

The Use of Academic English Vocabulary in the Writing of Chinese Students

By Marina Dodigovic, Hui Li, Yixiao Chen & Danni Guo

Academic Word List (AWL) comprises 570 words, which cover approximately 10% of most academic texts. The success of students of non-English speaking backgrounds in English medium education depends on the command of these words. This study investigated the level of command over AWL of Xi'an Jiaotong-Liverpool University (XJTLU) final year students. Overall, the students' writing was similar in profile to native speaker academic writing, with an approximately 12% AWL coverage. A total of 233 AWL-related errors found in the 453,801-word electronic corpus were mostly related to word form, its part of speech and context, with *analysis*, *focus* and *affect* being the most difficult words.

Introduction

The last decade has witnessed a growing awareness of the importance of academic vocabulary in university studies. It is widely acknowledged nowadays that foreign language students need to acquire a core of high frequency academic vocabulary in order to successfully engage in academic studies, and actively participate in various learning activities at English medium universities. Much research has been conducted in light of the compilation of academic vocabulary lists (Coxhead, 1998; 2000; 2002), vocabulary learning and teaching strategies (Cobb & Horst, 2001; Hiebert & Lubliner, 2008; Nagy and Townsend, 2012), and the usefulness of a generic academic word list across disciplines (Hyland & Tse, 2007; Sutarsyah, Nation & Kennedy, 1994). In contrast, only a handful of studies focuses on how students use those words (McCarthy & Odell, 2008; Mehrpour & Rahimi, 2010). With a purpose of improving our understanding of students' repertoires of academic vocabulary, this study investigates final year university students' dissertations to gauge how well

academic words are used. It is hoped that this research could shed light on the learnability of academic vocabulary in English as a foreign or second language, and the implications of academic vocabulary teaching at university level in particular.

Academic vocabulary is also called 'sub-technical vocabulary' (Cowan, 1974), which refers to words that are frequently used in a wide range of academic disciplines, but not commonly in texts of other genres (Coxhead & Nation, 2001). Following Coxhead's (2000) general academic vocabulary list, this paper tries to answer the following research questions:

1. How well are academic words used in XJTLU students' writing?
2. What salient errors can be identified in the use of academic words by Chinese learners?

Literature review

The list of 570 words, also known as AWL or Academic Word List, compiled by Coxhead

(2000), is a significant milestone in English vocabulary research. The words on this relatively short list cover approximately 10% of most academic texts and are deemed crucial to comprehension of such texts (Nation, 2006) and hence to the success of ESL (English as a Second Language) students in English medium education. Although the need for an ESL pedagogy based on such a list has been called into question as non-discipline specific (Neufeld, Hancioglu, & Eldridge, 2011), Li & Qian (2010) demonstrated that such a pedagogy is not only necessary, but can also be successful, especially in the Chinese context.

While research has mostly focused on the receptive needs of ESL students, e.g. the command of vocabulary sufficient for reading and listening, not much is known about the productive command of AWL, i.e. the use of such vocabulary in speaking and writing. Zhou (2010) noted that the receptive academic vocabulary of Chinese ESL learners was larger than their productive academic vocabulary, which is in line with the general receptive – productive vocabulary ratio of 2.2: 1 (Nation, 2001). Nagy & Townsend (2012) pointed out that productive knowledge of academic texts can enhance comprehension. Moreover, according to Coxhead (2012), productive command of AWL signifies that learners have become members of the academic community. A study by Storch & Tapper (2009) as well as the one by Deng, Lee, Varaprasad & Leng (2010) tracked the development of AWL in the writing of ESL students over the duration of an academic English course finding evidence of significant improvement. Therefore, productive command of AWL in a student population could be an important indicator of both their English learning success and the level of their integration into the international academic community.

For all of the above reasons, it is important to understand the level of productive command of academic vocabulary in an ESL English medium higher education environment, such as XJTLU. Since the use of AWL is more common in written academic prose than it is in spoken academic communication (Zhou, 2010), the writing produced by XJTLU students in their final year would be a good indicator of their English learning and their membership in the international academic community. It was

therefore the aim of this project to profile the vocabulary of final year projects (FYP) of XJTLU students and compare the frequency and accuracy of AWL use with that found in native speaker written academic prose.

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) distinguished 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.

In the language of Chinese English learners, Chang (1987) and Yip (1995) found several types of structural errors. Thus, Chang (1987) identified part of speech confusion in addition to verb form error, 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. Yip (1995) on the other hand claimed that Chinese learners mainly have problems with verb transitivity, as they use pseudo-passives (These sentences can analyse many ways), ergative construction (What is happened with these verbs?), tough movement (Never easy to be learned...) and existential construction (There are sentences cause learnability problems). Yip's (1995) conclusions are less revealing regarding the causes of the above errors.

Most of learner error research has focused on grammar or structure (Dodigovic, 2005; 2013), while relatively little attention has been paid to lexical errors. According to Agustin Llach (2011), despite the fact that lexical errors emerge as the most numerous in the available studies, the research in this area is still scarce. One of the most contentious issues in this kind

of research is the question of error taxonomy (Agustin Llach, 2011). Carrio Pastor (2004, cited in Agustin Llach, 2011), following James (1998), distinguished six lexical error types: 1) confusion of similar words, 2) wrong word formation, 3) lexical distortions, 4) use of a more or less specific word, 5) collocation error, 6) semantic choice error. Hemchua and Schmitt (2006) on the other hand, subdivided lexical errors into two large groups, based on form and meaning. Both of these taxonomies however, fail to account for the impact of lexical error on utterance structure. The taxonomy used in this study therefore relates to what is known about vocabulary command. Thus, proper command of a word requires the learner to know a number of facts about it. These include 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, and its part of speech and possible functions in an utterance (Schmitt, 2001; Folse, 2004; Nation, 2006).

Methodology

Data collection

In this study, 60 academic texts were collected as source for a small learner corpus, which contained the FYP of year four Chinese students studying in the Department of English, Culture and Communication at Xi'an Jiaotong-Liverpool University in China. While some of the participants were majoring in English, others were studying toward double majors in English and Finance or International Business. Prior to inclusion in the corpus, the texts were stripped of identifying information (name, ID number, supervisor, department etc), quotations, citations, page headers, abstracts, table of contents, proper nouns, translation, characters of other languages, references, graphics and appendices. However, commonly used abbreviations (e.g. EAP) were retained, whereas the authors' own creations (e.g. 'CE' for 'College English') were deleted. Finally, all of the edited samples were integrated into the electronic corpus by using the Corpus Builder software found at the

Compleat Lexical Tutor website (www.lextutor.ca).

Data analysis

In the process of analysis, the corpus was profiled and all AWL words were identified using the Vocabprofile component of the Compleat Lexical Tutor software. Each instance of academic word usage was assessed for accuracy by four independent raters. Erroneous uses of AWL vocabulary which were confirmed as such by native speakers were then categorized using a six-type taxonomy. The six types of errors are Context, Collocation, Word Form, Structure, PS (Part of Speech) and spelling. They are explained and exemplified in Table 1.

Results

In the corpus containing 453,801 words or tokens (single instances of word use) there were 12,656 types (different words). Out of the 570 AWL words, 555 were used by student writers with an 11.88% distribution. In total, 233 instances of AWL related errors involving 152 different academic words were identified. The frequency of AWL use in the corpus was found to be 1.88% higher than the expected 10% (Nation, 2006). The 233 errors were sorted both on error type and word to identify the most salient errors types committed by Chinese learners. The results are displayed in Figures 1 and 2 (page 17).

As can be seen in Figure 1, the 233 errors were classified using the six categories: Word Form, PS (Part of Speech), Context, Structure, Collocation and Spelling. Among the six error types, Word Form was found to occur most frequently, with a total of 96 occurrences. Consequently, Word Form error type alone makes up approximately 41% of the total AWL misuse. PS error follows with the raw frequency of 43 (approximately 18.5% of all AWL errors). Furthermore, the frequencies of Context, Structure and Collocation errors are 35, 31, and 24 respectively. Interestingly, only 4 errors were found in the category of Spelling (2%).

The results shown in Figure 2 revealed the most commonly misused academic words. Overall, 152 different words were misused. Seven of these stood out due to their relatively

Table 1: Error taxonomy with examples

Criteria	Explanation	Example	Correction
Context	Wrong word (e.g. “learn” instead of teach)	“writers will allocate the explanation part in an independent paragraph”	Allocate>>>>provide
Collocation	Words used together (e.g. “compelling reason”)	“Researchers study translation in perspective of social-linguistics/pragmatics and functional grammar”	In perspective of>>>>from the perspective/point of view
Word Form	The form of the word (e.g. “put” vs. “puts”)	“how movie titles are affect by cultural factors”	Affect>>>>affected
Structure	Sentence structure required by a particular word (“give something to somebody”)	“there may be other expressions which are polite in Chinese but impolite in English occur in daily life”	Occur>>>>occurring
PS	Part of speech	“it will use questionnaire and case study to analysis the direction of Employment Company”	Analysis>>>> analyze
Spelling	Correct or incorrect	“the brand names can also be used to simulate consumption”	Simuate>>>>stimulate?

frequent occurrence compared with other words. The following is the list of the most frequently misused academic words: analysis (frequency: 14), focus (10), affect (8), analyse/analyze (7), consist (5), emphasize (4), and emphasis (3).

Discussion

Overall, XJTLU students used over 97% of listed academic words in their final year projects, signaling their familiarity with this vocabulary list. The analysis of their AWL use suggests that the students had relatively good productive mastery of this vocabulary group, as the number of erroneous uses does not appear to be proportionally large. On the average, each word was used erroneously less than twice. Lexical error analysis indicates that almost one half (41%) of all

errors are errors of form, which raises the question whether these are real errors of word knowledge or just slips of pen (James, 1998). The fact that these were the final paper drafts would suggest that they were possibly edited and proofread with some care, potentially meaning that any inaccuracies identified are more likely to be errors proper (James, 1998). Another piece of evidence supporting such conclusion is the fact that less than 2% of all AWL related errors were in the spelling category. However, this may be an outcome of concurrent spell-checker use rather than that of careful proofreading and editing.

Majoring in English or having a substantial component of English in their major could have made a difference. It must be noted, however, that XJTLU is an English medium university, with all programmes being delivered in English.

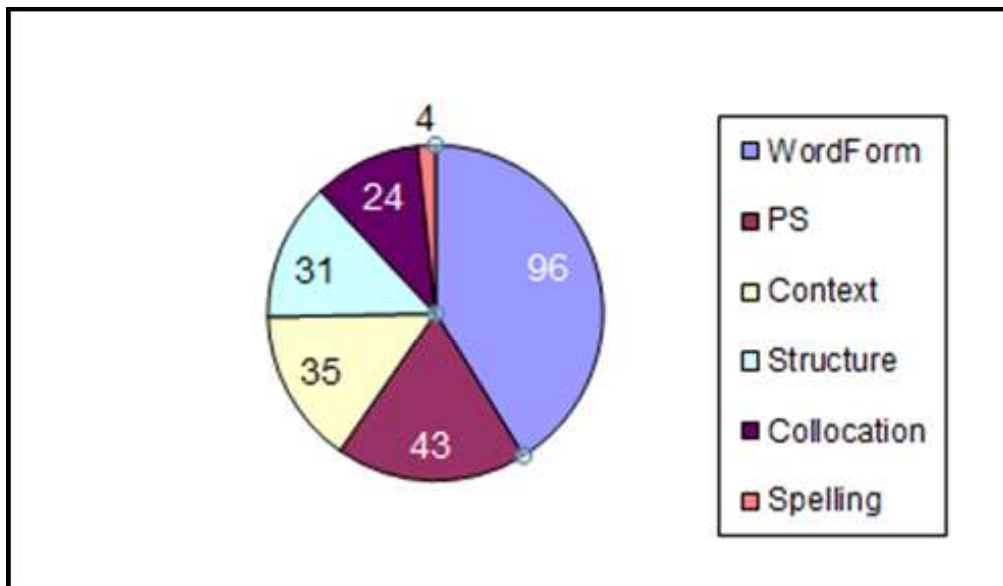


Figure 1: Error Type Distribution

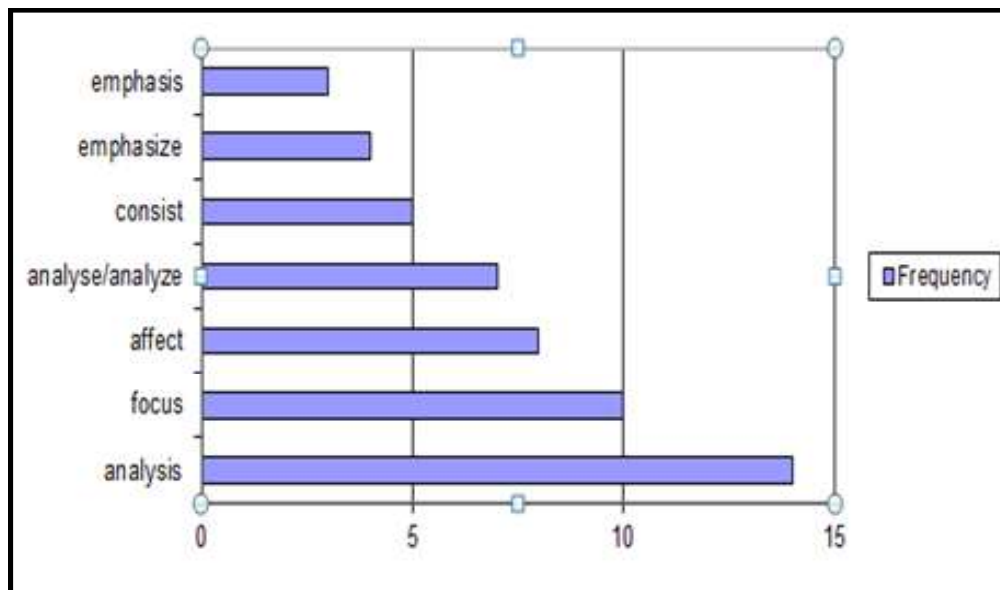


Figure 2: Most Frequently Misused Words

Therefore, there may be some grounds for the assumption that the results are representative of XJTLU population at large. In any case, the errors found in the population of English majors would seem to be errors proper (James, 1998), most likely identifiable in Chinese learners of English at every level and therefore worth analysing.

According to Chang (1987), word form errors would be indicators of first language (L1) interference or negative transfer, since Chinese

as the students' L1 is less inflected than English as their target language (L2). Based on Chang (1987), the same claim could be made regarding PS errors, which were the next most frequent category. Almost one fifth (18.5%) of all AWL errors were classified as PS, which is consistent with previous research (Dodigovic, 2005; 2013). What is surprising is the fact that only 15% of AWL errors had any bearing on sentence structure, although previous research (Yip, 1995) would suggest that structure errors are both

frequent and significant. A possible reason for the underrepresentation of structure-related errors could be the fact that only approximately 12% of the corpus vocabulary was investigated. The reason for this is the focus of the study on AWL, which generally accounts for approximately 10% of academic texts (Nation, 2001). In the learner corpus used in this study, the percentage of AWL was just under 12%. Lexical errors concerning vocabulary other than AWL are not the subject of this study. The same applies to structural errors of more general nature. The small size of corpus is another variable possibly contributing to the small size of the pool of errors.

Another unexpected outcome was the low percentage of collocation errors (10%). Literature identifies collocation, especially in EFL contexts (Yamashita & Jiang, 2010), as relatively difficult to master. Therefore, the low frequency of collocation errors may be indicative of a specific emphasis on AWL and collocation in English instruction at XJTLU. On the other hand, anecdotal evidence seems to suggest that academic words may not have been explicitly taught to the students. Rather, incidental acquisition of AWL words might have occurred through multiple encounters in everyday learning activities, involving extensive as well as intensive reading and writing tasks. Literature suggests that incidental learning may result in only partial mastery of words (Schmitt, 2010), and having an impact mainly on the receptive dimension of word knowledge (Brown, Waring & Donkaewbua, 2008). Therefore, failure to master all aspects of AWL vocabulary suggests that incidental learning may be a probable cause. One possible implication of this is that in addition to incidental learning of academic words, EAP instruction should facilitate deliberate learning of academic vocabulary, with a reasonable regard for word form, especially in Chinese contexts. This recommendation is echoed by Kirchner (2013), who studied the vocabulary size of a different group of students in a similar context.

Conclusion

This study has investigated academic vocabulary used in the final-year students' writing in the Department of English, Culture and Communication at XJTLU. Two research questions were addressed: 1) How well are

academic words used in XJTLU students' writing? 2) What salient errors can be identified in the use of academic words by Chinese learners? In response to the first question, the fact that only about a quarter of the identified AWL were used erroneously, less than two times each on the average, would suggest that relative mastery of AWL has been achieved. With respect to the second question, it appears that the large majority of all AWL related errors are word form errors. Based on previous research (Chang, 1987; Yip, 1995; Dodigovic, 2013), the salience of word form errors seems to point to the negative transfer from the students' first language, while the partial mastery of words suggests that these might have been acquired incidentally. The combined impact of negative L1 transfer and the possible incidental acquisition of vocabulary might best be countered by the provision for raising language awareness and increased opportunities for deliberate vocabulary learning within the context of EAP instruction.

References

- Agustin Llach, M. P. (2011). *Lexical errors and accuracy in foreign language writing*. Bristol: Multilingual Matters.
- Brown, R., Waring, R., & Donkaewbua, S. (2008). Vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, 20 (2), 136–163.
- Chang, J. (1987). Chinese speakers. In M. Swan, & B. Smith (Eds.) *Learner English* (2nd Edition) (pp. 224–237). Cambridge: Cambridge University Press.
- Cobb, T., & Horst, M. (2001). *Growing Academic Vocabulary with a Collaborative On-Line Database*. Retrieved from <http://elc.polyu.edu.hk/conference//papers2001/cobb.htm>.
- Cowan, J. R. (1974). Lexical and syntactic research for the design of reading materials. *TESOL Quarterly*, 8 (4), 389-399.
- Coxhead, A. (1998). *An academic word list*. LALS Occasional Publication. Victoria University of Wellington, Wellington.

- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34 (2), 213–238.
- Coxhead, A. (2002). The academic word list: A corpus-based word list for academic purposes. In B. Ketterman & G. Marks (Eds.), *Teaching and language corpora (TALC) conference proceedings* (pp. 73-89). Atlanta, GA: Rodopl.
- Coxhead, A. (2012). Academic vocabulary, writing and English for Academic Purposes: Perspectives from second language learners. *RELC Journal*, 43 (1), 137–145.
- Coxhead, A., & Nation, I. S. P. (2001). The specialized vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 252-267). Cambridge: Cambridge University Press.
- Deng, X., Lee, K. C., Varaprasad, C., & Leng, M. L. (2010). Academic writing development of ESL/EFL graduate students in NUS. *Reflections on English Language Teaching*, 9 (2), 119–138.
- Dodigovic, M. (2005). *Artificial Intelligence in Second Language Learning: Raising Error Awareness*, Clevedon: Multilingual Matters.
- Dodigovic, M. (2013). Intelligent Sentence Writing Tutor: A System Development Cycle, *International Journal of Artificial Intelligence in Education*, 22, 141–160.
- Folse, K. (2004). *Vocabulary myths*. Ann Arbor: University of Michigan Press.
- George, H. V. (1972). *Common Errors in Language Learning*. Rowley: Newbury House.
- Hemchua, S., & Schmitt, N. (2006). An analysis of lexical errors in the English compositions of Thai learners. *Prospect*, 21 (3), 3–25.
- Hiebert, E. H., & Lubliner, S. (2008). The nature, learning, and instruction of general academic vocabulary. In A. E. Farstrup & S. Samuels (Eds.), *What research has to say about vocabulary instruction*, (pp. 106-129). International Reading Association.
- Hyland, K., & Tse, P. (2007). Is There an "Academic Vocabulary"? *TESOL Quarterly*, 41(2), 235-253.
- James, C. (1998). *Errors in language learning and use. Exploring error analysis*. London: Longman.
- Kirchner, R. (2013). The results of a vocabulary levels test: Version 1 conducted with foundation year Chinese students. *English Teaching in China*, 2, 1–7.
- Li, Y. & Qian, D. D. (2010). Profiling the Academic Word List (AWL) in a financial corpus. *System*, 38, 402–411.



Marina Dodigovic is currently the Director of MA TESOL at XJTU. She has taught English, directed programs and trained ESL teachers in Europe, Australia and Asia. In addition, she has conducted research in applied linguistics and second language acquisition, resulting in a number of refereed publications.

marina.dodigovic@xjtu.edu.cn



Hui Li is currently teaching Interpreting and Translation at XJTU. Her research expertise is in vocabulary acquisition, vocabulary assessment, translation assessment, audio-visual translation, and interpreter and translator training. She is also a reviewer for some international academic journals, such as *International Assessment Quarterly*, *Asian ESP* or *Voices in Asia*.

hui.li@xjtu.edu.cn