A Corpus-Based Contrastive Study on Semantic Prosody of the Noun "Issue" in CLEC and COCA

Wentao Wang*

Lanzhou University of Technology, Lanzhou 730000, Gansu Province, China

***Corresponding author:** Wentao Wang, wangwtedu@sohu.com

Abstract: The study of semantic prosody has gradually become a focus of research in linguistics. By retrieving and comparing the noun "issue" in CLEC (Chinese Learner English Corpus) and COCA (Corpus of Contemporary American English), this paper intends to explore the similarities and difficulties on the semantic prosody of the noun "issue" between Chinese English learners and Americans.

Keywords: Semantic prosody; Corpus-based analysis; Contrastive interlanguage analysis

1. Research background

In the 20 years after Louw [1] first proposed the concept of semantic prosody, corpus linguists have conducted countless studies on semantic prosody. However, a contrastive study on the semantic prosody of English with different language groups is still lacking [2,3]. On the basis of adopting a corpus-based contrastive interlanguage analysis [4], this paper aims to explore the semantic prosody features of the noun "issue" used by Chinese English learners and Americans.

2. Standardized frequency of a node word

CLEC has 1,070,602 tokens, whereas COCA has 533,788,932 tokens. Hence, it is necessary to primarily determine the standardized frequency of the noun "issue." In order to standardize the frequency of a node word, the number of that node word in one corpus is first divided by the total token number of that corpus to obtain its frequency [5]. Subsequently, the outcome is multiplied by 1,000,000, thus obtaining the standardized frequency of that node word. The aforementioned standardization process is demonstrated as follows: standardized frequency of a node word = observation frequency of that node / token number of its corpus * 1,000,000 [6].

To begin with, the CLEC.txt file was imported into AntConc 3.5.9. After inputting "issue*" as a search term into AntConc 3.5.9, there was a total of 75 hits, with the results of "issue," "issues," and "issued." Considering that this thesis is a study of nouns, all lexical verb results of "issue" and "issued" were removed. After the manual deletion of only three verbs, 72 hits were left. Subsequently, "[issue*].[n*]" was typed into COCA's search line on the list page, yielding a total of 362,190 hits. The more specific results were "issue" with 189,257 hits and "issues" with 172,933 hits.

Abiding by the aforementioned computing method, which is shown in **Table 1** and **Table 2**, the standardized frequency of "issue" in CLEC and COCA can be obtained (the numbers shown below are rounded off to the nearest hundredth).

Table 1. The node word "issue" in CLEC

Word	Observation of the node word	Standardized frequency of the node word (per million tokens)
"issue"	72	67.25

Table 2. The node word "issue" in COCA

Word	Observation of the node word	Standardized frequency of the node word (per million tokens)
"issue"	362,190	678.53

It can be seen that the standardized frequency of the node word "issue" in COCA is nearly 10 times (678.53/68.19) as many as that in CLEC. There is a rather huge difference between Chinese English learners and Americans in regard to the usage of the noun "issue." In order to further clarify this difference, a chi-square test can be performed on the two groups by utilizing the Chi-Square and Log-Likelihood Calculator designed by Beijing Foreign Studies University, as shown in **Figure 1**.
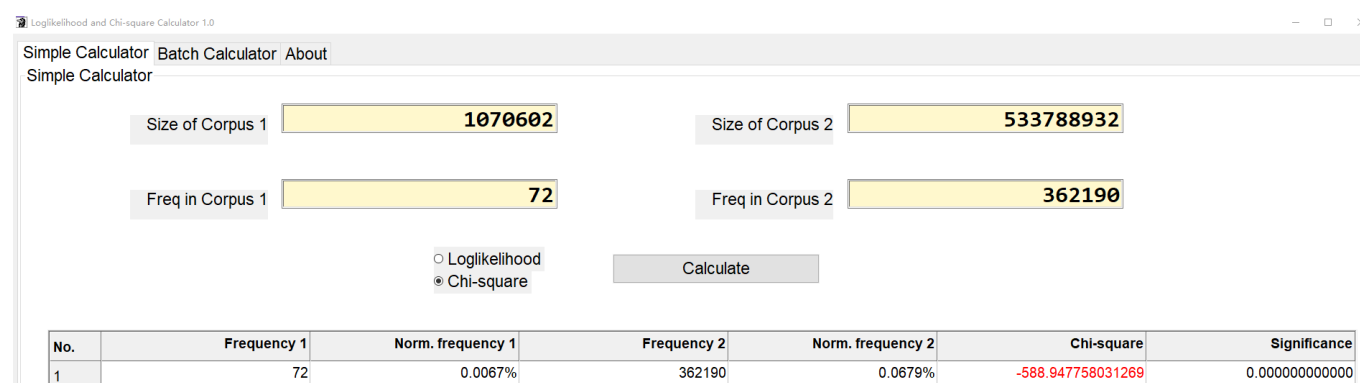

Figure 1. A chi-square test on the observation number of the node word "issue" in CLEC and COCA

In the chi-square test, a "-"value means that the usage of the node word "issue" is relatively less between the two compared corpora. Besides, when $|X|^2 > 10.83$ and $p < 0.001$, the credibility of the data is 99.99% [7]. It is believed that the difference between the two groups of comparative data is quite significant. We can conclude that there is an obvious variance in the usage of "issue" between Chinese English learners and Americans. Furthermore, Chinese English learners are less likely to use the noun "issue" compared to Americans.

3. Semantic prosody of the noun "issue" in CLEC

Looking back at the retrieved results of the noun "issue*" in CLEC on AntConc 3.5.9, there are 72 index lines of the noun "issue" in CLEC.

(1) "Euthanasia, or mercy killing, has long been a controversial *issue* all over the world." (Negative semantic prosody)
(2) "The *issue* also confronts Chinese." (Negative semantic prosody)
(3) "Only when we make it legalized can we map out detailed laws concerning this *issue* and establish special departments to perform it and thus preventing it from being abused." (Negative semantic prosody)
(4) "Yet it is a rather controversial *issue* because of the involvement of artificial actions." (Negative semantic prosody)
(5) "For instance, they may read newspapers and magazines about current *issues*." (Neutral semantic prosody)

Among the 72 index lines, 7 index lines are of positive semantic prosody, 22 index lines are of neutral semantic prosody, and 43 index lines are of negative semantic prosody. Obviously, among Chinese English learners, the noun "issue" is mostly used in expressions of negative semantic prosody, accounting for 59.72%, followed by neutral semantic prosody, which accounts for 30.56%. **Table 3** gives a clear summary of the distribution of the semantic prosody of the noun "issue" among Chinese English learners.

Table 3. Semantic prosody of the noun "issue" in CLEC

	Positive semantic prosody	Neutral semantic prosody	Negative semantic prosody
Frequency (sum = 72)	7	22	43
"issue(s)"	9.72%	30.56%	59.72%

With the aim of figuring out its semantic prosody, BFSU Collocator 1.0, a specialized corpus tools designed by Beijing Foreign Studies University, is required to analyze its collocation strength by retrieving collocates of the noun "issue."

The mutual information (MI) value reflects the mutual attractive relations between the node word and its collocates. The greater the MI value, the stronger the collocation strength [8]. The MI value can be positive or negative. A negative MI value indicates that the two words mutually exclude each other, implying that the two words do not frequently co-occur in the same context. According to Hunston [9], a MI value of 3 is considered as the critical value in practice, and collocates with a MI value equal or greater than 3 are considered to have substantial collocation force. Nevertheless, because CLEC is a relatively small corpus, it is acceptable to set a MI value of 2 as the minimum collocation frequency [10]. Therefore, this thesis sets a MI value of 2 as the minimum frequency with a span of 5 (-5/+5) to observe the collocation strength of the noun "issue." The high-frequency collocates of the noun "issue" are shown in **Table 4**.

Table 4. High-frequency collocates of the noun "issue"

Ranking	Collocates	F(c)	F(n,c)	MI
1	agreement	1	1	11.8521
2	contrive	1	1	11.8521
3	manufacture	1	1	11.8521
4	lastest	4	3	11.4371
5	latest	5	3	11.1151
6	attracts	2	1	10.8521
7	posting	2	1	10.8521
8	territory	3	1	10.2671
9	BBC	34	6	9.3496
10	employment	6	1	9.2671
…	…	…	…	…

Among the 49 collocates of the noun "issue" with MI greater than 2, 45 of them are of neutral semantic prosody, accounting for 91.84%. The number of collocates of the noun "issue" with neutral semantic prosody and negative semantic prosody is 2, accounting for 4.08%, respectively. For now, it can be concluded that the semantic prosody of the noun "issue" gravitates toward negative prosody and neutral prosody.

4. Semantic prosody of the noun "issue" in COCA

There is a total of 362,190 results of the noun "[issue].[n*]" in COCA; hence a random sampling method is adopted to select 100 index lines from 362,190 results [11]. According to Wei Naixing [12], it is unrealistic to perform random sampling in a large corpus, especially when the number of words and sentences in the corpus is increasing on a daily basis. As recommended, a target index line will be selected after every five index lines. Some examples of the selected index lines are shown below.

(1) "...when the e-mail arrives but it makes raising a simple *issue* quite frustrating these days..." (Negative semantic prosody)
(2) "And, folks, that is the crux of this whole *issue*." (Neutral semantic prosody)
(3) "This *issue* is more in the hands of men than women and in that sense men would be." (Neutral semantic prosody)
(4) "See the abortion *issue* has passed into history by the mainstream of society it is now entrenched in law." (Negative semantic prosody)
(5) "...or if it only is the kindred-requirement for some hytbold-pieces that is the *issue*." (Neutral semantic prosody)

Among the 100 index lines, five lines are of positive semantic prosody, 80 lines are of neutral semantic prosody, and the remaining five lines are of negative semantic prosody. The distribution of the semantic prosody of 100 index lines in COCA is shown in **Table 5**.

Table 5. Semantic prosody of the noun "issue" in COCA

	Positive semantic prosody	Neutral semantic prosody	Negative semantic prosody
Frequency (sum = 100)	5	80	15
"issue(s)"	5%	80%	15%

Overall, the semantic prosody of the noun "issue" in COCA gravitates toward neutral semantic prosody, accounting for 80%, which is quite different from Chinese English learners. Among Chinese English learners, negative semantic prosody has a dominant position in the use of the noun, rather than neutral semantic prosody, which only accounts for 30.56%.

The collocate strength of the noun "issue" in COCA is determined. "[issue].[n*]" is typed in the Word/Phrase search line, and "*" is typed in the collocates line, with a +5/-5 retrieving range selected. The minimum mutual information is set to 3. The high-frequency collocates of the noun "issue" are shown in **Table 6**.

Table 6. High-frequency collocates of the noun "issue"

	Word	FREQ	ALL	%	MI
1	address	6121	93062	6.58	4.17
2	related	3409	87540	3.89	3.42
3	environmental	3323	80344	4.14	3.50
4	raised	3018	82233	3.67	3.33
5	addressed	2290	25717	9.66	4.73
6	discuss	2483	47745	5.20	3.83
7	abortion	2290	38199	5.99	4.04

(Continued on next page)

	Word	FREQ	ALL	%	MI
8	addressing	1907	15699	12.15	5.06
9	facing	1880	40660	4.62	3.66
10	dealing	1875	43471	4.31	3.56
11	immigration	1818	39508	4.60	3.66
12	moral	1770	59107	2.99	3.04
13	resolve	1714	17459	9.82	4.75
14	discussed	1502	36721	4.09	3.49
15	controversial	1279	20906	6.12	4.07
16	resolved	1277	13486	9.47	4.70
17	ethical	1258	18373	6.85	4.23
18	sensitive	1118	28144	3.97	3.45
19	raises	1044	18867	5.53	3.92
20	pressing	1026	13916	7.37	4.34
…	…	…	…	…	…

In the top 100 high-frequency collocates of the noun "issue", 80 collocates are of neutral semantic prosody, 16 collocates are of negative semantic prosody, and 4 collocates are of positive prosody. The distribution of semantic prosody of the top 100 collocates of "issue" in COCA is shown in **Table 7**.

Table 7. Distribution of the semantic prosody of the top 100 collocates of "issue" in COCA

	Negative semantic prosody	Neutral semantic prosody	Positive semantic prosody
Frequency (sum = 100)	16	80	4
Percentage	16%	80%	4%

5. Conclusion

In conclusion, the collocates of the noun "issue" are mostly neutral semantic prosody, which is the same as Chinese English learners. However, in the index lines of the noun "issue" between CLEC and COCA, Chinese English learners tend to view the noun "issue" as a word with negative semantic prosody, such as "a controversial 'issue'," and "a serious moral 'issue'." However, Americans usually use this word with a neutral semantic prosody, such as "an 'issue' that is worth mentioning."

Disclosure statement

The author declares no conflict of interest.

References

[1] Louw B, 1993, Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies, in Baker M, Francis G, Tognini-Bonelli E (eds), Text and Technology: In Honour of John Sinclair, John Benjamins, Amsterdam.

[2] Sinclair J, 1996, Corpus, Concordance, Collocation, Oxford University Press, Oxford.

[3] Sinclair J, 1996, The Search for Units of Meaning. Textus, 9(1): 75-106.

[4] Hoey M, 2000, A World Beyond Collocation: New Perspectives on Vocabulary Teaching, in Lewis M (ed), Teaching Collocations, Language Teaching Publications, Hove.

[5] Sinclair J, (ed) 1995, Collins COBUILD English Dictionary, Collins, London.

[6] Nelson M, 2006, Semantic Associations in Business English: A Corpus Based Analysis. English for Specific Purpose, 25(2): 217-234.

[7] Hunston S, 2002, Corpora in Applied Linguistics, Cambridge University Press, Cambridge.

[8] Channell J, 2000, Corpus-Based Analysis of Evaluative Lexis, in Hunston S, Thompson G (eds.), Evaluation in Text: Authorial Stance and the Construction of Discourse, Oxford University Press, Oxford, UK, 38-55.

[9] Huston S, Francis G, 2000, Pattern Grammar – A Corpus-Driven Approach to the Lexical Grammar of English, John Benjamins, Amsterdam and Philadelphia, 5-137.

[10] Deng YC, Wang TS, 2005, Statistical Method and Computer Implementation of Word Collocation Extraction. Audio-visual Teaching of Foreign Languages, 2005(5): 25-28.

[11] Benson M, Benson E, Ilson RF, 1986, The BBI Combinatory Dictionary, John Benjamins Publishing House.

[12] Wei N, 2002, Corpus-Based and Corpus-Driven Approaches to the Study of Collocation. Contemporary Linguistics, 2: 101-114.