrowing down to incidental vocabulary learning through reading, few studies have considered individual differences as contributors to vocabulary gains as well as online processing of novel L2 words. In this section, three types of individual learner differences, namely L2 vocabulary size, working memory capacity, and morphological awareness, which may play critical roles in incidental vocabulary learning through reading—especially for compounds—are briefly reviewed.

L2 vocabulary size

Having a larger vocabulary size in L2 has been found to result in positive effects on learning words incidentally from reading. Horst et al. (1998) revealed that participants' L2 vocabulary size was moderately associated with vocabulary gains from reading ($r = .36$), after they were exposed to target words in authentic reading materials. Similarly, Webb and Chang (2015) demonstrated that EFL learners with a larger L2 vocabulary size learned significantly more words from extensive reading measured at immediate and delayed post-tests, compared with those with a smaller L2 vocabulary size. Zahar et al. (2001) found that EFL learners with a larger L2 vocabulary size needed to encounter a target word fewer times before they were acquired incidentally through reading, compared with those with a smaller L2 vocabulary size. A similar conclusion was also made in the meta-analysis by Uchihara et al. (2019, p.584). Despite the fact that the benefit of L2 vocabulary size for incidental vocabulary learning through reading seems indisputable, it remains unclear how L2 vocabulary size affects second language learners' processing of novel words in real time during reading.

Working memory capacity

Working memory capacity is another crucial individual learner difference that is worth considering for vocabulary learning. Baddeley (2000) conceptualized working memory as a multi-component cognitive system that consists of four elements: the central executive, the phonological loop, the visuo-spatial sketchpad, and the episodic buffer. The central executive is responsible for the control and regulation of cognitive processes, while the phonological loop and the visual-spatial sketchpad deal with the retention and manipulation of verbal and visual/spatial information, respectively. The verbal memory system is often divided into working memory and phonological short-term memory (PSTM)—the former entails both storage and manipulation of information, whereas the latter only concerns phonological storage. PSTM is typically measured by simple tasks, such as the digit span, word span, or nonword span test. By contrast, working memory capacity is often measured by complex tasks, such as the reading span or operation span test (Conway et al., 2005; Linck et al., 2014).

Evidence suggests both working memory and PSTM are correlated with language aptitude (S. Li, 2016), with research demonstrating that memory capacity plays a significant role in various domains of L2 acquisition (for a review, see Williams, 2012). Bisson et al. (2021) examined the role of working memory and PSTM in monolingual English native speakers' learning of Welsh words under incidental and intentional learning conditions. L1 participants were presented with the written and auditory forms of Welsh words as well as pictures depicting the meaning of the words in both conditions. Instead of being explicitly told to memorize the lexical items as in the intentional learning condition, participants were required to judge whether a letter presented to them appeared in a word or not under the incidental learning condition. The authors computed composite memory scores as an indicator of general memory processes, which included both working memory and PSTM.

Their results showed that general memory mechanisms were involved in vocabulary learning under incidental and intentional learning conditions. Moreover, they concluded that memory played a bigger role in the intentional learning situation. Compared with studies reporting significant effects of working memory and PSTM on L2 learners' vocabulary gains under intentional learning conditions (e.g., Martin & Ellis, 2012; Service & Kohonen, 1995; Speciale et al., 2004), research on the role of memory capacity in incidental L2 vocabulary learning—especially through reading—is still lacking. Montero Perez (2020) investigated the impact of working memory and PSTM on L2 learners' pickup of new words while viewing a French documentary that contained 15 pseudowords. She found that working memory—instead of PSTM—positively correlated with L2 learners' performance on immediate form and meaning recognition tests. Daneman and Hannon (2007) suggest that reading comprehension processes, including remembering new information, making inferences about new information, and integrating accessed knowledge with new information, all draw heavily on working memory capacity. Following this, working memory capacity may play a critical role in L2 learners' incidental learning of words from reading. Malone (2018) addressed L2 learners' incidental learning of English words with or without aural enhancement. Both PSTM and working memory capacity significantly correlated with form recognition, while only PSTM significantly correlated with meaning recognition. Furthermore, such relationships were moderated by treatment, with working memory capacity accounting for more variability in form recognition for participants receiving aural enhancement of the written input. Yi, Lu and DeKeyser (in press) also examined the influence of PSTM and working memory capacity on L2 learners' acquisition of novel English nouns in a sentence-reading experiment. However, neither PSTM nor working memory capacity predicted L2 learners' performance on any vocabulary knowledge tests. Yang et al. (2017) further explored the effects of post-reading word-focused activities on L2 learners' vocabulary learning through reading. After reading a short text, participants were assigned to one of three post-reading activities, in which they had to write a sentence using each of the eight target words, complete a gap-filling task by selecting the appropriate words from a list, or simply answer an essay question that did not require the use of the target words. Interestingly, working memory capacity was found to predict immediate vocabulary gains of those who received the gap-filling or question-answering task. Due to the lack of research and the inconsistencies in the current literature, only tentative predictions can be put forward with respect to the role of memory capacity in incidental L2 vocabulary learning through reading. Working memory has been found to be more strongly associated with language aptitude than PSTM (S. Li, 2016), and it is expected to be more likely to predict vocabulary gains under incidental learning conditions than PSTM (e.g., Peters, 2020). Given that vocabulary learning takes place incidentally during natural reading and that working memory/PSTM has been reported to play a bigger role in intentional learning than in incidental learning (e.g., Bisson et al., 2021), it is possible that neither working memory nor PSTM would significantly affect L2 learners' learning and processing of novel lexical items under reading-for-meaning conditions.

Morphological awareness

Research on morphological awareness and its influence on incidental vocabulary learning is lacking. Morphological awareness refers to one's grasp of the morphological structure of words and is usually operationalized as the ability to segment words into smaller, functionally identifiable units (Koda, 2000). Morphological awareness can be classified into three categories, namely inflectional awareness (e.g., Rothou & Padeliadu, 2015), derivational awareness (e.g., Kieffer & Box, 2013), and compounding awareness (e.g., McBride-Chang et al., 2008). Understanding how words are formed may be crucial for comprehending and learning new words (McBride-Chang et al., 2008; for a meta-analysis, see Ke et al., 2021). For languages such as English, in which morphologically complex words are created mainly through derivation or inflection, gaining derivational or inflectional knowledge helps L2 learners expand their lexical knowledge (Iwaizumi & Webb, 2021; Sasao & Webb, 2017). To diagnose L2 learners' weakness and improve their ability to learn words, Sasao and Webb (2017) developed a word part levels test (WPLT) to measure three aspects of derivational knowledge, namely recognition of affix forms, recognition of affix meanings, and recognition of the part of speech that an affix makes. Iwaizumi and Webb (2021) further investigated the relationship between L2 learners' derivational knowledge and their receptive vocabulary knowledge in English. In their study, L2 English learners at varying vocabulary levels were required to write as many derivatives (e.g., *artist/artistic*) as possible for each given headword (e.g., *art*). In line with earlier research, they found a positive association between L2 learner's ability to produce derivative forms and their receptive vocabulary knowledge. Similarly, Kieffer and Lesaux (2012) also found a strong positive relationship between the development of L2 derivational awareness and the growth of L2 vocabulary knowledge among fourth-to-seventh grade Spanish learners of English. For languages such as Chinese, in which compounding is the primary way of creating words, compounding awareness is critical for L1 children's early vocabulary development (McBride-Chang et al., 2008). Moreover, it has also been revealed to play an important role in L2 vocabulary growth, lexical inferencing, and reading comprehension. Chen (2018) found that L2 morphological awareness significantly contributed to lexical inferencing with and without contextual cues for skilled L2 learners. Using the structural equation modeling technique, Wu (2017) also showed that L2 Chinese learners' morphological awareness influenced reading comprehension through the mediation of vocabulary knowledge and lexical inferencing. Despite its importance for compound words, little research has investigated the impact of morphological awareness on the processing of novel compounds when encountered incidentally during reading.

The current study

To bridge the aforementioned gaps in the current literature, this study set out to investigate L2 learners' processing of novel L2 compounds across repeated exposures during reading, while considering the impact of lexical characteristics (i.e., semantic transparency) and individual learner differences (i.e., L2 vocabulary size,

working memory capacity, morphological awareness). In order to examine the processing of novel words without interrupting L2 learners' natural reading behavior, the eye-tracking technique was used. Specifically, for the purpose of this study, three eye-tracking measures, including first fixation duration, gaze duration, and total reading time, were used. To illustrate real-time processing of novel L2 words across repeated exposures, growth curve analysis was employed, following the practice of Godfroid and colleagues (Godfroid et al., 2018). The following research questions (RQs) were addressed:

1. Does semantic transparency impact L2 learners' processing of novel compounds during reading?
2. Do L2 learners' individual differences, including L2 vocabulary size, working memory capacity, and morphological awareness, affect L2 learners' real-time processing of novel compounds during reading? Moreover, do such influences differ between semantically transparent and opaque compounds?
3. How does L2 learners' processing of novel compounds change across repeated exposures? Do semantic transparency and individual learner differences (i.e., L2 vocabulary size, working memory capacity, morphological awareness) moderate such changing patterns?

Methodology

Participants

Sixty-one intermediate-level L2 Chinese speakers (43 females) were recruited from several universities in Beijing, China. Before being invited to participate in this study, they had to meet the following criteria. First, by the time the experiment began, they had to be enrolled in intermediate-level classes to study Chinese as a second language. Second, to ensure that participants had no extra amount of exposure to Chinese characters because of their first/heritage language, those speaking Korean, Japanese, or Vietnamese as their first or heritage language were excluded. Third, self-ratings of participants' communicative Chinese language ability were also collected, such that those achieving extremely low or high scores (see *Experimental tasks*) were screened out. Finally, to guarantee the quality of eye-tracking data, only those with normal or corrected-to-normal vision were invited to the lab. The participants were from various L1 backgrounds (i.e., Arabic, Bengali, Dutch, English, French, German, Hindi, Italian, Kazakh, Kyrgyz, Mongolian, Nepali, Polish, Romanian, Russian, Sesotho, Spanish, Ukrainian, Urdu, Uzbek), among which German (five participants) and Dutch (one participant) are said to be rich in compounds (Lieber & Stekauer, 2011). On average, the participants were 22.1 years old ($SD = 2.9$), and they reported to have received an average of 3.2 years ($SD = 3.1$) of formal instruction of Chinese by the time the experiment began.

Target compounds

Target compounds were chosen in a way such that L2 participants were highly familiar with the characters that constituted the compound words, whereas they

had no knowledge of the compounds. Twelve low-frequency disyllabic concrete Chinese compounds (see Appendix 1 in the supporting information online), including six semantically fully transparent words and six semantically fully opaque words, were selected based on the following procedures. First, 4,354 disyllabic Chinese nouns were retrieved from a word list downloaded from the CNCORPUS website (Jin et al., 2005), with frequency rankings provided. According to *Words and Characters for Chinese Proficiency Syllabus* (National Committee for Chinese Proficiency Test, 2001), the vocabulary size of intermediate-level Chinese learners generally does not exceed 5,000. To ensure that intermediate-level L2 Chinese participants were unlikely to have known the target compounds, a cutoff was set such that words in the pool with a frequency ranking below 5,000 (i.e., the most frequent 5,000 words) were excluded. D. Li (2003) reported that L2 Chinese learners at intermediate levels are able to read and write an average of 1,203 Chinese characters. To ensure that participants would process the characters that make up each target compound without difficulty, a threshold was set at 1,500, such that words consisting of Chinese characters with a frequency ranking beyond 1,500 were not considered. The author then examined the remaining words in the pool and further removed words that fell into the following categories: 1) words that were mistakenly tagged and turned out not to be nouns; 2) words that comprise two characters but are monomorphemic (e.g., 伯伯-father's elder brother); 3) words that may be perceived as phrases (e.g., 两腿-two legs); 4) words that are transliterations of foreign words (e.g., 雷达-radar); 5) words that are specific to Chinese culture (e.g., 气功 qigong) or certain fields of study (e.g., 心室-ventricular); 6) words that are ambiguous (e.g., 单元-unit in a textbook or building); 7) words that could function as either nouns or verbs (e.g., 保证-guarantee). After going through the above-mentioned steps, 387 compound words were kept. They were then grouped into three lists and rated by 15 native Chinese speakers, based on random assignment. For semantic concreteness, the raters were asked to judge whether a word is concrete based on whether the concept described by the word could be sensed. For semantic transparency, they were instructed to evaluate the semantic relationship between each constituent morpheme and the compound. Following Libben (1998), semantic transparency was operationalized as semantic compositionality. If a morpheme (e.g., 友-friend) relates to the meaning of the word (e.g., 友谊-friendship), then it should be rated as semantically transparent to the compound. By contrast, if a morpheme (e.g., 计-calculate) does not contribute to the meaning of the word (e.g., 伙计-shop assistant), then it should be rated as semantically opaque to the compound. In this study, only semantically fully transparent (i.e., both morphemes are semantically transparent to the compound, e.g., 果园-fruit garden, orchard) and fully opaque words (i.e., both morphemes are semantically opaque to the compound, e.g., 百合-hundred-combine, lily) were chosen. Following this, 228 compounds were selected as candidates for the experiment. A Qualtrics survey was then delivered to eight intermediate-level L2 Chinese speakers who did not participate in the experiment. Based on a four-point Likert scale (1: *I had never seen this word*; 2: *I had seen this word before, but I did not know its meaning*; 3: *I had already seen this word and knew its meaning*; 4: *I was already very familiar with this word, its meaning and usage*), they were instructed to rate their degree of familiarity for each compound as well as the constituent morphemes. Candidates

receiving average familiarity ratings higher than 1.5 for the compound and less than 2 for the constituent morphemes were removed. Following this, after matching lexical and sublexical characteristics (i.e., compound familiarity, compound frequency, first/second-morpheme familiarity, first/second-morpheme frequency, second-morpheme strokes) between semantically fully transparent and opaque compounds, 12 compounds were selected.

Reading materials

Given that it would be unrealistic to insert all the target words in a single readable story, the twelve compounds were broken down into two groups, based on random assignment. Each group contained three transparent compounds and three opaque compounds, which were then embedded in a story created by the author, with their presentation order randomized. Such a practice was carried out six times, resulting in six sets of stimuli embedded in a total of 12 stories (for the distribution of target compounds in the stories, see Appendix 2 in the supporting information online). By doing so, each target compound appeared in one of the two stories under the same set, adding up to six occurrences across all the stories. While creating the stories, the following principles were followed. First, to control for the length of the stories, each story comprised 20 sentences. Second, target compounds were not embedded in the first or last sentence. To control for the spacing of the target words, the remaining 18 sentences were divided into six blocks, each consisting of three consecutive sentences. Within each block, one target compound was randomly selected from the six candidates and inserted, after counterbalancing its sentential location (i.e., the first/second/third sentence) where it appeared. Third, to avoid any spillover or wrap-up effects, the target compounds were not inserted in the beginning or ending position in the sentences. Fourth, all stories were created by the author, built upon common themes. Moreover, based on the feedback collected from an experienced Chinese instructor, words that might be unfamiliar to intermediate-level L2 learners of Chinese were removed or replaced by more frequent lexical items, leading to an average word frequency of 266 times per million words. The twelve stories can be seen in Appendix 3 in the supporting information online. Using a ten-point Likert scale (1 → 10: least readable → most readable), four L1 speakers of Chinese were instructed to read each story and rate its readability, which was defined as the cohesion of the stories and the naturalness of the use of target words. The stories achieved an average readability rating of 8.0 ($SD = 0.6$), and revisions were made where necessary, based on the raters' feedback. On average, each story consisted of 210 words ($SD = 10$) and 300 characters ($SD = 4$). To statistically control for the contextual predictability of the target compounds in each sentence, a cloze test was administered, in which 10 Chinese L1 speakers were asked to read all the stories and fill in the blanks that replaced the target words based on their intuition. Contextual predictability was operationalized as the possibility one could figure out the semantic category of the target compound and calculated as the proportion of correct answers (i.e., the exact target word or synonyms of the target word) over the total number of responses for a given target word. To ensure that participants would focus on understanding the textual meaning, four yes/no comprehension questions were created for every story, each consisting of a statement

that does not include or address the target compounds. Two additional stories were created for practice purpose, in order to help participants become familiar with the eye-tracking reading procedure.

Experimental tasks

Screening surveys

A demographic questionnaire and a self-rating communicative Chinese language ability survey (adapted from Bachman & Palmer, 1989) were administered online through Qualtrics for screening purposes. The communicative language ability survey was conceptualized to measure L2 learners' grammatical, pragmatic, and socio-linguistic competence, in which participants were instructed to evaluate their L2 ability using a four-point Likert scale (1 → 4: bad → good). The score of this survey ranged from 0 to 84. To ensure that participants' L2 proficiency would not be too low or too high for the current study, those achieving scores below 36 or beyond 72 were not invited to the lab. Cronbach's alpha for this survey was .89; 95% CI [.85, .93].

Measurements of individual differences

Vocabulary knowledge assessment tools for Chinese—especially those that are brief and do not demand too much time to complete—are in short supply. Given this, a receptive Chinese vocabulary size test was borrowed from Pelzl et al. (2019) and delivered through Qualtrics (see Appendix 4 in the supporting information online). In this test, 105 Chinese words were sampled from seven frequency bands (frequency range: 1–7,000) based on the SUBTLEX-CH frequency list (Cai & Brysbaert, 2010). After random assignment, these words were grouped into seven blocks, each containing 15 words. The presentation order of each block, as well as the order of words within each block, was randomized. L2 Chinese participants were required to judge whether they knew the meaning of each word by responding "yes" or "no." While taking this test, they were not allowed to look up the words using dictionaries. Scores on this test ranged from 0 to 105. According to Pelzl et al. (2019), a cutoff score of 70% or above could be used for labeling participants as proficient L2 speakers of Chinese. Cronbach's alpha for this test was .96; 95% CI [.95, .98].

A shortened version of the operation span test was borrowed from the Georgia Tech Attention and Working Memory Lab (Forster et al., 2015) and run on the E-Prime software. This was a verbal working memory test that has been validated by many researchers. The participants were first exposed to a mathematical operation and had to judge whether it was correct or wrong. Following this, they were presented with an English letter. This math-letter sequence was repeated four to six times for each trial. By the end of each trial, participants were required to recall the letters in the correct order by selecting them from a matrix of twelve letters. The operation span test was automatized, with participants' responses timed. Participants were instructed to respond as accurately and as fast as possible. Working memory capacity scores were calculated by summing the number of letters correctly recalled in the correct order, ranging from 0 to 50. Cronbach's alpha for this test was .78; 95% CI [.67, .90].

Following the practice in the literature (e.g., Chen, 2018), a Chinese morphological awareness test (see Appendix 5 in the supporting information online) was developed by the author and delivered through Qualtrics. This test comprised 30 commonly used disyllabic Chinese words, which were piloted to ensure that intermediate-level Chinese L2 learners were familiar with them. Among the 30 lexical items, half were compounds (e.g., 鸡蛋 chicken-egg), and the other half were monomorphemic words (沙发 sofa). The compounds could be segmented into two smaller meaningful units (i.e., morphemes), with each unit directly contributing to the meaning of the word. By contrast, although the monomorphemic words were composed of two characters, they could not be divided into smaller meaningful units. In that case, neither of the two characters contributed to the meaning of the whole word. Participants were instructed to decide whether or not each word is segmentable. If they had difficulty making a judgment, they could choose "I don't know." Two examples—including one segmentable word (黑板 blackboard) and one non-segmentable word (蝴蝶 butterfly)—were provided. Each correct response received one point. Wrong answers or choosing "I don't know" received no points. Scores on this test ranged from 0 to 30. Cronbach's alpha for this test was .78; 95% CI [.69, .86].

Eye-tracking reading

Participants were exposed to the target compounds while reading the stories in front of a computer, with their eye movements recorded by an EyeLink 1000 Plus eye tracker (SR Research, Canada; sampling rate: 1,000 Hz). The stories were presented in a double-spacing manner on a 21-inch CRT monitor (resolution: 1,024 x 768 pixels; refresh rate: 150 Hz), each split into four screens (i.e., five sentences per screen). Based on random assignment, the stories were grouped into two lists, each containing six stories. The presentation order of the two lists was counterbalanced. Within each list, the presentation order of the stories was also randomized. Participants read the stories binocularly, yet only their right eye was monitored. To ensure the quality of eye-tracking data, nine-point calibrations were performed before the experiment, with additional calibrations carried out when necessary. Following the practice of previous studies (Elgort et al., 2018; Godfroid et al., 2018; Mohamed, 2018), drift corrections were set up at the beginning of each screen.

Prior lexical knowledge survey

After reading the twelve stories, a survey (see Appendix 6 in the supporting information online) was delivered to the participants online through Qualtrics. This survey was aimed to help filter out lexical items already known by the participants before taking the reading experiment (see *Statistical analysis*). The participants were asked to report how familiar they had been with each target compound before participating in the study, based on a four-point Likert scale (1: *I had never seen this word*; 2: *I had seen this word before, but I did not know its meaning*; 3: *I had already seen this word and knew its meaning*; 4: *I was already very familiar with this word, its meaning and usage*).

Procedure

Prior to the experiment, the demographic questionnaire and the communicative Chinese language ability survey were delivered online through Qualtrics. L2 Chinese learners who met the criteria as noted in the earlier section were invited and tested individually in an eye-tracking lab. On the first day, participants were instructed to read six stories naturally for meaning at their own pace. They were also notified that there would be comprehension questions following each story. To minimize effects of fatigue, compulsory three-minute breaks were taken after finishing reading every two stories, with recalibrations conducted after those breaks. Following the reading phase, they took a break for five minutes before proceeding to the morphological awareness test and the working memory test. The order of these two tests was counterbalanced, such that half the participants took the morphological awareness test before the working memory test, while the other half took them in the reversed order. On the second day, participants continued to read the remaining stories (six in total). As part of a broader project, they also took three surprise vocabulary knowledge tests (i.e., a form recognition test, a meaning recall test, and a meaning recognition test) immediately after the reading session, which was then followed by the prior lexical knowledge survey. The whole experiment took about two hours.

Statistical analysis

Data preparation

Two participants turned out to be Chinese heritage speakers. In addition, seven participants achieved accuracies of lower than 70% on the comprehension questions, indicating that they either did not follow the instructions and read the stories for meaning or had difficulty understanding the stories. After removing the nine participants, 52 participants were left for data analysis. Target compounds embedded in each story were defined as the area of interest, for which fixation durations (i.e., first fixation duration, gaze duration, and total reading time) recorded by the eye tracker were extracted. Given that real L2 words were used in this study, participants' self-reported ratings of prior lexical knowledge (see *Experimental tasks*) were checked against their performance on the vocabulary post-tests. Specifically, if they claimed to have already known the meaning of a target compound (i.e., selecting *3* or *4* in the survey) prior to the experiment, and indeed they responded correctly in terms of form and meaning recognition, then their eye-tracking data for this compound were excluded from analysis. Following this, 25% of the observations were removed (4.5% for opaque words, 20.5% for transparent words). Before analyzing the eye-tracking data, all trials where track loss happened were removed (1.5%). Fixation durations shorter than 80 ms were also excluded from analysis (1.3% for first fixation duration, 0.6% for gaze duration, and 0.1% for total reading time), as they are considered uninformative (Betancort et al., 2009). Due to technical issues, two participants did not complete the working memory test. As a result, they were excluded from statistical models that tested the effects of working memory capacity.

For the sake of the current study, the dependent variables of interest—namely first fixation duration, gaze duration, and total reading time—were transformed

Table 1. Correlations between participant-level predictors

	Vocabulary	Awareness	Memory	Comprehension
Vocabulary	–			
Awareness	0.36**	–		
Memory	0.20	−0.07	–	
Comprehension	0.40**	0.32*	0.10	–

Note. Pearson's correlation coefficients were computed, with missing values in memory removed using listwise deletion.
* $p < .05$; ** $p < .01$; *** $p < 0.001$.

using natural log. The independent variables (IVs) of interest included Exposure and Transparency: the former was a time-course predictor representing each occurrence (1-6) of the target compounds, while the latter was a dummy-coded categorical variable, with transparent compounds as the reference level. Individual differences of the participants, including L2 vocabulary size (Vocabulary), working memory capacity (Memory), and morphological awareness (Awareness), were rescaled to *z*-scores (calculated by subtracting the mean of the original variable before being divided by the standard deviation) and incorporated for statistical analysis. To control for the visual complexity of the target compounds, first-character strokes (Strokes) was standardized and treated as a covariate. Participants' accuracy rates on the comprehension questions (Comprehension), as well as the contextual predictability of each occurrence of the target compounds (Predictability), were transformed to the percentage scale before being standardized. Correlations between participant-level predictors are summarized in Table 1. As clearly shown in the table, only moderate correlations were found between participant-related predictors.

Growth curve analysis

To examine the real-time processing of novel L2 compounds across repeated exposures, growth curve analyses were carried out, following the recommendations of Mirman (2014). Given that changes over repeated exposures are often nonlinear, higher-order polynomial terms of Exposure—including quadratic (Exposure^2), cubic (Exposure^3), and quartic (Exposure^4) terms—were tested. Terms of Exposure are correlated, making it impossible to evaluate their effects independently. Following Mirman (2014), orthogonal polynomials were created, such that the correlations between terms of exposure were removed. The fixed effects consisted of terms of Exposure, Transparency, individual learner differences (i.e., Vocabulary, Memory, and Awareness), as well as covariates of interest (i.e., Strokes, Comprehension, Predictability). Additionally, interaction terms, including those between terms of Exposure and Transparency, between individual learner differences/covariates and Transparency, and between individual learner differences/covariates and terms of Exposure, were also considered.

Statistical models were built separately for first fixation duration, gaze duration, and total reading time, using the maximum likelihood technique. First, a maximal

model was fitted, with all the IVs (i.e., Exposure, Exposure^2, Exposure^3, Exposure^4, Transparency) and their interactions incorporated. To avoid failure of convergence of models due to complexity, only random intercepts for subjects and items were included. Following backward model selection procedures and through model comparisons, this maximal model was then gradually reduced to a parsimonious model, until all fixed effects remaining in the model were significant. Subsequently, effects of individual learner differences (i.e., Vocabulary, Memory, and Awareness) and the covariates (i.e., Strokes, Comprehension, Predictability), as well as their interactions with Transparency and terms of Exposure that remained in the so-far-best-fitted model, were tested, following forward model selection procedures through model comparisons. Finally, random slopes of fixed effects remaining in the so-far best models were tested for subjects and items. Statistical analyses were performed using the lme4 package (version 1.1-21, Bates et al., 2015) in R (version 3.6.2, R Core Team, 2020). Model comparisons were carried out using the *anova* function in the *lme4* package, based on log-likelihood tests. The significance level, *alpha*, was set at .017, after Bonferroni correction. For each best-fitted model, effect sizes measured by marginal R^2 (the proportion of variance explained by fixed effects) and conditional R^2 (the proportion of total variance explained by both fixed and random effects) were obtained using the *tab_model* function in the sjPlot package (version 2.8.2, Lüdecke, 2019), along with the *p*-values.

Results

Descriptive statistics

Participants achieved high accuracy rates on the comprehension questions ($M = 84.4\%$, $SD = 6.9\%$), indicating that they did read for meaning and had little difficulty understanding the stories. For the transparent compounds, L2 Chinese learners' average accuracy on form recognition, meaning recall, and meaning recognition was 92%, 54%, and 52%, respectively. For the opaque compounds, their average accuracy on these immediate vocabulary tests was 79%, 26%, and 48%, respectively. Such results indicate that L2 learners did process the orthographic and semantic information of the novel target compounds, albeit to different degrees depending on the type of vocabulary test and semantic transparency. First fixation duration, gaze duration, and total reading time for the novel transparent and opaque compounds averaged across six exposures were 336.7 ms ($SD = 197.9$ ms), 794.6 ms ($SD = 660.1$ ms), and 1558.9 ms ($SD = 1245.7$ ms), respectively. The average first fixation duration, gaze duration, and total reading time for the novel opaque compounds across six exposures were 323.9 ms ($SD = 178.2$ ms), 686.7 ms ($SD = 477.8$ ms), and 1499.1 ms ($SD = 1170.5$ ms), respectively. Correlation analyses revealed weak to moderate associations between the three eye-tracking measures. Specifically, first fixation duration was positively associated with gaze duration ($r = .38$, $p < .001$) and total reading time ($r = .17$, $p < .001$). Meanwhile, gaze duration moderately correlated with total reading time ($r = .49$, $p < .001$). Descriptive statistics for the eye-tracking measures for each occurrence of the opaque and transparent compounds are summarized in Table 2.

Table 2. Fixation durations for the target opaque and transparent compounds across exposures

Transparency	Exposure	FFD (M/SD)	GZD (M/SD)	TRT (M/SD)
Transparent	1	345.1 (204.7)	985.5 (927.0)	1954.5 (1539.4)
Transparent	2	359.9 (247.6)	869.9 (711.0)	1764.9 (1336.8)
Transparent	3	324.6 (158.8)	722.5 (538.0)	1515.1 (1164.2)
Transparent	4	332.7 (214.4)	801.2 (619.8)	1426.1 (1040.1)
Transparent	5	334.0 (168.0)	758.5 (559.6)	1409.2 (1015.6)
Transparent	6	323.9 (179.5)	623.2 (425.6)	1273.4 (1174.9)
Opaque	1	327.7 (169.4)	785.6 (556.9)	1964.1 (1498.9)
Opaque	2	337.2 (199.9)	745.9 (483.5)	1594.9 (1227.8)
Opaque	3	324.2 (197.8)	658.0 (461.6)	1454.0 (971.6)
Opaque	4	320.0 (168.6)	648.2 (500.4)	1311.0 (1020.8)
Opaque	5	320.2 (139.6)	649.6 (360.1)	1381.1 (1102.4)
Opaque	6	313.8 (187.2)	632.0 (464.4)	1283.8 (976.0)

Note. Means and standard deviations of fixation durations are in milliseconds and computed across participants.
FFD: first fixation duration. GZD: gaze duration. TRT: total reading time.

First fixation duration

Growth curve analysis for first fixation duration did not reveal any significant effect of the independent variables and covariates. Instead, only the intercept (*Estimate* = 5.66, *SE* = 0.02, *t* = 302.90, *p* < .001) was significant. In answer to the research questions, such results indicate that: 1) novel transparent and opaque compounds did not differ in first fixation duration received from L2 learners; 2) individual learner differences, including L2 vocabulary size, morphological awareness, and working memory capacity, had no impact on L2 learners' first fixation duration on the novel compounds; 3) for both novel transparent and opaque compounds, first fixation duration did not change significantly across exposures.

Gaze duration

Growth curve analysis for gaze duration (Table 3) did not find any significant effect of Transparency, indicating that gaze duration on the novel compounds did not vary by semantic transparency. When it comes to the impact of learner individual differences on gaze duration, a significant interaction between Vocabulary and Transparency was found (*Estimate* = −0.078, *SE* = 0.033, *t* = −2.387, *p* = .017), suggesting that participants with a larger L2 vocabulary size processed the novel opaque compounds faster than those with a smaller L2 vocabulary size. Both the linear (*Estimate* = −0.183, *SE* = 0.032, *t* = −5.780, *p* < .001) and quadratic terms of Exposure (*Estimate* = 0.093, *SE* = 0.032, *t* = 2.939, *p* = .003) were significant, indicating a curvilinear decrease in gaze duration (logged) across exposures. As illustrated in Figure 1,

Table 3. Growth curve analysis results for gaze duration

Parameters	Gaze Duration				
	Estimate	95% CI	SE	t	p
Intercept	6.333	6.263, 6.404	0.036	175.888	<0.001
Exposure	−0.183	−0.245, −0.121	0.032	−5.780	<0.001
Exposure: Vocabulary	−0.077	−0.139, −0.015	0.032	−2.418	0.016
Exposure^2	0.093	0.031, 0.156	0.032	2.939	0.003
Exposure^2: Strokes	0.079	0.016, 0.143	0.032	2.444	0.015
Vocabulary: Transparency	−0.078	−0.143, −0.014	0.033	−2.387	0.017
Random Effects					
Variance	0.04 $_{subject}$				
	0.00 $_{item}$				
Number of observations	2728				
Marginal R^2/Conditional R^2	0.024/0.112				

Note. Exposure and Exposure^2 represent the linear and quadratic term of Exposure, respectively. Model formula: Gaze Duration (logged) ∼ Exposure + Exposure: Vocabulary + Exposure^2 + Exposure^2: Strokes + Vocabulary: Transparency + (1|Subject) + (1|Item).

the decrease of gaze duration followed a wide U-shaped pattern, where at first there was a steep decrease until the third or fourth exposure, which was followed by a further, more gradual decrease. The absence of interactions between terms of Exposure and Transparency means that such a decreasing pattern—as well as the rate of decrease—did not differ significantly between novel transparent and opaque compounds. Additionally, Vocabulary was found to significantly interact with Exposure (*Estimate* = −0.077, *SE* = 0.032, *t* = −2.418, *p* = .016), indicating that the rate of decrease in gaze duration on novel compounds was faster for those who had a larger L2 vocabulary size. Overall, fixed effects in the best-fitted model explained 2.4% of the variance, whereas fixed effects and random effects together explained 11.2% of the variance.

Total reading time

Growth curve analysis for total reading time (Table 4) did not reveal any significant effect of Transparency, suggesting that L2 learners did not fixate on the novel opaque compounds significantly more time than the novel transparent compounds. For both types of compounds, individual learner differences did not show any significant impact on cognitive processing of compounds. After Bonferroni correction, the main effect of Strokes (*Estimate* = 0.067, *SE* = 0.032, *t* = 2.069, *p* > .017) was no longer significant. However, the significant interaction between Strokes and Transparency (*Estimate* = −0.175, *SE* = 0.049, *t* = −3.591, *p* < .001) indicates that orthographically more complex opaque compounds were processed
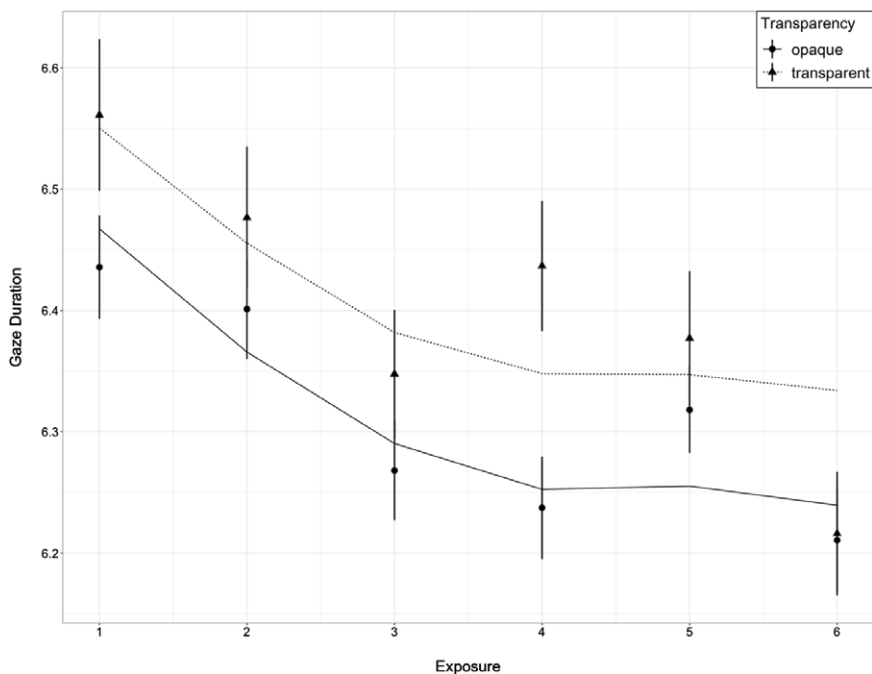
Figure 1. Growth curve model fits (lines) for effect of transparency on gaze durations (logged) for novel compounds across repeated exposures (dots and point ranges indicate means and standard errors, respectively).

faster than those with fewer strokes. The significant effects of the linear (*Estimate* = −0.348, *SE* = 0.030, *t* = −11.627, *p* < .001) and quadratic terms of Exposure (*Estimate* = 0.126, *SE* = 0.044, *t* = 2.849, *p* = .004) captured a curvilinear decrease in total reading time (logged) across exposures (Figure 2). This decrease followed a wide U-shaped pattern, beginning with an initial steep decrease until the fourth exposure, which was followed by a further, more gradual decrease. Such a pattern was similar to what was found in gaze duration, except that the decreasing rate was much faster. Given that the interactions between terms of Exposure and Transparency were not significant, this means that total reading time for transparent and opaque compounds followed similar changing patterns. Interestingly, a significant interaction between Exposure and Memory was found (*Estimate* = −0.081, *SE* = 0.030, *t* = −2.697, *p* = .007), indicating that the rate of decrease in total reading time across exposures was significantly faster for L2 learners with greater working memory capacity. Additionally, the significant interaction between Exposure and Comprehension (*Estimate* = −0.141, *SE* = 0.029, *t* = −4.859, *p* < .001) suggests that the rate of decrease in total reading time across exposures was also faster for those who comprehended the stories better. Overall, the fixed effects in the best-fitted model explained 5.4% of the variance, whereas the fixed effects along with the random effects together explained 31.7% of the variance.

Table 4. Growth curve analysis results for total reading time

Parameters	Total Reading Time				
	Estimate	95% CI	SE	t	p
Intercept	6.954	6.840, 7.068	0.058	119.571	<0.001
Exposure	−0.348	−0.407, −0.289	0.030	−11.627	<0.001
Exposure: Memory	−0.081	−0.139, −0.022	0.030	−2.697	0.007
Exposure: Comprehension	−0.141	−0.198, −0.084	0.029	−4.859	<0.001
Exposure^2	0.126	0.039, 0.213	0.044	2.849	0.004
Strokes	0.067	0.004, 0.130	0.032	2.069	0.039
Strokes: Transparency	−0.175	−0.270, −0.079	0.049	−3.591	<0.001
Random Effects					
Variance	0.14 $_{subject}$				
	0.00 $_{item}$				
	0.05 $_{subject:\ Exposure2}$				
Number of observations	2746				
Marginal R^2/Conditional R^2	0.054/0.317				

Note. Exposure and Exposure^2 represent the linear and quadratic term of Exposure, respectively. Model formula: Total Reading Time (logged) ~ Exposure + Exposure: Memory + Exposure: Comprehension + Exposure^2 + Strokes + Strokes: Transparency + (1+Exposure^2|Subject) + (1|Item).

Discussion

In answer to the research questions, this study revealed the following findings. First, semantic transparency showed no impact on L2 learners' online processing of novel compounds encountered incidentally during reading. Second, individual learner differences, including working memory capacity and morphological awareness, did not directly influence the processing of novel L2 compounds during reading. However, second language learners with a larger L2 vocabulary size processed novel opaque compounds faster in terms of gaze duration. Third, processing of novel transparent and opaque L2 compounds did not change across repeated exposures in terms of first fixation duration. Nevertheless, it followed similar curvilinear decreasing patterns in gaze duration and total reading time, with the decreasing rates being higher for those with a larger L2 vocabulary size and greater working memory capacity, respectively.

Processing of novel transparent and opaque compounds during reading

After controlling for sublexical, lexical, and contextual covariates, this study found no effect of semantic transparency on L2 learners' processing of novel compounds during reading. The lack of semantic transparency effect on first fixation duration is consistent with earlier research findings (e.g., Frisson et al., 2008; Juhasz, 2007) regarding L1 learners' processing of familiar compounds. First fixation duration is an eye-tracking measure indicative of L2 learners' familiarity check of orthographic forms (Juhasz & Rayner, 2003). Libben (1998) assumed that language users
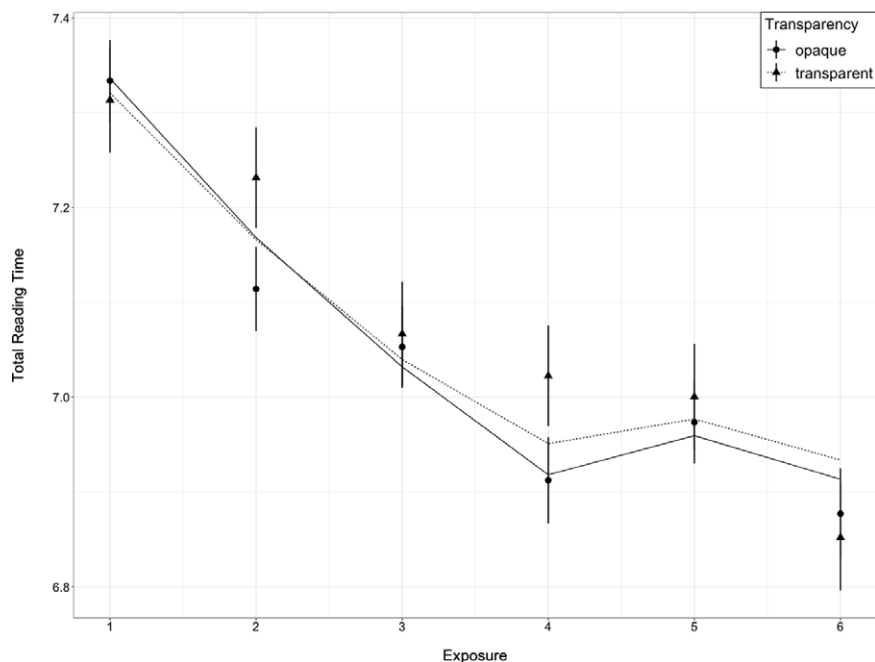
Figure 2. Growth curve model fits (lines) for effect of transparency on total reading time (logged) for novel compounds across repeated exposures (dots and point ranges indicate means and standard errors, respectively).

would have to parse novel compounds into constituents before checking the orthographic and lexical status of the morphemes. However, the absence of semantic transparency effect on first fixation duration, as revealed in this study, suggests that morphological parsing at the orthographic/stimulus level may not be compulsory for novel Chinese compounds when encountered during normal reading. The absence of semantic transparency effects on gaze duration and total reading time is unexpected. Following Libben's model (1998), for familiar compounds, effects of semantic transparency on gaze duration and total reading time indicate that language learners are taking a decomposition route by accessing the meaning of the compounds from their constituent morphemes. Previous studies (e.g., Frisson et al., 2008; Underwood et al., 1990) suggest that skilled L1 readers do not necessarily need to decompose familiar compounds into their constituents when retrieving their meaning during normal reading. Unlike familiar compounds, novel compounds do not have lexical representations in the mental lexicon. Therefore, it is likely that L2 learners might have to decompose both transparent and opaque compounds to infer their meaning, especially when contextual information is not employed. Based on the E-Z Reader Model (Reichle, Pollatsek, Fisher, et al., 1998; Reichle, Pollatsek, et al., 2006), gaze duration and total reading time might reflect L2 learners' attempt to derive the meanings of novel compounds on the basis of sublexical/lexical and contextual support, respectively. The lack of semantic transparency effect on total reading time, along with the absence of effects of

contextual predictability on reading times for the novel compounds (see Tables 3 and 4), suggests that L2 learners did not rely on contextual information to infer the meaning of novel target compounds. Furthermore, the absence of semantic transparency effect on gaze duration also indicates that L2 learners processed novel transparent and opaque compounds in the same way, probably adopting a decomposition route by deriving the meaning of the compounds from their constituent morphemes.

The roles of individual learner differences

This study demonstrated that L2 vocabulary size had no impact on second language learners' processing of novel compounds during reading, in terms of first fixation duration and total reading time. Given that first fixation duration and total reading time are assumed to reflect familiarity check of orthography and post-lexical integration at the discourse level, respectively, this indicates that second language learners' vocabulary knowledge in L2 may not contribute to these cognitive processes for novel lexical items during normal reading. Nevertheless, in the current research, a larger L2 vocabulary size was found to lead to faster processing of novel opaque compounds, in terms of gaze duration. Gaze duration is said to reflect lexical access for familiar words or form-meaning association for novel words. Consequently, such facilitation effect of L2 vocabulary size on the processing of novel opaque compounds suggests that the L2 mental lexicon might have been activated to help infer word meanings. Amenta et al. (2020) proposed that the meaning of novel compounds may be induced by extracting the distributional patterns of the constituent morphemes from past language experience. Following this, those with a larger L2 vocabulary might be more advantageous than those with a smaller L2 vocabulary when dealing with meaning resolution for novel opaque compounds, because their distributional semantic networks are more sophisticated due to higher lexical proficiency. By contrast, given the componential nature of novel transparent compounds, L2 learners could derive their meanings from the constituent morphemes, resulting in less need to activate the L2 mental lexicon to aid meaning inferencing.

The absence of effects of working memory capacity on L2 learners' processing of novel compounds during reading is also worth discussing. Earlier studies (e.g., Martin & Ellis, 2012; Service & Kohonen, 1995; Speciale et al., 2004) suggest that storage of novel lexical items is involved during intentional vocabulary learning, as evidenced by significant effects of PSTM. Working memory capacity has been argued to be critical for reading comprehension processes (Daneman & Hannon, 2007), such as memorizing and integrating information, as well as making inferences about new information. Surprisingly, in this study, no correlation was found between working memory capacity and reading comprehension (see Table 1). In addition, consistent with the findings in Yi et al. (in press), working memory capacity did not show any direct impact on L2 learners' cognitive processing of novel compounds during reading. The lack of correlation between working memory capacity and reading comprehension may relate to the possibility that the composed stories were not cognitively too demanding, as can be seen from participants' accuracy rates on the comprehension questions ($M = 84.4\%$, $SD = 6.9\%$). In a separate analysis focusing on the relationship between individual learner differences and

vocabulary gains, working memory capacity showed no impact on L2 learners' acquisition of any aspect of knowledge (i.e., form recognition, meaning recall, and meaning recognition) of the transparent or opaque compounds, which confirms the lack of involvement of memory processes when L2 learners were incidentally exposed to the novel lexical items during reading. Unlike the current research, Malone (2018) and Montero Perez (2020) reported significant correlations between form recognition of novel lexical items and working memory capacity. However, both studies involved audio-visual or multimodal input, with the former delivering the written materials to participants in a reading-while-listening way, whereas the latter exposed L2 learners to videos. Taken together, such patterns suggest that the demand of working memory during the processing of novel words in natural reading may not be as high as that in multimodal or intentional learning tasks. Under a reading-for-meaning context, L2 learners simply do not need to explicitly rehearse, memorize, and manipulate unfamiliar lexical items.

This study also did not find any effect of morphological awareness on the processing of novel compounds during reading. Morphological awareness—or more precisely, compounding awareness as in this study—refers to one's ability to evaluate the compositionality of compounds. The absence of the effect of morphological awareness as revealed here suggests that L2 learners do not segment novel compounds into constituent morphemes, when they are engaged in meaning-focused activities, such as reading. Given that each novel compound occurred six times in this study, this indicates that L2 learners' unawareness of the morphological structure of novel compounds was maintained throughout the experiment, even after repeated exposures. Such results were consistent with the conclusion made in the previous section, based on the absence of effects of semantic transparency on cognitive processing of novel L2 compounds.

Changing patterns of novel word processing across exposures

In line with earlier research findings (e.g., Godfroid et al. 2018; Joseph et al., 2014; Mohamed, 2018; Pellicer-Sánchez, 2016), cognitive processing of both transparent and opaque compounds in the current study decreased over repeated exposures, as observed in gaze duration and total reading time. Particularly, such changing patterns were curvilinear and roughly U-shaped, beginning with a sharp decrease until the third or fourth exposure and followed by a continuing, more gradual decrease until the sixth exposure. Such changing patterns echo the widely accepted idea that changes in cognitive processing over time are often nonlinear (Mirman, 2014), despite the fact that the exact curvilinear shape of changing patterns may not be identical across studies (such changing patterns may also vary across participants even within the same study). For instance, given more exposures, the decreasing pattern of fixation durations for novel lexical items encountered during reading might follow S-shaped curves (e.g., Godfroid et al., 2018). This study also found that the decreasing rates of gaze duration across exposures were faster for those with a larger L2 vocabulary size. Similarly, those with greater working memory capacity and higher reading comprehension accuracies also had faster decreasing rates of total reading time across exposures. Such results indicate that L2 vocabulary size, working memory capacity, and reading comprehension made independent yet probably distinct contributions to the processing of novel compounds during reading. Specifically, being exposed to the same novel compound repeatedly

through reading, learners with a larger L2 vocabulary size may be increasingly advantageous in terms of meaning inference, whereas those with greater working memory capacity may become increasingly efficient with respect to integrating the novel compounds into surrounding contexts. Interestingly, in the current research, early processing of novel L2 compounds, as captured by first fixation duration, did not change significantly across exposures. As already mentioned, first fixation duration mainly reflects L2 learners' familiarity check of orthographic information. Such a cognitive process is relatively fast (in the current study, 337 ms and 324 ms for transparent and opaque compounds, respectively). Consequently, it might be subject to a floor effect.

Concluding remarks

In summary, this study illustrates a multifaceted picture of online processing of novel L2 words during reading, by incorporating the influences of lexical characteristics, individual learner differences, and repeated exposures. Following a process-oriented approach, this study reveals that real-time processing of novel L2 words could be affected by the incidental nature of meaning-focused vocabulary learning activities, such as reading. Specifically, the absence of effects of semantic transparency, morphological awareness, and working memory capacity on the processing of novel compounds provides converging evidence supporting that L2 learners' primary focus was indeed not on the novel lexical items when engaged in natural reading: they might segment both novel transparent and opaque compounds into morphemes without awareness of the semantic relationship between the constituents (i.e., the morphological structure); moreover, they did not intentionally memorize such lexical items. From a theoretical point of view, these findings lend support to the widely accepted notion of incidental vocabulary learning (e.g., Hulstijn, 2001; Webb, 2020). By incorporating individual learner differences, the current study sheds important light on the impact of learner variables on the real-time processing of novel lexical items under incidental learning conditions. Finally, research findings in this study could also generate certain pedagogical implications. Given that intermediate-level L2 speakers of Chinese were not found to employ the morphological structure and semantic transparency of novel compounds encountered during reading to facilitate the inference of word meaning, instructional practice that raises learners' awareness of such information should be advocated to boost incidental vocabulary learning from extensive reading. Needless to say, the present study is not without limitations. For example, one might wonder whether the participants actually segmented novel compounds at all while reading, because neither interviews nor retrospective think-aloud protocols were carried out. In addition, the reading experiment was run in two consecutive days. Consequently, it is not sure whether results in this study are generalizable to research that undergoes longer periods of time. The small number of items encountered by each participant—along with data loss during the data preparation stage—might also reduce the chance of finding significant results, especially when it comes to individual differences. Future studies will be needed to examine and compare the processing of novel words under incidental and intentional learning conditions, while addressing the above-mentioned limitations.

References

Amenta, S., Günther, F., & Marelli, M. (2020). A (distributional) semantic perspective on the processing of morphologically complex words. *The Mental Lexicon*, **15**(1), 62–78. https://doi.org/10.1075/ml.00014.ame

Bachman, L., & Palmer, A. (1989). The construct of validation of self-ratings of communicative language ability. *Language Testing*, **6**(1), 14–29. https://doi.org/10.1177/026553228900600104

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Neurosciences*, **4**(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48. https://doi.org/10.18637/jss.v067.i01

Betancort, M., Carreiras, M., & Sturt, P. (2009). The processing of subject and object relative clauses in Spanish: An eye-tracking study. *Quarterly Journal of Experimental Psychology*, **62**(10), 1915–1929. https://doi.org/10.1080/17470210902866672

Bisson, M. J., Kukona, A., & Lengeris, A. (2021). An ear and eye for language: Mechanisms underlying second language word learning. *Bilingualism: Language and Cognition*, **24**(3), 549–568. https://doi.org/10.1017/S1366728920000723

Brusnighan, S., & Folk, J. (2012). Combining contextual and morphemic cues is beneficial during incidental vocabulary acquisition: Semantic transparency in novel compound word processing. *Reading Research Quarterly*, **47**(4), 426–446. https://doi.org/10.1002/RRQ.015

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PloS ONE*, **5**(6), e10729. https://doi.org/10.1371/journal.pone.0010729

Chen, T. (2018). The contribution of morphological awareness to lexical inferencing in L2 Chinese: Comparing more-skilled and less-skilled learners. *Foreign Language Annals*, **51**(4), 816–830. https://doi.org/10.1111/flan.12365

Chen, T. (2021). Word learning in L2 Chinese: From perspectives of learner-related and language-related factors. *Journal of Psycholinguistic Research*, **50**(3), 663–680.

Conway, A., Kane, M., Bunting, M., Hambrick, D., Wilhelm, O., & Engle, R. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, **12**(5), 769–786. https://doi.org/10.3758/BF03196772

Daneman, M., & Hannon, B. (2007). What do working memory span tasks like reading span really measure? In N. Osaka, R. Logie & M. D'Esposito (Eds.), *The Cognitive Neuroscience of Working Memory* (pp. 21–42). New York: Oxford University Press.

Dronjic, V. (2011). Mandarin Chinese compounds, their representation, and processing in the visual modality. *Writing Systems Research*, **3**(1), 5–21. https://doi.org/10.1093/wsr/wsr005

Elgort, I., Brysbaert, M., Stevens, M., & van Assche, E. (2018). Contextual word learning during reading in a second language an eye-movement study. *Studies in Second Language Acquisition*, **40**(2), 341–366. https://doi:10.1017/S0272263117000109

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. https://doi.org/10.1111/lang.12052

Ellis, R. (2004). Individual differences in second language learning. In A. Davies (Ed.), *The handbook of applied linguistics* (pp. 525–551). Oxford: Blackwell.

Forster, J., Shipstead, Z., Harrison, T., Hicks, K., Redick, T., & Engle, R. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, **43**(2), 226–236. https://doi:10.3758/s13421-014-0461-7

Frisson, S., Niswander-Klement, E., & Pollatsek, A. (2008). The role of semantic transparency in the processing of English compound words. *British Journal of Psychology*, (99), 87–107. https://doi.org/10.1348/000712607X181304

Fukkink, R. G., Blok, H., & de Glopper, K. (2001). Deriving word meaning from written context: A multicomponential skill. *Language Learning*, 51(3), 477–496. https://doi.org/10.1111/0023-8333.00162

Gan, H. (2008). Yuyi toumingdu dui zhongji hanyu yuedu zhong cihui xuexi de yingxiang [The effects of semantic transparency on vocabulary learning in intermediate Chinese reading]. *Yuyan Wenzi Yiingyong*, (1), 82–90.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., ... **Yoon, H.** (2018). Incidental vocabulary learning in a reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, **21**(3), 563–584. https://doi.org/10.1017/S1366728917000219

Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, **35**(3), 483–517. https://doi.org/10.1017/S0272263113000119

Hong, W., Feng, C., & Zheng, Z. (2017). Yuyi toumingdu, yujing qiangdu ji cihui fuxian pinlv dui hanyu eryu cihui xide de yingxiang [Semantic transparency, contextual strength, lexical repetition and their influence on L2 vocabulary learning]. *Xiandai Waiyu*, **40**(4), 529–539.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, **11**(2), 207–223.

Hulstijn. (2003). Incidental and intentional learning. In C. Doughty & M. Long (Eds.), *The handbook in second language acquisition* (pp. 349–381). Oxford: Blackwell.

Hulstijn, J. (2001). Intentional and incidental second language vocabulary learning: a reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–287). Cambridge: Cambridge University Press.

Institute of Language Teaching and Research. (1986). *Xiandai hanyu pinlv cidian* [A frequency dictionary of contemporary Chinese]. Beijing: Beijing Institute of Language Pres.

Iwaizumi, E., & Webb, S. (2021). Measuring L1 and L2 productive derivational knowledge: How many derivatives can L1 and L2 learners with differing vocabulary levels produce?. *TESOL Quarterly*. https://doi.org/10.1002/tesq.3035

Jin, G., Xiao, H., & Fu, L. (2005). Xiandai hanyu yuliaoku jianshe ji shenjiagong [Construction and deep processing of contemporary Chinese corpus]. *Yuyan wenzi yiingyong*, **2**, 111–120.

Joseph, H., Wonnacott, E., Forbes, P., & Nation, K. (2014). Becoming a written word: Eye movements reveal order of acquisition effects following incidental exposure to new words during silent reading. *Cognition*, **133**(1), 238–248. https://doi.org/10.1016/j.cognition.2014.06.015

Juhasz, B. J. (2007). The influence of semantic transparency on eye movements during English compound word recognition. In R. van Gompel, M. Fischer, W. Murrary & R. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 374–407). Amsterdam: Elsevier.

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

Ke, S., Miller, R. T., Zhang, D., & Koda, K. (2021). Crosslinguistic sharing of morphological awareness in biliteracy development: A systematic review and meta-analysis of correlation coefficients. *Language Learning*, **71**(1), 8–54. https://doi.org/10.1111/lang.12429

Kieffer, M., & Box, C. (2013). Derivational morphological awareness, academic vocabulary, and reading comprehension in linguistically diverse sixth graders. *Learning and Individual Differences*, **24**, 168–175. https://doi.org/10.1016/j.lindif.2012.12.017

Kieffer, M., & Lesaux, N. (2012). Development of morphological awareness and vocabulary knowledge in Spanish-speaking language minority learners: A parallel process latent growth curve model. *Applied Psycholinguistics*, **33**(1), 23–54. https://doi.org/10.1017/S0142716411000099

Koda, K. (2000). Cross-linguistic variations in L2 morphological awareness. *Applied Psycholinguistics*, **21**(3), 297–320. https://doi.org/10.1017/S0142716400003015

Laufer, B. (1988). The concept of "synforms" (similar lexical forms) in vocabulary acquisition. *Language and Education*, **2**(2), 113–132.

Laufer, B. (1990). Why are some words more difficult than others? Some intralexical factors that affect the learning of words. *IRAL-International Review of Applied Linguistics in Language Teaching*, **28**(4), 293–308.

Li, D. (2003). Zhonggaoji liuxuesheng shiziliang chouyang ceshi baogao [Report of a sampled test on the volume of lexical acquisition by intermediate and advanced learners of Chinese as a foreign language]. *Ji'nan daxue huawen xueyuan xuebao*, **2**, 12–19.

Li, J. (2011). Xiandai hanyu cidian de ciyi toumingdu kaocha [A quantification analysis of the transparency of lexical meaning of modern Chinese dictionary]. *Hanyu xuebao*, **35**, 54–62.

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, **38**(4), 801–842.

Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language*, **61**(1), 30–44. https://doi.org/10.1006/brln.1997.1876

Lieber, R., & Stekauer, P. (Eds.). (2011). *The Oxford handbook of compounding*. Oxford: Oxford University Press.

Linck, J., Koeth, J., & Bunting, M. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, **21**(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

Lüdecke, D. (2019). sjPlot: Data visualization for statistics in social science. doi: 10.5281/zenodo.1308157, R package version 2.7.2, https://CRAN.R-project.org/package=sjPlot

Malone, J. (2018). Incidental vocabulary learning in SLA: effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition*, **40**(3), 651–675. https://doi.org/10.1017/S0272263117000341

Martin, K., & Ellis, N. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, **34**(3), 379–413. https://doi.org/10.1017/S0272263112000125

McBride-Chang, C., Tardif, T., Cho, J.-R., Shu, H., Fletcher, P., Strokes, S., . . . **Leung, K.** (2008). What's in a word? Morphological awareness and vocabulary knowledge in three languages. *Applied Psycholinguistics*, **29**(3), 437–462. https://doi.org/10.1017/S014271640808020X

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton, FL: CRC Press.

Mohamed, A. (2018). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*, **40**(2), 269–293. https://doi.org/10.1017/S0272263117000092

Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, **42**(4), 749–773. https://doi.org/10.1017/S0272263119000706

National Committee for Chinese Proficiency Test. (2001). *Hanyu shuiping cihui yu hanzi dengji dagang [Words and Characters for Chinese Proficiency Syllabus]*. Beijing: Jingji kexue chubanshe.

Pellicer-Sanchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, **38**(1), 97–130. https://doi.org/10.1017/S0272263115000224

Pelzl, E., Lau, E. F., Guo, T., & DeKeyser, R. (2019). Advanced second language learners' perception of lexical tone contrasts. *Studies in Second Language Acquisition*, **41**(1), 59–86. https://doi.org/10.1017/S0272263117000444

Peters, E. (2020). Factors affecting the learning of single-word items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 125–142). New York: Routledge.

Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes*, **20**(1/2), 261–290. https://doi.org/10.1080/01690960444000098

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Reichle, E., Pollatsek, A., Fisher, D., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, **105**(1), 125–157. https://doi.org/10.1037/0033-295x.105.1.125

Reichle, E., Pollatsek, A., & Rayner, K. (2006). E–Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research*, **7**(1), 4–22. https://doi.org/10.1016/j.cogsys.2005.07.002

Rothou, K., & Padeliadu, S. (2015). Inflectional morphological awareness and word reading and reading comprehension in Greek. *Applied Psycholinguistics*, **36**(4), 1007–1027. https://doi.org/10.1017/S0142716414000022

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, **21**(4), 589–619. https://doi.org/10.1017/S0272263199004039

Sasao, Y., & Webb, S. (2017). The word part levels test. *Language Teaching Research*, **21**(1), 12–30. https://doi.org/10.1177/1362168815586083

Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, **48**(2), 281–317. https://doi.org/10.1111/1467-9922.00042

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, **12**(3), 329–363. https://doi.org/10.1177/1362168808089921

Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching*, **52**(2), 261–274. https://doi.org/10.1017/S0261444819000053

Service, E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics*, **16**(2), 155–172. https://doi.org/10.1017/S0142716400007062

Speciale, G., Ellis, N., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, **25**(2), 293–321. https://doi.org/10.1017/S0142716404001146

Swanborn MSL, de Glopper K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, **69**(3), 261–285. https://doi.org/10.3102/00346543069003261

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, **69**(3), 559–599. https://doi.org/10.1111/lang.12343

Underwood, G., Petley, K., & Clews, S. (1990). Searching for information during sentence comprehension. In R. Groner, G. d'Ydewalle, & R. Parham (Eds.), *From eye to mind: Information acquisition in perception*. Amsterdam: Elsevier.

Waring, R., & Nation, P. (2004). Second language reading and incidental vocabulary learning. *Angles on the English-speaking World*, **4**, 11–23.

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). New York: Routledge.

Webb, S., & Chang, A. (2015). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, **37**(4), 651–675. https://doi.org/10.1017/S0272263114000606

Williams, J. N. (2012). Working memory and SLA. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 427–441). London: Routledge.

Wu, S. (2017). Cihui zhishi, yusu yishi, cihui tuili yu eryu yuedu lijie: Laizi jiegou fangcheng moxing de zhengju [Contribution of vocabulary knowledge, morphological awareness and lexical inferencing ability to reading comprehension: Evidence from SEM]. *Shijie hanyu jiaoxue*, **31**(3), 420–432.

Yang, Y., Shintani, N., Li, S., & Zhang, Y. (2017). The effectiveness of post-reading word-focused activities and their associations with working memory. *System*, **70**, 38–49. https://doi.org/10.1016/j.system.2017.09.012

Yi, W., Lu, S., & DeKeyser, R. (in press). Orthographic, semantic, and contextual influences on initial processing and learning of novel words during reading: Evidence from eye movements. *Chinese Journal of Applied Linguistics*.

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review*, **57**(4), 541–572. https://doi.org/10.3138/cmlr.57.4.541

Zhou, X., & Marslen-Wilson, W. (2000). Lexical representation of compound words: Cross-linguistic evidence. *Psychologia*, **43**, 47–66.

Zhou, X., Marslen-Wilson, W., Taft, M., & Shu, H. (1999). Morphology, orthography, and phonology reading Chinese compound words. *Language and Cognitive Processes*, **14**(5–6), 525–565.