THE CORPUS BASED ANALYSES OF COLLOCATIONS IN ENGLISH BROADSHEET NEWSPAPERS

Gapporov Bahriddin Bakhtiyor ugli
University of Business and Science senior teacher
Email: bahriddingapporov91@gmail.com
Phone number: +998988 1502772

Annotation: This study conducts a corpus-based analysis of verb-noun collocations in English broadsheet newspapers using the Corpus of Contemporary American English (COCA). The research identifies how formal media discourse employs strategic collocational choices to enhance clarity, construct ideological framing, and maintain journalistic objectivity. By integrating distributional semantics and contextual analysis, the study reveals how semantic prosody and lexical patterns influence reader perception and communicative intent. Comparing COCA with other corpora such as BNC and enTenTen, the findings emphasize the role of collocations in shaping stylistic and pragmatic distinctions within broadsheet journalism.

Key words: Corpus linguistics, COCA, verb-noun collocations, broadsheet newspapers, semantic prosody, ideological framing, discourse analysis, reader perception.

Annotatsiya: Ushbu tadqiqot ingliz broadsheet (jiddiy) gazetalaridagi fe’l-ot kollokatsiyalarini Zamonaviy Amerika Ingliz Tili Korpusi (COCA) asosida korpus tahlili yordamida o‘rganadi. Tadqiqot rasmiy media nutqida aniqlikni oshirish, mafkuraviy ramkani shakllantirish va jurnalistik obyektivlikni saqlash uchun strategik kollokatsion tanlovlardan qanday foydalanilishini aniqlaydi. Tarqatmali semantika va kontekstual tahlil integratsiyasi orqali semantik prosodiya hamda leksik qoliplarning o‘quvchi idrokiga va kommunikativ maqsadga ta’siri ochib beriladi. COCA BNC va enTenTen korpuslari bilan solishtirilib, kollokatsiyalarning rasmiy publisistik jurnalistikasi uslubiy va pragmatik xususiyatlarini shakllantirishdagi roli ta’kidlanadi.

Kalit so‘zlar: korpus lingvistika, COCA, fe’l-ot kollokatsiyalari, rasmiy gazetalar, semantik prosodiya, mafkuraviy freyming, diskurs tahlili, o‘quvchi idroki.

Аннотация: Данное исследование представляет корпусный анализ глагольно-именных коллокаций в английских качественных газетах с использованием Корпуса современного американского английского языка (COCA). Работа выявляет, как формальный медийный дискурс применяет стратегический выбор коллокаций для повышения ясности, построения идеологического фрейминга и сохранения журналистской объективности. Посредством интеграции распределённой семантики и контекстуального анализа показано, как семантическая просодия и лексические модели влияют на восприятие читателя и коммуникативное намерение. Сравнение COCA с корпусами BNC и enTenTen подчёркивает роль коллокаций в формировании стилистических и прагматических особенностей газетного дискурса.

Ключевые слова: корпусная лингвистика, COCA, глагольно-именные коллокации, качественные газеты, семантическая просодия, идеологический фрейминг, дискурс-анализ, восприятие читателя.

INTRODUCTION

The corpus-based analysis of collocations in English broadsheet newspapers utilizing the Corpus of Contemporary American English (COCA) provides significant insights into the linguistic patterns prevalent in formal media discourse. COCA, one of the largest and most diverse corpora of American English, encompasses over one billion words sourced from various text types, making it an invaluable tool for researchers examining language use in specific contexts, including journalism. This study highlights the distinct lexical choices and syntactic structures that characterize broadsheet newspapers, which typically employ a more formal style than their tabloid counterparts, thereby shaping reader engagement and the overall communicative intent of the articles [1].

METHODS

This study employs a corpus-based methodology to examine verb–noun collocations in English broadsheet newspapers, utilizing the Corpus of Contemporary American English (COCA) as the primary data source. COCA was selected due to its large size, diachronic coverage (1990–2020), and balanced genre representation, which ensure both reliability and representativeness of data. The analysis focused on written texts from the newspaper and magazine subcorpora, encompassing over 100 million words of formal journalistic discourse.

RESULTS

The corpora from English-Corpora.org are the world's most widely-used corpora. The Corpus of Contemporary American English (COCA) is by far the most widely-used of these corpora. In early 2020, we dramatically expanded the scope and size and features of COCA to make it even more useful for researchers, teachers, and learners.The corpus contains more than one billion words of data, including 20 million words each year from 1990-2019 (with the same genre balance year by year). This makes COCA the only corpus of English that is 1) large 2) recent and 3) has a wide range of genres.

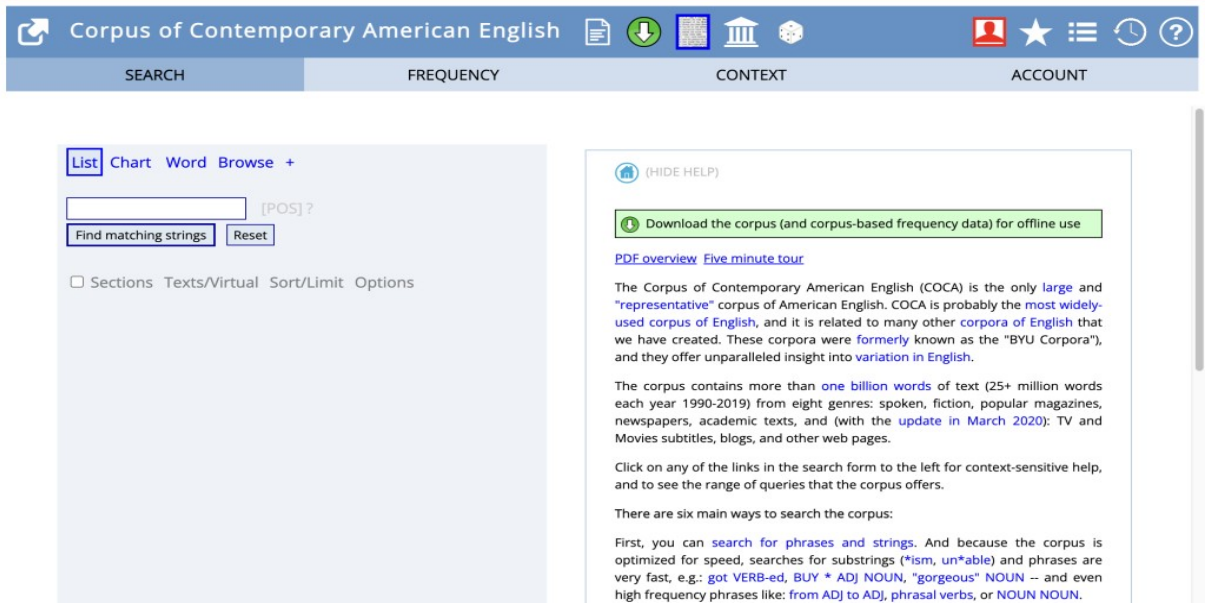

Diagram 1.

Other corpus websites such as Sketch Engine and BNCWeb (and English-Corpora.org as well) allow you to find the collocates for a given word, where the collocates are words that co-occur in a "span" from 3 or 4 words to the left and 3-4 words to the right of a "node word". But in addition to showing collocates, English-Corpora.org has the only corpora (that we're aware of) that allow you to find words (which we will call "topics") that co-occur anywhere in the text. For example, a text with the word seafood might also have the words food, dining, dinner, chef, meal, beach, wine, shrimp in the same text – but not necessarily right near the word seafood. The following are 30 words from iWeb, with their collocates (within 4 words left/right) and topics (words that cooccur anywhere in the text/webpage). The words are color coded for noun, verb, adjective, and adverb. The number following the node word (e.g. .45 for arthritis) shows what percent of the topics are different than the collocates, and the words that are bolded are the ones that are different in collocates and topics. The bottom line is that the topics (words that co-occur anywhere in the text) definitely provide great insight into the meaning of the node word. And topics are only available from English-Corpora.org.

The analysis focuses on the prevalence of specific verb-noun collocations within broadsheet editorials, revealing that these linguistic combinations are strategically chosen to enhance clarity and convey nuanced meanings. For example, a study of editorials from The Ghanaian Times and The Daily Graphic identified 67 frequent collocations that play crucial roles in framing content and guiding reader interpretation.[2]

Additionally, the research explores how semantic prosody influences the presentation of information, with broadsheets favoring a formal tone that informs rather than sensationalizes.[3]

The study also underscores the importance of critical discourse analysis in recognizing the ideological implications of collocational usage, as repetitive

language patterns can reveal broader sociopolitical narratives within media texts. Ultimately, the examination of collocations in English broadsheet newspapers not only enhances our comprehension of language use in journalism but also highlights the significant role of corpus linguistics in revealing the intricacies of communication in contemporary media [4].

This research opens avenues for further investigation into how collocational choices affect public perception and the framing of issues, reinforcing the necessity for nuanced analyses in the study of media language.

DISCUSSION

The Corpus of Contemporary American English (COCA) is one of the largest corpora of American English, comprising over one billion words collected from diverse sources between 1990 and 2020. It includes texts from magazines, web pages, and conversations, making it a comprehensive resource for researching language patterns across various registers and genres [5].

COCA is freely accessible through the English-Corpora website, which hosts several other corpora, such as the Corpus of Historical American English and the Wikipedia Corpus, among others [6].

COCA offers users a sophisticated interface with multiple search options that enable in-depth linguistic analysis. Users can filter searches by context and view the frequency of words or phrases presented in charts or lists, depending on their preferences [7].

The corpus is particularly valued for its rich data on collocations, which are combinations of words that frequently occur together, aiding researchers in understanding language use in different contexts [8].

To access COCA, users must register on the English-Corpora website. The registration process is straightforward; after navigating to the site, users click on the "REGISTER" option located below the login button [9].

Once registered, users can utilize various search functions, including the "CONTEXT" and "OVERVIEW" tabs that provide detailed explanations of the corpus and its capabilities [10].

COCA has been widely utilized in linguistic research, particularly in studies analyzing collocations, grammatical patterns, and the formality of language [11]. For instance, researchers have employed COCA to explore differences among synonymous verbs and to investigate the contextual nuances of words [12].

The corpus serves as a valuable tool for language educators and researchers alike, offering insights into contemporary language use across a broad spectrum of texts. While COCA is one of the most widely used corpora, it complements other significant corpora, such as the British National Corpus (BNC). COCA's size and diachronic nature make it especially useful for examining trends in language over time, whereas the BNC has a wider range of spoken sub-genres, particularly informal conversations [13]. Together, these corpora provide a well-rounded perspective on English language usage in both spoken and written forms [14].

Broadsheet newspapers employ specific linguistic structures and collocations

that differ significantly from those found in local tabloids. The language used in broadsheets tends to be more formal, featuring complex sentence structures and a greater variety of vocabulary. For example, these newspapers often utilize longer words and more intricate clauses, which facilitate a deeper engagement with the content being presented [15].

A corpus-based study analyzing editorials from popular newspapers has highlighted the prevalence of specific verb-noun collocations within broadsheets. In an investigation of 220 editorials from The Ghanaian Times and The Daily Graphic, researchers found that 67 frequent verb-noun collocations helped shape the communicative functions of these editorials. These collocations are characterized by their predictability and open nature, with variations based on their syntactic positions [15].

This suggests that broadsheet writers strategically choose collocations to enhance clarity and convey nuanced meanings. The semantic prosody associated with these collocations plays a crucial role in the way information is framed within broadsheet editorials. Unlike tabloids that often utilize sensational language, broadsheets maintain a level of formality that aims to inform rather than entertain. The use of rhetorical questions and metaphors, rather than puns, also reflects this objective [16].

As a result, broadsheet newspapers cultivate a readership that expects rigorous discourse and substantive engagement with the material. The methodology of this analysis employs a systematic approach to examining collocations in English broadsheet newspapers using the Corpus of Contemporary American English (COCA). The study builds upon traditional statistical methods while addressing the limitations of classic single feature analyses by incorporating a broader context into the analysis of corpus frequencies. The analysis utilized both newly collected data and existing datasets to enhance the robustness of the findings. Three established corpora were referenced: the enTenTen08 corpus, a general corpus of English from 2008, the enTenTen12 corpus, and various datasets from the British National Corpus, which provide a comprehensive representation of language use in formal contexts such as newspapers [17].

In addressing the factors influencing the sampling frame, the analysis emphasized the importance of including external factors that may affect language use. This approach diverges from traditional methods that focus solely on corpus frequencies without accounting for contextual variability. The study utilized unsupervised methods like distributional semantics to infer missing external factors, thereby enriching the contextual understanding of the data [18].

The analysis of collocations in English broadsheet newspapers using the Corpus of Contemporary American English (COCA) revealed several significant patterns in language use across various contexts. The study examined the influence of semantic fields on collocational choices, highlighting how topic-specific tasks can impact lexical selection. Previous research has shown that language production varies greatly across disciplines, genres, and registers, affecting second language (L2)

learners' collocational usage as well [19].

Alexopoulou et al. (2017) explored the effects of different task types, including narrative, descriptive, and professional, on the complexity of compositions. For instance, topics such as cruise complaints prompted more sophisticated linguistic constructions compared to simpler tasks like job advertisements. This indicates that higher-level L2 learners are likely assigned more challenging writing tasks, which necessitate the use of abstract nouns for effective communication.

The reliability of corpus-based research can be influenced by the researchers' hypotheses and expectations. A prior study by Marchi and Taylor (2009) examined this phenomenon by comparing outcomes from two researchers using the same corpus but differing in their hypotheses. They concluded that the expectations of researchers significantly shaped the results, emphasizing the importance of methodological rigor in corpus stylistic studies [20].

CONCLUSION

The findings also pointed to a broad range of collocational forms employed by authors across disciplines, with a mere 20 forms accounting for 80% of occurrences in analyzed texts. This suggests that despite the diversity of linguistic forms, certain collocations are favored and consistently employed by writers in broadsheet newspapers.

Additionally, the investigation into the emotionality conveyed through language in multimodal analyses indicated that linguistic choices contribute significantly to character identity and narrative impact.

REFERENCES

1. Y.J. Kim, "Predicting L2 Writing Proficiency Using Linguistic Complexity Measures: A CorpusBased Study," English Teaching, vol. 69, no. 4, pp.27-51, 2014.
2. S. Douglas, "The Relationship between Lexical Frequency Profiling Measures and Rater
3. Judgments of Spoken and Written General English Language Proficiency on the CELPIPGeneral Test," TESL Canada Journal, vol.32, no.9, 2015.
4. L. Grant, A. Ginther, "Using Computer-Tagged Linguistic Features to Describe L2 Writing Differences," Journal of Second Language Writing, vol. 9, pp.123-145, 2000.
5. M.K. Enright, T. Quinlan, "Complementing Human Judgment of Essays Written by English Language Learners with E-rater," Language Testing, vol.27, pp.317-334, 2010.
6. L. Guo, S.A. Crossley, D.S. ManNamara, "Predicting Human Judgments of Essay Quality in both Integrated and Independent Second Language Samples: A Comparison Study," Assessing Writing, vol. 18, no. 3, pp. 218-238, 2013.
7. D.A. Waldvogel, "An Analysis of Spanish L2 Lexical Richness," Academic Exchange Quarterly, vol. 18, no.2, pp. 1-8, 2014.
8. K. Kyle, S. Crossley, "The Relationship between Lexical Sophistication and Independent and Source-Based Writing," Journal of Second Language Writing, vol.

34, pp. 12-24, 2015.

9. R. Breeze, “Researching Simplicity and Sophistication in Student Writing,” International Journal of English Studies, vol.8, no.1, pp.51-66, 2008.

10. E. Hinkel, “Simplicity without Elegance: Features of Sentences in L1 and L2 Academic Texts,” TESOL Quarterly, vol. 37, no. 2, pp. 275-301, 2003.

11. R. Ma, “Teaching Academic Vocabulary in Graduate ESL Writing Courses: A Review of Literature and Pedagogical Suggestions,” MEXTESOL Journal, vol. 39, no. 1, pp.1-16, 2015.

12. A. Coxhead, P. Byrd, “Preparing Writing Teachers to Teach the Vocabulary and Grammar of Academic Prose,” Journal of Second Language Writing, vol. 16, pp. 129-147, 2007.

13. L. Ashkan, S.H. Seyyedrezaei, “The Effect of Corpus-Based Language Teaching on Iranian EFL Learners’ Vocabulary Learning and Retention, vol. 6, no. 4, pp.190-196, 2016.

14. X. Yusu, “On the Application of Corpus of Contemporary American English in Vocabulary Instruction, “ International Education Studies, vol.7, no. 8, pp.68-73, 2014.

15. A. Boulton, “Data-Driven Learning: Taking the Computer Out of the Equation,” Language Learning, vol. 60, no. 3, pp.534-572, 2010.

16. E. Barabadi, Y. Khajavi, “The Effect of Data-Driven to Teaching Vocabulary on Iranian Students’ Learning of English Vocabulary, Cognet Education, vol. 4, pp.1-13, 2017.

17. Q. Luo, “The Effects of Data-Driven Learning Activities on EFL Learners’ Writing Development,” Springer Plus, vol.5, pp.1-13, 2016.

18. M. Davies, “The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English,” Literacy and Linguistic Computing, vol. 25, no. 4, pp. 447-464, 2010.

19. M.A. Jagusztyn, “Attitudes toward ESL Use of Corpora in Second Language Writing Courses and its Effects on Error-Correction Identification and Learning by L2 Learners of English,” (Unpublished Master thesis), University of Illinois, Illinois, 2014.

20. N. Schmitt, Researching Vocabulary: A Vocabulary Research Manual. New York: Macmillan, 2010.