

# Corpus-based Study on the Use of Delexicalized Word “Thing” by Non-English Majors in L2 Writing

HU Xu-yuan

Wuhan University of Technology, Wuhan, Hubei, China

Recent years, interdisciplinary studies on English writing have become more and more prevalent, which is especially indicated by corpus-based studies, springing up. As the theoretical foundation, the Delexicalization Theory, which is a reasonably common language phenomenon, provides a more extensive perspective for English writing. According to the real linguistic data of EFL learners (The CET4 essays of four different classes), it can be easily found that students tend to use the word “thing”—considered to be one of the frequently-used delexicalized words, to express their opinions by means of the analysis of AntConc. However, the results indicate that Chinese non-English majors are inclined to overuse the word. Meanwhile, there are a large amount of inappropriate use of the word “thing” such as redundancy, substitution, collocation and semantic errors. Moreover, the study attempts to summarize the cause that the overuse may be attributed to improper writing strategies, and the effect of fossilization of interlanguage in order to provide some implications for English teaching promotion.

*Keywords:* corpus, delexicalization, thing, non-English majors, L2 writing, AntConc

## Introduction

In the traditional practice of English teaching, China’s teachers are always accustomed to paying great attention to the difference meaning of the word at the Chinese semantic level, and then try to regulate students’ use of English vocabulary through some typical situational examples (Wang, 2011), which is inefficient and causes some usage error. With the rise of corpus linguistics, the use of concordances can provide rich usage and context for words, enabling learners to compare and master the subtle semantic and pragmatic differences between different meaning of the words (Li, 2001). However, those delexicalized words such as “gain” and “run” are still a difficulty for English learners.

Reviewed those studies, it can be found that there is no such research focusing specifically on the phenomenon of learners’ improper use of delexicalized vocabulary. With the development of English teaching in China, currently, it can be observed that students’ writing indicate that they are inclined to overuse the word “thing(s)” to express opinions so that the study aims to investigate the use of “thing(s)” and attempts to analyse the causes.

## Research Design

This article adopts KWIC retrieval method by means of a corpus retrieval software—AntConc to search the keyword, and then to analyze the concordance to determine the different uses of keywords in contexts. What is more, the study carries out manual statistics, then compare and analyze the results from different aspects.

### Research Questions

- (1) What is the feature of Chinese non-English majors using the delexicalized word “thing”? How can those errors be classified?
- (2) What may be the cause of the inappropriate use?

### Data Collection

The study collected 143 essays from four classes of a key university of science and engineering in Central China Region. The students are freshmen majoring in Computer Science, Machinery Manufacturing, Artificial Intelligence and so on. Moreover, it is the first time to write a CET4 writing named “The challenges of studying abroad” without receiving any training since they enter the university. The length of the CET 4 essay is 100 to 200 words. Each class collected 35 essays on the average. Therefore, the essay can truly indicate the students’ real writing abilities. The table shows the class scale and collected text data.

Table 1

*The Number of Essays from Each Class*

Class	Class 1	Class 2	Class 3	Class 4
The number of essays	34	33	38	38
Textual data	4945	4671	5678	5953
File length	30344	28284	35502	19032
No. of Hits	44	92	65	23

### Research Method

The study adopts the corpus research method to investigate the use of the delexicalized word “thing(s)”.

Step 1. The original frequency of the occurrence of thing(s) in the corpus is extracted.

Step 2. The use of the word “thing(s)” by learners in different class is calculated and analyzed.

Step 3. The AntConc software is used to search the corpora separately, check the concordance obtained item by item, and irrelevant examples are manually removed, such as “for one thing” and other fixed phrases.

Step 4. Inductively classify the use of the word “thing(s)” in the learner’s corpus to determine the characteristics of the learner’s use of the word.

### Data Analysis

AntConc 3.4.1 is a free analysis tool that can be used to analyse texts in corpus linguistics. According to its creator-Laurence Anthony, AntConc is “a free, multiplatform tool for carrying out corpus linguistics research and data-driven learning” (Anthony, 2014). It has different functions, including word list, keyword list, collocation function and a concordance list. For the analysis of the current corpus, only the word list function and the concordance list function were used to determine how many times specific words occurred in the text and what these words refer to (the concordance list function provides the context in which the words are used).

## Results

### The Frequency of the Use of the Delexicalized Word “Thing(s)”

First, the 243 concordances retrieved from the non-English majors’ corpus are observed one by one. Since this study only focuses on the use of “thing(s)” by learners, the author does not pay attention to the grammatical errors in their use, but the frequency and significance of their use. In general, learners’ use of the word can be divided into two categories: acceptable usage and inappropriate or incorrect usage. Acceptable uses include fixed expressions composed of “thing(s)” such as “for one thing”, etc., as well as correct meaning expressions, including normal use of delexicalized words.

### The Error Types of the Inappropriate Use of the Delexicalized Word “Thing(s)”

Referring to the classification of Gui and Yang (2003), the inappropriate use of “thing(s)” in the learner’s corpus is classified into four categories: redundancy, substitution, collocation, and semantic ambiguity or error. The results show that among the 224 concordances, there are 17 acceptable concordances in the usage of thing, including 16 general usages and 11 delexicalized usages; 180 improper usages, including 59 semantic ambiguities or errors (32.7%), 57 redundant errors (31.6%), 50 collocation errors (27.7%) and 14 substitution errors (7.7%). It can indicate that non-English majors are mainly making mistakes in semantic ambiguities or errors (with the proportion of 32.7%) which may be related to the traditional English teaching method, emphasizing too much on the meaning of each word and then ignoring the significance of context. The following errors rank the second and the third are respectively redundant errors and collocation errors. The least mistakes EFL learner make are on substitution errors with the percentage of 7.7 %, reflecting that Chinese students grasp a relatively large number of vocabularies so that when it comes to the same content, they can have various choices of words to replace the original one.

Table 3

*Different Types of Errors and Corresponding Proportion*

Types of errors	redundant errors	collocation errors	semantic ambiguities or errors	substitution errors
The number of errors	57	50	59	14
Proportion	31.6%	27.7%	32.7%	7.7%

### Redundant Errors

In the process of information transmission, people often send more information than the minimum requirement for understanding. The part of the information that exceeds the minimum requirement is redundant. Here are the obvious examples. The EFL learners prefer to use “the superlative degree+ thing” to express the emotion, while the ideas could be expressed only by the superlative degree. Because “things, aspects, situations, events, etc.” are often used in Chinese to refer to the characteristics of things in general, causing learners to habitually use things after adjectives especially the superlative adjectives.

Example (1) *First, language is the most difficult thing. On abroad, communicating with others is a basic power.*

(2) *I’m an undergraduate, I can bet it’s one of the most exciting things of my life.*

### Collocation Errors

Collocation errors are mainly manifested when learners use the delexicalized word “thing(s)” with improper verbs or attributives, such as Example (3) and (4). The delexicalized feature of “thing” makes learners think that the word can be used with almost any expression, and thus excessively overuse of “thing”, which is also one of the reasons for collocation errors. To illustrate, the phrase “deal with” should be followed by problems. Also, the native speakers do not say “face many things”.

Example (3) *Therefore more should be focused on deal with the things of studying abroad.*

(4) *Although you face so many different things, I think you can do it well.*

### Semantic Ambiguities or Errors

When learners use the delexicalized word “thing(s)”, the characteristics of over-generalization are obvious. Under the circumstance that the meaning of “stuff, aspect, matter or problem” can not specifically expressed, EFL learners often use this word “thing(s)” to summarize. For example, if students want to express that the living habits of Chinese are different from foreigners, they will write “We cannot adjust quietly and we like some things that foreigners may dislike” because of the lack of specific and accurate vocabulary and the effects of Chinese negative transfer. At that time, “thing(s)” is always favorable to express general terms “aspect”, “problems”, “phenomena” and so on.

Example (5) *Because of such things, you may find it hard to make friends with students.*

(6) *First, cultural difference is a challenge. We like some things that foreigners may dislike.*

### Substitution Errors

Substitution errors mainly refer to errors made by other words when the learner cannot express himself. When learners encounter difficulties in expression, the solution is to use a general and safe word to replace the original one. Therefore, “thing” is often the first choice. If learners do not know how to express “big problem” and “various challenges”, they often choose “thing” for generalization. The main reason for substitution errors is that the learner’s vocabulary is too small to meet the needs of expression so that they often use delexicalized words like “thing” to sum up.

Example (7) *First of all, communication is a big thing. If you can't speak the native language fluently, you can't talk to people there normally.*

(8) *In the process of studying abroad, you will meet various things.*

## Discussion

As can be seen from the above, the frequency of use of “thing(s)” by learners is significantly high. Moreover, there is a phenomenon of overuse of the word. In addition, not only do learners overuse “thing(s)”, they often show semantic ambiguity, redundant errors, collocation errors, substitution errors and other errors.

### Writing Strategies

Secondly, it is closely related to the learning strategies used by learners. It can be seen clearly that learners prefer to use repetition strategies. When expressing the similar meaning, learners reuse words or collocations that they consider to be secure. Similarly, learners often use “risk aversion” learning strategies. “Thing” is a commonly used delexicalized word that learners learn early and are more familiar with, so it has become a

versatile “universal word”. When you need to express a certain meaning and cannot find a suitable word, learners usually subconsciously use well-known and generalized vocabulary, reducing the chance of making mistakes.

### **Fossilization of Interlanguage**

Interlanguage is a dynamic language system that gradually moves closer to the correct form of the target language with the development of learning (Selinker, 1972). It is a transitional language system between the mother tongue and the target language. During the learning process, learners constantly adjust their language behaviors, making this language behavior suitable for the idiomatic expression of the target language, and gradually transforming from errors to the right direction. Although the learner has mastered a certain amount of vocabulary at the intermediary stage, the learner often unconsciously chooses the most familiar and safest words, that is, the delexicalized words. In this study, it can be found that most collocational knowledge of delexicalized word is acquired between intermediate and advanced proficiency levels and at the same time fossilizes at this stage.

### **Conclusion**

Excessive and inappropriate use of “thing” is a phenomenon that should be attached great importance. It should be investigated from the aspects of learner’s writing strategies, and fossilization of interlanguage to find corresponding solutions. This research based on corpus provides a new perspective and means for vocabulary teaching, but also reveals some disadvantages of traditional English teaching for the reason that Chinese traditional English teaching has always focused on the analysis of grammar and the mechanical memory of vocabulary semantics. If the Chinese meanings of the target language are memorized in isolation without realizing their usage and semantic rhyme characteristics in the specific context, the language output will be inauthentic or even incorrectly expressed. Therefore, strengthening the study and application of corpus tools and introducing corpora as a supplementary method in teaching practice can improve the efficiency of language teaching. Furthermore, more causes of delexicalized words except for “thing” need to be further studied.

### **References**

- Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied linguistics*, 22(2), 173-195.
- Brems, L. (2003). Measure noun constructions: An instance of semantically-driven grammaticalization. *International Journal of Corpus Linguistics*, 8(2): 283-312.
- Cowie, A. P. (1988). Stable and creative aspects of vocabulary. *Vocabulary and Language Teaching*, 139.
- Ellis, R. (1997). *Second language acquisition*. The United States: Oxford.
- Gui, S. C. (2007). Uncertainty judgment and the acquisition of delexicalized verbs by Chinese learners of English. *Foreign Language Teaching & Research*, 39(1), 3-12.
- Mareček, D. (2016). Delexicalized and minimally supervised parsing on universal dependencies. *International Conference on Statistical Language and Speech Processing* (pp. 30-42).
- Nekrasova, T. (2016). EFL learners’ use of question constructions over time: Patterns and proficiency effects. *System*, 58, 82-96.
- Qin, W. (2009). Corpora research of language transfer in CEM. *Modern Educational Technology*, 11.
- Selinker, L. (1972). Interlanguage. *Applied Linguistics in Language Teaching*, 10, 209-232.
- Sinclair, J. (1988). A lexical syllabus for language learning. *Vocabulary and Language Teaching*, 140-160.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Shanghai Foreign Languages Education Press.
- Søgaard, A., & Wulff, J. (2012). An empirical study of non-lexical extensions to delexicalized transfer. *Proceedings of COLING 2012, Posters* (pp. 1181-1190).

Tognini, E. (2001). *Corpus linguistics at work*. Philadelphia, Amsterdam: J. Benjamins.

Yuan, H. (2014). A corpus-based study on the influence of L1 on EFL learners' use of prepositions. *Theory and Practice in Language Studies*, 4(12), 2513.