Synonymous variation of 'War' in the British national corpus using sketch engine: a linguistic analysis

Zoriana Rybchak *,†, Olha Kulyna *,†

Lviv Polytechnic National University, S. Bandery street 12, Lviv, 79000, Ukraine

Abstract

This article presents an intelligent system designed to analyze and control speech based on user-defined criteria, with the objective of enhancing communication skills through insightful data analysis. Leveraging Python libraries such as PyAudio, Vosk, Pandas, and Plotly, the system enables audio recording, speech-to-text conversion, data management, and visualization of speech patterns. The study explores effective speech recognition methods and algorithms for audio processing and text analysis, including keyword detection and segment analysis. Visualizations generated by the system offer users a clear understanding of their speech dynamics over time. The software features an intuitive interface to ensure widespread usability. Key functionalities include speech recording, processing, unwanted word management, audio playback, and chart creation. This research contributes a comprehensive speech analysis application utilizing modern techniques to provide actionable insights for improving spoken language proficiency.

Keywords

British National Corpus, synonym, linguistic analysis, Sketch Engine, CQL

1. Introduction

The field of corpus linguistics is a highly significant area within linguistics and related disciplines [1; 2; 3; 4; 5; 6; 7]

Studies from numerous fields adopt the term 'corpus' to refer to a collection of text data, but they treat entire texts as singular entities rather than systematically analysing collections of texts to generalize linguistic findings across the entire corpus or specific subsets within it [8].

The aim of this article is to conduct a linguistic analysis focused on synonymous variations of the word 'war' as found in the British National Corpus (BNC) using the Sketch Engine. The goal is to explore how this critical term is used across different contexts and to identify patterns or shifts in its usage.

The novelty of this research lies in its utilization of the Sketch Engine, a powerful tool for corpus linguistics, to delve into the nuanced variations of the word 'war' within a large and diverse corpus like the BNC. This method allows for a comprehensive examination of linguistic contexts where 'war' appears and offers insights into the semantic and pragmatic aspects of its usage.

The hypothesis of this study is that 'war' exhibits significant synonymy across the BNC, appearing in various linguistic forms and contexts that reflect the multifaceted nature of conflicts and warfare in the English language. Through this analysis, the researchers expect to uncover distinctive patterns of synonymous usage, shedding light on how language shapes and reflects attitudes towards conflict and related phenomena.

We have defined the following tasks for this study:

1. Identify a synonym series for the word "war" within the context of texts using Sketch Engine.
2. Analyze different synonymous options and their usage contexts to reveal nuances in the meaning of the word "war."
3. Determine the most commonly used synonyms.
4. Understand the contextual shades in which these synonyms are used and their impact on the perception of the texts in which they are employed.

One key benefit of corpus-based analyses is the ability to generalize findings across a diverse range of texts, providing a more representative and nuanced understanding of language usage. Unlike traditional approaches that rely on individual examples or limited data sets, corpus linguistics enables researchers to identify patterns that may not be immediately apparent, contributing to a more comprehensive picture of how language evolves and adapts over time. This holistic view is crucial for advancing our understanding of semantic variation and pragmatic nuances surrounding important concepts like 'war'.

Moreover, the incorporation of computational tools like the Sketch Engine enhances the efficiency and accuracy of linguistic analyses. By automating the process of data retrieval and analysis, researchers can focus more on interpreting results and drawing meaningful conclusions from the corpus. This synergy between computational techniques and linguistic inquiry underscores the interdisciplinary nature of corpus linguistics, which bridges theoretical insights with practical applications in diverse fields such as lexicography, discourse analysis, and sociolinguistics.

2. Methodology

Corpus analysis can be done by integrating computational methods of natural language processing.

J. Dunn stated that the use of text classification and text similarity models demonstrates how we can enhance our capabilities in conducting corpus linguistics on extensive databases [9]. These computational techniques are gaining significance as corpora expand beyond the scope of traditional linguistic analysis methods.

For our research we use keyword extraction which involves automatically extracting the most pertinent information from text using various tools and machine learning algorithms.

We can customize our software to identify keywords that align with our specific requirements. This way we can experiment with provided sample keyword extractor tool.

The British National Corpus (BNC) is a collection of 100 million words sampled from written and spoken English across various sources [10]. It aims to represent a diverse range of British English from the later part of the 20th century. For our survey we used the BNC XML Edition. British National Corpus works with Sketch Engine and offers a complete set of tools such as word sketch, thesaurus, keyword, word list, n-grams, concordance, trends and text type analysis. Our research was limited to word sketch, thesaurus and concordance. The word sketch examines the collocates and contextual words associated with a particular word. It provides a concise summary of the word's grammatical and collocation patterns on a single page. The findings are categorized into grammatical relations. The thesaurus in Sketch Engine automatically generates compilation of synonyms or words that belong to the same semantic category (semantic field). This list is created by analysing the context in which these words appear within the chosen text corpus. The concordance tool in Sketch Engine offers a wide range of search options and can locate words, phrases, tags, documents, tags types or corpus structures and presents the results in context as a concordance. Users can sort, filter, count and further process the concordance to achieve their desired outcomes.

The Corpus Query Language (CQL) was used as a specific code or query language in Sketch Engine. It enables users to search for lexical patterns and set search criteria that are beyond what the standard user interface allows.

3. Analysis and discussion

We enter Sketch Engine and select the British National Corpus (BNC) corpus. Initially, we created a profile for the word "war." To do this, in Word Sketch, I entered the lemma "war" specifying the part of speech as a noun. The noun "war" is used 21,541 times and functions as a modifier; the word "war" modifies another word; verbs used with the word "war" were obtained; as an object and as a subject; other nouns used with "war" with the conjunction "and"; prepositional phrases; adjectives.

Figure 1: Word sketch of 'War'

Next, we visualized the table using the "Show Visualization" button.

Figure 2: Visualisation

Next, we could see how the collocations were used in context and explore the metadata by clicking on the icon marked with "I."

Figure 3: Display and count metadata

We present usage examples for consideration:

A civil war in the United States in the final decade of this century leads to the formation of a breakaway group.

I think Mao was quite keyed up on the whole situation, I think he realized that to win the war they had to erm adjust the mass support very carefully, and I think that's basically what this I think that's why two months later they er they er gave up this document cos he was worried then they'd lose the middle peasants' support.

Across the country, more than 5m of Mozambique's 16m people have been displaced by the war between President Joaquim Chissano's government forces and the Renamo rebels. </s><s> Now, as the peace seems to hold, families are beginning to go home.

Next, our objective was to generate a synonym series for the word "war" using the Thesaurus button. We selected to display the first 50 results initially. Then, the search was narrowed down to 20 results.

Figure 4: Thesaurus

For convenience, we downloaded all the data into a folder on the desktop. We also created a visualization of the executed search.
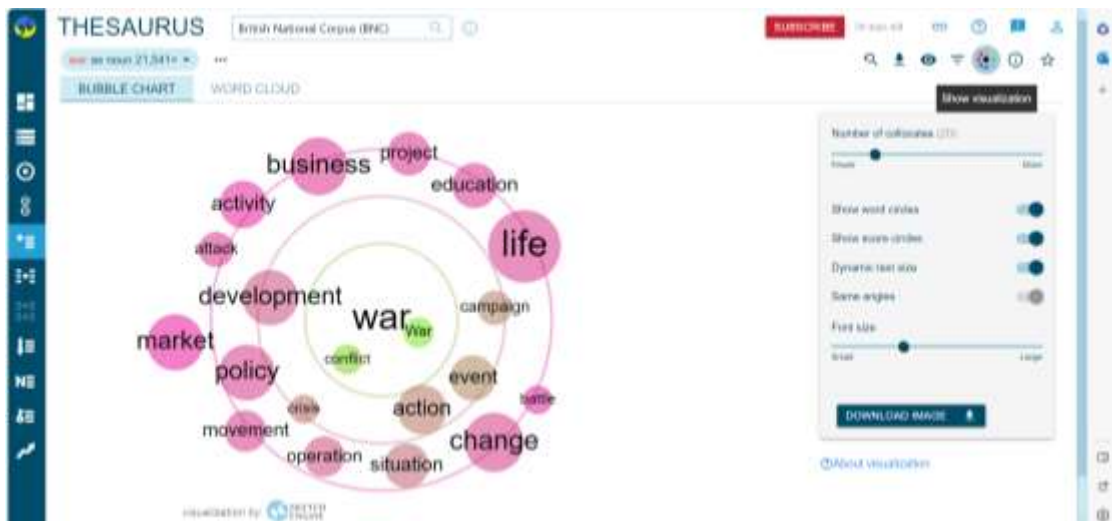

Figure 5: Visualization of Thesaurus

Let's look at examples in specific contexts:
The word "conflict" is used 7,075 times in various contexts. For analysis, we conduct an examination of the word "war" and each of the first 20 nouns from the synonym series

using Sketch Engine's Difference and Concordance tools.

Figure 6: Word Sketch of 'Conflict'

Let's illustrate with examples:

I do need, er, I do know that that the conflict between government, local government, the voluntary sector and all others who have an interest, can be quite prodigious, and the ways to resource can be also, quite considerable, and the whole thing does need to be debated and sorted out.

Circularisation along similar lines will also be required of the clients of all firms involved in a merger, and this exercise will be particularly useful in identifying any conflicts of interest.

The system itself had inherent 'contradictions provoking a conflict between private and public interest and hindering the proper operation of the planning machinery'

The profile of the word "conflict" expands our synonym series:

- Revolution
- Invasion
- Rebellion
- Struggle
- Contradiction
- Tension

We perform similar operations with the following words.

Figure 7: Word Sketch of 'Event'

Note that we are identifying the synonym series: "revolution," "invasion," "rebellion," which were found under "Conflict."

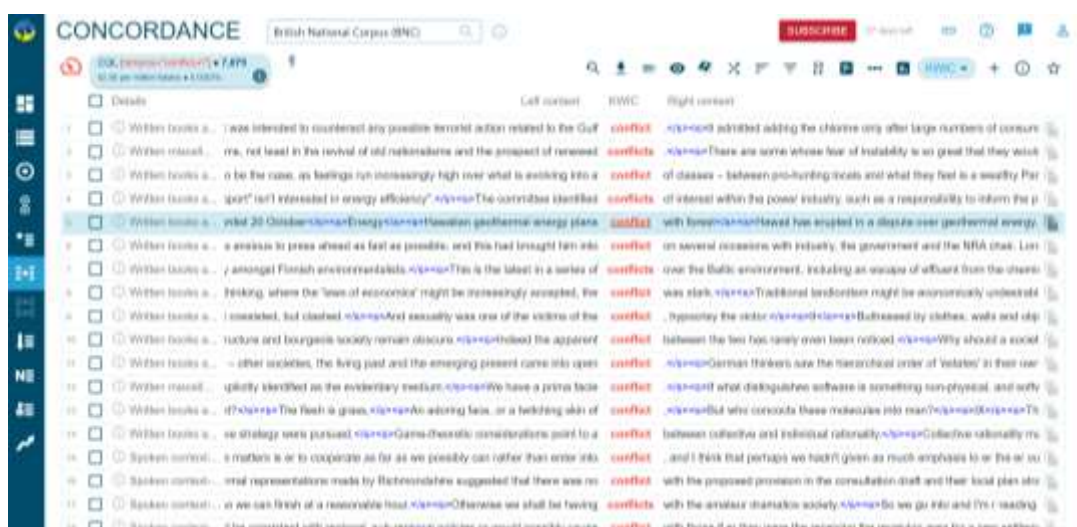It would be interesting to explore the Concordance:

Figure 8: Concordance of 'Conflict'

Examples:

Thanks to our parliamentary system and the stability that it has given us, the British people have been spared the horrors of revolution, civil war and invasion for more than 300 years. At the same time, it is also clear that there is not a strict and invariable relation between war, particularly defeat in war , and political revolution.

Therefore the exclusion of non-commercial ventures currently contained in the Transfer Regulations is in conflict with the EC Acquired Rights Directive and the exception is likely to be meaningless. Let's move on to the word "campaign," which is used 10,267 times.

Figure 9: Word Sketch of 'Campaign'

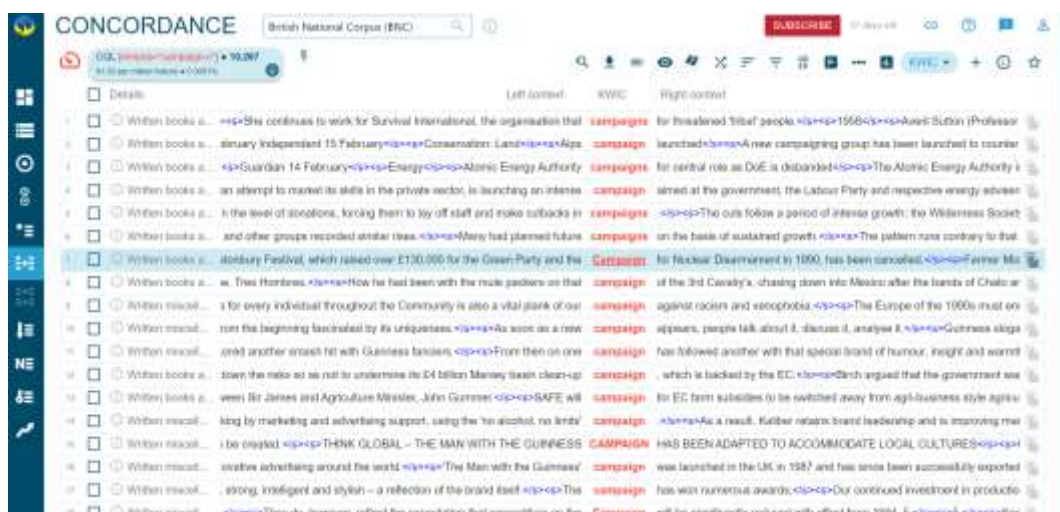Again, we encounter three synonyms: "revolution," "invasion," "rebellion."

Figure 10: Concordance of 'campaign'

Examples:
Even in eastern Europe the active anti-semitic campaigns , which were to stimulate the mass emigration of the Jews, still lay in the future
The campaign is fought on a national, and party, basis.
Despite this the two parties began almost immediately to undertake joint campaigns.
Let's move on to the next word "action," used 25,180 times, which also includes the synonymous series: revolution, invasion, rebellion.

Figure 11: Word Sketch of 'Action'

The word "crisis" is used 6,440 times and underscores its synonymy with revolution, invasion, rebellion, and conflict.

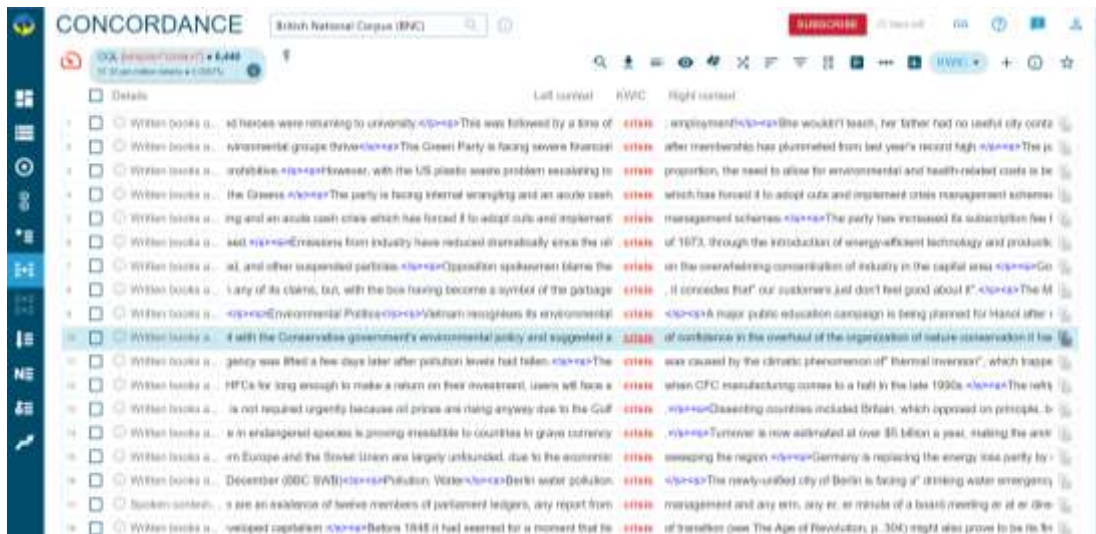**Figure 12:** Word Sketch of 'crisis'

Figure 13: Concordance of 'Crisis'

Examples:

Thus during the worst crisis in British industrial history neither the labour movement nor its radical Left were able to take advantage of the situation. The annual national rate of destruction of tropical rainforest has increased by 147 per cent since the Third World debt crisis began in 1982, according to an analysis by Friends of the Earth (FoE) of figures published by the UN Food and Agriculture Organisation (FAO). Other issues remained unresolved, including terms of trade and the debt crisis , action to combat global warming, and the means of safeguarding tropical forests. Let's move on to the word "situation," used 19,576 times, which offers additional synonyms: revolution, invasion, rebellion.

Figure 14: Word Sketch of 'Situation'

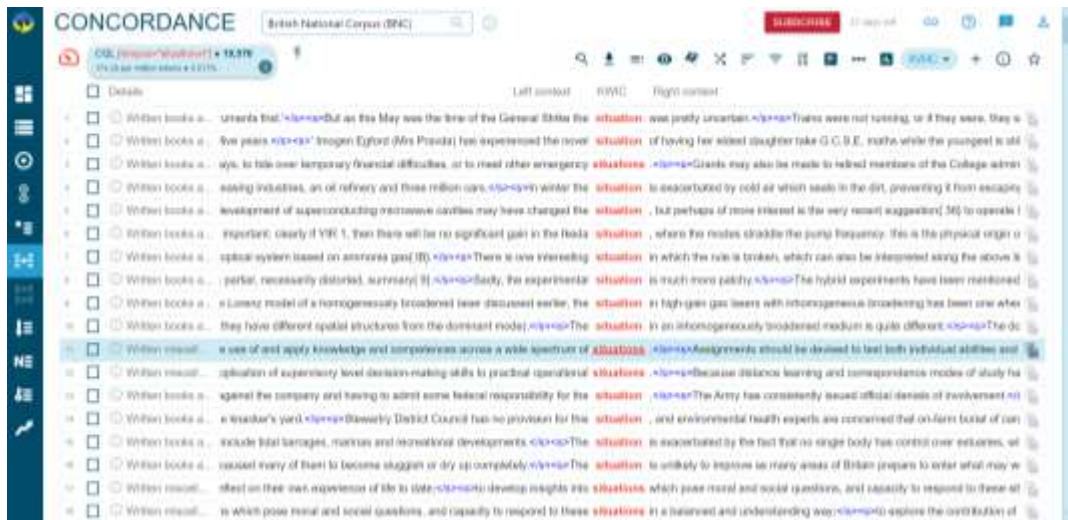We conduct a Concordance and proceed to the examples:

Figure 15: Concordance of 'Situation'

Examples:
There is one interesting situation in which the rule is broken, which can also be interpreted along the above lines.

Erm but the the point about today's discu discussions I don't call them interviews because er it's a self employed situation .

The climate is right, and we believe it could be sound financial management in a very difficult situation , it's been referred to, should we borrow.

The next word, "development," is used 32,898 times and complements the synonym series with the words "invasion," "revolution," and "rebellion."

Figure 16: Word Sketch of 'Development'

We analyze the concordance and extract examples.

Figure 17: Concordance of 'Development'

Examples:

In other regions we see scattered developments , again of figures which appear more or less subsidiary to the whole design.

The filling motifs – petals joined in twos, urn-peltae, squares with guilloche knots and floral scrolls – are all to be found, in various stages of development , elsewhere in the region.

Indeed, a close examination of the mosaics from Yorkshire and Humberside seems to reveal notable contrasts with the west in the number and importance of individual workshops as well as in the significance of planned developments (i.e. strategies of mosaic building desired by clients or, apparently, followed by mosaicists).

Let's look at and analyze "operation," used 15,564 times. It also highlights additional synonyms: revolution, invasion, rebellion.

Figure 18: Word Sketch of 'Operation'
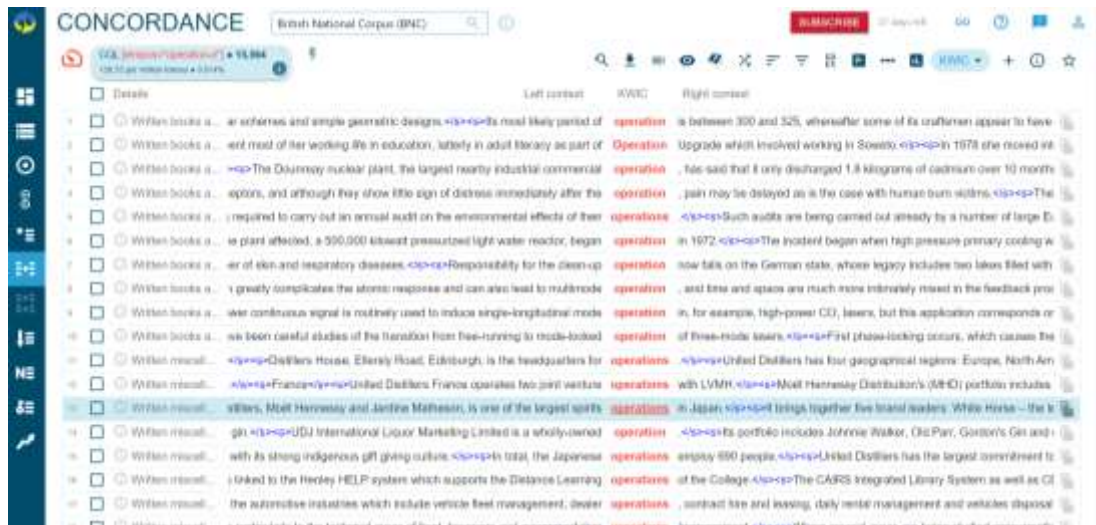
We extract examples from the Concordance.

Figure 29: Concordance of 'Operation'

Examples:

In total, the Japanese operations employ 690 people.

The company's oil-producing activities have been concentrated in the north, but operations are increasingly expanding in the south due to escalating militant action from tribal groups.

Ready for the op on the Saturday, and Friday I st started to sneeze, they took me temperature, I'd got a cold, between, it took us a fortnight and and, for the operation to come through it, common cold.

Similarly, we analyze words from the synonym series. All word searches can be entered using CQL (Corpus Query Language).

[lempos="war-n"] / [word="war"]
[lempos="conflict-n"] / [word="conflict"]
[lempos="event-n"] / [word="event"]
[lempos="campaign-n"] / [word="campaign"]
 [lempos="action-n"] / [word="action"]
[lempos="crisis-n"] [word="crisis"]
[lempos="situation-n"] / [word="situation"]
[lempos="development-n"] / [word="development"]
[lempos="operation-n"] / [word="operation"]
[lempos="change-n"] / [word="change"]
[lempos="policy-n"] / [word="policy"]
[lempos="life-n"] / [word="life"]
[lempos="movement-n"] / [word="movement"]
[lempos="attack-n"] / [word="attack"]
[lempos="education-n"] / [word="education"]
[lempos="business-n"] / [word="business"]
[lempos="project-n"] / [word="project"]
[lempos="activity-n"] / [word="activity"]
[lempos="battle-n"] / [word="battle"]
 [lempos="market-n"] / [word="market"]

Next, we selected the "Oneclick dictionary" option and gained access to dictionaries.

By analyzing and using examples, we identify words that, according to their meaning and contexts, do not correspond to the word "war": event, action, development, change, policy, life, movement, education, business, project, activity, market, situation.

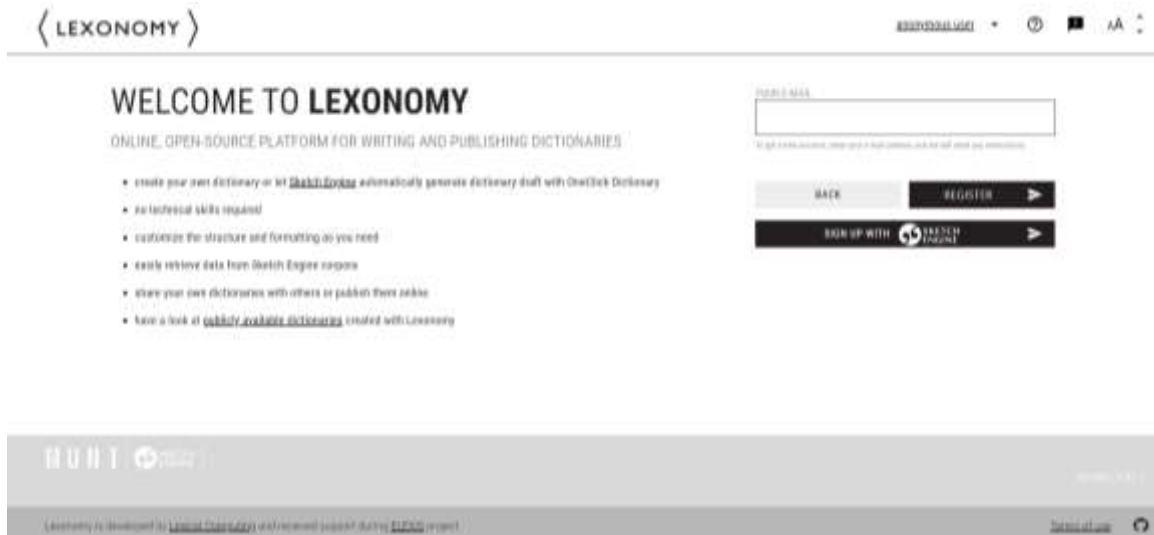Synonym series of the word "war":

Figure 20: Lexonomy

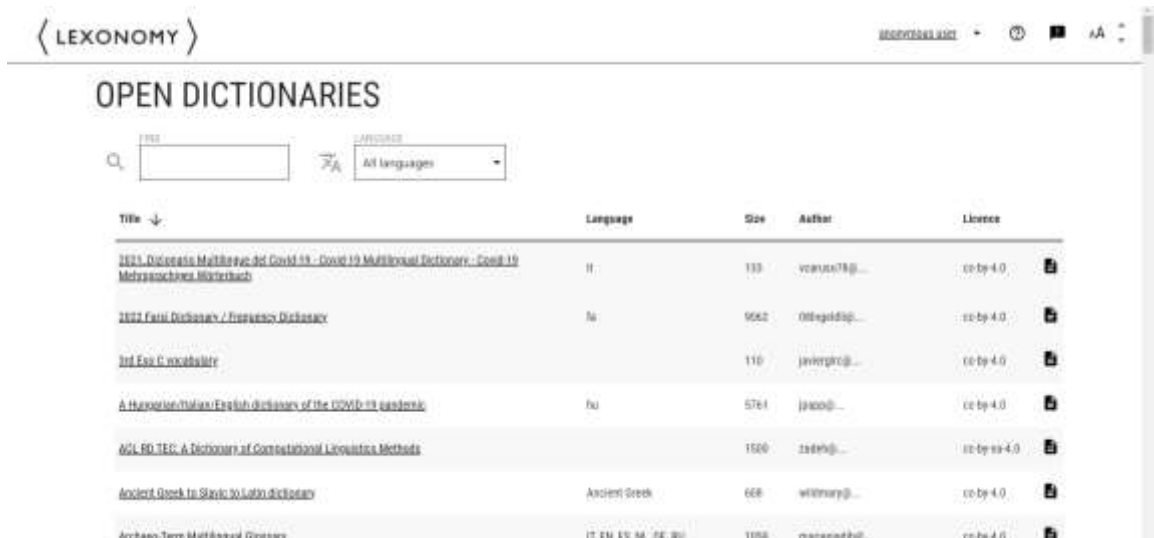We have excess to variety of dictionaries.

Figure 21: Open Dictionaries

We checked the meaning with Cambridge Dictionary.

4. Results

This pie chart provides a breakdown of ten distinct categories of synonyms for 'war' according to the findings:

Operation represents 21, 8 % of all recorded synonyms, indicating planned military actions or maneuvers with specific objectives.

Attack accounts for 15, 2 % of synonyms, reflecting offensive actions aimed at causing harm or damage.

Campaign represents 14,4 % of synonyms, signifying organized military operations with specific objectives.

War reached at 10, 1 %, reflecting large-scale armed conflicts between nations or groups.

Conflict accounts for 9,9 % of collected synonyms, representing various disputes or disagreements, ranging from interpersonal to societal issues.

Crisis reached 9 % of all synonyms, reflecting critical situations marked by instability and potential escalation.

Battle represents 6 % of synonyms, signifying engagements characterized by intense combat and strategic maneuvers.

Revolution comprises 5% of conflicts, denoting organized movements to overthrow established political or social systems.

Rebellion represents 1,5 % of synonyms, indicating acts of resistance or defiance against authority.
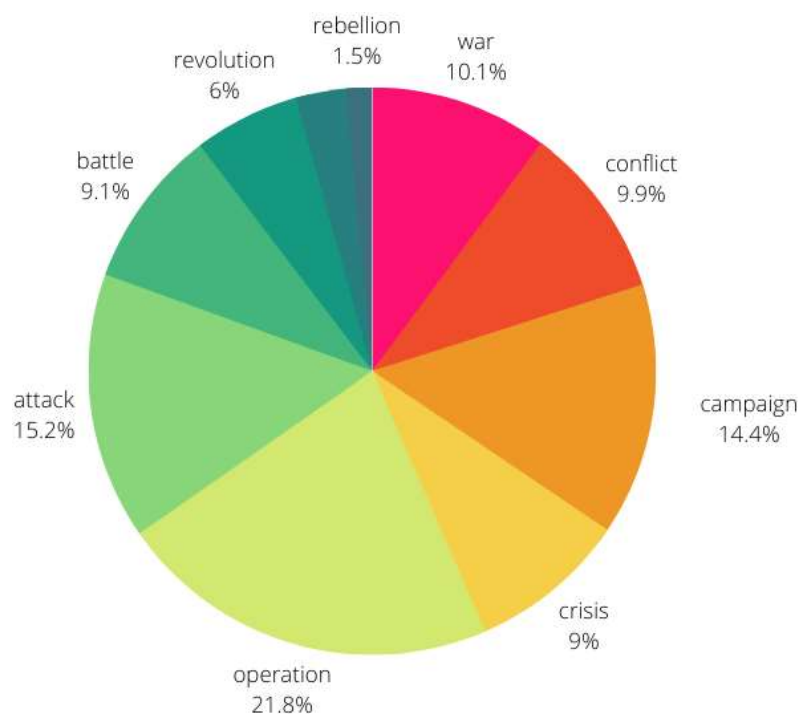

Figure 22: Synonyms of 'War' in British National Corpus

Accuracy and context are essential aspects of conducting linguistic analysis, particularly when examining the synonym series of the word "war" in the British National Corpus (BNC) using Sketch Engine. This analysis presents several challenges that require careful consideration:

Defining the parameters for constructing the synonym series of "war" poses a significant

challenge due to the diverse array of words used in contexts related to conflict, each with subtle variations in meaning.

Adjusting search parameters within Sketch Engine, such as selecting appropriate sub corpora, time frames, and refining constraints, is crucial. Improper settings can yield misleading or inaccurate results, impacting the integrity of the analysis.

Analyzing the usage of each word within the synonym series across different contexts is essential for grasping their semantics and nuanced meanings. However, aligning these contexts perfectly with the intended theme of "war" can prove challenging.

The vast amount of textual data contained in the British National Corpus necessitates thorough processing and analysis. This undertaking demands considerable time and patience to extract meaningful insights and draw valid conclusions. It is important to acknowledge that the examples provided in the analysis may not always align seamlessly with the thematic focus on "war," thus affecting the precision of the findings. Absolute accuracy cannot be guaranteed due to the inherent complexities of linguistic data analysis. Navigating these challenges requires a meticulous approach to ensure the reliability and validity of the research findings derived from linguistic analysis within large-scale corpora like the BNC.

5. Conclusions

In conclusion, this study not only contributes to the theoretical foundations of corpus linguistics but also offers practical insights into the usage and representation of 'war' within the English language. By elucidating the synonymous variations of this critical term, we aim to provide valuable perspectives on how language functions as a dynamic and adaptive system, reflecting and shaping human experiences of conflict and warfare. This research underscores the enduring relevance of corpus linguistics as a powerful tool for exploring the intricacies of language in all its richness and complexity.

The findings of this study are illustrated in the accompanying pie chart, which breaks down ten distinct categories of synonyms for 'war' based on their prevalence within the corpus analysis: operation represents 21.8% of all recorded synonyms, indicating planned military actions or maneuvers with specific objectives; attack accounts for 15.2% of synonyms, referring to offensive actions aimed at causing harm or damage; campaign represents 14.4% of synonyms, indicating organized military operations with specific objectives; war represents 10.1% of synonyms, reflecting large-scale armed conflicts between nations or groups; conflict contains for 9.9% of synonyms, representing various disputes or disagreements ranging from interpersonal to societal issues; crisis represents 9% of synonyms, reflecting critical situations marked by instability and potential escalation; battle - 6% of synonyms, signifying engagements characterized by intense combat and strategic maneuvers; revolution comprises 5% of synonyms, denoting organized movements to overthrow established political or social systems; rebellion possesses 1.5% of synonyms, indicating acts of resistance or defiance against authority.

This detailed analysis not only enhances our understanding of linguistic diversity but also highlights how language shapes our perception of complex societal issues (in our research 'war')..

References

[1] E. B. Soderqvist, Evidentiality across age and gender: A corpus-based study of variation in spoken British English, in Research in Corpus Linguistics, V. 5. doi: 10.32714/ricl.05.02.

[2] D. A. Kwary, A corpus and a concordancer of academic journal articles, in Data in Brief, V. 16, 2018. doi: 10.1016/j.dib.2017.11.023.

[3] V. Starko, A. Rysin, M. Shvedova, Ukrainian Text Preprocessing in GRAC, in 2021 IEEE 16th International Conference on Computer Sciences and Information Technologies, LVIV, Ukraine, 2021, doi: 10.1109/CSIT52700.2021.9648705.

[4] M. Shvedova, A. Rysin, V. Starko, Handling of Nonstandard Spelling in GRAC, in 2021 IEEE 16th International Conference on Computer Sciences and Information Technologies, LVIV, Ukraine. Vol. 2. P. 105-108, 2021. doi: 10.1109/CSIT52700.2021.9648834.

[5] V. Starko, A. Rysin. VESUM: A Large Morphological Dictionary of Ukrainian As a Dynamic Tool. COLINS-2022: 6th International Conference on Computational Linguistics and Intelligent Systems, May 12–13, 2022. URL: https://ceur-ws.org/Vol-3171/paper8.pdf

[6] M. Z. Lahjouji-Seppälä, A. Rabus & R. von Waldenfels, Ukrainian standard variants in the 20th century: stylometry to the rescue, in Russ Linguist, 2022. doi.org/10.1007/s11185-022-09262-9

[7] J. Egbert, B. Burch, D. Biber, Lexical dispersion and corpus design, in International Journal of Corpus Linguistics, V. 25, I. 1, 2020. doi: 10.1075/ijcl.18010.egb.

[8] P. Crosthwaite, S. Ningrum, M. Schweinberger, Research trends in corpus linguistics: a bibliometric analysis of two decades of Scopus-indexed Corpus Linguistics Reaserch in Arts and Humanities, in International journal of Corpus Lingustics, 2022. doi: 10.1075/ijcl.21072.cro.

[9] J. Dunn, Natural Language Processing for Corporus Linguatics. Cambridge University press, 2022. doi: 10.1017/9781009070447.

[10] British National Corpus. URL: www.natcorp.ox.ac.uk/corpus.