

ФИЛОЛОГИЯ

УДК 811.111 + 81'33 + 81'371

Д.В. Спиридонов, М.О. Гузикова

СТАТИСТИЧЕСКИЕ МЕТОДЫ И ПРОБЛЕМА ИЗУЧЕНИЯ ТЕКСТОВОГО КОНЦЕПТА (НА МАТЕРИАЛЕ ЗАЯВЛЕНИЯ ИСПАНИИ ПО ВОПРОСУ О САМОПРОВОЗГЛАШЕНИИ НЕЗАВИСИМОСТИ КОСОВО В МЕЖДУНАРОДНОМ СУДЕ ООН)

Рассматривается возможность применения количественных методов для определения лексических репрезентантов текстового концепта и их связи с идеологическим содержанием текста. На примере анализа заявления Испании по вопросу о самопровозглашении независимости Косово авторы анализируют семантические параметры концепта «суверенитет». Используемая методика основана на учете частотности как прямых, так и косвенных лексических репрезентантов концепта, а также их качественной семантической оценки.

Ключевые слова: политический дискурс; концепт; идеология; лексикостатистика; квантитативный анализ.

В лингвистической науке изучение текстовых концептов традиционно производится путем качественного анализа текстового материала (ср.: [1–3 и мн. др.]). В то же время возможности количественного, в том числе автоматизированного анализа, кажется, не получили должного внимания исследователей. Данная перспектива намечалась в различных работах, например [4], однако алгоритм такого анализа еще ждет своего построения.

Нас будет интересовать возможность применения количественных методов обработки текста применительно к концептам политического дискурса. В своем анализе мы будем исходить из следующих общих соображений:

1. Исследуемый текст является носителем некоторой идеологии, под которой мы понимаем связный образ идеальной реальности. Внутри исследуемого текста выделяются смысловые доминанты, логические отношения между которыми определяют структуру данной идеологии. При этом всякая идеология обладает определенным языковым коррелятом. В данном случае описание идеологии можно понимать как описание ее языка, т.е. системы понятий, структурирующих тот образ реальности, который лежит в основе этой идеологии. Такие базовые понятийные образования, обладающие регулярными средствами выражения в языке соответствующей идеологии, мы для удобства будем именовать «концепты».

2. Вслед за Майклом Фриденом [5, 6] мы будем исходить из того, что идеологии представляют собой комплексы или поля концептов, которые используются в рамках этих идеологий некоторым устойчивым, повторяющимся образом и фиксируют наиболее существенные фрагменты того идеального образа реальности, который лежит в основе данных идеологий. Внутри таких полей выделяются ядерные и периферийные концепты, кроме того, концепты могут вступать в отношения подчинения, быть равноправными или соположенными друг другу. Понять идеологию означает выделить центральные концепты и определить их содержание.

3. В конкретных политических (идеологических) текстах всякий концепт имеет разный набор конкретных лексических репрезентантов – лексем и фразеологических номинаций, называющих соответствующее понятие.

Трудность анализа заключается в том, что в коммуникации, в том числе политической, один и тот же тер-

мин может соотноситься с разными концептами (например, термин *справедливость* в дискурсе либералов и социалистов обладает разным понятийным содержанием). В этой связи уместно напомнить типологию идеологем А.П. Чудинова, выделявшего, с одной стороны, идеологемы, которые «используются только сторонниками определенных политических взглядов, соответствующие наименования передают специфический взгляд на соответствующую реалию», и с другой – идеологемы, смысл которых «неодинаково понимается сторонниками различных политических взглядов», но которые, однако, используются внутри разных идеологий [7. С. 92–93]. В последнем случае мы имеем дело с фактической (контекстуальной) многозначностью, репрезентирующей данный идеологический концепт лексемы. Эта многозначность не отражена в словаре, однако существенна при изучении идеологических различий и их фиксации в текстах политического характера.

Одним из идеологических концептов, активно используемых в современной политической коммуникации, является концепт «суверенитет». С одной стороны, лексема *суверенитет* относится к числу терминов международного права и, будучи термином, должна обладать семантической однозначностью и соответствующей межъязыковой эквивалентностью. С другой стороны, реальное ее употребление подчинено политическим интересам продуцентов разного рода текстов, в которых обсуждается суверенитет тех или иных государств или территориальных образований, и потому фактическое значение понятия в разных текстах различно (см. об этом: [8, 9 и др.]). Мы будем рассматривать это понятие в качестве идеологического концепта, фактическое значение которого подчинено соответствующей идеологической интенции.

Традиционными репрезентантами концепта «суверенитет», которые встречаются как в политико-юридическом, так в академическом дискурсе, выступают слова и словосочетания *независимость, автономия, территориальная целостность, монополия на применение силы и контроль над территорией*. При этом концептуальное наполнение понятия «суверенитет» зависит от преобладания тех или иных входящих в него семантических признаков, актуализируемых определенными лексическими репрезентантами.

Все лексические репрезентанты удобно при этом делить на два класса: «прямые» и «косвенные». К прямым относятся те репрезентанты, которые мы используем в рамках своей теоретической модели для именования соответствующего концепта. В нашем случае прямым репрезентантом концепта «суверенитет» является термин *суверенитет*. К косвенным репрезентантам мы отнесем все прочие потенциально возможные контекстные синонимы: *независимость, автономия, территориальная целостность, монополия на применение силы* и пр.

Мы рассмотрим репрезентанты концепта «суверенитет» на примере корпуса текстов, представляющих собой заявления 37 стран, направленные ими в Международный суд ООН в ответ на запрос суда о соответствии самопровозглашения независимости Косово международному праву [10]. Большие объемы текстового материала данного корпуса определяют неизбежность обращения к статистическим методам концептуального анализа, а также подталкивают к построению алгоритма такого анализа, который бы мог оказаться продуктивным при решении сходных задач на другом текстовом материале.

Предлагаемая схема анализа включает следующие этапы:

1. Выявление частотности «прямых» репрезентантов концепта в разных текстах корпуса с целью установления корреляции между позицией продуцента текста (и соответственно, его идеологической базой) и значимостью соответствующего концепта в продуцируемом им дискурсе.

2. Выявление в пределах отдельного текста частотных соответствий между прямыми репрезентан-

тами и прочими лексемами с целью установления «косвенных» репрезентантов концепта.

3. Анализ частотности отдельных лексем-репрезентантов и реконструкция на ее основе «идеологического» содержания соответствующего концепта.

Итак, обратимся к исследуемому корпусу. Все входящие в него тексты представлены на английском языке, их объем варьирует от одной до 372 страниц, от 375 до 575 838 знаков. Как известно, из 37 стран 22 страны высказались за предоставление Косово суверенитета, 15 заняли противоположную позицию. Табл. 1 иллюстрирует объем корпуса, количество вхождений искомым слов *sovereignty* и *sovereign*, выступающих в данном случае «прямыми» репрезентантами концепта.

Так как из табл. 1 видно, что объемы текстов стран различны, то для корректного сравнения количества употреблений искомым слов приведем абсолютную частоту к относительной. Для этого в лексической статистике используется показатель *ipm* – index per million, но при данных объемах показатель приведен к 100 тысячам слов (*ipht* – index per one hundred thousand).

Анализ частотности употреблений искомым лексем позволяет прийти к выводу, что обращение к концепту «суверенитет» в заявлениях отрицательно настроенных стран происходит в три (а точнее в 3,08 раза) чаще, чем в текстах стран – сторонниц независимости Косово. Так, в тексте Германии (высказалась «за»), содержащем 22 871 слово, показатель *ipht* – 104, а в тексте Румынии (голосовала «против»), содержащем 17 326 слов, этот показатель – 207, что при сопоставлении объема двух текстов почти в три раза выше.

Таблица 1

Количественные показатели объема исследуемого корпуса, количество вхождений искомым слов

Показатель	Всего в корпусе	Тексты стран «за»	Тексты стран «против»
Всего слов	569 687	265 946	303 741
Вхождения искомым слов <i>sovereignty</i> и <i>sovereign</i>	1 313	292	1 021
Слов в контекстах на расстоянии 5 слов вправо и влево / употреблений слов	13 130/1 794	2 920/566	10 210/1 228

Таблица 2

Частотность употребления искомым слов на 100 тыс. слов

Показатель	Тексты стран «за»	Тексты стран «против»
Вхождения искомым слов <i>sovereignty</i> и <i>sovereign</i>	109,8 ipht	336 ipht

Таблица 3

Количество вхождений слов *sovereignty* и *sovereign* (приведены только те страны, в текстах которых количество употреблений этих слов больше 10)

№	Страна	За / Против	Кол-во употреблений искомым слов	ipht
1	Сербия	Против	261	24
2	Испания	Против	147	49
3	Косово	За	88	18
4	Кипр	Против	65	21
5	Румыния	Против	51	29
6	США	За	50	10
7	Россия	Против	47	32
8	Аргентина	Против	34	20
9	Германия	За	29	13
10	Китай	Против	22	85
11	Албания	За	21	10
12	Австрия	За	17	21
13	Египет	Против	17	25
14	Финляндия	За	16	40
15	Великобритания	За	14	3
16	Швейцария	За	14	10
17	Чехия	За	11	19
18	Польша	За	10	10

Как видно из табл. 3, количество употреблений искомых слов нерепрезентативно для установления частотности употребления. Индекс *ipht*, напротив,

четко указывает на частотность употребления. Таким образом, исследуемые тексты могут быть ранжированы следующим образом.

Таблица 4

Частотность употребления искомых слов *sovereignty* и *sovereign*

№	Страна	За / Против	Кол-во употреблений искомых слов	<i>ipht</i>
1	Китай	Против	22	85
2	Испания	Против	147	49
3	Финляндия	За	16	40
4	Россия	Против	47	32
5	Румыния	Против	51	29
6	Египет	Против	17	25
7	Сербия	Против	261	24
8	Австрия	За	17	21
9	Кипр	Против	65	21
10	Аргентина	Против	34	20
11	Чехия	За	11	19
12	Косово	За	88	18
13	Германия	За	29	13
14	Албания	За	21	10
15	США	За	50	10
16	Польша	За	10	10
17	Швейцария	За	14	10
18	Великобритания	За	14	3

Из сравнения табл. 3 и 4 видно, что Испания является той страной, которая имеет наивысшие показатели и по количеству, и по частотности употреблений искомых слов *sovereignty* и *sovereign*. На примере текста заявления Испании [11] мы попытаемся определить, какие репрезентанты концепта «суверенитет» имеют приоритет для этой страны и насколько данный показатель сравним со средними значениями для двух групп стран. Для этого мы вновь обратимся к лексической статистике.

При анализе было выбрано примерно равное количество самых частотных лексических партнеров из подкорпуса текстов стран, высказавшихся в пользу

Косово, – это 10 слов (употребляются как минимум 9 раз), и из подкорпуса текстов негативно настроенных стран – 17 слов (употребляются как минимум 30 раз). Такая диспропорция объясняется тем, что, как мы установили ранее, искомые слова в последнем подкорпусе текстов употребляются как минимум в три раза чаще. Кроме того, из выборки были исключены все служебные части речи, а также все глагольные формы. Это обусловлено тем, что нас интересуют репрезентанты концепта «суверенитет», которые могут быть либо существительными, либо сочетаниями существительного и прилагательного.

Таблица 5

Наиболее частотные лексические партнеры искомых слов

№	Контексты стран «за»		Контексты стран «против»	
	Кол-во раз	Слово	Кол-во раз	Слово
1	132	<i>integrity</i>	589	<i>territorial</i>
2	129	<i>territorial</i>	568	<i>integrity</i>
3	109	<i>state(s)</i>	340	<i>state(s)</i>
4	59	<i>independent</i>	210	<i>principle(s)</i>
5	14	<i>political</i>	152	<i>respect</i>
6	14	<i>unity</i>	161	<i>Serbia(s)</i>
7	13	<i>Serbia</i>	85	<i>independent</i>
8	11	<i>accords</i>	83	<i>Kosovo</i>
9	10	<i>rights</i>	78	<i>equality</i>
10	9	<i>law</i>	70	<i>independence</i>
11			62	<i>political</i>
12			50	<i>unity</i>
13			40	<i>international</i>
14			34	<i>commitment</i>
15			33	<i>territory</i>
16			32	<i>resolution</i>
17			30	<i>member</i>

Данные табл. 5 показывают, что и по количеству «косвенных» репрезентантов концепта «суверенитет» страны, высказавшиеся положительно, уступают странам, выступившим против предоставления Косово независимости. Также можно заметить, что в обеих группах текстов «прямые» репрезентанты «связываются» со сходным набором лексем. Эти лексемы образуют класс «косвенных» репрезентантов концепта,

общих для обеих групп текстов. Так, в нашей выборке можно выделить такие «косвенные» репрезентанты концепта «суверенитет», как *territorial integrity*, *political unity* и *independence*.

В тексте заявления Испании словосочетание *territorial integrity* встречается 146 раз, *(political) unity* – 38, а *independence* – 34. В начале нашей статьи были приведены такие наиболее распространенные репре-

зентанты концепта, как *независимость, автономия, территориальная целостность, монополия на применение силы и контроль над территорией*. Из приведенного списка только два эквивалентных англоязычных репрезентанта встречаются в тексте Испании.

Помимо них, в данном случае особый интерес представляют еще два репрезентанта – *autonomy* и *border control*. Что касается слова *autonomy*, то в заявлении Испании оно встречается 10 раз, однако во всех случаях обозначает нечто противоположное государственному суверенитету, а именно – предоставление расширенных прав той или

иной территории при сохранении территориальной целостности государства (ср.: «the people of Kosovo can enjoy substantial autonomy within the Federal Republic of Yugoslavia»). Таким образом, в заявлении Испании лексема *autonomy* связана не с концептом «суверенитет», а с концептом «автономия», который относится к концепту «суверенитет» как его противоположность или альтернатива: территория выбирает между автономией и суверенитетом. Выступая против представления Косово суверенитета, Испания ратует за предоставление этой территории широкой автономии внутри Сербии.

Таблица 6

Частотность наиболее употребляемых слов в заявлении Испании в сравнении с группой стран, выступающих против предоставления Косово суверенитета

№	Слово	Количество употреблений	
		Страны «против»	Испания
1	territorial	589	152
2	integrity	568	132
3	state(s)	340	40
4	principle(s)	210	50
5	respect	152	63
6	Serbia(s)	161	45
7	independent	85	8
8	Kosovo	83	202
9	equality	78	7
10	independence	70	44
11	political	62	86
12	unity	50	38
13	international	40	240
14	commitment	34	22
15	territory	33	15
16	Resolution	32	221
17	member	30	24

Еще один распространенный репрезентант «суверенитета» – *border control* – встречается в тексте лишь однажды (ср.: «border control or customs control, are part of the core content of State sovereignty»). Единичный характер употребления данного словосочетания в качестве репрезентанта соответствующего концепта определяется периферийным (в данном конкретном тексте) положением фиксируемого им признака в семантике концепта «суверенитет».

Другим важным диагностическим инструментом количественного анализа должно быть сравнение частотности соответствующих коллокатов в исследуемом тексте и в языке в целом. Так, в подкорпусе Британского национального корпуса BNC (non-academic, law) [12] частотность коллокаций (4–4) для слова *sovereignty* выглядит следующим образом (включены слова с частотой биграммы не менее 10).

Таблица 7

Частотность коллокаций (4–4) лексемы *sovereignty* в BNC

№	Коллокат	Частотность	Всего употреблений	%
1	over	48	5 549	0,87
2	declaration	22	560	3,93
3	full	18	1 217	1,48
4	independence	17	1 089	1,56
5	territorial	14	280	5,00
6	republic	14	1 600	0,88
7	declarations	13	74	17,57
8	integrity	12	130	9,23
9	islands	10	433	2,31

Из приведенной выше таблицы видно, что *independence* и *territorial integrity* относятся к наиболее частотным коллокатам слова *sovereignty*, при этом *independence* встречается чаще. В случае заявления Испании частота репрезентанта *territorial integrity* выше, чем *independence*, что, вероятно, объясняется характером текста: Испания стремится защитить суверенитет Сербии, понимаемый как ее «территориальная целостность»; в то же время «независимость» в тексте относится прежде всего к стремлению Косово ее обрести – по этой при-

чине соответствующие репрезентанты хотя и соотносятся с одним и тем же концептом в семантической системе исследуемого текста, de facto наделяются в данном тексте разной смысловой нагрузкой и разной оценочностью: *territorial integrity* мыслится как то, что должно быть защищено, в то время как *independence* – это то, чему следует воспрепятствовать. Распределение идеологических акцентов определяет в конечном счете и количественные параметры употребления соответствующих лексем.

Думается, что проведенный анализ наглядно демонстрирует надежность и потенциальную значимость лексико-статистических данных как при изучении отдель-

ных концептов внутри политического дискурса, так и при решении задач комплексной лингвистической реконструкции разных политических идеологий.

ЛИТЕРАТУРА

1. *Современная политическая лингвистика* : учеб. пособие / Э.В. Будаев, М.Б. Ворошилова, Е.В. Дзюба, Н.А. Красильникова ; отв. ред. А.П. Чудинов ; Урал. гос. пед. ун-т. Екатеринбург, 2011. 252 с.
2. Демьянков В.З. Политический дискурс как предмет филологической науки // *Политическая наука. Политический дискурс: История и современные исследования*. 2002. № 3. С. 32–43.
3. Шейгал Е.И. Семiotика политического дискурса : дис. ... д-ра филол. наук. Волгоград, 2000. 431 с.
4. Баранов А.Н., Михайлова О.В., Сатаров Г.А., Шупова Е.А. Политический дискурс: методы анализа тематической структуры и метафоризма. М. : Фонд ИНДЕМ, 2004. 94 с.
5. Freedman M. Ideologies and political theory: a conceptual approach. Oxford : Clarendon Press ; New York : Oxford University Press, 1996. 592 p.
6. Freedman M. Liberal languages: ideological imaginations and twentieth-century progressive thought. Princeton, NJ : Princeton University Press, 2005. 271 p.
7. Чудинов А.П. Политическая лингвистика. М. : Флинта; Наука, 2006. 256 с.
8. Kutulus E. State sovereignty: concept, phenomenon and ramifications. N. Y. : Palgrave Macmillan, 2005. 232 p.
9. *Sovereignty in Fragments: Past, Present and Future of a Contested Concept* / ed. by H. Kalmo, K. Skinner. Cambridge : CUP, 2010. 269 p.
10. *Accordance with international law of the unilateral declaration of independence in respect of Kosovo (Request for Advisory Opinion)* // Cour internationale de justice = International Court of Justice. URL: <http://www.icj-cij.org/docket/index.php?p1=3&p2=4&k=21&case=141&code=kos&p3=0>
11. *Written Statement of Spain* // Cour internationale de justice = International Court of Justice. URL: <http://www.icj-cij.org/docket/files/141/15644.pdf>.
11. BYU-BNC: British National Corpus. URL: <http://corpus.byu.edu/bnc>.

Статья представлена научной редакцией «Филология» 9 ноября 2015 г.

STATISTIC METHODS IN THE STUDY OF A TEXTUAL CONCEPT (WITH REFERENCE TO THE STATEMENT OF SPAIN ON THE UNILATERAL DECLARATION OF INDEPENDENCE BY KOSOVO PRESENTED TO THE INTERNATIONAL COURT OF JUSTICE)

Tomsk State University Journal, 2015, 400, 6–10. DOI: 10.17223/15617793/400/1

Spiridonov Dmitry V. Ural Federal University (Ekaterinburg, Russian Federation). E-mail: dspiridonov@mail.ru

Guzikova Maria O. Ural Federal University (Ekaterinburg, Russian Federation). E-mail: mariagu@mail.ru

Keywords: political discourse; concept; ideology; lexical statistics; quantitative analysis.

The article explores the potential of statistical approaches for the retrieval of lexical representatives of a textual concept as well as for the study of their correlations with the ideological content of the text. Based on Michael Freedman's concept-oriented theory of ideology, the authors suggest to study the lexical representation of the concept "sovereignty" in the statements on the unilateral declaration of independence by Kosovo presented to the International Court of Justice, which implicitly reflect the ideological orientations of 37 countries as for the idea of sovereignty and its status in the international law. To do this, the authors calculate the frequency of both direct and indirect representatives of the concept within the statements made by 37 countries. This analysis shows that the frequency of words which can be regarded as potential representatives of the concept in question is three times higher in the texts of countries opposing Kosovo's independence than of those supporting it, this reveals the "argumentative potential" of the concept in the discussion on Kosovo's independence. The statement made by Spain, strongly opposed to the idea of the independence of Kosovo, has the highest number of words potentially representing the concept "sovereignty". The authors then proceed to the statistical analysis of collocations in Spain's statement which leads to retrieve the indirect representatives of the concept in question relevant for the particular text. These are: *territorial integrity*, *political unity* and *independence* (only nouns or combinations of nouns and adjectives were retained). Obviously, they correspond to the various aspects of sovereignty as defended by Spain. Besides these representatives, the authors analyze two statistically specific units: *autonomy* and *border control*. The authors show that the uniqueness of *border control* in the statement of Spain reflects the peripheral character of the corresponding meaning in the structure of the studied concept, while the qualitative analysis of *autonomy* reveals that it reflects the meaning opposing the idea of independence (autonomy is suggested as an alternative to independence) and cannot be considered as a representative of the concept "sovereignty" in this particular text. The authors conclude that the statistical methods and the techniques of automatic treatment of texts can be useful for the preliminary lexical and semantic analysis of political concepts as reflected in ideologically non-neutral texts.

REFERENCES

1. Budaev, E.V. et al. (2011) *Sovremennaya politicheskaya lingvistika* [Modern political linguistics]. Ekaterinburg: Ural State Pedagogical University.
2. Dem'yankov, V.Z. (2002) Politicheskii diskurs kak predmet filologicheskoy nauki [Political discourse as an object of philological science]. In: Gerasimov, V.I. & Il'in, M.V. (eds) *Politicheskaya nauka. Politicheskii diskurs: Istoriya i sovremennye issledovaniya* [Political science. Political discourse: history and modern studies]. Is. 3. Moscow: INION RAN.
3. Sheygal, E.I. (2000) *Semiotika politicheskogo diskursa* [Semiotics of political discourse]. Philology Dr. Diss. Volgograd.
4. Baranov, A.N. et al. (2004) *Politicheskii diskurs: metody analiza tematicheskoy struktury i metaforiki* [Political Discourse: the methods of analysis of thematic structure and metaphors]. Moscow: Fond INDEM.
5. Freedman, M. (1996) *Ideologies and political theory: a conceptual approach*. Oxford: Clarendon Press; New York: Oxford University Press.
6. Freedman, M. (2005) *Liberal languages: ideological imaginations and twentieth-century progressive thought*. Princeton, N.J.: Princeton University Press.
7. Chudinov, A.P. (2006) *Politicheskaya lingvistika* [Political Linguistics]. Moscow: Flinta; Nauka.
8. Kutulus, E. (2005) *State sovereignty: concept, phenomenon and ramifications*. New York: Palgrave Macmillan.
9. Kalmo, H. & Skinner, K. (eds) (2010) *Sovereignty in Fragments: Past, Present and Future of a Contested Concept*. Cambridge: CUP.
10. International Court of Justice. (2008) *Accordance with international law of the unilateral declaration of independence in respect of Kosovo (Request for Advisory Opinion)*. Available from: <http://www.icj-cij.org/docket/index.php?p1=3&p2=4&k=21&case=141&code=kos&p3=0>
11. International Court of Justice. (2009) *Written Statement of Spain*. Available from: <http://www.icj-cij.org/docket/files/141/15644.pdf>
12. BYU-BNC: British National Corpus. Available from: <http://corpus.byu.edu/bnc/>

Received: 09 November 2015