A NATURAL LANGUAGE PROCESSING APPROACH TO COMPARATIVE SENTIMENT AND TOPIC ANALYSIS OF ENGLISH NATIONAL ANTHEMS

Rezan Bakır[1] and Nurgul Ergul[2]

[1]Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Sivas University of Science and Technology, Turkey
[2]Department of Marketing and Advertising, Cumhuriyet University, Turkey

Abstract

National anthems are powerful symbols that capture a country's ideals, past events, and cultural identity; they are more than just songs. They are important in forming and reflecting a nation's collective consciousness because they are ceremonial and patriotic manifestations. Using cutting-edge natural language processing (NLP) techniques, this study compares the English national anthems of the US, UK, and Canada in order to examine their lyrical content in greater detail. This study aims to identify and clarify the underlying emotional tones, thematic consistency, and important references in these anthems by utilizing a variety of natural language processing (NLP) techniques, such as sentiment analysis, keyword extraction, named entity recognition, and topic modeling. Every anthem offers a rich tapestry of language that reflects the historical and cultural narratives of its country, providing a distinctive lens through which to view pride and identity. Insights into how these anthems convey national sentiments and values can be gained by using sentiment analysis, which enables us to measure the emotional undertones present in the lyrics. By highlighting the most important words and ideas, Keyword Extraction illuminates recurrent themes and focal points that characterize the meaning of each anthem. By recognizing and classifying noteworthy allusions to individuals, locations, and institutions, Named Entity Recognition (NER) provides insight into the historical and cultural backgrounds that influence the anthems. In order to map out the main concepts and motifs of each anthem, Topic Modeling identifies the main themes and subjects that run throughout the lyrics. In addition to identifying and contrasting these linguistic traits, the study looks into how they all work together to represent pride and national identity. This study offers a thorough understanding of how English-speaking countries express their collective identities through their anthems by comprehending the subtle ways in which these songs articulate and uphold national values. The results contribute to a more comprehensive understanding of the function of national anthems in cultural and national representation by providing insightful information about the emotional and thematic aspects of national pride.

Keywords:
National Enthem Analysis, NLP, Sentiment Analysis, Topic Modeling

1. INTRODUCTION

National anthems, which condense a country's historical, cultural, and ideological core into a few verses, are potent markers of national identity. Analyzing national anthems reveals how societies view themselves and their values, offering insight into the collective psyche of a country [1]. Research has shown that national anthems often convey themes of unity, resilience, and historical struggle, which are central to the construction of national identity. As an instance, the [2] study explored the major themes in English West African national anthems, focusing on unity, religion, freedom, and modesty. Based on colonial history, language, and geography, the article sampled five English-speaking nations: Ghana, Nigeria, Sierra Leone, Liberia, and The Gambia. The common perception of similar content and style was confirmed by the linguistic projection of the themes using literary figures and content lexical items. The report in another study [3] investigated the subjective identification of preferred topics and biases in national anthems using Tropes and Semantria software packages. Across continents and cultures, it showed consistent patterns in sentiments, relative weights, and preferred topics. According to the findings, sentiment scores range from neutral for Latin and Mediterranean anthems to more positive for Germanic, Slavic, and Central and Western Asian anthems.

In recent years, the application of Natural Language Processing (NLP) techniques in literary and linguistic studies has opened up new avenues for exploring the underlying emotions and themes in texts [4], [5]. For example, the [6] study evaluated text clustering methods for categorizing national anthems across 190 countries using English and other languages. Features like stop-words, stemming, corpus tokenization, noise removal, and TF-IDF were extracted using the Agglomerative Hierarchical Clustering technique. The results show that combining a clustering technique with an Agglomerative Hierarchical Clustering algorithm with TF-IDF properties is highly beneficial. In another study, [7] a Twitter dataset about the US election of 2016 was analyzed using NLP methods to identify patterns in the election. Topic models were applied to extract interesting topics, supervised models were used to categorize topics, and sentiment analysis was used to analyze tweets' attitudes. The analysis reveals interesting patterns and meaningful topics.

The purpose of this study is to compare the national anthems of the US, UK, and Canada using natural language processing (NLP) techniques, with an emphasis on thematic content and emotional tones. The purpose of national anthems is to elicit powerful emotional reactions and promote a sense of solidarity among the populace. According to[8], these anthems often reflect the historical contexts in which they were created, serving both as tools of statecraft and as expressions of national identity. The British national anthem, "God Save the King," for instance, is rooted in the monarchical tradition and reflects themes of loyalty and divine right [9] . In contrast, the Canadian national anthem, "O Canada," emphasizes themes of patriotism, natural beauty, and the protection of the homeland, reflecting Canada's colonial history and its subsequent emergence as an independent nation [10]

The American national anthem, "The Star-Spangled Banner," is steeped in the narrative of struggle and resilience, with its origins in the War of 1812 and its emphasis on freedom and perseverance [11]. In the context of modern linguistic studies, it is especially pertinent to use NLP techniques to examine the emotional and thematic content of these anthems. Through the

methodical analysis of texts made possible by NLP, scholars can spot patterns and trends that conventional literary analysis might miss [12]. By leveraging tools such as sentiment analysis and topic modeling, this study seeks to uncover the predominant emotions conveyed in each anthem, as well as the key themes that define them. Sentiment analysis, as explained by [13], is a technique used to determine the emotional tone of a text, categorizing it into positive, negative, or neutral sentiments. Numerous disciplines, such as literary studies, political discourse analysis, and social media analysis, have made extensive use of this approach. Sentiment analysis of national anthems can highlight the text's overall emotional appeal and indicate whether the anthem is meant to arouse pride, solemnity, or patriotism. For instance, research has examined how the anthems of various nations represent their emotional and cultural dynamics using sentiment analysis [14],[15]. In addition to sentiment analysis, topic modeling will be employed to identify the underlying themes within each anthem. Blei et al. [16] describe topic modeling as a method that uncovers the abstract topics that occur in a collection of texts, allowing for the examination of thematic structures within large bodies of text. Because it makes it possible to identify themes that are similar and different across various texts, this technique is especially helpful in comparative studies. This study will offer a thorough examination of the emotional and thematic content of the national anthems of the US, UK, and Canada by utilizing these NLP techniques. The study intends to advance knowledge of how national identity is expressed and interpreted through national symbols, as well as how these interpretations have changed over time, by comparing these anthems.

This work is novel because it takes an interdisciplinary approach, combining cultural analysis and natural language processing techniques to investigate how national anthems express national identity. This study offers insights that have not been thoroughly investigated using NLP techniques by comparing the English national anthems of the United States, Canada, and the United Kingdom. The study identifies underlying themes, sentiments, and cultural allusions in these anthems by using sentiment analysis, keyword extraction, named entity recognition (NER), and topic modeling. The integration of multiple NLP techniques into a single analytical framework makes this work unique. The combination of sentiment analysis, keyword extraction, NER, and topic modeling offers a more comprehensive and nuanced understanding of the anthems' content than traditional qualitative methods. The inclusion of word clouds as a visual tool further enhances the study's impact by providing an intuitive way to visualize key themes and linguistic patterns in the lyrics. This method makes the results more understandable to a larger audience and enables a more accurate interpretation of the linguistic components of the anthems. Additionally, the study advances our knowledge of how national symbols like anthems express pride and national identity. Through an analysis of the lyrics' emotional tone, major themes, and noteworthy allusions, this study illuminates how different countries use language to create and express their identities. This analysis's linguistic and cultural insights provide a useful viewpoint on the parallels and discrepancies in the ways that different countries view and construct their identities.

Overall, the novelty of this work lies in its innovative application of NLP techniques to conduct a detailed, data-driven analysis of national anthems. This approach provides new perspectives on the linguistic and cultural dimensions of national identity, making it a significant contribution to both linguistic and cultural studies.

2. METHODOLOGY

This study employs a comprehensive NLP approach to analyze and compare the English national anthems of the United Kingdom, the United States, and Canada. Several important NLP techniques are used in the methodology, including topic modeling, named entity recognition, sentiment analysis, and keyword extraction. Each method offers unique perspectives on the meaning and content of the anthems.

2.1 DATA COLLECTION

This study centers on the lyrical content of three significant English national anthems, each representing a distinct nation with its unique cultural and historical backdrop. The selected anthems are:

- **United Kingdom: "God Save the Queen":** "God Save the Queen" (or "God Save the King," depending on the current monarch) is the national anthem of the United Kingdom and is among the oldest and most well-known anthems in the world. It has been a major part of British national ceremonies and events since its inception in the 18th century. It is an essential text for comprehending national sentiment in the UK because of its lyrics, which express themes of devotion, patriotism, and divine favor.
- **United States: "The Star-Spangled Banner":** The national anthem of the United States, "The Star-Spangled Banner," was written during the War of 1812 by Francis Scott Key. The anthem's lyrics are inspired by the sight of the American flag flying over Fort McHenry after a night of intense bombardment. It symbolizes American resilience and pride, capturing themes of endurance and patriotism. The historical context of the anthem adds depth to its analysis, highlighting its role in American national identity.
- **Canada: "O Canada":** "O Canada" serves as the national anthem of Canada, with its origins tracing back to the early 20th century. The anthem's lyrics, originally in French and later translated into English, reflect Canadian values such as unity, freedom, and pride in the country's natural beauty. Its adoption as the national anthem in 1980 underscores its significance in representing Canadian identity and cultural heritage.

2.1.1 Data Acquisition:

- **Sources:** The lyrics for each national anthem were gathered from official and reputable sources to ensure both accuracy and authenticity. The sources included official national anthem websites, such as government portals and cultural organizations, as well as well-regarded lyrics databases and archives. Additionally, historical texts and scholarly articles were consulted to provide further validation and context.

2.1.2 Data Pre-Processing:

Preprocessing is a critical step in any data analysis or machine learning project, especially when working with textual data [17].

It acts as the cornerstone for guaranteeing that the data is clear, organized, and in a format that algorithms can use efficiently. Preprocessing is essential to NLP because it transforms unstructured, raw text into a format that computers can comprehend and analyze.To ensure that the data was clean, consistent, and prepared for complex natural language processing tasks, the collected lyrics underwent a rigorous pre-processing step before being analyzed.This included:

- **Cleaning Data:** This step involves the removal ofany extraneous formatting or non-lyrical content that may interfere with the analysis. Such content includes annotations, comments, advertisements, or any other metadata that are not part of the actual lyrics. By removing these elements, the dataset becomes more focused, ensuring that only the core lyrical content is analyzed, reducing noise and improving the quality of insights derived from the data.

- **Normalization:** Normalization is the process of standardizing the text to ensure uniformity across the dataset [18]. This involves converting the text into a consistent format, including ensuring uniform capitalization and punctuation. For example, words like "Love", "love", and "LOVE" should all be treated the same to avoid fragmentation in analysis. Normalization also involves standardizing abbreviations, correcting misspellings, and ensuring consistent usage of punctuation marks. This process reduces variability in the dataset, making it easier for algorithms to detect patterns and relationships within the text.

- **Tokenization:** Tokenization is the process of breaking down the lyrics into smaller, manageable units, typically words or phrases, that can be further analyzed [19]. In this stage, the text is divided into tokens—either individual words, sentences, or multi-word phrases—depending on the scope of the analysis. Tokenization is an essential step for preparing textual data for algorithms, as it helps transform unstructured text into a structured sequence that can be more easily processed by machines. For instance, each word can be treated as a feature for machine learning models, or the frequency of certain word patterns can be analyzed. Effective tokenization enables deeper insights into the lyrical content, allowing for sophisticated analyses such as word frequency counts, sentiment analysis, and syntactic structure investigation.

Ensuring the accuracy, comparability, and contextual understanding of the national anthems under analysis is the main goal of the data collection for this study. Reliable analysis requires accurate and authentic textual data because it ensures that the results accurately represent each anthem. The study ensures that the differences and similarities between various anthems are truly reflective of their content rather than data inconsistencies by employing consistent and high-quality text data. A thorough grasp of each anthem's function in expressing pride and national identity is also provided by gathering and preparing the lyrics with consideration for historical and thematic context. By collecting and preparing the lyrics of these anthems, this study aims to provide a robust basis for applying natural language processing techniques and deriving insights into national pride and identity.

2.2 USED NLP METHODS

This study employs several natural language processing techniques to analyze the lyrics of three English national anthems: "God Save the Queen," "The Star-Spangled Banner," and "O Canada." The following methods were applied to gain insights into the anthems' emotional tones, key themes, significant references, and overarching topics.

2.2.1 Sentiment Analysis:

To evaluate the emotional tone and subjective nature of the anthems' lyrics, known for its powerful yet user-friendly sentiment analysis tools, the TextBlob library was utilized. TextBlob offers a straightforward approach to determining the sentiment polarity and subjectivity of text data. Polarity, ranging from -1 (indicating the most negative sentiment) to +1 (indicating the most positive), provides insight into the overall emotional charge of the lyrics. Meanwhile, the subjectivity score measures how personal or opinionated the content is, with values from 0 (completely objective) to 1 (entirely subjective) [20].

This analysis provided a measurable understanding of the range of emotions found in each national anthem. Whether evoking sentiments of hope, defiance, unity, or patriotism, the sentiment polarity served to emphasize the anthems' emotional focus. The subjectivity score, on the other hand, demonstrated how much the lyrics represented individual or group experiences. A more thorough emotional profile of each anthem was provided by TextBlob's processing of the lyrical content, which also enabled us to identify minute changes in tone and expression that might not be immediately obvious. We could examine how these recognizable national symbols emotionally connect with their respective audiences by using this technique to compare the emotional impact of several anthems.

The goal of this analysis is to understand how each anthem expresses national pride and sentiment, revealing the emotional tone conveyed through the lyrics.

2.2.2 Keyword Extraction:

For extracting significant terms from the anthem lyrics, Python's collections. Counter was utilized. This method focused on isolating the most frequent and relevant words to uncover central themes and concepts within the lyrics. The process involved several key steps:

- **Tokenization and Stop Words Removal:** The lyrics were first broken down into individual words, or tokens, allowing for more granular analysis. Common stop words (e.g., "the," "and," "is"), which do not contribute meaningful information, were filtered out to ensure that the focus remained on the substantive content of the lyrics. This step was critical in refining the dataset and emphasizing the most important terms used in the anthems.

- **Frequency Analysis:** Once the lyrics were tokenized and stop words removed, the frequency of each remaining term was calculated. By analyzing word frequency, the most frequently occurring and contextually relevant words were identified, offering a snapshot of the key vocabulary employed in each anthem. Terms with higher frequencies typically signified their importance to the central message or theme of the anthem.

Through this process, significant insights into the thematic structure of the anthems were obtained. Repeated words and phrases often revealed common sentiments or ideas that the lyrics sought to convey, such as national pride, unity, or struggle. By focusing on the prominent terms, this analysis helped illuminate the core messages emphasized in each anthem, enabling a clearer understanding of the lyrical focus and the values they sought to express. This method also allowed for a comparative analysis across different anthems, highlighting any recurring motifs or distinctive keywords unique to each anthem's cultural or historical context.

2.2.3 Named Entity Recognition (NER):

SpaCy's powerful NER capabilities were utilized to identify and categorize named entities within the anthem lyrics. This process allowed for the automatic detection of key entities, such as specific people, places, and organizations, offering deeper insights into the historical and cultural references embedded in the text. The extracted entities were classified into several predefined categories, including:

- **Organizations (ORG):** References to institutions, groups, or formal organizations mentioned in the lyrics were identified and classified. This category helps uncover references to national institutions, alliances, or other bodies that may hold significance within the anthem.

- **Geopolitical Entities (GPE):** This classification included mentions of countries, cities, or political regions. GPE extraction is particularly valuable in understanding how the anthem refers to the nation itself or to foreign powers, allies, or territories, offering insights into its geopolitical stance or national identity.

- **Locations (LOC):** Locations such as mountains, rivers, or landmarks were recognized in this category, providing a geographical context to the anthem's themes. References to specific places often reflect national pride in natural landscapes or historical sites of significance.

- **Individuals (PERSON):** Personal names of key historical figures, leaders, or symbolic characters mentioned in the lyrics were extracted and categorized. Identifying references to individuals can shed light on the historical figures or heroes that an anthem celebrates, giving insight into the values and events it commemorates.

The objective of using NER was to systematically extract and categorize these significant references, helping to build a more nuanced understanding of the historical, political, and cultural contexts that are interwoven into the anthems. By highlighting specific entities, such as mentions of important figures or regions, this analysis facilitated a richer interpretation of the lyrics, providing insights into the national identity and values expressed through the anthem. Additionally, the classification of entities enabled comparative analysis, revealing differences or commonalities between the anthems based on their cultural and geopolitical references. This, in turn, contributed to a more comprehensive understanding of the messages and meanings that each anthem sought to convey.

2.2.4 Topic Modeling:

Topic modeling is a powerful technique used in natural language processing to automatically identify and discover the underlying themes or topics within large collections of text.

Unlike other methods that may focus on individual words or sentences, topic modeling captures the relationships between words across different documents to reveal patterns that represent broader themes. This approach is particularly useful when dealing with unstructured text data, as it helps to organize and summarize content by grouping similar words and phrases into coherent topics [21].

The primary importance of topic modeling lies in its ability to:

- **Extract Hidden Themes:** By analyzing word co-occurrence patterns, topic modeling can uncover the latent topics that dominate a body of text. This allows for a more structured understanding of complex and diverse content.

- **Facilitate Text Summarization:** Instead of manually reading through large volumes of text, topic modeling can provide an overview of the main topics, helping researchers or analysts quickly grasp the most important themes.

- **Improve Information Retrieval:** By categorizing content into topics, this technique enhances search functionality, allowing for more targeted and relevant retrieval of information based on specific themes or subjects.

- **Support Exploratory Data Analysis:** Topic modeling is particularly useful in exploratory analysis, where the aim is to discover trends or patterns without predefined categories. This makes it applicable to diverse domains, including social media analysis, document clustering, content recommendation, and more.

Common algorithms used for topic modeling include Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF). These models work by identifying sets of words that frequently appear together, grouping them into topics that reflect a specific theme or idea. To uncover the dominant themes and underlying topics present in the anthems' lyrics, Latent Dirichlet Allocation (LDA) was applied using the Gensim library. LDA is a widely used probabilistic topic modeling algorithm that uncovers hidden topics in large collections of text by assuming that documents are mixtures of multiple topics. Each topic, in turn, is characterized by a distribution of words. The goal of LDA is to infer these topic-word distributions and assign each document a set of topics, showing how different themes are represented across the text [22]. By analyzing word co-occurrence patterns, LDA can reveal the most prominent topics in a corpus, making it an effective tool for exploring and understanding thematic structures in unstructured data.This method enabled the systematic extraction of key topics, offering deeper insights into the recurring motifs and thematic elements within the lyrics. The process involved several key steps:

- **Tokenization and Bag-of-Words Representation:** The lyrics were first tokenized, breaking the text into individual words or terms. A Bag-of-Words (BoW) model was then constructed, capturing the frequency of each word in the corpus, without considering word order or context. This representation provided a structured format for further analysis, allowing the identification of key patterns and word distributions across the lyrics.

- **LDA Analysis:** LDA model identified the primary topics within the lyrics, providing insights into the overarching ideas and motifs expressed in each anthem.This analysis aimed to reveal and understand the principal themes

conveyed through the anthems' lyrics. By integrating these methods, the study aims to provide a comprehensive understanding of how each anthem reflects national identity and pride through its lyrical content.

2.3 COMPARATIVE ANALYSIS

The primary objective of the comparative analysis was to evaluate and contrast the thematic and emotional content of each anthem. By examining differences and similarities in sentiment, keywords, named entities, and topics, this analysis aimed to provide a deeper understanding of how each anthem articulates national identity and pride. This comparison sheds light on the distinctive ways in which the anthems reflect the values, history, and cultural significance of their respective countries. The findings offer valuable insights into how national anthems serve as expressions of national pride and identity, highlighting both commonalities and unique elements across different cultures.

The comparative analysis involved synthesizing results from multiple NLP techniques—sentiment analysis, keyword extraction, named entity recognition (NER), and topic modeling—across the three national anthems: "God Save the Queen," "The Star-Spangled Banner," and "O Canada." Each method provided unique insights into the lyrical content, which were then systematically compared to highlight distinctions and similarities.

2.3.1 Sentiment Scores Comparison:

Sentiment analysis results were compared to assess how each anthem conveys emotional tones. This included evaluating the polarity scores to understand the general positivity or negativity and subjectivity scores to gauge the extent of personal opinion or emotion in the lyrics. Differences in these scores were analyzed to determine how each anthem expresses national pride and sentiment.

Table.1. Comparative Overview of Applied NLP Techniques

NLP Technique	Tool/Library Used	Advantages	Limitations
Sentiment Analysis	TextBlob	-Easy to use -Provides polarity and subjectivity scores	-Lexicon-based; may miss nuanced emotions like solemnity or irony
Keyword Extraction	Python collections. Counter	-Simple and effective for small corpora - Highlights frequent terms	-Frequency alone doesn't capture contextual importance
Named Entity Recognition (NER)	SpaCy	-Accurate entity extraction Recognizes people,	-Limited historical or deep contextual interpretation
		places, and organizations	
Topic Modeling	Gensim (LDA)	-Captures hidden thematic structures -Suitable for small datasets	-Manual topic number selection -Potential topic overlap

2.3.2 Keyword Frequencies:

The most frequent and relevant keywords identified through keyword extraction were compared. This analysis focused on understanding which terms are most prominent in each anthem and what they reveal about the central themes and messages. By comparing keyword frequencies, we aimed to uncover the primary concepts emphasized in each anthem's lyrics.

2.3.3 Named Entities:

Named entities extracted from the lyrics were categorized and compared. This comparison included analyzing entities classified as organizations, geopolitical entities, locations, and individuals. The objective was to identify any common or distinct references that might reflect historical or cultural contexts unique to each anthem.

2.3.4 Topic Modeling Results:

The topics identified through Latent Dirichlet Allocation (LDA) were compared to uncover the main themes present in each anthem's lyrics. This involved examining the topics to see how they align or diverge across the anthems, providing insights into the overarching ideas and motifs expressed.

3. RESULTS AND DISCUSSION

3.1 SENTIMENT ANALYSIS RESULTS

The sentiment analysis, conducted using TextBlob, provided insights into the emotional tone and subjectivity of each anthem's lyrics:

- **"God Save the Queen"** from the United Kingdom demonstrated a positive sentiment with a polarity score of 0.314 and a high degree of subjectivity (0.657). This indicates that the anthem carries a positive emotional tone with significant personal or emotional content, reflecting deep national pride and reverence.
- **"The Star-Spangled Banner"** of the United States showed a polarity score of 0.266 and a subjectivity score of 0.516. The sentiment is somewhat less positive compared to the UK anthem, with moderate subjectivity. This suggests a tone of pride and patriotism tempered by a more restrained emotional expression.
- **"O Canada"** exhibited the highest positivity with a polarity score of 0.352 and a subjectivity score of 0.655, closely mirroring the UK anthem. The high sentiment score indicates a strong expression of national pride and emotional depth.

The sentiment analysis indicates that "O Canada" and "God Save the Queen" exhibit a stronger positive sentiment compared to "The Star-Spangled Banner." The higher positivity in "O Canada" (polarity score of 0.352) aligns with its focus on national pride and unity. Similarly, "God Save the Queen" reflects a positive tone, reinforcing themes of loyalty and reverence for the monarchy. The more restrained sentiment of "The Star-Spangled Banner" (polarity score of 0.266) may reflect the historical context of the anthem, which was born from a period of conflict and struggle. The moderate subjectivity in "The Star-Spangled Banner" suggests that while the anthem is patriotic, it carries a more objective tone compared to the more emotionally charged "O Canada" and "God Save the Queen."

3.2 KEYWORD EXTRACTION RESULTS

Keyword extraction, using Python's collections. Counter, highlighted significant terms used in each anthem:

- In **"God Save the Queen,"** frequent keywords included 'God,' 'save,' 'King,' 'Long,' and 'us.' These terms underscore the anthem's focus on divine and royal protection, reinforcing themes of loyalty and reverence towards the monarchy.

- For **"The Star-Spangled Banner,"** the prominent keywords were 'say,' 'star,' 'spangled,' 'banner,' and 'wave.' This reflects the anthem's emphasis on the visual and symbolic elements of the American flag, conveying resilience and patriotism.

- **"O Canada"** featured keywords like 'Canada,' 'thee,' 'land,' 'True,' and 'free.' These terms emphasize themes of national pride, freedom, and unity, portraying a strong connection to Canadian identity and values.

3.3 NAMED ENTITY RECOGNITION (NER) RESULTS

Using SpaCy's NER capabilities, the following notable entities were identified:

- **"God Save the Queen"** had no specific named entities, suggesting a focus on abstract concepts rather than particular individuals or locations.

- **"The Star-Spangled Banner"** included temporal references such as 'the night,' 'half,' and 'morning,' reflecting the historical context of the anthem's creation during a pivotal moment.

- In **"O Canada,"** entities included 'O Canada' (as an organization), and geographical references such as 'East' and 'Western Sea,' highlighting the anthem's focus on national identity and geographical pride.

3.4 TOPIC MODELING RESULTS

- **"God Save the Queen"** was associated with topics centered around divine and royal protection, as indicated by terms like 'God,' 'save,' and 'King.' This topic reinforces the anthem's themes of loyalty and reverence.

- **"The Star-Spangled Banner"** revealed topics related to themes of home, bravery, and freedom. Keywords such as 'home,' 'land,' and 'free' suggest a focus on resilience and patriotic pride.

- **"O Canada"** featured topics emphasizing national pride and the call to defend and stand for Canada, with terms like 'Canada,' 'guard,' and 'True,' reflecting themes of unity and strength.

3.5 WORD CLOUDS RESULTS

The most common words in the lyrics of each national anthem are visually summarized in word clouds, where word size denotes emphasis and frequency. By naturally emphasizing recurrent themes and emotional tones, these visualizations support the conclusions drawn from sentiment analysis, keyword extraction, and topic modeling.

God, "save," and "King" are prominently featured in the UK anthem's word cloud (Figure 1), which is consistent with the anthem's themes of divine authority, protection, and monarchical loyalty. These terms' size and central placement visually highlight the anthem's longstanding connotations of spiritual reverence and royal loyalty. This directly supports the sentiment analysis result, which highlighted themes of devotion and unity under a sovereign figure and revealed a strongly positive and subjective tone. The central themes of divine blessing and national unity via devotion to the crown were further validated by topic modeling.

The word cloud of "O Canada" features *"Canada"*, *"True"*, *"thee"*, and *"land"* as the most prominent terms. This reflects the anthem's strong nationalistic pride and its celebration of homeland and heritage. The prominence of inclusive and value-laden words such as *"True"* and *"free"* corresponds with the high polarity score (0.352) and high subjectivity, indicating a deeply emotional and patriotic message. The recurring keywords and their visualization in the word cloud validate the topic modeling result, which highlighted key themes such as protection of the homeland, loyalty, and unity.

The US anthem's word cloud highlights terms such as *"star"*, *"banner"*, *"wave"*, *"home"*, and *"free"*. These words evoke powerful imagery of the national flag and freedom, emphasizing the anthem's historical context and themes of endurance through struggle. The term *"banner"* symbolizes national unity and resistance, while *"wave"* and *"home"* suggest resilience and a yearning for peace. These visual elements correspond well with the sentiment polarity (0.266) and moderate subjectivity, which indicate a solemn yet hopeful tone. Topic modeling further supports these interpretations by identifying central themes of bravery, homeland, and patriotic defense. The Fig.1, Fig.2, and Fig.3 present the Word cloud for each of UK, Canada and US national anthem, respectively. These word clouds provide an intuitive visual summary of the central themes and keywords within each anthem, complementing the quantitative analyses and offering a more accessible understanding of the lyrics' thematic content.

Overall, the integration of sentiment analysis, keyword extraction, NER, topic modeling, and visual word clouds offers a comprehensive view of how these anthems express national identity and pride. The positive sentiment and recurring themes in "O Canada" and "God Save the Queen" emphasize strong national unity, while the symbolic focus in "The Star-Spangled Banner" reflects its historical and patriotic significance. The word clouds visually enhance the understanding of the key themes and messages conveyed by each anthem, highlighting their roles as powerful symbols of national identity.

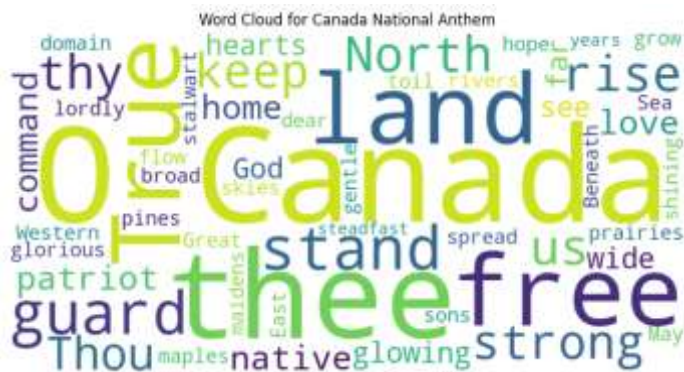Fig.1. Word cloud for the UK national anthem

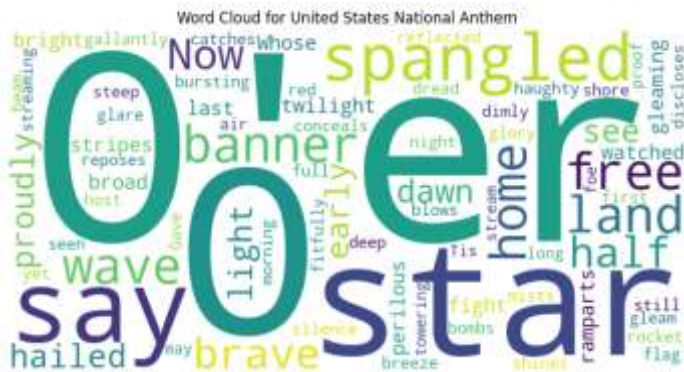Fig.2.Word cloud for the Canada national anthem

Fig.3. Word cloud for the US national anthem

4. CONCLUSION AND FUTURE WORKS

This study presents a comprehensive analysis of the English national anthems of the United Kingdom, the United States, and Canada using NLP techniques. By applying sentiment analysis, keyword extraction, NER, and topic modeling, the research uncovers the underlying themes, emotional tones, and significant cultural references within these anthems. The findings reveal distinct patterns in how each anthem articulates national pride, identity, and historical significance through its lyrics. The comparative analysis highlights key differences and similarities in the thematic and emotional content of the anthems. For instance, the UK anthem, "God Save the Queen," emphasizes loyalty and divine protection, reflecting a deep connection to

monarchy and tradition. In contrast, the US anthem, "The Star-Spangled Banner," focuses on resilience and freedom, encapsulating the country's values of independence and bravery. The Canadian anthem, "O Canada," emphasizes unity and national strength, portraying the country's diverse and inclusive identity.

Future research will aim to expand the analysis to include additional national anthems, particularly from Commonwealth countries, multilingual nations, and non-Western societies. This will enhance the diversity of cultural narratives examined and offer a more global perspective on how countries express national identity and collective values through anthem lyrics. Moreover, future work may incorporate historical variations in anthem lyrics to study how linguistic representations of identity have evolved over time. Additional enhancements may include the use of deep learning-based sentiment analysis tools to better capture nuanced emotional expressions, and the exploration of multimodal elements such as melody and rhythm, which also shape national perception. Moreovere, while this study focuses on the textual content of national anthems, future work could explore how audiences interpret and emotionally respond to these anthems through empirical methods such as surveys or sentiment perception studies. This would complement the NLP-based analysis with valuable human-centered perspectives

Overall, this study contributes to an interdisciplinary understanding of national identity by combining computational techniques with cultural analysis. The proposed extensions in future work aim to deepen this understanding and broaden its cross-cultural relevance.

REFERENCES

[1] K.A. Cerulo, "Symbols and the World System: National Anthems and Flags", *Sociological Forum*, pp. 243-271, 1993.

[2] C. Amenorvi and G.Y. Grumah, "A Thematic Analysis of the National Anthems of English West Africa", *Journal of Applied Linguistics and Applied Literature: Dynamics and Advances*, Vol. 7, No. 1, pp. 141-151, 2019.

[3] R. Silaghi-Dumitrescu, "Trends in the Texts of National Anthems: A Comparative Study", *Heliyon*, Vol. 9, No. 8, pp. 1-5, 2023.

[4] R. Ghanem and H. Erbay, "Context-Dependent Model for Spam Detection on Social Networks", SN Applied Sciences, Vol. 2, pp. 1-8, 2020.

[5] R. Bakır, H. Erbay and H. Bakır, "ALBERT4Spam: A Novel Approach for Spam Detection on Social Networks", *Bilişim Teknolojileri Dergisi*, Vol. 17, No. 2, pp. 81-94, 2024.

[6] P. Rai, N. Bista, K. Prasun and G. Sharma, "Text Classification of National Anthem using Agglomerative Hierarchical Clustering", *International Journal of Research Publications*, Vol. 146, No. 1, pp. 221-230, 2024.

[7] B. Kong, "Analysing Russian Trolls via NLP Tools", *Artificial Intelligence*, pp. 1-11, 2019.

[8] J.E. Morgan, "*America's National Anthem*", The Star-Spangled Banner in US History, Culture and Law, 2021.

[9] J. Black, "*English Nationalism: A Short History*", Oxford University Press, 2018.

[10] E. Mackey, "*House of Difference*", Cultural politics and National İdentity in Canada, pp. 1-216, 2005.

[11] J.R. Vile, "*America's National Anthem*", The Star-Spangled Banner in US History, Culture and Law, 2021.

[12] D. Jurafsky, "Speech and Language Processing", 2000.

[13] B. Liu, "Sentiment Analysis and Opinion Mining", 2022.

[14] R. Feldman, "Techniques and Applications for Sentiment Analysis", *Commun ACM*, Vol. 56, No. 4, pp. 82-89, 2013.

[15] R. Feldman, "Techniques and Applications for Sentiment Analysis: The Main Applications and Challenges of One of the Hottest Research Areas in Computer Science", *Commun ACM*, pp. 1-9, 2019.

[16] D.M. Blei, A.Y. Ng and M.I. Jordan, "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022, 2003.

[17] R. Ghanem and H. Erbay, "Spam Detection on Social Networks using Deep Contextualized Word Representation", *Multimedia Tools Applications*, Vol. 82, No. 3, pp. 3697-3712, 2023.

[18] A.A. Aliero, S.A. Bashir, H.O. Aliyu, A.G. Tafida, U.K. Bashar and M.D. Nasiru, "Systematic Review on Text Normalization Techniques and its Approach to Non-Standard words", *International Journal of Computer Applications*, Vol. 185, No.33, pp. 1-15, 2023.

[19] X. Song, A. Salcianu, Y. Song, D. Dopson and D. Zhou, "Fast Wordpiece Tokenization", *Proceedings of International Conference on Empirical Methods in Natural Language Processing*, pp. 2089-2103, 2020.

[20] S. Loria, "Textblob Documentation", *Journal of Data Analysis and Information Processing*, Vol. 2, No. 8, pp. 1-9, 2018.

[21] P. Kherwa and P. Bansal, "Topic Modeling: A Comprehensive Review", *EAI Endorsed Transactions on Scalable İnformation Systems*, Vol. 7, No. 24, pp. 1-11, 2019.

[22] H. Jelodar, "Latent Dirichlet Allocation (LDA) and Topic Modeling: Models, Applications, a Survey", *Multimedia Tools Applications*, Vol. 78, pp. 15169-15211, 2019.