

## Original Paper

# Epistemic Lexical Verbs Based on the COCA Corpus: A Study of “Argue”, “Note” and “Claim”

Bingshu Yin<sup>1</sup>

<sup>1</sup> Department of Foreign Languages, Inner Mongolia University of Technology, Hohhot, Inner Mongolia Autonomous Region, China 010051

Received: February 12, 2026

Accepted: April 28, 2026

Online Published: May 19, 2026

doi:10.22158/eltls.v8n3p80

URL: <http://dx.doi.org/10.22158/eltls.v8n3p80>

### **Abstract**

*This paper draws on the COCA corpus to conduct a comparative analysis of the usage characteristics of the epistemic lexical verbs “argue”, “claim” and “note”, examining dimensions such as frequency distribution, register differences, collocation features and colligations. The findings reveal that: (1) In terms of frequency and register distribution, “note” is the most frequently used, primarily concentrated in academic and expository registers; “argue” is active in both academic and interactional registers; whilst “claim” is more prevalent in media and online registers. (2) In terms of collocation features, “claim” frequently co-occurs with terms related to stance-taking groups and rights attribution; “argue” is more commonly paired with academic or professional subjects and argumentation-related vocabulary; whilst “note” primarily co-occurs with informative and explanatory vocabulary. (3) In terms of colligation, “claim” is commonly found in “that” clauses and noun phrases; “argue” primarily takes the form of “that” clauses and co-occurs with prepositional phrases; whilst “note” relies heavily on “that” clause structures. Overall, there are systematic differences in the usage patterns of these three verbs within discourse. Research indicates that the use of epistemic lexical verbs is constrained by multiple factors, including register, collocation and colligation.*

### **Keywords**

*Epistemic lexical verbs, Synonym differentiation, Second language writing, Corpus*

### **1. Introduction**

For second language learners, vocabulary acquisition is not merely a process of accumulating vocabulary, but also involves grasping the deep semantics and pragmatic features of words. The English language includes a considerable number of synonyms; despite the fact that they share conceptual similarity, in real communication their meaning may differ significantly depending on factors related to

collocation, register, and discourse. As Li Wenzhong states, the meaning of language is reflected in practical application, and the understanding of words must be combined with their co-occurrence relationship and context (Wenzhong, 2010). Whether the words are properly selected will directly affect the accuracy of meaning expression and the degree of discourse standardization. Hence, it is essential to grasp systematically the characteristics of word usage. In academic writing, epistemic lexical verbs have a vital role as they not only have the function of conveying information but can also serve to indicate the stance of the author, control the tone, and so forth. When it comes to the linguistic aspects of usage, there are differences between epistemic lexical verbs in terms of co-textual occurrence and distribution, hence varying in degree of subjectivity and certitude. This makes it meaningful to make distinctions between their usages so as to enhance the writing skills of second-language learners in academic English.

Traditional ways of telling synonyms apart often rely on introspection, which does not sufficiently reflect how words vary in actual use. In contrast, corpus-based methods for studying synonym differences use large authentic corpora to examine word combinations and co-occurrences in specific contexts. This leads to a better understanding of each word's meaning and usage, and offers a more empirical way to distinguish synonyms from one another. Following this approach, the current study uses the COCA corpus to identify and examine three epistemic lexical verbs in academic discourse: claim, argue, and note.

## 2. Literature Review

An academic writing has traditionally been regarded as an objective discourse form (Hyland, 2017). However, advances in discourse studies indicate that academic writing is not entirely objective; authors not only convey propositional information but also construct their subjective stance within the text. Stance can be realised through various linguistic resources, such as modal verbs (may, might, could), stance nouns (possibility, assumption, argument), and epistemic lexical verbs (believe, suggest, infer) (Biber, 2004; Xie & Patten, 2025). Epistemic lexical verbs (ELVs) are used to express the speaker's degree of commitment to a proposition or the strength of an assertion, thereby not only reflecting their cognitive mode but also revealing the source of the relevant knowledge. Such verbs not only participate in the construction of propositional meaning but also express certainty and uncertainty by modulating the intensity of tone, thus fulfilling the pragmatic function of either limiting ambiguity or reinforcing the argument. In terms of classification, epistemic lexical verbs can be divided into judgemental and evidential categories: the former includes conjectural verbs (e.g., believe, assume) and inferential verbs (e.g., conclude, infer), which are used to express subjective judgements; the latter includes citation verbs (e.g., argue, claim) and perceptual verbs (e.g., seem, appear), which are used to indicate the source of information.

Existing research on epistemic lexical verbs has primarily focused on variations in their distribution across different disciplines, cultural contexts and genres. At the disciplinary level, Szczygłowska

employed a corpus-based approach to examine medical and psychological research papers, finding differences in frequency of use and typological preferences between the two. Psychological papers utilised a greater number of epistemic lexical verbs than medical papers, and authors of medical papers showed a preference for deductive epistemic lexical verbs (deductive ELVs) (Szczygłowska, 2023); In terms of cultural context, Dontcheva-Navratilova analysed the selection and use of epistemic lexical verbs in linguistic and economic research papers written by native English-speaking and Czech scholars (Dontcheva-Navratilova, 2018). The study revealed that non-native English-speaking authors differed from native speakers in both the frequency of epistemic lexical verb usage and the types selected, reflecting variations in rhetorical norms across different academic cultures; Xie & Patten, through a comparative analysis of Chinese and English learner corpora, found that Chinese learners tend to express a higher degree of certainty when using epistemic lexical verbs, and exhibit less balance in grammatical structures and sentence patterns compared to native English speakers. With regard to genre, studies by Vass and Ni Xiyang indicate that there are significant differences in the frequency of epistemic lexical verbs across different genres (Vass, 2017; Xiyang, 2008). In the above research, the corpus-based approach was used to analyze legal scholarly articles and legal judgments, with notable disparities in the distribution of epistemic lexical verbs noted. These differences correlate closely with the communicative purposes of the genres. In the latter study, the researchers made a comparison between academic articles and news corpora, finding out that the use of epistemic lexical verbs is more prevalent in academia and that there are significant differences in the distribution of these verbs compared to news discourse. Overall, existing research describes the features of the use of epistemic lexical verbs in several ways. However, little attention has been paid to the semantics of synonyms of epistemic lexical verbs, and few comparisons have been carried out in this area.

The demarcation of synonyms has always been an issue when acquiring a second language. Conventional semantic analysis is largely dependent on the subjective evaluation of the researcher, and the comparison between the definition of the word in dictionaries and example sentences. Thus, such techniques lack objectivity and validity. Recently, the application of corpus linguistics has allowed for evidence-based research on synonyms through their use features such as word frequency and collocation. Thus, their differences become more objective and intelligible (Xuelian, 2015). For example, Lin Yating and Yuan Sen undertook a corpus-based comparison of “outcome” and “consequence”. They found that both words have neutral connotations but the former tends to be positive while the latter is negative (Yating & Sen, 2016).

To conclude, although there is an abundant amount of literature on epistemic lexical verbs in academic discourse from different perspectives, there is still some room left regarding the systematic comparative analysis of synonyms from this specific group of epistemic markers. Specifically, there is not enough empirical evidence on the linguistic characteristics of high-frequency epistemic lexical verbs used in academic writing, which include such words as argue, claim, and note. In other words, there is still insufficient research on the linguistic properties of these verbs in their natural context. This study aims

at applying the COCA corpus to conduct a comparative analysis of these epistemic lexical verbs in terms of frequency, register, collocation, and colligation.

### **3. Research Design**

#### *3.1 Research Questions*

Based on the aforementioned research objectives, this paper proposes the following three research questions:

- (1) What differences exist between “argue”, “claim” and “note” in terms of frequency and register?
- (2) What differences exist between “argue”, “claim” and “note” in terms of collocation?
- (3) What differences exist between “argue”, “claim” and “note” in terms of colligation?

#### *3.2 Research Corpus*

The Corpus of Contemporary American English (COCA), created by Mark Davies, is currently one of the most widely used large-scale balanced English corpora. The corpus contains authentic language material from a variety of domains—including spoken language, fiction, magazines, newspapers and academic journals—dating from 1990 onwards. With a corpus size exceeding one billion words, it is characterised by its richness, balanced stylistic distribution and continuous updates. Furthermore, COCA offers convenient online search functions and supports multi-dimensional analysis, including word frequency statistics, register distribution, collocation analysis and colligation. Consequently, it is widely used in the study of synonym differentiation and pragmatic features.

#### *3.3 Research Process*

First, the headwords *note*.[v\* ], *argue*.[v\* ] and *claim*.[v\* ] were retrieved from the “list” and “chart” columns to calculate their overall frequency of occurrence and distribution across different domains; subsequently, the “collocates” column was utilised to examine the collocation characteristics of each headword and extract high-frequency collocation patterns. Building on this, the “word” function was employed to analyse the co-occurrence patterns of the three terms.

### **4. Results and Discussion**

#### *4.1 Word Frequency and Register*

According to the statistical results from the COCA corpus (see Figures 4-1, 4-2 and 4-3), the three key terms “note”, “argue” and “claim” exhibit significant differences in their overall word frequencies. Among them, “note” has the highest total frequency, followed by “argue”, whilst “claim” has the lowest frequency of use. This result indicates that, in actual language use, “note” is the most widely used and possesses strong general applicability, whilst “claim” is relatively restricted in its usage.

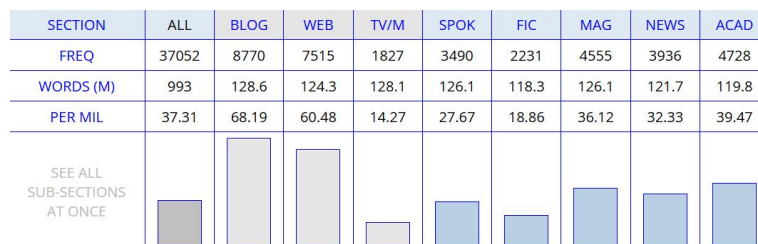


Figure 4-1. Word Frequency and Register Distribution of “Claim”

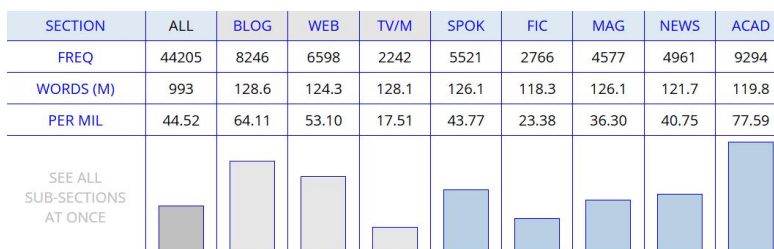


Figure 4-2. Word Frequency and Register Distribution of “Argue”

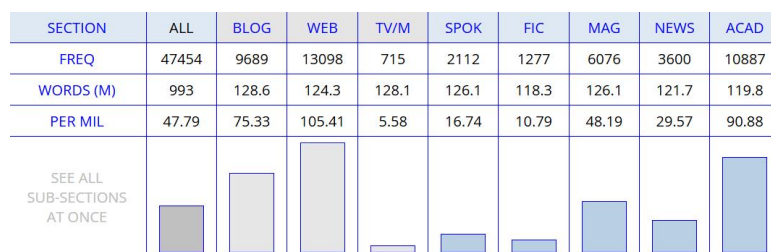


Figure 4-3. Word Frequency and Register Distribution of “Note”

Register is one of the major factors influencing vocabulary selection. The register mentioned here means the language variations resulting from different communicative contexts (conversational context, journalistic and academic writing, etc.) (Jidong & Ping, 2005). Concerning the register distribution, there are three different trends for the three terms. Specifically, the term “claim” is often used in blogs and online texts, with the lowest frequency being in the corpus of films and TV shows, suggesting its preference for general written language. “argue” appears mainly in academic and blog contexts, with the lowest frequency appearing in the corpus of films and TV shows, demonstrating its argumentative nature and high degree of interaction. By comparison, “note” can be used most frequently in online and academic discourse, and the least frequently used in the corpus of films and TV shows, suggesting the formality of its written language and presentation of information function. In general, though belonging to the same category of epistemic lexical verbs, the register distribution of the three terms varies significantly owing to their different pragmatic purposes.

The register distribution not only reflects the differences in the contexts where the vocabulary appears but also reveals their different semantic and pragmatic functions. As a whole, the differences of register distributions demonstrate that different vocabularies appear in different situations: “note” stresses objective descriptions and often appears in academic and expository writings; “argue” emphasizes

argumentative functions and often appears in academic discourses and interactions; “claim” is used mainly to provide information and positions and often appears in media or online discourses. This classification also suggests that synonyms have their own pragmatic functions and cannot be interchangeably used in practice.

#### 4.2 Collocation

Collocation refers to the habitual co-occurrence of lexical items in actual language use. Originally proposed by J.R. Firth as a relationship of association between words, it later developed into corpus-based linear co-occurrence and has gradually expanded to multi-dimensional research perspectives including syntax, discourse and cognition (Naixing, 2003). Based on this, this paper takes “claim”, “argue” and “note” as headwords, sets a span of four words in the COCA corpus, and selects collocations with an MI value greater than 3 for analysis.

From the high frequency collocations listed in Tables 4-1, 4-2, and 4-3, it is evident that there are notable distinctions between the collocational features of “claim”, “argue” and “note”. These distinctions are not only in relation to the types of collocates but also in their semantic and pragmatic roles. In general, the distribution of collocations in all the three cases shows that they are stable, showing both their semantic distinction and functional orientation.

**Table 4-1. The High-Frequency Combination of “Claim”**

Table 4- 1 The high-frequency combination of ‘claim’

No.	Collocation	Frequency	MI
1	credit	91182	4.26
2	victory	50295	4.44
3	critics	36,544	4.50
4	title	72,434	3.12
5	fame	21,568	4.58
6	ownership	18,062	4.75
7	prize	24859	4.27
8	represent	42,298	3.38
9	experts	51,527	3.09
10	Christians	27,463	3.84
11	Deduction	4052	6.41
12	advocates	17,960	4.25
13	supporters	28,263	3.60
14	falsely	3366	6.35
15	conservatives	25,462	3.26

**Table 4-2. The High-Frequency Combination of “Argue”**

Table 4- 2 The high-frequency combination of 'argue'

No.	Collocation	Frequency	MI
1	critics	36544	5.46
2	advocates	17,960	5.14
3	authors	38,512	3.91
4	lawyers	40,220	3.82
5	experts	51,527	3.43
6	proponents	6015	6.51
7	scholars	22008	4.58
8	opponents	20,801	4.60
9	supports	28,263	3.87
10	economists	11,402	4.98
11	lawyers	16,355	4.09
12	analysts	20160	3.77
13	conservatives	25,462	3.40
14	logic	24110	3.47
15	merits	6554	5.13

**Table 4-3. High-Frequency Collocation of “Note”**

Table 4- 3 High-frequency collocations of 'note'

No.	Collocation	Frequency	MI
1	important	331830	4.78
2	please	247051	5.05
3	interesting	117,181	5.03
4	medical	136,521	3.92
5	conditions	86,086	4.55
6	table	216,447	3.08
7	preformatted	15831	6.25
8	authors	38,512	4.26
9	self	45,155	3.75
10	research	62,832	3.06
11	email	46,566	3.45
12	infra	1341	8.19
13	purposes	31,902	3.53
14	tested	32,462	3.34
15	editors	14,405	4.28

First, collocations of the verb “claim” with a high frequency are mainly represented in groups that have a tendency for taking a stance, such as critics, supporters, proponents, conservatives, as well as names for certain objects connected to rights, possession or judgement, such as credit, property, prize. In addition, collocations with the verb “claim” may contain negative attributes, in particular, “falsely”. The frequent occurrence of the word claim proves the fact that this statement could still be unconfirmed or disputable since its ambiguity is reflected in its definition. Under some conditions, it becomes obvious that the word “claim” occurs in discourse games or in situations when multiple opinions have to be considered. Hence, the usage of the term implies making a statement that needs no comments about its credibility.

Conversely, the collocates associated with the verb “argue” tend to favor the fields of academics or profession (such as scholar, lawyer, economist, analyst) and often combine with opposing names such as “proponents” and “opponents”, and with other words that are related to the presentation of logical and reasonable points of view (such as “logic” and “merits”). This indicates that the verb “argue” is

mostly used to present the process of argumentation based on evidence and logic, highlighting the confrontation between rational analysis and viewpoints. From the perspective of collocation, “argue”, on the one hand, relies on reasons and evidence, and on the other hand, is often used in the context of discussion or debate on a certain issue. In this process, participants with different positions construct meaning by presenting arguments and responding to each other’s views. Therefore, “argue” is more focused on process and interaction, and its focus is not on stating a single conclusion, but on the promotion of the chain of argumentation and the formation of persuasiveness.

In the collocation of “note”, common descriptive or informative words such as “important”, “interesting”, “medical”, “conditions”, etc., also appear in the writing contexts of authors and editors, as well as descriptive contexts such as table, email, and preformatted. This reflects the function of “note” in objective statement, supplementary information or guiding readers’ attention. Compared to “claim” and “argue”, “note” has a more neutral tone; its primary function is not to express a stance or engage in debate, but rather to mark, emphasise or supplement existing information. In specific contexts, “note” is frequently used to organise the structure of information, for example, to highlight key content, provide background details or introduce supplementary explanations, thereby enhancing the text’s readability and clarity of information. Consequently, “note” is closer to a metadiscursive device, playing a significant role in academic writing, technical documentation and explanatory texts, with its primary functions being objectivity, guidance and support.

#### 4.3 Colligation

Colligation refers to the systematic co-occurrence relationship between words and their typical syntactic structures (Sinclair, 1991; Hunston & Francis, 2000). Based on data from the COCA corpus, a statistical analysis of the high-frequency syntactic constructions of “claim”, “argue” and “note” (see Tables 4-5, 4-6 and 4-7) reveals distinct differences in their colligational patterns.

**Table 4-5. Colligation of “Claim”**

Type	Frequency	Percentage
claim* +NP	95	47.5%
claim* +that-clause	92	46%
claim* +prep.	8	4%
claim* +to-infinitive	5	0.25%

**Table 4-6. Colligations of “Argue”**

Type	Frequency	Percentage
argue* +prep.	77	39.90%
argue* +noun	16	8.29%
argue* +that-clause	100	51.81%

**Table 4-7. Colligations of “Note”**

Type	Frequency	Percentage
note* +noun	36	19.15%
note* +prep.	48	25.53%
note* +that-clause	104	55.32%

First of all, “claim” mainly appears in claim that... and claim + noun phrases, and the proportions of the two structures are close, which are 46 % and 47.5 % respectively. This kind of structure is mostly used to quote a certain claim or position, but in the specific context, it often implies the speaker’s reservation attitude towards the claim, or suggests that it needs further verification. For instance, in the sentence “Some analysts claim that the data is inconclusive”, the word “claim that...” not only...proposes an opinion, but also shows that the speaker does not necessarily believe that it is true. In contrast, claim + noun phrases ( such as claim victory ) present results or positions through general expressions, which may also reflect a certain stance distance. On the whole, the claim’s class joins show a discourse strategy with a sense of distance in the report, and its semantic tendency is more cautious.

Secondly, “argue” mainly appears in argue that ( 51.81 % ) and argue + prep. ( 39.90 % ). Among them, argue that is more common in academic or argumentative style, which is used to introduce a complete proposition and expand the argument, like “The study argue that early intervention improves outcomes”. It shows strong logicity and analyticity; structures such as argue with / about... are more frequently used in interactive contexts to express disagreements or oppositions, such as in “They argue with the manager over the schedule”. It can be seen that the class connection of argue can not only serve logical argumentation, but also point to specific communicative behaviors.

Finally, the typical colligations of “note” include note that ( 55.32 % ) and several stylized expressions, such as “It should be noted that...” and “as noted above”. This type of structure is mainly used to introduce information, prompt key points or organize discourse, and less involved in position opposition or controversy. For example, “It should be noted that all procedures complied with ethical guidelines”. This sentence not only states the facts, but also reminds the reader to focus on key information. Compared with “claim” and “argue”, “note” usually does not strengthen the speaker’s subjective stance, but presents a more objective and restrained way of expression, and helps to enhance the coherence of the text.

To conclude, although each of “claim”, “argue” and “note” may introduce propositional statements, the specific connectivity of the words indicates the distinct discourse-functional orientation: “claim” refers to statements of an uncertain and controversial character while preserving some distance from the standpoint, “argue” involves the creation of an argument, where its communicative value alternates between logical reasoning and personal disagreement according to circumstances, and “note” aims at

offering informative cues and structuring discourse, remaining largely impartial in terms of expression. In summary, although “claim”, “argue” and “note” can be used to introduce propositional content, their respective types of colligation reflect different discourse function orientations: “claim” tends to cite uncertain or controversial propositions and maintain a certain position distance; “argue” focuses on construction and demonstration, and its expression effect varies with context between rational reasoning and interpersonal conflict and “note” is mainly responsible for information prompt and discourse organization function, showing a more neutral expression characteristics.

## 5. Conclusion

Based on the COCA corpus, this paper conducts a comparative analysis of the epistemic lexical verbs “argue”, “claim” and “note” across multiple dimensions, including word frequency distribution, register differences, collocation features and colligations. From the results above, it is clear that with respect to word frequency and register distribution, “note” is the most common verb occurring in the academic and expository register; “argue” occurs mostly in the academic and interactive register while “claim” is mostly used in the media and online registers. In relation to the collocative properties, “claim” is commonly used together with stance-taking groups and rights attribution. On the other hand, “argue” is more common when used with academic and professional topics and argumentation-related words. “note”, on the other hand, tends to occur together with vocabulary that informs and explains. In terms of colligation, “claim” usually appears in that-clauses and noun phrases. “argue” is mostly in that-clauses and together with different preposition phrases, whereas “note” uses many that-clauses. In general, there exist systematic distinctions between these three verbs regarding their usages in discourse. In particular, “claim” often leads into statements with which speakers take a position or convey controversial ideas; “argue” stresses the process of arguing by emphasizing the logic reasoning of the speaker based on some reasons and evidences; whereas “note” plays a role of conveying information or suggestion and prompting readers to notice important issues or provide further information. The results of this study reveal that the use of epistemic lexical verbs is connected with various aspects like discourse distributions, collocations, and syntactic patterns and thus cannot be considered as equivalent semantic substitution. It means that in practice, these verbs need to be differentiated according to the contexts.

This study supplements the corpus-based analysis of the usage of epistemic lexical verbs and provides a reference for second language learners to distinguish and use near-synonymous verbs in academic writing. At the same time, this study also has some limitations. For example, only three epistemic lexical verbs are selected as the analysis objects, and the research scope is relatively limited. Further research can be conducted on the basis of expanding the scope of lexical items, combining more types of corpora and quantitative methods to further explore the characteristics of the use of epistemic lexical words.

## References

- Biber, D. (2004). Historical Patterns for the Grammatical Marking of Stance: A Cross-Register Comparison. *Journal of Historical Pragmatics*, 5(1), 107-136.
- Dontcheva-Navratilova, O. (2018). Intercultural and Interdisciplinary Variation in the Use of Epistemic lexical verbs in Linguistics and Economics Research Articles. *Linguistica Pragensia*, 28(2), 154-167.
- Hunston, S., & Francis, G. (2000). *Pattern Grammar*. Amsterdam: John Benjamins Publishing Company.
- Hyland, K. (2017). Metadiscourse: What is it and where is it going?. *Journal of Pragmatics*, 113, 16-29.
- Jidong, Z., & Ping, L. (2005). Corpus-based Approaches to the Differentiation of English Synonyms. *Foreign Languages Bimonthly*, 2005(06), 53-56+100.
- Naixing, W. (2003). Fifty Years of Collocational Study: Conceptual Changes and Methodological Developments. *Foreign Languages Bimonthly*, 2003(02), 11-15
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Szczygłowska, T. (2023). Lexical Verbs of Epistemic Modality in Academic Written English: Disciplinary Variation. *Linguistica Silesiana*, 43, 91-111.
- Vass, H. (2017). Lexical Verb Hedging in Legal Discourse: The Case of Law Journal Articles and Supreme Court Majority and Dissenting Opinions. *English for Specific Purposes*, 48, 17-31.
- Wenzhong, L. (2010). The Research Perspectives of Corpus Linguistics. *Foreign Languages Bimonthly*, 2010(02), 37-40+72+127.
- Xie, F., & Patten, A. (2025). Epistemic lexical verbs in Academic Writing: A Corpus-Based Comparative Study of L1 English and L1 Chinese Writers of English. *Applied Corpus Linguistics*, 5(3), 100144.
- Xiying, N. (2008). Analysis on one hedging strategy-epistemic lexical verbs in academic research articles. *Journal of Dalian Maritime University(Social Science Edition)*, 2008(03), 186-188.
- Xuelian, L. (2015). An Analysis of a Pair of English Synonyms “Capable” and “Competent” Based on COCA. *Journal of Chengdu Normal University*, 2015(02), 60-64.
- Yating, L., & Sen, Y. (2016). Comparative Study of Synonyms Based on the COCA Corpus: A Case Study of “Outcome” and “Consequence”. *China Terminology*, 2016(01), 180-183.