

РИГОВАНОВА ВІКТОРІЯ АНАТОЛІЇВНА

*Національний університет «Одеська юридична академія»,
доцент кафедри іноземних мов та соціально-гуманітарних дисциплін
Київського інституту інтелектуальної власності та права,
кандидат філологічних наук, доцент*

ЖОРНОУЙ УЛЯНА ВАСИЛІВНА

*Національний університет «Одеська юридична академія»,
старший викладач кафедри іноземних мов
та соціально-гуманітарних дисциплін
Київського інституту інтелектуальної власності та права,
кандидат філологічних наук*

EXPLORING NEWSPAPER ARTICLES THROUGH CORPUS METHODS: RESEARCH PATTERNS REVIEW

Introduction. Facing relatively new area of linguistic research, known as computational linguistics, we are used to the terms ‘corpus enrichment ‘and’ corpus analyses’. Since the first annotated corpus was produced in 1978, the interest and activity in corpus analysis has grown at a considerable speed. Many linguists favor the methods of corpus linguistics, although there is some disagreement about whether corpus linguistics is a methodology or a theory of language (or both) [1].

In order to conduct corpus analysis linguists require tools, as a corpus is not particularly valuable unless it is used in conjunction with computer software that can quickly and accurately perform manipulations on its contents. Some tools can be built in corpora, others are web-based, allowing users to search a corpus online, some tool functions appear as standalone platforms, capable of carrying out analysis on any text or corpus which the user specifics. Tools listed hereafter are free of charge: 1. Antconc (creator: Laurence Anthony); 2. Xaira (creator: Lou Burnard); 3. View (creator: Mark Davies); 4. Sketch Engine (creator: Adam Kilgarriff) has 30 day free trial.

Tognini-Bonelli makes a distinction between corpus-based (uses a corpus as a source of examples, to check researcher intuition or to examine the frequency and/or plausibility of the language contained within a smaller data set) and corpus- driven research (the corpus itself is the data and the patterns in it are noted as a way of expressing regularities (and exceptions) in language) [2]. McEnery and Partington, however, refer to the corpus-assisted analysis which can involve using a corpus as data in order to carry out linguistic analysis, but can also involve other forms of data or analysis occurring simultaneously (e.g. interviews, or etymological or historical research) [5]. Additionally, we may refer to existing linguistic frameworks or categories while doing corpus research, thus, in critical discourse analysis the corpus analysis should be used in conjunction with other forms of analysis which take social, historical and political context into account as well as drawing on existing theories.

Main body and discussion. Taken the requirements of the corpus approaches and variety of tools into consideration we suggest the overview of some patterns to conduct the newspaper articles analyses. Not only does the corpus-assisted analysis provide the linguists with the large amounts of existing data that enable quick and accurate analysis, but also it does practical consideration to avoid simplification, stereotyping or researcher bias conclusion.

Using the online database (e. g. Newsbank UK, GRAC: General Regionally Annotated Corpus of Ukrainian or others) the researcher may specify the time-frame or/and newspaper itself and then conduct the search of the node word(s). The overall search results in the compilation of sub-corpora for each of the key terms. In more general terms, it is possible to identify the dominant semantic fields. On this stage the first comparative study can be done. On the basis of the keyword lists, it is possible to develop hypotheses that can then be explored in greater depth using concordance analyses. Overall the stages for the deep analysis can be as follows: tagging the analysed corpus/corpora, tokenize the tagged corpus, count frequencies. As an example of such analysis George R S Weir and Nikolaos K. Anagnostou found out the preferences for part of speech in leading Scottish newspaper in comparison with reference BNC. As we have mentioned above the results can be unprojected, despite the consideration of the fact that the newspaper corpus should resemble the characteristics of the BNC reference corpus, «the newspaper corpus has a greater proportion of nouns than the BNC (32 % against 26 %); the BNC has a greater contribution from ‘non-specified’ parts-of-speech than does the newspaper corpus (22 % against 12 %); there is a wide difference in the proportion of determiners, with the newspaper showing almost 5 times the number of the BNC» [4].

By doing the corpus-driven analysis of the news reports and newspaper articles about COVID-19 linguists can contribute to better understanding how the disease is conceptualized metaphorically, public's reactions and concerns and government's reactions and responses to these disease. The research starts with corpus compiling, the next stage is indentifying the key words and analysing collocations (here it is possible to involve such measures: the frequency of the node and the frequency of the collocation). On this stage the ten top theme frequencies can be revealed. On the next stage the part of speech collocations is analysed. It is also beneficial to conduct the comparative research of the relevant topics. As an example of this comparison the Malaysian linguists discover the mismatching in the conceptualisation of the 2014–2015 Ebola outbreak; SARS, reported in local English newspapers; and COVID-19. While the similarities can be found in accepting the virus in terms of war lexemes, the differences are observed in relation of COVID-19 to politics and economy themes [3].

Conclusion. Obtaining a corpus on which to carry out research is a basic step in corpus linguistic research. If there is no corpus existence the researcher needs to build it, however, the ethical and copywrite issues may occur for newspaper article usage, thus it will make the research time-consuming or costly. Annotation of the texts under analysis should be

relevant to the research question. Semantic tags are useful in helping researchers to spot patterns of meaning that go beyond lexis. Searching on combinations of grammatical tags can enable to identify grammatically complex uses of language. The benefits of such analysis are in speed, accuracy, large amount of data for linguistic investigation, prevention of researcher's subjectivity and intuition.

References:

1. McEnery, T., Xiao, R. and Tono, Y. (2006) *Corpus-Based Language Studies: An Advanced Resource Book*. London : Routledge.
2. Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work*. Amsterdam : John Benjamins.
3. Nor Fariza Mohd, Nor Adlyn Syahirah Zulcafli *Corpus Driven Analysis of News Reports about Covid-19 in a Malaysian Online Newspaper*: Journal of Language Studies. Volume 20 (3), August 2020. On-line resource: <http://doi.org/10.17576/gema-2020-2003-12>
4. George R. S. Weir, Nicolaos K. Anagnostou Exploring newspapers: a case study in corpus analysis. On-line resource: <https://www.researchgate.net/publication/241820191>
5. Paul Baker (2010). *Sociolinguistics and Corpus linguistics*. Edinburgh University Press. – 189 p.

Ключові слова: корпусний аналіз, газетна стаття, корпусні інструменти.

Ключевые слова: корпусный анализ, газетная статья, корпусные инструменты.

Key words: corpus analysis, newspaper article, corpus tools.

КІСЕЛЬОВА АНАСТАСІЯ АНДРІЇВНА

*Національний університет «Одеська юридична академія»,
доцент кафедри прикладної лінгвістики, кандидат філологічних наук*

ЗАВАЛЬСЬКА ЛЮБОВ ВОЛОДИМИРІВНА

*Національний університет «Одеська юридична академія»,
доцент кафедри прикладної лінгвістики, кандидат філологічних наук*

ФУНКЦІОНАЛЬНІ ТИПИ ХЕШТЕГІВ, ПРИСВЯЧЕНИХ ТЕМІ КОРОНАВІРУСУ

У сучасних умовах життя людей дедалі більше переходить в онлайн-простір. Зараз в галузі вивчення медіадискурсу і медіатексту на перший план виходить той вид медіатексту (традиційно розділеного на радіо-, теле-, газетні та інтернет-тексти), який створюється в Інтернеті. Він, таким чином, створюється не тільки журналістами в електронних версіях різних видань, але і відомими людьми або звичайними користувачами на своїх сторінках в соціальних мережах.