# A Computer Corpus-based Study of FOREIGN in English Newspapers and its Pedagogical Implications

Fei Deng
School of Foreign Studies, South China Agricultural University, Guangzhou, CHINA

Xiang Li
English Department, Guangdong Police College, Guangzhou, CHINA

**ABSTRACT**
This study adopts the methodology of corpus-based analysis and Contrastive Interlanguage Analysis (CIA) and uses two corpora (Chinese English Newspapers Corpus and American English Newspapers Corpus) as the source of data to make an empirical study on the use of the word, *foreign*. Major findings show that firstly Chinese English newspapers tend to overuse *foreign* and collocations of *foreign* are more diversified in Chinese English newspapers than those in native English newspapers, and secondly Chinese English newspapers focus on the neutral sense of *foreign* while neglecting its derogative meanings. Based on the above findings, pedagogical implications have been stated at last.

**Keywords:** corpus-based study, *foreign*, English newspapers, Contrastive Interlanguage Analysis, pedagogical implication

## INTRODUCTION

A corpus is defined as "a collection of examples of language in use that are selected and compiled in a principled way" (Huang & Yao, 2015; Duan, 2017; Sadi, 2017) and corpus linguistics as linguistic studies of such corpora. Corpus linguistics is considered a new scholarly enterprise in current years. Over the last forty years, there has been a great development in the compilation and analysis of corpora stored in computerized databases. According to Biber (2000), a corpus-based approach to linguistic studies and language teaching possess many advantages with which the traditional approaches cannot compete. Firstly, computers make it possible to identify and analyze complex patterns of language use; secondly, they allow the storage and analysis of a larger database of natural language than could be dealt with by hand; thirdly, they provide consistent, reliable analyses---they do not suffer from mental fatigue or indecisiveness; fourthly, they can be used interactively, allowing the human analyst to make difficult linguistic judgments while the computer manages all the record-keeping.

With its rapidly growing economy, China nowadays is more deeply involved in global cooperation and development than ever before. Therefore, a window for the world to learn more about China is greatly needed. Today Chinese English newspapers have become one of the major ways for the world to know more about China, so scholars now are more concerned about how Chinese English newspapers can express the intended messages effectively.

Yu Xi (2006) conducted a study on the usage of a specific word *foreign* to see whether there were differences in the use of *foreign* between Chinese English news reports and native English news reports. His research found the

**Correspondence:** Fei Deng, *School of Foreign Studies, South China Agricultural University, Guangzhou, China.*
✉ faydra@126.com

**Contribution of this paper to the literature**

- The present study uses corpus approach to analyze the word FOREIGN based on a large quantity of data, which is innovative in that the findings are more reliable compared with the previous studies in this field which usually used a small number of data to make analysis. The research findings can extend the scope of earlier studies of English lexis.
- The findings explored by the present study can offer pedagogical implications that English vocabulary teaching should focus on its all-round meaning so that learners can use words in appropriate situations.
- The corpus-based methodology used by the present study can be applied to help learners with English learning. Although textbooks and handbooks can offer meanings and usage of English words, there is little guidance as to how often and in what specific situations they should use these words to be able to approximate good native English writing. A large collection of native English corpus and software such as AntCont used in this study can fill this gap.

overuse of *foreign* in Chinese English news reports. However, there were some limitations in his study. He did not find out the specific reasons behind the overuse of *foreign*. More importantly, the material (English news report articles) which he used in the research has been out-of-date, and cannot reflect the current situation of the English news reports. Thus, there is an urgent need to collect the most updated data to conduct the research.

The present study aims to investigate the differences in the use of adjective *foreign* by China's English news reporters and native English news reporters. A corpus-based approach and Contrastive Interlanguage Analysis (CIA, hereafter) are adopted in the present study. Besides, a further study is conducted to explore the possible causes behind these differences. The present study is intended to help raise both teachers' and learners' awareness of the understanding and proper use of the word *foreign* in English writing.

## RESEARCH DESIGN

### Material

In this study, the researchers compile two new corpora to study the use of *foreign*. One is the American English newspapers corpus (hereafter AEC), which contains 366,221 words of the news reports from five major American English newspapers (The New York Times, The Washington Post, The Boston Globe, Los Angeles, Chicago Tribute). The other one is the Chinese English newspapers corpus (hereafter CEC), which gathers a total number of 304,682 words from four major Chinese English newspapers (China Daily, Beijing Review, Xinhua News, 21st Century). All the news reports were collected from 2011 to 2016, which the researchers believe can reflect the current trend of the word usage. The following are the descriptive data of the two corpora used in the present study.

### Instrumentation

**Table 1.** The Descriptive Data of the Two Corpora Used in the Present Study

| Corpus | AENC | CENC |
|---|---|---|
| Newspapers | The New York Times, The Washington Post, The Boston Globe, Los Angeles, Chicago Tribute | China Daily, Beijing Review, Xinhua News, 21st Century |
| Size of corpus | 366,221 words | 304,682 words |

In this research, AntConc3.4.3 is used as the search tool in linguistic analysis. The program functions include wordlist, concordance, collocates, keyword list, concordance plot, file view and clusters. Moreover, online calculation instrument---Log-likelihood Calculation is used to test the keyness of the word *foreign*.

**Table 2.** Overall Frequencies of FOREIGN in CENC and AENC

|  | CENC | AENC |
|---|---|---|
| Total Number of Words in Corpus | 304,682 | 366,221 |
| Tokens | 397 | 178 |
| Tokens/10 000 Words | 13.03 | 4.86 |
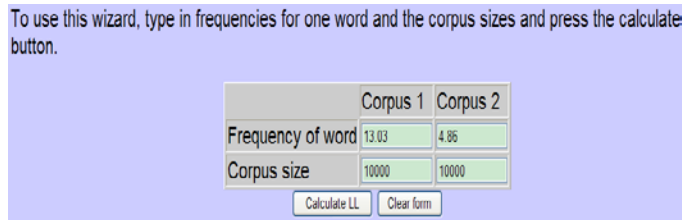


**Figure 1.** Log-likelihood Calculator



**Figure 2.** Log-likelihood Calculator Results of FOREIGN

## RESULTS AND DISCUSSIONS

### Overall Frequencies of FOREIGN

The overall frequency of *foreign* in the two corpora—CENC and AENC is first examined. **Table 2** shows the number of types and tokens of *foreign* in the two corpora. Since the corpora differ in size, the number of tokens per 10,000 words is also given.

**Table 2** shows that the Chinese English newspapers tend to use *foreign* more frequently than the native English newspapers (13.03 vs. 4.86 per 10,000), which might indicate that Chinese English newspapers overuse the word *foreign*. The researchers will then conduct a test from a statistical perspective, to see whether the observed overall frequencies in the two corpora do reveal something significant.

The current study employs an online test, Log-likelihood Calculation. Following the instructions, the researchers enter the plain numbers respectively shown in **Figure 1**, then press the button "Calculation LL", and the result appears within a second as is presented in **Figure 2**. **Figure 3** tells how this tool works.

**Figure 2** shows that G2 is 3.87, larger than G2 of 3.84 at the level of $p<0.05$ (as shown in **Figure 4**). And the "+" in **Figure 2** indicates the overuse in Corpus 1, the CEC, relative to Corpus 2, the AEC. The instructions in **Figure 4** show that the higher the G2 value is, the more significant is the difference between two frequency scores.

With the statistics shown above, the current research can draw a conclusion that the difference between word frequencies of *foreign* in the two corpora is of great significance (Ho et al., 2017). Next, a further study on *foreign* between the two corpora will be carried out.
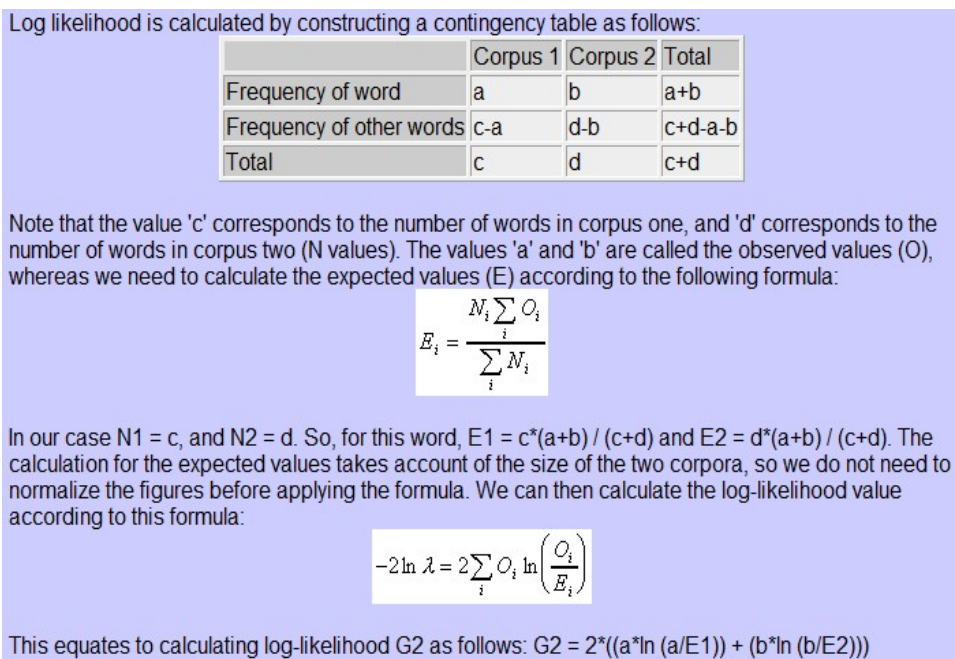
Log likelihood is calculated by constructing a contingency table as follows:

|  | Corpus 1 | Corpus 2 | Total |
| --- | --- | --- | --- |
| Frequency of word | a | b | a+b |
| Frequency of other words | c-a | d-b | c+d-a-b |
| Total | c | d | c+d |

Note that the value 'c' corresponds to the number of words in corpus one, and 'd' corresponds to the number of words in corpus two (N values). The values 'a' and 'b' are called the observed values (O), whereas we need to calculate the expected values (E) according to the following formula:

$$E_i = \frac{N_i \sum_i O_i}{\sum_i N_i}$$

In our case N1 = c, and N2 = d. So, for this word, E1 = c*(a+b) / (c+d) and E2 = d*(a+b) / (c+d). The calculation for the expected values takes account of the size of the two corpora, so we do not need to normalize the figures before applying the formula. We can then calculate the log-likelihood value according to this formula:

$$-2 \ln \lambda = 2 \sum_i O_i \ln \left( \frac{O_i}{E_i} \right)$$

This equates to calculating log-likelihood G2 as follows: G2 = 2*((a*ln (a/E1)) + (b*ln (b/E2)))

**Figure 3.** The Way Log-likelihood Calculator Works

**Note:** (thanks to Chris Brew) To form the log-likelihood, we calculate the sum over terms of the form x*ln(x/E). For strictly positive x it is easy to compute these terms, while if x is zero ln(x/E) will be negative infinity. However the limit of x*ln(x) as x goes to zero is still zero, so when summing we can just ignore cells where x = 0. Calculating ln(0) returns an error in, for example, MSExcel and the C-maths library.

The higher the G2 value, the more significant is the difference between two frequency scores. For these tables, a G2 of 3.8 or higher is significant at the level of p < 0.05 and a G2 of 6.6 or higher is significant at p < 0.01.

- 95th percentile; 5% level; p < 0.05; critical value = 3.84
- 99th percentile; 1% level; p < 0.01; critical value = 6.63
- 99.9th percentile; 0.1% level; p < 0.001; critical value = 10.83
- 99.99th percentile; 0.01% level; p < 0.0001; critical value = 15.13
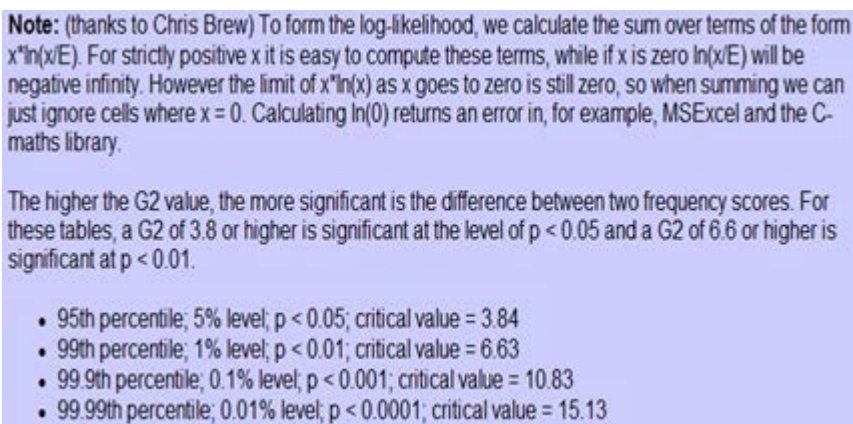
**Figure 4.** The Way Log-likelihood Calculator Works

## The Collocations of Adjective FOREIGN in CENC and AENC

This section will examine the collocations of *foreign* in the two corpora and try to find out whether there is any difference between the two corpora in the collocation of this word.

First of all, AntCon3.4.3 is used to search the top ten collocations of *foreign* in CENC and AENC. The top ten collocations of *foreign* in CENC and AENC are listed in the following table.

**Table 3.** The Top Ten Collocations of Adjective *FOREIGN*

| List | CEC | Tokens | % | AEC | Tokens | % |
|------|-----|--------|---|-----|--------|---|
| 1 | exchange(s) | 44 | 11.1 | minister(s) | 22 | 12.4 |
| 2 | minister(s) | 39 | 9.8 | policy | 15 | 8.4 |
| 3 | investment(s) | 27 | 6.8 | ministry/ ministries | 13 | 7.3 |
| 4 | company/companies | 24 | 6.0 | exchange | 10 | 5.6 |
| 5 | ministry | 23 | 5.8 | journalists | 8 | 4.5 |
| 6 | country/countries | 20 | 5.0 | investment | 7 | 3.9 |
| 7 | affair(s) | 18 | 4.5 | governments | 6 | 3.4 |
| 8 | policy/policies | 16 | 4.0 | relation(s) | 5 | 2.8 |
| 9 | investor(s) | 13 | 3.3 | secretary | 4 | 2.2 |
| 10 | trade | 11 | 2.7 | affair investor(s) | 3 | 1.6 |

From the above table, it can be see that there are some similarities in the use of frequent collocations of *foreign* in the two corpora. Among the top ten collocations, there are six collocations appearing in both corpora. They are: *foreign* minister(s), *foreign* exchange(s), *foreign* policy / policies, *foreign* investment(s), *foreign* affair(s), and *foreign* investors. The researchers also find that in CEC and AEC, most of *foreign*'s collocations are noun phrases. But when it comes to the tokens and types, it is found that CEC's three frequently used collocations: *foreign company (companies), foreign country (countries)* and *foreign trade* are rarely seen in AEC. Furthermore, in CEC there are 87 types of nouns which are modified by *foreign*, while in AEC, there are 46 types of nouns modified by *foreign*. Generally speaking, native English speakers tend to use words in a more flexible way, and the collocations of the words may appear to be more varied. But with regard to the types of nouns modified by *foreign* in the two corpora, the Chinese English newspapers collocate *foreign* with more various kinds of nouns. Unexpectedly, the native English newspapers use fewer types of nouns. Thus, in terms of the meaning of *foreign*, it is assumed that the native English newspapers might replace *foreign* with other words to express the similar meaning. So in the next step, a comparison on the most frequent collocations of *foreign* will be made between CEC and AEC.

## Specific Analysis of the Top Ten Most Frequent Collocations of FOREIGN in CENC and AENC

As discussed before, the collocations *foreign exchange(s), foreign minister(s), foreign ministry/ministries, foreign policy/policies, foreign affair(s)* and *foreign investor(s)* appear in the top ten frequent collocation lists in both corpora. In these high frequent collocations, except for the fixed phrase *foreign exchange* which means a system of buying and selling, the word *foreign* in the other five collocations is defined in *Oxford Advanced Learner's Dictionary* (Hornby, 2005) as follows:

1. of, in or from a country or an area other than one's own;
2. dealing with or involving other countries.

However, as mentioned before, CEN's top three most frequent collocations *foreign company/companies, foreign country/countries*, and *foreign trade*, are rarely seen in AEC. Why do native English newspapers seldom use these collocations? According to *Chamber Maxi Paperback Dictionary* (Schwarz, 1992), *foreign* has other definitions: alien; extraneous; not belonging; unconnected; not appropriate.

Apparently, the meanings have a derogative tincture, which can be found in the following example in AEC:

*But this small, working-class neighborhood, known for its sometimes kitschy embrace of Old World pride, seems determined to restore a connection with a country few of its residents have visited, a language that fewer speak, and a culture that has already grown increasingly foreign.*

The sentence above is taken from a news report: *Reclaiming An Identity Washed Out In A Flood* in *New York Times* on March 20, 2011, which is about a small Czech community living in Iowa and trying to protect their own culture from disappearing. Considering the context of this news report, "*foreign*" means that the Czech culture has become strange when people have lost the sense of belonging to their own culture.

From the example above, it can be seen that native English newspapers sometimes use *foreign* with negative connotations, while in CEC the researchers cannot find such kinds of usage. This is one of the differences on the use of *foreign* between the two corpora in terms of the word's semantic meaning. In the next step, the researchers will continue to analyze another difference on the use of *foreign* between CEC and AEC.

## Specific Analysis of the Frequent Collocations of FOREIGN unique in CEC

As mentioned before, in CEC, the collocations *foreign company/companies, foreign country/countries*, and *foreign trade* are widely used, while in AEC they are rarely seen. So, here come the questions: why do Chinese English newspapers overuse these collocations? What other expressions do native English newspapers use when they want to express similar meanings? The following part will focus on the above questions and make a specific analysis of these three collocations.

1. The first collocation is "*foreign* company/companies", which refers to companies that belong to other countries or companies with investments by other countries in China. In China, "foreign company/companies" has become a common phrase after 1978 due to economic policies. From then on, companies from developed countries flooded into China to develop new markets. The large amount of *foreign* capital brought by these companies has greatly boosted China's economy. Since then, "*foreign* company/companies", which is the English translation of the Chinese term "waiguo gongsi" or "waizi gongsi", has been widely used around China. But such a popular phrase is rarely found in AEC. In order to get more details, the researchers study the synonyms of *company*: *firm* and *corporation*, to see whether there are "*foreign* firm/firms" or "*foreign* corporation/corporations" used in AENC. The searching result displays one for "*foreign firms*". Therefore, it is clear that native English news reports seldom use this collocation. After trying another search words for the similar meaning with "*foreign company/companies*" in AEC, the researchers get 3 instances of "*international firms*", 3 instances of "i*nternational company/companies*", and 1 instance of "*international corporations*", which have similar meaning to "*foreign company/companies*". According to Collins Cobuild Dictionary (Sinclair, 2000), "international" is defined as "*between or involving different countries*". Clearly, this definition is neutral, and by referring to other dictionaries, the researchers get similar definitions. Thus it is easy to understand why native English newspapers prefer to use "international".
2. The second collocation is "*foreign country/countries*" in CEC, which cannot be found in AEC. So how do native English newspapers express the same meaning? the researchers search for the frequent collocations of "country/countries" in AENC, and find that there are 30 "other country/countries" used to express the same meaning. It seems that "other country/countries" might have less subjective bias.
3. As for "*foreign* trade", the researchers find that there are several similar expressions in CEC. They are "i*nternational trade*", "*global trade*", "*multilateral trade*", and "*world trade*". However, in AEC, there are only 3 instances of "*international trade*", which might be a substitution for "*foreign* trade". The cause of overuse of "*foreign* trade" is similar to the case of "*foreign company/companies*". With the rapid development of economy and new technology in China, especially the One Belt, One Road Initiative, international trade has become very popular among countries, which, to some extent, leads to the overuse of "*foreign* trade" in China.

After the above analysis of "*foreign company/companies*", "*foreign country/countries*" and "*foreign trade*", the current study concludes that there are external and internal factors which lead to the overuse of these collocations. As to the external factors, the Reform and Opening-up policy allows other countries' investments in China, and the Chinese people may have greater chances to know more about "*foreign*" things and use this word based on these

cross-border economic activities. For the internal factors, they may come from a linguistic perspective. In Chinese English dictionaries, the word *foreign* does not suggest anything negative. However, in some native English dictionaries, this word has negative meanings.

## Possible Causes of the Results

The possible reasons for the overuse of *foreign* in Chinese English newspapers might be explained from the two perspectives.

Social background is considered as the first reason. It is well known that China nowadays has much more frequent contact with other countries in many fields, especially in the areas of economy and culture. Since there are much exchange and international cooperation with the outside world, the Chinese phrase "waiguo" (other countries), which is translated as "*foreign*", is commonly seen in Chinese mass media when they describe the things dealing with or involving other countries. As a result, the wider the Chinese phrase "waiguo" is used, the more frequently *foreign* might appear in Chinese English newspapers, hence the overuse of it.

The second reason is explained from the linguistic perspective. In most English-Chinese bilingual dictionaries, the first definition of *foreign* is "waiguode", which does not carry any negative meaning. After studying all the collocational patterns of *foreign* in CEC, the researchers found that *foreign* is mostly used in the first definition. That is to say, Chinese English newspapers focus on the core meaning of *foreign* while overlooking its other meanings. Thus, the current study can conclude that the lack of all-round understanding of *foreign* may result in its overuse in Chinese English news reports.

## PEDAGOGICAL IMPLICATIONS

The major findings are summarized as follows:

Firstly, Chinese English newspapers tend to use *foreign* more frequently than English native newspapers and they focus on the neutral meaning of *foreign* while neglecting its derogative meanings. Secondly, Chinese English newspapers use *foreign* in more variety of situations to express the meaning of between or involving different countries, while English native newspapers use international to express the similar meaning.

These findings bring some pedagogical implications for English teaching in China.

Firstly English teaching should call the learners' attention to the all-round meaning of *foreign*, especially its derogative meaning; Secondly learner should be made aware to use more appropriate word like international to express the intended meaning in some situations. Lastly but least, English learner can be taught to use native English corpus to guide their writing. Textbooks and handbooks can offer the meanings and usage of the English words, but there is little guidance as to how often and in what specific situations they should use these words to be able to approximate good native English writing. A large collection of native English corpus and software such as AntCont used in this study, makes complements for the textbooks. While the concordance from the learner corpus provides the actual illustration for the learners' typical problems, the concordance from the native corpus provides resources of reference when the teachers or learners need information concerning the actual use of the focused usage by the native English speakers.

## ACKNOWLEDGEMENTS

## REFERENCES

Biber, D., Conrad, S., & Rappen, R. (2000). *Corpus linguistics: Investigating language structure and use*. Beijing: Foreign Language Teaching and Research Press, 4.

Duan, Q. (2017) A Study of the Influence of Learning Organization on Organizational Creativity and Organizational Communication in High Tech Technology. *Eurasia Journal of Mathematics Science and Technology Education, 13*(6) 1817-1830. doi:10.12973/eurasia.2017.00699a

Ho, S., Chen, W., & Hsu, W. (2017). Assessment System for Junior High Schools in Taiwan to Select Environmental Education Facilities and Sites. *Eurasia Journal of Mathematics Science and Technology Education, 13*(5), 1485-1499. doi:10.12973/eurasia.2017.00681a

Hornby, A. S. (2005). *Oxford Advanced Learner's Dictionary*. Oxford: Oxford University Press, 454.

Huang, C., & Yao, Y. (2015). Corpus Linguistics. *International Encyclopedia of the Social & Behavioral Sciences*, 949-953. doi:10.1016/B978-0-08-097086-8.52004-2

Jacobs, G. J., & Durandt, R. (2017) Attitudes of Pre-Service Mathematics Teachers towards Modelling: A South African Inquiry. *Eurasia Journal of Mathematics Science and Technology Education, 13*(1), 61-84. doi:10.12973/eurasia.2017.00604a

Sadi, O. (2017) Relational Analysis of High School Students' Cognitive Self-Regulated Learning Strategies and Conceptions of Learning Biology. *Eurasia Journal of Mathematics Science and Technology Education, 13*(6) 1701-1722. doi:10.12973/eurasia.2017.00693a

Schwarz, C., & Klein, V. (1992). Chambers maxi paperback dictionary. *Chambers*, 678.

Sinclair, J. (2000). Collins cobuild English dictionary. *Shanghai Foreign Language English Press*, 822-823.

Xi, Y. (2006). A corpus based study of foreign in the chinese English news report. *Foreign Language Teaching, 6*, 23-26. doi:10.16362/j.cnki.cn61-1023/h.2006.06.019

# http://www.ejmste.com