

УДК 930:81'37

СЕМАНТИЧЕСКИЙ КОНТЕНТ-АНАЛИЗ ВЫСТУПЛЕНИЙ ДЕПУТАТОВ ГОСУДАРСТВЕННОЙ ДУМЫ РОССИЙСКОЙ ИМПЕРИИ: МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ

А. В. Сметанин

Пермский государственный национальный исследовательский университет, 614990, Пермь, ул. Букирева, 15
smetanin.av@gmail.com

Предлагается методика контент-анализа, дающая возможность преодолеть ограниченность частотного разложения текста. Методика основана на инструментарии сетевого анализа, что позволяет исследователю выявлять связи между значимыми понятиями в нарративе. Методика верифицирована на материалах выступлений депутатов Государственной Думы Российской империи в начале XX в., продемонстрирован её эвристический потенциал. Предлагаемая методика может быть использована для реконструкции системы политических категорий конкретных депутатов.

Ключевые слова: семантический анализ текста, сетевой анализ, история идей, Государственная Дума Российской империи, фракция кадетов.

Одним из традиционных направлений квантификации исторической науки является количественный анализ текстов, успешно применяемый при исследовании разных периодов истории. Первый опыт изучения депутатского нарратива дореволюционной Государственной Думы средствами компьютеризированного контент-анализа относится к 1996 г., ко времени работы семинара в МГУ под руководством Н.Б. Селунской и Л.И. Бородкина [Становление..., 1996]. Исследователи анализировали выступления депутатов Думы I созыва (1906 г.) и пришли к выводу, что в дискуссиях прослеживается лишь одна оформленная позиция – «кадетская». Работа выполнялась с помощью канадской программы ТАСТ и базировалась преимущественно на частотных характеристиках текста с некоторыми элементами семантического анализа.

Зарубежные исследователи справедливо отмечают, что в настоящее время в общественных науках количественный анализ текста связан преимущественно с подсчётом изолированных слов и редким использованием более сложных методов реконструкции семантических связей [Handbook of Multimethod..., 2006, p.155]. Между тем частотный анализ имеет ограничения, поскольку суть его заключается в подсчёте количества употребленных лексем (или групп лексем), при этом предполагается прямая связь между значительным количеством употреблений лексемы и её важностью для автора. Подобная методика даёт возможность пренебречь вероятностью того, что некоторые слова автор упоминает автоматически ввиду небольшого лексического запаса или привычки, а некоторые важные для него категории, наоборот, могут использоваться лишь в особых случаях, чтобы подчеркнуть их смысловую нагрузку. Семантический анализ позволяет оценить важность каждой лексемы в общей системе понятий, её место во взаимосвязи с другими речевыми единицами.

В качестве понятного и наглядного инструмента семантического анализа в данной работе предлагается сетевой анализ. Идея использования сетевого контент-анализа на теоретическом уровне была разработана в течение 1990-х гг. [Roberts, 2000, p.265], однако объём практических работ в данном направлении, особенно по истории, всё ещё необходимо признать скудным. В качестве положительного примера применения этого метода можно указать на исследование Р.В. Топки, посвящённое крестьянским наказаниям депутатам при выборах в Государственную Думу в 1906 г. [Топка, 2001]. Он проанализировал взаимную встречаемость конкретных пожеланий в наказах и выявил некоторые закономерности крестьянского политического мышления начала XX в. Впрочем, автор не производил расчёта сетевых характеристик и ограничился созданием самой сети.

В настоящем исследовании сетевому контент-анализу подвергнуты материалы выступлений депутатов Государственной Думы Российской империи. Тексты выступлений являются важнейшим источником по политической истории 1906–1917 гг., который поможет уточнить наши представления об идеях и взглядах отдельных депутатов, фракций, политических течений, социальных и этнических групп, а также общества в целом. Тексты всех выступлений сосредоточены в 32 томах

стенографических отчётов Думы [Россия. Государственная..., 1906–1917], которые и послужили основой исследований.

Алгоритм предлагаемой методики содержит общую часть и вариативную, зависящую от целей изучения. Общая часть традиционна для частотного контент-анализа.

1) Подготовка корпуса текстов. Сетевой анализ не предполагает каких-либо дополнительных требований к обработке или кодированию текстов, однако эти требования могут диктоваться особенностями источника. Например, в стенограммах Думы помимо речи оратора фиксировались выкрики с мест, и эти ремарки необходимо отделять от основного текста, поскольку они зачастую содержат лексемы, совершенно несовместимые с позицией говорящего. Также остро стоит проблема репрезентативности выборки. Выступления депутатов значительно различаются по объёму: зачастую это короткие мнения, но встречаются речи, длившиеся до четырёх часов¹, соответственно различаются и степень информативности текстов. Стоит учитывать и то, что если в Думе I и II созывов преобладали выступления общеполитического характера, то с 1907 г. основная часть ее работы приходится на рассмотрение узкоспециальных вопросов и регламента работы Думы. Очевидно, что при обсуждении частных законопроектов личная позиция оратора проявляется довольно слабо.

2) Составление и обработка частотного словаря корпуса текстов. Данную процедуру выполняют любые программы компьютеризированного контент-анализа. Текст разбивается на речевые единицы, и подсчитывается количество упоминаний той или иной единицы в тексте.

3) Выделение категорий. Именно этот пункт является вариативной частью алгоритма. Под категорией понимается лексема или группа лексем, объединённых общим смыслом; в англоязычной литературе «категории» соответствует понятие *theme* (тема). Категории формируются исследователем в зависимости от поставленных задач и особенностей текста. Например, в настоящей статье употребляется категория «национальность» – термин, который употреблялся в начале XX в. для обозначения этносов, соответственно при формировании категории в неё включаются понятия, которые в тот период считались синонимичными: национальность, народность, этнос со всеми словоформами. Термин «народ» включить в категорию нельзя, поскольку он, как правило, употреблялся в значении «население». Наличие категорий помогает сократить частотный словарь до ограниченного числа смысловых единиц.

Предлагаемая методика предусматривает два возможных сценария исследования. В первом случае ставится задача реконструировать систему значимых понятий (категорий) на основе имеющегося корпуса текстов. Следование этому сценарию позволяет выявить ключевые идеи автора нарратива, их взаимосвязь, даже если мы ничего не знаем о ценностях этого автора. Второй сценарий предполагает предварительное составление списка интересующих категорий и в дальнейшем поиск в нарративе отражения этих категорий, определение степени их значимости для автора. Выбор сценария зависит от целей исследования.

4) Создание матрицы взаимной встречаемости категорий (*co-occurrences matrix*). Данная матрица является основой для создания семантической сети. Взаимная встречаемость в исследовании фиксируется в том случае, если любые лексемы двух категорий располагаются на расстоянии не более семи слов друг от друга, при желании этот интервал может быть как увеличен, так и уменьшен. Предполагается, что если слова часто встречаются в тексте поблизости друг от друга, то они для оратора составляют смысловую общность, т.е. можно говорить о семантической связи лексем и категорий.

Для создания матрицы подсчитываются все случаи встречаемости каждой категории с каждой, и матрица принимает определенный вид (табл. 1).

Подобный формат матрицы бывает неудобен, поскольку объём сравниваемых текстов может сильно различаться, и чем больше текст, тем больше фиксируется взаимных совпадений категорий. Предпочтительней использовать относительные показатели (меры сходства), которые не зависят от объёма текста, например, распространённый в исследованиях коэффициент Жаккара [Гайдышев, 2004, с.222].

5) Проведение сетевого анализа. Полученная симметричная матрица позволяет создать сеть и рассчитать необходимые сетевые характеристики. Для этого используется специальное программное обеспечение, в данной работе – американская программа для научной работы UCInet 6.0. Конкретные примеры сетевых моделей и сетевых расчётов приведены далее.

Таблица 1

Пример матрицы взаимной встречаемости (абсолютные значения).

Категории	Категория			
	1	2	3	4
1				
2	4			
3	0	6		
4	5	11	2	

Поскольку предложенный алгоритм имеет вариативный компонент, т.е. два сценария выделения категорий, то для каждого из сценариев далее демонстрируется пример из исследовательской практики.

Сценарий 1. Исследование текстового корпуса с целью выделения системы категорий.

Если исследование подразумевает максимально полную реконструкцию взглядов автора нарратива, то стоит придерживаться данного сценария. Кроме того, он может быть полезен в ситуации, когда взгляды автора (или группы авторов) изучены мало и нет предварительной гипотезы о возможных ценностях данного человека.

В качестве примера реализации подхода без предварительного выделения категорий рассмотрим анализ выступлений депутата И.И. Гайдарова. Депутат был избран в Думу III созыва от Дагестанской области. Будучи дворянином и человеком с высшим образованием [Государственная Дума..., 2008, с.119], он, однако, вступил во фракцию социал-демократов, а спустя два года перешёл в Мусульманскую группу. Ситуация примечательна тем, что социал-демократы занимали в Думе крайние левые скамьи, в то время как Мусульманская группа держалась умеренного центра. Одно время даже обсуждался вопрос о ее слиянии с октябристами². Таким образом, целью анализа выступлений И.И. Гайдарова является политическая идентификация депутата.

Для анализа были отобраны стенограммы первой сессии Думы III созыва (1907 – 1908 гг.), когда И.И. Гайдаров находился в составе социал-демократов. За этот период депутат успел выступить семь раз, суммарный текстовый корпус составил 5,6 тыс. слов. В результате экспертной редукции частотного словаря было выделено 40 смысловых категорий. Установление связей между ними позволяет выявить существенные характеристики взглядов оратора.

Во-первых, можно определить категории, которые лидируют не только по частоте встречаемости, но и согласно сетевым характеристикам, т.е. важнейшие для депутата.

В качестве основного показателя при определении значимости категорий в сети использован индекс сетевой промежуточности (betweenness). Он дает возможность оценить, насколько важен тот или иной узел в качестве транзитного пункта, для связи между сегментами сети³ [Градосельская, 2004, с. 69–70]. Поскольку в текстовый корпус входит несколько речей депутата, посвящённых разной тематике, подобные промежуточные узлы позволяют отслеживать категории (идеи), появившиеся во всех выступлениях этого депутата, связывающие отдельные смысловые блоки между собой.

На схеме размер узла зависит от частоты упоминания соответствующей категории в тексте. Форма узла отражает значимость категории в рамках сети: крупнейшие узлы (индекс промежуточности не менее 50% от индекса категории-лидера) обозначены кругами, все остальные (менее значимые) – квадратами. Длина линии связей зависит от частоты взаимной встречаемости двух категорий; чем короче линия, тем чаще эти категории употребляются в тексте совместно. Например, у И.И. Гайдарова категории «земля» и «налоги» находятся очень близко, следовательно, они употреблялись в речи депутата в одних и тех же ситуациях, и в системе взглядов эти категории выражают интересы одного порядка. В качестве иного примера можно рассмотреть категории «политика» и «правительство», связь между которыми достаточно слабая.

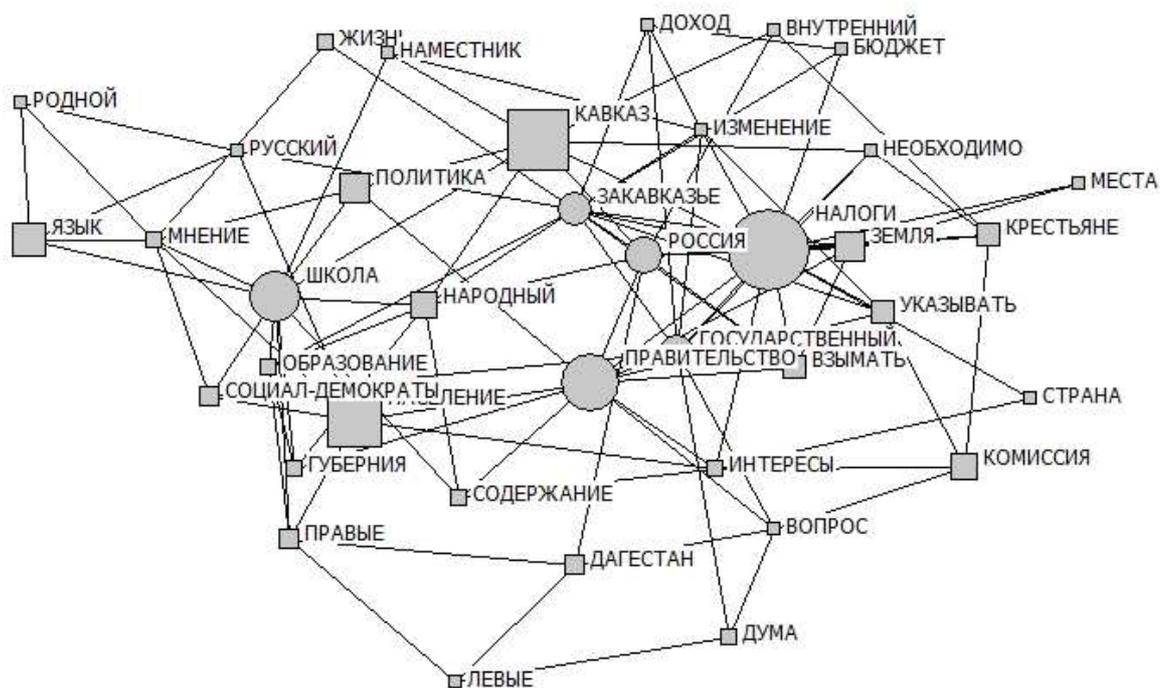


Рис. 1. Сетевая модель нарратива И.И. Гайдарова (1907 – 1908 гг.)

Нетрудно убедиться в том, что частота употребления категории в тексте не свидетельствует о ее важном месте в сетевой модели, иными словами, в системе взглядов автора.

Таблица 2
Наиболее часто употребляемые И.И. Гайдаровым категории (на 1000 слов)

Категория	Частота	Коэффициент промежуточности
Налоги	7,1	106,3
Правительство	5,5	81,2
Кавказ	5,4	40,6
Школа	5,0	71,8
Население	4,6	35,4

К важнейшим элементам исследуемого нарратива нужно отнести категории «налоги», «школа», «правительство». Они имеют как высокий частотный показатель, так и высокий коэффициент промежуточности. Данные приоритеты не позволяют напрямую идентифицировать И.И. Гайдарова только как социал-демократа или исключительно «мусульманина». Темы налогов и действий правительства были важнее для левых, тогда как вопросы образования являлись предметом постоянных забот Мусульманской группы. С учетом того, что подготовка речей ораторов социал-демократической фракции находилась под строгим партийным контролем, неудивительно, что в данном списке преобладали социал-демократические интересы.

Помимо явных приоритетов методика позволяет выявить латентные приоритеты, т.е. редко обсуждаемые, но занимающие важное место в понятийной системе автора нарратива. Такие категории имеют относительно низкую частотность и достаточно высокий коэффициент промежуточности. Так, вторым по значимости узлом в сети И.И. Гайдарова является категория «Закавказье», занимающая в частотном словаре лишь девятое место. Сюда же причислим категории «Россия» и «государственный». Иными словами, на латентном уровне у Гайдарова озвучивается повестка центр-периферийных отношений, которая в большей степени характеризует Мусульманскую груп-

пу, а не социал-демократов. Вероятно, сохранение подобного габитуса облегчило переход в умеренно-либеральный центр Думы после двух лет пребывания на левом фланге.

Предполагаемая методика способна определить малозначимые, но часто упоминаемые понятия, которые условно можно назвать «фигурами речи». Для автора текста эти категории не наполнены важным содержанием. Таковых категорий у И.И. Гайдара две: «Кавказ» и «население». Интересно, что по сравнению со столь значимым «Закавказьем» понятие «Кавказ» для И.И. Гайдара оставалось в большей степени абстрактным географическим и административным определением, связанным с деятельностью наместника. Второстепенность категории «население» и вовсе противоречит социал-демократической идеологии. Две категории, «время» и «деньги», в системе взглядов И.И. Гайдара оказались изолированы, не обнаружили связи ни с одной другой категорией и, как следствие, не нашли отражения на рис. 1.

Моделирование системы взглядов И.И. Гайдара позволяет заключить, что в публичных выступлениях депутата затрагивались типично левые вопросы, тогда как на латентном уровне сохранялись приоритеты всех мусульманских депутатов, которые в конечном счёте качнули чашу весов в свою сторону. Двойственность политического мышления И.И. Гайдара согласуется с наблюдением Д.М. Усмановой, отметившей, что данный депутат в 1907 – 1909 гг. инициировал практически все запросы, исходившие от мусульман и касавшиеся мусульманских проблем, но делал это при поддержке социал-демократической фракции [Усманова, 2005, с. 272].

Сценарий 2. Исследование корпуса текстов с целью определения выраженности заранее выделенных категорий.

Если в первом случае мы отказывались от априорного выделения категорий и составляли их перечень на основе частотного анализа текста, то для данного класса задач применяется обратный алгоритм. Подобный подход продуктивен при необходимости сравнительного анализа депутатских взглядов.

В качестве примера разберём раскол кадетской фракции в 1912–1914 гг. Этот раскол зрел несколько лет и не привёл к окончательному разделению фракции и партии, однако в исследуемый период стало очевидно, что противоречия во фракции достигли апогея. В апреле 1914 г. октябристы констатировали «распад кадетской фракции на два крыла» [Партия «Союз 17 октября»..., 2000, с.463]. Традиционно выделяют левоопозиционное крыло, придерживавшееся радикальных взглядов на ситуацию в стране, и правое умеренное крыло, готовое к компромиссам с более консервативными силами. И правые, и левые кадеты имели ярких лидеров. Для прояснения идейных противоречий проанализируем взгляды двух из них.

Признанным идеологом левого кадетизма в период Думы IV созыва был томский депутат Н.В. Некрасов, на это указывали даже официальные отчёты фракции [Четвёртая Государственная..., 1914, с. 11]. Николай Виссарионович был инженером-железнодорожником, преподавателем Томского технологического института, поэтому основная его деятельность в Думе была связана с рассмотрением вопросов путей сообщения. Примерно в 1913 г. начинается стремительный рост его влияния во фракции, и он нередко выступает в качестве представителя кадетов при оглашении мотивов голосования [Государственная Дума: Указатель..., 1914, с.186].

В правом крыле фракции в качестве визави Н.В. Некрасова стоит назвать В.А. Маклакова. Известный московский юрист был одним из наиболее последовательных критиков заигрывания кадетов с левыми партиями. В 1912 г. даже появились слухи о том, будто сторонники В.А. Маклакова готовятся сместить П.Н. Милюкова, лидера фракции, дабы «восстановить её престиж»⁴. Впрочем, в 1914 г. правительственный наблюдатель резюмировал, что сторонники Маклакова во фракции малочисленны [Донесения Л.К. Куманина..., 1998, с.14].

Оба депутата являлись активными ораторами, что позволяет оценить их личные взгляды на репрезентативной выборке. В исследовании не учтены выступления по формальным поводам (протесты, краткие заявления, объяснения по личным вопросам, комментирование поправок и т.п.), а также доклады комиссий, поскольку их содержание согласовывалось с большинством в соответствующей комиссии и не всегда отражало личные взгляды докладчика. В итоге был сформирован корпус выступлений Н.В. Некрасова и В.А. Маклакова за 1912 – 1914 гг. Объём текстов оказался неравноценен: у Маклаков – около 37 тыс. слов, у Некрасова – около 11 тыс. Диспропорция связана с тем, что Некрасов чаще выступал докладчиком комиссий либо участвовал в прениях по узкоспециальным законопроектам, а подобные речи не способствуют достижению адекватности результа-

тов, потому они не вошли в корпус текстов.

С целью получения сопоставимых результатов необходимо задать ограниченный перечень категорий, позволяющий оценить отношение депутатов к ключевым политическим вопросам. Перечень был сформирован на основе Программы конституционно-демократической партии в редакции 1906 г. [Российские либералы..., 1996, с.51–58]. В документе чаще всего упоминаются категории «закон» (32 случая) и «государство» (31). Было выделено 32 категории, имеющие два и более упоминания в тексте программы.

В данном случае пришлось ограничить количество словоформ в каждой категории. Например, слово «прав» нельзя однозначно идентифицировать с категорией «права, правовой», так как это слово может иметь в речи множество значений в зависимости от контекста – как слово «права» в родительном падеже, так и краткое прилагательное, обозначающее правоту человека. Вторая важная ремарка: категории используются в том контексте, в каком употреблены в программе партии. Например, в категории «религия / вероисповедание» не учитываются названия конкретных конфессий, лишь сам термин.

Для начала проведём традиционный частотный анализ выраженности категорий. Поскольку массивы текстов несопоставимы, воспользуемся относительным показателем.

Таблица 3.

Частота употребления категорий депутатами (на 1000 слов)

Категория	Н.В. Некрасов	В.А. Маклаков
Закон	1,46	6,42
Государство	1,28	1,13
Суд	0,82	8,26
Права	1,64	2,65
Свобода	1,19	0,35
Гражданский	0,73	0,37
Парламент	0,82	0,25
Запрет	0,37	0,48
Земля	0,46	0,12
Самоуправление	4,02	0,37

Таблица приводится не полностью, только для 10 категорий, наиболее часто цитируемых в Программе партии кадетов. Но даже из этих данных видны существенные расхождения между депутатами. Во-первых, у В.А. Маклакова есть супер-категории, постоянно фигурирующие в выступлениях, – «суд» и «закон», что неудивительно, так как юрист Маклаков выступал в Думе в основном по вопросам, связанным с деятельностью Министерства юстиции и Сената. Ведущее место в риторике его однопартийца Н.В. Некрасова занимают категории «власть» (5,02), «самоуправление», «народ» (3,29) и «Россия» (3,29), причём три из перечисленных категорий не входят в первую десятку по частоте встречаемости в программе кадетов.

У депутатов обнаружились и общие черты. Не зафиксированы в их речи категории «союзы», «налоги». Категория «труд» в значении «производственная деятельность» употреблена обоими ораторами лишь однажды, как и категории «имущество», «восстановление». Слабо выражены этническая тема («народность – национальность») и категория «империя».

Дальнейший семантический анализ покажет, что из виду выпускается ряд важнейших моментов, и простой количественный срез не позволяет создать адекватную систему депутатских взглядов.

Партия народной свободы, по крайней мере, во взглядах Н.В. Некрасова эти понятия были наполнены важным смыслом. Другое контринтуитивное наблюдение связано с местом категории «власть». Несмотря на частое упоминание в речи, эта категория оказалась практически ни с чем не связанной и не играла решающей роли во взглядах депутата.

В целом Н.В. Некрасов апеллировал к общедемократическим ценностям, его риторика базировалась на популистских, предельно широких лозунгах, которые можно встретить и у социалистических партий. В данном случае можно вспомнить ремарку С.И. Шидловского, который при рассмотрении деятельности другого левого кадета заметил, что тот «принадлежал приблизительно к эсерам» [Шидловский, 1923, с.63]. Что касается самого Н.В. Некрасова, то недалёкой от истины, хоть и пристрастной, оказалась его оценка В.Н. Коковцовым: премьер-министр охарактеризовал выступления левого кадета как «самую безудержную демагогию» [Коковцов, 1992, с.276], ориентированную на возбуждение публики.

Взгляды же В.А. Маклакова носят более узкий характер. Наиболее цитируемые категории являются и центральными в сети, т.е. депутат постоянно озвучивал ключевые для себя идеи: закон, суд, права, власть. Маклаков декларирует в чистом виде идею правового государства. В этом контексте необходимо отметить самое важное противоречие во взглядах двух депутатов: противопоставление народа и государства. Ранее говорилось, что категория «народ, население» является наиболее весомым понятием риторики Н.В. Некрасова, у В.А. Маклакова также фиксируется значимость данной категории, но добавляется другая, более тяжеловесная. Категория «государство» при очень невысокой частоте употребления в речи (ниже, чем у Некрасова) имеет невероятно высокий индекс промежуточности – 37,0, т.е. это понятие употребляется редко, но всегда в качестве важного смыслового.

На основании семантического контент-анализа выступлений Н.В. Некрасова и В.А. Маклакова можно заключить, что противостояние левого и правого крыла кадетской фракции было связано с конфликтом двух идей программы партии. С одной стороны, революционная идея народного освобождения, с другой стороны, реформистская идея правового государства. Как отмечал сам В.А. Маклаков, левый фланг партии «в революцию верил и с революционной идеологией не разрывал» [Маклаков, 1936, с.501]. В свою очередь П.Н. Милюков обвинял коллегу в том, что тот «не всегда разделял мнения» кадетов [Милюков, 1990, с.18].

Определяя в заключение возможности предложенной методики, отметим, что она в наибольшей степени пригодна для решения частных задач, связанных с прояснением социополитических позиций отдельных исторических деятелей. Методика обладает большим эвристическим потенциалом, так как позволяет дополнить частотный анализ пониманием значимости той или иной категории в системе взглядов автора.

Применение этой методики при исследовании истории парламентаризма представляется весьма перспективным. Во-первых, для этого не требуется каких-либо априорных сведений о взглядах автора нарратива, что немаловажно при изучении речей малоизвестных депутатов. Во-вторых, данная методика позволяет связать речи, произнесённые в разное время, в единую систему без ущерба для адекватности результатов. В-третьих, исследователь получает возможность реконструировать латентный, неявный уровень политического сознания, что затруднительно сделать при помощи традиционных методов.

Таким образом, семантический сетевой анализ позволяет получить наглядный, объективизированный срез политического сознания на основе корпуса текстов. Немаловажно, что метод остаётся в пределах количественного подхода к анализу источника. Как следствие этого, результаты в меньшей степени зависят от интерпретации эксперта. Появляется возможность сравнивать полученные результаты по разным текстовым корпусам между собой, используя количественные показатели.

Примечания

¹ Голос Москвы. 1912. 1 марта.

² ГАРФ. Ф.115, оп.2, д.4, л.7(об).

³ Индекс сетевой промежуточности относится к индексам, позволяющим оценить «центральность» того или иного узла в сети. Данный индекс равен количеству кратчайших путей между всеми узлами сети, проходящих через данный узел.

⁴ Голос Москвы. 1912. 14 февр.

Библиографический список

- Handbook of Multimethod Measurement in Psychology. Washington, 2006.
- Roberts C.W. A Conceptual Framework for Quantitative Text Analysis // Quality & Quantity. 2000. №34.
- Гайдыйшев И.П. Решение научных и инженерных задач средствами Excel, VBA и C/C++. СПб., 2004.
- Государственная Дума: Указатель к стенографическим отчётам. Четвёртый созыв. Сессия 2 (1913–1914). Заседания 1–111. СПб., 1914.
- Государственная Дума Российской империи: 1906 – 1917: Энциклопедия. М., 2008.
- Градосельская Г.В. Сетевые измерения в социологии. М., 2004.
- Донесения Л.К. Куманина из Министерского павильона Государственной думы, декабрь 1911 – февраль 1917 года // Вопр. истории. 1999. №8.
- Коковцов В. Н. Из моего прошлого: Воспоминания (1903 – 1919 гг.). М., 1992. Кн. 1.
- Маклаков В.А. Власть и общественность на закате старой России. Париж, 1936. Т.3.
- Милюков П.Н. Воспоминания (1859 – 1917). М., 1990. Т.2.
- Партия «Союз 17 октября»: Протоколы съездов, конференций и заседаний ЦК. М., 2000. Т.2.
- Российские либералы: кадеты и октябристы / сост. Д.Б. Павлов, В.В. Шелохаев. М., 1996.
- Россия. Государственная Дума: Стенографические отчёты. Созыв 1–4. СПб., 1906–1917.
- Становление российского парламентаризма начала XX в. / Н.Б. Селунская, Л.И. Бородкин, Ю.Г. Григорьева, А.Н. Петров. М., 1996.
- Топка Р.В. Контент-анализ: семантический или документалистический? Опыт применения на материале крестьянских наказов от южноукраинских губерний в I Государственную Думу // Круг идей: историческая информатика в информационном обществе: Тр. VII конф. АИК. М., 2001.
- Усманова Д.М. Мусульманские представители в российском парламенте. 1906 – 1916. Казань, 2005.
- Четвёртая Государственная Дума: Отчёт. Фракция народной свободы. СПб., 1914.
- Шидловский С.И. Воспоминания. Берлин, 1923. Т.2.

Дата поступления рукописи в редакцию 18.07.2014

**SEMANTIC TEXT ANALYSIS OF PARLIAMENTARY SPEECHES
IN THE STATE DUMA OF THE RUSSIAN EMPIRE:
METHODOLOGICAL ASPECTS**

A. V. Smetanin

Perm State National Research University, Bukirev str., 15, 614990, Perm, Russia
smetanin.av@gmail.com

The essay focuses on contemporary method of text analysis of political narrative. Traditional approach to content analysis assumes creation of frequency dictionaries which means that each word is considered as an isolated element of the text out of connection with other words. It raises a problem of incorrect interpretation of the results which can be solved by using semantic text analysis and network analysis in particular. Network technologies can be useful in two different cases: the first one is the opportunity of detailed reconstruction of an author's views. In this case semantic categories (nodes of network) are retrieved from frequency dictionary, then, co-occurrence matrix is created on which network model is based. Another way of research is connected with preliminary determination of the list of categories. This can be productive in a comparative analysis of different texts or authors. In our opinion, betweenness centrality is the best network metric for detection of key categories because it displays nodes close to the majority of network segments. The chosen methodology was verified through two examples from Russian parliament history. In the first example, the author considered the case of a deputy's change of the faction and concluded on the basis of his speeches that ideologically he was ready for the change of the group. In the second example the author investigated the views of ideological opponents within the same faction and made the conclusion that they took different conflicting ideas from the party program. In general, semantic text analysis proved its practical applicability.

Key words: semantic text analysis, network analysis, history of ideas, the State Duma of the Russian Empire, the Cadet faction.

References

- Handbook of Multimethod Measurement in Psychology. Washington, 2006.
- Roberts C.W.* A Conceptual Framework for Quantitative Text Analysis // *Quality & Quantity*. 2000. №34.
- Gaydyshev I.P.* Reshenie nauchnykh i inzhenernykh zadach sredstvami Excel, VBA i C/S++. SPb., 2004.
- Gosudarstvennaya Duma: Ukazatel' k stenograficheskim otchetam. Chetvertyy sozyv. Sessiya 2 (1913 – 1914). Zasedaniya 1 - 111. SPb., 1914.
- Gosudarstvennaya Duma Rossiyskoy imperii: 1906 – 1917: Entsiklopediya. M., 2008.
- Gradosel'skaya G.V.* Setevye izmereniya v sotsiologii. M., 2004.
- Doneseniya L.K. Kumanina* iz Ministerskogo pavi'ona Gosudarstvennoy dumy, dekabr' 1911 – fevral' 1917 goda // *Vopr. istorii*. 1999. №8.
- Kokovtsov V.N.* Iz moego proshlogo: Vospominaniya (1903 – 1919 gg.). M., 1992. Kn. 1.
- Maklakov V.A.* Vlast' i obshchestvennost' na zakate staroy Rossii. Parizh, 1936. T.3.
- Milyukov P.N.* Vospominaniya (1859 – 1917). M., 1990. T.2.
- Partiya «Soyuz 17 oktyabrya»: Protokoly s'ezdov, konferentsiy i zasedaniy TsK. M., 2000. T.2.
- Rossiyskie liberaly: kadety i oktyabristy / sost. D.B. Pavlov, V.V. Shelokhaev. M., 1996.
- Rossiya. Gosudarstvennaya Duma. Stenograficheskie otchety. Sozyv 1–4. SPb., 1906–1917.
- Stanovlenie rossiyskogo parlamentarizma nachala XX v. / N.B. Selunskaya, L.I. Borodkin, Yu.G. Grigor'eva, A.N. Petrov. M., 1996.
- Topka R.V.* Kontent-analiz: semanticheskiy ili dokumentalisticheskiy? Opyt primeneniya na materiale krest'yan-skiikh nakazov ot yuzhnoukrainskiikh guberniy v I Gosudarstvennuyu Dumu // *Krug idey: istoricheskaya informatika v informatsionnom obshchestve: Tr. VII konf. AIK*. M., 2001.
- Usmanova D.M.* Musul'manskie predstaviteli v rossiyskom parlamente. 1906 – 1916. Kazan', 2005.
- Chetvertaya Gosudarstvennaya Duma. Otchet. Fraktsiya narodnoy svobody. SPb., 1914.
- Shidlovskiy S.I.* Vospominaniya. Berlin, 1923. T.2.