

An Analysis of English Lexico-grammatical Features in Sunday Times Newspaper Integrated Corpus

Dilhan Rajapaksha

සාරසංකේපය

මෙම ලිපිය ශ්‍රී ලාංකීය සන්දර්භය තුළ භාවිතා වන ඉංග්‍රීසි වාග්කෝෂීය-ව්‍යාකරණ ලක්ෂණ, විශේෂයෙන් 2022 මාර්තු සහ අප්‍රේල් මාසවල “Sunday Times” පුවත්පතේ ව්‍යාපාර සම්බන්ධ ලිපිවල, විශේෂයෙන් ගොනු කරන ලද වාග් සංහිතයක් භාවිතා කරමින් විමර්ශනය කිරීම අරමුණු කරයි. තෝරාගත් කාල රාමුවට අනන්‍ය වූ වාග් සංහිතාත්මක ලක්ෂණ හඳුනාගැනීමට සහ ඉංග්‍රීසි ලිඛිත භාවිතයන් සඳහා අවශ්‍ය ප්‍රතිදානය සැපයීමේදී වාග් සංහිතාත්මක මෙවලම්වල සඵලතාවය විමර්ශනය කිරීමට වාග් සංහිත භාවිතයෙන් අධ්‍යයනය සිදු කෙරේ. ක්‍රමවේදයට ද්‍රව්‍ය තෝරාගැනීම, වාග් සංහිතයක් නිර්මාණය සහ විශ්ලේෂණය ඇතුළත් වේ. උපවාග් සංහිතාත්මක විවෘත-මූලාශ්‍ර “Concordancer AntConc” 3.5.9 සහ 4.14 අනුවාද භාවිතයෙන් විශ්ලේෂණය කෙරේ. අධ්‍යයනයෙන් සොයාගෙන ඇත්තේ ආර්ථික අර්බුදය දැඩි ලෙස පැවති සන්දර්භයක් තුළ සාමාන්‍යමය අර්ථයක් දැක්වීම සඳහා “due” යන වචනය බොහෝ දුරට “to” යන පූර්විකාව සමඟ භාවිතා කරන බවයි. විශ්ලේෂණය කරන ලද උප වාග් සංහිතයෙහි සකර්මක සහ උපකාරක ක්‍රියා පදයක් ලෙස “they have” යන පදය භාවිතා වන බව අධ්‍යයනයෙන් ද සොයාගෙන ඇත. ශ්‍රී ලංකාවේ භාවිතා වන ඉංග්‍රීසි භාෂාවේ වාග්කෝෂීය-ව්‍යාකරණ ලක්ෂණ හඳුනා ගැනීමට සහ කෙණ්‍රයේ සම්මත වාග් සංහිතාත්මක මෙවලම් හුරු කරවීමට අධ්‍යයනය දායක වේ.

යතුරු පද:- ශ්‍රී ලාංකීය පුවත්පත් ආශ්‍රිත වාග් සංහිත, වාග්කෝෂීය ලක්ෂණ, Sunday Times පුවත්පත, උපවාග් සංහිතාත්මක මෙවලම්

Abstract

This article aims to investigate the lexico-grammatical features of English used in the Sri Lankan context, particularly in business-related articles of **the Sunday Times newspaper** in March and April 2022, using a specifically assembled corpus. The study is conducted using corpus tools to recognize lexical features unique to the selected time frame and to investigate the effectiveness of corpus tools in providing necessary output for English written practices. The methodology includes material selection, corpus creation, and analysis. The sub-corpus is analyzed using the open-source concordancer AntConc 3.5.9 and 4.14 versions. The study finds that the word "due" is mostly used with the preposition "to" to indicate a negative connotation in the context of an economic downfall. The study also finds that "they have" is used as both a transitive and auxiliary verb in the analyzed sub-corpus. The study contributes to identifying the lexico-grammatical features of English used in Sri Lanka and familiarizing the industry standard corpus tools.

Keywords: - Sri Lankan Newspaper Corpora, Lexical Features, Sunday Times, Corpus Tools, Practicing English

Introduction

The purpose of this research is to investigate the Lexico-grammatical features of English used in Sri Lankan context from the business-related articles of **the Sunday Times newspaper** in months of March and April 2022 by generating specifically assembled corpus. The analysis of lexis is done by using corpus tools. Corpus will facilitate the analysis of written occurrences in the business context that is unique to the selected time frame.

As specific objectives,

- This will help in recognizing lexical features utilized in local English newspapers, which serve as the leading voice of legitimate written English usage in Sri Lanka.
- To investigate the effectiveness of corpus tools in providing necessary output for English written practices.

Objectives of Corpora in Language research

Corpus linguistics have long been established as a valid method of examining language via teaching, learning and researching. According to Nelson (2000) there are key reasons to use corpora in language analysis as,

- Authenticity
- Verifiability
- Retrievability
- Exposure to Large Amounts of Data
- New Insights into Language
- Enhancement of Learner Motivation

Use of Corpus Linguistics:

Corpus linguistics is a methodology of studying language that is based on a large and organised collection of text. It has been widely employed in a variety of domains, including language instruction, language learning, and sociolinguistics. Corpus analysis has been used to explore many elements of language use in a variety of situations, including business contexts. Business communication is a language genre that employs certain lexis, grammar structures, and discourse elements.

Review of Literature

One type of corpus that has been widely used in linguistic research is the newspaper corpus, which is a collection of articles from various newspapers.

One example of a newspaper corpus is the "British National Corpus" (BNC), which includes a variety of texts from different genres, such as newspapers, books, and academic journals. This corpus has been used in a wide range of linguistic studies, including investigations of lexical and grammatical patterns in different genres of text (**Leech et al., 1994**).

Another example is the "Corpus of Contemporary American English" (COCA), which includes more than 560 million words of text from a wide range of sources, including newspapers, magazines, and fiction. This corpus has been used in studies of linguistic variation across different regions and time periods (**Davies, 2010**).

Newspaper corpora have also been used to investigate specific linguistic features, such as discourse markers and modality. For example, Gries (2013) analyzed the use of the discourse marker "well" in the BNC and found that it is used more frequently in spoken than written texts. Similarly, Thompson and Ye (1991) used the BNC to investigate the use of modal verbs in news reports and found that they are often used to indicate uncertainty and speculation.

One recent study that used a newspaper corpus is "An Analysis of English Lexico-grammatical Features in Sunday Times Newspaper Integrated Corpus" by **Aliakbari and Kargar (2021)**. This study used a corpus of articles from the Sunday Times newspaper to investigate patterns of lexical and grammatical usage in different sections of the newspaper. The

authors found that the use of passive voice and complex sentences was more frequent in the news section, while the opinion section showed a higher frequency of modals and adverbial clauses.

In conclusion, newspaper corpora are a valuable resource for linguistic research, providing insights into patterns of language use in different genres and contexts. They have been used in a wide range of studies, from investigations of linguistic variation to analyses of specific linguistic features. With the increasing availability of digital texts, corpus linguistics is likely to continue to play an important role in the study of language in the years to come.

Lexico-grammatical Features in Business Context:

A number of studies have examined the use of Lexico-grammatical features in business contexts using corpus linguistics. For example, **Biber et al. (2013)** used corpus analysis to examine the differences in Lexico-grammatical features between spoken and written business communication. They found that written business communication was more formal and used more passive voice and nominalization than spoken communication. Similarly, **Chua et al. (2018)** used corpus analysis to investigate the use of hedging devices in business communication. They found that the use of hedging devices was related to the degree of certainty in the information being conveyed.

In addition, several studies have investigated Lexico-grammatical features of business communication in specific countries or regions. For instance, **Zeng et al. (2015)** used corpus analysis to investigate the use of modality in Chinese business news discourse. They found that modality was used to express the source and degree of information. Similarly, **Brown and Chua (2019)** examined the use of lexical bundles in Singaporean business communication. They found that lexical bundles were

used to convey specific meanings related to business activities and outcomes.

Aims of the Research

1. To analyze the Lexico-grammatical features of English used in business-related articles of **the Sunday Times newspaper** in Sri Lanka during March and April 2022.

2. To identify the lexical features utilized in local English newspapers that serve as the leading voice of legitimate written English usage in Sri Lanka.

3. To investigate the effectiveness of corpus tools in providing necessary output for English written practices in a specific context.

Overall, the research aims to contribute to the existing literature on corpus linguistics by examining the use of Lexico-grammatical features in the business context of Sri Lanka. The study aims to provide insights into the language practices and patterns in the local context, and explore the potential of corpus linguistics in analyzing language use in a specific context.

Methodology

Material selection

At the first stage, written materials were selected. After careful observation of authentic written material available in Sri Lanka with the emphasis on prevailing patterns of text, newspapers were selected as the suitable material for corpus. From there, weekend newspapers were selected as they contain descriptive and hefty articles compared to daily published news items. The selection of Sunday Times newspaper permitted with cases of having an esteemed collection of columnist and

professional correspondents, and the availability of properly maintained news article archive. The field of business is selected for article compilation. The prevailing local states permits the repetition of similar lexical and lexico-grammatical structures. Economy being the widespread topic of discussion in the previous year (**Anon., 2022**) it was selected as the field. Months of March and April were the beginning of the crisis time and that was the reason for selecting these months. Therefore, related articles showcase these repetitions.

After the initial compilation, graphic and other non-lexical items were removed from the articles. Then ten articles were produced with 2000-2200-word range per each document. This was done to avoid any misalignments and discrepancies when executing in the corpus tools. Next the meta data including details of the material and authors' data were saved in a separate Excel sheet for future references. After that, every document was saved separately in .txt format to make them compatible with corpus tools. In this way the sub-corpus was formed.

Analysis

Analysis of the created sub corpus was carried out using the open source Concordancer - AntConc 3.5.9 and 4.14 versions (**Anthony, 2023**).

3.1 Concordance

According to **Aswini (2019)** Concordance is depicted as “In corpus linguistics, concordance refers to a list of words which are arranged in alphabetical order. This is an additional resource that modern computer-based corpora offer. With the application of suitable software tools of concordance, lists of words and phrases of a corpus can be identified and retrieved with references to its immediate context.”

AntConc 3.59 tool was used to analyze the concordance. The lexical item “due” is used to designate the features related to it. “Due” can take the form of an Adjective, noun and adverb. Here the word is mostly used with the preposition “to” which makes “due to” a prepositional phrase. The meaning of “due to” defines; “as a result of : BECAUSE OF” (**webster, 2023**). The prepositional phrase “due to” is used to enunciate both negative and positive meanings. Yet in the formulated corpus (Diagram 3.1) the meaning gives a negative connotation. This is because the context is related to an economical downfall and all “due to” phrases are used to conceptualize the reasons for downfall and there is a tendency to showcase past actions mostly in this context

Examples

- “due to the direct downward revision of”
- “due to the workers not plucking”.

When the phrase “due to” is analyzed according to the alignment in the context taking it as the center point, the left side is used to exemplify the effect of the sentence and right side is to express the cause. According to the diagram 3.1 the lexical item comes next with “due to” is definite article “the” (highlighted in green colored section) and it follows a noun (highlighted in purple colored section).

Examples

- “due to the loss”
- “due to the inability”
- “due to the pandemic”

Concordance Results 1: 42 hits

Hit	KWIC	File
1	g tax scheme etc). In addition due to the direct downward revisi	DRSTMC13_01.txt
2	akdown of their supply chain due to lockdowns and curfew. Wo	DRSTMC20_01.txt
3	maintaining competitiveness due to the workers not plucking th	DRSTMC20_01.txt
4	estion of containers continue due to delays then vessels will by	DRSTMC20_02.txt
5	power generation decreases due to droughts, the share of coal	DRSTMC20_03.txt
6	onomic response is different due to differences in their vulnera	DRSTMC13_01.txt
7	ment has suffered drastically due to the loss of tax revenue.	DRSTMC13_01.txt
8	of the flow of import duties due to restrictions on the import c	DRSTMC13_01.txt
9	scales are grinding to a halt due to the inability to import agri-	DRSTAPL17_01.txt
10	and need to be addressed in due course. \x93We see 2022 as a	DRSTMC13_02.txt
11	he prices had to be increased due to the highly depreciated exci	DRSTAPL10_01.txt
12	usinesses struggling initially due to the pandemic and now a	DRSTMC20_01.txt
13	nability to import agri- inputs due to the non- availability of fore:	DRSTAPL17_01.txt
14	ardise the consumer interest due to a price hike by the	DRSTMC13_01.txt
15	rupee-dollar exchange rate is due to depreciate further , there is	DRSTAPL24_01.txt
16	in local fuel prices, that is due to domestic taxes and subsidi	DRSTAPL24_01.txt
17	liquidity and cash flow issues due to loss of sales and the	DRSTAPL10_01.txt
18	which has not been legislated. Due to the revision of the VAT	DRSTMC13_01.txt
19	the end of their useful life, due to the release of heavy metals	DRSTMC20_03.txt
20	bypassing Colombo mainly due to delays in other ports and	DRSTMC20_02.txt

Diagram 3.1 Concordance results of “Due”

The 20th hit suggests an another frequently used type of lexico-grammatical item. Here “due” is taken as a time adjunct.

3.2 Multi-word expressions

Concordance Results 5: 7 hits

Hit	KWIC	File
1	on a lease of 53 years and they have another 23 years mo	DRSTMC20_01.txt
2	from high cost of living as they have already lost their incc	DRSTAPL10_01.txt
3	vulnerability to these shocks. If they have built up their balance	DRSTMC13_01.txt
4	n-performing facilities. As such they have to consider these faci	DRSTMC06_02.txt
5	t 40-50 years pointing out that they have no electricity for abou	DRSTAPL17_01.txt
6	to lack of capital, even though they have a good idea and a	DRSTMC20_01.txt
7	it into their profits even though they have not been actually rec	DRSTMC06_02.txt

Diagram 3.2 Concordance results of “they have”

In this analysis the multi-words are searched to make different meanings. “They have” expression has come up with 07 hits. Here the broad usage is depicted. The utility of the word “have” can be identified. Here (Diagram 3.2) “have” is used as a transitive verb and an auxiliary verb. Coupling with “they”, the word “have” produces results mostly of auxiliary meaning. Then the following word usually is a past participle term.

Examples,

- They have built = Have is an auxiliary verb.
- They have already lost = Have is an auxiliary verb and here the adverb already has come up between past participle.

But in this result of the corpus, “they have” phrase is used in a transitive meaning which occurs more than the auxiliary meaning.

3.3 Collocation

The concept of collocation is based on the notion that each word in a language prefers certain lexical contexts over others, i.e. that any given word tends to co-occur with certain words more often than it does with others. For example, the word grass is often used together with green, and the lexeme LETTER is often used together with the lexemes WRITE AND READ (Lehecka, 2015).

In this analysis AntConc 4.14 version is used for its advance collocative figures. The result suggests 5 collocate types with the word supply.

KWIC Plot File View Cluster N-Gram Collocate Word Keyword Wordcloud									
Collocate Types 5 Collocate Tokens 18 Page Size 100 hits 1 to 5 of 5 hits									
	Collocate	Rank	Freq(Scaled)	FreqL	FreqR	Range	Likelihood	Effect	
1	shortages	1	140	0	7	4	40.568	5.535	
2	domestic	2	130	3	1	2	19.271	4.834	
3	chains	3	30	0	3	1	21.586	6.535	
4	economics	4	20	1	1	1	14.382	6.535	
5	force	4	20	0	2	1	14.382	6.535	

Diagram 3.3 Collocation of “supply”

Here the words are sorted in the order of frequency. In this set-up word combinations are not always adjacent to each other but statistically connected with each other. When considering the word “shortage”, it located next to “supply”, yet the word “economics” gives results of “economics of energy supply” which is not adjacent (Diagram 3.4). Yet there is a statistical significance to be gathered.

Total Hits: 2 Page Size 100 hits 1 to 2 of 2 hits				
File	Left Context	Hit	Right Context	
1 DRSTMC20_...	renewables need to be quantified in making this a reality.	Economics	of energy supply	Economics pose the biggest hurdles in
2 DRSTMC20_...	antified in making this a reality. Economics of energy supply	Economics	pose the biggest	hurdles in making this transition from

Diagram 3.4 Collocation of “supply”

The Likelihood and effect of collocated words shows the statistical significance. So, for collocates, the likelihood measure regulates if the collocate pair comes together based on chance or not. The effect function measures the strength between the collocated words. Here the words of “shortage” and “domestic” (Diagram 3.3) statistically closer hence making them top of the list.

3.4 Frequency list

A frequency list is simply a list of all the types (words) in a corpus together with the number of occurrences of each type. Frequency is one of the most significant concepts which form the foundation for the examination of corpora (**Brindle, 2016**). For this analysis, 4.1.4 version of AntConc was used. Frequency list is denoted by word list function in AntConc.

	Type	Rank	Freq	Range	NormFreq	NormRange
1	the	1	1421	10	69667.108	1.000
2	to	2	704	10	34514.880	1.000
3	and	3	607	10	29759.278	1.000
4	of	4	575	10	28190.420	1.000
5	in	5	434	10	21277.639	1.000
6	a	6	371	10	18188.949	1.000
7	is	7	259	10	12697.946	1.000
8	that	8	208	10	10197.578	1.000
9	as	9	187	10	9168.015	1.000
10	for	9	187	10	9168.015	1.000
11	ane	11	146	10	7157.915	1.000
12	with	12	140	10	6863.754	1.000
13	it	13	139	10	6814.728	1.000
14	on	13	139	10	6814.728	1.000
15	this	15	131	10	6422.513	1.000

Diagram 3.5 Word list results

The word list produced the results in the usual form where definite article “the” came as the most frequently used word with 1421 level of frequencies within the formulated corpus.

Then the wildcard option was used to take custom results. Here the results were taken with the

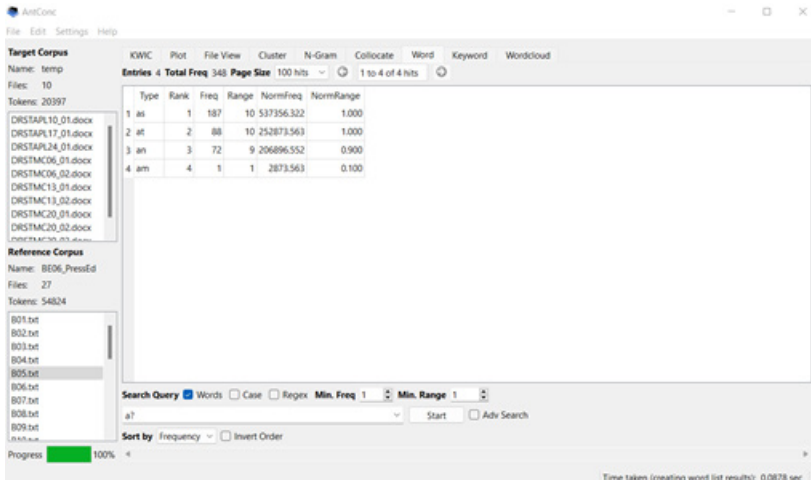


Diagram 3.6 Wildcard word list.

Words starting from the letter “a”. This gave specific results other than the usual word list results that we opted to. From this option, word “as” topped with 187 frequencies. Results can be filtered either as frequency list or in case based alphabetical order.

3.5 Keyword list

Keyword lists are formed by evaluating word occurrences in the target(local) corpus with a reference corpus which is much larger than the local corpus. Here the result words will denote as keywords with more frequency of occurrences from the target corpus.

The keywords list in AntConc is, as the name suggests, a tool to create a list of keywords. To do this your target corpus is compared to a reference corpus. The target and reference corpora do not need to be of the same size. The comparison is then done statistically (Tang, 2023).

The keyword list analysis was done using the 4.1.4 version of the AntConc. Here as reference corpus, **British English 2006: Press Editorial** corpus was used. The results depict the

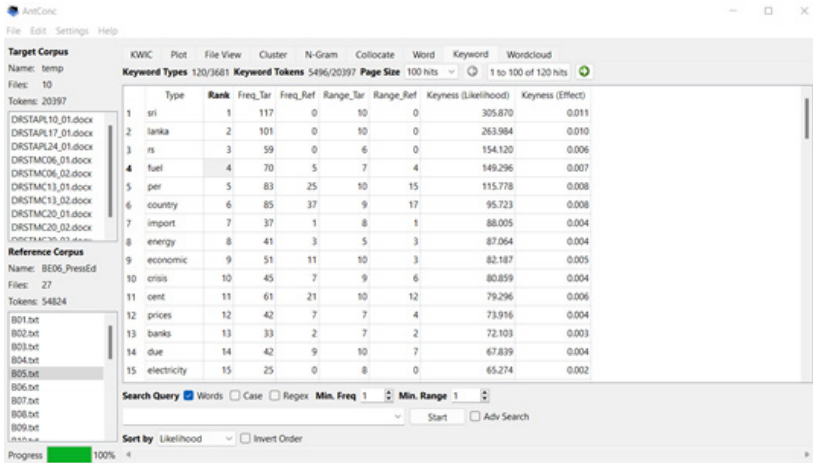


Diagram 3.7 Keyword list results

context of the source material of the corpus. The prevalent words suggested the crisis of middle quarter of the 2022 which was denoted from the keyword listing. “sri” and “lanka” came as top results as the corpus being local. The keyword “rs” came third referring to the local currency of Rupee (Diagram 3.7). Key words of “fuel”, “energy”, “economic”, “crisis”, “electricity” e.c.t depict the lexical terms related to the Sri Lankan economic crisis that were in news articles of relevant time. In quantitative terms, all the top 15 keywords occurred more than 25 in frequency and “Keyness” scored to be above 65.274. Keyword list is sorted through “likelihood” to evaluate most favorable occurrences. The resulted keywords illustrate the discourse patterns in the corpus with its related time period.

Important of the Research

The importance of corpus research lies in its ability to provide valuable insights into language use and structure. Corpora allow researchers to analyze and quantify patterns in language that may not be immediately obvious through intuition or observation alone. This data can be used to inform language teaching, improve machine translation and natural language processing, and even shed light on social and cultural phenomena.

Corpus research is particularly important in fields such as computational linguistics, where the ability to analyze and process large amounts of text data is crucial. By building and analyzing corpora, researchers can develop more accurate language models and algorithms, which in turn can lead to more effective natural language processing tools.

In addition to its practical applications, corpus research also has important theoretical implications. By analyzing patterns in language use across different contexts and genres, researchers can gain a better understanding of how language works and how it is used to convey meaning. This understanding can help us to develop more nuanced theories of language and communication, and can even shed light on broader questions about human cognition and culture. Overall, corpus research is a powerful tool for exploring and understanding the complexities of human language.

Conclusion

This section sums up the paper with findings of the analysis. The research work aimed mainly at finding lexicogrammatical features of English in local context with Business related fields. That objective was successfully accomplished. Local contexts of English were able to find in Keywords,

collocation and concordances. The business-related fields were targeted to elicit local lexical items that is presented in the given time frame where fuel shortages, riots, crisis and many form of commotions and economical barriers were present. The economic situation was the viral subject of all news media. Those said features were able to find out through created corpus analysis.

Specific objectives were also met in this analysis.

- Lexical features that have been used in local English newspapers were recognized. Concordance with multi word expressions were able to analyses the said point in ample amounts.
- The industry standard corpus tools were familiarized. Here the newest versions of open source corpus tools were tested and utilized for meaningful results. Some new options were learned. Wildcard analysis statistically based analytical tools like likelihood and effect were adapted.
- Effectiveness of corpus tools in providing necessary output for English written practices were investigate. Corpus tools are well equipped to analyze common writing patterns that were used in creating corpus which was sourced on authentic news items of everyday

practice. Incorporating these lexical patterners could help local learners to advance in their written practices.

References

1. Aliakbari, M., & Kargar, A. A. (2021). *An Analysis of English Lexico-grammatical Features in Sunday Times Newspaper Integrated Corpus. Theory and Practice in Language Studies*, 11(3), 309-317.

2. Anon., 2022. *The World Bank*. [Online]
3. Available at: <https://www.worldbank.org/en/country/srilanka/overview>
4. Anthony, L., 2022. <https://www.laurenceanthony.net/software/antconcl/>. [Online]
5. Available at: <https://www.laurenceanthony.net/software/antconcl/>
6. Aswini, 2019. *Enhancing Oral Communication Skills in English Using Authentic Online*
7. Audio Corpora Materials A Study at the Tertiary Level.
8. Brindle, A., 2016. *The Language of Hate: A Corpus Linguistic Analysis of White Supremacist Language*. s.l.:Routledge.
9. Davies, M. (2010). *The Corpus of Contemporary American English: 520 million words, 1990-present*. Retrieved from <https://corpus.byu.edu/coca/>
10. Gries, S. Th. (2013). *50-something years of work on collocations: What is or should be next.... International Journal of Corpus Linguistics*, 18(1), 137–166.
11. Leech, G., Conrad, S., & Culpeper, J. (1994). *The British National Corpus: a skelet on grammar and pragmatics*. Longman.
12. Thompson, S. A., & Ye, Y. (1991). *Evaluation in the reporting verbs used in academic papers. Applied Linguistics*, 12(4), 365–382.
13. Lehecka, T., 2015. *Handbook of Pragmatics*. s.l.:John Benjamins Publishing Company.
14. Nelson, 2000. *Corpus-based study of the lexis of business English and business English teaching materials*. s.l.:University of Manchester.
15. Zeng, J., Zhang, H., & Wei, W. (2015). *A corpus-based study of modality in Chinese business news discourse. English for Specific Purposes*, 38, 52-64.