CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Exploring multiple constraints on second language development of English polysemous phrasal verbs

Xiaopeng Zhang* and Ju Wen

North Minzu University
*Corresponding author. Email: zhxpnwnu@126.com

Abstract

The present study examined Chinese speakers' knowledge of English polysemous phrasal verbs (PVs) and factors that may constrain the development of PVs. The intermediate and advanced learners judged the acceptability of 100 senses of 50 PVs. Results indicate that both the intermediate and the advanced learners tended to favor the high-frequency senses (51.2%∼67.2%) of PVs but disfavor the low-frequency senses (32.9%∼46.3%) of PVs. PV frequency, semantic transparency, and time spent reading books and watching films/TV could predict the advanced learners' mastery of the high-frequency senses, while PV frequency and preemption could predict their mastery of the low-frequency senses. Semantic transparency, PV frequency, and preemption could predict the intermediate learners' knowledge of the high-frequency senses, while semantic transparency, frequency of high-frequency senses, and preemption could predict their acceptance of the low-frequency senses. No reliable relationship was detected among the learners' PV knowledge, entrenchment, time spent in second language immersion, listening to music, and communicating with others for the two groups.

Phrasal verbs (PVs), one kind of particular multiword units, are ubiquitous in English. English PVs are used approximately 2000 times per million words (Biber, Johansson, Leech, Conrad, & Finegan, 1999), and learners encounter one PV in every 150 words on average in daily communication (Gardner & Davies, 2007). Serving as ready-made chunks, PVs can help people use language fluently and naturally. As noted by Sinclair's (1991) idiom principle, people employ semi-preconstructed phrases that are conducive to fluency in speaking or writing. Failure to use PVs in the intended context makes language use sound unnatural and nonidiomatic (Garnier & Schmitt, 2015; Siyanova & Schmitt, 2007). PVs are therefore important in language learning. However, they are daunting for second language (L2) learners. It has been argued that unlike children, adults L2 learners

who already know words have different knowledge of the unit of language than children do and tend to pay attention to individual words; therefore, they are less likely to undersegment an L2 into chunks (Arnon & Christiansen, 2017). The fact that adults are easy to segment linguistic input into individual words is associated with literacy, which has been detected to impact their awareness of words and decrease the possibility to divide input into chunks (Kurvers & Uri, 2006). The misinterpretation, misuse, and avoidance of PVs by L2 learners are extensively documented and could be attributed to multiple constraints, such as cross-linguistic difference (Dagut & Laufer, 1985; Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993), syntactic complexity (Garnier & Schmitt, 2015; Liao & Fukuya, 2004; Siyanova & Schmitt, 2007), semantic transparency (Dagut & Laufer, 1985; Garnier & Schmitt, 2015; Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993; Liao & Fukuya, 2004), frequency of use (Chen, 2013; Garnier & Schmitt, 2015, 2016; Schmitt & Redwood, 2011), and time spent in social networking (Garnier & Schmitt, 2016). Nevertheless, none of these factors could explain the reduced L2 competence of PVs in its entirety. Moreover, it is not yet clear how these factors function at different stages of PV development. This study was thus motivated to further examine how these factors together with two unexplored potential constraints (entrenchment and preemption) influence the Chinese speakers' acquisition of English PVs.

L2 Development of English PVs

Albeit differently defined from the syntactic and semantic perspectives (Biber et al., 1999; Gardner & Davies, 2007; Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993; Siyanova & Schmitt, 2007), PVs contain two or more lexical words. In most cases, a PV has a verb that is intermediately followed by either an adverbial particle (e.g., *put off*) or a verb and particle intervened with one or more words (e.g., *come up with*). Research evidence suggests that PVs are challenging for L2 speakers. For instance, by comparing the use of PVs between the learner corpora and native corpora, Waibel (2007) found that the frequency of L2 learners' use of PVs in many subcorpora of the International Corpus of Learner English is significantly lower than that of the Louvain Corpus of Native English Essays. Some PVs were overused (Chen, 2013), while a lot of PVs were underused or avoided (Dagut & Laufer, 1985) in learners' language. The difficulty in learning PVs has been attributed to orthographic properties, frequency of PVs, semantic transparency, L1-L2 difference, influence from equivalent one-word verbs, learners' exposure to L2, and so on. The roles of these factors are reviewed in the following paragraphs.

Influence from orthographic properties

English PVs consist of two or more orthographic words, making it difficult for L2 learners, especially those whose L1 does not possess PVs, to consider them as monolithic units (Siyanova & Schmitt, 2007). Psycholinguistic evidence indicates that when processing an L2, learners tend to focus on individual words instead of PVs as a whole (Foster, 2001). In the learning process, unless learners are informed that the PVs they encounter are holistic semantic units, they are likely to treat them

as separate words. The separation of the components has two possible consequences, the first being that learners prefer to use individual words to substitute for PVs in communication (Siyanova & Schmitt, 2007), and the second being that they often misunderstand the meaning of PVs that are less decomposable from individual words.

Influence from semantic transparency

PVs vary in terms of semantic transparency. Some are easy to work out from the individual words, conveying literal meaning, while others are not, expressing figurative meaning (Siyanova & Schmitt, 2007). For instance, the meaning of *take in* in *They will certainly need to take in plenty of liquid* (meaning that they allow liquid to enter their body through drinking) is more transparent than its meaning in *She was taken in by his charm* (meaning that she was deceived by his charm). The figurative meaning spells more difficulty for L2 learners due to semantic incompatibility between PVs and their individual words. By using a multiple-choice test, a verb translation test, and a verb-memorizing test, Dagut and Laufer (1985) examined Israeli learners' use of English PVs, and found that most Israeli learners tended to avoid using PVs, *inter alia*, figurative PVs. Hulstijn and Marchena (1989) adopted Dagut and Laufer's (1985) elicitation task, finding that Dutch learners were prone to avoid figurative PVs that they perceived as Dutch-like. These PVs are semantically opaque and Dutch learners may consider them as "language specific and not transferable to L2" (Liao & Fukuya, 2004, p. 200). Such an effect of semantic complexity was also detected by Liao and Fukuya (2004), who investigated the Chinese learners' avoidance of English PVs. Results demonstrate that semantic transparency and test type affected the learners' avoidance of PVs. However, Garnier and Schmitt (2016) did not observe the effect of semantic complexity in testing the Chilean learners' development of English polysemous PVs. Such a null effect in Garnier and Schmitt (2016) may be due to two factors. The first has to do with the way semantic complexity was identified. They distinguished literal and figurative meaning senses purely based on two English native speakers' intuition, which might be problematic, because native speakers' judgments on the figurative or literal sense were sometimes divergent from L2 learners'. The second involves the fact that they did not treat figurative meaning senses as a graded category. Figurative meaning as a construct consists of instances differing in the degrees of semantic transparency. For instance, *take out* in "*I had to take out a loan to cover all my expenses* is more opaque in meaning than in *You should take her out to this new Chinese restaurant*," although *take out* in these two sentences expresses the figurative meaning according to the conceptual approach (Dirven, 2001; Kurtyka, 2001). Such varying degrees of semantic complexity may cause different difficulties in L2 learning. However, this was not measured by Garnier and Schmitt (2016).

Influence from frequency of PVs

Usage-based approaches assume that the more frequent a PV, the more likely it is to be acquired (Ellis, 2002; Nation, 2001; Schmitt, 2014). However, to date, only a few studies have been done to examine the relationship between frequency and

knowledge of PVs. Corpora were often used to estimate the frequency of PVs that learners may encounter, although the frequency counts in native language (L1) corpora sometimes do not fit the real amount of L2 learners' exposure to the target construction (Schmitt, 2014; Zhang & Mai, 2018). Schmitt and Redwood (2011) tested the role of PV frequency obtained from the British National Corpus (BNC) complete, BNC written, and BNC spoken, as well as the Corpus of Contemporary American English (COCA) in L2 learners' reception and production of English PVs. Results suggested significant positive correlations between the test scores and PV frequencies. The authors concluded that corpus frequency could reliably predict the participants' receptive and productive knowledge of PVs. Likewise, Chen (2013) examined the relationship between corpus frequency of the 50 most frequent PVs in both the BNC and the COCA and Chinese speakers' knowledge of English PVs, finding a positive correlation between PV frequency and the learners' production of PVs, with r^2 variance being 17.1% and 11.9% based on COCA and BNC frequencies. This was verified by Garnier and Schmitt (2016), who found that frequency from the COCA could significantly predict Chilean learners' knowledge of high-frequency English polysemous PVs, confirming the basic tenet of usage-based approaches.

Influence from L1-L2 difference

L1-L2 difference is another key factor that may be responsible for L2 learners' avoidance of English PVs. Dagut and Laufer (1985) attributed avoidance of producing PVs by Israeli learners to the structural difference between Hebrew and English. Laufer and Eliasson (1993) tested Swedish English-L2 learners' knowledge of English PVs. Unlike Hebrew, Swedish has PVs. After comparing the results of their study with those of Dagut and Laufer (1985), they found that Hebrew learners rather than Swedish learners avoided PVs. Laufer and Eliasson (1993) assumed that it is the L1-L2 difference that could be the best predictor of avoidance. However, Liao and Fukuya (2004) argued that the avoidance of PVs was the result of interlanguage development instead of the L1–L2 difference, as they found that some Chinese intermediate English-L2 learners produced English PVs less frequently than advanced learners and English natives, while advanced learners did not manifest a clear pattern of avoidance behavior. Similar findings were obtained by Hulstijn and Marchena (1989), who investigated the advanced and intermediate Dutch English-L2 learners' avoidance behavior, and observed that both the advanced and the intermediate learners did not avoid PVs categorically when doing the multiple-choice, memorization, and translation tests.

Influence from other verbs

L2 learners' difficulty of PV use is in some cases caused by the equivalent one-word verb. In English, most PVs are synonymous with one-word verbs, although one-word verbs are always used in more formal register while PVs often occur in somewhat colloquial register (Biber et al., 1999; Freeborn, 1995). Results of Dagut and Laufer's (1985) investigation manifest that most of the Hebrew learners preferred one-word verbs to PVs. This observation has also been found in Liao and Fukuya (2004), where Chinese intermediate Enlgish-L2 learners tended to use more one-word verb than PVs in the translation and recall tasks, respectively. Siyanova

and Schmitt (2007) explored the likelihood of using PVs and one-word verbs by native speakers and English learners whose L1s were Arabic, Russian, Italian, and Chinese at the University of Nottingham. Results showed that L2 learners were less likely to use PVs than native speakers in informal spoken contexts.

Influence from learners' exposure to L2

Schmitt and Redwood (2011) found that corpus frequency is moderately related to language learning gains. That said, corpus cannot exactly mirror one's exposure to the target language (Schmitt 2014). Learners may have different amounts of exposure to an L2. It has been found that learners' intuition on the amount of time devoted to an L2 use may act as a good predictor in the development of PVs (Garnier & Schmitt, 2016; Schmitt & Redwood, 2011; Zhang & Mai, 2018). Specifically, Schmitt and Redwood (2011) found that time spent in extensive reading and watching English films and TV had positive effects on the acquisition of PVs. Garnier and Schmitt (2016) also observed that time L2 learners spent in reading and social networking per week was significantly related to their PV knowledge, while time spent listening to music and watching films in English was not. Siyanova and Schmitt (2007) detected that the quality of exposure to an L2 (i.e., how often they interact with English native speakers, and the degree to which L2 learners adapted to the target culture) had great effects on learners' using of PVs. L2 learners' engagement in social activities can significantly enhance the acquisition of PV knowledge.

Difficulty in learning different meanings

Another factor that makes the PV learning much more difficult is that PVs are in most cases polysemous. Gardner and Davies (2007) estimated that the 100 most frequent English PVs on average had 5.6 meaning senses. The conceptual approach (Kurtyka, 2001) holds that there is no direct link between a linguistic form and what it refers to; rather, in the interpretation of meaning, they are linked through the mediation of concepts in people's mind. With continued use, words become polysemous, namely, secondary or figurative meanings evolve from the primary or literal meaning. For instance, *board* originally refers to a flat piece of wood, and it gradually takes on a meaning of a group of people who manage a company. Likewise, the meaning of PVs becomes polysemous via continued use, with secondary meanings being derived from the literal meaning, as Garnier and Schmitt (2016) argued that it would make more sense to distinguish "literal" from "figurative" meanings (p. 31). Polysemous senses of PVs may complicate the form–meaning relationship for L2 learners. For instance, learners should map *bring up* onto different meanings like *carry something up, nurture a child, or mention a suggestion* (Biber et al. 1999). Only two studies have measured knowledge of PV meaning senses in L2 learners as of late. Schmitt and Redwood (2011) found that the intermediate L2 learners were able to recognize about 65.2% of their tested PVs, and produce about 48.2% of them, demonstrating that intermediate L2 learners had good knowledge of the selected English PVs. However, Schmitt and Redwood only examined a single PV meaning sense. Garnier and Schmitt (2016) directly looked at L2 learners' knowledge of high-frequency English polysemous PVs, and detected that on average learners knew 40%

of PV senses, but they knew around 20% of all different meaning senses of each tested PV, suggesting that they had difficulty in producing the polysemous PVs.

Interim summary

Despite extensive separate research into effects of the aforementioned factors, it remains less clear about how these factors conspire to influence L2 development of PVs. When the above potential factors are considered at the same time, which are the most significant factors that could predict the L2 learners' English PV knowledge? This issue is both theoretically and practically significant, as the primary mission for L2 learning studies is to extensively research all factors that influence the L2 learning process, helping educators take advantage of effective strategies to maximize the learning potential. Moreover, language is characterized as polysemy. The above literature review reveals that most previous research into PV development, with the exception of Garnier and Schmitt (2016), did not take account of polysemy. Studies in this line directly counted the number of PV occurrences from corpus, omitting different senses of PVs, so it is significant to explore whether the observed effects of the above-mentioned factors on L2 acquisition of PV knowledge also obtain when different senses of PVs are investigated.

The conceptual approach (Kurtyka, 2001) holds that the figurative meaning of a word including PVs is derived from the literal meaning. The derivation of figurative meanings makes the form–meaning mapping much more complex (Biber et al., 1999). From the learning perspective, different meaning senses may compete for being mapped onto the same form, which may constrain the development of PV knowledge. Cognitive studies have found that a form repeatedly used to convey a given meaning may restrict language users from using this form to convey other meanings (Brooks, Tomasello, Dodson, & Lewis, 1999; Langacker, 1987). This effect, termed as entrenchment that may have a role to play in L2 development (Zhang & Mai, 2018), has yet to be tested in the process of PV development. Although L2 learners have been found to prefer one-word verbs to PVs, the competing effect from the alternative verbs and phrases, which may result from preemption, has been detected to play differing role in distinct stages of both L1 (Ambridge, 2013; Ambridge et al., 2015; Ambridge, Pine, & Rowland, 2012; Matthews, Lieven, Theakston, & Tomasello, 2005) and L2 (Zhang, 2017; Zhang & Mai, 2018) learning. To date, effects of preemption among different meaning senses of a PV and other competing constructions on PV learning remain an uncharted territory. In what follows, entrenchment and preemption are discussed briefly.

Two potential factors: Entrenchment and preemption

Entrenchment is closely related to frequency effect. Langacker (1987) argued that "every use of a structure has a positive impact on its degree of entrenchment" (p. 57). When employed to explain language development, it predicts that the formation of a construction is predicated on repeated occurrences, and the repetition of a form in one structure reduces the possibility that this form could be used in others (Braine & Brooks, 1995; Tomasello, 2003). Such statistical inference based on the attested use is in accordance with the associative learning theory that people are

less likely to pair C with B that has been paired with A (Kording & Wolpert, 2011). Entrenchment is at play in children's avoidance of ungrammatical verb argument structures (Ambridge et al., 2012, 2015; Ambridge, Pine, Rowland, Jones, & Clark, 2009; Ambridge, Pine, Rowland, & Young, 2008; Brooks et al., 1999) and L2 learners' restriction of overgeneralization of *un*-prefixation (Zhang, 2017). It was also found to play a role in the restriction of L2 development of grammatical denominal verbs (Zhang & Mai, 2018): the Chinese college students' judgments on the acceptability of denominal verbs were negatively correlated with the frequency of nominal forms of the denominal verbs, namely, the nominal forms that are frequently used to restrict the verbal use that is less frequently used. This study is intended to explore whether the use of verbs constrain the L2 learning of different senses of PVs.

Different from entrenchment, preemption was originally theorized to inhibit the formation of derivational morphemes. Well-attested lexical words gradually block children's own coinages that convey the same meaning (Clark & Clark, 1979). For instance, *went* blocks *goed. Clark and Clark's (1979) preemption account predicts that "if a potential innovative use would be precisely synonymous with a well-established use, the innovative use is normally preempted by the well-established use, and is therefore considered ungrammatical" (p. 798). This is found to be psychologically real when children recovered from the overgeneralized errors of verb argument structures (Ambridge et al., 2008, 2009, 2012, 2015) and *a*-adjective errors (Boyd & Goldberg, 2011), and L2 learners restricted the overgeneralized errors of *un*-prefixation (Zhang, 2017). Preemption is also at play in the L2 acquisition of denominal verbs: the frequency of competing verbs was significantly negatively correlated to the advanced L2 learners' judgments on the acceptability of denominal verbs (Zhang & Mai, 2018). Based on the preemption account, we assume that the higher frequency of the alternatives, the less likely the PV is to be used, and the less acceptable is this PV.

Most research into effects of entrenchment and preemption aims to test how learners make use of these two mechanisms to avoid incorrect constructions so as to venture explanations for the learnability problem (Bowerman, 1988). Nevertheless, these two factors may restrict L2 development due to the following considerations. The entrenchment and preemption accounts predict that when a form is repeatedly used in a construction, learners may gradually build an inference that it seldom occurs in other novel constructions, and that if learners have mapped a particular form onto a given meaning, this form–meaning relationship is likely to block them from mapping other novel forms into this meaning (Braine & Brooks, 1995; Langacker, 1987). In the learning process, grammatical constructions could also be novel when learners have never met them before, so well-attested constructions may restrict the development of constructions that are novel to L2 learners. Another purpose of this study is to explore whether, and to what extent, these variables influence L2 acquisition of English PVs.

Research question

As reviewed above, L2 learning of PVs is a process replete with multiple constraints. The present study attempted to look at the development of English polysemous PVs and constraints involved in this process. By comparing effects of these possible

variables, we would be in a better position to pinpoint the most significant contributing factors involved in the PV learning process. To this end, this study is guided by two research questions. First, how good is L2 learners' knowledge of both high- and low-frequency meaning senses of the target PV? Second, what are the factors (e.g., frequency of PVs, entrenchment, preemption, semantic transparency, and exposure to L2, including years of English learning, time spent in English-speaking countries, and time spent in reading books, listening to music, watching films and TV programs, and communicating with others in English) that significantly restrict the L2 development of PV knowledge?

Method

Participants

One hundred and twenty Chinese English-L2 learners were recruited. Group A was made up of 60 third-year English major students (12 males, 48 females; $Mean_{age} = 21.03$ years, $SD = 0.45$), labeled as intermediate level. They had received more than 10 years of formal English learning. Group B was composed of 16 postgraduate students majoring in linguistics and 44 Chinese English teachers (8 males, 52 females; $Mean_{age} = 36.8$ years, $SD = 6.13$) from two universities in China, labeled as advanced level. They had received more than 14 years of formal English education. In addition, the 39 English teachers had had at least 6 months of learning experience in English-speaking countries. Based on years of English learning, we presumed that all participants had a relatively high level of proficiency. This assumption had been confirmed by their performance in doing the Y_Lex test of English vocabulary size (Meara, 2005). Considering that the administration of a proficiency test is not possible in our study for practical reasons, we made reference to previous studies that looked at language proficiency using various approximations, such as the participants' self-rating task (Siyanova-Chanturia et al., 2011) or the vocabulary size task (Gyllstad & Wolter, 2016). We used these two tasks as a proxy for measuring the participants' general proficiency, as Alderson (2005) has found that vocabulary size is highly correlated with measures of writing, listening, and reading skills in a language. The Y_Lex test version was chosen on the grounds that all participants in the present study had known all the words in the X_Lex version (0–5K). The Y_Lex test has been designed to test receptive knowledge of vocabulary between the 6K and 10K levels. All participants were also required to self-evaluate their English ability in terms of speaking, listening, reading, and writing on a 10-point scale ($1 = none$, $10 = nativelike$). A t test on the scores of vocabulary size demonstrated that the postgraduate students and teachers achieved higher scores than the third-year English major students ($t = 18.658$, $df = 118$, $p < .001$). Self-rating scores revealed the same pattern as that of the English vocabulary test. The participants' biodata are provided in Table 1.

Materials

The target construction was selected from the PHaVE List (Garnier & Schmitt, 2015). This list contains 150 high-frequency PVs, among which were 91 PVs with

Table 1. Participants' biographical information: Means (standard deviations)

Groups	Intermediate	Advanced
Age (year)	21.03 (0.54)	36.8 (6.13)
Gender (M vs. F)	12 vs. 48	8 vs. 52
Self-evaluations: Speaking	5.03 (1.28)	6.42 (1.06)
Self-evaluations: Listening	5.22 (1.14)	6.57 (0.85)
Self-evaluations: Reading	6.08 (1.12)	7.07 (0.84)
Self-evaluations: Writing	5.35 (1.05)	6.52 (0.75)
Score of Y_Lex test	2268.58 (577.25)	4025.83 (446.12)

Note: The max score of the Y_Lexvocabulary test is 5,000.

at least two meaning senses. Fifty PVs were randomly selected from the 91 polyse-mous PVs, with 55% sampling rate being applied (i.e., 55% × 91). These 50 ran-domly selected PVs covered the full range of frequencies and varied in degrees of semantic opacity. The target PVs appear at least 10 times per million words in either the COCA or the BNC (Liu, 2011), suggesting that they are high-frequency PVs. Two meaning senses of each PV were chosen. One is frequently used while the other is less frequently used based on their occurrences in the COCA. For instance, one sense of *put out* (*make information or products known or accessible to the public*) occurred 3,900 times, while the other sense (*place something somewhere in order for it to be seen or used*) occurred 830 times.[2] Each sense of the PV was presented in a sentence, serving as a prompt. These prompts were either from the PHaVE List or from the Collins COBUILD English Dictionary (See online Appendix A). For in-stance, "Police have **_put out_** a warning about thieves in the area" and "The bus is coming. Paula has **_put out_** her luggage for the bus." Given that the target PVs were polysemous, the two meaning senses for a PV together with 60 filler sentences were randomly presented in the test to avoid priming effects. As with Garnier and Schmitt (2016), in the test participants were told that some sentences contained the same PV as others so as not to cause potential confusion. This explicit explana-tion and random presentation of test sentences were expected to minimize the po-tential biased due to the two usages of the same verb. The test instructions were written in English and explained by the first author. The 60 fillers (30 ungrammati-cal vs. 30 grammatical), including the usage of prepositions, the plurality of nouns, adjective–noun phrases, and so on, may help to avoid the strategies used by parti-cipants when only grammatical PVs were included. Furthermore, we may know whether our participants were influenced by extra factors, such as fatigue, according to the participants' judgments on the acceptability of the fillers. The PVs and fillers were underlined in order to make them more noticeable.

Task for measuring semantic transparency

Semantic transparency of PV meaning senses was identified differently from Garnier and Schmitt (2016), who classified senses of a PV into a binary category,

namely, literal versus figurative, according to the authors' and another English native speaker's judgments. English native speakers' intuition about the semantic opacity may not compatible with that of L2 learners, as found in our pilot study. Figurative meanings per se, in contrast, vary on a cline of semantic opacity (Howarth, 1996; Taylor, 2004), so Garnier and Schmitt's binary classification may omit the subtle differences between figurative meanings in terms of semantic opaqueness. For instance, the figurative meaning (Sense 1: *Decide not to do or say something*, Sense 2: *Prevent somebody/something from reaching their full potential*, and Sense 3: *Contain an unwanted physical manifestation, such as tears, laughter, sigh, or sneeze*) of *hold back* in Garnier and Schmitt (2016) obviously differed in meaning transparency. Apparently, Sense 2 is more opaque than the other two senses. To remedy this problem, we operationalized semantic transparency as the degree of easiness for learners to work out the meaning of PVs from the individual words, so transparency is a continuum and "must be quantified by subjective self-assessment ratings" (Macis & Schmitt, 2017, p. 328). A transparency task was designed to see how easy or difficult it was to guess the meaning of PV according to the component of the PV. Participants would rate the meaning of each PV with its Chinese translation equivalent in a bracket in a sentence on a 5-point Likert scale, where 1 = *very easy to guess* (very transparent), 5 = *very difficult to guess* (very opaque).

Questionnaire for learners' exposure to English

Previous studies have suggested that participants' everyday exposure to English is one of the important factors in the development of lexical knowledge (Garnier & Schmitt, 2016; Schmitt, 2014). The present study used a questionnaire adapted from Garnier and Schmitt (2016) to collect the participants' biodata to account for the variation in their ratings of the PV (See online Appendix B). Participants were required to provide the time they spent in reading (i.e., English books, magazines, and newspapers, and visiting English websites); watching films, videos, or TV in English; and using English to keep in contact with people. They were also asked to provide the number of years spent studying general English, and the time spent in countries where the L2 is spoken.

Pilot test

To guarantee the well formedness of all testing items, two native speakers of English were asked to judge the acceptability of all testing items. Results showed that all testing items were acceptable. These items were then administered to 10 third-year English major students at a Chinese university. Follow-up interviews showed that there were no strange words and they did not have any confusion on the test, indicating that these testing items were ready to be used in the main study. After that, these 10 students were asked to identify the degree of easiness to work out the meaning senses of PVs. They first received a detailed explanation of PVs and transparency, and then were provided with each PV (followed by its Chinese translation equivalent in a bracket) in a sentence and asked to rate the degree of easiness to guess the meaning of each PV according to the individual words on a 5-point Likert scale. Finally, a 5-point-scale familiarity test adapted from Wang and Koda (2005) for the PVs was administered to these 10 students.[1] Results revealed

that the averaged familiarity score for each PV was above 2.5, so they were included in this study.

Administration of main test

The test administration consisted of three phases: the main test and the test about the participants' exposure to English, the Y_Lex vocabulary size test, and the semantic transparency test. For the main test, all participants were informed that they should rate the acceptability of the 160 underlined phrases based on the sentences where the phrases occurred. The instructions were as follows: "There are 160 sentences, each with some part underlined. To complete the task, please score each underlined part of the sentence from 1 to 5 according to the following criteria: 1 = *totally unacceptable*, 2 = *very poor English*, 3 = *poor English*, 4 = *almost correct English*, 5 = *completely acceptable*. Please judge the acceptability of the underlined part based on the meaning conveyed by the sentence, then select a related number standing for your judgment and mark it out." All testing items were randomly presented for each participant. After finishing the judgment task, they were required to finish the questionnaire about their exposure to English and the Y_Lex vocabulary size test. For this test, all participants were tested individually on the computer, with testing scores automatically recorded. For the semantic transparency test, half of the participants were asked to identify the degree of easiness to work out the meaning senses of PVs, with the testing procedures the same as those for the pilot test.

Operationalization of predictor variables

To test effects of potential factors involved in the development of English PVs, linear mixed-effects models were used to fit the data. The predictor variables are detailed in the following paragraphs.

Frequency of polysemous senses

The frequency for different senses of each PV was obtained from the PHaVE List (Garnier and Schmitt, 2016). Because the frequencies of difference senses for each PV were not normally distributed, they were subject to log $(N+1)$ transformation.

Entrenchment

In principle, the verb used in a PV to express a given meaning may have a negative influence on the learners' acceptance of this PV. That is, for each sense of the target PV, we predict that the higher frequency of verbs used, the less acceptable learners judge each sense of this PV. In the present study, we used the frequency of verbs to measure the effect of entrenchment. Because the frequencies of verbs were not normally distributed, they were log $(N+1)$ transformed.

Preemption

For each sense of the target PVs, 8 English native speakers were asked to suggest at least one competing verb or phrase that conveyed the same meaning as this sense

based on the sentences in the acceptability judgment task. The most commonly suggested competing verb or phrase was chosen. The lemma frequency of this competing form obtained from the COCA was used to constitute the preemption measure. Remember that the frequency of meaning senses of PVs in the PHaVE List were obtained from the COCA, so frequency of competing forms from the COCA were used. Because the lemma frequencies of competing forms were not normally distributed, they were log $(N+1)$ transformed.

Interference from high-frequency meaning sense

Studies regarding the learning of L2 polysemous verbs have suggested that the learning of low-frequency meaning senses of a verb may be constrained by the use of high-frequency meaning senses of this verb, that is, the frequency of a commonly used sense is negatively correlated to the participants' judgments of the use of the less commonly used sense of this verb (Zhang, 2010; Zhao, 2010). In line with this observation, interference from the high-frequency meaning sense of a PV on the participants' judgment of use of the low-frequency meaning sense of this PV was included as a control variable.

Semantic transparency

As discussed previously, senses of a PV vary on a cline of semantic transparency (Howarth 1996; Siyanova & Schmitt, 2007; Taylor, 2004). Semantic transparency in the present study was operationalized as the degree of easiness that the meaning can be worked out from the individual words. Another 10 participants judged the degree of easiness that the meaning can be worked out from the individual words on a 5-point scale. The instructions for the participants to judge the semantic transparency were as follows: "Some meaning of phrasal verbs is quite easy to work out from the individual words. For instance, *bring up the box from the first floor* (meaning carry them up) is somewhat easy to infer from the individual words *bring* and *up*; *bring up children* (meaning *nurturing the children*) is difficult to infer the meaning from *bring* and *up*; while *bring up a suggestion* (meaning *making a suggestion*) is the most difficult to infer the meaning from *bring* and *up*." The ratings obtained for each sense were averaged by participants and then were log transformed due to skewed distribution.

Exposure to English

This involved a group of variables, such as years of L2 learning, time spent in English-speaking countries, time spent reading English books, listening to music with English lyrics, and using English in communication with others obtained from the questionnaire. They were included as an estimation of different individuals' exposure to English.

L2 proficiency

According to the years of English learning and their performance in the vocabulary size test, participants were divided into the intermediate and advanced groups, so L2 proficiency was a categorical variable.
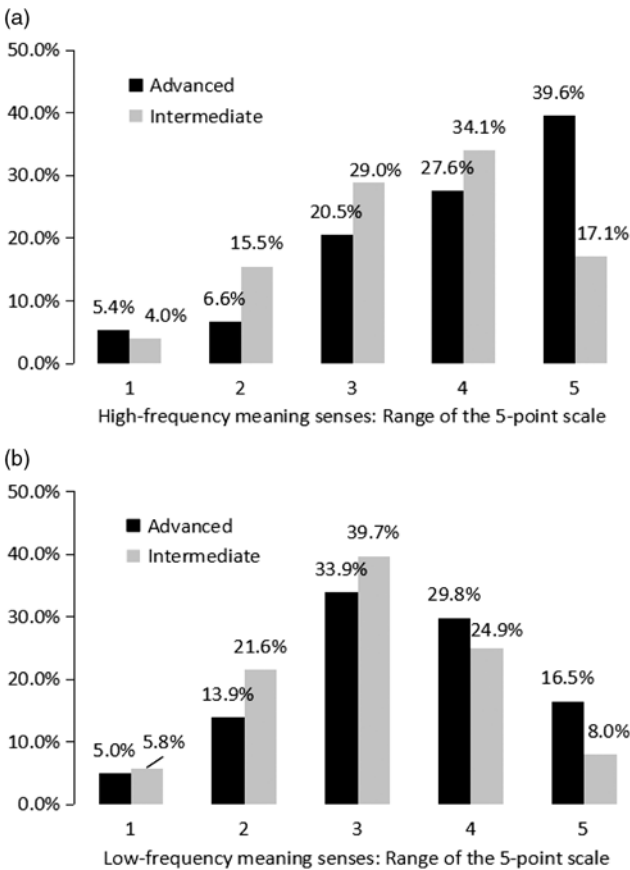
Figure 1. The participants' ratings on (a) high- and (b) low-frequency meaning senses.

Results

Learners' acceptance of different PV meaning senses

The participants' ratings on the high- and low-frequency meaning senses are presented in Figure 1. For high-frequency meaning senses, the advanced learners judged 39.6% of them as *completely acceptable*, 27.6% of them as *almost correct English*, the 27.1% of them as *poor* or *very poor English*, and 5.4% of them as *completely unacceptable*. The intermediate learners treated 17.1% of them as *completely acceptable*, 34.1% of them as *almost correct English*, the 44.5% of them as *poor* or *very poor English*, and 4.0% of them as *completely acceptable*. If we take the ratings on *completely acceptable* and *almost correct English* as acceptance of PVs, we find both the advanced and the intermediate learners tended to accept the high-frequency meaning senses, with 67.2% and 51.2% of high-frequency meaning senses judged as acceptable, respectively. For low-frequency meaning senses, the pattern seems different. The advanced learners judged 16.5% of them as *completely acceptable*, 29.8% of them as *almost correct English*, and 55.9% of them as *poor* or *very poor English*. The case was even worse for the intermediate learners. They treated 61.3% of them as *poor* or *very poor English*. Comparatively, these two groups were more prone to disfavor the low-frequency meaning senses.

To sum up, the analysis of the participants' ratings on the acceptability of the two types of meaning senses reveals that the two groups of learners had a better mastery of high-frequency meaning senses than of low-frequency meaning senses of PVs. Note that our L2 learners received formal English learning for more than 14 years on average. It is valuable to examine what factors significantly restricted the participants from learning the senses of PVs.

Factors constraining learners' acceptance of PV meaning senses

As discussed previously, factors constraining the development of different meaning senses of PVs may differ. Specifically, semantic transparency, frequency of PVs, entrenchment, preemption, years of English learning, and time spent reading books, watching films, listening to music in English, and using English to communicate with others are all potential factors influencing the learning of the high- and low-frequency meaning senses. In addition, the learning of low-frequency meaning senses of PVs may be constrained by the use of high-frequency meaning senses, so the participants' ratings on the acceptability of the two types of meaning senses were analyzed separately.

Analysis of high-frequency meaning senses

Although our participants had received formal English education for many years and the target constructions were the most frequent PVs, the two groups of participants still considered some high-frequency senses of PVs less acceptable, so it is intriguing to examine what factors may influence their judgments. The mixed-effects regression modeling was built to fit the participants' judgment data via the statistics package *lme4* in R environment (Bates, Maechler, Bolker, & Walker, 2015; R Development Core Team, 2015). The correlations among all predictors were weak or moderate ($|rs| \leq .621$). Collinearity diagnostic analyses for the PVs showed that no multicollinearity existed among all predictor variables, with the condition index value less than 30 (David, 1991). The maximal random-effects structure for which the model would converge was used (Barr, Levy, Scheepers, & Tikly, 2013). Frequency of PVs, semantic transparency, entrenchment, preemption, amount of exposure to English (years of English learning, time spent in L2-speaking countries, time spent reading books in English, listening to music in English, watching films/TV programs in English, and communicating with others in English), and L2 proficiency were treated as fixed effects, and participants and dependent variables (DVs) as random effects. L2 proficiency (intermediate vs. advanced) was sum coded (–1, 1) so as to make interactions interpretable.

We calculated *p* values by likelihood ratio tests of the full model with the fixed effect in question against the model without the fixed effect in question. The linear mixed-effects model yielded significant effects for semantic transparency, frequency of meaning senses, time spent watching films, L2 Proficiency × Preemption, L2 Proficiency × Time Spent Reading Books in English, L2 Proficiency × Time Spent Watching Films in English, and L2 Proficiency × Time Spent Communicating With Others in English, suggesting that preemption and the time spent in reading books, watching films, and communicating with others in English had different roles to play in the two groups of participants' development of the high-frequency meaning senses of PVs.

Given these significant interaction effects, the mixed-effects model was established to fit the two groups' performance in the acceptability judgment task, respectively. The maximal random-effects structure for which the model would converge was also employed, with frequency of PVs, semantic transparency, entrenchment, preemption, years of English learning, time spent in L2-speaking countries, time spent reading books in English, listening music in English, watching films/TV programs in English, and communicating with others in English as fixed effects, and participants and DVs as random effects: Model = lmer (AJTScores ~ (1|participant) + (1|PV) + SemanticTransp + PVfrequcy + Entrenchment + Preemption + YearsEnglishLearn + TimeL2-speakingCoun + ReadingBook + WatchingF + ListeningM +UsingEC. Correlations among all fixed factors were weak, as presented in Tables 2 and 3.

We calculated *p* values by likelihood ratio tests of the full model with the fixed effect in question against the model without the fixed effect in question. As shown in Table 4, the effects for semantic transparency and frequency of PVs were significant for both the intermediate and the advanced learners, demonstrating that the more opaque the meaning sense, the less acceptable this meaning sense was judged, and the more frequent the meaning sense, the more acceptable it was judged. In addition, years of English learning, and time spent reading books and watching films in English significantly predicted the advanced learners' knowledge of the high-frequency meaning senses. The alternative competing verbs were a significant factor that may influence the process in which the intermediate judged the acceptability of high-frequency meaning senses. That is, the higher the frequency of the alternative competing verbs or phrases, the less likely the intermediate learners were to accept the corresponding use of PVs.

Analysis of low-frequency meaning senses
The mixed-effects regression modeling was used to fit the participants' judgments on the use of PVs. The procedures of data analyses were the same as those for the high-frequency meaning senses. Entrenchment, frequency of PVs, semantic transparency, preemption, interference from the high-frequency sense, amount of exposure to English, and L2 proficiency were fixed effects, and participants and DVs were random effects. Collinearity diagnostic analyses yielded no multicollinearity among all predictor variables, with the condition index value less than 30, so all factors were included in the model. Significant effects for semantic transparency, frequency of PVs, L2 Proficiency × Semantic Transparency, and L2 Proficiency × Frequency of PVs were found, suggesting that semantic transparency and frequency of PVs had different roles to play in the two groups of participants' learning the low-frequency meaning senses of PVs.

Considering these significant interaction effects, the mixed-effects model was utilized to fit the two groups' performance in the acceptability judgment task, respectively. Data analysis procedures for the low-frequency PVs were the same as those for the high-frequency PVs: Model = lmer (AJTScores ~ (1|participant) + (1|PV) + SemanticTransp + PVfrequcy + Entrenchment + Preemption + highfr_senseInterfer + YearsEnglishLearning + TimeL2-speakingCoun + ReadingBook + WatchingF + ListeningM + UsingEC. We calculated *p* values by likelihood ratio tests of the full model with the fixed effect in question against the model without

Table 2. Correlation matrix of the fixed variables for high-frequency meaning senses: Advanced learners

	1	2	3	4	5	6	7	8	9	10
1. Semantic transparency	1.000									
2. Frequency of PVs	0.175**	1.000								
3. Preemption	−0.263***	0.130**	1.000							
4. Entrenchment	−0.621***	−0.047	−0.353***	1.000						
5. Years of English learning	<0.001	<0.001	−0.001	<0.001	1.000					
6. Time spent in L2-speaking countries	0.001	<0.001	<0.001	<0.001	0.085**	1.000				
7. Time spent reading books in English	<0.001	<0.001	0.001	<0.001	0.341***	0.019*	1.000			
8. Time spent watching films in English	−0.001	<0.001	<0.001	<0.001	−0.345***	0.053**	0.125**	1.000		
9. Time spent listening to music in English	<0.001	<0.001	0.002	<0.001	−0.388***	0.002	−0.401***	−0.147**	1.000	
10. Time spent using English	−0.001	<0.001	−0.001	<0.001	−0.324***	−0.090*	−0.420***	−0.091*	0.109**	1.000

Note: PVs, phrase verbs. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 3. Correlation matrix of the fixed variables for high-frequency meaning senses: Intermediate learners

	1	2	3	4	5	6	7	8	9	10
1. Semantic transparency	1.000									
2. Frequency of PVs	0.175**	1.000								
3. Preemption	−0.263***	0.130**	1.000							
4. Entrenchment	−0.608***	−0.047*	−0.353***	1.000						
5. Years of English learning	<0.001	<0.001	<0.001	−0.284	1.000					
6. Time spent in L2-speaking countries	<0.001	<0.001	<0.001	<0.001	0.088**	1.000				
7. Time spent reading books in English	<0.001	<0.001	<0.001	<0.001	0.300***	0.009	1.000			
8. Time spent watching films in English	<0.001	<0.001	<0.001	<0.001	−0.366***	0.044*	0.113**	1.000		
9. Time spent listening to music in English	<0.001	<0.001	<0.001	<0.001	−0.222***	0.011	−0.224***	−0.122**	1.000	
10. Time spent using English	<0.001	<0.001	<0.001	<0.001	−0.339***	−0.017**	−0.136***	−0.033*	0.022*	1.000

Note: PVs, phrase verbs. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 4. Selected mixed model of fixed effects on the participants' acceptability of high-frequency senses

Fixed effects	Advanced			Intermediate		
	β	SE	χ^2	β	SE	χ^2
(Intercept)	0.714	1.461		1.860	1.112	
Semantic transparency	**−0.199**	**0.090**	**5.063***	**−0.157**	**0.069**	**4.762***
Frequency of PV	**0.328**	**0.147**	**6.880****	**0.455**	**0.112**	**10.615****
Preemption	−0.107	0.088	3.002	**−0.141**	**0.066**	**4.211†**
Entrenchment	0.133	0.125	1.110	0.004	0.103	0.002
Years of English learning	**0.023**	**0.009**	**6.955****	0.022	0.037	0.349
Time spent in L2-speaking countries	0.011	0.007	2.335	−0.205	0.233	0.701
Time spent reading books in English	**0.020**	**0.008**	**4.699***	−0.019	0.014	1.553
Time spent watching films in English	**0.077**	**0.029**	**7.003***	0.002	0.009	0.039
Time spent listening to music in English	0.002	0.014	0.022	0.003	0.011	0.040
Time spent using English	0.039	0.020	2.488	0.055	0.041	2.455
	$R^2_m = .160$			$R^2_m = .201$		

Note: SE, standard error. R^2_m refers to variance explained by fixed factors. $^\dagger p$ = marginally significant. *$p < .05$. **$p < .01$.

the fixed effect in question. Correlations among all fixed factors were less than moderate, as shown in Tables 5 and 6.

As shown in Table 7, the effects of semantic transparency, interference from high-frequency sense, and preemption were significant for the intermediate learners, demonstrating that the more opaque in semantics, the less likely the intermediate learners were to accept the PVs, that the more frequently the PV expressed one meaning, the less likely the intermediate learners were to accept this PV than express another meaning, and that the more frequent the competing verbs, the less likely the PVs were judged acceptable. Effects of PV frequency and preemption were significant for the advanced learners, indicating that they were more likely to accept the high-frequency meaning senses than the low-frequent ones in doing the acceptability judgment task, and that the more frequent the competing verbs, the less acceptable the low-frequency meaning senses of PVs were judged. However, entrenchment, years of English learning, time spent in English-speaking countries, reading books in English, watching English films, listening to music in English, and using English to communicate with others were not at play in the process where all learners acquired knowledge of low-frequency meaning senses.

To recap, factors constraining the advanced learners' learning of high-frequency meaning senses were semantic transparency, frequency of PVs, years of English

Table 5. Correlation matrix of the fixed variables for low-frequency meaning senses: Advanced learners

	1	2	3	4	5	6	7	8	9	10	11
1. Semantic transparency	1.000										
2. Frequency of PVs	0.609***	1.000									
3. Interference from high-frequency sense	−0.351***	−0.619***	1.000								
4. Preemption	0.100*	0.016	0.198**	1.000							
5. Entrenchment	0.083**	0.196**	−0.371***	−0.336***	1.000						
6. Years of English learning	−0.458***	−0.565***	0.039	−0.323***	−0.001	1.000					
7. Time spent in L2-speaking countries	<0.001	−0.001	0.001	<0.001	−0.001	0.085**	1.000				
8. Time spent reading books in English	<0.001	0.001	0.002	−0.001	0.001	0.341***	0.019	1.000			
9. Time spent watching films in English	<0.001	−0.002	−0.002	<0.001	−0.001	−0.345***	0.053**	0.125**	1.000		
10. Time spent listening to music in English	<0.001	<0.001	0.002	−0.001	<0.001	−0.388***	0.002	−0.401***	−0.147**	1.000	
11. Time spent using English	<0.001	0.001	0.000	<0.001	<0.001	−0.324***	−0.090**	−0.420***	−0.091*	0.109**	1.000

Note: PVs, phrase verbs. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 6. Correlation matrix of the fixed variables for low-frequency meaning senses: Intermediate learners

	1	2	3	4	5	6	7	8	9	10	11
1. Semantic transparency	1.000										
2. Frequency of PVs	0.609***	1.000									
3. Interference from high-frequency sense	−0.351***	−0.619***	1.000								
4. Preemption	0.100*	0.016	0.198**	1.000							
6. Entrenchment	0.083	0.196**	−0.371***	−0.336***	1.000						
6. Years of English learning	<0.001	<0.001	<0.001	<0.001	<0.001	1.000					
7. Time spent in L2-speaking countries	<0.001	<0.001	<0.001	<0.001	<0.001	0.008*	1.000				
8. Time spent reading books in English	<0.001	<0.001	<0.001	<0.001	<0.001	0.300***	0.009	1.000			
9. Time spent watching films in English	<0.001	<0.001	<0.001	<0.001	<0.001	−0.166**	0.044*	0.113**	1.000		
10. Time spent listening to music in English	<0.001	<0.001	<0.001	<0.001	<0.001	−0.022***	0.011	−0.204***	−0.122**	1.000	
11. Time spent using English	<0.001	<0.001	<0.001	<0.001	<0.001	−0.200***	−0.017**	−0.106***	−0.033*	0.022*	1.000

Note: PVs, phrase verbs. *$p < .05$. **$p < .01$. ***$p < .001$.

Table 7. Selected mixed model of fixed effects on the participants' acceptability of low-frequency senses

Fixed effects	Advanced			Intermediate		
	β	SE	χ^2	β	SE	χ^2
(Intercept)	2.771	1.156		2.226	0.822	
Semantic transparency	**−0.251**	**0.059**	**5.447***	**−0.157**	**0.054**	**7.392****
Frequency of PVs	**0.557**	**0.221**	**5.622***	0.288	0.146	2.463
Entrenchment	−0.091	0.092	1.049	−0.020	0.062	1.531
Preemption	**−0.144**	**0.057**	**4.915***	**−0.113**	**0.044**	**6.337***
Interference from high-frequency sense	−0.059	0.145	0.177	**−0.222**	**0.114**	**5.399***
Years of English learning	−0.010	0.009	1.333	0.009	0.030	0.188
Time spent in L2-speaking countries	0.002	0.006	0.188	−0.022	0.201	0.012
Time spent reading books in English	0.002	0.008	0.103	0.002	0.006	0.359
Time spent watching films in English	0.019	0.030	0.212	−0.014	0.011	0.538
Time spent listening to music in English	0.008	0.007	0.975	0.002	0.012	0.313
Time spent using English	−0.008	0.021	0.109	0.032	0.033	1.227
	$R^2_m = .188$			$R^2_m = .223$		

Note: SE, standard error. R^2_m refers to variance explained by fixed factors. *p < .05. **p < .01.

learning, and time spent reading books and watching films in English, while factors restricting the intermediate learners' development of high-frequency meaning senses were semantic transparency, frequency of PVs, and preemption. Factors constraining the advanced learners' development of low-frequency meaning senses were semantic transparency, frequency of PVs, and preemption, while factors restricting the intermediate learners' learning of low-frequency meaning senses were semantic transparency, interference from high-frequency sense, and preemption.

Discussion

Recent decades have witnessed scholars' persistent efforts to investigate factors that influence the L2 development of English PVs, with the hope of proposing possible accounts of why L2 learners are less likely to achieve nativelike ability in using PVs (Chen, 2013; Dagut & Laufer, 1985; Laufer & Eliasson, 1993; Liao & Fukuya, 2004; Garnier & Schmitt, 2016; Hulstijn & Marchena, 1989; Siyanova & Schmitt, 2007; Schmitt & Redwood, 2011). This study, based on the previous work, examined effects of semantic transparency, frequency of PVs, entrenchment, preemption, and exposure factors such as years of English learning, time spent in L2-speaking

countries, reading books in English, watching English films/TV, listening to English music, and communicating with others in English. Our research yielded the important findings outlined in the following paragraphs.

For the development of PV knowledge, the intermediate and advanced learners tended to accept the use of high-frequency meaning senses, while they were prone to reject the use of low-frequency meaning senses of PVs. For factors that may constrain the development of PV knowledge, the patterns for the two groups were different. Frequency of PVs, semantic transparency, and time spent reading books and watching films in English could reliably predict the advanced learners' knowledge of high-frequency meaning senses, while semantic transparency, frequency of PVs, and preemption could reliably predict their knowledge of low-frequency meaning senses. Semantic transparency, frequency of PVs, and preemption could significantly predict the intermediate learners' knowledge of high-frequency meaning senses, while semantic transparency, interference from high-frequency meaning senses, and preemption could significantly predict the intermediate learners' knowledge of low-frequency meaning senses.

The role of linguistic factors

Our results obtained from the two groups of English-L2 learners were in conformity with the basic tenet of usage-based approaches that the more frequent a construction, the more likely for learners to acquire it. Specifically, we have found a positive relationship between L2 learners' receptive knowledge of PVs and frequency of PVs, supplementing the existing evidence that PV frequency is positively correlated with the L2 learners' productive knowledge of PVs (Chen, 2013; Garnier & Schmitt, 2016; Schmitt & Redwood, 2011). Although our study and previous ones used frequency of PVs from the COCA and BNC, which are, in some cases, not considered as a good estimation of L2 learners' exposure to some particular constructions, we found that corpus frequency was a reliable predictor of L2 knowledge of PVs. This may be due to the fact that PVs are frequently used constructions in spoken and written discourses (Biber et al., 1999; Gardner & Davies, 2007). As pointed out by Garnier and Schmitt (2016), the effect of frequency on the L2 development of PVs observed from the mixed-effects model may help to interpret the relation between frequency and the L2 learners' PV knowledge with more confidence than based on a correlation coefficient, which further lends support to the assumption that frequency is one important contributing factor in L2 acquisition.

Different PVs vary in terms of semantic transparency. Some meanings of PVs are difficult to decipher from their individual components, and some are not. A few studies have demonstrated that it is the semantic complexity that makes L2 learners find it difficult to deal with PVs that convey idiomatic meanings, and therefore reject them (Dagut & Laufer, 1985; Hulstijn & Marchena, 1989; Laufer & Eliasson, 1993). Possible accounts are that unaware L2 learners tend to treat PVs as different orthographic words rather than recognize them as single semantic units in the process of L2 learning (Arnon & Christiansen, 2017; Foster, 2001; Garnier & Schmitt, 2015), unless they are informed that the PV they are learning is a holistic unit. The resultant consequence is that they may decode the meaning of PVs from individual words, and therefore misinterpret them. In doing the task, such misinterpretation

was likely to induce learners to reject the use of target PVs in the perceived sentential context. Note that our study used most of the PVs that were the same as those in Garnier and Schmitt (2016), where the effect of semantic transparency was not detected in Chilean students of English. The inconsistency between these two studies may be attributed to two factors. The first one is associated with knowledge being tested. Garnier and Schmitt (2016) tested the L2 learners' productive PV knowledge while ours examined learners' receptive PV knowledge, which may be differently influenced by semantic complexity. This possible influence, to the best of our knowledge, has seldom been explored in the L2 research agenda. The second inconsistency concerns the identification of literal and figurative meanings. Garnier and Schmitt's (2016) binary classification of PVs based on native speakers' intuition may not be applicable to L2 learners. The results, based on a separate mixed model for both high-frequency meaning senses, suggest that both the advanced and the intermediate learners' judgments on the acceptability of PVs were influenced by semantic complexity. This speaks to the fact that PVs are difficult for learners although they have been exposed to the L2 for many years.

The effect of preemption was significant in the process in which the intermediate learners judged the use of high-frequency meaning senses of PVs. The absence of preemption in the advanced learners was possibly due to the fact that they tended to accept the high-frequency meaning senses (67.2%), which may minimize the effect of preemption. However, the preference of using high-frequency meaning senses in the intermediate learners was comparatively low (51.2%), which made the effect of preemption possible. Note that in the present study, the alternative competing forms for the PV use were in most cases one-word verbs. This is consistent with previous findings (Dagut & Laufer, 1985; Liao & Fukuya, 2004; Siyanova & Schmitt, 2007) that the intermediate English-L2 learners prefer one-word verbs to PVs. The more frequently they encounter or use the one-word verbal equivalents, the more likely they consider the related PVs as less acceptable. That is, well-attested lexical words constrain the use of the target PVs (Clark & Clark, 1979). For low-frequency meaning senses, both the intermediate and the advanced learners were less likely to encounter and use them. Chances were that the well-attested synonymous one-word verbs were easily retrieved and used when learners wanted to express the intended meaning. When they were unsure whether the target PV fit the sentential context in the judgment task, they may replace this PV with the frequently used one-word competing verb. Robenalt and Goldberg (2016) found that L2 learners are conservative in L2 use, and they tend to use well-attested familiar constructions instead of less familiar ones. This may point to the possibility that L2 learners use familiar constructions as "safe bets," hence the underuse of formulaic sequences (AlHassan & Wood, 2015, Boers & Lindstromberg, 2009). It was also in line with the recent finding that L2 learners' ratings on the acceptability of denominal verbs were negatively affected by the frequency of competing verbs (Zhang & Mai, 2018).

The effect of interference from high-frequency meaning senses was significant when the intermediate learners judged the acceptability of the low-frequency meaning senses of PVs. Such an effect has also been detected by Zhao (2010) and Zhang (2010), that the most commonly used sense of a word often blocks the learning of less commonly used senses of this word in novice and intermediate L2 learners. With increasing exposure to PVs that express the high-frequency meaning senses,

the intermediate learners may believe that these PVs seldom express other meanings. Bayesian statistical inference assumes that in the initial stage, learners acquire a construction via analyzing its frequency distribution, and repeated use of a form in Context A constitutes indirect negative evidence for its occurrence in others (Braine & Brooks, 1995; Dowman, 2000; Pinker, 1989).

The effect of entrenchment of verbs was not observed in our current study. This seemingly goes against recent findings concerning L2 learners' avoidance of generalized errors of *un*-prefixation and acceptance of English denominal verbs (Zhang, 2017; Zhang & Mai, 2018). Specifically, frequency of bare forms (e.g., *close*) was a factor associated with the preference of the L2 learners to reject ungrammatical *un*-verbs (e.g., *unclose*), and the frequency of nominal forms of the denominal verbs was significantly negatively correlated to the L2 learners' judgments on the acceptability of all denominal verbs. The null effect of entrenchment in the current study may be attributable to the fact that PVs in this study were selected from the PHaVE List, which is made up of the 150 most common English PVs, and the high-frequency and well-attested use of PVs may minimize the effect of entrenchment that potentially constrains the development of PVs, as we found that both groups of participants' acceptance of different meaning senses of PVs was significantly positively correlated with the PV frequency.

The role of extra-linguistic factors

Different from Garnier and Schmitt (2016), time spent in watching films/TV in English in our study significantly predicted the advanced learners' knowledge of high-frequency meaning senses of PVs. This is because most of our participants were teachers of English who often used film clips in their teaching. As several of them commented, they often emphasized some common multiword units in their teaching, and therefore paid much attention to these constructions, which could strengthen the retention of these units in the long-term memory. In addition, Garnier and Schmitt (2016) combined all the polysemous senses in their data analysis, which might mitigate this effect. Therefore, the effect of watching films/TV in English in the development of different type of senses calls for further attention. As found by Garnier and Schmitt (2016), time spent reading books, journals, magazines, and newspapers in English was a reliable factor that could predict the advanced learners' knowledge of high-frequency meaning senses. This is because L2 learners learned English mainly from books. PVs are pervasive in both spoken and written language. The more learners read, the more likely learners encounter PVs, and the more likely learners acquire them. The effect of listening to music in English was ineffective, because learners often focus on the melody of English songs instead of the lyrics. Moreover, unlike daily communication, this kind of aural input is very difficult for L2 learners to understand, let alone to remember the target construction thereof.

In the present study, no effect of years spent in English-speaking countries was found on the development of PV knowledge. On average, the advanced learners had spent 5.4 months in the United Kingdom or the United States ($N = 60$; Min $= 0$; Max $= 18$; $SD = 3.77$), while the intermediate learners had less than 1 month of experience in the English-speaking countries ($N = 60$; Min $= 0$; Max $= 6$; $SD = 1.04$).

Such a short period of living or working experience in the English-speaking countries is less likely for learners to benefit from L2 immersion. This was not observed by Garnier and Schmitt (2016) and Siyanova and Schmitt (2007). The acquisition of complex constructions that do not exist in learners' L1 like English PVs is predicated upon the learners' full adaptation to the L2 community (Adolphs & Durow, 2004; Dornyei, Durow, & Zahran, 2004). Too little time spent in L2 countries made this chance too faint.

The effect of time spent reading books in English was not significant for the intermediate learners. This contrasts with previous findings that reading has been found to reliably facilitate lexical development (Horst, Cobb, & Meara, 1998; Pigada & Schmitt, 2006) including PVs (Garnier & Schmitt, 2016; Gonzalez Fernandez & Schmitt, 2015; Schmitt & Redwood, 2011). A scrutiny of our participants' data obtained from the questionnaire revealed that the intermediate learners spent almost the same amount of time in reading books in English. To be specific, they spend time mainly reading their English coursebooks with little time devoted to English newspapers, journals, and magazines, which made the effect of reading books ineffective in our mixed model.

The invalid effect of time spent communicating with others in our study may be attributed to the fact that our participants spent little time in communicating with others in English, because they were living in China and had little chance to use English, even if they used QQ or We-chat to communicate with foreign teachers. In this sense, our findings do not mean that time devoted in reading and communication made no difference in PV learning. On the contrary, rich exposure to L2 in different readings and via communicating with native speakers could benefit learners substantially in L2 lexical development.

Conclusion

The present study examined multiple factors that may constrain the Chinese speakers' learning of English PVs, finding that both the intermediate and the advanced Chinese English-L2 learners tended to favor the high-frequency meaning senses of English PVs but disfavor the low-frequency meaning sense of PVs, and that the major constraints that restricted the development of PVs seemed different in distinct stages of PV learning. For the advanced learners, frequency of PVs, semantic transparency, and time spent reading books and watching films/TV in English could reliably predict their knowledge of high-frequency meaning senses, while frequency of PVs, and preemption could reliably predict their development of low-frequency meaning senses. For the intermediate learners, semantic transparency, frequency of PVs, and preemption were significant factors that could predict their acceptance of high-frequency meaning senses, while semantic transparency, interference from high-frequency meaning senses, and preemption were effective in the prediction of their judgments on the acceptability of low-frequency meaning senses. What needs to be stressed is that preemption was significant in the distinct stages of development of two types of meaning senses, while entrenchment was ineffective in the intermediate learners' development of low-frequency meaning senses. Based on these findings, we recommend that to explain the learners' limited ability in L2,

attention should be paid to the effect of multiple potential variables, *inter alia*, preemption.

Albeit informative, we should be cautious in generalizing our findings to other constructions and languages due to several shortcomings. The first one is that we did not set a standard for splitting high-frequency and low-frequency meaning senses for PVs, although there is no established standard for identifying the high-frequency PVs. More rigorous selecting procedures should be used in future studies to take into consideration the effect of frequencies on L2 learning gains. The second one has to do with the fact that we only used a receptive task to measure learners' PV knowledge. Conclusions may be more valid if both receptive and productive tasks were utilized, because Liao and Fukuya (2004) found that L2 learners' use of PVs tended to vary in different testing tasks. The third one is probably related to testing bias that in doing the judgment task, our participants had seen two usages of the same verb, which are likely to bring about bias in their judgments in the same test session, so future studies are expected to obtain judgment scores for the two senses of PVs in distinct test sessions in order to avoid such bias.

Author ORCIDs. (iD) Xiaopeng Zhang 0000-0002-5409-6293

Notes

1. The range on Wang and Koda's (2005) 5-point scale familiarity test are 5 = *known* (*I think each of us knows this phrase and can use it productively*), 4 = *very familiar* (*I think most of us are familiar with this phrase*), 3 = *familiar* (*I think many of us are familiar with this phrase*), 2 = *not likely familiar* (*I think many of us are not familiar with this phrase*), and 1 = *not at all familiar* (*I don't think most of us have seen this phrase before*). If the averaged score for a PV was below 2.5, then this PV would be excluded.

2. For the detailed procedures of tallying the occurrences of different meaning senses for the polysemous sense of PVs, readers are referred to Garnier and Schmitt (2015).

References

Adolphs, S., & Durow, V. (2004). Social-cultural integration and the development of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 107–126). Amsterdam: Benjamins.

Alderson, C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment.* London: Continuum.

AlHassan, L., & Wood, D. (2015). The effectiveness of focused instruction of formulaic sequences in augmenting L2 learners' academic writing skills: A quantitative research study. *Journal of English for Academic Purposes*, **17**, 51–62.

Ambridge, B. (2013). How do children restrict their linguistic generalizations? An (un-)grammaticality judgment study. *Cognitive Science*, **37**, 508–543.

Ambridge, B., Bidgood, A., Twomey, E., Pine, J. M., Rowland, C. F. & Freudenthal, D. (2015). Preemption versus entrenchment: Towards a construction-general solution to the problem of the retreat from verb argument structure overgeneralization. *PLOS ONE*, **10**, e0123723.

Ambridge, B., Pine, J. M., & Rowland, C. F. (2012). Semantics versus statistics in the retreat from locative overgeneralization errors. *Cognition*, **123**, 260–279.

Ambridge, B., Pine, J. M., Rowland, C. F., Jones, R. L., & Clark, V. (2009). A semantics-based approach to the "no negative evidence" problem. *Cognitive Science*, **33**, 1301–1316.

Ambridge, B., Pine, J. M., Rowland, C. F., & Young, C. R. (2008). The effect of verb semantic class and verb frequency (entrenchment) on children's and adults' graded judgments of argument structure overgeneralization errors. *Cognition*, **106**, 87–129.

Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1-L2 differences. *Topics in Cognitive Science*, **9**, 621–636.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, **68**, 255–278.

Bates, D., Maechler, M., Bolker, B., &Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, **67**, 1–48.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, UK: Longman.

Boers, F., & Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition*. Basingstoke: Palgrave Macmillan.

Bowerman, M. (1988). The "no negative evidence" problem: How do children avoid constructing an overly general grammar? In J. A. Hawkins (Ed.), *Explaining language universals* (pp. 73–101). Oxford: Blackwell.

Boyd, J. K., & Goldberg, A. E. (2011). Learning what not to say: The role of statistical preemption and categorization in a-adjective production. *Language*, **87**, 55–83.

Braine, M. D. S., & Brooks, P. J. (1995). Verb argument structure and the problem of avoiding an overgeneral grammar. In M. Tomasello, and W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 352–376). Hillsdale, NJ: Erlbaum.

Brooks, P. J., Tomasello, M., Dodson, K., & Lewis, L. B. (1999). Young children's overgeneralizations with fixed transitivity verbs. *Child Development*, **70**, 1325–1337.

Chen, M. (2013). Overuse or underuse: A corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics*, **18**, 418–442.

Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language*, **55**, 767–811.

Dagut, M., & Laufer, B. (1985). Avoidance of phrasal verbs: A case for contrastive analysis. *Studies in Second Language Acquisition*, **7**, 73–79.

David, B. (1991). *Conditioning diagnostics: Collinearity and weak data in regression*. New York: Wiley.

Dirven, R. (2001). English phrasal verbs: Theory and didactic application. In M. Putz, S. Niemeier, and R. Dirven (Eds.), *Applied cognitive linguistics: II. Language pedagogy* (pp. 3–27). Berlin: de Gruyter.

Dornyei, Z., Durow, V., & Zahran, K. (2004). Individual differences and their effects on formulaic sequence acquisition. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, and use* (pp. 87–106). Amsterdam: Benjamins.

Dowman, M. (2000). Addressing the learnability of verb subcategorizations with Bayesian inference. In L. Gleitman, and A. Joshi (Eds.), *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 101–112). Austin, TX: Cognitive Science Society.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, **24**, 143–188.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, and M. Swain (Eds.), *Researching pedagogic tasks: Second language learning. Teaching and testing* (pp. 75–93). Harlow, UK: Longman.

Freeborn, D. (1995). *A course book in English grammar*. London: Macmillan.

Gardner, D., & Davies, M. (2007). Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly*, **41**, 339–359.

Garnier, M., & Schmitt, N. (2015). The PHaVE List: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research*, **19**, 645–666.

Garnier, M., & Schmitt, N. (2016). Picking Up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System*, **59**, 29–44.

Gonzalez Fernandez, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *ITL International Journal of Applied Linguistics*, **166**, 94–126.

Gyllstad, H., & Wolter, B. (2016). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning*, **66**(2), 296–323.

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

Howarth, P. (1996). Phraseology in English academic writing: Some implications for language learning and dictionary making. Lexicographica Series Maior, Volume **75**. Tübingen: Niemeyer.

Hulstijn, J. H., & Marchena, E. (1989). Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition*, **11**, 241–255.

Kording, K. P., & Wolpert, D. M. (2011). Bayesian statistics and utility functions in sensorimotorcontrol. In K. Doy, S. Ishii, A. Pouget, and R. P. N. Rao (Eds.), *Bayesian brain: Probabilistic approaches to neural coding* (pp. 299–316). Cambridge, MA: MIT Press.

Kurtyka, A. (2001). Teaching English phrasal verbs: A cognitive approach. In M. Putz, S. Niemeier, and R. Dirven (Eds.), *Applied cognitive linguistics: II. Language pedagogy* (pp. 29–54). Berlin: de Gruyter.

Kurvers, J., & Uri, H. (2006). Metalexical awareness: Development, methodology or written language? *Journal of Psycholinguistic Research*, **35**, 353–367.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. **1**). Stanford, CA: Stanford University Press.

Laufer, B., & Eliasson, S. (1993). What causes avoidance in L2 learning: L1-L2 difference, L1-L2 similarity, or L2 complexity? *Studies in Second Language Acquisition*, **15**, 35–48.

Liao, Y., & Fukuya, Y. J. (2004). Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning*, **54**, 193–226.

Liu, D. (2011). The most frequently used English phrasal verbs in American and British English: A multi-corpus examination. *TESOL Quarterly*, **45**, 661–688.

Macis, M., & Schmitt, N. (2017). The figurative and polysemous nature of collocations and their place in ELT. *ELT Journal*, **71**, 50–59.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2005). The role of frequency in the acquisition of English word order. *Cognitive Development*, **20**, 121–136.

Meara, P. (2005). *Y_lex [computer software]*. Swansea, UK: Lognostics.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, **18**, 1–28.

Pinker, S. (1989). *Learnability and cognition*. Cambridge, MA: MIT Press.

R Development Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at http://www.R-project.org/

Robenalt, C., & Goldberg, A. E. (2016). Nonnative speakers do not take competing alternative expressions into account the way native speakers do. *Language Learning*, **66**, 66–94.

Schmitt, N. (2014). *Researching vocabulary: A vocabulary research manual*. Beijing: Foreign Language Teaching and Research Press.

Schmitt, N., & Redwood, S. (2011). Learner knowledge of phrasal verbs: A corpus-informed study. In F. Meunier, S. de Cock, G. Gilquin, and M. Paquot (Eds.), *A taste for corpora*. Amsterdam: Benjamins.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word vs. one-word verbs. *International Review of Applied Linguistics*, **45**, 119–139.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom-processing by native and non-native speakers. *Second Language Research*, **27**, 251–272.

Taylor, J. R. (2004). Why construction grammar is radical. *Annual Review of Cognitive Linguistics*, **2**, 321–348.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Waibel, B. (2007). *Phrasal verbs in learner English: A corpus-based study of German and Italian students* (Unpublished doctoral dissertation, Albert-Ludwigs-Universität Freiburg).

Wang, M., & Koda, K. (2005). Commonalities and differences in word identification skills among learners of English as a second language. *Language Learning*, **55**, 71–98.

Zhang, X. (2010). Function of sense typicality in the representation and retrieval of English polysemous words. *Modern Foreign Languages*, **33**, 395–402.

Zhang, X. (2017). Second language users' restriction of linguistic generalization errors: The case of English un-prefixation development. *Language Learning*, **67**, 569–598.

Zhang, X., & Mai, C. (2018). Effects of entrenchment and preemption in second language learners' acceptance of English denominal verbs. *Applied Psycholinguistics*, **39**, 413–436.

Zhao, C. (2010). The representation of ambiguous words in Chinese EFL learners' mental lexicon: A developmental model. *Modern Foreign Languages*, **33**, 55–63.