

ТРИАНГУЛЯЦИЯ В КОНТЕНТ-АНАЛИЗЕ. ВОПРОСЫ МЕТОДОЛОГИИ И ЭМПИРИЧЕСКАЯ
ПРОВЕРКА

Автор: А. Н. ОЛЕЙНИК

ОЛЕЙНИК Антон Николаевич - доктор социологии (PhD), кандидат экономических наук, старший научный сотрудник Института экономики РАН и ассоциированный профессор (associate professor) университета "Мемориал", Канада (E-mail: aoleynik@hotmail.com).

Аннотация. В статье обсуждается новый подход к триангуляции результатов количественного и качественного контент-анализа, инструментальной и репрезентационной интерпретации данных. Предлагаемая модель триангуляции включает в себя три типа словарей: корреляционный, основанный на замещении и коды количественного контент-анализа. Для количественной оценки степени конвергенции результатов

анализа, осуществленного с помощью этих словарей, используется коэффициент корреляции Пирсона между значениями коэффициента Косинус в каждом из случаев. Потенциал модели показан на примере анализа двух совокупностей транскриптов интервью, проведенных с представителями российской элиты в 2005 - 2007 годах исследователями Левада-Центра (N = 43) и в рамках исследования, посвященного особенностям властных отношений в постсоветском контексте (N = 40). Эта модель триангуляции может быть использована для оценки надежности и валидности контент-анализа как в случае, когда он проводится одним исследователем, так и в качестве дополнения к обычно используемому в таких случаях коэффициенту согласия между кодировщиками. В последнем случае появляется возможность количественной оценки двух факторов, обуславливающих неполное совпадение результатов качественного и количественного контент-анализа: различия в подходах и субъективность оценок кодировщика.

Ключевые слова: контент-анализ • триангуляция • словарь, основанный на замещении • корреляционный словарь • интерпретация

Социальные и гуманитарные науки, в отличие от естественных, изучают человека и его действия не непосредственно, а опосредованно, через призму текста. Во всяком случае, именно так определяют специфику этих наук специалисты в области семиотики в самых различных ее версиях. Исследовательская программа, основы которой были заложены М. Бахтиным, сводится к тому, что "метод (и предмет) работы гуманитария - философский в том смысле, что гуманитарий общается не с эмпирическим субъектом, но с возможным, предполагаемым субъектом (автором), предметно представленным только в его ино-бытии - в тексте" [1]. Ю. Лотман тоже соглашается с центральным местом текста в социальном анализе, подчеркивая, что именно наиболее полная и доступная информация принимает в этой отрасли знаний текстовой формат [2]: литературные произведения, газетные и журнальные публикации, произведения эпистолярного жанра, записи речей и транскрипты углубленных интервью.

Использование информации в текстовом формате особенно плодотворно при анализе социального взаимодействия, ибо текст всегда имеет адресата: потенциально неограниченный круг читателей; узкий круг знающих автора лиц; наконец, "альтер-эго" самого автора (интимный дневник). Иными словами, при изучении социального действия, определяемого через учет поведения окружающих и корректировку на этой основе собственных поступков [3], без внимательного анализа текстов не обойтись.

Взгляд на окружающий мир через призму текста создает определенные трудности для исследователя. Формула динамично развивающейся науки, по версии Р. Коллинза, включает в себя три элемента: эмпирический характер, технология исследования и способность достигать высокого уровня абстракции [4]. Эмпирика текста иного рода, чем, скажем, физического объекта - камня или льда. Ее требуется реконструировать, причем единство в используемых методах и результатах отсутствует. Не случайно в приведенной выше цитате акцент делается на возможном и предполагаемом.

Технологии анализа текста - контент-анализа - принято разделять на количественные и качественные. В количественном случае на первом плане находятся такие понятия, как характер выборки (слов, текстов или их частей), частотность, статистические ассоциации. В качественном - символы, их контекст и интерпретация. М. Д. Уайт и Э. Марш резюмируют отличия между двумя техниками с помощью таблицы 1 [5].

В подтверждение непарадигмального характера социальных и гуманитарных наук [6], отсутствует согласие относительно сфер применимости качественных и количественных методов контент-анализа (см. подборку статей "Спор о методах" [7]). Основная часть публикаций в научных журналах подготовлена либо с помощью количественных, либо качественных методов. Например, среди 1195 статей, опубликованных

Таблица 1 Сравнительные характеристики количественного и качественного контент-анализа

Параметр сравнения	Количественный	Качественный
Исследовательский подход	Дедуктивный	Индуктивный
Исследовательская ориентация	Позитивизм	Герменевтика, гуманистическая традиция
Цель	Осуществление надежных (reliable) обобщений, анализ	Интерпретация
Выборка	Систематическая, преимущественно случайная	Преднамеренная (purposeful)
Процесс кодирования	Схема кодирования разрабатывается до начала анализа	Схема кодирования разрабатывается в процессе анализа
Коды	Объективные и четко сформулированные критерии	Существенен субъективный компонент
Основания проверки гипотез, вопросов исследования	Частотность, тесты статистической значимости	"Укорененность" в данных и глубина их понимания

между 1993 и 2002 гг. в трех ведущих западных журналах в сфере маркетинга, 46,3% написаны по результатам количественных исследований, 38,4% имеют теоретический характер и в 6,5% обсуждаются результаты качественных исследований [8].

В этой связи попытки избежать крайностей и развить подход, включающий элементы как количественного, так и качественного анализа, представляются и перспективными, и проблематичными одновременно. Такой "серединный" методологический подход получил название триангуляции. Он заключается в комбинации различных методов анализа и подчеркивании их взаимодополняющего, а не взаимоисключающего характера. Подобно тому, как использование множества ориентиров повышает точность в определении координат на местности, сравнение результатов анализа одного и того же текста с использованием различных методов позволяет увеличить их валидность. Если комбинируют качественные и количественные методы контент-анализа, то речь идет о триангуляции "между методами" (between/across methods), в отличие от триангуляции "внутри метода" (within method) с использованием нескольких количественных или качественных методов [9]. Триангуляция первого типа особенно эффективна для увеличения внешней валидности (external validity) результатов анализа.

При всей простоте идеи триангуляции ее четкая технология и процедура до сих пор не разработана. Каждый автор решает эту задачу по-своему, что снижает востребованность контент-анализа на основе триангуляции. В данной статье предпринята попытка предложить процедуру триангуляции результатов количественного и качественного контент-анализа текста, которая, с одной стороны, имеет теоретическое обоснование и, с другой, достаточно легка и "технологична" в применении разными исследователями в различных контекстах. Иными словами, предложенный анализ нацелен на создание методики триангуляции контент-анализа, которая была бы одновременно валидна и надежна.

Характеристики двух выборок исследуемых текстов. В качестве объекта анализа выступают транскрипты серии углубленных интервью, проведенных сотрудниками Аналитического центра Юрия Левады (Левада-Центр) в рамках исследования элитарных групп в постсоветской России [10]. Целью данного исследования был анализ властвующей элиты, а также оценка стабильности складывающейся социально-экономической системы в целом. Транскрипты 43 углубленных интервью (общим объемом 148369 слов) с экспертами ($N = 14$), представителями бизнеса ($N = 8$), законодательной власти ($N = 13$) и высокопоставленными государственными чиновниками ($N = 8$) были любезно предоставлены для вторичного анализа Л. Гудковым и Н. Зоркой. Таблица 2 содержит информацию о половозрастном распределении респондентов. Все интервью состоялись в Москве в 2005 - 2006 гг.

Таблица 2

Распределение респондентов по полу и возрасту (% от общего числа)

	Исследование Левада-Центра	Исследование "Особенности власти в постсоветском контексте"
<i>Пол</i>		
Мужской	83,7	80
Женский	16,3	20
<i>Возраст</i>		
До 41 года	0	32,5
41 - 50	27,9	40
51 - 60	44,2	17,5
Старше 61	27,9	10

Вторым объектом анализа, позволившим частично проверить ряд выводов и увеличить их надежность, стали транскрипты 40 углубленных интервью (общим объемом 246075 слов), проведенных в рамках исследовательского проекта "Особенности власти в постсоветском контексте: теоретические соображения и эмпирические исследования бюрократии", поддержанного Канадским советом по социальным и гуманитарным наукам (некоторые его результаты представлены в работе [11]). Исследование нацелено на изучение внутренней среды российского чиновничества, а также ограничений, в условиях которых оно действует. Среди собеседников - эксперты ($N = 9$), чиновники среднего (начальники отделов, заместители начальников департаментов $N = 15$) и верхнего среднего (начальники департаментов, заместители министров $N = 16$) уровня исполнительных органов федеральной и региональной власти. Интервью были проведены в 2006 - 2007 гг. в столице, Санкт-Петербурге, Ленинградской и Кемеровской областях¹.

Обе выборки имеют преднамеренный характер: в них попали собеседники, чье мнение исследователи сочли наиболее релевантным для решения исследовательских задач. Поэтому элементы индуктивного (inferential) статистического анализа, используемого ниже, приведены для иллюстрации потенциала предложенного подхода. Акцент же делается на анализе статистических связей внутри выборки текстов.

Реплики и вопросы интервьюеров были исключены из контент-анализа. Для его осуществления была использована программа *QDA Miner v. 2.0.8* с модулем *WordStat v. 5.1.12*, разработанные компанией *Provalis Research* (Монреаль).

Инструменты контент-анализа. Ключевым моментом контент-анализа является кодирование: соотнесение частей текста с категориями, с помощью которых осуществляется анализ. Список категорий образует своеобразный "словарь" для чтения и интерпретации транскриптов. Различают несколько типов словарей, в частности основанные на корреляции и на замещении. В основе "корреляционного" словаря лежит критерий частоты и порядка совместного расположения слов в тексте. Такой словарь в наибольшей мере соответствует специфике количественного контент-анализа и "репрезентационной" интерпретации текста. Корреляционная модель "проистекает из стремления рассматривать текст в его собственных терминах" [12]. Отсюда становится понятным термин "репрезентационный": задача анализирующего текст заключается в адекватной репрезентации намерений его автора. Согласно этой логике, аналитик ("читатель") должен привносить в текст как можно меньше новых идей и слов, стремясь взглянуть на него с позиции источника ("автора") [13]. Семантический анализ письменных обращений в Красноярский горсовет и публикаций в СМИ о руководителях региона показывает богатый потенциал использования корреляци-

¹ В проведении интервью участвовали Н. Апарина, О. Гаман-Голутвина, Е. Гвоздева, С. Глинкина, К. Клеман и автор статьи.

онных словарей. Анализ порядка слов в этих текстах подчас говорит о специфике властных отношений в России больше, чем развернутые философские и политологические рассуждения [14].

Словарь, основанный на замещении, имеет иную логику, хотя он тоже применим в количественном контент-анализе. Предполагается, что каждое из слов, отнесенных к определенной категории словаря, является взаимозаменяемым. Так, любое из встречающихся по тексту слов "жарко", "солнце" или "пляж" может быть отнесено к категории "лето". Эти слова не обязательно встречаются в тексте вместе. Такой словарь представляет собой "эвристическим образом образованный список слов, каждое из которых можно заменить другим благодаря их отнесению к одному и тому же коду" [12]. Стоит обратить внимание на эвристический характер словаря, основанного на замещении: список кодов (категорий) и слов отражает не только специфику самого текста, но и интересы "читателя", в том числе и теоретические конструкции, которые сложились в его голове до чтения текста. В этом смысле словари, основанные на замещении, подчинены логике не репрезентационной, а "инструментальной" интерпретации. Картина еще более усложняется, когда "читателей" несколько, причем каждый со своими интересами и подходами к прочтению текста.

Наконец, в списках кодов для качественного контент-анализа акцент делается не на частоте упоминания тех или иных слов, а на темах или сюжетах, которые присутствуют в тексте. Если с помощью корреляционного словаря можно уловить лишь явным образом сформулированные смыслы, то коды качественного контент-анализа отражают и неявным образом присутствующие в тексте идеи - те, что требуется вычитывать "между строк". В результате перспектива "читателя" еще более явно выражена. По этой причине качественный контент-анализ часто критикуют за субъективизм и отсутствие объективных критериев в кодировании: ведь кодом "лето" тогда отмечается не столько явным образом упомянутый "пляж", сколько настроение, возникшее при просмотре фотографий. Выражаясь более строго, качественный контент-анализ характеризуется сравнительно низкой надежностью.

Обычно для решения проблемы надежности при осуществлении качественного анализа формулируются эксплицитные правила кодирования, в котором принимает участие несколько исследователей. После завершения ими работы коды сравниваются и рассчитывается индекс согласия между кодировщиками, равный отношению между числом совпадений кодов к сумме числа совпадений кодов и числа несовпадений кодов. Значения индекса согласия, превышающие 0.7, рассматриваются в качестве показателя приемлемой надежности [15].

Выбор типа словаря стратегии контент-анализа не в последнюю очередь зависит от типа анализируемого текста. Ю. Лотман различает риторические и стилистические тексты [2]. Первые написаны с целью генерирования новых идей, вторые - для логичного и лишеного противоречий изложения существующих. Соответственно, инструментальная интерпретация представляется уместной прежде всего в отношении риторических текстов, а репрезентационная - в отношении стилистических. Сходную идею, но с несколько иным акцентом, формулируют Р. Хогенраад с соавторами: "чем менее структурирован сам текст, тем более структурированным и основанным на четко определенных категориях должен быть его анализ" [12]. Несколько упрощая, допустимо предположить, что для анализа учебников и академических текстов в парадигмальных науках лучше всего подходит корреляционный словарь, тогда как текст "полифонического" романа, включающий множественность равноправных сознаний с их "правдами" [16], лучше читать и "достраивать", используя техники качественного контент-анализа.

Углубленные структурированные интервью представляют собой некий промежуточный случай. С одной стороны, они проводятся на основе общего "гайда" и включают ответы на одни и те же вопросы. С другой стороны, они отражают уникальное мнение респондента, а респондент зачастую вправе задавать уточняющие вопросы и вступать в диалог. Налицо комбинация элементов как риторических, так и стилисти-

ческих текстов, что открывает редкую возможность легитимного применения различных типов словарей в контент-анализе одного и того же текста. Исследовательская задача, решению которой посвящена настоящая статья, заключается в разработке техники триангуляции результатов контент-анализа транскриптов углубленных интервью, осуществленного с помощью трех типов словарей: (i) корреляционного, (ii) основанного на замещении и (iii) списка "качественных" кодов.

Словари, используемые в контент-анализе транскриптов. *Словарь (iii)* - список кодов - произведен от основного исследовательского вопроса, которому были посвящены интервью с представителями властвующей элиты. Иными словами, словарь (iii) обусловлен интересами прежде всего исследователя ("читателя") и имеет инструментальный характер. В соответствии с задачами двух исследовательских проектов были созданы два словаря качественных кодов: C_{iii} и B_{iii} ("С" означает исследование Левада-Центра, посвященное вопросам стабильности складывающейся в России социально-экономической системы; "В" - исследование особенностей власти в постсоветском контексте).

Словарь C_{iii} включает 12 кодов, организованных в 3 "ядра" (nodes): "Нестабильность", "Факторы нестабильности" и "Факторы стабильности". Для оценки внешней валидности кодов, в соответствии с принятыми в литературе рекомендациями, каждый из них иллюстрируется тремя примерами (табл. 3). Обычно рекомендуется включение в отчет "как минимум трех независимых примеров интерпретации каждого кода, особенно если речь идет о кодировании латентных переменных" [15].

В качестве иллюстрации приводится определение кода "Действия "просвещенных" бюрократов" как стабилизирующий фактор, а также три примера его интерпретации. В истории нередки случаи, когда стабильное развитие и модернизация страны осуществлялись усилиями "просвещенных" бюрократов. "Великие" реформы 1860-х годов в России, равно как и реформы конца XVIII в. в странах Юго-Восточной Германии, были подготовлены именно такими людьми [17].

"Это довольно сильные люди, которые способны определять стратегию - много таких людей, которые реально могут предлагать интересные, нестандартные решения, и хорошо владеют и макроэкономическими показателями, и понимают взаимосвязь явлений" (бизнес, ж., 41 - 50²). "...Идея служения. Если эта идея захватывает человека, она связана с тем, чтобы служить чему-то большему, нежели то, что ты делаешь в обыкновенной жизни. Тогда это оказывается очень мощным фактором, который помогает государству - или государственным структурам - привлечь людей, для которых эта идея служения близка и важна" (бизнес, м., 41 - 50). "В высшее руководство естественным путем отбора проникают люди, заинтересованные в судьбах государства. Хищник такого масштаба не проникает. Или проникает, но у него двойственная позиция... И вот много таких людей, у которых внутри и хищник, и государственный человек. Если государственный человек победит - все произойдет, если победит хищник - то безнадежно" (законодательная власть, м., 51 - 60).

Словарь B_{iii} включает 46 кодов, сгруппированных в шесть "ядер" (табл. 4): "Бизнес" (отношения между представителями власти и бизнеса), "Конституция" (термин используется в смысле, предложенном Э. Гидденсом [18] - для описания совокупности норм и практик, структурирующих повседневную жизнь российского чиновничества), "Ограничения" (факторы, способные ограничить своекорыстное поведение чиновников, они подробно обсуждаются в [19]), "Критические ситуации" (проективные ситуации, в которых не существует единственного социально одобряемого варианта поведения), "Песочные часы" (метафора, предложенная Р. Роузом с соавторами для описания разрыва между властвующей элитой и простыми россиянами на уровне повседневности [20]) и "Власть" (техники навязывания воли).

² Здесь и далее в скобках обозначены сфера деятельности респондента, пол, возрастная категория.

Таблица 3 Коды для качественного контент-анализа (C_{iii})

Ядро	Код	Число кодов	% кодов	Число текстов	% текстов	Число слов	% слов
Дестабилизирующие факторы	Дефицит идей	31	5,2	21	48,8	1214	0,6
	Состояние экономики и инфраструктуры	42	7,1	20	46,5	2375	1,3
	Массовые протесты	33	5,6	21	48,8	1499	0,8
	Действия властвующей элиты	170	28,7	39	90,7	7217	3,8
	Проблемы передачи власти в 2008 году	29	4,9	19	44,2	1424	0,8
	Вертикаль власти	59	10,0	30	69,8	3500	1,9
Нестабильность	Риск	16	2,7	13	30,2	998	0,5
	Неопределенность	92	15,5	37	86,0	4150	2,2
Стабилизирующие факторы	Действия "просвещенных" бюрократов	25	4,2	14	32,6	1076	0,6
	Манипуляция	25	4,2	15	34,9	929	0,5
	Массовая поддержка	23	3,9	14	32,6	828	0,4
	Ресурсная рента	47	7,9	26	60,5	1455	0,8

Характеризующийся одной из самых высоких частотностей код "Команды" использован в качестве иллюстрации. Под командой понимается группа чиновников, объединенная не по формальному признаку работы в одном и том же учреждении, а узлами личной лояльности, общностью пережитого, личностным доверием и другими неформальными связями. В этом смысле команда имеет много общего с сетью как "устойчивой совокупностью контактов или схожих социальных связей между индивидами" [21].

"Своих приводят очень много. Я много своих приводил. Приводишь своего человека на малозначительные должности советника, консультанта и говоришь - посмотри, вот, на человека" (средний уровень, м., до 41). "Команда может распасться на какое-то время, в том смысле, что люди остаются в разных ведомствах, а потом снова собираются... Я видел даже разные примеры, когда людей специально рассеивали - вот большой начальник уходит, и, казалось бы, он собрал под собой большую команду... Вот и опять же изучаем телефонный справочник: он собрал около себя команду, там были разные люди - были его однокашники по X, были совершенно молодые люди из Y... Они работали несколько лет вместе, скажем, до 2001 года где-то... После этого каким-то образом ситуация изменилась ... Ну, собственно говоря... у нас появился новый президент, пришел Z, новые кадровые возможности, перетурбации в Правительстве... и часть людей разлетелась - по другим ведомствам, министерствам, департаментам Белого Дома... А потом я смотрю, раз - в W все и собрались, как только Q стал министром" (верхний средний уровень, м., до 41). "Солидарность ограничена ну очень узким кругом... действительно идет речь о круге людей, то, что можно назвать командой - там еще какая-то солидарность возможна" (эксперт, ж., 41 - 50).

Словари (ii) основаны на замещении. Те же самые категории, которые присутствуют в словарях C_{iii} и B_{iii} , были заменены списками из конкретных слов и фраз. Иначе говоря, присутствие в тексте любого(ой) из этих слов или фраз рассматривалось в качестве достаточного основания для его кодирования соответствующим образом (табл. 5 и 6). Роль исследователя свелась к составлению списков C_{ii} и B_{ii} , а программа автоматически и не обращая внимания на латентный смысл (если только он не предполагался самой структурой ключевых фраз) осуществила кодировку по заданным алгоритмам.

Словари (i) имеют корреляционный характер. В них были автоматически включены все слова, встречающиеся как минимум пять раз, но менее, чем в 75% транскриптов, всего 2386 слов в словаре C_i (табл. 7) и 3500 слов в словаре B_i (табл. 8).

Модель триангуляции. Для оценки надежности и валидности инструмента качественного контент-анализа, а именно словарей (iii) была разработана и протестирована-

Таблица 4 Коды для качественного контент-анализа (B_{iii})

Ядро	Код	Число кодов	% кодов	Число текстов	% текстов	Число слов	% слов
Бизнес	Бюджет	43	1,3	23	57,5%	2981	0,9
	"Ноги" документа	21	0,7	9	22,5%	2252	0,7
	Коррупция	129	4,0	34	85,0%	11435	3,3
	Сети	98	3,1	31	77,5%	6786	2,0
	"Вращающиеся двери"	84	2,6	34	85,0%	4846	1,4
Конституция	Общее благо	79	2,5	34	85,0%	5501	1,6
	Конфликт интересов	87	2,7	31	77,5%	6527	1,9
	Согласование	78	2,4	28	70,0%	6193	1,8
	Функционал	154	4,8	37	92,5%	8513	2,5
	Поколения	53	1,7	19	47,5%	3108	0,9
	Информационные потоки	52	1,6	21	52,5%	3489	1,0
	Родственники	63	2,0	27	67,5%	3729	1,1
	Уровни иерархии	178	5,6	37	92,5%	12140	3,5
	Силовые ведомства	44	1,4	16	40,0%	3920	1,1
	Администрация президента	44	1,4	19	47,5%	2544	0,7
	Карьерный рост	75	2,4	25	62,5%	6161	1,8
	Ресурсы	65	2,0	33	82,5%	4440	1,3
	Команды	246	7,7	39	97,5%	20470	5,9
	Доверие	52	1,6	33	82,5%	3078	0,9
	Призвание	148	4,6	36	90,0%	10315	3,0
Ограничения	Административная реформа	41	1,3	26	65,0%	3069	0,9
	Харизма	32	1,0	19	47,5%	1731	0,5
	Пересечение интересов	76	2,4	31	77,5%	5454	1,6
	Выборы	36	1,1	14	35,0%	3444	1,0
	Международное давление	51	1,6	25	62,5%	2294	0,7
	Идеология	40	1,3	24	60,0%	1936	0,6
	Закон	113	3,5	32	80,0%	7425	2,1
	Контроль со стороны коллег	32	1,0	19	47,5%	1541	0,4
	Массовые протесты	63	2,0	23	57,5%	4653	1,3
	СМИ	34	1,1	16	40,0%	3072	0,9
	Частная собственность	25	0,8	14	35,0%	1927	0,6
	Регламенты	103	3,2	32	80,0%	6658	1,9
	Религия	3	0,1	3	7,5%	118	0,0
	Соперники в борьбе за власть	16	0,5	11	27,5%	919	0,3
	Традиции	49	1,5	28	70,0%	2674	0,8
	Вертикаль власти	164	5,1	36	90,0%	10055	2,9
Критич. ситуац.	Работать ли с родственниками	31	1,0	21	52,5%	1645	0,5
	Почему нужно доверять	29	0,9	23	57,5%	1537	0,4
Песоч. часы	Обмен информацией	84	2,6	33	82,5%	7945	2,3
	Мы и Они	84	2,6	31	77,5%	5329	1,5

Власть	Принуждение	86	2,7	29	72,5%	4962	1,4
	Власть как самоцель или инструмент	37	1,2	25	62,5%	1655	0,5
	Позитивные стимулы	36	1,1	20	50,0%	2636	0,8
	Легитимация	29	0,9	18	45,0%	1774	0,5
	Выбор меньшего зла	60	1,9	24	60,0%	5090	1,5
	Манипуляция	40	1,3	20	50,0%	2629	0,8

вана на двух выборках следующая модель (рис. 2). Словари (iii), (ii) и (i) условно расположены на континиуме от инструментальной до репрезентационной интерпретации, причем словарь, основанный на замещении, представляет собой промежуточную форму, находясь примерно посередине: его категории были взяты из "инструментального" словаря (i), но наполнены в логике корреляционного анализа. Если между словарями существует ассоциация - связь в той или иной форме, то это было бы серьезным аргумен-

Таблица 5 Словарь S_{ib} основанный на замещении

Код (категория)	Слова и фразы	Частота	% от общего числа	Число текстов	% текстов	TF-IDF
Дефицит идей	низк* AND интеллект*; иде* AND нет; стратег* AND нет; понимай* AND нет; неординарн*	10	1,2	9	20,9	6,8
Состояние экономики и инфраструктуры	*инфляц*; налог* AND политик*; рост* AND замедл*; инвестиц*; неконкурент*	82	9,5	21	48,8	25,5
Массовые протесты	бунт*; простеет*; революц*	129	15,0	29	67,4	22,1
Действия властвующей элиты	борьб* AND клан*; конфликт* AND интерес*; *компетентн* AND мало; слом* AND систем*; админи-страт* AND реформ*; временщ*; раскол*_элит*; личн*_преданн*; некомпетентн*; несамостоят*	122	14,2	31	72,1	17,3
Проблемы передачи власти в 2008 г.	интриг*; преемник*; поколен*	81	9,4	29	67,4	13,9
Вертикаль власти	вертикал* AND власт*; концентрац* AND власт*; концентрац* AND ответствен*; обратн*_связ*; исполнен*; цен*_ошибк*	48	5,6	21	48,8	14,9
Риск	вероятност*; просчит*; прогноз*	49	5,7	23	53,5	13,3
Неопределенность	неизвестн*; неожидан*; непредсказуем*; неопределен*; случайн*	74	8,6	31	72,1	10,5
Действия "просвещенных" бюрократов	благородн*; умны*; служен*	28	3,3	18	41,9	10,6
Манипуляция	телев* AND показ*; пропаганд*; *цензур*; фальсифи*	35	4,1	18	41,9	13,2
Массовая поддержка	довер* AND президент; поддержк*; рейтинг*	81	9,4	27	62,8	16,4
Ресурсная рента	нефт*	120	14,0	30	69,8	18,8

Примечание: * означает любую букву или набор букв; AND присутствие двух ключевых слов в одной и той же фразе; _ два ключевых слова следуют строго одно за другим; TF-IDF (term frequency - inverse document frequency) рассчитывается как произведение отношения частоты конкретной категории к совокупной частоте всех категорий в документе и логарифма отношения общего числа текстов к числу текстов, в которых присутствует конкретная категория.

том, свидетельствующим о надежности и валидности инструментов. Нахождение такой связи является практической реализацией триангуляции "между методами".

Модель имеет несколько "измерений", в зависимости от количества "читателей" исходных текстов с присущими каждому из них специфическими интересами (или тестируемых теорий). "Хотя можно предположить, что число возможных тем в тексте и их вариаций, конечно, не представляется возможным точно определить" [12]. В данном случае количество измерений ограничено двумя, по числу исследовательских проектов, в рамках которых были проведены интервью. Иными словами, вначале интервью исследования "С" были проанализированы с точки зрения его исследовательских задач, а затем - с точки зрения исследовательских задач проекта "В". Обоснованием данного решения является тот факт, что в обоих случаях объектом анализа является властвующая элита современной России, а именно "административный класс" [22]. То же самое проделано и с интервью исследования "В" (рис. 3).

Наиболее простой формой существования возможной связи является измерение расстояний между транскриптами, проранжированными по критерию сродства числа и конфигурации кодов (категорий) с использованием каждого из словарей. Затем достаточно рассчитать корреляции между рангами транскриптов или коэффициентами

Таблица 6 Словарь V_{ib} основанный на замещении

Код (категория)	Слова и фразы	Частота	% от общего числа	Число текстов	% текстов	TF-IDF
Бюджет	распил*; связ* AND бюджет*	8	0,0	6	15,0	6,6
Коррупция	взятк*; откат*	130	0,1	30	75,0	16,2
Общее благо	государствен*_интерес*; государствен*_мышлен*; общ*_дел*; общ*_благ*; общ*_интерес	106	0,0	30	75,0	13,2
Конфликт интересов	интерес* AND конфликт*; интерес* AND противореч*; противостоян*	19	0,0	9	22,5	12,3
Согласование	согласован*	55	0,0	20	50,0	16,6
Поколения	поколен*; молод*	159	0,1	29	72,5	22,2
Уровни иерархии	центр AND власт*; муницип*; регион*_уровн*; федеральн*; регион*_власт*	142	0,1	31	77,5	15,7
Силловые ведомства	военн*; арми*; милиц*; фсб*; силовик*	120	0,0	26	65,0	22,5
Администрация президента	администрац*_президент*; кремл*	66	0,0	20	50,0	19,9
Карьерный рост	повышен* AND служб*; продвижен* AND служб*; карьер*	121	0,0	26	65,0	22,6
Ресурсы	баррел*; нефт*; газ*	70	0,0	19	47,5	22,6
Команды	скамейк*; команд*; клан*; коллектив*	530	0,2	38	95,0	11,8
Административная реформа	административ*_рефор*; рефор* AND аппарат*	83	0,0	26	65,0	15,5
Пересечение интересов	олигарх* AND власт*; рын* AND власт*; монопол* AND доминир*	4	0,0	2	5,0	5,2
Выборы	*выборн*; выборы	32	0,0	11	27,5	17,9
Международное давление	угроз* AND внешн*; угроз* AND военн*; америка*; европ*; штаты; нато	43	0,0	19	47,5	13,9
Идеология	идеологическ*; коммунист*_иде*; либерал*_иде*; прав*_иде*	10	0,0	8	20,0	7,0
Закон	закон* AND уважен*; закон* AND выполнен*; закон* AND подчинен*; закон* AND треб*; конститу*	25	0,0	15	37,5	10,6
Массовые протесты	бунт*; протест*; революц*	46	0,0	14	35,0	21,0
СМИ	газет* NOT цензур*; газет* NOT контр*; телев* NOT цензур*; телев* NOT контр*	56	0,0	21	52,5	15,7
Частная собственность	част*_собственн*	1	0,0	1	2,5	1,6
Вертикаль власти	вертик*_власт*	42	0,0	19	47,5	13,6
Песочные часы	мы AND они; нас AND они	72	0,0	20	50,0	21,7
Принуждение	насил*; примен*_сил*; тюр*	12	0,0	8	20,0	8,4
Власть как самоцель или инструмент	власт* AND сохранен*; власт* AND удержат*	1	0,0	1	2,5	1,6

Манипуляция	пиар*; манипулир* NOT цен*	17	0,0	9	22,5	11,0
-------------	----------------------------	----	-----	---	------	------

Примечание: NOT означает присутствие первого ключевого слова, когда второе отсутствует в этой же фразе.

Косинус (Cosine), равными косинусу угла между векторами кодов каждого из транскриптов. Требуется лишь определить с точкой отсчета - случайно выбранная точка отсчета (вектор кодов транскрипта X_i) увеличивает риск ошибки. При выборе точки отсчета наиболее разумным представляется анализ порядка агломерации транскриптов по критерию коэффициента Косинус. В качестве точки отсчета можно выбрать любой из пары, с которой начинается процесс агломерации. В проекте "С" таковым оказался транскрипт интервью с бывшим заместителем федерального министра, ныне руководителем банка (м., 41 - 50). В проекте "В" - транскрипт интервью с

Таблица 7 Корреляционный словарь C_i - (извлечение: 30 наиболее часто встречающихся слов)

	Частота	% от общего числа	Число текстов	% текстов	TF-IDF
Бизнес	162	0,4	30	69,8	25,3
Деньги	148	0,4	31	72,1	21,0
Система	142	0,3	32	74,4	18,2
Достаточно	140	0,3	31	72,1	19,9
Сегодня	138	0,3	28	65,1	25,7
Зрения	137	0,3	31	72,1	19,5
Никто	129	0,3	31	72,1	18,3
Точки	127	0,3	28	65,1	23,7
Почему	120	0,3	32	74,4	15,4
Ты	120	0,3	27	62,8	24,3
Происходит	113	0,3	31	72,1	16,1
Через	101	0,2	32	74,4	13,0
Один	101	0,2	30	69,8	15,8
Страны	100	0,2	32	74,4	12,8
Ему	100	0,2	30	69,8	15,6
Поскольку	98	0,2	30	69,8	15,3
Абсолютно	97	0,2	29	67,4	16,6
Далее	97	0,2	27	62,8	19,6
Ситуация	96	0,2	29	67,4	16,4
Элиты	96	0,2	27	62,8	19,4
Закон	96	0,2	22	51,2	27,9
Свои	94	0,2	32	74,4	12,1
Стране	94	0,2	32	74,4	12,1
Теперь	94	0,2	29	67,4	16,1
Часть	93	0,2	29	67,4	15,9
Уровне	92	0,2	31	72,1	13,1
Сказал	92	0,2	30	69,8	14,4
Проблема	92	0,2	27	62,8	18,6
Бизнеса	91	0,2	26	60,5	19,9
Решения	90	0,2	30	69,8	14,1

начальником отдела федерального министерства, ранее длительный период работавшего в налоговой полиции (м., 41 - 50). Приведенные коэффициенты корреляции рассчитаны между списками значений коэффициентов Косинус между указанным транскриптом и остальными транскриптами при использовании словарей (i), (ii) и (iii).

Коэффициенты корреляции рассчитаны как для "радиусов", то есть между словарями, сконструированными в рамках одного и того же исследовательского проекта (B_i , B_{ii} и B_{iii}), так и для "хорд", то есть между однотипными словарями, разработанными в рамках различных проектов для анализа одного и того же текста (B_i и C_i , C_{ii} и B_{iii}). Что касается расчета корреляций для "хорд", то это позволяет обеспечить сравнение одноуровневых понятий и конструкций [15].

Значения коэффициентов корреляции Пирсона свидетельствуют о наличии средне-сильной связи между результатами контент-анализа для "радиусов" во всех случаях, кроме одного (C_i и C_{ii} при анализе транскриптов проекта "В"). Эти значения можно интерпретировать двояким образом. Во-первых, как элемент описательной (descriptive) статистики: в рамках данной совокупности транскриптов результаты контент-анализа характеризуются "избирательным средством". Средне-сильный характер ассоциации может быть отнесен на счет, с одной стороны, концептуальных различий между качественным и количественным контент-анализом (в частности, учетом латентных и явных смыслов, инструментальной и репрезентационной ориентацией и т.д.) и, с другой стороны, последовательности в осуществлении исследователем кодирования (надежности его или ее анализа).

Таблица 8 Корреляционный словарь B_i (извлечение: 30 наиболее часто встречающихся слов)

	Частота	% от общего числа	Число текстов	% текстов	TF-IDF
Допустим	198	0,1	25	62,5	40,4
Деле	189	0,1	29	72,5	26,4
Тебя	183	0,1	25	62,5	37,4
Знаете	167	0,1	27	67,5	28,5
Чиновник	154	0,1	27	67,5	26,3
Департамента	153	0,1	26	65,0	28,6
Зрения	142	0,1	28	70,0	22,0
Кажется	138	0,1	25	62,5	28,2
Далее	133	0,1	22	55,0	34,5
Точки	132	0,1	29	72,5	18,4
Министерство	117	0,0	20	50,0	35,2
Министерства	116	0,0	24	60,0	25,7
Естественно	114	0,0	26	65,0	21,3
Понимаете	ПО	0,0	24	60,0	24,4
Пришел	109	0,0	29	72,5	15,2
Сразу	109	0,0	29	72,5	15,2
Управления	109	0,0	29	72,5	15,2
Абсолютно	108	0,0	28	70,0	16,7
Чиновников	108	0,0	28	70,0	16,7
Сложно	105	0,0	28	70,0	16,3
Момент	103	0,0	28	70,0	16,0
Которое	102	0,0	28	70,0	15,8
Отдела	100	0,0	22	55,0	26,0
Образом	99	0,0	27	67,5	16,9
Какие	98	0,0	29	72,5	13,7
Году	97	0,0	27	67,5	16,6
Должность	97	0,0	29	72,5	13,5
Многие	95	0,0	28	70,0	14,7
Госслужбе	94	0,0	27	67,5	16,0
Замминистра	94	0,0	20	50,0	28,3

Для более четкого разграничения этих двух причин требуется учет в предложенной модели стандартного коэффициента согласия между кодировщиками. Значения коэффициентов корреляции варьируются в интервале от 0,39 до 0,61 при том, что ориентиром является величина 0,7 (минимально приемлемое значение коэффициента согласия). Иными словами, для оценки вклада концептуальных различий между качественным и количественным контент-анализом можно привести коэффициенты корреляции к общему знаменателю, равному 0,7³. Тогда, например, коэффициент корреляции между C_i и B_{ii} равный 0,61, означает минимальные различия между двумя формами количественного контент-анализа (нормализованный коэффициент корреляции, равный 0,87).

Во-вторых, значительно более осторожно следует подходить к интерпретации полученных данных с позиций индуктивного статистического анализа. Если бы выборка транскриптов была случайной (например, каждое пятое из всех публичных выступлений официального лица), то статистически значимые результаты контент-анализа можно было распространить на всю совокупность транскриптов. В рассматриваемом же случае статистическая значимость в лучшем случае позволяет предположить возможность обобщения результатов на всю совокупность транскриптов, анализируемых (кодировемых) автором.

³ Данная идея была предложена Н. Пеладом (N. Peladeau) - одним из разработчиков *QDA Miner* (личная переписка с автором статьи, апрель 2008 г.).

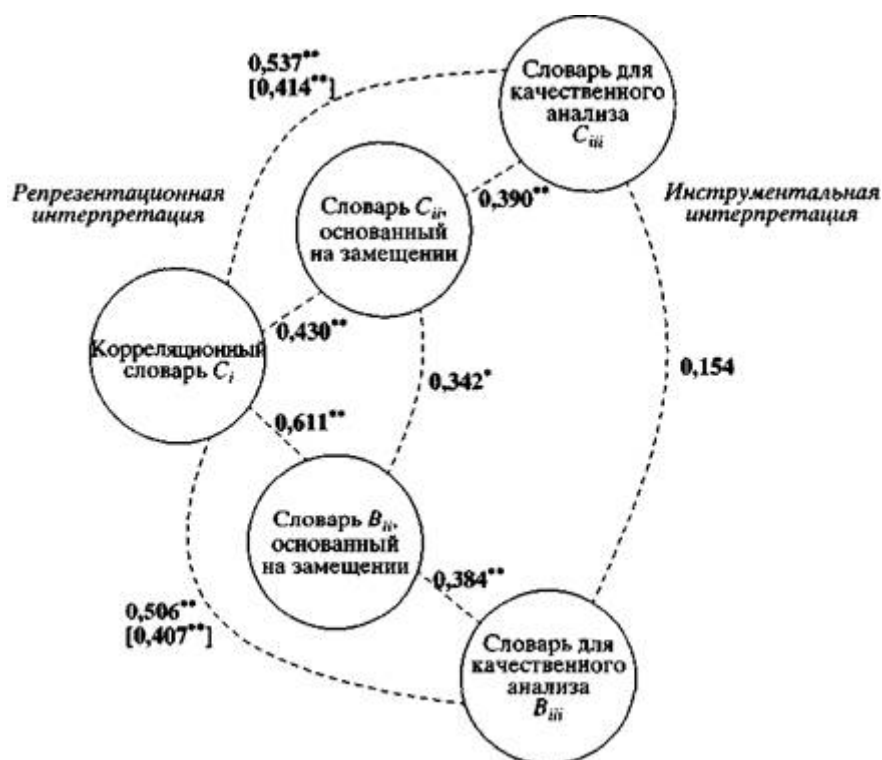


Рис. 1. Модель триангуляции результатов контент-анализа с использованием трех словарей транскриптов интервью, проведенных Левада-Центром.

Примечание: Коэффициенты корреляции, значимые на уровне $\alpha = 0,01$ (двухсторонний), отмечены **, на уровне $\alpha = 0,05$ - *.

Стоит заметить, что результаты "чисто" качественного и "чисто" количественного контент-анализа коррелируют между собой сильнее, чем в случае опосредованной словарем, основанным на замещении, связи. Можно предположить, что анализ с использованием словаря, основанного на замещении, имеет свою собственную логику, не сводимую ни к первой, ни ко второй форме контент-анализа. Расчет частичного коэффициента корреляции между C_i и C_{iii} - когда влияние словаря C_{ii} контролируется - подтверждает данный факт: $r_{i iii} = 0,537$, $r_{i iii ii} = 0,414$ (значения частичных коэффициентов корреляции приведены на рисунках в квадратных скобках). Словарь (i) эффективен в улавливании явных смыслов слов и фраз, словарь (ш) - латентных, а словарь (ii) - в улавливании явных смыслов не всех слов и фраз, а лишь тех, что определены словарем (iii). Напрашивается аналогия с делением в институциональной экономике всех организационных форм на рынок, фирму и "гибрид", или отношенческую контрактацию. Последняя, занимая промежуточное положение между первыми двумя, тем не менее имеет собственную логику и динамику [23].

Стоит обратить внимание на различия в коэффициентах корреляции для "радиусов" в зависимости от того, какие интересы преследует читатель. К каждой из выборок транскриптов применялись как словари, отражающие заложенные в "гайд" принципы, так и словари, построенные для решения других исследовательских задач. Логично предположить, что ассоциативные связи для "радиусов" в первом случае будут сильнее, чем во втором.

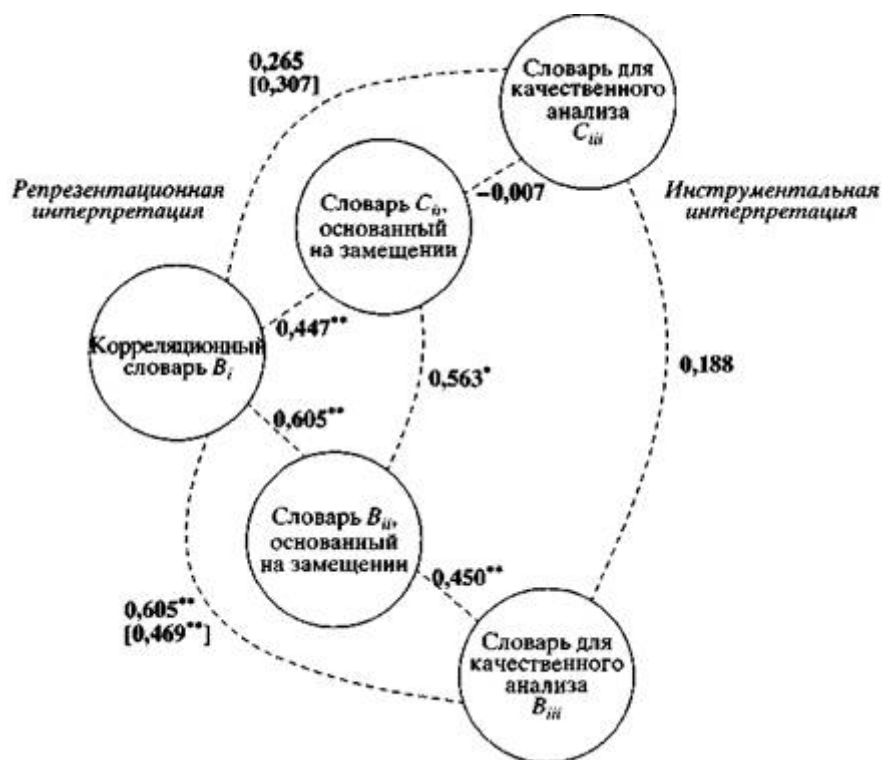


Рис. 2. Модель триангуляции результатов контент-анализа с использованием трех словарей транскриптов интервью, проведенных в рамках проекта "Особенности власти в постсоветском контексте"

Однако ассоциативные связи между словарями, созданными в рамках проекта "В", оказались сильнее ассоциативных связей между словарями проекта "С" при анализе и транскриптов "В", и транскриптов "С". Данный факт может быть интерпретирован двояко. С одной стороны, как свидетельство большей аутентичности словарей "В" дискурсу самих чиновников, даже когда последние говорят на специально заданные - в соответствии с программой исследования "С" - темы. Действительно, в проекте "В" исследователи в большей мере полагались на методы индукции и укорененного (grounded) теоретизирования, чем в проекте "С". С другой стороны, обсуждаемый эмпирический факт может свидетельствовать о наличии у текста ("автора") внутренней логики, на которую собеседник может влиять лишь в ограниченных масштабах. Подобно музыке, композиция которой подчиняется определенным канонам, дискурс автора может обладать внутренней логикой вне зависимости от обсуждаемых сюжетов и конкретных собеседников.

СПИСОК ЛИТЕРАТУРЫ

1. Библер В. Михаил Михайлович Бахтин, или Поэтика культуры. М.: Прогресс, 1991. С. 72.
2. Lotman Y. Universe of the Mind: A Semiotic Theory of Culture. Bloomington and Indianapolis: Indiana University Press, 1990. P. 2.
3. Weber M. Economy and Society: An Outline of Interpretative Sociology. N. Y.: Bedminster Press, 1968. P. 4.
4. Collins R. The Sociology of Philosophies: A Global Theory of Intellectual Change. Cambridge, MA: The Belknap Press of Harvard University Press, 1998. P. 535.
5. White M. D., Marsh E. E. Content analysis: a flexible methodology // Library trends. 2006. N 1 (55). P. 35 - 36.
6. Kuhn T. Scientific Paradigms // Barnes B., editor, Sociology of Science: Selected Readings. Harmondsworth: Penguin Books, 1972. P. 86.
7. Неприкосновенный запас. 2004. N 3 (35).

8. *Hanson D., Grimmer M.* The mix of qualitative and quantitative research in major marketing journals, 1993 - 2002 // *European Journal of Marketing*. 2007. N 1/2 (41). P. 63.
9. *Jick T. D.* Mixing qualitative and quantitative methods: triangulation in action // *Administrative Science Quarterly*. 1979. N 4 (24). P. 603; *Bryman A., Teevan J.* *Social Research Methods*. Canadian Edition. Don Mills, ON: Oxford University Press, 2005.
10. *Гудков Л., Дубин Б., Левада Ю.* Проблема "элиты" в современной России: размышления над результатами социологического исследования. М.: Фонд "Либеральная миссия", 2007.
11. *Административные реформы в контексте властных отношений: опыт постсоциалистических трансформаций в сравнительной перспективе / Под ред. Олейника А., Гаман-Голутвиной О.* М.: Росспэн, 2008.
12. *Hogenraad R., Mckenzie D., Peladeau N.* Force and influence in contentanalysis: the production of new social knowledge // *Quality & Quantity*. 2003. N 37. P. 224.
13. *Roberts C.* A conceptual framework for quantitative text analysis // *Quality & Quantity*. 2000. N 34. P. 262.
14. *Ермаков С., Ким И., Михайлова Т., Осетрова Е., Суховольский С.* Власть в русской языковой и этнической картине мира. М.: Знак, 2004.
15. *Gray J., Densten I.* Integrating quantitative and qualitative analysis using latent and manifest variables // *Quality & Quantity*. 1998. N 32. P. 423.
16. *Бахтин М.* Проблемы поэтики Достоевского. М.: Советская Россия, 1979. С. 6 - 7.
17. *Lincoln W. B.* *The Great Reforms: Autocracy, Bureaucracy and the Politics of Change in Imperial Russia*. DeKalb, IL: Northern Illinois University Press, 1990; *Цвайнерт Й.* Роль государства в "догоняющей модернизации": реформы после 1789 года в Германии и "Великие реформы" в России в сравнительной перспективе // *Административные реформы в контексте властных отношений*. Указ. раб.
18. *Giddens A.* *Constitution of Society: Outline of the Theory of Structuration*. Cambridge: Polity Press, 1984.
19. *Олейник А.* Потенциальные и реальные ограничения своекорыстного поведения государственных служащих в России // *Административные реформы в контексте властных отношений*. Указ. раб.
20. *Rose R., Mishler W., Haerpfer C.* Getting real: social capital in postcommunist societies // *Studies in public policy*. 1997. N 278. P. 9 - 10.
21. *Granovetter M., Swedberg R.* Introduction // *The Sociology of Economic Life / edited by Granovetter M., Swedberg R.* Boulder, CO: Westview Press, 1992. P. 9.
22. *Ледяев В.* Власть, авторитет и господство в России: основные характеристики и формы // *Административные реформы в контексте властных отношений*. Указ. раб.
23. *Менар К.* Теория организаций: разнообразие соглашений в развитой рыночной экономике // *Институциональная экономика / Под ред. А. М. Олейника*. М.: Инфра М, 2005.