

УДК 81'42:32

Мальмаков Владислав Андреевич
магистрант кафедры перевода
Российский государственный педаго-
гический университет имени
А. И. Герцена
г. Санкт-Петербург, Россия

Vladislav Malmakov
Master Student of the Translation
Department
Herzen State Pedagogical University
of Russia
Saint Petersburg, Russia
vmalmakov@mail.ru

КОРПУСНЫЙ АНАЛИЗ ПОЛИТИЧЕСКОЙ РИТОРИКИ В РЕЧАХ ДОНАЛЬДА ТРАМПА И ДЖО БАЙДЕНА

В статье проводится корпусный анализ политической риторики Дональда Трампа и Джо Байдена на материале их выступлений перед Конгрессом США (2024–2025 гг.). Исследование фокусируется на выявлении лингвистических особенностей и способов речевого воздействия, таких как использование местоимений, тональность текста и частотность ключевых слов. Применяются методы морфологической и синтаксической разметки с использованием библиотек spaCy и textblob, что позволяет объективно оценить различия в риторике двух политиков. Результаты показывают, что Д. Трамп чаще использует позитивную лексику и субъективные оценки, тогда как Д. Байден делает акцент на коллективных ценностях. Исследование демонстрирует эффективность корпусных методов для анализа политического дискурса и может быть полезно для лингвистики, политологии и медиаисследований.

Ключевые слова: корпусный анализ; политическая риторика; Дональд Трамп; Джо Байден; речевое воздействие.

CORPUS-BASED ANALYSIS OF POLITICAL RHETORIC IN DONALD TRUMP'S AND JOE BIDEN'S SPEECHES

The article presents a corpus-based analysis of the political rhetoric of Donald Trump and Joe Biden, focusing on their addresses to the U.S. Congress (2024–2025). The study examines linguistic features and persuasive techniques, including pronoun usage, sentiment analysis, and keyword frequency. Methods such as morphological and syntactic tagging (using spaCy and textblob) provide an objective comparison of the two leaders' rhetorical styles. The results reveal that D. Trump employs more positive and subjective language, while J. Biden emphasizes collective values. The research highlights the effectiveness of corpus linguistics in political discourse analysis and contributes to the fields of linguistics, political science, and media studies.

Key words: corpus analysis; political rhetoric; Donald Trump; Joe Biden; persuasive techniques.

Актуальность. Исследование дополняет корпусные работы с точки зрения политической риторики.

Цель исследования. Выявить и сопоставить лингвистические особенности, способы речевого воздействия в речах Дональда Трампа (на совместном заседании Конгресса, 4 марта 2025 г.) и Джо Байдена («О положении страны», 7 марта 2024 г.) с использованием методов корпусного анализа.

Объект и предмет исследования. Объектом исследования является политическая риторика президентов США к Конгрессу как специфический жанр политической коммуникации. Предметом исследования являются лингвистические особенности Д. Трампа и Д. Байдена на материале собственного корпуса.

Политическая риторика – один из видов риторики, которая изучает речь как средство получения и осуществления государственной власти, а также оказания влияния на властные структуры [1]. Политики зачастую в своей речи прибегают к способам речевого воздействия на аудиторию. Одним из них является применение лингвистического образа «своих» и «чужих» [4, с. 36]. Они реализуются посредством инклюзивных местоимений, например в речи Д. Трампа: *We won all seven swing states, giving us an electoral college victory of 312 votes. We won the popular vote* [5]. Использование данного воздействия применимо к Д. Байдену: *We will not walk away. We will not bow down. I will not bow down* [6]. В приведенных случаях прослеживается тенденция того, что оба спикера не отделяют себя от публики, ведь они являются ее частью. Благодаря этому устанавливается положительный контакт с аудиторией, поскольку происходит непосредственное единение с народом. Что касается образа «чужих», то оба президента не сдерживаются во взаимной критике. Так, Д. Трамп заявляет следующее: *In comparison, under Joe Biden, the worst President in American history, there were hundreds of thousands of illegal crossings a month, and virtually all of them, including murderers, drug dealers, gang members, and people from mental institutions and insane asylums, were released into our country* [5]. В свою очередь, Д. Байден более сдержанным образом критикует на тот момент бывший кабинет президента Д. Трампа: *They (the last Administration) added more to the national debt than any Presidential term in American history* [6]. Данный образ строится в противостоянии республиканцев с одной стороны и демократов с другой. Американские лидеры способны оценивать ценностную систему своего народа и свободным образом опираются на нее в своих речах. Так, Д. Трамп апеллирует лозунгом *make America great again* [5], в то время как Д. Байден прибегает к применению слов *freedom, democracy, equality* [6]. Приведенные способы речевого воздействия являются лишь частью огромного множества методов влияния на аудиторию.

Корпусная лингвистика, в свою очередь, являясь разделом компьютерной лингвистики, рассматривает общие принципы построения и использования лингвистических корпусов с применением компьютерных технологий [2, с. 11]. Под языковым корпусом подразумевается большой, структурированный и размеченный массив данных, который предназначен для решения конкретных лингвистических задач. Именно поэтому корпус в области политической риторики способствует изучению речевых особенностей политических лидеров, поскольку он включает в себя несколько речей, а также позволяет определить, какими аспектами общественной жизни они апеллировали в то или иное время.

Составленный собственный корпус является одноязычным (английский язык), специализированным [3]. Он создавался на платформе Google Colab с применением встроенной библиотеки pandas (для сохранения результатов в формате Excel), встроенный модуль collections. Counter (для подсчета частотности слов), re (для фильтрации стоп-слов). Использовались сторонние библиотеки spaCy (для морфологической и синтаксической разметки) [7], textblob (для анализа тональности) [8]. Они являются сторонними, поскольку для их применения необходимо загрузить соответствующие библиотеки. Эмпирический материал включал в себя 9940 слов (речь Д. Трампа) и 8270 слов (речь Д. Байдена). Поскольку для анализа были отобраны стенограммы, слова, которые не относятся к обоим президентам, не были учтены, например, реакция публики, слова спикера Джеймса Джонсона. Применение морфологической и синтаксической разметок реализуется в рисунках 1 и 2.

```
with open(f'results/{filename}', 'w', encoding='utf-8') as f:
    for token in doc:
        f.write(f"Текст: {token.text}\n")
        f.write(f"Лемма: {token.lemma_}\n")
        f.write(f"Часть речи: {token.pos_}\n")
        f.write(f"Тег: {token.tag_}\n")
        f.write(f"Зависимость: {token.dep_}\n")
        f.write(f"Родитель: {token.head.text}\n")
        f.write("---\n")
```

Рис. 1. Код для морфологической разметки

```
with open(f'results/{filename}', 'w', encoding='utf-8') as f:
    for sent in doc.sents:
        f.write(f"Предложение: {sent.text}\n")
        for token in sent:
            f.write(f"{token.text} ({token.dep_}) -> {token.head.text}\n")
            f.write("\n" + "="*80 + "\n")
```

Рис. 2. Код для синтаксической разметки

Морфологическая разметка (part-of-speech tagging) включает не только признак части речи, но и признаки грамматических категорий, свойственных данной части речи [2, с. 36]. Разметка (annotation), в свою очередь, реализуется посредством приписывания текстам и их компонентам специальных тегов, а именно лингвистических, которые описывают лексические, грамматические и иные характеристики элементов текста, и экстралингвистических, включающих в себя информацию об авторе, название текста, год и место издания, жанр, тематику [2, с. 34]. Экстралингвистическая разметка не применялась в настоящем исследовании, поскольку она не является актуальной в отношении рассмотрения политической риторики.

В рамках корпуса код для морфологической разметки реализуются посредством команд *token.text* (определяет исходный текст токена), *token.lemma_* (определяет лемму, т. е. словарную форму слова); *token.pos_* (отвечает за часть речи, например, *VERB* – глагол; *token.tag_* (более детальный тег, например, *VBD* – глагол в прошедшем времени); *token.dep_* (определяет зависимость); *token.head.text* (определяет родительский токен). Токеном принято считать фрагмент текста, который считывается и понимается компьютером. Данный процесс проходит с каждым словом, как, например, со словом *America* в предложении *America is Back*:

Текст: *America*

Лемма: *America*

Часть речи: *PROPN*

Тег: *NNP*

Зависимость: *nsubj*

Родитель: *is*

Синтаксическая разметка состоит из команд *token.text* (текст токена); *token.dep_* (указывает тип синтаксической связи); *token.head.text* (указывает текст управляющего слова). Благодаря синтаксической разметке можно описать синтаксические связи между лексическими единицами и различные синтаксические конструкции, например, придаточное предложение, именное сказуемое и т. д. [2, с. 45]. Полученная разметка выглядит следующим образом:

Предложение Д. Грампа: *Who would want to do that?*

Who (*nsubj*) -> *want*

would (*aux*) -> *want*

want (*ROOT*) -> *want*

to (*aux*) -> *do*

do (*xcomp*) -> *want*

that (*dobj*) -> *do*

? (*punct*) -> *want*

Предложение Д. Байдена: *War was raging in Europe*

War (*nsubj*) -> *raging*

was (*aux*) -> *raging*

raging (*ROOT*) -> *raging*

in (*prep*) -> *raging*

Europe (*pobj*) -> *in*

. (*punct*) -> *raging*

Применение анализа тональности текста реализуется посредством библиотеки *textblob*. *Textblob* является библиотекой Python и применяется для обработки естественного языка (NLP), исправления орфографии, перевода.

Данная библиотека, как и `sraCy`, применяются в Google Colab, поскольку он основан на языке Python. Код анализа тональности текста рассмотрен в рисунке 3.

```
def analyze_sentiment(text):  
    analysis = TextBlob(text)  
    return {  
        'polarity': analysis.sentiment.polarity, # опре  
        'subjectivity': analysis.sentiment.subjectivity  
    }
```

Рис. 3. Код для анализа тональности

Анализ тональности текста происходит благодаря функциям `polarity` (полярность) и `subjectivity` (субъективность). Полярность измеряется от -1 до 1, где -1 – максимально негативная лексика, 0 – нейтральная лексика и 1 – максимально положительная лексика. Приведем пример предложений с положительной и негативной оценками из риторики Д. Трампа:

Положительная оценка. [1.00] *Now we're having the best results, just about, that we've ever had. [1.00] This will be our greatest area. [0.91] I believe that my life was saved that day in Butler for a very good reason.*

Негативная оценка. [-1.00] *We suffered the worst inflation in 48 years but perhaps even in the history of our country. [-1.00] That morning, Laken was viciously attacked, assaulted, beaten, brutalized, and horrifically murdered. [-0.91] It's demeaning for women, and it's very bad for our country.*

Стоит также привести примеры таких предложений в риторике Д. Байдена:

Положительная оценка. [1.00] *We have the best economy in the world. [0.80] Here tonight is UAW President Shawn Fain, a great friend and a great labor leader. [0.70] I thank Vice President Harris for being an incredible leader defending reproductive freedom and so much more.*

Негативная оценка. [-0.75] *Hate, anger, revenge, retribution are the oldest of ideas. [-0.62] Folks at home, that's why the banks are so mad. [-0.50] Why are you angry with me providing you the opportunity for the best-educated workforce in the world?*

В отношении частотности слов применяется следующий код в рисунках 4 и 5.

```
def is_significant(token):
    if token.is_stop or token.is_punct or len(token.text) <= 2:
        return False
    if token.tag_ == "VBG" and token.pos_ == "VERB":
        return False
    unwanted_patterns = [
        r"'s$", r"n't$", r"'re$", r"'ll$",
        r"^going$", r"^gonna$", r"^wanna$",
        r"^would$", r"^could$", r"^should$"
    ]
    allowed_pos = {"NOUN", "PROPN", "ADJ", "VERB"}
    return (
        token.pos_ in allowed_pos and
        not any(re.search(patt, token.text.lower()) for patt in unwanted_patterns)
    )
```

Рис. 4. Код для подсчета частотности слов

```
def get_top_words(doc, n=15):
    words = [token.lemma_.lower() for token in doc if is_significant(token)]
    return Counter(words).most_common(n)

top_trump_words = get_top_words(doc_trump)
top_biden_words = get_top_words(doc_biden)
```

Рис. 5. Продолжение кода для подсчета частотности слов

Прежде чем подсчитать частотность слов посредством модуля *re*, необходимо задать ряд условий (рисунок 4). Первое условие не принимает в учет слова, которые являются стоп-словами, знаком пунктуации или если длина токена равна или меньше двух. Второе условие не учитывает глаголы в форме герундия или причастия. Третьим условием выступают сокращения ('s, n't, 're, 'll), разговорные формы (*going, gonna, wanna*), модальные глаголы (*would, could, should*). В случае если не реализуется какое-либо из условий, происходит подсчет частотных значимых слов (15 слов) в нижнем регистре (*token.lemma.lower*).

Результаты исследования

Получены результаты размеченных речей обоих президентов с точки зрения морфологии и синтаксиса. В отношении тональности, речь Д. Трампа стремится к субъективизму (0,49) и употреблению более положительной лексики (0,16) чем Д. Байден (0,42) и (0,14) соответственно. Самыми частотными словами у Д. Трампа оказались *thank* (69), *country* (48), *people* (47) и т. д. У Д. Байдена – *America* (44), *year* (40), *american* (34) и т. д. Такая частотность слов отображает стремление обоих политиков заручиться большей поддержкой со стороны американского народа, а также их стремление не

отдаляться от него. Данные результаты применимы для анализа особенностей риторики обоих американских президентов, а также для статистических данных.

Заключение

Корпус применим для проведения исследования по таким лингвистическим вопросам, как морфология и синтаксису, для анализа тональности текста и определения частотности значимых слов в риторике Д. Трампа и Д. Байдена во время обращения к Конгрессу США. Для наиболее подробного исследования настоящего вопроса возможно дальнейшее рассмотрение иных речей президентов, а также рассмотреть особенности применения таких риторических приемов, как метафоры, сравнения (*simile*), эпитеты.

ЛИТЕРАТУРА

1. Данилина В. В. Политическая риторика как научное направление // Вестник Московского университета. Серия 21. Управление (государство и общество). 2005. №2. URL: <https://cyberleninka.ru/article/n/politicheskaya-ritorika-kak-nauchnoe-napravlenie> (дата обращения: 10.06.2025).
2. Захаров, В. П. Корпусная лингвистика : учебник / В. П. Захаров, С. Ю. Богданова. – 3-е изд., перераб. – Санкт-Петербург : СПбГУ, 2020. – 234 с. – ISBN 978-5-288-05997-1. – Текст : электронный // Лань : электронно-библиотечная система. – URL: <https://e.lanbook.com/book/139576> (дата обращения: 10.06.2025).
3. Собственный корпус [Электронный источник]. URL: <https://colab.research.google.com/drive/1yFj56EkEh8PAMMbCmpqufexPRAEKJVbR?usp=sharing> (дата обращения: 11.06.2025).
4. Чернявская В. Е., Молодыхенко Е.Н. История в дискурсе политики: лингвистический образ «своих» и «чужих». Изд. Стереотип. – М.: ЛЕНАНД, 2018. – 200 с.
5. Donald J. Trump. 5th Article II Sec. 3 Address to a Joint Session of Congress [Электронный ресурс]. Режим доступа: <https://www.americanrhetoric.com/speeches/stateoftheunion2025.htm> (дата обращения: 11.06.2025).
6. Joe Biden. Third Presidential State of the Union Address [Электронный ресурс]. URL: <https://www.americanrhetoric.com/speeches/stateoftheunion2024.htm> (дата обращения: 11.06.2025).
7. SpaCy. Industrial-Strength Natural Language Processing [Электронный ресурс]. URL: <https://spacy.io/> (дата обращения: 11.06.2025).
8. TextBlob: Simplified Text Processing [Электронный ресурс]. URL: <https://textblob.readthedocs.io/en/dev/> (дата обращения: 11.06.2025).