The Misuse of Academic English Vocabulary in Chinese Student Writing

Marina Dodigovic
Xi’an Jiaotong-Liverpool University, Suzhou, China

Sisi Wang
Jiangsu Telecommunication Services Corporation Limited, Suzhou, China

Academic vocabulary, otherwise known as Academic Word List or AWL is a part of core vocabulary in English medium education, crucial to both understanding of textbooks and academic writing. While much has been written about this group of words, little is known about how Chinese learners studying at English medium universities use these words. This paper focuses on the AWL related errors found in a corpus of final year projects of Chinese students at an English medium university. In particular, it discusses the most common error type, which pertains to word form, its possible causes and implications for the teaching practice. Instruction, in addition to the first language transfer seems to be among the most likely causes. Focus on contextualized attention to word form as part and parcel of vocabulary learning is indicated for the teaching practice.

Keywords: Academic Word List, word form, lexical error, vocabulary learning

Introduction

Lexical approaches to second and foreign language teaching (Richards & Rodgers, 2001) have fuelled a renewed interest in vocabulary in both English language teaching (ELT) and research. The notion of core vocabulary has become a part of ELT needs analysis, as well the teaching practice itself. Thus, it has become unthinkable to discuss ELT lexis without referring the 2,570 most useful words, consisting of 2,000 of the most frequent general words and 570 academic words shared across the academic disciplines (Nation, 2006). In particular the academic words, known as Academic Word List (AWL) (Coxhead, 1998, 2000, 2002) seem to be of most use to non-English-speaking background (NESB) students at English medium universities. While much research has revolved around vocabulary learning and teaching strategies (Cobb & Horst, 2001; Hiebert & Lubliner, 2008; Nagy & Townsend, 2012), few studies have focused on the productive vocabulary knowledge (McCarthy & O’Dell, 2008; Mehrpour & Rahimi, 2010). This study is an effort in this direction.

Literature Review

While research focus has predominantly been on the receptive command of AWL, which enables learners to read and listen with comprehension, not much is known about the productive command of academic vocabulary, which enables them to speak and write proficiently. Studies such as those by Augustin Llach (2011), Storch and Tapper (2009), and Deng, Lee, Varaprasad, and Leng (2010) tracked the development of
academic vocabulary in the writing of ESL (English as a Second Language) students over the duration of an academic English course finding the evidence of significant improvement. However, lexical errors in the area of academic vocabulary have largely remained unexplored.

The lack of accuracy, otherwise known as language error is significant in three respects: It informs the teacher about what should be taught; it informs the researcher about the course of learning; it is an outcome of the learner’s target language hypothesis testing (James, 1998). The sources of error are deemed to be the redundancy of code (intralingual), various sources of interference (interlingual), and unsuitable presentation (George, 1972). Similarly, James (1998) identified three main sources of L2 errors: (1) interlingual (transfer), (2) intralingual (developmental), and (3) induced (through instruction).

The most common approaches to learner error are anchored in the paradigms of error analysis (EA) and transfer analysis (TA). While EA compares the learner language with the target language (L2), TA compares the learner language with the first language (L1) (Chan, 2010). A number of small-scale studies have analyzed the errors of Chinese speakers in Hong Kong (Chan, 2010), Singapore (Zhu, 2007), and mainland China (Liu, 2012). Xiao (2007) pointed out the importance of contrastive corpus linguistics for EA, i.e., TA as its subprocedure according to Chan (2010). In fact, a number of error focused studies concerning Chinese learners of English have been conducted on learner corpora (Chan, 2010; Liu, 2012; Edmunds, 2013).

The interlanguage of Chinese learners of English is sometimes called “Chinglish” (W. Li, 2014). Its description is frequently based on the ways in which it deviates from the standard TL (Target Language). Thus, Chang (1987) identified verb form error as one of its most common features, including time, tense, and aspect. These were seen as negative transfer from Chinese, which is not an inflected language (Chang, 1987). Negative transfer occurs when learners introduce features unique to their first language into their second language (Selinker, 1972). Thus, Chinese learners for example, may show a tendency toward not inflecting words in English, in cases in which inflection would be required.

This could be explained by transfer from Chinese (Crompton, 2011), as with the exception of some nouns that can under circumstances take a plural form (e.g., “pengyoumen” = friends, used as address form), Chinese nouns generally do not have a plural form. Instead, numerals and counting words are used to express the meaning of plural (Cook & Singleton, 2014). This would make noun-related form errors potentially transfer errors. Similarly, verbs in Chinese are also not inflected. To express the past, the particle “le” is added at the end of the sentence, in addition to the inclusion of time adverbials, which contribute to disambiguation. Sometimes, the particle “guo” can alternatively be used, especially in negative statements (e.g., “Wo meizouguozuo ye.” = I haven’t done my homework [yet].) The lack of verb inflection in Chinese would suggest transfer as the source of verb form errors (Liu, 2012). If this were the case though, transfer would possibly result in rule relaxation in English, leading to the use of predominantly uninflected forms.

Intralingual or developmental errors, which are sometimes called redundancy errors (James, 1998), could have a similar effect, leading to either overgeneralization or undergeneralization (Han & Odlin, 2006). In fact, such errors have been found in Thai medical students’ writing (Sattayatham & Honsa, 2007), suggesting that Chinese as L1 may not be responsible for them. However, Dulay and Burt (1974) found that in case of developmental errors, grammatical morphemes were missing altogether.

Induced errors according to James (1998) might have something to do with language instruction or learning environment. Thus various sources report the overuse of certain linguistic features in the language of Chinese learners of English. Field (1994) for example found that students in Hong Kong overused certain
connectives, as they were explicitly taught to use them. Edmunds found that Chinese students overused the definite article, which is partly explained by previous instruction. On the other hand, the lack of explicit instruction might have resulted in the underuse of features such as hedging devices (Chan, 2010) or nominalization in academic English (Dodigovic, 2013). Hoey (2005) found that instruction can cause cracks in lexical or grammatical priming. Priming means that a word or a morpheme is learned through encounters, the knowledge of it being the cumulative effect of all of these encounters (Hoey, 2005). Such education induced cracks have the potential to destabilize the knowledge of aspects of language.

The above cases mostly pertain to word form error, as Folse (2004) identified word form as one of the fundamental aspects of word knowledge, which makes it relevant to research in lexical errors. Other areas in which errors might be identified can be derived from the concept of word knowledge depth. Depth concerns everything a learner knows about a word, including the possible contexts in which it can occur, other words that often accompany it, the idiomatic expressions in which it is used, the connotations that it can have, the spelling conventions and variations, the pronunciation and its variations, the sentence structure required to accommodate this word, its part of speech, and possible functions in an utterance (Folse, 2004).

Not all errors have the same gravity. Thus, those errors that interfere with comprehension are deemed to have more gravity (Chan, 2010). However, those that appear more frequently have according to Chan (2010) more prevalence. Interlanguage theory differentiates between errors based on their resistance (Han & Selinker, 1999). Thus those errors which defy correction and persist despite repeated instruction are called fossilized errors (Gass & Selinker, 2008). Fossilisation (Han & Odlin, 2006) 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 error fossilisation. James (1998) further elaborated on this concept by distinguishing between a slip, an odd mistake or a systemic error. A slip is expected to result in self-correction, a mistake calls for feedback, while error requires full correction of the erroneous utterance.

**Research Questions**

In light of the literature review, this study seeks to answer the following questions:

1. What is the most frequent type of academic vocabulary errors?
2. What are the possible causes of the most frequent errors?
3. What are the implications of these errors for the teaching practice?

**Methodology**

**Data Source**

The source of data was a 453,801-word corpus of 60 final year projects of year four Chinese students studying English, Communication and Business at an English medium university in China.

**Error Classification**

Errors were classified based on the aspects of word knowledge depth. Consequently, six types of errors were included in the classification: word form, part of speech, context, structure, collocation, and spelling.

**Results**

There were 233 instances of AWL related errors involving 152 different academic words. The 233 errors
were sorted on error type to identify the most frequent errors types committed by Chinese learners, with the word form leading at approximately 41% of all AWL related errors. Therefore, form errors are further discussed and analyzed in this study. Table 1 below contains the breakdown of errors by category and frequency.

Table 1
Error Breakdown by Category and Frequency

<table>
<thead>
<tr>
<th>Error type</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word form</td>
<td>41</td>
</tr>
<tr>
<td>Part of speech</td>
<td>18</td>
</tr>
<tr>
<td>Context</td>
<td>15</td>
</tr>
<tr>
<td>Structure in which it is used</td>
<td>13</td>
</tr>
<tr>
<td>Collocation</td>
<td>10</td>
</tr>
<tr>
<td>Spelling</td>
<td>2</td>
</tr>
</tbody>
</table>

Discussion

Word Form Error Classification

Nine error sub-categories have been identified, namely noun number, verb 3rd person singular, preposition + V-ing, be + other than -ing, to + inflected verb, passive (inf. Instead of pp), modal + inflected verb, have + V (present participle), and past tense (S. Wang, 2014) (see Table 2). This way of categorization is based on morphology, more specifically on inflectional errors. The discussion of each group is found below. Table 3 includes a list of examples by sub-category.

Table 2
Ranking of the Frequency of Errors

<table>
<thead>
<tr>
<th>Sub-categories</th>
<th>Frequency</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun number</td>
<td>34</td>
<td>35.8</td>
</tr>
<tr>
<td>Verb 3rd person sg</td>
<td>23</td>
<td>24.2</td>
</tr>
<tr>
<td>Prep. + V-ing</td>
<td>11</td>
<td>11.6</td>
</tr>
<tr>
<td>Be + other than -ing</td>
<td>8</td>
<td>8.4</td>
</tr>
<tr>
<td>To + inflected verb</td>
<td>8</td>
<td>8.4</td>
</tr>
<tr>
<td>Passive (inf. Instead of pp)</td>
<td>5</td>
<td>5.3</td>
</tr>
<tr>
<td>Modal + inflected verb</td>
<td>4</td>
<td>4.2</td>
</tr>
<tr>
<td>Have + V (present participle)</td>
<td>1</td>
<td>1.05</td>
</tr>
<tr>
<td>Past tense</td>
<td>1</td>
<td>1.05</td>
</tr>
</tbody>
</table>

Table 3
Examples of Word Form Categories

Noun number
E.g. Thus, not only adult but also the children don’t like to watch domestic animation.

Verb 3rd person sg
E.g. Thus, the results of this group indicates that the Chinese stock markets are not semi-strong efficient.

Prep. + V-ing
E.g. After compared and contrast the variances of two companies in the taste part…

Be + other than -ing
E.g. Newspapers in the UK are targeted particular social classes and political affiliations.
### Commentary

While approximately 36% of all form related errors are noun-related, a cumulative total of 64% are connected with the verb, which conforms with the findings of previous research (Liu, 2012; Dodigovic, 2013). The presence of inflectional morphemes which are not always misplaced in this study, sometimes found in the same sentence in which an error was made, suggests that the students have been instructed in grammar and are at least aware of the existence of these morphemes. When they misuse them, it is without a clear pattern. It is therefore possible that these errors could have been induced, for instance through misunderstood instruction. Thus, Xu claimed that 70% of English teachers in China were still using grammar-translation method (as cited in Edmunds, 2013). This method however, might have resulted in better command of grammatical morphemes. On the other hand, C. Li (2014) found that the students belonging to the same population as those in this study were mainly exposed to communicative language teaching, where form is not the focus of instruction. Similarly, deliberate vocabulary learning also did not seem to have been encouraged, whether contextualized or decontextualized. Literature suggests that incidental learning may result in only partial mastery of words (Schmitt, 2010). In addition, according to Wei (2007), Chinese students have been found to underutilize contextualized activation and management as a vocabulary learning strategy, which also may contribute to the lack of accuracy in contextualized inflection. Extensive reading and communicative practice might have therefore led to the noticing of the entire range of inflectional morphemes, but may not have been sufficient to make successful hypotheses about their use (Swain, 1998).

On the other hand, each student might have arrived at a different conclusion regarding how English inflection works. It is also possible that early instruction, although form focused, was not effective as it did not provide room for the meaningful application of the grammar rules learned by rote (Cook & Singleton, 2014). The rather careless, almost whimsical choice of inflectional affixes, in addition to the awareness of these forms, may reflect the attitude that these are less important for successful communication (Y. Wang, 2014). Moreover, it can signal that the new style of instruction might have caused cracks in the colligational or grammatical priming for theses students. According to Hoey (2005), such cracks can become codified, leading to long term grammatical insecurity. Given the evidence of almost random use of grammatical affixes, this is a likely scenario for the subjects in this study. Cognitive language learning theories rooted in psychology would possibly identify inflectional affixation as an area of lower mental activity, pointing to the lack of problem-solving orientation in learning and possibly teaching (Ellis, 2001). Thus, rote learning of rules, decoupled from the context of use, might have also contributed to the formation of these errors.

<table>
<thead>
<tr>
<th>Table 3 (continued)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun number</td>
</tr>
<tr>
<td>To + inflected verb</td>
</tr>
<tr>
<td>E.g. Moreover, the Chinese financial analysts’ ability to analyzing and predicating future…</td>
</tr>
<tr>
<td>Passive (inf. instead of pp)</td>
</tr>
<tr>
<td>E.g. …the null hypothesis should be reject.</td>
</tr>
<tr>
<td>Modal verb + V. (inflected)</td>
</tr>
<tr>
<td>E.g. Translators could also involved tension.</td>
</tr>
<tr>
<td>Have + V (present participle)</td>
</tr>
<tr>
<td>…Chinese regions have involving…</td>
</tr>
<tr>
<td>Tense mismatch</td>
</tr>
<tr>
<td>This qualitative study shows support and supplements to such argument, and also confirmed…</td>
</tr>
</tbody>
</table>
Pedagogical Implications

With the advent of English medium universities in China, which often facilitate educational migration to English speaking countries, found to be prestigious by major stakeholders (C. Li, 2014), English instruction would need to be able to address both fluency and accuracy from the very beginning. Extensive reading, including the use of graded readers (Nation, 2006), may be one way of compensating for the lack of opportunities for genuine meaningful communication in the target language. Watching movies, TV series, listening to the news or popular songs with lyrics is another (Nation, 2006). While the awareness of morphology needs to be raised in context (Cook & Singleton, 2014; Wei, 2007), priming for colligation (Hoey, 2005) would also be important through exposure to rich linguistic data as evidenced in the above readings or audio-visual materials. Therefore, the word form itself needs to be learned in its context, as an aspect of the word itself, rather than as a part of an abstract rule. In light of cognitivist and interactionist language learning theories, two ingredients might be necessary toward the improvement of learning outcomes: (1) opportunities for problem solving and (2) feedback. While inflection practice needs to be embedded in problem solving tasks, rather than in mechanical transformation drills, feedback strategies could be naturalistic, including repetitions, confirmation checks, comprehension checks, and clarification requests.

Error correction as a form of feedback might also be useful, since it has a twofold purpose. One is Consciousness Raising (CR) or explanation and the other is Language Awareness (LA) or explication (James, 1998). According to James (1998), the former means helping learners to notice what they do not know, while the latter means providing explicit insight into something that the learners already know implicitly. CR occurs when errors are identified as such. By offering an array of reactions, the teachers need to try to decide whether this is a slip, an odd mistake or a systemic error (James, 1998). A slip is expected to result in self-correction, a mistake calls for feedback, in this case a clue to the required form, while error calls for full correction of the erroneous word form. Since the errors found in this study are likely to be genuine errors, as the corpus includes only the edited final drafts of papers, based on feedback received on the first drafts, correction might be an adequate response.

Suggestions for Further Research

It is suggested to expand the sample of the research, including participants from other universities in China in order to obtain a more representative sample. Furthermore, the proficiency in English or the year of study of the participants should not be constrained to the same level. Rather, it could be expanded to include other levels at university in order to get a more comprehensive view of AWL related lexical issues. In addition, other error types as well as other core vocabulary could also be made the subject of further research.

Conclusion

This paper presents the taxonomy of AWL related errors, identified in a corpus of English writing of Chinese university students using word form, part of speech, context, structure, collocation, and spelling for classification purposes. The most frequent of those was found to be the word form type (41%). Nine sub-categories of AWL word form errors were identified and ranked according to frequency of occurrence in the learner corpus. The sub-categories included noun number, verb 3rd person singular, preposition + V-ing, be + other than -ing, to + inflected verb, passive (inf. Instead of pp), modal + inflected verb, have + V (present participle), and the past tense. The analysis indicates that verb inflection might be the most common amongst
the above problems. While this and other word form error types seem to have their roots in L1, as Chinese is an uninflected language, intermittent although not correct use of inflectional affixes suggests sources other than just L1. The lack of regularity in affixation errors excludes intralingual or developmental sources. Thus, instruction as a source of induced errors appears to be the most probable cause. Emphasis on contextualizing vocabulary learning as well as the learning of its forms in order to raise the awareness of colligation and facilitate the priming of words for context-appropriate inflection might be some of the ways in which instruction could be structured to counter the tendency toward making word form errors in Chinese contexts.

References
Coxhead, A. (2002). The academic word list: A corpus-based word list for academic purposes. In B. Kettermann & G. Marks (Eds.), *Teaching and language corpora (TALC) conference proceedings* (pp. 73-89). Atlanta, GA: Rodopli.
Edmunds, K. (2013). *Chinese ESL learners overuse of the definite article: A case study* (Unpublished dissertation). Emory University, Atlanta, GA.


