



Castledown

 OPEN ACCESS

Australian Journal of Applied Linguistics

ISSN 2209-0959

<https://www.castledown.com/journals/ajal/>

Australian Journal of Applied Linguistics, 4 (3), 82–102 (2021)
<https://doi.org/10.29140/ajal.v4n3.545>

What constitutes vocabulary learning difficulty? A classroom-based study of semantic relatedness and L1 familiarity effects on L2 word learning



HANZHONG SUN ^a

SHAOHUA FANG ^b

^a *Quzhou University, China*
sunhanzhong@foxmail.com

^b *University of Pittsburgh, USA*
shf64@pitt.edu

Abstract

There has been a plethora of studies investigating the effect of semantic relatedness on second language (L2) word learning. However, most prior studies failed to control for the lexical properties of target words, which may be responsible for the mixed results yielded. This study, therefore, sets out to revisit this issue by controlling for L1 familiarity, one of the lexical factors confirmed to impact L2 learnability. Another goal of this study is to explore the extent to which semantic relatedness interacts with L1 familiarity, thus helping us determine the suitable condition under which word learning takes place. Towards these two aims, four sets of English target words matched in length were created (i.e., words of high/low L1 familiarity placed in semantic related/unrelated sets). For each set, after the timed learning session, an immediate posttest and an unannounced one-week delayed posttest, both of which measured the receptive knowledge of the target words, were administered to forty-one English as foreign language (EFL) learners. The results showed that (a) semantic relatedness negatively affected L2 word learning on the delayed posttest only, regardless of L1 familiarity status, suggesting a robust hindrance effect; (b) L1 familiarity persistently served as a facilitative force, as it aided the acquisition of both semantically related and unrelated word sets on both posttests; (c) similar L1 familiarity levels might result in additional interference. Our findings highlight the complex interplay between these two input-related variables.

Keywords: semantic relatedness, L1 familiarity, L2 word learning, interference

Copyright: © 2021 Hanzhong Sun & Shaohua Fang. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within this paper.

Introduction

One of the essential tasks involved in second language (L2) word learning is to map lexical forms to meanings¹ (Henriksen, 1999; Jiang, 2002; Schmitt, 2008). Given the opaque relationship between these two aspects (DeKeyser, 2016), this process can sometimes be exceedingly challenging, especially for beginning learners, susceptible to a variety of item-, learner-, and context-related variables (Housen & Simoens, 2016; Suzuki, Nakata & DeKeyser, 2019). In terms of item difficulty, Housen and Simoens (2016) made a fine distinction between intrinsic feature properties and input properties. Of these two, the investigation of input-related features may likely yield more pedagogical suggestions regarding how to best promote learning outcomes, as these features can be manipulated with relative ease even in natural classroom settings.

Learners usually receive multiple exposures to the target language to master a word. One of the things weighed up in this process is whether new words should be introduced in semantically related versus unrelated sets. Interestingly, despite extensive scholarly attention (Erten & Tekin, 2008; Finkbeiner & Nicol, 2003; Hong & Feng, 2014; Hoshino, 2010; Ishii, 2015; Jiang, Jong, Tse & Chai, 2020; Nakata & Suzuki, 2019; Papathanasiou, 2009; Tinkham, 1993, 1997; Waring, 1997; Wilcox & Medina, 2013; Zhang & Sheng, 2009), results to date remain inconclusive, with all possibilities easily tracked down in relevant literature, leaving language practitioners more baffled. The mixed results generated from most earlier studies may be largely due to their failure to control for the lexical properties of the target words across each set (Nakata & Suzuki, 2019), making one set linguistically easier while adding extra burdens to the other.

The present study is an attempt to revisit this controversial issue of whether semantic relatedness affects L2 word learning in a natural classroom setting. Unlike most prior research, we controlled for L1 familiarity, a lexical factor confirmed to impact L2 word learnability (Chapman & Gilbert, 1937; Tagashira, Kida & Hoshino, 2010). Additionally, given the complexity of this learning endeavour, where multiple factors likely conspire to determine learning difficulty, our study also concerned how semantic relatedness interacted with L1 familiarity in influencing the ultimate learning gains. This empirical examination of these two input-relevant features may hopefully bring valuable pedagogical insights into vocabulary teaching/learning and coursebook writing alike.

Literature Review

Theoretical underpinnings regarding the semantic relatedness effect

Semantic relatedness refers to the semantic relation between two or more lexical items. Words that are semantically related to each other form a semantic set, such as *orange*, *apple* and *pear* (FRUIT). There is a widespread belief among language researchers and practitioners that presenting new vocabulary in semantically related sets, also known as semantic clustering, will speed its acquisition. This preferred instruction is supported by semantic fields theory (Lehrer, 1974) and gains a sound psycholinguistic basis. First, semantic priming studies (Anderson, 1995; Perea & Rosa, 2002) consistently demonstrate that a word (e.g., *apple*) is recognized faster when preceded by a semantically related prime (e.g., *pear*) compared to when preceded by an unrelated counterpart (e.g., *car*), which leads to a conclusion that words are semantically connected. This is also evidenced by word association tasks where participants tend to produce a number of semantically related words in response to the cue words (Aitchison, 1987), due to spreading activation over this network (Collins & Loftus, 1975). Therefore, it can be reasonably deduced that packaging new words into semantically related sets presents an effective method to teach and learn vocabulary, because it truly

reflects the way words are stored in the mental lexicon and helps students see how words can be organized (for more reasons, see Nation, 2001).

In contrast, some researchers have been questioning the effectiveness of this pedagogical practice based on interference theory (Baddeley, 1997), claiming that semantic clustering causes confusion within the set due to the similar concepts the words share, and therefore increases the learning difficulty. Distinctiveness hypothesis (Hunt & Mitchell, 1982), on the other hand, predicts that the dissimilarity of meaning may contribute to the ease with which semantically unrelated words are acquired.

These two theoretically informed competing positions have been well corroborated by a number of studies with learners from various L1 backgrounds, in distinct learning settings (i.e., natural classroom and lab), as well as with a wide range of outcome measurements. In what follows, we will report the different results from prior research.

Empirical evidence regarding the semantic relatedness effect

A broad inquiry into prior studies yields three contradictory results, that is, positive, negative, and null. We mainly review L2 studies pertaining to each of these results.

Positive effect of semantic relatedness on L2 word learning

Empirically, the superiority of semantically related sets, as opposed to unrelated sets, was borne out by a handful of studies. Hoshino (2010) tested the relative learning efficacy of five types of word pairs, including a categorical set and a mixed set, in a natural classroom setting. The learning gains were defined as the number of words whose translation or meaning could be correctly recalled. The analysis revealed that grouping words in the same category proved to be an effective way. With a different approach, Hashemi and Gowdasiaei (2005) adopted a between-subject design and employed the same set of words to create two conditions, that is, related and unrelated. Utilizing the Vocabulary Knowledge Scale as the assessment tool (Wesche & Paribakht, 1996), they found that the participants receiving semantically related set vocabulary instructions were more successful in terms of the gains of both breadth and depth knowledge, than another batch of students learning the same words but immersed in an unrelated condition.

Both these studies which were performed in real classrooms administered a one-shot testing session (Hoshino: three or four days after learning; Hashemi & Gowdasiaei: one week after learning). In doing so, little was known about the dynamics of learning outcomes. The current study addressed this shortcoming by using two posttests to capture if the variable behaved differently as a function of timing. This rich source of information not only enables us to gain a clearer picture regarding how this factor works, but also helps to provide targeted recommendations towards how to lessen learning burdens.

Negative effect of semantic relatedness on L2 word learning

Closely replicating Tinkham's (1993) studies targeting the English native-speaking population, Waring (1997) performed an experiment where Japanese EFL learners were instructed to learn artificial nouns assigned to L1 (Japanese) words in both semantic related and unrelated sets. In each set, an individually administered test continued until the participants could utter all the words correctly corresponding to L1 in a preset time limit (for each word) to meet the criterion of learning. Then the number of trials was recorded to determine the difficulty of these two sets. The results of

this replication study were quite similar to those of Tinkham (1993, 1997) which showed that it took learners a longer time to learn the words semantically linked to each other, providing strong evidence of the disadvantage of semantic clustering. Finkbeiner and Nicol (2003) asked English natives to learn L2 pseudowords, but they used oral translation tasks in both directions (L1-L2 and L2-L1). They were found to spend more time on translation for words learned in semantic sets in comparison with those in random sets, indicating a robust semantic interference effect.

These studies were undertaken in lab settings using pseudowords as materials, which might undermine ecological validity. Therefore, the conclusions may not be readily applicable to natural classroom settings. As a response, researchers also conducted a series of classroom-based studies employing real words. Erten and Tekin (2008) investigated this topic among Turkish four-graders. A picture-word matching task appropriate for this younger population was employed and the test completion time was recorded. It was found that learning in semantic unrelated conditions not only yielded better results, but also required much shorter time to complete the test. With Greek EFL learners as participants and L1 translation test as measurement, Papathanasiou (2009) also found that grouping words in semantic sets had a deleterious effect on vocabulary learning. By the same token, Zhang and Sheng (2009) also found the negative effect of semantic clustering among Chinese undergraduates. Another two studies conducted in Chinese context examined the collocations instead of single items. Hong and Feng (2014) asked learners of Chinese to acquire Chinese collocations presented in both semantic clustering and random sets. The immediate posttest, a blank-filling test and a multiple-choice test measuring the productive and receptive knowledge respectively, suggested that participants were placed greater burden when learning the highly confusable collocations grouped according to their semantic relation. The other replication work by Jiang et al. (2020) generated mixed results. They found that semantic clustering impeded the reception of English collocations but facilitated the production.

Though situated in real classrooms with relatively high external validity, these inquiries suffer from at least one limitation. Even though some studies attempted to control for L2 variables (Erten & Tekin, 2008; Zhang & Sheng, 2009), such as L2 word length and part of speech, they did not consider L1-related variables. Echoing Nakata and Suzuki (2019), we interpret these findings, if not all, as partially reflecting the lexical difficulty deriving from L1, not the learning condition alone (related/unrelated). In other words, L1 lexical features might constitute additional learning burdens when not well controlled for. One of these considerations should be L1 familiarity (Tagashira et al., 2010; also see our detailed discussion in the following section). To the best of our knowledge, few earlier studies along this line have recognized it as a potentially confounding variable, except for Nakata and Suzuki (2019). Motivated by this lacuna, the present study intends to explore the semantic relatedness effect on L2 word learning in a natural classroom by controlling for L1 familiarity. Teasing apart the influence of this lexical variable may provide us with a more clean and compelling result.

Null effect of semantic relatedness on L2 word learning

More interestingly, despite the predominance of adverse effects found in earlier studies, a more recent study revealed a lack of such an effect. Nakata and Suzuki's (2019) study is different from previous ones by rigorously controlling for a suite of L1-related and L2-related variables that might confound L2 word learning, such as L1 familiarity, L2 word frequency and L2 word pronounceability, thus making the items lexically equivalent across related and unrelated sets. They asked Japanese participants to perform a self-paced computer-delivered word learning task, followed by two posttests. Their results showed that semantic relatedness did not impose any effect on L2 word learning, in both immediate and one-week delayed posttests. To determine potential

interference, they also analyzed within-set errors in both sets, and found that more confusion occurred in semantically related sets than unrelated sets, regardless of the timing of the posttest. They then proposed a tentative argument to account for their differential findings, that is, the semantic relatedness both positively and negatively (seen in within-set errors) affected L2 word learning, and these two competing effects neutralized each other.

Although offering a fresh perspective, their study was not naturally implemented. Two aspects are worth noting. First, all sessions were completed on a computer screen. Second, the number of encounters with each item was strictly limited to four. This might preclude the detection of such an effect. Therefore, the conclusion derived from this tightly controlled experiment might not be transferred to authentic contexts where students often learn words outside class in their own manner. In addition, their study included a list of words whose L1 translation equivalent were all familiar based on a database. It might generate different results if items with unfamiliar L1 concepts were incorporated.

Effect of L1 familiarity on L2 word learning

In traditional EFL textbooks, new words are often paired with their L1 translations, and hence it is not surprising that EFL learners frequently memorize these new words by resorting to corresponding L1 translations. A substantial body of research has demonstrated that the use of L1 can facilitate L2 word learning, particularly at the initial stage of learning (Groot, 1998; Jin & Webb, 2020; Lotto & Macaro & Lee, 2013; Prince, 1996). However, a caveat should be made that characteristics of L1 need to be taken into careful consideration, such as L1 familiarity. This was acknowledged by Nation (2013) who proposed a general principle of learning burden: "...the more a word represents patterns and knowledge that the learners are already *familiar* with, the *lighter* its learning burden. These patterns and knowledge can come from *the first language*..." (emphasis added, pp 44-45). Chapman and Gilbert's (1937) study is probably the first of this kind to lend support to this principle. They asked their English-speaking participants to learn Hindustani words, which were arbitrarily paired with English words, creating two learning conditions: L1 familiar and L1 unfamiliar. The results revealed that L1-L2 associations were developed more quickly and permanently when L1 words were familiar to learners.

Pushing this L1 familiarity effect further, Tagashira *et al.* (2010) claimed that in L2 antonym pair learning, similar L1 familiarity levels might cause additional interference. They argued that when learners were exposed to an L2 antonym pair with two corresponding L1 forms all familiar to them, L1 forms tended to compete with each other, and as a result, this would increase more interference. By contrast, when learners encountered these words with half of L1 forms being familiar and the other half unfamiliar to them, L1 did not likely compete against each other. Put differently, the L1 familiarity effect may be weakened depending on whether L1 familiarity levels of the to-be-learned words are similar. They then conducted two studies to experimentally test this hypothesis. Their participants were instructed to learn three groups of L2 adjective antonym pairs with varying degrees of L1 familiarity (High-High, High-Low, Low-Low). With no time constraint during learning (Study 1), they found that one week after learning, Low-Low adjective pairs were the most difficult to learn, due to both low and similar L1 familiarity degrees. More intriguingly, when a time restriction was imposed at learning phase (Study 2), the three-week delayed posttest showed that Low-Low pairs and High-High pairs exhibited similar learning rate, both worse than High-Low combination. Of particular interest here is that the L1 familiarity effect did not emerge, standing in stark contrast to the principle of learning burden (Nation, 2013). Though no direct sound evidence was obtained to explain the absence of L1 familiarity effect, they deemed that the interference effect caused by the

similar familiarity (High-High) might be larger than the superior effect contributed by higher familiarity.

To recap, we found conflicting results regarding whether L1 familiarity contributes to L2 word pair learning, an issue that apparently warrants further investigation. It also remains to be seen whether the results of L2 adjective antonym pairs (a special type of semantically related words) from Tagashira et al. (2010) can be extended to words that are (a) nouns; (b) presented in a set, larger than a pair; (c) semantically connected to each other.

Research aims

We identified two lexical factors contributing to the learnability of L2 words by reviewing both previous empirical evidence and theories underlying these results. It has now become evident to us that the impact of each variable is still far from clear. As a response, the current study was carried out to examine the effect of semantic relatedness and L1 familiarity on L2 vocabulary learning, in search of a method that aims to enhance vocabulary learning efficiency.

Our first aim is to examine whether semantic relatedness has a significant impact on L2 word learning. Similar to earlier research (Erten & Tekin, 2008; Nakata & Suzuki, 2019; Papathanasiou, 2009; Wilcox & Medina, 2013; Zhang & Sheng, 2009), we conducted two testing sessions, that is, immediate posttest and one-week delayed posttest, allowing valid comparison between the current study and the previous ones. However, our study differed from most previous ones by creating two conditions, that is, high L1 familiarity and low L1 familiarity, therefore affording us the chance to explore the impact of semantic relatedness on learning purposefully selected L2 words (i.e., concrete non-cognate nouns) under each condition. Controlling for L1 familiarity may return more direct results, since it has been found to affect learning outcomes (Tagashira et al., 2010). Unlike previous studies (Nakata & Suzuki, 2019; Tagashira et al., 2010) that consulted a database for L1 familiarity ratings, we obtained these data by administering a subjective self-rating task to a batch of learners with similar backgrounds to the participants, such as age and education level.

We also probed how L1 familiarity exerted an influence on L2 vocabulary gains. Along this research strand, we extended Tagashira et al.'s (2010) research by including noun word sets as target items, as opposed to adjective word pairs in their study. Our study also attempted to investigate how semantic relatedness and L1 familiarity jointly functioned in determining our participants' lexical performances. It should be noted that we adopted a within-subject design, meaning that our participants were exposed to all word stimuli. This thus allows us to exclude possible learner-related factors and ensures valid comparisons across different word sets.

To sum up, inspired by existing studies stated above, and responding to the call for attending to the identified gaps, our study sought to address the following three research questions (RQs) in a natural classroom setting where our participants performed the tasks in a pen-and-pencil format with a sufficient amount of time. Under this design, semantic relatedness and L1 familiarity were considered as independent variables and L2 lexical performances in the form of test scores as dependent variables.

RQ1: To what extent does semantic relatedness impact L2 lexical performances on the immediate and delayed posttests?

RQ2: To what extent does L1 familiarity impact L2 lexical performances on the immediate and delayed posttests?

RQ3: To what extent does the interaction of semantic relatedness and L1 familiarity impact L2 lexical performances on the immediate and delayed posttests?

Method

Participants

A total of forty-six university first-year students in China participated in the present study. However, forty-one of them were retained in the final statistical analysis based on the following inclusion criteria. First, they had no prior knowledge of the target words (meaning, in this study), as measured in a pre-test. Second, they received all the treatments at the learning phase. Third, they completed all the tests at the testing phase. Fourth, they did not receive additional exposure to the target words during the interval between immediate posttest and delayed posttest, according to their self-report. And lastly, they provided all of the background information needed in the current study, which enabled us to grasp a better picture of them.

The remaining participants comprised twenty-five male and sixteen female students. They ranged in age from 17 to 20 ($Mean = 18.71$, $SD = 0.72$). All the participants were native speakers of Chinese, and they learned English mainly in classroom settings with an average length of 7.90 years ($SD = 1.88$). All of them sat the National Matriculation English Test to gain admission into the university they have been attending, with a self-reported average score of 85.61 ($Min = 41$, $Max = 103$, $SD = 14.49$) out of 150. They also sat the College English Test, Band Four (CET-4)² one and a half months after participating in this experiment. The test scores, available upon request for teaching/research purposes from the Teaching Affairs Office of the University, indicated that none of them exceeded 425 out of 710. This means that none met the basic goal stipulated in the College English Curriculum Requirements. Despite the temporal interval, we argue that the scores they later obtained in CET-4 could reflect their English language proficiency at the time of experimenting. Therefore, they could be reasonably classified as low-to-intermediate learners of English. Additionally, none of them had any self-reported learning disorders.

Materials

Four word sets were created in the current study: semantically related L2 words with high L1 familiarity (Related-High) and low L1 familiarity (Related-Low), and semantically unrelated L2 words with high L1 familiarity (Unrelated-High) and low L1 familiarity (Unrelated-Low). Each set consisted of ten target words, and all of the words were of the same grammatical category, i.e. concrete nouns.

The two-step procedure was followed in the selection of the target words. Firstly, a pool of eighty L2 words, twenty for each set, were selected from three sources: (a) previous studies (Zhang & Sheng, 2009), (b) *Oxford Advanced Learner's Dictionary* (2020), and (c) word association by both authors. The preliminary words were non-cognate for the participants. Besides, none of them were compound or derivative nouns. These considerations were important because we intended not to introduce other confounding variables. For the semantically related sets, we focused on two categories, vegetable for high L1 familiarity set and musical instrument for low L1 familiarity set. The selection of these two categories was based on our experience that L1 equivalents under the vegetable category were assumed to be highly familiar, whereas most of the L1 characters falling into the latter category were of relatively low familiarity but could be easily recognized as a music instrument. We included in each of the semantically unrelated sets a variety of categories, such as fruit, human organ and ball. The L1 characters of these categories were of varying degrees of familiarity.

Secondly, a screening task was performed to further ensure the appropriateness of word selection. We asked seventy-nine students (aged 18.68 on average), sampled from the same native population but not attending the main experiment, to rate the familiarity of the L1 translation equivalents of the eighty L2 words on a seven-point scale, with “1” indicating minimum familiarity and “7” pointing to the highest degree of familiarity. Besides, given that the length of words has been shown to be a reliable predictor of visual word recognition (Ellis & Beaton, 1993), L2 word length therefore should be equivalent across the four word sets under investigation. In the current study, the length of word was operationalized as both the number of letters (New, Ferrand, Pallier, & Brysbaert, 2006) and the number of syllables (Ellis, 1995; Lovatt, Avons, & Masterson, 2000).

With these concerns in mind, forty L2 words, ten for each set, were determined as target words in the current work. The lexical properties of each set are presented in Table 1 and fully listed in Appendix A is the detailed information about each word. Two one-way ANOVA tests revealed that no significant differences were found in the number of English letters among the four sets ($F(3, 39) = 0.534, p = .662$), and there were also no significant differences in the number of L2 word syllables across these four conditions ($F(3, 39) = 0.479, p = .699$). In terms of L1 familiarity, an array of paired sample t-tests were conducted, which showed that (a) no significant differences emerged between the two high L1 familiarity sets ($p = .544$) and between the two low L1 familiarity sets ($p = .756$); but (b) significant differences were indeed found between the two semantically related sets ($p < .001$) and between the two semantically unrelated sets ($p < .001$). Therefore, words that were said to be of high familiarity were significantly more familiar to the participants in comparison with those of low familiarity.³

Table 1 Mean (SD) of the number of letters and syllables, and L1 familiarity ratings

Set (n=10)	No. of letters	No. of syllables	L1 familiarity	Example words
Related-High	5.60 (1.26)	1.90 (0.57)	5.93 (0.40)	celery, turnip
Unrelated-High	5.50 (1.27)	1.67 (0.67)	5.80 (0.41)	blade, aisle
Related-Low	5.90 (1.45)	2.00 (0.82)	2.63 (0.36)	tuba, valve
Unrelated-Low	6.20 (1.48)	2.10 (0.99)	2.57 (0.45)	plaid, avocado

Instruments

We devised a meaning-recall test (i.e., pre-test), two sets of meaning-recognition tests (i.e., immediate posttest and delayed posttest), a self-perceived difficulty questionnaire and a short background questionnaire.

Meaning Recall Test. The meaning recall test, served as the pretest, was designed to measure prior knowledge of the form-meaning connection of forty target words used in this study (See Appendix B). We also included in this test the same number of filler words which we assumed were quite familiar to the participants. The inclusion of these words was to ensure that they could be well motivated to complete this test. If only the target words were included, they were more likely to become discouraged, and as a consequence, they might not take this test seriously. However, only the responses to the target words were considered in the analysis. During this test, the participants were asked to circle the words familiar to them, and then to translate them into Chinese or explain the meanings of them in English.

Meaning Recognition Tests. For each word set, two sets of meaning recognition tests were used as posttests, that is, immediate and delayed, to assess the effects of variables in question. We administered a total of eight meaning recognition tests. Of note, we used recognition test format (See the example test in Appendix C), that is, giving multiple options to choose from (Laufer & Aviad-Levitzky, 2017), rather than the recall test (as the one used in pre-test), because it might be overly challenging to our low proficiency participants if no options were provided, which might obscure the learning effects. In order to minimize the guessing effect, we added another five distractors to the option pool. Hence, the test of each set comprised ten target words and fifteen options (Chinese characters). The two sessions of tests (immediate and delayed) were identical, except for the sequencing. We reordered the target items and the options to avoid the possible influence of memory trace.

Self-perceived Difficulty Questionnaire. To triangulate the experimental data, we also asked the participants to rank the difficulty level of the four word sets in ascending order, with “1” representing the least difficulty level and “4” the highest level.

Short Background Questionnaire. This questionnaire asked for demographic information (i.e., gender and age), student identity number⁴, time spent in learning English, score of National Matriculation English Test and a yes-or-no question concerning their learning disorder.

Procedures

The whole experiment was undertaken in a natural classroom setting in China. Typically, there were four periods (two periods for each time) of English classes per week. At the beginning, the participants were briefed on the purpose of this study, namely, to examine how well they could memorize the words presented to them within a limited amount of time. The study spanned three consecutive weeks, with the second week being the learning and immediate posttest sessions and the third the delayed posttest session. It should be noted that we divided our participants randomly into four groups with each represented by eleven or twelve students, and thus we were able to adopt Latin Square Design, displayed in Table 2. The purpose of it was to avoid possible influence of the presentation order.

In the first week, all the participants took the pre-test in the form of meaning-recall. Those demonstrating any knowledge about at least one target word were excluded from data analysis, though they were involved in this experiment throughout. This was to ensure that the participants had no knowledge prior to the learning sessions. No time constraint was imposed on this test.

The participants were officially engaged in the learning session in the second week. They were given a word card containing ten target words as well as their Chinese translation. They then were asked to memorize these words within ten minutes⁵ in whatever way that might suit them. They were also informed of the presence of a test which measured these words immediately after the timed learning session, such that they could make every endeavor to commit these words to memory. Ten minutes later, we collected the word card, and handed out to them the meaning-recognition test. Although no time limit was imposed, most of them completed each test within three minutes. However, they were not aware of the delayed test which took place one week later (the third week). This unannounced test allowed us to examine the retention of the learning gains. This cycle covering the learning, immediate testing and delayed testing sessions repeated four rounds. At the end of this experiment, the participants filled in a self-perceived difficulty questionnaire and a short biographical questionnaire. The former one aimed to explore whether their perception of the difficulty level of the

word sets was positively associated with their performances on these word tests, and the latter one allowed for a better understanding of the participants.

The scoring of these tests was solely done by the first author. One point was awarded to one correct response and 0 point was given to the incorrect or the missing responses. The total score of each set was 10 points.

Table 2 *Procedures of the experiment*

		Day 1	Day 2	Day 3	Day 4
Week 1	All groups	Pre-test			
Week 2 (learning + immediate posttest)	Group 1	Related-High	Unrelated-High	Related-Low	Unrelated-Low
	Group 2	Unrelated-Low	Related-High	Unrelated-High	Related-Low
	Group 3	Related-Low	Unrelated-Low	Related-High	Unrelated-High
	Group 4	Unrelated-High	Related-Low	Unrelated-Low	Related-High
Week 3 (delayed posttest)	Group 1	Related-High	Unrelated-High	Related-Low	Unrelated-Low
	Group 2	Unrelated-Low	Related-High	Unrelated-High	Related-Low
	Group 3	Related-Low	Unrelated-Low	Related-High	Unrelated-High
	Group 4	Unrelated-High	Related-Low	Unrelated-Low	Related-High
	All Groups	(Day 4) Self-perceived Difficulty Questionnaire Short Background Questionnaire			

Results

In this section, we report the results of immediate post-test and delayed post-test separately. The data analyses were conducted by using the statistical software R (R Development Core Team, 2016). The *lme()* function in the *nlme* package was used to build statistical models. In doing so, we first constructed a baseline model that only contained the intercept of the test score. Then the investigated variables, that is, relatedness and L1 familiarity, and their interaction were incrementally added to this model to further construct three models. We used *anova()* function to determine the main effects of these two variables and their interaction by comparing these three models. If the interaction effect was found, *emmeans()* function was employed to decompose this effect. We also tabulated the participants' subjective rank of the difficulty level of the four word sets in an effort to triangulate these experiment results. The descriptive statistics of the test scores was presented in Table 3 and graphically displayed in Figure 1.

Table 3 *Mean score (SD) of the participants' performance on the tests (N = 41)*

Test	Related-High	Unrelated-High	Related-Low	Unrelated-Low
Immediate Posttest	9.32 (1.60)	9.78 (0.57)	9.27 (1.45)	8.90 (1.89)
Delayed Posttest	5.54 (2.64)	7.85 (2.47)	3.73 (2.82)	4.83 (2.53)

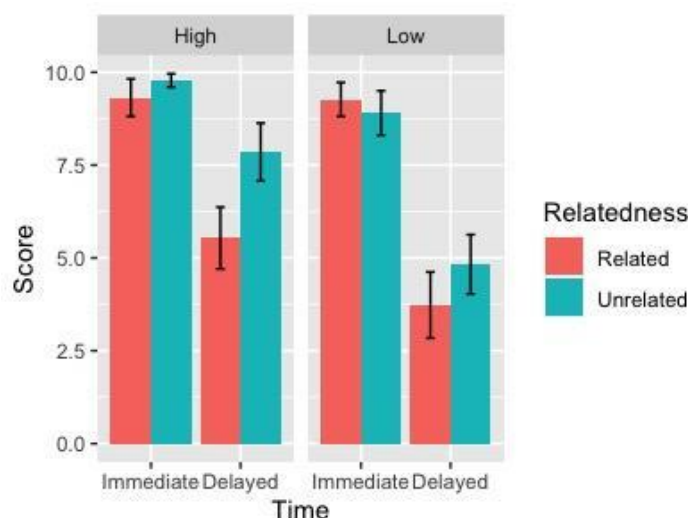


Figure 1 Mean scores of the four word sets by L1 familiarity and semantic relatedness (The error bars indicate 95% CIs.)

Results of the immediate posttest

The results revealed that the effect of semantic relatedness was not significant ($\chi^2(1) = 0.05, p = .8167$), which indicates that immediately after learning, semantic relatedness neither facilitated nor inhibited L2 word learning when high L1 familiarity and low L1 familiarity items were combined. However, a significant effect of L1 familiarity was found ($\chi^2(1) = 4.95, p = .0262$), suggesting that L1 familiarity affected L2 word learning. Words of high L1 familiarity were learned better compared with those with low L1 familiarity without taking into consideration the semantic relatedness. More importantly, the interaction between semantic relatedness and L1 familiarity also emerged ($\chi^2(1) = 4.11, p = .0427$).

To decompose this interaction effect, simple effect analyses were then conducted, showing that on the immediate posttest, no statistically significant difference was found between the semantically related and unrelated words either in high L1 familiarity condition ($\beta = -0.46, SE = 0.29, t = -1.60, p = .1132$) or in low L1 familiarity condition ($\beta = 0.37, SE = 0.29, t = 1.26, p = .2103$). However, words of high L1 familiarity were better learned than those of low L1 familiarity only in semantically unrelated sets ($\beta = 0.88, SE = 0.29, t = 3.02, p = .0031$), but not in related sets ($\beta = 0.05, SE = 0.29, t = 0.17, p = .8669$).

Taken together, these findings indicate that on the immediate posttest, (a) semantic relatedness did not produce any significant effects on learners' lexical performances across two L1 familiarity levels, and (b) L1 familiarity effect was contingent upon semantic relatedness, since a positive effect was observed only within the group of L2 words which bear no semantic relation.

Results of the Delayed Posttest

A series of similar statistical procedures were performed to examine the main effect of semantic relatedness and L1 familiarity, and the interaction effect between these two factors. The results of the delayed posttest displayed a significant effect for semantic relatedness ($\chi^2(1) = 17.05, p < .0001$). Words in semantically unrelated sets were retained better in comparison with the words learned in related settings irrespective of L1 familiarity. Besides, the main effect of L1 familiarity also reached significance in that high L1 familiarity words were scored higher than their counterparts with low L1

familiarity ($\chi^2(1) = 43.40, p < .0001$). Furthermore, the interaction between semantic relatedness and L1 familiarity was marginally significant ($\chi^2(1) = 3.36, p = .0666$).

Subsequent simple effect analyses showed that in the delayed posttest (a) the mean score of semantically related words were significantly lower than that of semantically unrelated set in both high L1 familiarity ($\beta = -2.32, SE = 0.47, t = -4.90, p < .0001$) and low L1 familiarity condition ($\beta = -1.10, SE = 0.47, t = -2.32, p = .0219$), and (b) words of high L1 familiarity received higher scores in comparison with words of low L1 familiarity in semantically related condition ($\beta = 1.80, SE = 0.47, t = 3.82, p = .0002$) as well as in unrelated condition ($\beta = 3.02, SE = 0.47, t = 6.40, p < .0001$).

Overall, the reported scores in delayed posttest suggest a robust hindrance effect of semantic relatedness and also a facilitative role of L1 familiarity.

Results of the Self-Perceived Difficulty

To gain more insights into the effects of these variables on the learning outcomes, we asked forty-one valid participants to rank the difficulty level of each word set, with “1” signaling the least difficult. The results from three participants were discarded, because they did not complete this questionnaire as required.

A close examination of the remaining ordinal data reveals that their perception was partially in line with the above experimental results. Similar patterns could be identified within the two sets with high L1 familiarity, as well as within the other two sets whose L1 equivalents were less familiar. For instance, the participants were more likely to approach the two sets of high L1 familiarity with relative ease, and they considered the other two sets as relatively more difficult. As observed in Table 4, the figures from Related-High and Unrelated-High sets were quite close. Nevertheless, the two related sets with different degrees of L1 familiarity (related-high versus related-low) displayed the opposite pattern, with the low L1 familiar one regarded as more difficult. This dissimilar pattern also applied to the other two sets with no semantic relation (unrelated-high versus unrelated-low).

Based on these results, we safely conclude that their judgement of difficulty might be largely driven by L1 familiarity rather than semantic relatedness. This concurs with the results in the immediate posttest where only L1 familiarity effect was found.

Table 4 Participants' ranking of their perceived difficulty level of the four word sets ($N = 39$)

	1 (easy)	2	3	4 (difficult)
Related-High	17 (44%)	8 (21%)	5 (13%)	9 (23%)
Unrelated-High	16 (41%)	13 (33%)	6 (15%)	4 (10%)
Related-Low	3 (8%)	7 (18%)	16 (41%)	13 (33%)
Unrelated-Low	3 (8%)	11 (28%)	12 (31%)	13 (33%)

Discussion

The present work probed the effects semantic relatedness and L1 familiarity had on L2 vocabulary learning. The participants were tested both immediately and one week after learning. Since we obtained differential results in each test session, the implications will be discussed separately in the following text. Regarding the immediate learning outcomes, our findings showed that packaging newly-learned L2 words into semantically related sets did not produce any impeding or facilitating

learning effect, whether these words were of high or low L1 familiarity. This conflicts with a number of earlier studies, most of them exhibiting a hindering effect. We claim that these differences might be partly attributed to various outcome measurements, including receptive translation test (i.e., translating L2 words into L1, Hoshino, 2010; Papathanasiou, 2009; Wilcox & Medina, 2013; Zhang & Sheng, 2009), trial-to-criterion test (i.e., recording the number of learning trials until certain criteria are reached, Higa, 1963; Waring, 1997), Vocabulary Knowledge Scale (Hashemi & Gowdasiaei, 2005), picture-word matching task (Erten & Tekin, 2008), contextualized blank-filling/multiple-choice test (Hong & Feng, 2014), among others. However, our study employed a meaning recognition test as a more direct method to tap into L2-L1 connection, which simply required the participants to select an appropriate L1 meaning from a pool based on a given L2 word. Our receptive task, we argue, demands fewer cognitive efforts than most tasks mentioned above, particularly the L2-L1 translation task. In addition to the facility of the measurement used here, our participants were also given ample time for the memory tasks and meanwhile totally aware of the upcoming vocabulary tests. Putting these factors together, our null-effect result came as no surprise because most participants performed extraordinarily well, which is displayed by our data clearly showing that the average score of each set reached 9 out of 10 (90%), leading to a ceiling effect.

Despite the high accuracy rate for each set, L1 familiarity was found to play a role, since our analysis showed such a main effect. It is particularly interesting to note that this effect was only present in semantically unrelated conditions. We speculate that the absence of such an effect in two related sets might be the result of additional interference caused by similar L1 familiarity levels (Tagashira et al., 2010). In our study, when presented with the related-high set, the participants were advantaged due to highly familiar L1 translation, and on the other hand, these L1 characters might compete against each other, resulting in more confusion. Exposed to related-low sets, however, they perhaps experienced less or even no competition because of low activation level of L1 translation. In related-high sets, therefore, it is argued that the negative effect brought about by the competition among L1 characters due to their high familiarity might be larger than the positive L1 familiarity effect, thus making its performances comparable to those of related-low sets (slight difference, but not significant).

For the two semantically unrelated sets, we did observe direct evidence in favor of positive L1 familiarity effect, as the performances in unrelated-high sets were significantly better than those in its unrelated-low counterparts. This speaks directly to the principle of learning burden (Nation, 2013), stating that learners possessing more knowledge about a certain word will likely experience lighter burden in learning that word. In the current study, learners were asked to associate L2 forms with L1, or the other way around. When L1 were quite familiar to them, the associations tended to be formed more rapidly and strongly. On the contrary, loose connections would be created when L1 were not so familiar, because learners probably struggled and spent efforts to familiarize themselves with the concept embodied in L1.

Encouragingly, our results on the immediate posttest are also backed by a self-perception questionnaire utilized to examine how the participants viewed the difficulty level of these four word sets globally. The results showed that despite individual variations, they were inclined to regard L1 familiarity as the major source of learning difficulty, with words of lower L1 familiarity being more challenging to acquire. While these perceptions were elicited at the end of the experiment (after they completed the delayed posttest), they were not prone to delayed posttest performances. We claimed that presumably the perceptions of difficulty were formed during the learning sessions, and therefore likely reflected their actual performances.

Then, one might wonder why L1 familiarity emerged as an influencing factor whereas semantic relatedness did not. We argue that these two variables function differently. L1 familiarity serves as an intra-lexical factor, which means that the impact, if found, may be associated with every individual lexical item, while semantic relatedness acts as an inter-lexical factor, meaning that this effect can only be obtained in word pairs or sets that include at least two lexical items. In our study, with enough time input to memorize the words in each set, and nearly no time lag between learning and immediate posttest session, we could expect a stronger effect, if not equal, for L1 familiarity than for semantic relatedness. It also makes sense that the effect of semantic relatedness did not appear.

We also tested the same batch of participants using the same test with different sequencing one week after the immediate posttest to investigate the retention of the vocabulary knowledge. The results of the delayed posttest are informative about how these factors worked. Contrary to the results reported on the immediate posttest, our analysis of the delayed posttest showed a robust impeding effect on the retention of the words that were packed into a semantic set, whether their L1 counterparts were familiar or unfamiliar to them. Theoretically, these findings offer further evidence supporting interference theory (Baddeley, 1997) and distinctiveness hypothesis (Hunt & Mitchell, 1982), both of which postulate that shared word knowledge places heavier cognitive burdens on learners. These also overlap some of the previous research (Erten & Tekin, 2008; Papathanasiou, 2009; Wilcox & Medina, 2013; Zhang & Sheng, 2009), though most used different assessment methods to measure the retention. Our results, however, differed from Hashemi and Gowdasiaei (2005) who found a facilitative effect of presenting words in a lexical set one week after the learning phase (they did not administer an immediate posttest). This discrepancy is open to multiple interpretations. Firstly, employing a between-subject design, they used the same words for two learning conditions, and these words were placed under different conditions (related/unrelated), so there was no need to control for lexical variables. Though efforts, such as proficiency level test and vocabulary pre-test, were made to ensure the homogeneity of the target population, other individual learner factors (e.g., learning ability) might be at play. Secondly, in their study, the target words were embedded in meaningful sentential contexts allowing the students to understand both the meaning and the use, while we presented the words in isolation. This difference is crucial because context has been shown to result in better learning (Ellis, 2013; Nassaji, 2003). Therefore, it remains unclear which treatment of the study or the supporting context contributed to learning. Finally, their study used a Vocabulary Knowledge Scale to detect learning gains embracing several vocabulary knowledge components (i.e., form, meaning and use). Our study only tapped into the meaning component. Our results also run counter to Nakata and Suzuki's (2019) study which did not reveal a significant effect, even though they strictly controlled for all the possible lexical factors. One plausible explanation concerns the delivery of the materials. In their study, learners performed all sessions on screen which resembled a lab setting and studied the words trial-by-trial with each word appearing four times only. Participants in our study were engaged in the learning sessions naturally with a traditional pen-and-pencil format in whatever way suitable for them. For example, they could write anything (e.g., copy the words) on the word cards during learning sessions, and with learning period at their disposal, the encounter frequency for each single word was not strictly controlled for. This can possibly explain why the retention rate was much lower in Nakata and Suzuki's (2019) study (related = 21.95%, unrelated = 22.54%) than in our study (related = 46.35%, unrelated = 63.40%). Future studies may obtain fruitful results if the number of encounters increases, and a more natural design is employed.

Analyzing the impact of L1 familiarity on retention showed that L1 familiarity played a facilitative role. Words containing higher L1 familiarity information were retained better. Different from the performances on the immediate posttest in our study, we found this positive effect held irrespective of whether they were learned in a semantic set or semantically related set. These findings again lend credence to the theory of learning burden (Nation, 2013). As depicted in preceding text, words with

higher L1 familiarity might secure stronger and more stable connections in lexical networks. Concerning semantically related sets, our finding is consistent with the first experiment of Tagashira et al. (2010) who conducted the test phase one week after participants learned three types of L2 antonym pairs with similar or varying degree of L1 familiarity at their own pace (with no time restriction). Different from their study, we used noun word sets. In this sense, our study also serves as an extension to theirs. However, our result contradicts their second experiment which did not obtain an obvious L1 familiarity effect when a time limit was imposed on the learning phase and the test phase was implemented three weeks after the learning phase. The differential results may be ascribed to cognitive load imposed by both varied concepts and delay of testing. For one thing, within a limited amount of time, pair learning in their second experiment might be more demanding than learning in word set (group learning). They used 15 pairs of antonyms. This means their participants should be involved in more different categories of L1 concepts than our participants did (either vegetable or musical instrument in related condition). For another, delaying the timing of the test made the participants more prone to forget. These two factors may largely explain the floor effect in their results.

Then for the semantically related sets in our study, why was the L1 familiarity effect detectable in the delayed posttest only whereas the immediate posttest did not show it? We interpret such an effect possibly as the result of the positive effect caused by higher L1 familiarity being greater than the interference effect caused by similar L1 familiarity levels, if any. Future empirical investigation is clearly warranted to elucidate this possibility with a more elaborate research design.

One thing that should be pointed out is that the experiment reported in the current study, like many others of this kind, is behavioral in nature, and the obtained results are likely a composite of various cognitive processes. Therefore, though we found a strong negative effect of semantic clustering in delayed posttest overall, it is premature to exclude the possibility of a smaller facilitating effect induced by semantic relations unless more sophisticated methods are adopted, such as event-related brain potentials. Theoretically, future research can move towards this direction in the hope of shedding more light on the intricate interaction between semantic relatedness and L1 familiarity.

Conclusion, Implications and Limitations

The purpose of the present study was to investigate how semantic relatedness and L1 familiarity impacted L2 word learning. The results revealed that semantic relatedness impeded L2 lexical performances on the delayed posttest only, whether the words were of high L1 familiarity or not. L1 familiarity, on the other hand, played a facilitative role in learning both semantically related sets and unrelated sets on both posttests. In addition, similar L1 familiarity levels might induce an interference effect, though this clearly needs to be further clarified. In summary, we add to the existing body of research supporting interference theory (Baddeley, 1997), distinctiveness theory (Hunt & Mitchell, 1982) and learning burden principle (Nation, 2013). Our study manifests the complexity of these two input-related lexical variables.

Since our study was implemented in a natural classroom setting with relatively high ecological validity, it has direct pedagogical implications which can be productively applied to vocabulary teaching/learning and material developing. First, teachers should be aware of the hazardous effect semantic clustering has on L2 word learning, and thus when introducing new words with semantic relations, it will be beneficial to strategically set them apart as much as possible. Compilers of word lists should also be encouraged to spread them out to avoid confusion. Second, as L1 information was confirmed as an important contributing factor at the early stage of learning, it is beneficial to use it as a stepping stone. In traditional classroom settings, encountered with lexical items with unknown

L1 concepts, learners (especially those with limited L2 proficiency) are strongly advised to familiarize themselves with the L1 referents first, which later serve as a hook to hang L2 forms on (Fraser, 1999).

Despite the usefulness of these results, a few limitations should be acknowledged, and thus the results should be interpreted with caution. The first limitation pertains to the scope of vocabulary knowledge. Our study only assessed form-meaning mapping. Whether the findings of this study can be generalized to other aspects of word knowledge, such as collocation, remains unknown. The second limitation concerns test timing. It is interesting to explore if our results are still applicable when an additional delayed posttest is used (e.g., three weeks after learning), because research in cognitive psychology has found that conditions inducing more cognitive loads (e.g., interference) improve learning over time (Pulido & Dussias, 2020), which is termed *Desirable Difficulties* (Bjork & Kroll, 2015). Following this logic, reversed patterns are expectable if longer-term retention is examined. Future studies also benefit from the inclusion of learner-related variables, such as pre-existing vocabulary size, for research has shown that students with denser vocabulary networks gained more new words (Kasahara & Yanagisawa, 2021; Webb & Chang, 2015).

Acknowledgements

The authors would like to extend sincere gratitude to the students for their voluntary and active participation, who in return were informed of the final results. This study was sponsored by a research project from Quzhou University (Number: JG201915).

References

- Aitchison, J. (1987). *Words in the mind: An introduction to the mental lexicon*. Oxford: Basil Blackwell.
- Anderson, J.R. (1995). *Cognitive psychology and its implications (4th ed)*. New York: W.H. Freeman.
- Baddeley, A. D. (1997). *Human memory: Theory and practice (rev. ed.)*. East Sussex, UK: Psychology Press.
- Bjork, R., & Kroll, J. (2015). Desirable difficulties in vocabulary learning. *The American Journal of Psychology*, 128(2), 241–252. <https://doi.org/10.5406/amerjpsyc.128.2.0241>
- Chapman, F. L., & Gilbert, L. C. (1937). A study of the influence of familiarity with English words upon the learning of their foreign language equivalents. *Journal of Educational Psychology*, 28(8), 621–628. <https://doi.org/10.1037/h0063528>
- Collins, A.M., & Loftus, E.F. (1975). A spreading activation theory of semantic processing. *Psychological Review*, 82(6), 407–428. <https://doi.org/10.1037/0033-295X.82.6.407>
- DeKeyser, R. (2016). Of moving targets and chameleons: Why the concept of difficulty is so hard to pin down. *Studies in Second Language Acquisition*, 38(2), 353–363. <https://doi.org/10.1017/s0272263116000024>
- Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, 43(4), 559–617. <https://doi.org/10.1111/j.1467-1770.1993.tb00627.x>
- Ellis, R. (1995). Modified oral input and the acquisition of word meanings. *Applied Linguistics*, 16(4), 409–441. <https://doi.org/10.1093/applin/16.4.409>
- Ellis, N. C. (2013). Second language acquisition. In G. Trousdale & T. Hoffmann (Eds.), *Oxford handbook of construction grammar* (pp. 365–378). Oxford: Oxford University Press.

- Erten, I. H., & Tekin, M. (2008). Effects on vocabulary acquisition of presenting new words in semantic sets versus semantically unrelated sets. *System*, 36(3), 407–422. <https://doi.org/10.1016/j.system.2008.02.005>
- Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics*, 24(3), 369–383. <https://doi.org/10.1017/s0142716403000195>
- Fraser, C.A., (1999). Lexical processing strategy use and vocabulary learning through reading. *Studies in Second Language Acquisition*, 21(2), 225–241. <https://doi.org/10.1017/s0272263199002041>
- Hashemi, R. H., & Gowdasiaei, F. (2005). An attribute-treatment interaction study: Lexical-set versus semantically unrelated vocabulary instruction. *RELC Journal*, 36(3), 341–361. <https://doi.org/10.1177/0033688205060054>
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303–317. <https://doi.org/10.1017/s0272263199002089>
- Higa, M. (1963). Interference effects of intralist word relationships in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 2(2), 170–175. [https://doi.org/10.1016/S0022-5371\(63\)80082-1](https://doi.org/10.1016/S0022-5371(63)80082-1)
- Hong, W., & Feng, C. (2014). Facilitatory or inhibitory? The effect of semantic relatedness on L2 collocation learning. *Modern Foreign Languages*, 37(1), 96–106. (in Chinese)
- Hunt, R. R., & Mitchell, D. B. (1982). Independent effects of semantic and nonsemantic distinctiveness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(8), 81–87. <https://doi.org/10.1037/0278-7393.8.1.81>
- Hoshino, Y. (2010). The categorical facilitation effects on L2 vocabulary learning in a classroom setting. *RELC Journal*, 41(3), 301–312. <https://doi.org/10.1177/0033688210380558>
- Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition*, 38(2), 163–175. <https://doi.org/10.1017/S0272263116000176>
- Ishii, T. (2015). Semantic connection or visual connection: Investigating the true source of confusion. *Language Teaching Research*, 19(6), 712–722. <https://doi.org/10.1177/1362168814559799>
- Jiang, M.Y., Jong, M. S., Tse, C., & Chai, C. (2020). Examining the Effect of Semantic Relatedness on the Acquisition of English Collocations. *Journal of Psycholinguistic Research*, 49(2), 199–222. <https://doi.org/10.1007/s10936-019-09680-9>
- Jiang, N. (2002). Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, 24(4), 617–637. <https://doi.org/10.1017/s0272263102004047>
- Jin, Z., & Webb, S. (2020). Incidental vocabulary learning through listening to teacher talk. *The Modern Language Journal*, 104(3), 550–566. <https://doi.org/10.1111/modl.12661>
- Kasahara, K. & Yanagisawa, A. (2021). Learning new verbs with known cue words: The relative effects of noun and adverb cues. *Language Teaching Research*. <https://doi.org/10.1177/13621688211001613>
- Laufer, B., & Aviad-Levitzky, T. (2017). What type of vocabulary knowledge predicts reading comprehension: Word meaning recall or word meaning recognition? *The Modern Language Journal*, 101(4), 729–741. <https://doi.org/10.1111/modl.12431>
- Lehrer, A. (1974). *Semantic fields and lexical structure*. Amsterdam: North Holland.
- Lotto, L., & de Groot, A.M.B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, 48(1), 31–69. <https://doi.org/10.1111/1467-9922.00032>
- Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word-length effect and disyllabic words. *The Quarterly Journal of Experimental Psychology*, 53(1), 1–22. <https://doi.org/10.1080/713755877>

- Macaro, E., & Lee, J. H. (2013). Teacher language background, codeswitching, and English-only instruction: Does age make a difference to learners' attitudes? *TESOL Quarterly*, 47(4), 717–742. <https://doi.org/10.1002/tesq.74>
- Nakata, T., & Suzuki, Y. (2019). Effects of massing and spacing on the learning of semantically related and unrelated words. *Studies in Second Language Acquisition*, 41(2), 287–311. <https://doi.org/10.1017/s0272263118000219>
- Nassaji, H. (2003). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *The Modern Language Journal*, 87(2), 261–276. <https://doi.org/10.1111/1540-4781.00189>
- Nation, I.S.P. (2000). Learning vocabulary in lexical sets: Dangers and guidelines. *TESOL Journal*, 9(2), 6–10. <https://doi.org/10.1002/j.1949-3533.2000.tb00239.x>
- Nation, I.S.P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I.S.P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.
- New, B. , Ferrand, L. , Pallier, C. , & Brysbaert, M. (2006). Re-examining word length effects in visual word recognition: New evidence from the English lexicon project. *Psychonomic Bulletin & Review*, 13, 45–52. <https://doi.org/10.3758/bf03193811>
- Oxford Advanced Learner's Dictionary (10th ed.)*. (2020). Oxford: Oxford University Press.
- Papathanasiou, E. (2009). An investigation of two ways of presenting vocabulary. *ELT Journal*, 63(4), 313–322. <https://doi.org/10.1093/elt/ccp014>
- Perea, M. & Rosa, E. (2002). The effects of associative and semantic priming in the lexical decision task. *Psychological Research*, 66(3), 180–194. <https://doi.org/10.1007/s00426-002-0086-5>
- Prince, P. (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *The Modern Language Journal*, 80(4), 478–493. <https://doi.org/10.1111/j.1540-4781.1996.tb05468.x>
- Pulido, M., & Dussias, P. (2020). Desirable difficulties while learning collocations in a second language: Conditions that induce L1 interference improve learning. *Bilingualism: Language and Cognition*, 23(3), 652–667. <https://doi.org/10.1017/s1366728919000622>
- R Development Core Team. (2016). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org>
- Schmitt, N. (2008). Review article: instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363. <https://doi.org/10.1177/1362168808089921>
- Suzuki, Y. , Nakata, T. , & Dekeyser, R. (2019). The desirable difficulty framework as a theoretical foundation for optimizing and researching second language practice. *The Modern Language Journal*, 103(3), 713–720. <https://doi.org/10.1111/modl.12585>
- Tagashira, K., Kida, S., & Hoshino, Y. (2010). Hot or gelid? The influence of L1 translation familiarity on the interference effects in foreign language vocabulary learning. *System*, 38(3), 412–421. <https://doi.org/10.1016/j.system.2010.03.015>
- Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System*, 21(3), 371–380. [https://doi.org/10.1016/0346-251x\(93\)90027-e](https://doi.org/10.1016/0346-251x(93)90027-e)
- Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of a second language vocabulary. *Second Language Research*, 13(2), 138–163. <https://doi.org/10.1191/026765897672376469>
- Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System*, 25(2), 261–274. [https://doi.org/10.1016/s0346-251x\(97\)00013-4](https://doi.org/10.1016/s0346-251x(97)00013-4)
- Webb, S., & Chang, A.C.S. (2015). How does prior 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>

- Wesche, M. B., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13–40. <https://doi.org/10.3138/cmlr.53.1.13>
- Wilcox, A., & Medina, A. (2013). Effects of semantic and phonological clustering on L2 vocabulary acquisition among novice learners. *System*, 41(4), 1056–1069. <https://doi.org/10.1016/j.system.2013.10.012>
- Zhang, L., & Sheng, Y. (2009). A study of the influence of semantically-relatedness of word-sets on L2 vocabulary acquisition. *Foreign Language Learning Theory and Practice*, 29(4), 18–23. (in Chinese)

Author biodata

Hanzhong Sun is currently a lecturer teaching intensive English reading and linguistics courses at Quzhou University in China. One of his research areas lies in instructed second language (L2) acquisition. Along this line, he is particularly interested in L2 vocabulary acquisition, written corrective feedback and teacher talk. He also takes a profound interest in L2 lexical processing and various lexical features associated with this process.

Shaohua Fang is currently a PhD student in linguistics at the University of Pittsburgh. His research interest lies in second language acquisition from the perspectives of theoretical linguistics and psycholinguistics. He is particularly interested in examining the differences between native and non-native speakers in the processing mechanisms involved.

-
1. Word knowledge consists of multiple components. In this study, we were solely concerned with the factors affecting form-meaning link. In the remainder of this paper, we used word learning interchangeably with form-meaning connection.
 2. CET-4 is a large-scale nationwide English proficiency test with well-established reliability and validity taken by college students in China, which measures their writing, listening, reading and translation skills as well as their vocabulary and grammar knowledge in English.
 3. Ideally, all the words grouped into low L1 familiarity sets should be rated no more than “3” on a seven-point scale by each rater. By the same token, all of the high L1 familiarity words should receive a score well above “4”. However, considering the potential differences in their life experiences, this requirement was notoriously difficult to meet. We used the mean rating scores instead, and meanwhile ensured that at least 90% of the raters met this requirement for each word.
 4. The identity number was uniquely assigned to each student when he/she was admitted to the university. They were required to provide this information in all the tests and questionnaires. It was used for the matching purpose only, because each participant completed ten tests/questionnaires successively. We did not use this information to evaluate their performances for personal reasons.
 5. We ran a small pilot study, asking three students who had similar backgrounds but did not attend the study to memorize one set of words. A subsequent interview with them revealed that ten minutes was a proper time span.

Appendixes

Appendix A. Lexical properties of target words used in the study

Target words	L1 translation	Number of letters	Number of syllables	L1 translation familiarity rating
Related-High				
celery	芹菜	6	3	5.92
turnip	萝卜	6	2	6.44
laver	紫菜	5	2	6.14
lettuce	生菜	7	2	5.99
chilli	辣椒	6	2	5.40
yam	山药	3	1	6.38
ginger	生姜	6	2	5.86
leek	韭菜	4	1	5.19
lentil	扁豆	6	2	6.21
spinach	菠菜	7	2	5.81
Unrelated-High				
blade	刀片	5	1	5.37
poultry	家禽	7	2	5.33
kidney	肾脏	6	2	6.55
aisle	走廊	5	1	5.63
racket	球拍	6	2	6.15
cob	玉米棒子	3	1	6.00
antenna	天线	7	3	5.68
meteor	流星	6	2	6.06
gear	齿轮	4	1	5.27
durian	榴莲	6	2	5.96
Related-Low				
maracas	响葫芦	7	3	2.78
dulcimer	大扬琴	8	3	2.24
zither	扁琴	6	2	2.06
snare	小军鼓	5	1	3.00
cymbal	铜钹	6	2	2.44
clarinet	黑管	8	3	2.26
tuba	低音号	4	2	2.60
fife	横笛	4	1	3.05
valve	电子管	5	1	2.94
bongos	小队鼓	6	2	2.96
Unrelated-Low				
juniper	杜松	7	3	2.03
polo	马球	4	2	2.23
plaid	格子呢	5	1	2.99
cape	岬角	4	1	2.08
calcite	方解石	7	2	2.22
avocado	鳄梨	7	4	2.96
garter	吊袜带	6	2	2.65
cloister	修道院	8	2	3.09
claves	响棒	6	1	2.29
zucchini	小胡瓜	8	3	3.18

Appendix B. Meaning Recall Test

Instruction: Please circle the words you can recognize, and then write down the Chinese translation, or English explanation/synonym.

Student Identity Number: _____

butter	email	fife	durian
juniper	yam	carrot	polo
chilli	dulcimer	antenna	cymbal
avocado	plaid	garter	nature
wealth	gear	soldier	cave
maracas	stick	clarinet	artist
golf	bongos	zither	spinach
lentil	fence	snare	ginger
turnip	poultry	musician	heat
leaf	accident	meteor	valve
tuba	laver	scene	police
hobby	cinema	lettuce	mistake
claves	blade	cape	celery
wind	branch	offer	wire
achieve	racket	cob	aisle
tiger	wood	hat	university
calcite	cabbage	cloister	pond
kidney	post	world	edge
result	leek	zucchini	noise
dust	drawer	potato	problem

Appendix C. Sample Posttest

Instruction: Please choose the Chinese translation that best matches each word.

Student Identity Number: _____

芹菜、**南瓜**、紫菜、**丝瓜**、生姜、扁豆、菠菜、萝卜、**茄子**、辣椒、韭菜、生菜、山药、**西兰花**、**大葱**

lettuce _____ turnip _____ leek _____ spinach _____ yam _____
 chilli _____ ginger _____ celery _____ laver _____ lentil _____

Note: This test was used as the immediate post-test of the Related-High set. The five Chinese characters in bold are added as the distractors, which belong to the same vegetable category. They did not appear in bold form in the original test.