

An Analysis of the Lexical Characteristics in English Abstracts of Chinese Academic Journal Articles: Take Intelligence Journals Articles as an Example

Xiaoyu Yao

School of Foreign Language, Ningbo University, Ningbo 315211, China;

318467909@qq.com

Keywords: Corpus, lexical characteristics, abstracts, cognition.

Abstract. This paper explores lexical characteristics of abstracts of English research articles (RAs) published in Chinese top academic journals (indexed by CSSCI) through corpus method from the perspectives of lexical complexity, lexical diversity and lexical density. The results showed that Chinese journals abstract had below lexical characteristics: the lexical diversity and lexical density in CAC were lower than EAC, while the lexical complexity of both corpora were similar.

1. Introduction

As a main mean to exchange thoughts with worldwide researchers, writing persuasive Research Articles (RAs) is considered to be important. As important parts of RAs, abstracts summarize contents of papers and are usually open in most databases. The quality of abstract is determined by Lexical proficiency (Grabe, 2016; Harley & King, 1989; Engber, 1995; Lauger & Nation, 1995). In order to study the lexical characteristics of abstract in Chinese journals, corpus approach is applied with considering three indicators (lexical complexity, lexical diversity, lexical density).

2. Literature Review

Lexical Complexity: The average word length is an important factor in determining the difficulty and complexity of texts. The length of a word could affect the memories of other words (Novak & Spelman, 2003: 74-79). The formula for calculating the average word length (AWL) is: number of letters / number of words.

Lexical Diversity: Laufer(1991) defined it as the ratio of number of words to the number of different unique word stems (Type/Token Ratio(TTR)=Types x 100%/Tokens). The bigger the lexical diversity is, the more word types are used, indicating that the learner has better lexical proficiency. Through applying WordSmith 5, Standardized Type/Token Ratio is used to decrease negative effect of high repeat frequency of words in estimating lexical diversity through TTR.

Lexical Density: Lexical words meant noun, notional verb, adjective and adverb (Read, 2000: 203) and Lexical Density means the ratio of lexical words to the total words of the text (Ure, 1971).

In this research, differences of lexical characteristics in English abstracts between Chinese journals and international journals are discussed to fulfill related gap in Chinese context.

3. Research Methodology

Two corpora were constructed to describe the lexical characteristics of Chinese researchers. One was called as CAC (short for Chinese Abstracts corpus), an observed corpus, including 500 abstracts from 10 Chinese top journals. Another one was called as EAC (short for English Abstracts Corpus), a reference corpus, including 500 abstracts from international journals. Author randomly selected 1000 abstracts published between 2013 to 2017 from 10 Chinese journals and 10 international journals respectively. Two research instruments including WordSmith 5 and Log-likelihood ratio calculator were employed in the study.

4. Results and Discussion

WordSmith 5 showed that two corpora had 54303 tokens and 91357 tokens in CAC and EAC respectively. There were 4480 and 7534 types in CAC and in EAC respectively. Mean word length was 5.76 and 5.61 in each respectively.

4.1 Lexical Complexity

The length of a word closely links with the text style and genre, reflecting the lexical complexity. Standardizing the collection of data helps to compare two corpora comparatively. The numbers of letters in every 10 thousand words were included as: The span of the word length in CAC was little bigger than that in EAC. The number of words containing more than 15 letters occupied 0.34% in CAC, and 0.28% in EAC. The words (2-4 letters) appeared more frequently in the both two corpora (CAC:43.45%; EAC:42.95). 79.14% of CAC and 77.40% of EAC words had 2-8 letters. Sorted by frequency, the ranking of frequency of words with different numbers of letter was: 3 letters > 2 letters > 7 letters > 4 letters > 8 letters > 5 letters > 6 letters.

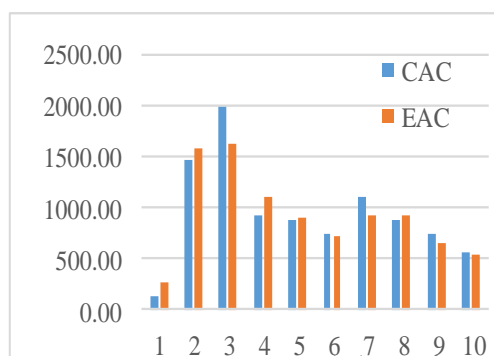


Fig.1 The number of 1-10 letters words

In Figure 1, CAC consisted of 49720 words that contained less than 11 letters (91.56% of total words) and EAC had 83964 words that contained less than 11 letters (91.91% of total words). The log likelihood ratio was 0.45 ($p=0.503>0.05$), indicating no significant difference between the two corpora. Words with 1-10 letters were the most common words in abstracts. The results showed that the average word length is 5.71 (42.75%) and 5.61 (40.25%) in CAC and EAC respectively. The average length of English word was 4 letters, and it was considered to be difficult to understand if the average length was over 4 letters (Laufer, 1990a). A study showed that the number of words influences storage of vocabulary of human brain (Peter L., 2000:1-22) in three word-length effects: letters, syllables and morphemes. Word frequency, lexical ambiguity, extraction time, the structure of the morpheme, lexical features and semantic priming affected the extraction of words in human brain when languages learners used language to communicate (Tian & wang, 2010:120 ~ 122). The cognitive theory holds that the higher the frequency of a word, the lower the threshold of recognition. Short words are often considered as polysemes, and often used by any group of learners in any types of genre. Research articles related to social issues belong to professional genre. Professional and scientific vocabulary have more number of letters usually. Researchers always use professional vocabulary to write RAs. As a result, the average word length is longer than ordinary articles.

4.2 Lexical Diversity

The Lexical Diversity provided an indication of whether the authors have a good proficiency in producing vocabulary (Laufer, 1991). The TTR was 8.31 and 8.32 in CAC and EAC. The STTR was 39.75 and 43.48 respectively. The STTR derivation was 57.88 and 55.28 respectively. There were some differences between the two corpora: EAC's STTR was 43.48, 3.73 higher than that of CAC. It indicated that the author in EAC used more derivative and inflectional forms (such as proposal and richer, abandons). EAC's STTR deviation was 55.28%, 2.60% than that of CAC, indicating that the data in EAC was less discrete and fluctuated than CAC. There were no significant ($p=0.764>0.05$). Lexical diversity was closely related to word derivation and inflection, usage of synonyms and three types of vocabulary storage models were proposed by cognitive psychologists (Huang, 2001:430-436): First one was Word Access Model: They took a word as the basic unit when language learners storing a word. A word, no matter how many morphemes, were stored as separate entry in the mental

lexicon. Every variant of a word (such as a girl, girls, girlish) was independent on the word, and they all had their own mental representations. Second one was Morpheme Access Model: This model assumed that the word was made up of morphemes, the basic units of the word. Language learners stored English words in the form of morpheme. Third one was Augmented Addressed Morphology Model: It held that learners' familiar words were stored in whole words, while unfamiliar words are stored by basic unit morphemes.

Three models are complementary and hierarchical. In the preliminary learning stage, most of language learners stored and extracted the word in a whole word way. With the improving of the level of language proficiency, language learners' vocabulary size was larger and they accumulated enough affixes. The storage model was gradually transit to the second and the third model, which caused that high-frequency words might be stored and extracted in a whole word way; and the low-frequency words might be stored and extracted separately from the roots and the affixes. Chinese authors knowing the affixes could recognize English words and construct new words, effectively making use of cognitive resources and saving processing time. Since there was no significant difference between Chinese authors and native authors in derivation of word-formation, the lexical diversity would be explained by another reason. Namely, Chinese authors' ability to use synonymous expressions was weaker than native authors. Through this result, English affixes should be paid attention in teaching.

4.3 Lexical Density

The ratio of the lexical words to the total amount of words is influenced by the length of the text. WordSmith 5 calculates the lexical density in every 1000 characters, improving the reliability of this indicator. The rankings of frequency of types of the words were same in the two corpora. The most used word class was Noun. It accounted for 38.92% and 33.02% respectively. Next was Adjective, Notional Verb. The less used word class was Adverb. Log likelihood and p value ($p=0.000$) showed that there were differences between these four lexical words in the two corpora except Notional Verb ($P=0.145>0.05$). The lexical density was 58.23% and 53.57% in the CAC and EAC respectively. The p value ($p=0.000<0.05$) was calculated only to find that there was significant difference between two corpora. In CAC, the number of Notional Verb, Adjective and Adverb was less than it in EAC a little bit. On the contrary, the number of Noun in CAC was more than it in EAC. Text capacity and text difficulty were determined by lexical density. The more lexical density was, the more capacity and difficulty the text are. Ure(1971) studied lexical density on written text and spoken text and drew a conclusion that the lexical density in written text was more than 40%, and in spoken text was less than 40%. The present study found that the lexical density was 58.23% and 53.57% in CAC and in EAC respectively, which were higher than Ure's result. The lexical density in CAC was a little more than it in EAC (58.23% and 53.57%), which showed that CAC was slightly more difficult than EAC. However, In CAC, the number of Notional Verb, Adjective and Adverb was less than it in EAC a little bit. So, the impact was from the number of Noun. This showed that the author of Chinese articles does have a high level of English and they used nominalization to express more information. At the same time, there was still a certain gap in other aspect except noun compared with the abstracts in EAC, which showed that the Chinese authors have some room for improvement.

5. Conclusion

The results showed that lexical characteristics in two corpora had difference in lexical complex, diversity and density and this result could provide a guidance to improve skills of teaching academic writing for teachers and improve writing skills for researchers. Comparing with international journals, Chinese journals had below characteristics: the lexical diversity and lexical density were less.

References

- [1] Engber, Cheryl A. "The relationship of lexical proficiency to the quality of ESL compositions." *Journal of Second Language Writing* 4.2(1995):139-155.
- [2] Grabe, W, and C. Zhang. "Reading-Writing Relationships in First and Second Language Academic Literacy Development. " *Language Teaching* 49.3(2016):339-355.

- [3] Harley, B and M. L. King. "Verb Lexis in the Written Compositions of Young L2 Learners." *Studies in Second Language Acquisition* 11.4(1989):415-439.
- [4] Huang Yuanzhen. "On the relationship between morphological motivation and lexical acquisition." *Foreign Language Teaching & Research* (2001).
- [5] Laufer, B. Why are some words more difficult than others? ---- some intralexical factors that affect the learning of words [J]. *IRAL*. Vol.18.4,(1990a): 293-307.
- [6] Laufer, B. The development of L2 lexis in the expression of the advanced learner. *The Modern Language Journal*,.75.4 (1991):440-448.
- [7] Laufer, B and P Nation. "Vocabulary Size and Use: Lexical Richness in L2 Written Production." *Applied Linguistics* 16.3(1995):307-322.
- [8] Novak, C and L Spelman. "List composition and the word length effect in immediate recall: A comparison of localist and globalist assumptions." *Psychonomic Bulletin & Review* 10.1 2003):74-79.
- [9] Peter Lovatt, S.E. Avons, and Jackie Masterson. "The Word-length Effect and Disyllabic Words." *Quarterly Journal of Experimental Psychology A Human Experimental Psychology* 53.1 2000):1-22.
- [10] Read, John Arthur Stanley. *Assessing vocabulary. The Companion to Language Assessment.* John Wiley & Sons, Inc. 2000.
- [11] Tian and Wang. "Psycho-cognitive theories and foreign language teaching." *Peking University Press*, (2010):120-122.
- [12] Ure, J. Lexical density and register differentiation. In G. Perren and J.L.M. Trim (eds), *Applications of Linguistics*, London: Cambridge University Press. (1971) 443-452.