ПРОБЛЕМАТИЗАЦИЯ КОРРУПЦИИ В ПУБЛИЧНОМ ПОЛИТИЧЕСКОМ ДИСКУРСЕ: ОПЫТ КОЛИЧЕСТВЕННОГО ИССЛЕДОВАНИЯ

Я.Ю. Старцев
Институт философии и права
Уральского отделения РАН

Аннотация: *В статье проведён количественный контент-анализ массива высказываний семи лидеров общественного мнения, связанных с проблемой коррупции. Объектом исследования послужили интернет-блоги и тексты официальных интервью и выступлений (блоги М. Каца, А. Навального, Н. Старикова, И. Стрелкова, С. и А. Удальцовых, стенограммы видеолекций С. Кургиняна, стенограммы пресс-конференций и посланий Президента РФ). Построены частотные семантические модели каждого из изученных текстовых массивов, выявлена их структура и соотношение друг с другом. Результатом исследования стала сравнительная характеристика структуры антикоррупционного дискурса в отношении изучавшихся текстовых массивов.*

Ключевые слова: *коррупция, политический дискурс, общественное мнение.*

В современной России коррупция в органах государственной власти и местного самоуправления, в государственных и муниципальных учреждениях часто рассматривается как одна из ключевых политических тем [Актуальные проблемы… 2016]. Постоянная актуализация коррупционной проблематики характерна как для официального политического дискурса, так и для представителей оппозиционных партий и движений; рассуждения о финансовых злоупотреблениях, взяточничестве и казнокрадстве регулярно формируют повестку большинства средств массовой информации, хоть сколько-нибудь ориентирующихся на политическую жизнь. Наконец, уже достаточно длительное время без антикоррупционных лозунгов и программ не обходится ни одна избирательная кампания.

Такое многообразие политических высказываний относящихся к коррупции вызывает у исследователя естественное стремление структурировать информационный поток и выяснить, каким образом организован политический дискурс, связанный с этой тематикой. В частности, несомненный интерес могут представлять различия в трактовке коррупции разными политическими силами или лидерами общественного мнения — особенно через их публичные высказывания, которые в значительной степени формируют как установки их сторонников, так и структуру политической повестки в целом. Гигантский объём информации с неизбежностью требует использования количественных методов обработки больших масси-

вов данных и, даже при наличии и доступности таких данных, формирования значимой для исследуемого феномена выборки.

Между тем, даже пробный и поверхностный обзор высказываний такого рода позволяет констатировать, что суждений о коррупции в целом, безотносительно событийного контекста, крайне мало и они достаточно банальны. Содержательные характеристики, которые как раз и могут что-то сказать о месте коррупционной проблематики в структурировании политического поля, связаны, как правило, с частными событиями, персонажами, информационными поводами. С одной стороны, учёт этого контекста обязателен, если мы хотим выявить позиции разных политических сил именно в связи с актуальной политикой, а не их философские умозаключения, не привязанные к современности. С другой стороны, разнообразие конкретных поводов, провоцирующих антикоррупционные высказывания, достаточно велико для того, чтобы сделать сравнительное исследование весьма затруднительным. Как следствие, возникает методологическая проблема формирования исследовательских средств, которые позволили бы структурировать (анти)коррупционный дискурс разных политических сил безотносительно конкретных ситуаций, но с сохранением связи с событийным контекстом[1].

Постановка проблемы и методика исследования

Исходная гипотеза исследования заключалась в том, что установки и подходы к пониманию и трактовке коррупции разными политическими акторами могут быть выявлены методами количественного контент-анализа. Выявление различий в таких трактовках даёт основания для построения специфичной для того или иного субъекта модели политического дискурса, связанного с проблемой коррупции. Специфическая задача — поиск таких лексем и их сочетаний, которые отражали бы подход автора (авторов) к проблеме безотносительно конкретных событий, лиц и ситуаций, спровоцировавших высказывания, относящиеся к коррупционной проблематике. Результатом должна стать модель, характеризующая не событийный или персональный аспект (анти)коррупционного дискурса, т. е. не тематизацию данного явления в рамках текущей повестки, а его проблематизацию — повторяющиеся суждения, оценки и дискурсивные паттерны вне зависимости от конкретного повода, персоналий или типа коррупционного поведения.

Выборка текстов для количественного анализа формировалась исходя из следующих соображений:

- представленность разных позиций в рамках современного российского идеологического поля, в частности — оппозиций «патриоты\западники», «левые\правые», «либералы\государственники»;
- известность и цитируемость автора в информационном пространстве;
- мобилизационный потенциал (способность стимулировать свою аудиторию к тому или иному политическому действию за пределами собственно информационной среды).

В качестве формата бытования текстов были выбраны блоги, — в силу их однозначной связанности с тем или иным лидером общественного мнения и представляемым им политическим течением, ориентации на актуальную политическую жизнь и сравнительной доступности, в т. ч. — в рамках достаточно длительного временного отрезка.

С учётом этих соображений выборка включает в себя тексты, размещённые в следующих блогах:

- блог А. Навального за 2011–2017 гг. [Навальный 2017];

[1]Мы уже проводили подобное исследование применительно к значительно меньшим текстовым массивам, позволявшим ручной подсчёт, без систематического использования методов компьютерной лингвистики. См. [Старцев 2006].

- блог М. Каца за 2010–2017 гг. [Кац 2017];
- блог Н. Старикова за 2009–2017 гг. [Стариков 2017];
- блоги С. Удальцова за 2012–2014 гг. [Удальцов 2014] и А. Удальцовой за 2006–2016 гг. [Удальцова 2016] как единый текстовый массив;
- блог И. Стрелкова за 2016–2017 гг. [Стрелков 2017];
- стенограммы видеолекций С. Кургиняна «Суть времени» [Кургинян 2011].

Кроме того, ввиду необходимости представить и официальный дискурс, в выборку были включены стенограммы ежегодных пресс-конференций и посланий Федеральному Собранию Президента РФ за 2000–2016 гг. Публичные обращения и интервью главы государства не могут рассматриваться как прямой аналог интернет-блога, однако это наиболее доступная и ориентированная на наиболее широкую аудиторию прямая речь Президента.

Тексты всех блогов, включая размещённые там же транскрипты или резюме видеороликов, if any, взяты за весь период их существования (различный для разных блогов), без тематической или хронологической фильтрации; последние учтённые записи относятся к апрелю 2017 г. или к моменту завершения публикаций, если он наступил раньше. Кроме того, в выборку были включены комментарии читателей (зрителей) блогов: на наш взгляд, это позволяет избежать анализа индивидуальной стилистики конкретного автора и в большей степени отражает как раз представляемый им способ проблематизации коррупции, в т. ч. — через оценки сторонников и противников. Разумеется, такой подход опирается на презумпцию модерируемости комментариев, освобождающей их от информационного шума («спама») и, возможно, от высказываний, не вписывающихся в доминирующий на данной информационной площадке дискурс. Кроме того, наблюдение за реакциями читателей (зрителей) этих блогов показывает, что комментарии сторонников безусловно доминируют количественно во всех случаях.

Собственно использованный метод анализа заключается в сравнении частотности значимых слов, соседствующих с терминами, определяющими коррупционное поведение, в выбранных для исследования текстах.

Вместе с тем, компьютерный анализ текстов на естественных языках, особенно на русском, сталкивается с целым рядом технических проблем, способ решения которых в значительной степени влияет на результаты [Большакова и др. 2011]. В нашем случае построение семантической модели анализируемых текстовых массивов осуществлялось путём последовательного применения к каждому элементу выборки нескольких фильтров (после удаления служебных пометок интернет-интерфейса):

- устранение служебных частей речи (классических «стоп-слов»): фильтр формировался на основе грамматических словарей русского языка, с необходимыми дополнениями;
- устранение имён собственных, отражающих в большей степени ситуативный контекст, а не структуру взглядов на коррупцию, а следовательно иррелевантных для нашего исследования.

Сформированная таким образом семантическая модель каждого элемента выборки, представляющая собой упорядоченную в соответствии с исходным текстом последовательность значимых лексем, была лемматизирована и стала объектом собственно частотного анализа. В рамках настоящего исследования учитываются только 1-граммы. Составленный на основе существующих словарей список синонимов слова «коррупция» был использован в качестве шаблона для поиска ближайшего контекста каждого из входящих в список терминов: для каждого из терминов и для всех случаев их употребления были выявлены конкордансы (текстовые фрагменты, содержащие данное слово) с шагом в 3–4 лексемы в обе стороны, без учёта границ предложений. Итоговый список конкордансов был подвергнут обычному частотному анализу: 100 наиболее частотных лексем рассматриваются как основа для частотно-

семантической модели (анти)коррупционного дискурса применительно к каждому элементу выборки (отдельному авторскому массиву).

Синонимический ряд, для элементов которого производилось построение конкордансов, включал в себя следующие термины (все они лемматизированы, как и текстовые массивы, в которых производился поиск; дополнительные морфологические анализаторы не использовались):

взятка
взяткоемкий
взяткоемкость
взяточник
взяточничество
вор
воровство
вымогательство
казнокрад
казнокрадство
коррумпированность
коррумпированный
коррупциогенность
коррупциогенный
коррупционер
коррупционный
коррупция
краденый
красть
мздоимство
откат
поборы
подкуп
подкупность
попил
продажность
продажный
разложение
распил
украденный
украсть.

Лексемы, связанные с кражей (краденый, украденный, красть, украсть) широко используются в русском языке и вне коррупционного контекста, однако в анализируемых текстовых массивах таких случаев было лишь два: в остальном эта терминология применялась именно к коррупционной проблематике. То же относится к лексемам «вор» и «воровство».

Для обработки текстов использовались следующие инструменты: NLTK 3: Natural Language Toolkit [Bird, Klein, Loper 2014; Hardeniya 2015], MyStem [Яндекс 2011], частотные словари Национального корпуса русского языка [НКРЯ 2017].

Анализ эмпирических данных

Поиск конкордансов позволил сформировать перечень строк (фрагментов текстов-семантических моделей), содержащих все случаи вхождения этих лексем. Количество строк-конкордансов для разных текстовых массивов существенно различалось (от 2906 для блога М. Каца до 79 для стенограмм видеолекций С. Кургиняна). Ниже в качестве иллюстрации приведены некоторые конкордансы для текстового массива С. и А. Удальцовых:

«пенсионный возраст обижать труженик пахать глупо власть взятка аренда раздавать иль сотня забирать

многочисленный коррупционный преступление получать взятка особо крупный размер совершать состав организованный

дедовщинный расформировывать новый народный милиция взятка злоупотребление служебный положение чекистский выборы

свободный счастливый жулик вор сказать человек брать взятка оскорбление апчехов человек прекрасно лицо

давать чиновник полный свод действие свобода брать взятка творить беспредел регион страна чиновник

кафедра выводить сотрудник милиция взять крупный взятка день экзамен платить разгильдяй кризис терять

ходячий овно кстати взять поличное получение взятка возглавлять парламентский комитет заводить

уровень каменный образование место мир воровство взятка чекистский фашистень неоспоримый лидер мир чекистский

городок коррупционер разводиться понимашь взятка шоколадка сволочь взять семерка новый купить подлюга

письмо извинение патриарх покаяние работа адвокат вид взятка судья весенний план оппозиция молодец

возражать популистка пытаться контролировать ловить взятка сажать самокат детский сад нравиться фонтанировать идея

управление человек развивать фирма ноль государство взятка отсюда рассуждать научаться видеть цена болтовня

работать госструктура активно использовать провокация взятка цель выявление жулик вор госструктура максимально

личный задание координатор отлично известно объявлять взяточник еврей денежный мешок конгресс заседание конфликтный

рнургалий спина шутка сюрприз прошлое оставаться взяточничество злоупотребление служебный полномочие коррупция

житель обвинять директор дженерал центр произвол взяточничество становиться скандировать стройка таганок взятка блок

сторона строить новый союз параллельно разбираться жулик вор правильно юго восток справляться начинаться

установление личность главный простой подозрение похожий вор убийца прикопаться то особо пониматься грубый

черный полотнище крупный буква надпись ленэкспо сходка вор проспект растяжка провисеть час прибывать

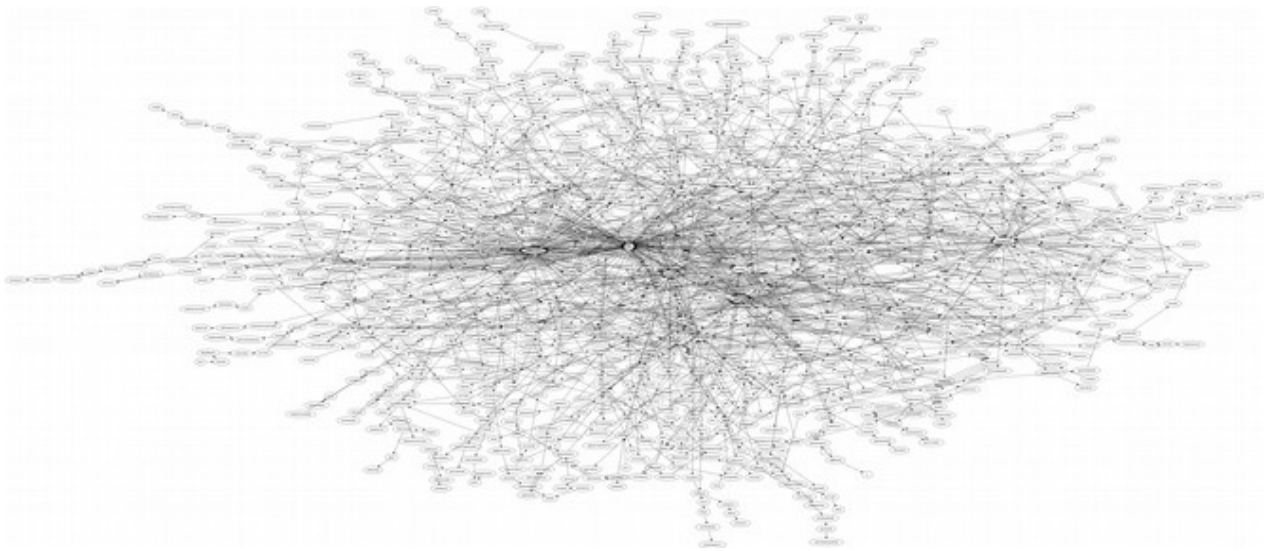
становиться давать повод возмущение тема лидер правый вор расхищать богатство родина дедушка бабушка

отсталый негодяй правитель редкостный негодяй лицемер вор правильный требование оправдывать проведение несогласован »

Анализ сравнительной частотности всех лексем, формирующих конкордансы для каждого текстового массива позволил говорить о наличии семантического ядра, т. е. значительного количества высокочастотных слов, не входивших в первоначальный синонимический ряд, служивший основой для поиска. Исследовательская гипотеза на этом этапе заключается в том, что наиболее частотные в конкордансах лексемы — т. е., значимые слова, которые с высокой вероятностью оказываются рядом с терминами из «коррупционного» списка, — во многом формируют специфичный для данного дискурса (представляющего его текстового массива) семантический контекст, отражающий установки (коллективного) автора по отношению к коррупции.

Наличие подобного семантического ядра в каждом случае может быть продемонстрировано через построенные на основе конкордансов графы. Вершины графа представляют отдельные лексемы, формирующие конкордансы, рёбра — связи между непосредственно соседствующими словами; несколько рёбер, соединяющих одну и ту же пару вершин отражают многократность совместного появления соответствующих терминов (в качестве средства визуализации использовался программный комплекс Graphviz [Graphviz 2017]).

Рисунок 1.
Конкордансы для «коррупционного» списка в текстовом массиве И. Стрелкова



При большем количестве конкордансов (прежде всего — в связи с объёмом исходного текстового массива) визуализация отдельных элементов становится затруднительной, но наличие семантического ядра оказывается более наглядным, — как в случае с текстовым массивом А. Навального.

Рисунок 2.
Конкордансы для «коррупционного» списка в текстовом массиве А. Навального

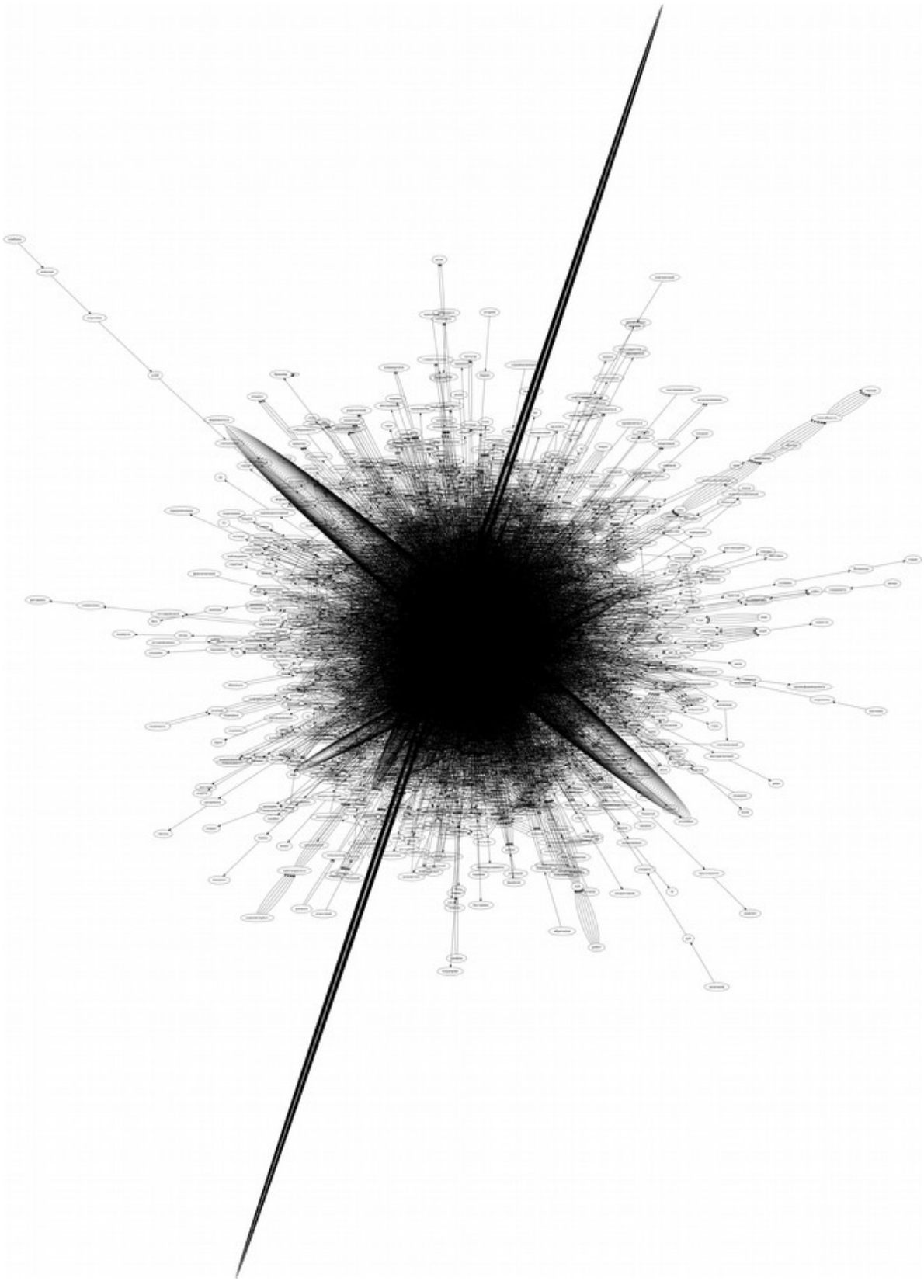


Рисунок 3.
100 наиболее частотных лексем в конкордансах для текстового массива М. Каца

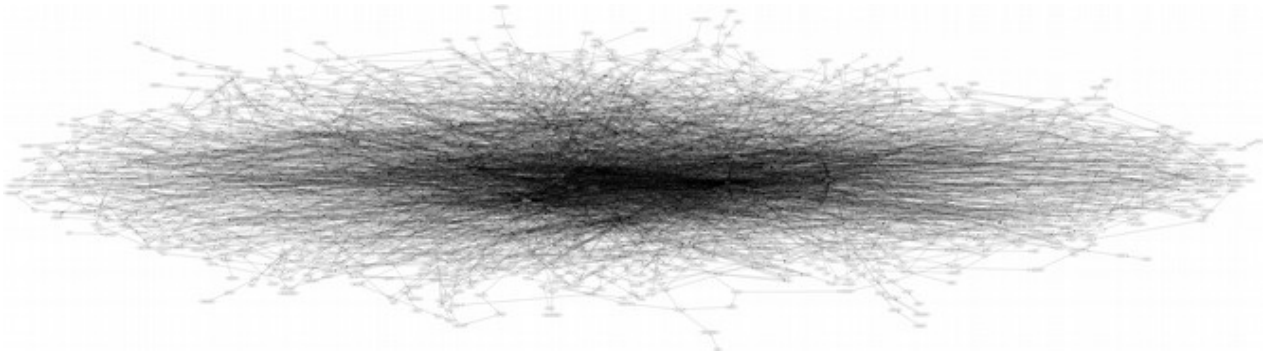
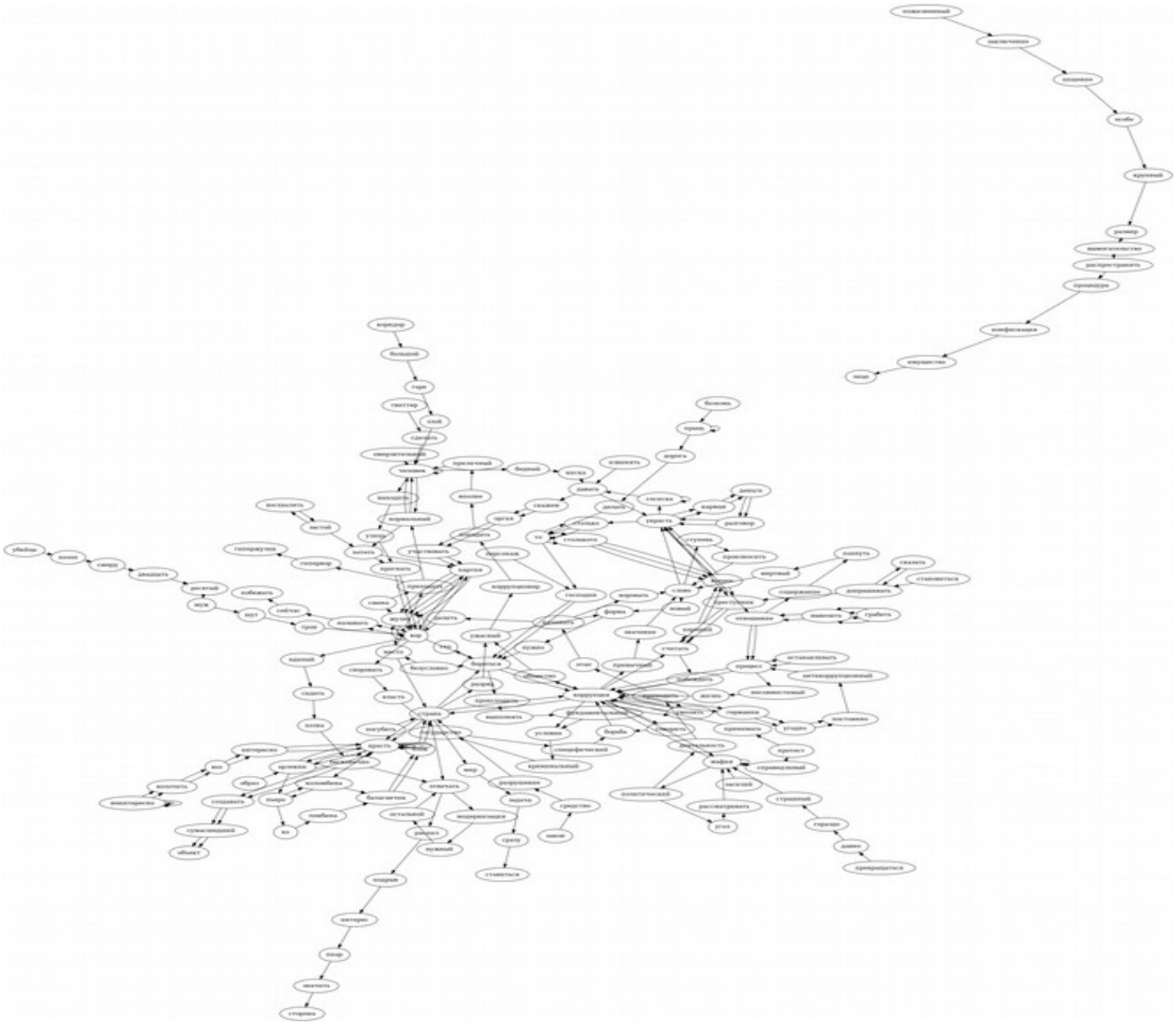


Рисунок 4.
100 наиболее частотных лексем в конкордансах для текстового массива С. Кургиняна



Анализ всех лексических сочетаний в рамках конкордансов представляется технически возможным, но эвристически бесполезным: нас интересуют прежде всего наиболее устойчивые сочетания (случаи «соседства») терминов, характеризующих коррупционное поведение и регулярно сопутствующих им других значимых слов. В составленных для каждого набора конкордансов частотных списках была отсечена верхняя сотня (сто наиболее частотных для каждого набора 1-грамм) — семантическое ядро, которое и стало предметом дальнейшего анализа. В свою очередь, это позволило создать более наглядные визуализации, позволяющие проследить взаимосвязь формирующих семантическое ядро лексем.

Рисунок 5.
100 наиболее частотных лексем в конкордансах для текстового массива А. Навального

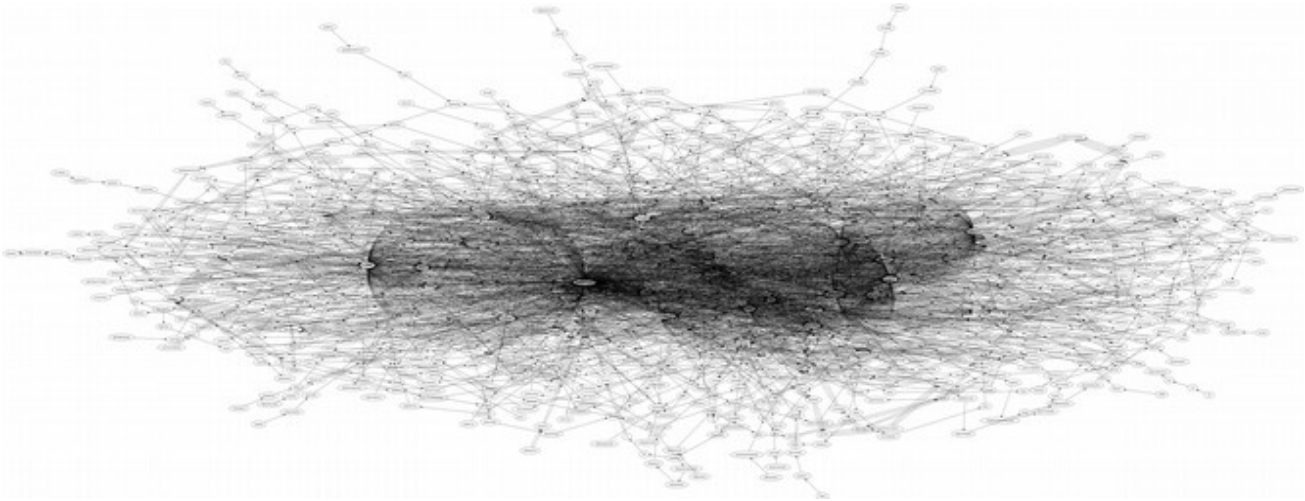


Рисунок 6.
100 наиболее частотных лексем в конкордансах для текстового массива Н. Старикова

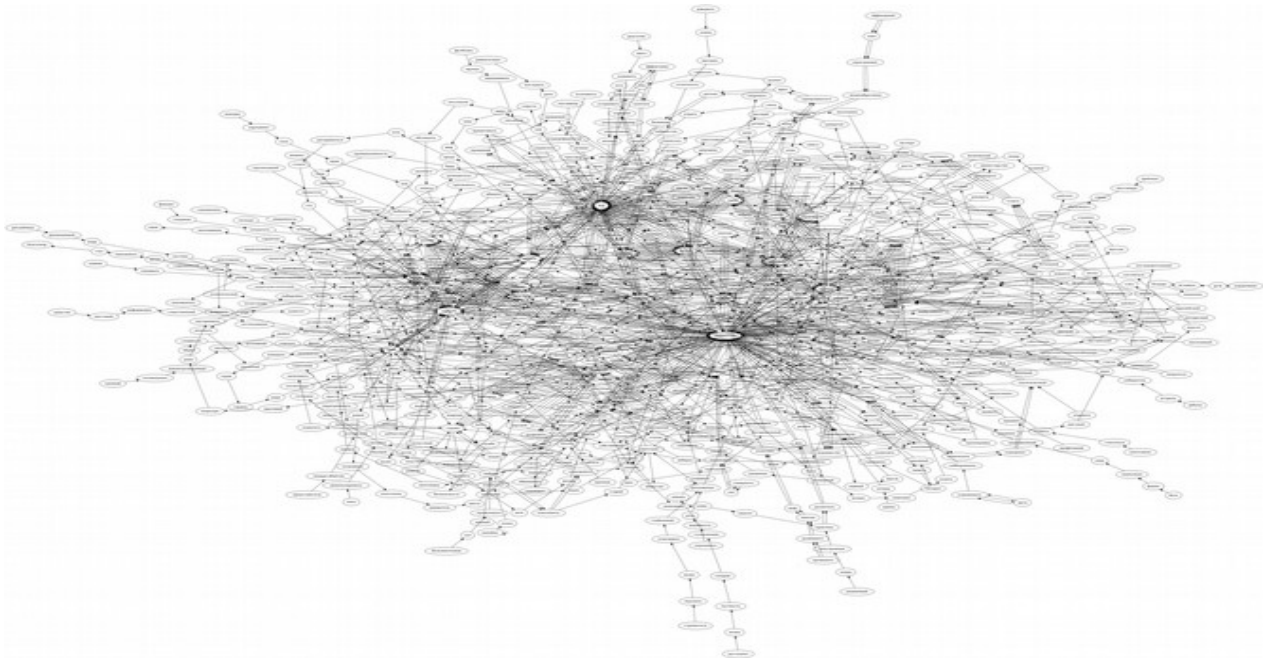


Рисунок 7.

100 наиболее частотных лексем в конкордансах для текстового массива И. Стрелкова

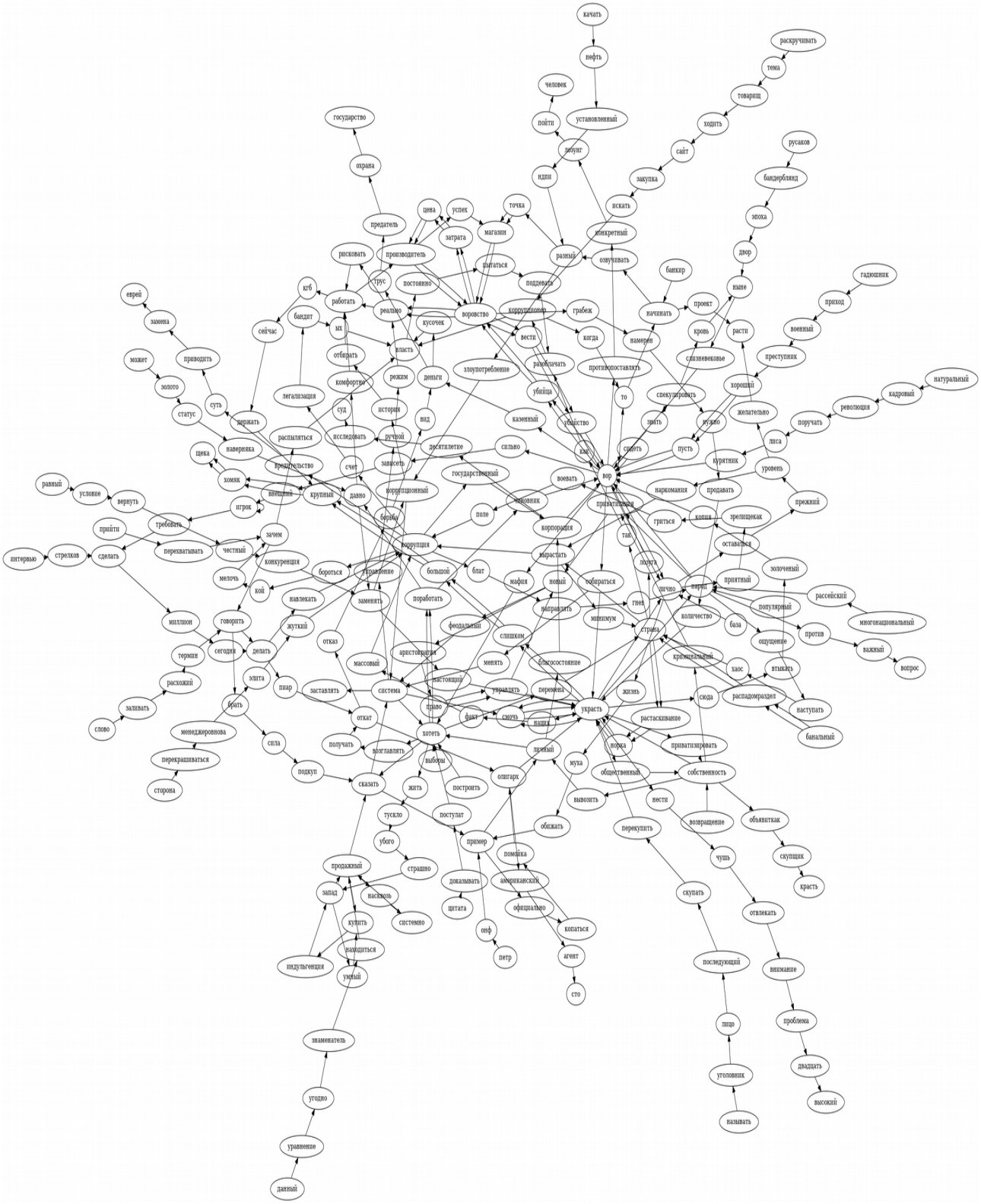
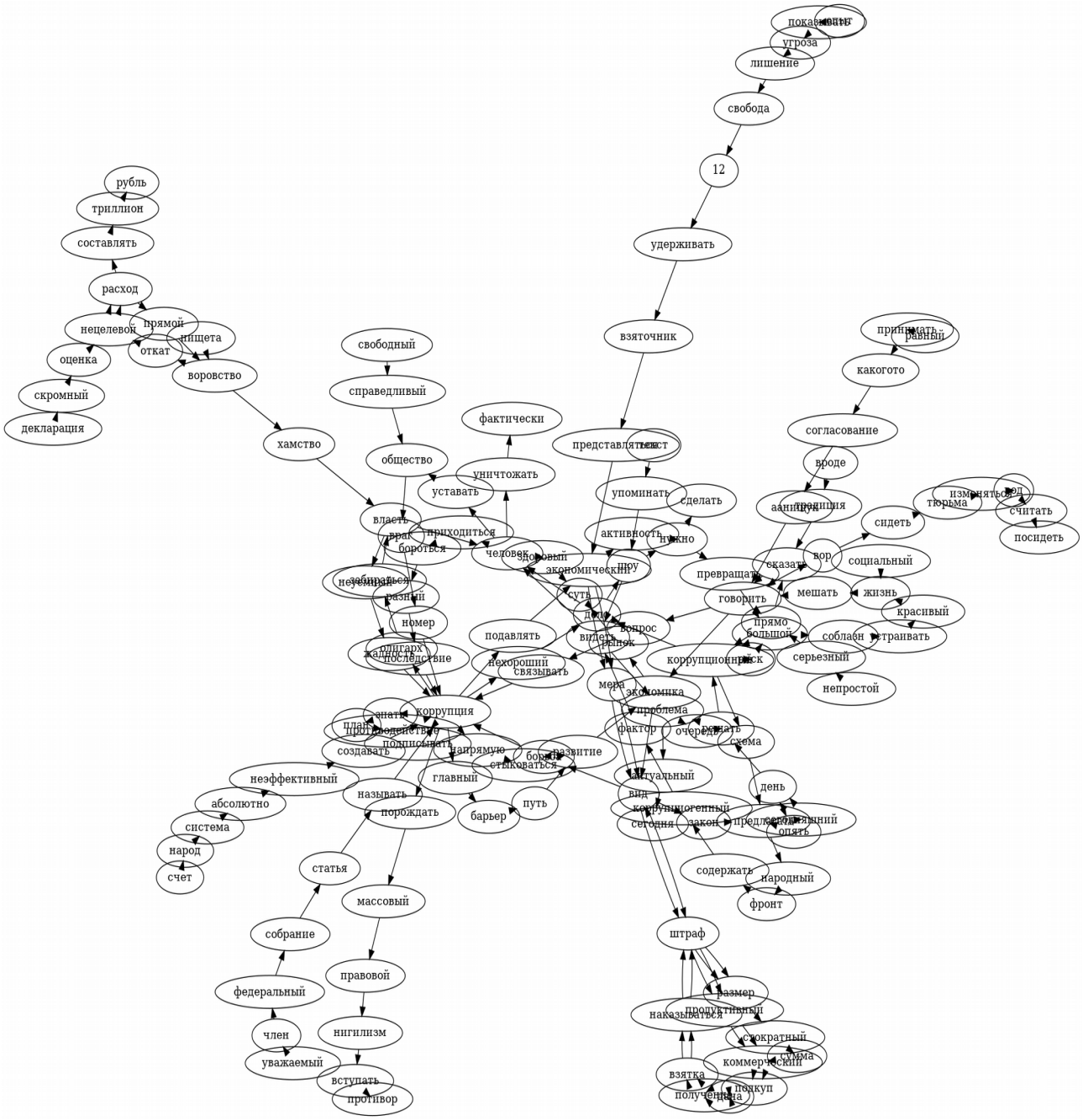


Рисунок 8.
100 наиболее частотных лексем в конкордансах для текстового массива Президента РФ



Сформированные таким образом частотные списки были прорейтингованы — каждой лексеме-элементу семантического ядра присвоен рейтинг, соответствующий её частотности по шкале от 1 до 100: наиболее частотное для данного массива слово получило рейтинг 100, следующее за ним — 99 и т. д., до завершающего список наименее частотного в рамках сотни термина с рейтингом равным 1. Это позволило как осуществить представленные ниже кумулятивные расчёты, так и сопоставить сравнительный вес одних и тех же лексем в разных тек-

стовых массивах — т. е., сравнить вес того или иного значения в соответствующих антикоррупционных дискурсах. В качестве примера такого рейтингования ниже представлена сотня наиболее частотных лексем для конкордансов из текстового массива Президента РФ:

100 коррупция
99 борьба
98 коррупционный
97 дело
96 говорить
95 вопрос
94 воровство
93 штраф
92 проблема
91 власть
90 человек
89 противодействие
88 государственный
87 взятка
86 мера
85 орган
84 общество
83 вид
82 становиться
81 сегодня
80 развитие
79 деньги
78 нужно
77 знать
76 сфера
75 ответственность
74 хотеть
73 закон
72 система
71 общественный
70 решать
69 украсть
68 суд
67 считать
66 предлагать
65 сказать
64 результат
63 большой
62 народ
61 схема
60 размер
59 расход
58 коммерческий
57 создавать
56 откат

55	оставаться
54	правовой
53	прямой
52	экономика
51	суть
50	уровень
49	просить
48	социальный
47	коллега
46	называть
45	жизнь
44	проверка
43	лицо
42	подкуп
41	экономический
40	сумма
39	нецелевой
38	фактически
37	препятствие
36	неэффективный
35	работа
34	организация
33	управление
32	давать
31	наказываться
30	барьер
29	среда
28	правоохранительный
27	уважаемый
26	вор
25	подписывать
24	число
23	отношение
22	реализация
21	шоу
20	получение
19	рубль
18	необходимый
17	сейчас
16	ресурс
15	вести
14	составлять
13	преступный
12	сделать
11	бороться
10	питательный
9	последний
8	сидеть

7 серьезный
6 чиновник
5 высокий
4 федеральный
3 предприниматель
2 направленность
1 государство

Построение и сравнительный анализ семантических моделей антикоррупционного дискурса осуществлялся в три этапа:

1. Структурирование наиболее частотных лексем (семантического ядра) на основе универсального для всех моделей шаблона.

2. Определение кумулятивного веса отдельных проблемно-тематических блоков для каждого текстового массива и сравнение структуры и веса однотипных блоков у разных авторов.

3. Сравнение веса совпадающих лексем в разных моделях. На этом же этапе анализировалась роль локальных гапаксов — терминов, встречающихся лишь в одной из моделей.

Для структурирования частотных лексем исходя из их значения, в контексте коррупционной деятельности, были сформированы пять проблемно-тематических блоков:

- действия, предпринимаемые или планируемые автором дискурса;
- субъекты, о деятельности которых говорит автор дискурса;
- именования коррупционного поведения, различные способы его идентификации;
- оценочная характеристика предмета дискурса — фактически, перечисление свойств;
- пространственно-хронологическая и предметная среда, релевантная для данного дискурса;
- термины, относящиеся к самой ситуации высказывания (часто — специфичные для данного массива или универсальные сорные слова, либо слова с недостаточно определённым значением).

С некоторой долей условности мы можем сказать, что первый блок («Действия») выполняет программную функцию, отвечая на вопросы «что делать», «что должно быть сделано» или «что мы делаем», второй («Субъекты») связан с идентификацией ключевых акторов и, возможно, назначением ответственных, отвечая на вопрос «кто», третий («Поведение») выполняет идентификационную и номинативную функции, четвёртый («Характеристики») связан с экспликативной функцией, пятый («Среда») отвечает за локализацию (анти)коррупционной активности в том или ином хронотопе, в социальной и предметной среде, отвечая на вопросы «где», «когда», «по поводу чего», а шестой («Контекст») обеспечивает выполнение коммуникативной функции, формируя специфическую связь автора с аудиторией.

На наш взгляд, составленные таким образом семантические кластеры позволяют не только исчерпать набор значений анализируемых лексем, но и представить вполне убедительную модель антикоррупционного дискурса, позволяющую как углублять исследование в каждом случае, так и проводить сравнительный анализ. Во всех спорных случаях омонимия снималась вручную для каждого корпуса с учётом контекста (так, слово «акция» встречается в частотных списках С. и А. Удальцовых, А. Навального и Н. Старикова, однако в первом случае всегда означает «протестная акция», в двух последних — «акция предприятия (компании)»; слово «средство» у С. Кургиняна используется применительно к средствам производства в марксистском смысле, у Н. Старикова — для обозначения финансовых средств, в остальных случаях используется как служебное слово, — и пр.). Итоги данного этапа представлены ниже в табличной форме.

Таблица 1.

Структурирование семантического ядра корпуса М. Каца

Действия	**Субъекты**	**Поведение**	**Характеристики**	**Среда**	**Контекст**
сидеть (вор должен) ответ голосовать предлагать проект заниматься называть работа написать видеть писать думать против работать считать сделать вопрос сказать хотеть бороться знать говорить борьба делать	бизнес оппозиция государство партия продажный народ коррупционер чиновник власть человек (люди) жулик вор	поборы брать (взятку) красть воровать коррупционный распил откат взятка воровство украсть коррупция	правда ситуация повод цена пример большой честный факт уровень проблема	фонд жизнь цель политика право дорога бюджет жить велосипед суд город выборы закон сейчас страна деньги	приходить пусть реальный день плохой оставаться согласный вещь новый сколько сторона плохо слово нужный тема потом то начинать место получать пост становиться система нужно хороший давать дело
1308	783	835	344	731	1049

Замыкающие таблицу численные значения — сумма рейтинговых показателей всех лексем, вошедших в ту или иную категорию. На данном этапе этот показатель позволяет увидеть, что в блоге М. Каца при рассуждениях о коррупционной проблематике значительное место занимают указания на разные виды и конкретные случаи коррупции (суммарный вес 835), гораздо меньший вес имеют оценочные и иные характеристики этой деятельности (причины, условия, факторы, моральная оценка и пр. — суммарный вес 344) и преобладают — больше, чем в большинстве других текстовых массивов — коллоквиальные обороты и указания на контекст высказывания (суммарный вес 1049). Что касается содержания каждого блока, то ещё до систематического сравнения можно отметить, что указание на «бизнес» как

субъекта, активно присутствующего в антикоррупционном дискурсе почти уникально для этого массива: другие авторы большей частью говорят о коррупции, не упоминая предпринимательство. В графе «действия» значительное место занимают действия скорее информационного характера — «знать», «хотеть», «видеть» и пр. Это в значительной степени может быть связано с особенностями блога как жанра, — однако, специфика есть и тут: Кац — чуть ли не единственный из анализируемых авторов с такой настойчивостью предлагающий думать.

Таблица 2.

Структурирование семантического ядра корпуса С. Кургиняна

Действия	**Субъекты**	**Поведение**	**Характеристики**	**Среда**	**Контекст**
допрашивать	приличный (человек, люди)	доля	превращаться (в мафию, в балаган)	политика	побежать
побеждать	власть	воровать	опасный	нэп	ставиться
прекращать	нормальный (человек, люди)	разрушать	нехороший	империализм	киска
отвечать	общество	сумасшедший (объект роскоши)	наивный (отношение к К.)	средство (производства)	форма
прогнать	лицо	оргия	антикоррупционный	сейчас	воз
интересно	господин	грабить	степень	германия	арлекин
сказать	продажный	вывозить	омерзительный	мир	отношение
борьба	государство	объект (роскоши)	бросовый	кибернетика	пьеро
протест	преступник	прыщ	цена	политический	место
неинтересно	мафия	простота (хуже воровства)	условие	процесс	сторона
хотеть	партия	карман	постоянно	страна	приходиться
говорить	жулик	криминальный	чистый		нужно
вопрос	человек (люди)	давать (взятку)	уровень		волочить
считать	вор	воровство	справедливый		образ
называть		красть	происходить		девка
бороться		украсть	плохой		угодно
		коррупция	бесконечно		нужный
			устанавливать (НЭП)		сосиска
			деньги		слово
					простой
					дело
					создавать
					то
920	867	932	717	553	1061

Таблица 3.

Структурирование семантического ядра корпуса А. Навального

Действия	**Субъекты**	**Поведение**	**Характеристики**	**Среда**	**Контекст**
внимание	судья	строительство	нарушать	жизнь	часть
обвинение	компания	откат	крупный	город	главный
жалоба	глава	брать (взятку)	денежный	фильм	далеко
работать	взяточник	махинация	миллион	поселок	место
проект	продажный	акция	миллиард	бюджет	хороший
рассказывать	партия	воровство	рубль	дом	календарь
писать	чиновник	схема	деньги	сейчас	приходиться
ответ	борец	потратить		сегодня	нос
сказать	власть	нападение		дерево	нужный
кампания	человек (люди)	давать		цель	становиться
стенка (поставить к)	коррупционер	коррупционный		суд	тема
говорить	жулик	взятка		гектар	единый
повесить	вор	украсть		тайна	привет
сделать		коррупция		выборы	пост
отвечать				страна	получать
задерживать				соцопрос	дело
хотеть				фонд	день
участвовать					новый
работа					
выходить					
знать					
митинг					
считать					
поддерживать					
вопрос					
видеть					
делать					
находить					
против					
расследование					
борьба					
1344	825	755	416	769	851

Таблица 4.

Структурирование семантического ядра корпуса Президента РФ

Действия	**Субъекты**	**Поведение**	**Характеристика**	**Среда**	**Контекст**
сидеть (вор должен)	предприниматель	направленность (коррупционная)	серьезный	государство	последний
бороться	высокий (эшелон)	получение	преступный	федеральный	питательный (среда)
сделать	чиновник	давать (взятку)	необходимый	сейчас	составлять
вести (борьбу)	лицо	нецелевой	шоу	рубль	ресурс
реализация	народ	подкуп	вор	среда	отношение
подписывать	общественный	откат	барьер	сумма	число
правоохранительный	общество	схема	препятствие	экономический	уважаемый
наказываться	орган	украсть	уровень (высокий)	жизнь	фактически
управление	государственный	решать	правовой	социальный	коллега
организация	человек (люди)	взятка	расход	экономика	суть
работа	власть	воровство	размер	коммерческий	прямой
неэффективный		коррупционный	большой	суд	оставаться
проверка		коррупция	проблема	система	нужно
называть				закон	становиться
просить				сфера	вид
создавать				деньги	вопрос
результат				развитие	дело
сказать				сегодня	
предлагать					
считать					
хотеть					
ответственность					
знать					
мера					
противодействие					
штраф					
говорить					
борьба					
1437	628	770	530	882	802

Таблица 5.

Структурирование семантического ядра корпуса Н. Старикова

Действия	**Субъекты**	**Поведение**	**Характеристики**	**Среда**	**Контекст**
называть	общество	воровать	честный	история	нужный
выходить	правый	мздоимство	большой	смысл	приходить
сделать	преступник	акция (украсть)	размер	идея	любить
хотеть	глава	подкуп	факт	средство (финансовые)	оставаться
решение	правительство	казнокрадство	мнение	рубль	сколько
комментарий	лицо	откат	опасность	фильм	место
делать	чиновник	взяточничество	источник	период	путь
пытаться	партия	получать (взятку, доход)	проблема	миллион	тема
сказать	государство	коррупционный		лента	гость
тюрьма (вор должен сидеть в)	жулик	разложение		мир	точный
побеждать	продажный	воровство		экономика	ребенок
против	народ	взятка		государственный	становиться
вопрос	коррупционер	украсть		статья	подлинный
знать	человек (люди)	коррупция		доллар	новый
бороться	власть			либерализм	слово
думать	вор			запад	нужно
сидеть (вор должен)				царский	главный
говорить				суд	давать
борьба				сейчас	дело
				советский	
				индивидуализм	
				закон	
				деньги	
				страна	
989	895	957	331	986	892

Таблица 6.

Структурирование семантического ядра корпуса И. Стрелкова

Действия	**Субъекты**	**Поведение**	**Характеристики**	**Среда**	**Контекст**
возвращение (собственности)	мафия	заставлять (система)	пример	военный	подымать
разоблачать	общество	распил	уровень	мир	остров
сидеть (вор должен)	бывший (должностное лицо, номенклатурщик)	хомяк (прячет украденное)	база (партноменклатура как)	экономика	суть
сказать	предатель	щека (хомяк прячет за)	большой	суд	вести
отбирать	вояка	растаскивание	отвратительный	перспектива	хватить
сделать	противник	норка (прятать украденное)	честный	процесс	нужно
смочь	еврей	красть	расти	сейчас	реально
собираться	борец	брать (взятку)	крупный	ныне	точка
знать	солдафон	индульгенция	высокий	магазин	новый
работать	олигарх	купить (человека, индульгенцию)	слишком	выборы	нужный
вопрос	сми	откат	система (К как)	давно	то
говорить	дурак	коррупционный		деньги	начинать
борьба	коррупционер	взятка		собственность	сторона
против	жулик	разложение		страна	создавать
хотеть	взяточник	давать (взятку)			лично
	получать	воровство			пусть
	человек (люди)	украсть			настоящий
	чиновник	коррупция			часть
	государство				тема
	народ				
	продажный				
	власть				
	вор				
807	1245	1050	551	651	746

Таблица 7.

Структурирование семантического ядра корпуса С. и А. Удальцовых

Действия	**Субъекты**	**Поведение**	**Характеристика**	**Среда**	**Контекст**
держать	гражданский	произвол	честно	миллион	последний
надеяться	левый	очередной	основной	политика	пост
думать	лидер	править («понятия», воры)	факт	двор	часть
движение	судья	политический	итог	жизнь	нужно
бороться	властный (жулик)	коррупционный	уровень	режим	место
контроль	массовый	взятка	честный	свобода	осенний
протест	коррумпированный	давать (взятку)		начинаться	прийти
поддерживать	оппозиция	воровство		результат (выборов)	день
сделать	гражданин	украсть		город	сторона
работать	участник	коррупция		образование	нужный
требование	коррупционер			сейчас	главный
отвечать	чиновник			жить	сколько
начинать	партия			право	оставаться
тюрьма	продажный			слово	новый
выходить	народ			суд	хороший
митинг	человек (люди)			закон	система
наступление	власть			страна	дело
лозунг	жулик			выборы	становиться
сидеть	вор				
делать					
сказать					
знать					
голос					
вопрос					
говорить					
борьба					
против					
хотеть					
акция					
1474	1106	601	261	812	796

В целом сравнительный анализ семантических моделей показывает, что во всех текстовых массивах преобладают (имеют больший суммарный вес) действия и намерения, — как информационные, так и направленные на физический мир. Напротив, перечисление каких-то свойств, обстоятельств, признаков явлений, упоминающихся в (анти)коррупционном контексте или формулировка прямого отношения к коррупции существенно уступают по частотно-

сти всем остальным блокам. В значительной степени это может быть связано с некоторой «самоочевидностью» характеристик коррупции в рамках данного дискурса, которая не требует дополнительной экспликации: популярность темы убивает объяснение.

Наконец, сопоставление рейтингов частотных лексем, составляющих семантическое ядро каждого корпуса, позволяет говорить об относительном весе разных значений — т. е., о той сравнительной роли, которую они играют в рамках соответствующего дискурса. Во всех случаях безусловно лидируют лексемы «коррупция», «вор», «жулик», «воровство», «украсть», «борьба», что в значительной степени предопределено организацией и методикой нашего исследования. Вместе с тем, различия между разными корпусами не менее существенны и не менее интересны. Если рассматривать два десятка наиболее частотных для каждого корпуса терминов как показательные для позиции автора, то мы можем наблюдать, наряду с общими для всех частотными лексемами, следующую картину:

- М. Кац: малоинформативная «двадцатка», отличающаяся лишь тем, что именно там концентрируются практически все виды коррупционного поведения, упоминаемые автором — поборы, взятки, откаты, распилы и пр. В целом это свидетельствует о доминировании идентификационной и номинативной функции;

- С. Кургинян: специфично упоминание в топ-20 политических партий, мафии, политических процессов, кибернетики (в рамках развёрнутой метафоры) и страны (термин, присутствующий и у других авторов, но обычно с очень маленьким весом). В значительной степени это говорит об ориентированности на экспликативные задачи и о некоторой масштабности подхода, что подтверждается и одним из самых высоких индексов по соответствующему проблемному блоку, 717;

- А. Навальный: «фонд», «расследование», «дело», «находить», «миллиард», «рубль», «соцопрос» — резюме публичной деятельности этого политика, на которой строится его политический дискурс;

- Президент РФ: наиболее частотные термины свидетельствуют об институциональной природе дискурса — речь идёт об органах власти, штрафах, предпринимаемых мерах, государственной организации противодействия, проблемах развития. Один из самых высоких показателей по соответствующему проблемному блоку («Действия») — 1437;

- Н. Стариков: наименее информативный частотный список, как будто собирающий наиболее общие для всех массивов лексемы, с отсутствием хоть сколько-нибудь выраженной специфики;

- И. Стрелков: гораздо больший, чем у других авторов, вес слов «страна», «государство», «народ», «собственность». Уникальное, и тем более интересное в силу большого веса, упоминание о «разложении» и о коррупции как системе, — в целом, ориентация на общенациональный и общенародный масштаб рассмотрения проблем;

- С. и А. Удальцовы: выборы, протестные акции, политические партии и акцент на становлении процессов и институтов явным образом подчёркивают политический активизм как доминанту этого дискурса.

Тот же сравнительный подход применим не только к весу отдельных лексем в рамках одного дискурса, но и при сопоставлении разных дискурсивных моделей. Например, при сравнении семантических ядер дискурса Н. Старикова и С. и А. Удальцовых мы видим 46 совпадающих лексем с существенно различным распределением весов: имеют практически один и тот же вес (и, в рамках нашей гипотезы, играют одинаковую роль в политическом дискурсе) термины вор, новый, власть, воровство, знать, коррупция, украсть. Напротив, значение лексемы «думать» для Старикова на 67 пунктов из 100 больше, чем для Удальцовых; для них, в свою очередь, категории «честный» и «выходить» (на митинг), соответственно, на 68 и 51 пункт более значимы, чем для Старикова. Точно так же — и весьма парадоксальным образом — Президент РФ гораздо чаще,

чем И. Стрелков говорит об обществе (разрыв в 81 пункт), а тот, в свою очередь, гораздо чаще, чем Президент говорит о государстве (разрыв в 87 пунктов). Вместе с тем, они уделяют практически равное внимание (соответствущие веса почти не различаются) деньгам, воровству, коррупции и взяткам, — при незначительном показателе корреляции между ядрами их дискурсов, 34 совпадающих лексемы.

Вариативность моделей антикоррупционного дискурса

В целом анализируемые варианты антикоррупционного дискурса чрезвычайно схожи, ориентированы на единую парадигму и демонстрируют много общего. Формальный анализ лишь отчасти показывает это сходство: из 700 лексем, формирующих семантические ядра рассматриваемых текстовых массивов, только 13 являются общими для всех корпусов (борьба, власть, вопрос, вор, воровство, говорить, давать, коррупция, сейчас, сказать, украсть, хотеть, человек/люди). Содержательное исследование семантических моделей даёт гораздо больше материала для обобщения.

Все называемые анализируемыми авторами действия суммируются в простой политической логике «желание — делание», отсутствует рефлексия (кроме Каца и, с отставанием, Удальцовых и Старикова): лишь в редких случаях появляется термин «знать», практически никогда — «думать», «осознавать», «размышлять» или «изучать», почти всегда «хотеть — говорить — делать». В целом, авторы анализируемых дискурсивных практик не претендуют на статус мыслителей. Результатом предпринимаемых или желательных действий большая часть авторов видит коррупционеров, сидящих в тюрьме — за исключением корпуса С. Кургиняна, где речь идёт о «победе» и «прекращении», и корпуса А. Навального, с высокой частотностью радикального «к стенке».

Для всех текстовых массивов характерно общее коммуникативное пространство: коррупция — главный термин и мы не видим попыток терминологически переопределить связанные с ней явления; расхожая фраза «вор должен сидеть в тюрьме» выступает как общая референция, отражая, вероятно, уже не привязанность к советскому кинематографу, а ориентацию на современный фольклор, так же как повсеместное использование жаргонного термина «откат» или сетевого мема «жулики и воры». Везде «вор» (вариант — «жулик и вор») — наиболее частотный термин: с одной стороны, это свидетельство популярности слогана, запущенного А. Навальным, с другой — представление о коррупции преимущественно в связи с преступлениями против собственности: коррупционер покушается не на государство, мораль или закон, а на собственность, он не изменник, бандит или дегенерат, а вор, его действия — не злоупотребление или разрушение, а кража. Закон и право упоминаются далеко не всеми авторами, а правовая («преступник», «нарушитель») или моральная («омерзительно») оценка действий коррупционеров или связанных с ними ситуаций появляются лишь в качестве исключения. Коррупция для большинства авторов — это экономический феномен, вопрос денег, собственности, рублей, гектаров и перераспределения доходов. Вместе с тем, как ни парадоксально, в большинстве массивов отсутствуют упоминания о бюджете или бюджетной деятельности в рамках коррупционной проблематики.

В контексте коррупционного поведения и противодействия коррупции взаимодействуют одни и те же субъекты: народ (общество, люди), чиновники, власть, государство, жулики, воры и коррупционеры. Крайне редко появляются представители правоохранительной системы (правоохранители в президентском дискурсе и судьи у Навального), предприниматели (Президент РФ, Кац, «компании» у Навального). Напротив, все анализируемые авторы, кроме Президента РФ, политизируют коррупцию, включая в число субъектов связанной с ней деятельности политические партии; олигархи в коррупционном контексте упоминаются только у

И. Стрелкова. Характеристика участников антикоррупционной деятельности как граждан, а их действий — как гражданских появляется лишь в виде исключения; точно также не попадают в частотные списки политики, как представители определённой категории.

Характеристики и свойства явлений, так или иначе связанных с коррупцией, мало представлены во всех массивах, как отмечалось выше. Объединяет их разве что единогласное признание серьёзности проблем, высокого уровня коррупции, обращение к фактам и примерам.

Термины, относящиеся к среде коррупционного поведения отличаются наибольшим разнообразием, однако и тут есть общие черты: большая часть авторов концентрируются на современности (даже если рассматривать лексемы «сегодня» и «сейчас» как своего рода вводные слова, их частотность говорит сама за себя); экскурсы в прошлое появляются только у Н. Старикова и С. Кургиняна, не исключая упоминаний о дне сегодняшнем. Напротив, терминология, относящаяся к будущему в выборке практически отсутствует: будущее лишь косвенно представлено через встречающееся единожды «развитие» и трижды — «проекты». Локализация коррупционных проблем, несмотря на некоторую вариативность, почти всегда включает в себя страну как предельную социогеографическую рамку (лишь в двух случаях речь идёт о городах и посёлках, никогда — о регионах, в одном случае — о мире в целом, наряду с проблемами страны).

Наконец, лексемы, характеризующие коммуникативный контекст самого высказывания в значительной степени отражают индивидуальные (реже — институциональные) стилистические привычки автора: так, массив И. Стрелкова содержит маркеры разговорной и достаточно авторитарной речи, тексты Президента РФ — официозные речевые обороты, корпус С. Кургиняна — большое количество метафор и сравнений и пр.

Различия между проанализированными дискурсивными моделями не менее существенны, и многие из них могут быть выявлены преимущественно с помощью формальных методов. Выявление лексических совпадений между семантическими ядрами, вокруг которых строятся изучаемые дискурсивные модели, а также различий в весе совпадающих лексем позволяет рассчитать своего рода негативную корреляцию — семантическую дистанцию, разделяющие разные дискурсивные модели. Ниже эти дистанции представлены в табличной форме; в качестве единицы измерения используются те же условные единицы, «пункты», которые применялись для расчёта веса отдельных лексем: минимальная возможная дистанция составляет 0 (полное совпадение лексем и их весов), максимальная — 10000 (ни одной совпадающей лексемы).

Таблица 8.

Семантическая дистанция между дискурсивными моделями

Авторы	**Дистанция**
Кургинян <=> Президент РФ	8197
Кургинян <=> Навальный	8044
Навальный <=> Президент РФ	7907
Кургинян <=> Стрелков	7837
Президент РФ <=> Стрелков	7738
Кургинян <=> Удальцов	7545
Кургинян <=> Стариков	7503

Президент РФ <=> Удальцов	7250
Кац <=> Кургинян	7199
Навальный <=> Стрелков	7194
Президент РФ <=> Стариков	7154
Кац <=> Президент РФ	6969
Стрелков <=> Удальцов	6837
Стариков <=> Стрелков	6815
Навальный <=> Стариков	6769
Навальный <=> Удальцов	6344
Стариков <=> Удальцов	6295
Кац <=> Стрелков	6085
Кац <=> Навальный	6047
Кац <=> Стариков	5758
Кац <=> Удальцов	5413

Как видно из таблицы, все дистанции укладываются в сравнительно небольшой диапазон (около 0.25 от теоретически возможного), но разрыв между крайними позициями достаточно заметен: дискурс М. Каца ближе к дискурсу С. и А. Удальцовых почти в полтора раза по отношению к дистанции, разделяющей антикоррупционные дискурсы С. Кургиняна и Президента РФ. Некоторые из этих дистанций, особенно в сравнении с соседними, кажутся контринтуитивными — прежде всего это относится как раз к крайним позициям: С. Кургинян пользуется репутацией пропрезидентского политика, хотя дистанция в этом случае максимальна, а М. Кац и С. и А. Удальцовы соответственно воспринимаются как представители противоположных идеологических течений, — но семантические ядра их высказываний наиболее близки друг к другу. Между тем, приведённые расчёты небезосновательны и позволяют продемонстрировать некоторые особенности организации антикоррупционного дискурса.

Количество лексических совпадений между семантическими ядрами президентского дискурса и текстового массива С. Кургиняна крайне невелико — 25 лексем. Содержательно все совпадающие лексемы являются максимально общими — по большей части это неспецифичные термины, связывающие все изучаемые массивы. Различия, с другой стороны, формируются за счёт преобладания в корпусе Кургиняна экспрессивной, метафоричной лексики, а в корпусе Президента РФ — официальных формул. Таким образом, значительная дистанция тут определяется в первую очередь стилистикой текстов и, во многом, разными адресатами.

В свою очередь, лексические совпадения для корпусов М. Каца и супругов Удальцовых составляют 56 лексем (один из самых высоких показателей для всех попарно сравниваемых семантических ядер). Большая часть этих совпадений, наряду с всё теми же общими терминами, формируется за счёт перечисления различных действий (у обоих корпусов для этого блока весьма высокие показатели, соответственно 1308 и 1474) и за счёт высокой степени политизации дискурса (большая, чем в других массивах концентрация таких терминов, как «политика», «партия», «выборы», «оппозиция», «власть», «система»). Наконец, именно два эти корпуса содержат практически полный перечень именований коррупционного поведения, за исключением архаизмов — в отличие от других текстовых массивов, в каждом из которых присутствуют лишь некоторые из них. В целом, можно предположить, что близкая дистанция

отражает тут уже не стилистику, а организацию деятельности: повседневный политический активизм, формирующий соответствующие лексические паттерны, ориентированные на разговор с улицей и на мобилизацию улицы.

В целом два эти случая представляются достаточно показательными для того, каким образом складывается антикоррупционный дискурс: несмотря на очевидное присутствие идеологических маркеров, он в значительной степени оказывается ориентирован на коммуникативную ситуацию и коммуникативные задачи, что сказывается и на его структуре. Ту же тенденцию можно увидеть сравнивая дискурсы М. Каца и А. Навального, или С. Кургиняна и Н. Старикова: ожидаемое и с очевидностью наблюдаемое идеологическое родство не ведёт механически к сокращению семантической дистанции между дискурсами, которые производятся в разных коммуникативных ситуациях и отражают разную организацию деятельности. Как следствие, построенные нами модели не поддаются однозначной кластеризации: структурные характеристики, которые выявляет формальный анализ, дают для этого некоторые основания, но они не совпадают ни с идеологическим посылом, ни с адресной группой разных дискурсивных проектов, как и с рядом других характеристик, которые могут быть прояснены а рамках содержательного анализа.

Основные особенности отдельных вариантов антикоррупционного дискурса уже обозначались в ходе изложения, остаётся их резюмировать. Структура действий, предлагаемых разными авторами, весьма точно отражает организацию их политической деятельности: для М. Каца это максимальное разнообразие средств уличной политики, от обдумывания и предложения проектов, написания текстов до голосования и протестных акций; для С. Кургиняна это преимущественно общие термины, не конкретизирующие способы деятельности; А. Навальный в равной степени ориентируется на мобилизацию общественного мнения, выборы, протестные акции и антикоррупционные расследования; Президент РФ выступает прежде всего в институциональном качестве, говоря о действиях, подразумевающих властные полномочия; Н. Стариков, как и С. Кургинян, оперирует максимально общими терминами с акцентом на знаниях и размышлении; И. Стрелков в большей степени, чем другие авторы, ориентируется на результат желательных действий, зачастую весьма радикальный.

Идентификация основных акторов, связанных с коррупционной или антикоррупционной деятельностью, представляет интерес в тех случаях, когда она отклоняется от стандартного лексикона и указания на абстрактных «чиновников»: указание на глав территориальных единиц (А. Навальный), предпринимателей (А. Навальный, Президент РФ, М. Кац), олигархов и военных (И. Стрелков) и пр. Точно так же указание на различные формы коррупционного поведения становится специфичной для данного дискурса характеристикой, когда перечень этих форм максимально подробен (М. Кац, И. Стрелков, С. и А. Удальцовы). Здесь, как и при определении среды (анти)коррупционных взаимодействий, можно увидеть достаточно устойчивую зависимость: ориентированность на повседневность и постоянное внимание к каким-то конкретным сферам общественной жизни (городская жизнь, образование, дороги, строительство и пр.) делает антикоррупционный дискурс более разнообразным и предметным.

Наконец, несмотря на частный характер коррупционной проблематики по отношению к масштабным идеологическим противостояниям и к основной политической повестке, идеологические предпочтения достаточно ярко проявляются в каждом из проанализированных массивов. Это касается и сохранения, даже применительно к частному сюжету, важных понятий-маркеров (средства производства, подчёркивание роли евреев, закона, собственности или индивидуализма), и предлагаемых форм деятельности (консерваторы не призывают к уличным протестам) и того веса, который приобретают одни и те же лексемы в разных моделях.

В целом проведённое исследование даёт основания для нескольких выводов:

1. В рамках публичного дискурса лидеры общественного мнения по-разному трактуют проблему коррупции; наиболее существенные различия связаны с коммуникативной ситуацией, в рамках которой производится дискурс, с идеологическими установками авторов, индивидуальными стилистическими особенностями.

2. Общее информационное пространство способствует формированию общих черт антикоррупционного дискурса, независимо от особенностей авторства (выявленные общие характеристики представлены в тексте статьи).

3. Особенности антикоррупционного дискурса, его семантическая структура могут быть выявлены методами количественного анализа текстов на естественном языке, в т. ч. с помощью предложенной методики выборочного частотного анализа и структурирования данных по проблемно-тематическим блокам.

4. Предложенная методика позволяет сформировать достаточно полное и дифференцированное представление об объекте исследования, в том числе — ввиду дальнейшего его изучения, на основе большого массива слабо структурированных текстовых данных.

Данное исследование планировалось в первую очередь ввиду апробации методики количественного анализа дискурсивных практик. На наш взгляд, оно позволило выявить существенные возможности и перспективы подобных исследований, даже на предложенном, достаточно базовом уровне. Вместе с тем, изучение антикоррупционного дискурса на материале сформированных лидерами общественного мнения текстов представляет и самостоятельное значение; основные итоги тем самым могут способствовать пониманию этого феномена.

Актуальные проблемы… 2016. Актуальные проблемы научного обеспечения государственной политики Российской Федерации в области противодействия коррупции. — *Сборник научных трудов по итогам Второй Всероссийской научной конференции* / Институт философии и права Уральского отделения РАН; ответственный редактор В.Н. Руденко. — Екатеринбург: УрО РАН.

Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. 2011. *Автоматическая обработка текстов на естественном языке и компьютерная лингвистика*. — М.: МИЭМ.

Кац М. 2017. *Кац предлагает победить*. — Доступно: http://maxkatz.livejournal.com. — Проверено: 20.08.2017.

Кургинян С. 2011. *Суть времени. Виртуальный клуб*. — Доступно: http://eot.su/essence_of_time/. — Проверено: 20.08.2017.

Навальный А. 2017. *Финальная битва между добром и нейтралитетом*. — Доступно: http://www.navalny.com. — Проверено: 20.08.2017.

НКРЯ. 2017. *Национальный корпус русского языка*. — Доступно: http://www.ruscorpora.ru/corpora-freq.html. — Проверено: 20.08.2017.

Стариков Н. 2017. *Николай Стариков. Блог*. — Доступно: https://nstarikov.ru/blog. — Проверено: 20.08.2017.

Старцев Я.Ю. 2006. Политическая система в зеркале дискурса: опыт реконструкции логики президентских посланий. — *ПОЛИТЭКС*. — Т. 3. — № 1. — С. 175–190.

Стрелков И. 2017. *Игорь Стрелков. Руководитель ОД «Новороссия»*. — Доступно: http://strelkov-i-i.livejournal.com. — Проверено: 20.08.2017.

Удальцов С. 2014. *Сергей Удальцов*. Доступно: http://echo.msk.ru/blog/udaltsov/. — Проверено: 20.08.2017.

Удальцова А. 2016. *Анастасия Удальцова*. — Доступно: http://udaltsova.livejournal.com. — Проверено: 20.08.2017.

Яндекс. 2011. *Все технологии. MyStem*. — Доступно: https://tech.yandex.ru/mystem/. — Проверено: 20.08.2017.

Bird S., Klein E., Loper E. 2014. *Natural Language Processing with Python*. — Доступно: http://www.nltk.org. — Проверено: 20.08.2017

Graphviz. 2017. *Graph Visualization Software*. — Доступно: http://www.graphviz.org/. — Проверено: 20.08.2017

Hardeniya N. 2015. *NLTK Essentials. Build cool NLP and machine learning applications using NLTK and other Python libraries*. — Packt Publishing.