

УДК 004.912

**DIGITAL METHODS FOR STUDYING AND VISUALISING THE SEMANTIC
STRUCTURE OF ENGLISH-LANGUAGE TEXTS OF NATIONAL ANTHEMS IN A
SOCIAL CONTEXT**

Оганисян Е.К. (Национальный исследовательский университет ИТМО)

**Научный руководитель – к.культур.н., доцент (квалификационная категория
«ординарный доцент») Пучковская А.А.**

(Национальный исследовательский университет ИТМО)

The research aimed to identify the most definite correlations between the selected features of the texts, as well as visualise the results in the most informative way. During the study, a comprehensive analysis of textual data extracted from national anthems was completed. The emphasis was set on the semantics of the texts studied and the usage of nontrivial methods of data processing.

Введение. Nowadays people express themselves through a huge variety of ways, still, the language remains the most common. Thus, an assumption was made that national anthems might reflect the general state and sentiment of the nation. Furthermore, national anthems can be studied to identify the interrelations between different nations. In addition, this task might help to solve if there are potential problems with the ways national anthems are portraying the people. e.g. racism, discrimination, violence, etc.

Основная часть. At the first step, the author has stated hypotheses, which must have been proven or disproven. Among these assumptions were the following: common religion affects the texts; current political state of the country puts a certain sentiment to the anthem; the country's historical background and the overall sentiment of the anthem are interdependent; the most frequent words include information about self-characteristics of a nation. Afterwards, a linguistic corpus of 205 texts of the official versions of national anthems was collected via a special concordance tool – AntConc. To identify the keywords, word clouds in R programming language were made for each text. Subsequently, the KWICs (keywords in context) were extracted via LancsBox. After extracting the necessary data, a dataset with the following attributes was created: religion, politics, reference to God and the national portrayal (where given) for each of the texts in the corpus. Lastly, the author created interactive JavaScript maps using the previously processed information from the dataset. This step was made to show the correlations found.

Выводы. At the end of the research, it was found that the mention of God in the text of the anthem depends on the most popular religion in the country, and it is common for the whole geographic region, not just for a specific country. Countries faced tragic events in history usually reflect this in the general sentiment of the text of the national anthem. In countries located in a region for which a common history is similar for all countries, people are characterized by the same set of words. Demo and results can be observed here: <http://anthemtalk.tilda.ws/>. This solution is found relevant for those who study linguistics and for children: it is visually understandable and interactive. Thus, it might be used as an additional tool in education.

Оганисян Е.К. (автор)

+7 921 366 52 37

eugenia.k.oganisian@gmail.com

Пучковская А.А. (научный руководитель)

УДК 004.912

ЦИФРОВЫЕ МЕТОДЫ ИЗУЧЕНИЯ И ВИЗУАЛИЗАЦИИ СЕМАНТИЧЕСКОЙ СТРУКТУРЫ АНГЛОЯЗЫЧНЫХ ТЕКСТОВ НАЦИОНАЛЬНЫХ ГИМНОВ В СОЦИАЛЬНОМ КОНТЕКСТЕ

Оганисян Е.К. (Национальный исследовательский университет ИТМО)

Научный руководитель – к.культур.н., доцент (квалификационная категория «ординарный доцент») Пучковская А.А.

(Национальный исследовательский университет ИТМО)

Цель исследования – выявить наиболее четкие корреляции между выбранными особенностями текстов, а также визуализировать результаты наиболее информативным способом. В ходе исследования был проведен комплексный анализ текстовых данных, извлеченных из англоязычных версий национальных гимнов. Акцент был сделан на семантику изученных текстов, а также использование нетривиальных методов обработки данных.

Введение. Несмотря на многообразие существующих средств самовыражения и коммуникации, естественный язык остается наиболее распространенным. Таким образом, было сделано предположение, что национальные гимны могут отражать общее состояние и настроения нации. Кроме того, национальные гимны могут быть изучены для выявления взаимосвязей между различными народами. Также изучение текстов гимнов может помочь узнать, есть ли потенциальные проблемы с тем, каким образом национальные гимны отображают людей (например, есть ли признаки расизма, дискриминации, насилия и т. д.).

Основная часть. На первом этапе были сформированы гипотезы, которые следовало доказать или опровергнуть. Были установлены следующие предположения: самая распространённая в стране религия влияет на тексты гимнов; текущее политическое положение страны придает гимну определенную эмоциональную окраску; история страны и общее настроение текста взаимосвязаны; наиболее часто встречающиеся слова включают информацию о характеристиках нации. Лингвистический корпус из 205 текстов официальных версий национальных гимнов был собран с помощью специального инструмента-конкорданса AntConc. Ключевые слова для каждого текста были выделены и сформированы в облака слов на языке R. Далее KWIC (ключевые слова в контексте) были извлечены через LancsBox. После извлечения необходимых данных был создан набор данных со следующими атрибутами: религия, политика, упоминание Бога и национальный портрет, где это было доступно. Интерактивные JavaScript карты были созданы на основе обработанной информации из набора данных. Этот шаг был сделан, чтобы показать найденные корреляции.

Выводы. По итогам исследования было выявлено, что упоминание Бога в тексте гимна зависит от распространённой религии в стране, это действует для всего географического региона, не только для конкретной страны. Страны с трагическим историческим прошлым обычно отражают это в тональности текста гимна. В регионе, где у стран общая история, люди характеризуются одинаковым набором слов. Результаты доступны по ссылке: <http://anthemtalk.tilda.ws/>. Это решение актуально для тех, кто изучает лингвистику и для детей: оно наглядно и интерактивно. Таким образом, оно может быть использовано в качестве дополнительного образовательного инструмента.

Оганисян Е.К. (автор)

+7 921 366 52 37

eugenia.k.oganisian@gmail.com

Пучковская А.А. (научный руководитель)