

Lexical transfer in the writing of Chinese learners of English

Marina Dodigovic*

American University of Armenia, Armenia

Chengchen Ma

Xi'an Jiaotong-Liverpool University, China

Song Jing

Australian National University, Australia

Abstract

This study aims to further the understanding of first language (L1) lexical transfer within the context of L1 Chinese learners of English. Previous transfer research has often focused on a small subset of grammar errors, without examining how lexical choices, especially in collocations and multi-word units (MWU), might have been influenced by L1 or L1-based assumptions about vocabulary use. There is therefore a need to look for evidence of L1 transfer or word-for-word translation from the native language in L2 production at each of the three levels: individual words, collocations and MWU. Such errors points to subordinate bilingualism, which is rooted in translation as a teaching/learning method (Cook, 2014), which is common in China (Edmunds, 2013). Therefore this paper addresses the following research questions: 1) To what extent does the transfer of L1 word polysemy, collocations, and MWU impact Chinese learners' English vocabulary use? 2) Are more advanced learners as prone to L1 lexical transfer errors as the less advanced ones? The approach used here is corpus-linguistic. The main research task is to examine an existing corpus of Chinese student writing in English and analyze and classify the identified lexical transfer errors. The findings indicate that the most common of these are errors caused by L1 polysemy in individual words, followed by MWU and collocation errors. More advanced learners appear to be slightly but not significantly less prone to lexical transfer errors. Instruction which follows the recommendations made in this paper is likely to prevent the onset of such errors.

Keywords: lexical transfer, polysemy, collocation, multi-word unit, subordinate bilingualism

Introduction

First language (L1) transfer, sometimes also called L1 interference or 'cross-linguistic influence' (Jarvis & Pavlenko, 2008), has been found at different linguistic levels, from phonology and spelling to discourse. Although transfer research has so far mainly focused on grammar, MacWhinney (1992) suggested that a significant number of L1 transfer cases were found at the lexical level as well. However, only a modest amount of research has been conducted to investigate the impact of L1 transfer on L2 vocabulary acquisition. Yet, vocabulary is of tremendous importance, as 'without vocabulary nothing can be conveyed' (Wilkins, cited in Thornbury, 2002). Words build language structures and convey L2 learners' intended meaning, but only when they are appropriately selected and used (Shalaby, Yahya & Kl-Komi, 2009). Inappropriate use of words could lead to errors and miscommunication. Such errors are called lexical errors (Augustin Llach, 2011).

Errors have served as indices of writing quality in formal contexts. In other words, there is 'a negative correlation between quality writing and linguistics errors in general and lexical errors in particular' (Augustin Llach, 2011, pp.67). In order to be able to deal adequately with lexical errors found in learner writing, language teachers should be aware of the sources of such errors. This is particularly important in Chinese contexts, where

*Tel.: +374 60612-740; Email: mdodigovic@aua.am; 40 Marshal Baghramyan Ave., Yerevan, 0019, Armenia

corpus studies have already recognized a significant presence of lexical errors (Chan, 2010; Liu, 2011; Edmunds, 2013).

Relevant research (Hemchua & Schmitt, 2006; Zhou, 2010; Xia, 2013) suggests that lexical errors do not only occur at single word level, but also at collocation and multi-word unit (MWU) levels (Gray & Biber, 2013). Among those are lexical transfer errors which have been identified at every level (Yang, Ma & Cao, 2013; Li, 2005; Yamashita & Jiang, 2010), although not necessarily all within the same study. In short, lexical transfer errors are lexical errors caused by L1 transfer. The identification of lexical transfer errors in the English output of Chinese learners coincides with the finding that grammar-translation method happens to be the prevalent L2 instruction approach in China (Edmunds, 2013). If confirmed, such findings could have profound implications for language teaching practice in Chinese contexts. Hence, the present study aims to identify and analyze the cases of negative lexical transfer from Chinese to English caused by 1) Chinese word polysemy at single word level, 2) Chinese collocations, and 3) Chinese MWU.

Previous Research on Lexical Transfer

The literature on L2 lexis is based on two broad traditions: one is applied linguistics and the other is psycholinguistic (Schmitt & Meara, 1997). Two types of lexical studies are generally found in applied linguistics, one being the study of vocabulary size and the other being that of vocabulary depth. Vocabulary size, or the number of words known (Schmitt, 2000), is indeed a critical indicator of a learner's lexical competence, but a number of researchers now believe that vocabulary depth is the crucial factor in measuring L2 vocabulary knowledge (Folse, 2006). This includes the meaning(s) of a word, its written form, spoken form, grammatical behavior, the collocations it can build, the register it is associated with, the associations it attracts, as well as its frequency (Schmitt, 2000).

Central to the psycholinguistic study of lexis is the notion of 'mental lexicon,' which refers to the representation of word knowledge in permanent memory (Carroll, 2007). According to Jiang (2000), two terms are used to describe a lexical entry. One is lexeme (also called lexical unit or lexical item), and the other is lemma. A lexeme consists of morphological and phonological information, which functions as a single meaning unit, while the lemma contains semantic and syntactic information (Jiang, 2000). In L1, the four aspects of a lexical item (semantics, syntax, morphology, and phonology) are highly integrated with each other, so that one activated feature could activate the rest (Jiang, 2000). However, L2 lexical representation and processing in L2 learners' mind is different (Cook, 2014).

According to Jiang (2000), there are two constraints in L2 vocabulary learning. The first is the lack of both adequate quantity and quality of L2 input, and the other one is the influence of the well-structured L1 lexical system and its associated conceptual system. The former constraint makes it difficult for L2 learners to construct 'semantic, syntactic, and morphological specifications ... [of] a word' (Jiang, 2000, p. 49), while the latter makes it easy for L2 learners to rely on their L1 knowledge while building the L2 language system. There are three stages in Jiang's (2000) L2 lexical processing model, namely the initial stage, the L1 lemma mediation stage and the L2 integration stage. In particular at the second stage, the link between L2 words and their associated L1 translation equivalents becomes stronger (Jiang, 2000). Due to the above constraints, many L2 learners never attain the third stage, their L2 vocabulary thus remaining dependent on L1 lexical features.

Similarly, Wolter (2006) suggested that L2 learners' misunderstanding of L2 vocabulary is partly due to their lack of L2 lexical knowledge, and partly because of excessive dependence on L1 lexis. Furthermore, Lewis (1997) assumed that it is natural for L2 learners to fall back on their L1 when unable to express themselves in L2. Finally, according to Cook (2014), an L2 word is often associated not only with a concept, but with the entire L1 word, which leads to subordinate bilingualism. Moreover, it is also deemed that less proficient learners are more likely to resort to connecting L2 words with the L1 ones (Kroll & Stewart, 1994; Jiang, 2002).

However, Pienemann et al. (2005) claimed the opposite, i.e. that advanced learners are more likely to exhibit the signs of subordinate bilingualism in terms of L2 lexis than novice learners. These authors hold the view that learners do not transfer elements from their L1 until they have acquired an adequate amount of L2

knowledge. This does not surprise in view of Pienemann's (2003) processability theory, which posits that L2 acquisition is beholden to predictable developmental sequences, allowing learners to acquire only those linguistic forms which are appropriate for their developmental stage. Thus, Pienemann et al. (2005) are more inclined to believe that early errors are a consequence of stagewise L2 development, i.e. that they are developmental.

What Cook (2014) calls subordinate bilingualism is akin to the notion of L1 transfer. The concept of 'transfer,' a product of behaviorism, was used extensively in the early years of SLA to refer to the process in which the learners' L1 influences L2 in a positive or negative way (Gass, 2013). 'Transfer' or 'cross-linguistic influence' is preferred by linguists in SLA, while the term 'interference' is used more commonly in psycholinguistic approaches to SLA (Jarvis & Pavlenko, 2008). More recently, there have been attempts to redefine the terms. Thus, Grosjean (2001) pointed out that the term 'transfer' should be used to refer to the permanent influence of L1 on L2, while 'interference' should be used to describe L1 features occurring from time to time in L2. In this study, transfer is used in the sense of the definition rendered by Gass (2013). In the same vein, L2 errors caused by L1 transfer are called transfer errors.

Transfer errors as well as developmental errors are deemed to contribute to the state of interlanguage (Yip, 1995) or the language of L2 learners which is found somewhere along the continuum between the L1 and L2 (Han & Selinker, 1999). Within the interlanguage framework, those errors which defy correction and persist despite repeated instruction are called fossilized errors (Gass & Selinker, 2008). Fossilisation resembles Grosjean's (2001) narrow definition of transfer. It is connected with the Multiple Effects Principle (MEP), "which predicts that when language transfer works in tandem with one or more second language acquisition processes" (Han & Selinker, 1999, p. 248) interlanguage structures are more likely to stabilize, leading to a permanent influence of L1 on L2 (Han & Odlin, 2006).

Cases of L1 transfer can be exacerbated by polysemous L2 words or such words that have more than one meaning sense (Schmitt, 2000). For example, Lennon (1996) found that in speech production, L2 English learners, even the advanced ones, frequently made errors while using polysemous words such as 'go,' 'put' and 'take.' Morimoto and Loewen (2007) conducted a quasi-experimental study involving 58 Japanese high-school L2 English learners to compare the effectiveness of two kinds of vocabulary instruction, the image-schema-based instruction and translation-based instruction, on learning of English polysemous words. The results revealed that image-schema-based instruction is more effective than the translation-based instruction.

Conversely, few studies have addressed the influence of L1 word polysemy on L2 vocabulary acquisition. Amongst the few is the study by Duan and Qin (2012), which argues that, unlike English, Chinese makes use of the same word (character) to express a number of meanings, which allows the Chinese language to be economical. However, this misleads Chinese learners to believe that English follows the same pattern, which could result in L1 lexical transfer errors.

Another lexical area with the potential for L1 transfer is collocation. Collocations are described as 'the combinations of words which occur naturally with greater than random frequency' (Lewis, 1997, pp.25). Collocational knowledge as a parameter which 'distinguishes native speakers from nonnative speakers' (Schmitt, 2000, pp. 79) is not easy for L2 learners to acquire. Even advanced learners appear to have difficulty with L2 collocations (Yamashita & Jiang, 2010).

L1 influence was found to play a critical role in a number of studies of L2 collocation, in which researchers either asked learners to take elicitation tests of collocations or collected and analyzed collocations in learner production. The former type is more common. Thus, a translation test was developed by Biskup (1990, cited in Nesselhauf, 2005) to investigate Polish learners' knowledge of English collocations, and it was discovered that participants seldom made errors in translation from L2 collocations to L1, but a considerable number of errors occurred when translating collocations from L1 to L2. Yamashita and Jiang (2010) conducted a study involving 47 Japanese learners of English, in which participants were shown some collocations on a computer screen, after which they needed to judge whether or not those collocations were acceptable English collocations. The findings showed that English collocations which are congruent with Japanese collocations are more easily learned than those that are incongruent. Nesselhauf (2003) analyzed the use of English collocations in a learner

corpus consisting of 32 essays written by L2 learners of English whose L1 was German. The study discovered that 56% of collocation errors could be attributed to negative L1 transfer.

Studies with Chinese speakers have yielded similar results. Based on a cloze test administered to Chinese learners, Shei (1999, cited in Nesselhauf, 2005) concluded that advanced Chinese learners found it more difficult to learn English collocations when compared with their European counterparts. The research by Lombard (1997, cited in Nesselhauf, 2005) which investigated the collocations produced by Chinese learners also found that 10% of non-native-like collocations might be due to L1 transfer, other sources being the incorrect use of English synonyms. Wang (2011) used three collocation tests to investigate language transfer in the acquisition of light verb + noun collocations by Chinese learners. He found that 61.84% of the participants' uses of English light verb + noun collocations may be traced to either positive or negative transfer from Chinese. Hence he concluded that the influence of Chinese on the acquisition of English collocation was obvious and significant. He also suggested that the priority in teaching L2 collocations should be given to the L1 incongruent ones. In Chen and Lin's (2011) study, 355 first-year college students from three different universities in Taiwan were asked to complete a 50-item multiple-choice collocation test and questionnaire. The results showed that L1 transfer, together with overgeneralization and misapplication of synonyms, was one of the top three factors leading to collocation errors.

Other scholars analyzed learner writing to explore common collocation errors made by Chinese learners (Li, 2005; Duan & Qin, 2012; Yang, Ma, & Cao, 2013). For example, Duan and Qin (2012) found that some common English collocation errors such as 'eat medicine,' 'find an object,' 'pay time,' are due to negative transfer from Chinese collocations. More evidence of transfer of Chinese collocations comes from Yang, Ma and Cao (2013). They argue that due to L1 transfer, Chinese learners of English frequently produce unacceptable collocations in English, such as 'learn knowledge' or 'strong competition.'

Beyond collocations, which are usually restricted to two words, there are longer strings of words called multi-word units or MWU (Schmitt, 2000). They have been categorized into four linguistic categories: 'phrasal verbs,' 'fixed phrases,' 'idioms' and 'proverbs' (Schmitt, 2000). From the point of view of language production, MWU are regarded as 'formulaic expressions,' 'lexical phrases,' or 'lexical chunks' which are stored in long-term memory and can be easily activated (Pawley & Syder, 1983). Therefore, they are one of the key elements of accurate, fluent and efficient linguistic production, playing a critical role in SLA (Biber et al., 1999; Erman and Warren, 2000; Hyland, 2008).

Furthermore, except for idioms, whose meaning does not equal the meanings of their component parts, most MWU are to some extent compositional (Nation, 2013). In other words, the meaning of each component part in an MWU contributes to the meaning of the whole. As the selection of words in an MWU is not arbitrary, the knowledge of its parts is conducive to the understanding of the whole MWU (Bogaards, 2001; Boers & Lindstromberg, 2009). MWU are also regarded as being transferable because they are semantically and syntactically compositional. L2 MWU studies (e.g. Rafee, Tavakoli & Amirian, 2011; Adel & Erman, 2012; Karabacak & Qin, 2013) indicate that, compared with native speakers, L2 learners use fewer MWU, while also being likely to overuse certain MWU and underuse others.

Few studies have so far addressed the L1 influence on the acquisition of L2 MWU. One of the exceptions is Peromingo (2012), who explored the L1 influence on L2 learners' production of both correct and incorrect English MWU by analyzing argumentative writing from several learner corpora. The findings suggest that L2 learners tend to overuse the MWU which are similar to the L1 ones.

Paquot (2013) analyzed the writing by French learners of English from the first version of International Corpus of Learner English (ICLE) with special focus on their use of English 3-word sequences with a lexical verb (marked). The results indicated that French learners made few errors in using English 3-word sequences with a lexical verb. However, more errors were found in their selection of English unmarked word combinations, whose French translation equivalents are easy to trace.

In conclusion, studies of lexical transfer have so far examined this phenomenon in three different contexts: single word polysemy, collocation and MWU, although not necessarily all at once. These contexts coincide with the scope of the term *lexis*, which subsumes not only single words, but also collocations and MWU

(Schmitt, 2000; Thornbury, 2002). Hence, any comprehensive studies of lexical transfer should not be restricted to individual words, but would need to observe the effects of L1 on words in both collocational and MWU contexts (Gray & Biber, 2013). In accordance with this, Dodigovic, Wei and Jing (2015) proposed the following taxonomy of the contexts for lexical transfer from Chinese to English 1) Chinese word polysemy, 2) Chinese collocations, 3) Chinese MWU. This taxonomy is followed in the current study.

Cook (2014) identifies the grammar-translation approach to teaching as one of the causes of subordinate bilingualism or lexical transfer. This approach appears to be the dominant L2 instruction methods in China (Edmunds, 2013). Vocabulary teaching in this method encourages establishing links between L1 and L2 single words, but does not necessarily pay attention to collocations or MWU, which sets the stage for word-for-word translation and hence lexical error. Hence, it is important to investigate the evidence of lexical transfer in China, in all of its lexical contexts. The findings could have significant implications for L2 teaching practice in the Chinese speaking world. Hence, the current study aims to identify and analyze the cases of negative lexical transfer from Chinese to English caused by 1) Chinese word polysemy at single word level, 2) Chinese collocations, and 3) Chinese MWU. In doing so, it addresses the need for a comprehensive approach to lexis in lexical transfer research, since corpus studies have to some extent examined individual aspect, but not necessarily the entire scope of lexis in Chinese-English interlanguage.

Research Questions

To obtain a better understanding of the influence of Chinese as L1 on English vocabulary learning, the present study attempts to address the following research questions: 1) To what extent does the transfer of L1 word polysemy, collocations, and MWU impact Chinese learners' English vocabulary use? 2) Are more advanced learners as prone to L1 lexical transfer errors as the less advanced ones? Both, Chinese learners of English and their teachers, stand to gain from the answers to these questions and the ensuing implications for the teaching practice. It is hoped that the insights gained through this study will contribute to the increase of English language proficiency in Chinese contexts.

Methodology

Data

The learner corpus used in the present study consists of 100 samples of writing (541,482 words in total). Fifty of those were written by first year students and 50 by fourth year students at a Sino-British English medium university in China. The students whose writing was included were native speakers of Chinese from the department of English. They were aged between 18 and 23, and had been learning English for at least 6 years (three years in middle school and three years in high school) prior to university enrolment. Research ethics was abided by as per university protocol.

The genre and topics of year one and year four student writing are somewhat different. First year students were required to write a 1,000-word essay to demonstrate their appreciation of a given movie, a poem or a novel chapter, on which they were able to work for several weeks. In contrast, year four students presented their Final Year Projects (FYP), on which they had worked for an entire year. Each FYP was at least 10,000 words long. Despite the considerable differences, each type of assignment was level-appropriate, thus being a true indicator of the writers' ability in written English.

In the present study, each instance of negative lexical transfer from Chinese to English, in word, collocation, or MWU, was counted as one error. Instances of L1 lexical transfer, highlighted and marked in the corpus, were grouped into the three pre-defined categories: 1) those caused by Chinese word polysemy, 2) Chinese collocations and 3) Chinese MWU. The average length of writing by first year students was 1,000 words, while that by final year students was 10,000. Due to the considerable difference in word count, it was deemed that the raw frequencies could be misleading. In order to make the results better comparable across sub-samples,

there was a need to ‘norm’ the raw frequency (Biber, Conrad & Reppen, 1998). In the present study, the counts were normed to a basis of 1,000 words using the following formula:

Raw error frequency = (number of errors / total word count) * 1,000

The data was subsequently processed using the IBM SPSS Statistics Version 20 for both descriptive and inferential analysis.

Procedure

The corpus of student writing was manually analysed for lexical errors by initially three and later two Chinese-English bilinguals, who used dictionaries and the English corpora available through lextutor.ca to assess the accuracy of their linguistic judgement. The accuracy of their judgement was subsequently verified by a native English speaker. Based on their native speaker Chinese competence, the analysts also decided whether the context causing error is the polysemy of the Chinese equivalent of the target English word, an underlying Chinese collocation or a Chinese MWU. Each instance was counted as one error. For example, in cases where the error was based on an underlying Chinese collocation which also included a polysemous Chinese word, the error was counted as collocation error. Thus, larger lexical units took precedence over smaller ones in error count. The accuracy of the Chinese aspect of this research was subsequently verified by an L1 Chinese rater.

In order to investigate whether the polysemous Chinese words identified in the learner corpus as the cause of lexical transfer are frequently used by native speakers of Chinese, the corpus of ‘Texts Of Recent Chinese,’ whose acronym is ToRCH (TORCH2009, Texts of Recent Chinese, Brown family, 2009, 2013 summer edition) available from <http://111.200.194.212/cqp/torch09/> was used in the study. The ToRCH project was initiated under the name of CC2009 meaning Chinese corpus 2009 by Corpus Research Group at Beijing Foreign Studies University. The current version was finalized on 20 July 2014 after the removal of some duplicated portions of texts. The corpus contains 1,066,347 tokenized Chinese words or 1,670,356 Chinese characters from texts of 15 types (Press: Reportage, Press: Editorial, Press: Reviews, Religion, Skill and hobbies, Popular lore, Belles-lettres, Miscellaneous: Government and house organs, Learned, Fiction: General, Fiction: Mystery, Fiction: Science, Fiction: Adventure, Fiction: Romance, and Humor). While the polysemy of the Chinese words was judged by at least two native speakers of Chinese and later verified by an L1 Chinese rater, the frequency was measured in terms of the word’s frequency ranking within the ToRCH corpus.

Quantitative Research Results

The 100 writing samples in the learner corpus analyzed in the current study yielded a total of 395 lexical transfer errors (199 caused by Chinese word polysemy, 87 by Chinese collocations, and 109 by Chinese MWU). In order to answer the first research question (To what extent does the transfer of L1 word polysemy, collocations, and MWU impact Chinese learners’ English vocabulary use?), the number of lexical errors caused by the three categories of Chinese transfer was counted. As can be seen in the Figure 1 below, a clear difference was found in the proportion of lexical errors caused by 1) Chinese word polysemy, 2) Chinese collocations and 3) Chinese MWU.

The greatest number of errors was caused by Chinese word polysemy (50%), which accounts for one half of the total number of L1 lexical transfer errors. It is followed by errors resulting from Chinese MWU (28%) and Chinese collocations (22%).

The frequency of the Chinese words identified in the study as those causing transfer because of their polysemy is displayed in the Table 1. As can be seen in the table, the Chinese character ‘有(you)’ appears 6,539 times in the corpus of ToRCH, ranking as the first, followed by the character ‘都(dou)’ with a frequency of 3,747. The third most frequent Chinese character is ‘大(da)’ with a frequency of 3,078. It would appear that the more frequent the word, the more often it causes L1 transfer in L2.

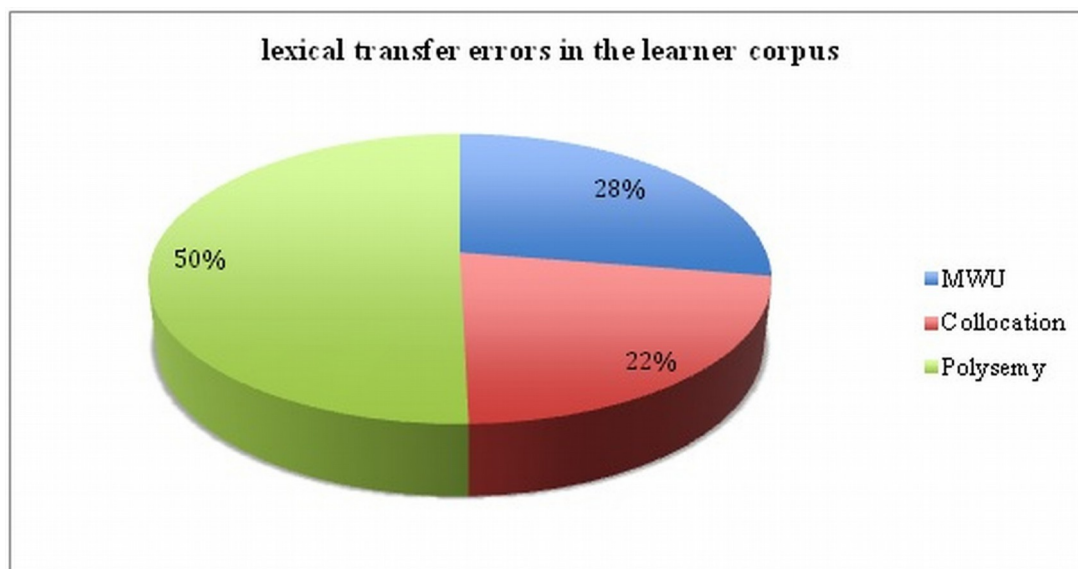


Figure 1. Errors in three categories

Table 1

The Frequency of Some Chinese Polysemous Words in the Corpus of Texts of Recent Chinese

NO.	Chinese Words/characters	Frequency
1	有	6,539
2	都	3,747
3	大	3,078
4	还	2,581
5	好	2,160
6	后	1,893
7	看	1,811
8	用	1,548
9	做	1,365
10	国家	1,023
11	重要	940
12	需要	825
13	通过	681
14	主要	650
15	方面	632
16	情况	567
17	提高	564
18	作用	508
19	变化	500

The second research question (Are more advanced learners as prone to L1 lexical transfer errors as the less advanced ones?) was answered by comparing the error frequency in the writing of year one and year four students. Compared with the writing of year one students (less advanced learners), fewer lexical transfer errors were identified in the writing of year four students (more advanced learners).

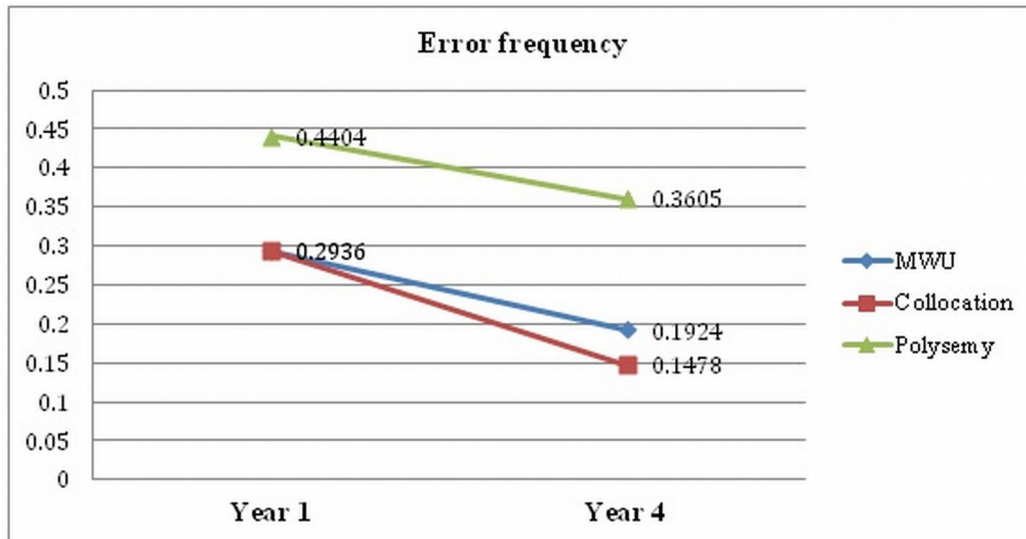


Figure 2. Error frequency in writing by year one and year four students

Figure 2 compares the frequency of lexical errors in the writing of first year and final year students respectively in terms of the three transfer categories on a 1,000-word basis. Specifically, in the papers written by first year students, in every 1,000-word unit, 0.4404 error could be attributed to Chinese word polysemy, and the number of errors caused by the transfer of Chinese collocations was the same as the number of those caused by MWU (0.2936 and 0.2936 respectively). However, the figures decreased in the writing of final year students. While 0.3605 errors were caused by the transfer of Chinese word polysemy, 0.1478 were caused by Chinese collocations and 0.1924 by Chinese MWU.

An independent sample t-test was conducted using SPSS statistics software. The difference in the frequency of lexical transfer from Chinese at all levels was not statistically significant, $t(56.609) = 1.788$, $p = .079$. The corresponding medium effect size $r = .231$ suggests however that the additional years of tertiary studies that year four students had over year one students might have had a moderate bearing on lexical transfer error decline.

Qualitative Research Results

As presented in the Results of Quantitative Research, among the three error types, those caused by Chinese words polysemy account for exactly one half of all lexical transfer errors identified in the corpus. Here are some examples based on the polysemous verb ‘看(kan),’ which with the frequency of 1,811 in the ToRCH Corpus is one of the ten most frequent Chinese words and has several translation equivalents in English: look, see, watch, view, judge, assess, decide:

- 1) I would like to see fiction
*I would like to read fiction
我想读小说
- 2) see herself in the mirror
*looks at herself in the mirror
她看镜子里的自己

- 3) everything she looked
 *everything she sees
 她看到的任何一个东西
- 4) it does not only see whether the translation is faithful to the original text
 *it does not only decide whether the translation is faithful to the original text
 它不仅仅看翻译是否与原文相符合
- 5) She see Berry
 *She visits Berry
 她看望 Berry

A medium frequency word is ‘认识(ren shi),’ which can be used both as a verb and a noun. Its English translation equivalents include: know, realize, acquaint oneself with, be familiar with, recognition, understanding. It has a frequency of 360 in the ToRCH Corpus and has caused two errors in the present study.

- 1) have his/her own recognition of ...
 *[sb.] has his/her own understanding of
 [某人]对...有自己的认识
- 2) they have the ability to realize these western goods.
 * They have ability to get to know/get familiar with these western goods.
 他们有能力去认识这些西方的商品。

Chinese collocations caused over one fifth of the instances of lexical transfer. Three types have been identified: verb + noun, adjective + noun and noun + noun. An example of the first is ‘学习(xue xi)’ and ‘知识(zhi shi),’ which always collocate with each other to make up ‘学习知识(xue xi zhi shi),’ whose English word-for-word translation could be ‘learn knowledge.’ However, the correct English collocation should be ‘gain knowledge’ or ‘acquire knowledge.’ Here are two instances of this transfer type in the corpus:

Verb + Noun Collocations

- 1) The most important task for student in university should be learning
knowledge.
 *The most important task for students in university should be gaining
knowledge.
 大学生的主要任务是学习知识。
- 2) to learn the real practical knowledge
 *to gain the practical knowledge
 去学习知识

Another collocation type being discussed here is adjective + noun collocations.

Adjective + Noun Collocation

- black sky
 *dark sky.
 天黑了(tian hei le)

Finally, errors caused by Chinese noun + noun collocations are discussed here.

Noun + Noun Collocations

- 1) tool books
 *reference books
 工具书
- 2) development space
 *room for improvement
 发展空间

In the present study, close to one third of transfer errors were caused by Chinese MWU. Some typical examples are discussed in this section. Many English MWU that indicate the author's point of view were negatively transferred from the Chinese ones. For instance:

- 1) standing in perspective of...
*walk in the shoes of.../ from the point of view of.../from the perspective of .../in the perspective of...
站在...的角度
- 2) from this point to consider
*from this point of view
从这点考虑
- 3) from this point to see
*from this point of view
从这点看

Another type of transfer error found appeared to be caused by the difference in the order of language elements in MWU of Chinese and English.

- 1) have some extent impact
*have impact to some extent
在某种程度上有影响
- 2) have some degree of influence
*have some influence to some degree
在某种程度上有影响
- 3) In the present day world
*In the world today/in the contemporary world
在当今世界
- 4) I not only can
* I can not only
我不仅能够

Discussion

The results show that Chinese word polysemy caused the most transfer errors, followed by Chinese MWU and Chinese collocations. The frequency of transfer errors in the three categories was lower in the writing of year four students than in the writing of year one students. This indicates that while writing in English, more advanced learners tend to make fewer connections to the Chinese lexical network. This finding is consistent with Kroll and Stewart's (1994) as well as Jiang's (2002) argument that in the minds of learners who are less proficient, L2 words are directly connected to their L1 equivalents. This finding, however, appears to counter Pienemann et al. (2005) view that more advanced learners are more prone to L1 transfer than the less advanced ones.

Despite the fact that final year students made fewer transfer errors, the difference in error frequency between the two groups was not statistically significant. As statistical significance is dependent on sample size, a large enough sample size would yield a small enough p value so that the desired level of statistical significance could be achieved. This was not the case in the current study. Therefore it is more noteworthy that the additional years of English-medium instruction seem to have had a moderate impact on the decline in the number lexical transfer errors.

However, the more advanced learners persisted with lexical transfer. This could be partly explained by 'fossilization,' which is a feature of L2 interlanguage (Yip, 1995). Moreover, as argued by Han and Selinker (1999, pp. 248), 'there is a greater tendency for interlanguage structures to stabilize, leading to possible fossilization in spite of repeated pedagogical intervention.' In addition, the finding is consistent with Jiang's

(2000) L2 processing model, which stipulates that the transition from the L1 lemma mediation stage to the final stage could hardly happen due to the cessation of lexical development, or, more specifically, due to fossilization.

Lexical Transfer Errors Caused by Chinese Word Polysemy

The errors caused by L1 word polysemy make up 50% of the identified lexical transfer errors. One of the underlying reasons for this could be that in Chinese, many high-frequency words are used in different contexts to convey similar senses and meanings without any change in word form. L2 learners are likely to assume that the situation is the same in English, which might cause overgeneralization and hence lexical errors. Thus, for example a high frequency Chinese verb ‘看(kan), whose meaning can be ‘look,’ ‘see,’ ‘watch,’ ‘read’ and ‘visit,’ was found to have caused 5 instances of lexical errors in the current study. This finding supports the study conducted by Yang, Ma and Cao (2013), who also concluded that ‘看(kan)’ is a typical Chinese word that causes negative transfer.

Based on the rest of the results, it can also be concluded that high frequency Chinese words are more polysemous than the low frequency words, thus causing a significant number of transfer errors. Translation as an L2 instruction method, relying on the most common L1 lexical inventory might have been a precipitating factor. In fact, literature (Edmunds, 2013) suggests that grammar-translation method is the preferred English teaching approach in China. Hence, there seems to be a link between Chinese word polysemy caused transfer errors and the excessive use of translation in L2 instruction and learning.

Although the students were not exposed to the grammar-translation method at the English-medium university where this study was conducted, the six years of exposure to this approach in primary and secondary school might have established the habit of linking L2 words not with the concept, but with the equivalent L1 word (Cook, 2014). Research evidence from Dodigovic (2014) shows that students tend to study words using bilingual dictionaries and write Chinese equivalents of unknown English words on the margins. Some confessed that when tasked with writing an essay in English they first compose it in Chinese and then translate into English. Thus, the neural networks established in the process (Lightbown & Spada, 2011) might have been too strong for the new instructional context to re-wire, especially as the students remained immersed in Chinese in their daily lives.

Lexical Transfer Errors Caused by Chinese Collocations

Chinese collocations caused 22% instances of lexical transfer and have been classified into three categories: verb + noun, adjective + noun and noun + noun. The first example represents the verb + noun type. Thus, in Chinese, ‘学习(xue xi)’ and ‘知识(zhi shi)’ always collocate with each other to make the collocation ‘学习知识(xue xi zhi shi), whose English word-for-word translation could be ‘learn knowledge.’ However, the correct English collocation should be ‘gain knowledge.’

Apart from the errors found in the present study, ‘学习(xue xi)’ has been found to cause other errors while collocating with ‘技能(ji neng),’ ‘文化(wen hua),’ and ‘经验(jing yan),’ whose literal English translations are ‘to learn skills,’ ‘to learn culture,’ and ‘to learn experience’ respectively (Li, 2005). However, the correct English collocations should be ‘to master skills,’ ‘to acquire education,’ and ‘to gain experience.’

The second category of collocations exhibiting signs of transfer is the adjective + noun type. Chinese people usually say ‘天黑了(tian hei le),’ whose literal English translation could be ‘sky has become black.’ ‘黑(hei)’ is used to refer to the dark color of the sky. In English however, ‘sky’ is usually collocated with ‘dark.’ Similarly, when describing a person who has blond hair, Chinese people are likely to say ‘golden hair’ rather than ‘blond hair.’ One probable explanation could be that the Chinese collocation ‘金发(jin fa)’ is directly transferred into English. Another reason might be the learners’ lack of English vocabulary knowledge or awareness of cultural difference.

The third and final category of transfer affected collocations is the noun + noun type. For instance, instead of saying ‘reference books,’ a learner apparently separated the two elements in the Chinese phrase ‘工具书(gong ju shu)’ into ‘工具(gong ju)’ and ‘书(shu)’ and then literally translated them into English, making the incorrect English collocation ‘tool books.’

Based on the discussion above, it appears that the findings of the present study are in agreement with those made by Yamashita and Jiang (2010). They concluded that L2 collocations which are not congruent with L1 collocations are more likely to cause negative transfer. In other words, the L2 collocations that cannot be accurately represented through word-for-word translation from L1 would lead to transfer errors. This in turn points to translation in the English language classroom as one of the likely precipitating factors in the case of Chinese collocation transfer. The other one is possible lack of attention to collocations as such.

Lexical Transfer Errors Caused by Chinese MWU

In the present study, 28% transfer errors were found to have been caused by Chinese MWU. Some typical examples are discussed in this section. Many English MWU that indicate the author's point of view were negatively transferred from their Chinese equivalents. For instance, the Chinese MWU '站在...的角度(zhan zai...de jiao du)' was expressed as 'stands in perspective of...' in English, in which '站在(zhan zai)' was literally translated as 'stand.' However, compared with Chinese, which prefers the use of concrete language, English usually uses more abstract expressions, such as 'from the point view of...', 'from the perspective of...', and 'in the perspective of...'

Another type of transfer error appeared to be caused by the difference between Chinese and English word order within MWU. For instance, 'have some extent impact' is a literal translation from Chinese 在某种程度上有影响 with the word order unchanged. In Chinese, '某种程度上(mou zhong cheng du shang),' meaning 'to some extent/to some degree,' is used to express the scale or range, and is usually put prior to the thing being described. Thus, the appropriate English equivalent of the MWU 在某种程度上有影响 is 'have impact to some extent.'

In the case of Chinese MWU transfer, grammar-translation methodology again seems to be a probable cause. This is not only likely due to the reliance on L1 translation to decipher the L2 meaning, but also to the underlying belief that language is lexicalized grammar (Schmitt, 2000). Such a belief does not take into consideration the lexical constraints which govern the proximity of words in connected discourse, but assumes that any word, given that it is the required part of speech, can fill any syntactic slot (Thornbury, 2002).

Conclusions

Intrigued by the role of L1 in L2 vocabulary acquisition, and the paucity of corpus-based research focusing on L1 lexical transfer in Chinese contexts, the present study attempted to explore the lexical transfer errors caused by 1) Chinese word polysemy, 2) Chinese collocations, and 3) Chinese MWU. A learner corpus containing 100 writing samples by 100 Chinese learners of English who were at the time studying at a Sino-British university in China was compiled and manually analyzed. The results show that the majority of lexical transfer errors could be attributed to Chinese word polysemy. Although less advanced learners made overall more lexical transfer errors than the more advanced ones, the difference was not statistically significant.

The fact that more advanced learners did not significantly outperform the less advanced ones could be explained by fossilization. Two possible underlying reasons for this were considered. The first reason could be the Chinese learners' lack of adequate depth of English vocabulary knowledge, due to the lack of extensive exposure to English and the lack of awareness of the lexical features of English vocabulary. The second reason is the over-reliance on the Chinese conceptual network while learning English, which is exacerbated by the grammar-translation approach to English instruction.

Implications for Teaching Practice

The findings from the present study have several pedagogical implications in terms of L2 vocabulary teaching and learning. First of all, the focus of English vocabulary teaching should shift from size to depth, which means that learners should not be encouraged to merely memorize the meanings and forms of as many English words as possible, but should be able to understand the concepts and meaning senses represented by those words, as

well as the associated registers and contexts in which they can be used. All of the above also require English teachers to have in-depth English vocabulary knowledge.

Secondly, teachers should make the learners aware of the fact that there are no exact overlaps between translation equivalents across languages. Moreover, in order to reduce the negative transfer from L1, the use of bilingual dictionaries should decrease, especially for intermediate or advanced L2 learners. In contrast, the use of monolingual learners' English dictionaries should be encouraged since they could provide L2 learners with more accurate and in-depth lexical knowledge, and offer them the contexts in which the words are used.

Thirdly, as argued by Ellis (2008), production could facilitate acquisition only if the learner is pushed, so teachers should require learners to produce L2 as frequently as possible. For instance, learners should be encouraged to try to think in English while writing English papers. In this manner, the role played by Chinese could be reduced, thus preventing negative L1 transfer.

Fourthly, different approaches to teaching L2 lexis should be employed with learners at different levels. Novice L2 learners most likely correspond to the initial stage in Jiang's (2000) L2 lexical processing model. L2 learners at this stage are hardly able to establish a direct connection between the concept and the L2 word. Instead, they connect the L2 word with its L1 translation equivalent. Therefore, Jiang (2000) suggests that an interlingual teaching approach, namely the use of L1 translation, could be used in moderation to help the novice L2 learners establish the forms and core senses of L2 words. However, lexical teaching strategies should change with intermediate or advanced learners, who are already at the L1 mediation stage. In order to help intermediate or advanced learners overcome the lexical or semantic fossilization, which leads to subordinate bilingualism, the use of L1 equivalents should be avoided, and authentic and contextualized L2 materials should be used.

In addition, as suggested by Shalaby, Yahya and El-Komi (2009), word lists containing L2 words that are difficult to acquire could be very helpful in L2 teaching. This is in particular the case with the multiple English equivalents of the high-frequency polysemous Chinese words.

Similarly, since L2 collocations which are not congruent with L1 are found to cause transfer errors, lists of English collocations that cannot be directly deduced from their L1 translation equivalents should be generated. Furthermore, English collocations should be taught as unified wholes rather than as separate words. This is especially important for beginners who are vulnerable to the negative influence of L1 collocations. Finally, learners should be made aware of MWU, especially the ones that do not translate to English word for word.

Teachers and learners could turn to English language corpora for help concerning many aspects of vocabulary, in particular collocation and context of use. The Compleat Lexical Tutor available at <http://www.lextutor.ca> is a website enabling access to several corpora and analytical tools, which could be successfully used for this purpose.

Limitations and Desiderata for Further Research

Although the present study has yielded findings that have important pedagogical implications, it is not without limitations with regard to its methodology. The first limitation concerns the learner corpus used in this study. The 100 texts contained in the learner corpus might lack representativeness, as they were written by the students from the same department at only one university. Therefore, studies that involve more learners who have different educational backgrounds might yield more generalisable results. Although the error count was normed, the considerable difference in word length might have still led to some potential problems. In future research, variables such as the length and topic of texts should be constrained to increase the validity of the findings. Furthermore, sample size needs to be enlarged to include texts written by a larger number of learners from a variety of backgrounds.

Finally, the present study is a cross-sectional one as it explores lexical transfer by examining data from 100 different learners at the same point in time. The lexical fossilization phenomenon discovered in the present study might be further investigated by approaching the issue from a developmental perspective, which calls for longitudinal studies. Considering the current dearth of studies focusing on lexical transfer, any further research in this area would help illuminate the issues at hand.

References

- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach, *English for specific purposes*, 31(2), 81-92.
- Bialystok, E., Craik, F. I., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia, *Neuropsychologia*, 45(2), 459-464.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (p. 1204) Harlow: Pearson Education Limited.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing, *International Journal of Corpus Linguistics*, 14(3), 275-311.
- Bogaards, P. (2001). Lexical units and the learning of foreign language vocabulary, *Studies in second language acquisition*, 23(03), 321-343.
- Boers, F., & Lindstromberg, S. (2009) *Optimizing a lexical approach to instructed second language acquisition*. Basingstoke: Palgrave Macmillan.
- Carroll, D. (2007). *Psychology of language*. Belmont: Thomson Higher Education.
- Chan, A. Y. W. (2010). Toward a Taxonomy of Written Errors: Investigation Into the Written Errors of Hong Kong Cantonese ESL Learners, *TESOL Quarterly*, 44 (2), 295 – 319.
- Chen, M.-H. & Lin, M. (2011). Factors and Analysis of Common Miscolllocations of College Students in Taiwan. *Studies in English Language and Literature*, 28, 57 – 72.
- Cook, V. (2014). How Do Different Languages Connect in Our Minds? In Cook, V. & Singleton, D. (Eds.) *Key Topics in Second Language Acquisition*. (pp. 1-16) Bristol: Multilingual Matters.
- Dodigovic, M. & Wang, S. (2015). The misuse of academic English vocabulary in Chinese student writing, *US-China Foreign Language* 13(5), 349-356.
- Dodigovic, M. (2005). *Artificial intelligence in second language learning: Raising error awareness*. Clevedon: Multilingual Matters.
- Duan, M. & Qin, X. (2012). Collocation in English Teaching and Learning, *Theory and Practice in Language Studies*, 2(9), 1890-1894.
- Edmunds, K. (2013). Chinese ESL Learners' Overuse of the Definite Article: A Corpus Study, BA thesis; Emory University.
- Ellis, R. (2008). *The study of second language acquisition*. 2nd edition. Oxford University Press.
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse*, 20(1), 29-62.
- Folse, K. S. (2006). The Effect of Type of Written Exercise on L2 Vocabulary Retention. *TESOL Quarterly: A Journal For Teachers Of English To Speakers Of Other Languages And Of Standard English As A Second Dialect*, 40(2), 273-293.
- Gass, S. (2013). *Second Language Acquisition: An Introductory Course (4th Edition)*. New York: Routledge/Taylor Francis.
- Gass, S. & L. Selinker (2008). *Second Language Acquisition: An Introductory Course (3rd Edition)*. New York: Routledge/Taylor Francis.
- Gray, B. & Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, 18 (1), 109-135.
- Grosjean, F. (2001). The bilingual's language modes. In Nicol, J. (Ed.). *One Mind, Two Languages: Bilingual Language Processing* (pp. 1-22). Oxford: Blackwell.
- Han, Z. & Odlin, T. (Eds.), *Studies of Fossilization in Second Language Acquisition*. Clevedon, U.K.: Multilingual Matters, 134-156.
- Han, Z., & Selinker, L. (1999). Error resistance: Towards an empirical pedagogy. *Language Teaching Research*, 3(3), 248-275.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation, *English for specific purposes*, 27(1), 4-21.

- Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. New York: Routledge.
- Jiang, N. (2000). Lexical representation and development in a second language, *Applied linguistics*, 21(1), 47-77.
- Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech, *Bilingualism: Language and Cognition*, 9(02), 119-135.
- Kroll, J. F. & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations, *Journal of Memory and Language*, 33, 149 – 174.
- Karabacak, E., & Qin, J. (2013). Comparison of lexical bundles used by Turkish, Chinese, and American university students, *Procedia-Social and Behavioral Sciences*, 70, 622-628.
- Lennon, P. (1996). Getting ‘easy’ verbs wrong at the advanced level, *IRAL-International Review of Applied Linguistics in Language Teaching*, 34(1), 23-36.
- Lewis, M. (1997). *Implementing the lexical approach: putting theory into practice*. Hove: Language Teaching Publications.
- Li, S. (2005). *An Investigation into Lexical Misuses by Chinese College Students under the Negative Influence of Their First Language*. Master’s thesis: Zhejiang University.
- Liu, Z. (2011). Negative Transfer of Chinese to College Students’ English Writing, *Journal of Language Teaching and Research*, 2 (5), 1061-1068.
- Llach, M. P. A. (2011). *Lexical errors and accuracy in foreign language writing*. Claredon: Multilingual Matters.
- MacWhinney, B. (1992). Transfer and competition in second language learning, *Advances in Psychology*, 83, 371–390.
- Morimoto, S., & Loewen, S. (2007). A comparison of the effects of image-schema-based instruction and translation-based instruction on the acquisition of L2 polysemous words, *Language Teaching Research*, 11(3), 347-372.
- Nation, I. S. P. (2013). *Learning vocabulary in another language. 2nd edition*. Cambridge: Cambridge University Press.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins Publishing.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching, *Applied linguistics*, 24(2), 223-242.
- Pawley, A. & Syder, F.H. (1983). Two puzzles for linguistic theory: Native-like selection and native-like fluency. In J. Richards & R. Schmidt (Eds.) *Language and Communication*. (pp. 121 – 225) London: Longman.
- Pienemann, M., Di Biase, B., Kawaguchi, S., & Håkansson, G. (2005). Processing constraints on L1 transfer. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 128-153) Oxford: Oxford University Press.
- Paquot, M. (2013). Cross-linguistic influence and formulaic language: French EFL learners’ use of recurrent word sequences under scrutiny, *Learner Corpus Research* 18(3), 391-417.
- Peromingo, J. P. (2012). Corpus analysis and phraseology: Transfer of multi-word units, *Linguistics and the Human Sciences*, 6(1-3), 321-343.
- Rafee, M., Tavakoli, M. & Amirian, Z. (2011). Structural Analysis of Lexical Bundles Across Two Types of English News Papers Edited by Native and Non-native speakers, *The Modern Journal of Applied Linguistics*, 3 (2), 1 – 15.
- Schmitt, N. (2000) *Vocabulary in language teaching*. Ernst Klett Sprachen.
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition*, 19, 17 – 36.
- Shalaby, N. A, Yahya, N., & KI-Komi, M. (2009) ‘Analysis of lexical errors in Saudi college students’ compositions,’ *Ayn, Journal of the Saudi Association of Language and Translation*, 2(3), pp. 65-92.
- Schmitt, N., & Hemchua, S. (2006). An analysis of lexical errors in the English composition of Thai learners, *Prospect: an Australian journal of TESOL*, 21(3), 3-25.
- Thornbury, S. (2002) *How to teach vocabulary*. Harlow : Longman.
- Wang, D. (2011). Language Transfer and the Acquisition of English Light Verb + Noun Collocations by Chinese Learners, *Chinese Journal of Applied Linguistics*, 34 (2), 107 – 125.

- Wray, A. (2002) *Formulaic sequences and the lexicon*. New York: Cambridge University Press.
- Wolter, B. (2006). Lexical network structures and L2 vocabulary acquisition: The role of L1 lexical/conceptual knowledge, *Applied Linguistics*, 27(4), 741-747.
- Xia, L. (2013). A Corpus-Driven Investigation of Chinese English Learners' Performance of Verb-Noun Collocation: A Case Study of Ability, *English Language Teaching*, 6 (8), 119 – 124.
- Yip, V. (1995). *Interlanguage and learnability: from Chinese to English*. Philadelphia: John Benjamins Publishing.
- Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations, *TESOL Quarterly*, 44(4), pp. 647-668.
- Yang, L., Ma, A. P., & Cao, Y. (2013). Lexical Negative Transfer Analysis and Pedagogic Suggestions of Native Language in Chinese EFL Writing. *The proceedings of the 2013 Conference on Education Technology and Management Science (ICETMS 2013)*. (pp. 669 – 672) Atlantis Press.
- Zhou, S. (2010). Comparing receptive and productive academic vocabulary knowledge of Chinese EFL learners. *Asian Social Sciences*, 6(10), 14–19.

About the authors

Marina Dodigovic, PhD, is a professor at the American University of Armenia. Previously, she served as the director of both the MA TESOL program and the Research Centre for Language Technology at Xi'an Jiaotong-Liverpool University. Her research interests have gravitated toward vocabulary teaching, learning and assessment.

Chengcheng Ma, MA in TESOL, is a graduate from Xi'an Jiaotong-Liverpool University and University of Liverpool. She served as a research assistant at Research Centre for Language Technology at Xi'an Jiaotong-Liverpool University. Currently, she is teaching English in Kunming, PR China.

Song Jing is a graduate from Xi'an Jiaotong-Liverpool University. He served as a research assistant on the project titled *Lexical Transfer from Chinese to English in the Writing of XJTLU Students* at Xi'an Jiaotong-Liverpool University. At present, he is pursuing an MA degree at the Australian National University.