



Journal of Arts & Humanities

Volume 06, Issue 08, 2017, 47-53

Article Received: 21-08-2017

Accepted: 04-09-2017

Available Online: 07-09-2017

ISSN: 2167-9045 (Print), 2167-9053 (Online)

DOI: <http://dx.doi.org/10.18533/journal.v6i8.1255>

A Corpus-based Study of English Vocabulary in Art Research Articles

Ping Wang¹

ABSTRACT

The learning of English as a foreign language is an additional burden for art majors. This study aimed to examine high frequency words in art research articles to improve the efficiency of art majors' English learning, especially their academic reading and writing. For this aim, the study built a corpus, analyzed data from art research articles and compared data with three base word lists. We found that the General Service List (GSL) and the Academic Word List (AWL) had a high coverage in our corpus, and there was a different high frequency word order in the Art Research Article Corpus (ARAC). These findings provide some implications for teaching English for art majors.

Keywords: Art Majors, English Teaching, High Frequency Words, Word List.

This is an open access article under [Creative Commons Attribution 4.0 License](#).

1. Introduction

In China, art majors have much more difficulty in the acquisition of English as a foreign or second language. For them, there is a specialized entrance examination of art colleges with the emphasis on candidates' professional skills. And for a better score in the entrance examination, art students pay more attention to develop their enterprise in art. Their English language proficiency is comparatively lower than other majors (Zhang et al., 2008; Wang, 2012; Li, 2014; Yuan & Wang, 2015). But, once they are admitted to art colleges, they have to take the common core course, i.e. English communication. Thus, it is a challenge for English educators to teach English communication in art colleges. How to teach the course effectively is not only essential for students to learn about foreign culture, communicate in master classes delivered by foreign artists, read research articles in international journals, but also strategic for the whole country to build a strong cultural nation, realize cultural renaissance, and involve in more international culture interactions.

Many studies (e.g. Zhang, 2010; Li, 2012) indicate that the mainstream English pedagogy is still the traditional grammar-translation approach and the class is still teacher-centered in Chinese art colleges. In the author's college, English teachers use the common textbooks, which is similar to those

¹ Lecturer, Department of Common Core Courses Teaching, Shandong University of Arts, People Republic of China. E-mail: margaretpwang@163.com

of other colleges or majors. Many students show little interest in English communication class, partly due to the old-fashioned learning materials, and partly because of the unmatched and unsuitable issue between the textbooks and art majors' current English level. Liu (2010) maintains art majors' small vocabulary size is the primary cause of their lower English proficiency. And she also states that due to the inadequate vocabulary, art majors cannot understand some simple texts and articles, let alone speak English, listen to English and write in English. These statements echo what Hu and Nation (2000) note: vocabulary size predicts how much can be understood. This means vocabulary is essential in the acquisition of English, and the bigger a reader's vocabulary, the better the reader could understand the reading materials.

Lexicology researchers find that there is a special need and learning focus for different majors at different learning stages (Coxhead, 2000; Nation, 2001). The vocabulary need varies in different disciplines. For instance, Ward (2009) created a basic engineering English word list and suggested that there should be a priority of lexis learning for less proficient learners. Valipouri and Nassaji (2013) analyzed a corpus of chemistry research articles and suggested that students should learn vocabulary according to their discipline-specific needs. However, few scholars established an art English word list and examined the lexical need of art major to read research articles. This study aims to study English vocabulary in art research articles, examine the coverage of the GSL and the AWL words in our corpus, and help them to master the needed vocabulary to understand art research articles better for further study and involvement in the Eastern and Western culture communication.

1.1 Lexical threshold necessary for reading comprehension

Vocabulary may be a good predictor of reading comprehension (Laufer, 1992; Nation, 2006). If a reader has a bigger vocabulary size, the reader can understand better the reading materials. According to Alderson (1984), there is a lexical threshold below which readers cannot employ their reading strategies to read L2. Thus, it is vital for researchers and educators to know the lexical threshold for various reading comprehension.

Researchers (e.g. Laufer, 1989; Nation & Waring, 1997; Hu & Nation, 2000) have investigated the relation between L2 vocabulary knowledge and reading comprehension. They claimed that readers needed to know a certain percentage of the running words in a text for better comprehension. For instance, Laufer (1989) stated that 95% of running words were needed for reasonably understanding a text. This indicates that if a reader's lexical threshold is below the 95%, the reader could not read the text effectively without any help. To read for pleasure, Hirsh and Nation (1992) suggested learners master 98% of the running words in a text. More studies recommended that 95% of the running words be necessary for minimal comprehension with 4000-5000 word families and 98% be great for optimal comprehension with 8000 word families (Nation & Waring, 1997; Nation, 2006; Laufer and Ravenhorst-Kalovski, 2010).

1.2 Previous studies on word lists

Investigations of word lists and vocabulary are abundant due to computer science and internet technology. Michael West (1953) established the General Service List (GSL) which included the most frequent 2000 words. Studies showed that the GSL occupied a large proportion in various reading materials. For example, Sutarsyah et al. (1994) found the coverage of the GSL was 82.5% in English for Special Purposes (ESP) corpus and 78.4% in English for Academic Purposes (EAP) corpus. Billuroglu and Neufeld (2005) got the similar finding that the GSL accounted for about 80% of running words in any written text. For the beginners of English learning and less inefficient learners, it is generally recommended that the GSL should be paid more attention for its high frequency occurrence in written texts.

Further studies showed that the GSL is not sufficient for intermediate and advanced learners who have more reading tasks, especially academic reading. Coxhead and Nation (2001) stated that learners should focus on academic vocabulary when they master the GSL words. Coxhead (2000) developed her Academic Word List (AWL) with 570 word families (i.e. headwords and their inflectional and derivational family members), with 10% of the running words in her corpus. This indicates that if a reader masters the AWL, the reader can understand 10% of their academic materials, such as textbooks

or research articles. Since it was formed on the basis of a wide range of corpus, the AWL is widely cited and cross-disciplinary. There are Medical Word lists (Wang et al., 2008; Hsu, 2013), Engineering English Word Lists (Ward, 2009; Hsu, 2014), Agriculture Word List (Martinez et al., 2009), and Chemistry Academic Word List (Valipouri and Nassajo, 2013). However, there are few studies on word lists for art majors, especially for them to understand research articles in their discipline.

1.3 The purpose of the present study

This study aimed to examine the frequency and usage of the GSL and the AWL words in our art corpus. We explored the following two questions:

- i. What are the profiles of the GSL and the AWL in our art research article corpus?
- ii. Are there any overlaps and differences among the GSL, the AWL and our corpus?

2. Method

2.1 Corpus compilation

Art Research Article Corpus (hereafter ARAC) was built for the study. All the research articles in the corpus were downloaded from two prestigious databases: Wiley Online Library (<http://onlinelibrary.wiley.com/>) and ScienceDirect Online (<http://www.sciencedirect.com/>). We selected art research articles as the single genre for articles' reading and writing are two fundamental skills for art majors. They will benefit a lot from the mastery of the high frequency words in art research articles. Considering art subjects' establishment and the accessibility of research articles in the author's university, we chose four sub-disciplines of art: media and communication, design, museum and heritage studies, and art history and theory. For the representation of the samples in the ARAC, we consulted two art professors who work at the author's university, and they gave lists of high quality art journals. We chose four international journals for each subject. From each journal, we randomly selected twenty articles with full text, which were published between 2010 and 2016. The result showed that the ARAC contains 320 texts with the running words of 1.95 million running words.

2.2 Data processing and analysis

All the downloaded samples were originally in PDF formats and were converted into text forms (i.e. *.txt) to be processed by the computer software program Range, which can be downloaded for free at <https://www.victoria.ac.nz/lals/about/staff/paul-nation>. Range is a program that "can be used to find the coverage of a text by certain word lists, to create word lists based on frequency and range, and to discover shared and unique vocabulary in several pieces of writing" (Nation, 2005, p.2). We used Range for lexical analysis for it consists of three base lists: the first and second 1000 most frequent words in the GSL, and the AWL Words, so that we could examine the coverage of the GSL and the AWL in our corpus. For the standardization of the research articles in the corpus, figures, tables, notes, acknowledgments, bibliographies, references and appendices in the articles were removed to be readable by the software. For the normalization of the words, a headword and all its inflections and derivations were counted as a word family. For instance, a word *achieve* and other forms such as *achieves*, *achieved*, *achievable*, *achievement* and *achievements* were counted as one word by the software. If they know a headword, learners will master its family members much easier (Coxhead, 2000). In this study, a word family means a word token.

To analyze the data, we first determined the frequency and distributions of word tokens and types in the ARAC. The data was processed by the software Range, and Range could show the coverage of the words in the ARAC. Then, we examined the words which were not on the lists of the GSL and the AWL but occurred with high frequency, and analyzed the usage and coverage of those words. We followed Coxhead's (2000) procedure, so we only included the words which occurred in the four fields of sub-corpora and at least 5 times in each discipline.

3. Results and Discussion

3.1 The coverage of base word lists in the art research article corpus

In the corpus, there were 1 950 693 running words, 78457 word types and 4112 word families. A total number of 2287 word families were met the word selection criterion (≥ 20 times occurrence). Table 1 shows the total coverage of the base word lists in ARAC. The first 1000 frequent words account for 73.21% of lexical coverage and had 593 word families in the Art Research Articles Corpus. The second 1000 frequent words had coverage of 6.74% with 412 word families in the ARAC. In total, the first 2000 GSL words were counted in 79.95% coverage. The results show that mastery of the top 2000 GSL words had a significant meaning in reading and writing art research articles. If a learner has the mastery of 1005 word families, he or she could understand almost 80% of the running words in art research articles.

Table 1: The coverage of three base word lists in ARAC

Word List	Tokens	% In The ARAC	Families
one	1385069	73.21	593
two	126775	6.74	412
three	198678	10.57	485
Proper noun	43519	2.35	234
Not in the list	138131	4.14	563
Total	1892172	97.01	2287

Note: Word list One is the first 1000 frequent GSL word list, List Two is the second 1000 frequent GSL word list, and List Three is the AWL. Not in the list means the words were non-GSL and non-AWL.

For accumulative coverage point, learners should learn more words. Word list Three is the AWL, which accounted for 10.57% of lexical coverage in the ARAC with 485 word families. This percentage was higher than the average 10% (Coxhead, 2000) and that of Chemistry Academic Word's 9.96% (Valipouri & Nassaji, 2013). According to Wang et al. (2008), the AWL accounted for 12.24% of lexical coverage in their medical academic corpus, which was much higher than that of coverage in the ARAC. If the learner knows the meaning and usage of the AWL, his or her vocabulary would be expanded by a wide margin (i.e. 10% on average). In addition to the former 79.95%, the learner would have 90.52% of lexical coverage in the ARAC. The results indicate when mastering the 485 word families, the learner could understand 90.52% of the running words of art research articles after the training of reading strategies.

In order to have a more lexical coverage, it is vital for art majors to learn more proper nouns, because they covered 2.35% with 234 headwords in the ARAC. Many proper nouns have very high frequencies in the corpus such words as countries for their culture heritage and art achievements, like China (326 times), America (673 times), Italy (267 times), and popular themes or times like Islam (274 times), Catholic (310 times), and Renaissance (410 times). Additionally, learners' vocabulary size would cover 92.87% of lexical coverage in the ARAC.

From Table 1, we could figure out there were words which were not in the lists and we referred to these words as non-GSL and non-AWL words. They accounted for 4.14% of lexical coverage in the ARAC corpus. It is essential for art majors to know more about these words for better reading comprehension. If they know these non-GSL and non-AWL words, they could reach 97.01% of lexical coverage in reading art research articles, and the percentage is more than that of 95% for a reasonable understanding.

We identified 2242 word families with high frequency, covering 95% of lexical coverage in the ARAC. This means that art majors need to master these 2242 word families to understand art research articles.

3.2 Comparison of the GSL and the AWL in the ARAC

We compared the GSL words in the ARAC with ones in the GSL itself. Both of the lists of the top 30 words are the function words, such as *the*, *to*, *that*, *with*, *this* and *have*. There are some exceptions. One is the word *which* in the top 30 list in the ARAC with 6524 occurrences while it was listed 40th with 2286 times in the GSL. That is because our corpus consisted of research articles which were more academic texts with more complicated and longer sentences, and the word *which* is high frequent one to be used to modify antecedents.

Table 2: Comparison of top 10 GSL words in the ARAC and in the GSL

Top 10 words in the ARAC	Top 10 words in the GSL
which	know
social	which
time	think
study	like
new	make
most	see
different	people
use	year
experience	well
however	because

We removed the function words which are considered too general, such as articles (*a, an* and *the*), some prepositions (*in, on, at, of* and *to*), pronouns (such as *it, I, we*, and *they*), and numbers (*one, two, three*, etc). We examined top 10 words as showed in Table 2. There was only the word *which* with high frequency occurrences in both lists, while other words were totally different. It is very urgent for art majors to know high frequency words in art research articles for efficient vocabulary acquisition.

Beside the GSL words, we compared top 10 academic words in the ARAC and in the AWL. Table 3 shows the top 10 words in the AWL on the right side, and these words have highest frequencies in a cross-disciplinary corpus with humanities and sciences texts (Brown et al., 2013). The words on the left side of Table 3 are the highest frequency ones in the ARAC. All the words on the both sides were quite different from each other. In total, we identified 485 academic word families with high frequency in our corpus. We compared our words with AWL, and found that many of the AWL words did not occur, or had low frequency in our corpus. Even the high frequency AWL words in the ARAC have different frequency orders. This indicates that it is necessary for educators and researchers to establish discipline-specific vocabulary lists so that students could master to read and write in their field (Hyland & Tse, 2007; Wang et al., 2008).

Table 3: Comparison of top 10 academic words in the ARAC and in the AWL

Top 10 AWL words in the ARAC	Top 10 words in the AWL
design	repertoire
media	obtain
research	distribution
process	parameter
data	aspect
theory	dynamic
approach	impact
role	domain
context	publish
individual	denote

4. Conclusion

The study was conducted on the basis of the art research article corpus. We built the art research article corpus by collecting articles from sixteen journals. We used the software Range to analyze data, and compared word frequencies with the three base word lists. We identified 2242 word families with 95% of lexical coverage in the ARAC. We also figured out there are overlaps and differences among the GSL, the AWL and our corpus. These findings offered implications for language educators and material designers to establish discipline-specific word lists and students will benefit a lot from specific word list and acquire the words efficiently.

The study is only a preliminary one on the vocabulary of art research articles. Due to the accessibility of data downloaded from the internet, we only collected data from four main sub-disciplines of art, so the corpus is comparatively small and confined to research articles. For further

study, a larger corpus of more art research articles and art textbooks would be significant. These efforts will provide more implications for language teachers, especially for those in the ESP field, and students would benefit much from the specific word list for better reading and writing in their fields.

References

- Alderson, J. (1984). Reading: a reading problem or a language problem? In J. C. Alderson, & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 1-24) London: Longman.
- Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography*, 6, 253-279.
- Billuroglu, A., & Neufeld, S. (2005). The necessities in lexis: a new perspective on vocabulary profiling. http://www.lextutor.ca/vp/tr/BNL_Rationale.doc.
- Coxhead, A. (2000). A new academic word list *TESOL Quarterly*, 34, 213-238.
- Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, 33(66-76), 66.
- Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 689-696.
- Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. *Language Teaching Research* 17(4), 454-484.
- Hsu, W. H. (2014). Measuring the vocabulary load of Engineering textbooks for EFL undergraduates. *English for Specific Purposes*, 33, 54-65.
- Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language* 13, 403-430.
- Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly*, 42(2), 235-253.
- Li, F. X. (2014). A Study on the English Learning Status of Art Students. *Journal of University of Shanghai for Science and Technology*, 36(3), 277-281.
- Li, H. (2012). On Present Situation and Reform of University English Teaching in Professional Arts Institutions. *Journal of Nanchang Teaching College*, 27 (8), 156-157.
- Li, Y., & Qian, D. D. (2010). Profiling the academic word list (AWL) in a financial corpus. *System*, 38, 402-411.
- Liu, J. T. (2010). Analysis on the Status Quo and Countermeasures of Art Majors' English Studying. *Hundred Schools in Arts*, 117 (8), 429-432.
- Liu, N. A., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal*, 16, 33-42.
- Martinez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes*, 28, 183-198.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63(1), 59-82.
- Nation, L. S. P., & Heatley, A. (2007). Range. <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx>.
- Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. In N. M. Schemitt, M. (Ed.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge: Cambridge University Press.
- Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based study. *RELC Journal*, 25(2), 34-50.
- Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes*, 12, 248-263.
- Wang, L. Y. (2012). Study on Effective Teaching of English for Art Major Students. *Value Engineering*, 8, 209-210.
- Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, 7, 442-458.
- Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, 28, 170-182.
- West, M. (1953). *A general service list of English words*. London: Longman, Green & Co.
- Yuan, J., & Wang, L. M. (2015). The Study of Knowledge Structure of College English Textbooks for Arts

and Sports Majors: From the Perspective of Connectionism. *Heilongjiang Researches on Higher Education*, 25(3), 174-176.

Zhang, L. (2010). Strategies and Survey of College English Teaching to Art Students. *Heilongjiang Education*, 6, 91-92.

Zhang, Y. J., Wang, L. J., Shen, T. T., & Liu, C. X. (2008). Survey and Discussion on ESP Teaching and Learning of the Fashion Design Major. *Journal of Zhejiang Sci-Tech University*, 25(6), 771-775.