System 107 (2022) 102825

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

Incidental learning of semantically transparent and opaque Chinese compounds from reading: An eye-tracking approach

Wei Yi[a,*], Robert DeKeyser[b]

[a] Peking University, China
[b] University of Maryland, College Park, USA

A R T I C L E I N F O

Keywords:
Incidental vocabulary learning
Reading
Semantic transparency
Compound words
Eye-tracking

A B S T R A C T

Using eye-tracking, this study examined the role of semantic transparency in L2 learners' incidental learning of compounds from reading and how real-time processing of novel compounds contributes to vocabulary gains. L2 Chinese speakers encountered 12 novel compounds embedded in stories, each occurring six times. Their knowledge of the compounds was tested immediately after the reading task. Results confirmed the advantage of transparent compounds over opaque compounds (in terms of form recognition and meaning recall), as well as the positive association between reading time summed across exposures and vocabulary gains (in terms of form and meaning recognition, regardless of semantic transparency). Our findings add to the literature on incidental vocabulary learning in that semantic transparency was found to moderate the acquisition order of different aspects of vocabulary knowledge, which highlights the importance of lexical characteristics for future research.

1. Introduction

Vocabulary plays an important role in the development of L2 skills. In recent years, it has been widely acknowledged that incidental vocabulary learning, especially via reading, is an important and necessary supplement to classroom instruction and promotes the growth of L2 vocabulary (Schmitt, 2008). Research interests in incidental L2 vocabulary learning from reading have been steadily growing during the past two decades (e.g., Chen & Truscott, 2010; Elgort, Brysbaert, Stevens, & Van Assche, 2018; Elgort & Warren, 2014; Godfroid, Boers, & Housen, 2013; Godfroid et al., 2018; Mohamed, 2018; Pellicer-Sánchez, 2016; Pellicer-Sanchez & Schmitt, 2010; Pigada & Schmitt, 2006; Tekmen & Daloğlu, 2006; Waring & Takaki, 2003; Webb, 2007; Webb & Chang, 2015; Yi, 2022; Zahar, Cobb, & Spada, 2001). Vocabulary learning is an incremental process (Barclay & Schmitt, 2019; Fukkink, Blok & de Glopper, 2001), and L2 learners often need to encounter novel words multiple times before the words can be fully acquired. Consequently, it is unsurprising to see that most studies on incidental vocabulary learning from reading have involved repeated exposures to novel lexical items. However, current literature on this topic is limited in that very few studies have looked into how lexical characteristics—especially semantic transparency—affect incidental acquisition of words. Furthermore, most studies have focused on the product of incidental vocabulary acquisition, leaving largely unknown whether real-time processing of novel lexical items is linked to vocabulary learning outcomes. To bridge these gaps, the present study used eye-tracking to investigate: a) whether semantic transparency impacts incidental vocabulary learning through reading; and b) whether attentional processing of compounds—quantified by

* Corresponding author. School of Chinese as a Second Language, No. 5 Yiheyuan Road, Haidian District, Beijing, 100871, China.
E-mail address: weiyisla@pku.edu.cn (W. Yi).

https://doi.org/10.1016/j.system.2022.102825
Received 17 August 2021; Received in revised form 24 May 2022; Accepted 24 May 2022
Available online 26 May 2022
0346-251X/

eye-tracking measures—predicts L2 learners' incidental learning of such lexical items.

2. Literature review

2.1. Incidental L2 vocabulary learning from reading

Vocabulary knowledge is crucial for L2 development (Hulstijn, 2015). Nation (2006) estimated that L2 learners of English need to have a vocabulary as large as 6000 to 7000 word families for listening, and 8000 to 9000 word families for reading. Learning such a large number of lexical items is definitely a daunting challenge. According to a survey by Laufer (2000, as cited in Schmitt, 2008), even after receiving 1800–2400 h of classroom instruction, English major students in China on average achieved a vocabulary size of only 4000 word families. Intentional vocabulary learning, delivered for example through classroom instruction, is generally regarded as being more effective than incidental vocabulary acquisition (Schmitt, 2008). However, given the limits of formal instruction in terms of time and resources, it is almost impossible for L2 learners to acquire a vocabulary as large as that estimated to be necessary by Nation (2006) via instruction alone. According to Nagy, Anderson, and Herman (1987), L1 English primary school students come across as many as 16,000–24,000 new vocabulary items per year, of which one third are unfamiliar words encountered incidentally during reading (Nagy & Anderson, 1984). Following this, incidental learning of words, especially via reading, has been widely accepted as an essential supplement to intentional, classroom-based vocabulary acquisition (Schmitt, 2008).

Incidental L2 vocabulary learning from reading is conceptualized as a process in which L2 learners are attending to the meaning of reading materials, with vocabulary knowledge being gained as a by-product (Hulstijn, 2003). Under a reading-for-meaning condition, L2 learners often need to encounter a novel word multiple times before it is fully acquired. A recent meta-analysis by Uchihara, Webb, and Yanagisawa (2019) confirmed the positive impact of repeated exposures on incidental vocabulary learning through meaning-focused activities (including but not limited to reading). After synthesizing 26 empirical studies that investigated incidental vocabulary learning through different input modes (i.e., reading, listening, reading while listening, and viewing), Uchihara and colleagues concluded that frequency of exposure had a medium non-linear effect (r = .34) on incidental vocabulary learning. Moreover, repeated exposures turned out to be more important for incidental vocabulary learning through reading (r = .41), compared with that through listening (r = .39) or viewing (r = .22).

The benefit of repeated exposures for incidental vocabulary learning has been demonstrated by many L2 studies. For instance, Pellicer-Sanchez and Schmitt (2010) carried out a study in which proficient Spanish learners of English were exposed to 34 words from an African language embedded in an unmodified English novel. After more than 10 exposures, 84% and 76% of the target words were recognized for meaning and spelling respectively, whereas meaning and word class were recalled at 55% and 63% respectively. Waring and Takaki (2003) found that after being exposed to target pseudowords in an English graded reader 15 to 18 times, Japanese EFL learners acquired 61%, 42%, and 18% of the target items in terms of form recognition, meaning recognition, and meaning production (translation), respectively. In another cross-sectional study, Rott (1999) selected 12 German words and embedded them in six passages. After encountering the target words six times, English learners of German successfully recognized 61% of the word definitions. Meanwhile, they could correctly translate 45% of the target words. As can be seen clearly, the above-mentioned studies vary substantially in the choice of stimuli (i.e., real L2 words vs. pseudowords vs. non-target-language words), the aspect of vocabulary knowledge (e.g., form vs. meaning, recognition vs. production), and the number of exposures, which may partially explain the variation in the reported vocabulary gains.

2.2. Semantic transparency and vocabulary acquisition

Words can be monomorphemic or multimorphemic—differing in terms of the number of morphemes that make up a lexical item. Studies have found that morphologically complex words are represented and processed differently from monomorphemic words (e.g., Zhou & Marslen-Wilson, 2000). For languages such as English, morphologically complex words are created mainly through inflection (e.g., *walk-walks-walked*) and derivation (e.g., *happy-happiness-happily-unhappy*). However, this is not the case for all languages. Chinese, for example, is morphologically poor and relies heavily on compounding to generate words. More than 70% of words in Chinese are compounds (Institute of Language Teaching and Research, 1986). Compounds consist of multiple morphemes, and the semantic relationship between the morphemes and the compound (i.e., the semantic transparency), varies. For semantically transparent compounds (e.g., *blueberry*), the meaning of the compound can be predicted by the meaning of its constituent morphemes. By contrast, for semantically opaque compounds, the meaning of the compound cannot (e.g., *bighorn*) or can only partially (e.g., *strawberry*) be deduced from its constituents (Libben, 1998).

Semantic transparency is a crucial factor to consider when the representation of compounds in the mind is examined. According to Libben (1998), semantically fully transparent compounds are represented through decomposed morphemes, whereas semantically fully opaque compounds are stored as full forms, similar to monomorphemic words. Transparent and opaque compounds differ in the degree of transparency in the semantic relationship between constituent morphemes and the compound. Therefore, it is likely that transparent compounds may be easier to acquire than opaque compounds, because learners can compute the meaning of the compound from its constituent morphemes (Mok, 2009). Although little research has investigated the influence of semantic transparency on vocabulary acquisition, results in the few studies to date are consistent. Chen (2020) reported that L2 learners of Chinese performed better when recalling transparent compounds than opaque compounds, after explicitly learning these lexical items without any contextual support. Brusnighan and Folk (2012) examined the incidental learning of novel compounds by native speakers of English. They found that participants were able to derive word meanings for both transparent and opaque compounds after a single exposure from reading sentences. Moreover, meaning recognition for novel transparent compounds (91%) was significantly better than novel opaque compounds (66%), regardless of whether the context was informative or not.

2.3. Capturing attentional processing with eye-tracking for prediction of vocabulary gains

Although interest in incidental L2 vocabulary learning from reading has been steadily growing in recent years, studies examining the relationship between online processing and incidental acquisition of L2 words, especially from the perspective of attention, are still lacking. Attention is necessary for second language acquisition (Schmidt, 1995, 2010). Attention is also generally regarded as a cognitive resource that can be allocated, quantified and modeled (e.g., Man & Harring, 2019, 2021). Broadly speaking, attention can be divided into two categories—overt and covert (Carrasco, 2011). Overt attention is accompanied with eye movements, whereas covert attention can be deployed without directing one's gaze toward the target of interest. According to the E-Z Reader model (Reichle, Pollatsek, & Rayner, 2006), overt attention—indexed by eye movements—is assumed to be allocated sequentially during reading, processing one word at a time. Consequently, by recording the eye movements during reading, attentional processing of lexical items can be largely captured by eye-tracking measures. Recently, a small group of studies (e.g., Godfroid et al., 2018, 2013; Mohamed, 2018; Pellicer-Sánchez, 2016) have found that attentional processing of novel lexical items—measured by various eye-tracking measures—is positively associated with incidental vocabulary gains from reading. Godfroid et al. (2013) had advanced learners of English read short passages that contained pseudowords while their eye movements were recorded. They found that participants who spent more total reading time on the pseudowords were more likely to recognize them in a gap-filling test administered immediately after the reading session. Following a design similar to that of Godfroid et al. (2013), Pellicer-Sánchez (2016) also reported a positive relationship between total reading time spent on target pseudowords and immediate vocabulary recall test scores. In Mohamed (2018), advanced L2 learners of English were exposed to 20 target pseudowords embedded in a graded reader. He found that total reading time spent on the pseudowords had greater effects than number of exposures on all kinds of vocabulary knowledge tests (i.e., form recognition, meaning recall, and meaning recognition). Additionally, first fixation duration and gaze duration were found to predict form recognition and meaning recall, respectively. Godfroid and colleagues (Godfroid et al., 2018) also carried out a study in which they examined L2 English learners' processing of unfamiliar Dari words embedded in an authentic English novel. Similar to Mohamed (2018), they found total reading time contributed to meaning recognition and meaning recall of target words, independently from effects of exposure frequency. Lastly, Elgort et al. (2018) used eye-tracking to examine Dutch-speaking L2 learners' acquisition of low-frequency English words from reading. By comparing the reading time of novel low-frequency words against those of familiar high-frequency words in semantically neutral contexts, they found that first fixation duration of novel and familiar words decreased over exposures, thus supporting the establishment of orthographic representation of novel lexical items.

3. The current study

The literature to date on incidental L2 vocabulary learning from reading is limited in the following ways. First, little research has investigated how semantic transparency impacts L2 learners' incidental learning of compounds. Second, evidence supporting the association between real-time processing of novel words and incidental vocabulary gains is still scarce. To bridge these gaps, the present study examined Chinese-*as*-a-second language learners' incidental acquisition of semantically transparent and opaque compounds while they were engaged in silent reading. To facilitate causal inferencing, part of speech, concreteness, lexical and sublexical (i.e., Chinese character) frequencies, lexical and sublexical (i.e., Chinese character) familiarity, as well as orthographic complexity (i.e., number of strokes) and prior knowledge, were controlled (see Table 1 under *Stimuli*). A repeated-measures design was adopted, in which learners were exposed to 12 target compound words embedded in 12 short stories written by the first author distributed on two consecutive days, with each target word occurring six total times across the stories. The decision to allow six exposures for each target word was based on the following concerns. First, previous studies (e.g., Rott, 1999) suggest that measurable vocabulary gains can be achieved after six exposures to novel L2 words. Second, with six exposures, experimental control for contextual confounds (especially the distribution of target words) was made possible (see *Stimuli*). To ensure strong ecological validity, L2 real words—instead of pseudowords—were used. Following the practice of Godfroid et al. (2018), Mohamed (2018), and Pellicer-Sánchez (2016), summed

Table 1
Characteristics of target Chinese compounds.

Characteristics	Transparent	Opaque
Word familiarity	1.00 (0.46)	0.60 (0.25)
Word frequency	1.85 (0.10)	1.50 (0.45)
Word strokes	15.50* (3.45)	10.83* (1.60)
First-character familiarity	1.97 (0.27)	2.25 (0.34)
First-character frequency	2.49 (0.30)	2.93 (0.37)
First-character strokes	9.17* (1.72)	6.00* (1.90)
Second-character familiarity	2.17 (0.30)	2.39 (0.17)
Second-character frequency	2.47 (0.48)	2.88 (0.29)
Second-character strokes	6.33 (2.66)	4.83 (0.75)

Note. Standard deviations are provided in the parentheses following the means. The asterisk * indicates significant differences between semantically transparent and opaque compounds.

total reading time (summedTRT) was calculated by summing total reading time across the six exposures, and was used to predict learners' vocabulary gains. To tap into the multifaceted nature of vocabulary knowledge, a form recognition test, a meaning recall test, and a meaning recognition test were administered immediately after the final reading session. Specifically, the current study sought to address the following research questions (RQs):

RQ_1. After encountering novel L2 compounds six times, how much vocabulary knowledge can be acquired incidentally from reading?
RQ_2. Do semantically transparent and opaque compounds differ significantly in terms of the amount of vocabulary knowledge acquired incidentally from reading?
RQ_3. Is attentional processing of novel compounds, quantified by summedTRT, predictive of the amount of vocabulary knowledge acquired by L2 learners?

4. Methodology

4.1. Participants

Sixty-one college students (43 females) were recruited from several universities in Beijing, China. To participate in this study, they had to be enrolled in intermediate-level Chinese-*as*-a-second-language classes. Additionally, their scores in a self-reported communicative language ability survey had to be between 36 and 72 (see *Experimental tasks*). To ensure that participants had no experience with Chinese characters other than that from learning Chinese as a second language, heritage speakers of all varieties of Chinese, as well as L2 Chinese learners who were speakers of Korean, Japanese, or Vietnamese as their first or second languages, were excluded. The participants' average age was 22.1 years ($SD = 2.9$). On average, they received 3.2 years of formal instruction of Chinese ($SD = 3.1$). Based on a six-point Likert scale (1 → 6: never → very frequently), their average self-reported frequency of interaction with Chinese native speakers was 4.4 ($SD = 1.1$), and their average self-reported use of Chinese when interacting with Chinese native speakers was 4.8 ($SD = 1.1$). All participants had normal or corrected-to-normal vision.

4.2. Stimuli

Target compounds were selected and used as stimuli based on the following procedure. To begin, a frequency-ranked word list containing 4354 disyllabic Chinese nouns was downloaded from the CNCORPUS website (Jin, Xiao, & Fu, 2005). To ensure that intermediate-level L2 Chinese participants would not know the target lexical items, words in the list with a frequency ranking below 5000 were excluded. Additionally, to ensure that learners would be familiar with the characters that constituted each target word, words consisting of individual Chinese characters with a frequency ranking beyond 1500 were not considered either. Following this, 387 nouns were kept in the list and subjected to a norming study in which their semantic concreteness and transparency were evaluated. Fifteen native Chinese speakers were randomly assigned to one of three lists, each containing 129 nouns. They were instructed to decide whether each word's referent was concrete or abstract based on whether the concept described by the word can be perceived via one of the five senses (i.e., sight, smell, hearing, taste, and touch). Candidate words perceived as concrete by no less than 80% of the raters were kept. With respect to semantic transparency, raters were instructed to evaluate the semantic compositionality of each target noun (Libben, 1998). Specifically, a constituent character/morpheme's meaning is considered transparent in a compound where it directly contributes to the meaning of the whole word (e.g., *you-deng*, oil-lamp, oil lamp). By contrast, a constituent character/morpheme's meaning is considered opaque in a compound where it has no direct relation to the meaning of the whole word (e.g., *shui-mu*, water-mother, jellyfish). For the sake of this study, only semantically fully transparent (i.e., both characters/morphemes relate to the meaning of the compound word) and fully opaque compounds (i.e., neither character/morpheme relates to the meaning of the compound word) were chosen. Subsequently, an online Qualtrics survey was delivered to eight intermediate-level Chinese-*as*-L2 learners, in which they were required to rate their degree of familiarity with the candidate compounds as well as the constituent characters, based on a four-point Likert scale (1: *I have never seen this word before*; 2: *I have seen this word before*; 3: *I know the meaning of this word*; 4: *I am familiar with this word*). Candidate compounds with an average familiarity rating higher than 1.5 were not kept. To make sure that participants would have a minimum amount of knowledge of the constituent characters/morphemes, those consisting of characters with an average familiarity rating of less than 1.5 were also removed. Following this, six semantically transparent compounds and six semantically opaque compounds were selected (see Appendix A in the supporting information online). Non-parametric Wilcoxon rank-sum tests showed that semantically transparent and opaque compounds were matched on various lexical and sub-lexical characteristics (Table 1), except for word strokes ($W = 1.5$, $p = .01$) and first-character strokes ($W = 3.5$, $p = .02$).

Given that it would be extremely difficult to embed all the target compounds in a single readable story, the twelve compounds were split into two groups (abbreviated as G) based on random assignment, but each containing three transparent compounds and three opaque compounds. For each group of compounds, a story was created. In order to control for the length of the stories, each story was designed to comprise 20 sentences. Target compounds were not embedded in the first or last sentence within each story. To control for the spacing of the target compounds, the remaining 18 sentences were divided into six blocks, each containing three consecutive sentences, and each block containing one compound randomly selected from the six candidates embedded. Within each block, the sentential location of the target compounds was also counterbalanced, so that the chance for a compound to appear in the first/second/third sentence was equal. Such a practice was repeated six times so as to generate six sets (abbreviated as S) of stories (i.e., S1G1-S6G2), allowing each target compound to occur six times (once in each story).

All stories were created by the first author following a narrative style of writing, on topics easily accessible to the participants (e.g., S1G1-Our physics teacher, S1G2-Xiaoming's grandfather). The stories were created following the pre-specified sentential locations for the target compounds (see Appendix B in the supporting information online). The target compounds did not appear in the initial or final position of the sentences. In addition, grammatical structures and words that might be unfamiliar to intermediate-level L2 learners of Chinese were avoided. All the stories were proofread by four native speakers of Chinese. Each story was rated in terms of readability by four native Chinese speakers based on a ten-point Likert scale (1 → 10: least readable→ most readable), with revisions made where necessary. On average, the stories achieved a high readability rating ($M = 8.0$, $SD = 0.6$).

The stories (see Appendix C in the supporting information online) were then tokenized (i.e., split into words) using Wordless (Ye, 2019). Overall, these stories consisted of 3600 characters, 2600 word tokens, and 700 word types. On average, each story contained 210 words ($SD = 10$) and 300 characters ($SD = 4$), with lexical coverage rates (i.e., 1- number of target compounds/total number of word tokens) ranging from 97.1% to 97.4%. Text characteristics of the stories are presented in Table 2.

For each story, four comprehension questions were created. Each comprehension question consisted of a statement, and participants had to judge whether each statement was correct based on their understanding of the stories. Comprehension questions did not include or address the target compounds. Three L2 speakers of Chinese at intermediate levels were recruited to pilot the comprehension questions, and revisions were made based on their responses. To estimate the contextual predictability of the target compounds in each sentence, a cloze test was also administered, with the target compounds replaced by blanks. Given that Chinese L2 participants in this experiment were supposed to have no knowledge of the target words, 10 native speakers of Chinese were recruited to fill in the blanks with appropriate words based on their intuitions. Their answers, including the exact target words or synonyms of the target words, as well as words that share the same semantic category with the target word (e.g., *lily* vs. *rose/orchid*), were treated as correct responses. Contextual predictability was calculated as the proportion of correct answers to the total number of responses for a target word, ranging from 0 to 1 ($M = 0.57$, $SD = 0.35$).

4.3. Experimental tasks

Screening Surveys. Participants were screened before being invited to the lab to ensure they met the participation criteria outlined in the previous section. Specifically, a demographic survey was delivered through Qualtrics, in conjunction with a self-rated communicative Chinese language ability survey (adapted from Bachman & Palmer, 1989). The communicative Chinese language ability survey consisted of 21 items distributed across three sections, targeting grammatical, pragmatic, and socio-linguistic competence respectively. Participants responded to the survey using four-point Likert scales, with their total scores ranging from 0 to 84. To ensure that participants' L2 proficiency would not be too low or too high for the purpose of this study, those achieving a total score below 36 or beyond 72 were excluded. Cronbach's alpha for this survey was .89, 95% CI [.85, .93].

Eye-tracking While Reading. The twelve stories were read by participants on two consecutive days. Based on random assignment, two lists (list A and list B) were first generated, each containing three sets of stories (i.e., six stories). The presentation order of the two lists were counterbalanced, such that half of the participants received list A on the first day, followed by list B on the second day, while the other half of the participants began with list B on the first day, followed by list A on the second day. Within each list, the presentation order of the stories was randomized across the sets and within each set. Each story was split into four screens, with five sentences on each screen. Before reading the stories, a practice text was provided to help familiarize the participants with reading in front of the eye tracker. Participants were instructed to read the stories naturally for meaning at their own pace. They were also notified that there would be comprehension questions following each story. Compulsory breaks were taken after finishing reading every two stories.

Participants' eye movements were automatically recorded by an EyeLink 1000 Plus eye tracker (SR Research, Canada; sampling rate: 1000 Hz). Participants read the stories binocularly, but only the right eye was monitored. The stories were presented in a normal, double-spacing manner on a 21-inch CRT monitor (resolution: 1024 × 768 pixels; refresh rate: 150 Hz) connected to a Dell PC. Chinese

Table 2
Text characteristics of the stories.

Story	Word tokens	Word types	Type-token ratio	Number of Characters	Average sentence length	Average readability rating	Lexical coverage rate
S1G1	216	118	0.55	297	10.3 (3.2)	7.6	97.2%
S1G2	222	118	0.53	304	11.1 (1.8)	8.9	97.3%
S2G1	234	134	0.57	300	11.7 (2.5)	7.9	97.4%
S2G2	224	125	0.56	304	10.7 (3.1)	7.9	97.3%
S3G1	217	122	0.56	308	10.9 (2.2)	8.0	97.2%
S3G2	213	121	0.57	302	10.7 (2.9)	7.6	97.2%
S4G1	220	124	0.56	295	11.0 (2.3)	8.0	97.3%
S4G2	221	124	0.56	301	11.1 (1.8)	8.7	97.3%
S5G1	206	129	0.63	298	10.3 (1.9)	7.7	97.1%
S5G2	206	118	0.57	297	10.3 (2.0)	8.9	97.1%
S6G1	208	116	0.56	301	10.4 (2.4)	7.0	97.1%
S6G2	213	133	0.62	293	10.7 (2.4)	8.3	97.2%

Note. Average sentence length was measured by number of word tokens, with standard deviations provided in the parentheses. The stories were labelled in the following way: set number-group number (e.g., S1G1).

characters were displayed in black, using Song 20-point font on a white background. Participants sat at a viewing distance of 70 cm from the computer monitor, with their head stabilized by means of a chin rest and a forehead rest. Nine-point calibrations were done at the beginning of the experiment and after each break. Additional calibrations were carried out for each participant when needed. Drift correction, in which a fixation point was displayed to allow the eye tracker to correct for small drifts in the calculation of gaze position, was set up at the beginning of each screen (Elgort et al., 2018; Godfroid et al., 2018; Mohamed, 2018). After the system was calibrated, the maximal gaze-position error was less than 0.5° in visual angle. The two reading sessions took about 60 min in total.

Vocabulary Knowledge Tests. A form-recognition test (see Appendix D in the supporting information online) was administered to test participants' receptive knowledge of the orthographic form of the target compounds. For each target word, two orthographically similar words were used as distractors, created by replacing the first/second character of the target word. Take the target word "天平 (scale)" for instance: by replacing either the first character "天" or the second character "平", two distractors, namely, "天使 (angel)" and "太平 (peace)" were created. The distractors and target compounds were matched for part of speech and word strokes (i.e., first-character strokes + second-character strokes). By adopting such a design, participants responded correctly only when they are able to retrieve the exact orthographic form of a given target compound. The 12 target compounds were then mixed with 24 distractors and divided into three blocks, each containing two transparent compounds, two opaque compounds, and eight distractors. Participants were instructed to choose the words that had appeared in the stories they had read. The order of the three blocks, as well as the order of the words within each block, were randomized. Participants received a point if they correctly chose a word that had appeared in the stories. If they missed a word that had occurred in the stories, they would receive a zero score instead. Cronbach's alpha for this test was .59, 95% CI [.43, .75]. This test took about 5 min and was delivered through Qualtrics.

A meaning-recall test (see Appendix E in the supporting information online) was created to measure participants' productive knowledge of the target words. Participants were asked to type out the meaning of the 12 compounds embedded in the stories. To encourage L2 learners' production and address the partial nature of their semantic knowledge, participants were instructed either to explain the exact meaning of the target words or provide synonyms or words semantically related to the target words, using Chinese or English. The presentation order of the target words was randomized. Participants received a score of 1 or 0, depending on whether they described the meaning of the target words correctly. A native speaker of Chinese was recruited to rate participants' responses based on the above-mentioned criteria, along with the first author. Inter-rater reliability computed by Cohen's Kappa was .85. Cronbach's alpha for this test was .79, 95% CI [.71, .87]. This test took about 10 min and was delivered through Qualtrics.

A meaning recognition test (see Appendix F in the supporting information online) was created to measure participants' receptive semantic knowledge of the target compounds. This was a multiple-choice test. For each target word, there were four choices, with "I don't know" also included as an option. Participants were instructed to choose the best option that explained the meaning of each word. Careful attention was given to the design of distractors, on the basis of the following principles and procedure. First, using a lexical inferencing task, ten L2 Chinese learners comparable to the participants in this study were instructed to guess the meaning of the target compounds, without contextual support. Incorrect responses were collected from them and considered as potential candidates for distractor choices. Unlike generating distractors solely dependent on researchers' intuition, such a practice had the benefit of ensuring that distractor choices would indeed be distracting to participants. These candidates were then evaluated and screened, so that they were of the same word class and were semantically consistent with at least one sentence context in which the target compound occurred. For the opaque compounds, at least one distractor was created based on their literal meaning. The choices were arranged in such a way that the number of correct answers for each choice (i.e., A, B, C, D) was the same. This test was delivered online through Qualtrics, with the test items randomized. This test took about 5 min. Cronbach's alpha for this test was .62, 95% CI [.47, .76].

4.4. Procedure

Participants who completed the screening surveys and met the criteria to participate in the study were invited to the lab. They were tested individually. The experiment was split into two sessions over two consecutive days. On day one, each participant read the first three sets of stories (6 stories total). As part of the design for a larger project, they also took additional cognitive tests following the reading phase. On day two, the participants read the remaining three sets of stories (6 stories). Immediately after the reading session, they took a series of surprise vocabulary knowledge tests. To avoid transfer effects, the participants started with the form recognition test, then the meaning recall test, and finally, the meaning recognition test. After completing the vocabulary knowledge tests, the participants received a brief survey, in which they reported how familiar they had been with each target compound before participating in the study. Participants were instructed to respond based on a four-point Likert scale (1: *I have never seen this word before*; 2: *I have seen this word before*; 3: *I know the meaning of this word*; 4: *I am familiar with this word*). The whole experiment took about 2 hours across the two days.

5. Statistical analysis

5.1. Data cleaning

Two participants turned out to be Chinese heritage speakers and started learning Chinese at very young ages. Seven participants had accuracy rates for the comprehension questions lower than 70%, indicating they either had difficulty understanding the stories or did not follow the instructions and read the stories for meaning. After removing these participants, 52 participants were kept for data analysis. Given that real L2 words—instead of pseudowords—were used in this study, self-reported familiarity of the target compounds collected from the participants was checked against their performance on the immediate vocabulary knowledge posttests. Specifically,

Table 3
Mixed-effects modeling results for vocabulary gains.

	Form Recognition					Meaning Recall					Meaning Recognition				
Parameters	Estimate	SE	z	p	exp(β)	Estimate	SE	z	p	exp(β)	Estimate	SE	z	p	exp(β)
Intercept	3.61	0.54	6.65	<.001	36.97	0.45	0.40	1.11	.266	1.57	0.52	0.45	1.15	.249	1.68
Transparency	−1.89	0.52	−3.64	<.001	0.15	−2.10	0.50	−4.19	<.001	0.12					
summedTRT	0.58	0.22	2.61	.009	1.79						0.38	0.14	2.72	.007	1.46
Comprehension	1.18	0.37	3.15	.002	3.25	1.12	0.25	4.46	<.001	3.06	0.17	0.18	0.94	.347	1.19
Comprehension:Transparency	−0.73	0.37	−1.98	.048	0.48						0.74	0.24	3.08	.002	2.10
Prior Exposure											−0.59	0.26	−2.25	.024	0.55
Random Effects															
Variance	0.80 $_{subject}$					1.88 $_{subject}$					0.29 $_{subject}$				
	0.05 $_{item}$					0.47 $_{item}$					1.84 $_{item}$				
Marginal R^2/Conditional R^2	0.330/0.467					0.286/0.584					0.150/0.485				

Note. Transparency was dummy-coded, with opaque compounds as the reference group. Similarly, Prior Exposure was dummy-coded, with compounds previously not encountered as the reference group. summedTRT and Comprehension were standardized. Marginal R^2 and Conditional R^2: refer to the proportion of variance explained by fixed effects and all effects (fixed and random effects), respectively. Model formulas include:

- Form Recognition ~ Transparency + summedTRT + Comprehension + Comprehension:Transparency + Prior Exposure + (1|subject) + (1|item)
- Meaning Recall ~ Transparency + Comprehension + (1|subject) + (1|item)
- Meaning Recognition ~ summedTRT + Comprehension + Comprehension:Transparency + Prior Exposure + (1|subject) + (1|item)

if a participant claimed to have already known the meaning of a target compound (i.e., selecting *3* or *4* in the survey) prior to participating in the experiment, and indeed he or she responded correctly in the semantic knowledge tests, then this compound was excluded from analysis for this person. Following this, 25% of the responses in the vocabulary knowledge tests were removed (i.e., 4.5% for opaque compounds, and 20.5% for transparent compounds). In order to examine whether attentional processing of novel compounds predicts L2 learners' vocabulary gains, the target compounds were defined as areas of interest on the eye tracker, and total reading time for each compound was extracted at each occurrence. The data cleaning procedure applied to vocabulary knowledge tests was also adopted for screening eye-tracking data, so that target compounds claimed and confirmed to have been known by the participants were not analyzed. Additionally, total reading time shorter than 80 ms were excluded from analysis (0.1%), as they are considered to be uninformative (Betancort, Carreiras, & Sturt, 2009).

5.2. Mixed-effects modeling

To assess how L2 learners incidentally acquire novel compounds from reading, three binary variables, namely, form recognition, meaning recall, and meaning recognition, were treated as dependent variables. The independent variables included semantic transparency and summedTRT. Semantic transparency (Transparency) was dummy-coded, with transparent compounds as the reference group. In order to quantify the amount of overall cognitive effort spent on processing each target compound, summedTRT was computed by summing total reading time across the six exposures, which was then rescaled through standardization and transformed to *z*-scores. Transparent and opaque compounds were matched for second-character strokes, but not for first-character strokes and word strokes. Given that word strokes and first-character strokes were highly correlated ($r = .84$), only first-character strokes (Strokes) were incorporated as a covariate and standardized. Additionally, participants' accuracy rates for the comprehension questions (Comprehension) were also included as a covariate, which was first transformed to a percentage scale and then standardized. Similarly, to control for the contextual predictability of each target compound (Predictability), contextual predictability ratings across each occurrence of a target compound were summed and then averaged, before being transformed to the percentage scale and standardized. Depending on whether participants reported to have previously encountered the target compounds before participating in the experiment, a binary variable Prior Exposure was created as a covariate and dummy-coded, such that words without prior exposure was treated as the reference group.

Mixed-effects logistic models were fit for form recognition, meaning recall, and meaning recognition separately, using the *lme4* package (version 1.1.21, Bates, Maechler, Bolker, & Walker, 2015) in R (version 3.6.2, R Core Team, 2019). Statistical models were implemented using a maximum likelihood technique, following forward model-selection procedures. Specifically, initial models with random intercepts for subjects and items were fitted, with effects of independent variables (i.e., Transparency, summedTRT, as well as the interaction between Transparency and summedTRT) and covariates (i.e., Strokes, Prior Exposure, Comprehension, Predictability, and interactions between each covariate and Transparency) added sequentially. Model comparisons were carried out using the *anova* function in the *lme4* package, based on log-likelihood tests. Finally, random slopes of fixed effects remaining in the so-far-best models were tested for subjects and items. For each best-fitted model, effect sizes measured by conditional R^2 (i.e., the proportion of total variance explained by both fixed and random effects) and marginal R^2 (i.e., the proportion of variance explained by fixed effects),

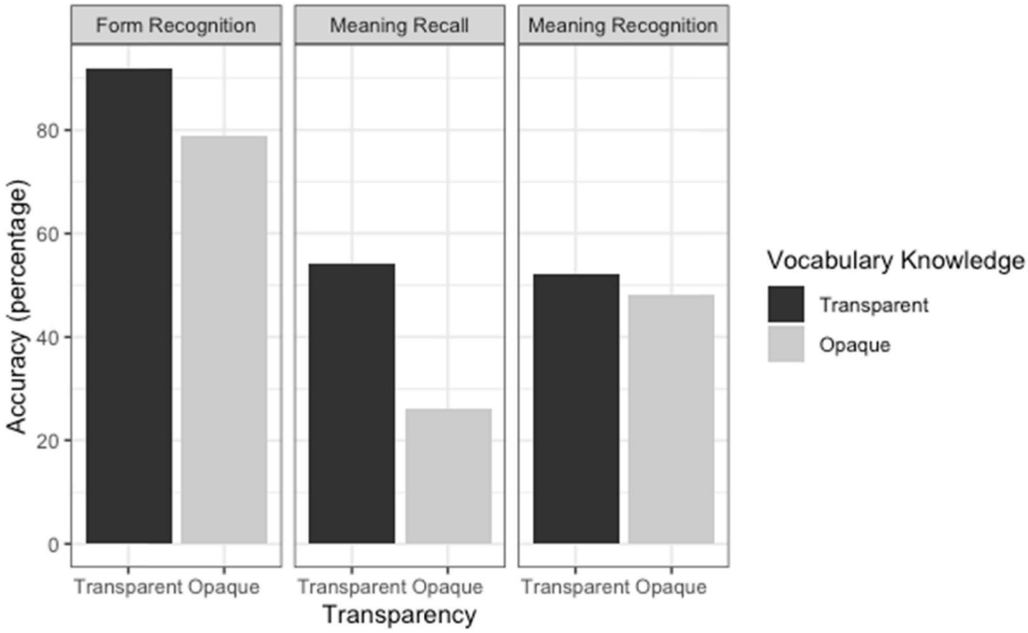

Fig. 1. Vocabulary gains acquired incidentally from reading.

along with *p*-values, were calculated using the *tab_model* function in the sjPlot package (version 2.8.2, Lüdecke, 2019). For logistic regression models, regression coefficient β corresponds to logged odds ratios (ORs). To interpret the results on the original scale, exponentiated coefficients exp(β), which referred to changes in the likelihood of achieving correct responses in terms of form recognition/meaning recall/meaning recognition, were reported. Specifically, ORs smaller than 1 indicate a decrease in the odds of achieving a correct answer, while ORs larger than 1 signal an increase in the odds of responding correctly. The significance level, *alpha*, was set at .05.

6. Results

Mixed-effects logistic modeling results for form recognition, meaning recall and meaning recognition are summarized in Table 3. Overall, fixed effects in the best-fitted models explained 33.0%, 28.6% and 15.0% of the variance for form recognition, meaning recall and meaning recognition, respectively. Taken together, fixed effects and random effects in the best-fitted models explained 46.7%, 58.4% and 48.5% of participants' variability in form recognition, meaning recall and meaning recognition, respectively.

6.1. *Vocabulary Gains Acquired Incidentally from reading*

In answer to the first research question, after six exposures under a reading-for-meaning condition, for the transparent compounds, the average accuracy on form recognition, meaning recognition, and meaning recall was 92% ($SD = 27\%$), 52% ($SD = 50\%$), and 54% ($SD = 50\%$), respectively. By contrast, for the opaque compounds, intermediate-level L2 learners of Chinese achieved an average accuracy of 79% ($SD = 41\%$), 48% ($SD = 50\%$), and 26% ($SD = 44\%$), respectively. Notably, for meaning recall (*Estimate* = 0.45, $SE = 0.40$, $z = 1.11$, $p = .266$) and meaning recognition (*Estimate* = 0.52, $SE = 0.45$, $z = 1.15$, $p = .249$), the intercepts in the best-fitted models were not significant, indicating that L2 learners' accuracies in these tests were not significantly higher than 50%. Average scores of form recognition, meaning recall and meaning recognition across conditions are visually presented in Fig. 1.

6.2. *Semantic transparency affects incidental vocabulary learning*

For both form recognition (*Estimate* = −1.89, $SE = 0.52$, $z = -3.64$, $p < .001$, exp(β) = 0.15) and meaning recall (*Estimate* = −2.10, $SE = 0.50$, $z = -4.19$, $p < .001$, exp(β) = 0.12), semantic transparency significantly impacts incidental learning of novel compounds from reading, with opaque compounds being much more difficult to acquire than transparent compounds. Specifically, for semantically opaque compounds, the odds of recognizing the orthographic form and recalling the meaning were 85% and 88% lower than semantically transparent compounds. However, incidental learning of compounds, as measured by meaning recognition, was not significantly predicted by semantic transparency, meaning that L2 learners achieved similar accuracies for transparent and opaque compounds in this test.

6.3. *Contextual influences on incidental vocabulary learning*

Two contextual covariates, namely, reading comprehension and contextual predictability, were incorporated for data analyses. For the opaque compounds, reading comprehension turned out to significantly predict L2 participants' form recognition (*Estimate* = 1.18, $SE = 0.37$, $z = 3.15$, $p = .002$, exp(β) = 3.25) and meaning recall (*Estimate* = 1.12, $SE = 0.25$, $z = 4.46$, $p < .001$, exp(β) = 3.06) performance. Reading comprehension also predicted L2 participants' vocabulary gains for the transparent compounds, yet only in terms of form recognition (evidenced by the significant effect of comprehension and the interaction between comprehension and transparency in Table 3) and meaning recognition (*Estimate* = 0.74, $SE = 0.24$, $z = 3.08$, $p = .002$, exp(β) = 2.10). Interestingly, contextual predictability, operationalized as the likelihood for a target compound to appear in any given sentence in the stories, was found to have no impact on L2 participants' vocabulary gains.

6.4. *The relationship between attentional processing and incidental learning of novel compounds*

On average, total reading time summed across six exposures was 9175.5 ms ($SD = 4314.7$) and 8779.2 ms ($SD = 4237.5$) for the transparent and opaque compounds, respectively. In answer to the third research question, summedTRT was found to significantly predict incidental learning of compounds from reading, in terms of form recognition (*Estimate* = 0.58, $SE = 0.22$, $z = 2.61$, $p = .009$, exp(β) = 1.79) and meaning recognition (*Estimate* = 0.38, $SE = 0.14$, z = 2.72, p = .007, exp(β) = 1.46). One standard deviation of increase in summedTRT was expected to enhance the odds of recognizing the orthographic form and meaning of novel compounds by 79% and 46%, respectively. Interestingly, summedTRT was not predictive of L2 learners' performance on the meaning recall test.

7. Discussion

7.1. *Incidental learning of transparent and opaque compounds from reading*

After six exposures, L2 learners obtained sizable gains in form recognition, for both transparent (92%) and opaque compounds (79%). For transparent compounds, vocabulary gains in meaning recognition (52%) and meaning recall (54%) were comparable. However, for opaque compounds, meaning recall (26%) was found to be much more difficult than meaning recognition (48%). Mixed-effects modeling revealed that semantic transparency significantly impacted incidental learning of L2 compounds from reading, with opaque compounds being much more difficult to acquire than transparent compounds, in terms of both form recognition and meaning recall.

The vocabulary gains reported in this study confirm the widely acknowledged benefit of repeated exposures for incidental vocabulary learning (Uchihara et al., 2019). In the present study, L2 Chinese learners were exposed to the novel compounds six times over two consecutive days, with the occurrences of the same novel compound in different contexts. Perfetti and colleagues (Bolger, Balass, Landen, & Perfetti, 2008; Reichle & Perfetti, 2003) put forward an instance-based framework to explain how word knowledge accrues after encountering a word repeatedly in context. According to this framework, each exposure to a novel lexical item creates a memory trace that encodes the word and its accompanying context; after experiencing the word in a variety of contexts, one's word-meaning knowledge becomes increasingly decontextualized. Acquiring a word does not only mean achieving substantial vocabulary gains measured by offline measures, but also robust lexical representations in the mental lexicon. From this perspective, needless to say, L2 learners' knowledge of the novel compounds is still partial and incomplete, and additional exposures in contexts are required before such lexical items become fully consolidated.

Current research on incidental vocabulary learning from reading has frequently incorporated target lexical items with different lexical characteristics, making it unrealistic to compare pickup rates of vocabulary knowledge across studies. However, it is still meaningful to compare the rates of uptake among different aspects of vocabulary knowledge within a study. Barclay and Schmitt (2019) argue that various aspects of vocabulary knowledge can hardly be acquired simultaneously and may progress at different rates. Using structural equation modeling, González-Fernández and Schmitt (2020) confirmed that recognition and recall knowledge are two distinct latent constructs, with the former acquired before the latter across all components of word knowledge. In line with this, recent studies suggest that incidental learning of L2 words from reading takes place sequentially, starting with form recognition, followed by meaning recognition, and finally meaning recall. For instance, Mohamed (2018) reported that average vocabulary gains of L2 pseudowords with varying numbers of exposures were highest in form recognition (42%), followed by meaning recognition (30%), and finally meaning recall (13%). Similar to Mohamed, after eight exposures to L2 pseudowords during reading, Pellicer-Sánchez (2016) found that L2 learners' accuracies were highest in form recognition (86%), followed by meaning recognition (75%), and finally meaning recall (55%). Results of the present study are partially consistent with these studies, in that the acquisition of initial vocabulary knowledge of opaque compounds did follow the above-mentioned order. Nevertheless, this study also adds to these previous findings, in that such a sequence may not always be strictly followed. Different from opaque compounds, transparent compounds can be morphologically decomposed by L2 learners, with their meanings derived by combining the constituent morphemes (Brusnighan & Folk, 2012; Chen, 2020; Mok, 2009). Therefore, it is unsurprising that meaning recall was not significantly more difficult than meaning recognition for transparent compounds.

The impact of semantic transparency on incidental L2 vocabulary acquisition is also worth examining in further detail. Little effort has been spent on studying the contribution of semantic transparency to the acquisition of compounds, despite the fact that compounding is a prominent way of generating words for languages such as Chinese, and that semantic transparency is crucial for understanding the acquisition of compounds. In a separate analysis focusing on the processing of novel L2 compounds across repeated exposures, no significant difference was found between the transparent and opaque compounds in terms of the amount of attention attracted from L2 learners. Therefore, one can hardly attribute the acquisition advantages of transparent compounds over opaque compounds (in form recognition and meaning recall) to potential differences in the allocation of cognitive resources. Instead, we believe such result patterns might largely result from the componential nature of transparent compounds, as mentioned in the beginning section of this manuscript. To be specific, L2 learners might have less difficulty combining the orthographic forms of the constituent morphemes of transparent compounds and maintaining them in their memory. Similarly, given that transparent compounds are semantically componential, L2 learners would also have less difficulty inferring their meaning from the constituent morphemes, even when contextual support is not available (as in the case of meaning recall). The absence of transparent compounds' advantage over opaque compounds in meaning recognition (52% vs. 48%), however, may largely result from the nature of the test. Unlike what happens in the meaning recall test, participants were presented with four choices—one correct answer and three distractors—in the meaning recognition test. The availability of multiple choices may help L2 learners to retrieve episodic memory of sentential contexts in which the compounds occurred, which is particularly necessary for the opaque compounds.

Lastly, as reported in the previous section, we also found contrasting patterns of two types of contextual influences on the acquisition of the target compounds. Specifically, L2 learners' vocabulary gains for both transparent and opaque compounds were positively associated with their comprehension of the reading materials, whereas no relationship was found between L2 learners' vocabulary gains and the contextual predictability of the target compounds. These results suggest that L2 learners did make use of contextual information to acquire the orthography and meaning of novel compounds (including those of transparent compounds), yet it was passage-level comprehension—rather than sentence-level predictability of the target compounds—that played a role. Such a pattern further confirms that L2 learners in the present study spent more effort comprehending the stories and did not explicitly attend to the target compounds (either by proactively predict them or infer their meaning using contextual predictability), which corresponds to the incidental nature of the reading task.

7.2. Linking attentional processing and incidental vocabulary acquisition using eye-tracking

In answer to the third research question, attentional processing of novel lexical items, operationalized as summed total reading time across repeated exposures, significantly predicted L2 learners' recognition knowledge of both transparent and opaque compounds, in terms of form and meaning. Meanwhile, summed total reading time was not found to be associated with meaning recall. Overall, such results confirmed the widely-accepted assumption that attention is associated with L2 acquisition (Schmidt, 1995, 2010). In line with previous studies (Godfroid et al., 2013, 2018; Mohamed, 2018; Pellicer-Sánchez, 2016), our results support the practice of using eye-tracking to capture and quantify the amount of attention allocated to novel lexical items, which helps establish the link between online lexical processing and incidental vocabulary gains.

However, the association patterns between summed total reading time and various aspects of vocabulary knowledge acquired incidentally from reading are not consistent. For example, results in Mohamed (2018) and the current study both supported a positive relationship between summed total reading time and form recognition, yet such a pattern was not found in Godfroid et al. (2018) and Pellicer-Sánchez (2016). When it comes to meaning recognition, Godfroid et al. (2018), Mohamed (2018) and the current study all revealed that it can be predicted by summed total reading time on novel lexical items, yet such results were not replicated by Pellicer-Sánchez (2016). Similarly, although Godfroid et al. (2018), Mohamed (2018) and Pellicer-Sánchez (2016) all reported a positive relationship between total reading time and meaning recall, the present study failed to corroborate the finding that longer reading time led to better meaning recall for transparent and opaque compounds. Mohamed (2018) is the first and only study to date to report that summed total reading time is predictive of all types of vocabulary knowledge, including form recognition, meaning recognition, and meaning recall.

Notably, the above-mentioned studies differ substantially in various aspects, such as the use of target items (L2 words vs. pseudowords vs. non-target-language words), the administration of vocabulary knowledge tests (e.g., whether test items are randomized), and the statistical analyses used. Despite these differences, given that the vocabulary knowledge tests in these studies follow similar designs, preliminary explanations for the aforementioned inconsistencies are not impossible. To begin with, the absence of the association between summed total reading time and recognition knowledge of form and meaning of L2 pseudowords in Pellicer-Sánchez (2016) could probably be due to the lack of statistical power as well as their statistical analysis. Pellicer-Sánchez only had six target items and 37 s language speakers. Moreover, instead of using more sophisticated statistical techniques such as mixed-effects modeling, she ran Wilcoxon rank-signed tests to obtain correlations between total reading time and vocabulary gains, without considering other confounding variables. Second, although Godfroid et al. (2018) did not find a significant impact of summed total reading time on form recognition, in a similar study (Godfroid et al., 2013), Godfroid and colleagues did report such an association even when participants only had one exposure to target L2 pseudowords. Therefore, the lack of predictive power of total reading time for form recognition in Godfroid et al. (2018) may largely result from their research design. Unlike the earlier study, Godfroid et al. (2018) used Dari words as target items and embedded them in English reading materials. Since L2 English participants had no systematic knowledge of Dari spellings, a form recognition test with Dari words as distractors may have caused unexpected difficulties, thus distorting the relationship between form recognition and total reading time. Finally, the lack of significant relationship between meaning recall and summed total reading time in the present study, standing in contrast to previous findings, is likely not attributable to factors such as experimental control and statistical analysis. One possible explanation lies in the choice of target stimuli. Mohamed (2018) and Pellicer-Sánchez (2016) both created pseudowords (e.g., *blef-kill*, *holter-house*), while Godfroid et al. (2018) presented Dari words (*harami-illegitimate child*) in the texts. Under such circumstances, to perform correctly in the meaning recall test, L2 learners could only rely on contextual information retrieved from memory—which is moderated by the amount of attention allocated. However, for transparent and opaque compounds, when asked to recall their meaning, L2 learners might take advantage of constituent morphemes to describe the meaning of the compound, though such a strategy was not helpful when dealing with opaque compounds. Such a strategy could still help, however, when the meaning recognition test was performed. the availability of multiple choices, as mentioned earlier, might drive participants to rely more heavily on another path, by retrieving contextual information stored in their memory.

8. Conclusion

This study adds to the L2 vocabulary acquisition literature by examining L2 learners' incidental learning of semantically transparent and opaque compounds from reading, using the eye-tracking technique. Results confirm that different aspects of vocabulary knowledge develop at different rates, yet the acquisition order of various aspects of vocabulary knowledge may be slightly different between transparent and opaque compounds. Semantic transparency was found to significantly impact the incidental learning of novel compounds from reading, with transparent compounds turning out to be much easier to acquire than opaque compounds in terms of form recognition and meaning recall. Finally, the positive association between online processing and incidental learning of novel lexical items is also confirmed. Nevertheless, given the componential nature of the meaning of transparent compounds as well as the nature of the meaning recall test, total reading time summed across repeated exposures only predicts L2 learners' form and meaning recognition, but not meaning recall. Considering that compounding plays a crucial role in many languages other than English and that opaque compounds cause more difficulties for L2 learners, more research effort is needed to promote the understanding and instruction of such lexical items in the future.

Author statement

Wei Yi: Conceptualization, Methodology, Investigation, Data curation, Writing-Original draft preparation. Robert DeKeyser: Supervision; Writing-Reviewing and Editing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.system.2022.102825.

References

Bachman, L., & Palmer, A. (1989). The construct of validation of self-ratings of communicative language ability. *Language Testing, 6*(1), 14–29.
Barclay, S., & Schmitt, N. (2019). Current perspectives on vocabulary teaching and learning. In G. Gao (Ed.), *Second handbook of English language teaching* (pp. 1–22). Cham: Springer.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.
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.
Bolger, D. J., Balass, M., Landen, E., & Perfetti, C. (2008). Context variation and definitions in learning the meanings of words: An instance-based learning approach. *Discourse Processes, 45*(2), 122–159.
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), 426-426.
Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research, 51*(13), 1484–1525.
Chen, T. (2020). Word learning in L2 Chinese: From perspectives of learner-related and language-related factors. *Journal of Psycholinguistic Research, 1–18*. https://doi.org/10.1007/s10936-020-09740-5
Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics, 31*(5), 693–713.
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.
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.
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.
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.
González-Fernández, B., & Schmitt, N. (2020). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied Linguistics, 41*(4), 481–505.
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. (2015). *Language proficiency in native and non-native speakers: Theory and research*. Amsterdam: John Benjamins.
Institute of Language Teaching and Research. (1986). *Xiandai hanyu pinlv cidian [A frequency dictionary of contemporary Chinese]*. Beijing: Beijing Institute of Language Press.
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.
Laufer, B. (2000). *Task effect on instructed vocabulary learning: The hypothesis of 'involvement'. Selected Papers from AILA '99 Tokyo* (pp. 47–62). Tokyo: Waseda University Press.
Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language, 61*(1), 30–44.
Lüdecke, D. (2019). *sjPlot: Data visualization for statistics in social science*. https://doi.org/10.5281/zenodo.1308157. R package version 2.7.2. https://CRAN.R-project.org/package=sjPlot.
Man, K., & Harring, J. R. (2019). Negative binomial models for visual fixation counts on test items. *Educational and Psychological Measurement, 79*(4), 617–635.
Man, K., & Harring, J. R. (2021). Assessing preknowledge cheating via innovative measures: A multiple-group analysis of jointly modeling item responses, response times, and visual fixation counts. *Educational and Psychological Measurement, 81*(3), 441–465.
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.
Mok, L. W. (2009). Word-superiority effect as a function of semantic transparency of Chinese bimorphemic compound words. *Language & Cognitive Processes, 24*(7–8), 1039–1081.
Nagy, W., & Anderson, R. (1984). How many words are there in printed school English? *Reading Research Quarterly*, 304–330.
Nagy, W., Anderson, R., & Herman, P. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal, 24*(2), 237–270.
Nation, P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review, 63*(1), 59–82.
Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition, 38*(1), 97–130.
Pellicer-Sanchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do *things fall apart*? *Reading in a Foreign Language, 22*(1), 31–55.
Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language, 18*(1), 1–28.
R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.
Reichle, E., & Perfetti, C. A. (2003). Morphology in word identification: A word experience model that accounts for morpheme frequency effects. *Scientific Studies of Reading, 7*, 219–237.
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.
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.
Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (vol. 9, pp. 1–63). Honolulu, Hawaii: National Foreign Language Resource Center.
Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. Chan, S. Chi, K. Cin, J. Istanto, M. Nagami, J. Sew, et al. (Eds.), *Proceedings of CLaSIC 2010* (pp. 721–737). Singapore: National University of Singapore.
Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363.
Tekmen, E. A., & Daloğlu, A. (2006). An investigation of incidental vocabulary acquisition in relation to learner proficiency level and word frequency. *Foreign Language Annals, 39*, 220–243.
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.
Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language, 15*(2), 130–163.
Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(1), 46–65.
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.
Ye, L. (2019). *Wordless* [Computer Software]. Retrieved from Version 1.1.0. https://github.com/BLKSerene/Wordless.
Yi, W. (2022). Processing of novel L2 compounds across repeated exposures during reading: A growth curve analysis. *Applied Psycholinguistics*, 1–29. https://doi.org/10.1017/S0142716422000017
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.
Zhou, X., & Marslen-Wilson, W. (2000). Lexical representation of compound words: Cross-linguistic evidence. *Psychologia, 43*, 47–66.