Нижегородский государственный университет им. Н.И.Лобачевского
Национальный исследовательский университет

Учебно-научный и инновационный комплекс
«Социально-гуманитарная сфера и высокие технологии:
теория и практика взаимодействия»

Иудин А.А.
Рюмин А.М.

Контент-анализ текстов: компьютерные технологии

(Учебное пособие)

Мероприятие 1.2. Совершенствование образовательных технологий,
укрепление материально-технической базы учебного процесса

Учебная дисциплина: Методы анализа документов в социологии

Специальности, направления: Социология – 040201,
Социальная работа – 040101

Нижний Новгород – 2010

УДК 316.77
ББК 60.56

Традиционные и компьютерные методы анализа документов в социологии. Учебное пособие. Нижний Новгород, ННГУ, 2010. 37 с.

Учебное пособие подготовлено в соответствии с государственным образовательным стандартом высшего профессионального образования. В нем изложены теоретические основы и практические рекомендации, описывающих работу с документами с использованием традиционных и новых методов анализа. Предназначено для студентов дневной и заочной формы обучения по специальности 040201 (Социология) и 040101 (Социальная работа).

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Самым распространенным видом информации является информация, представленная в виде текстов на языке данной страны, поэтому анализ текстов является одним из наиболее распространенных видов научного и научно-практического анализа. Более того, существуют науки, которые целиком или в основном описаются именно на анализ текстов. Наиболее распространенным направлением анализа текстов является сжатие информации – возможность выделить из совокупности текстов наиболее существенные, концептуальные моменты, важные для данного конкретного направления исследований. Традиционные формы сжатия информации – аннотирование, конспектирование, реферирование – уже давно не считаются какими-то специфическими видами работы с информацией и ими владеют любые специалисты.

Понятие анализ текстов иногда употребляется как синоним понятия контент-анализ, хотя последнее понятие шире. *Контент-анализ* относится к более широкой области исследований, затрагивающих не только текст, но информацию различного рода — изображения, аудио и видеоинформацию. Кроме того, контент-анализ, в отличие от других способов изучения документов, позволяет вписать содержание документа в социальный контекст, осмыслить его как проявление или как оценку социальной жизни. Понимание социального контекста документа предполагает выявление того, что именно получило в нем отражение, какой резонанс этот документ получил или может получить в общественной жизни и, наконец, степень оригинальности документа, отличие его от других документов такого рода.

Контент-анализ — это метод *количественного систематического* подхода к изучению текста. Важным является то, что он формализован. Формализованность, систематичность и строгость контент-анализа проявляется в том, что исследование проводится на основании методологически обоснованной программы, по определенным процедурам и служит для получения информации, отвечающей некоторым критериям качества.

С помощью контент-анализа изучались религиозная символика и популярные песни, устанавливались отличия эротических кинокартин от порнографических, устанавливалась мера эффективности политических слоганов, реклам и вражеской пропаганды, определялись особенности суицидального поведения, проявившиеся в предсмертных записках самоубийц, стереотипы сознания различных социальных групп, выявлялась направленность демонстрации людей определенной национальности на телеэкранах, идеологическая подоплека передовиц газет, отличия в трактовках одного и того же события в разных СМИ, исследовались многие другие темы.

В последние десятилетия данный социологический метод заимствовали и активно используют представители социогуманитарных наук, заинтересованные в установлении объективных признаков разнообразных человеческих коммуникаций. Сюда следует отнести юристов, историков, журналистов, языковедов, литературоведов, культурологов, политологов, психологов, экономистов, социальных работников. Среди множества профессиональных областей использования метода можно выделить прикладную лингвистику, историю, искусствоведение, антропологию, связи с общественностью, коммуникативистику, педагогику, криминологию, этнографию, нарратологию.

Разработка методов анализа текстов в настоящее время идет в четырех направлениях:

- определение соотношения и взаимодействия информационных методов с методами опроса и другими методами сбора данных при постановке исследуемых задач;
- разработка технических разновидностей методик анализа информации применительно к специфике текстовых источников в различных отраслевых социологиях;
- обогащение методов анализа информации методологическими и методическими принципами смежных наук с развитыми методами анализа различных видов источников;
- разработка специализированного программного обеспечения для проведения контент-анализа.

В данном учебном пособии описаны общие принципы работы с методом контент-анализа, представлена история его развития, теоретические и практические сведения о нём. Важной особенностью современного этапа генезиса как качественных, так и количественных методов работы является их компьютеризация. Повышая эффективность работы, скорость обработки данных, увеличивая точность анализа, позволяя затрачивать меньше усилий на механические этапы исследований, постоянно развиваясь и создавая ряд других важных возможностей для исследователя, такая тенденция ставит и ряд дополнительных актуальных задач. Среди них особо стоит выделить необходимость совершенствования навыков работы с компьютерным программным обеспечением, в силу чего в методическом пособии предоставлено описание особенностей обработки текстовых массивов на примере компьютерного пакета ЛЕКТА. Программа позволяет производить контент-анализ текстового материала, прослеживая основные эксплицитные сюжетные линии, идентифицировать латентные идеи, стереотипы и т.д. Её исключительно важной и оригинальной особенностью является не простой подсчёт частотности использования индикаторов, дающий сравнительно мало данных для анализа, а установление групп корреляций между ними, определяемой возможностью проведения факторного анализа инстру-

ментами пакета. Также в приложении к пособию приведены краткое описание функционала нескольких иностранных пакетов, предназначенных для аналитической работы с текстовыми массивами.

Сегодня специалисту-социологу необходимо знать теоретические основы контент-анализа, обладать навыками работы с описанным в пособии и аналогичным программным обеспечением, в силу широкой востребованности метода и очевидных перспектив расширения областей его использования и развития функционала.

Курс носит обязательный характер. Освоение курса требует знания программ университетского курса по дисциплинам «Методика и техника социологических исследований», «Статистика и теория вероятностей», «Информатика», «Статистические методы обработки экспериментальных данных», «Социальное моделирование и программирование». Курс предназначен для освоения студентами основных навыков анализа вербальной информации. Лекционные и практические занятия направлены на формирование у студентов целостного понимания анализа информационных потоков и освоения ими навыков контент-анализа. В результате изучения курса студент должен:

- знать основные этапы развития анализа документов и вклад различных исследовательских школ в развитие контент-анализа;
- изучить основные теоретические и методологические направления изучения документов в социологии;
- иметь представления о типах методов анализа документов и применяемом программном обеспечении;
- уметь на практике использовать изученные методы;
- провести от начала до конца один учебный проект.

В рамках курса проводится серия лабораторных работ. Она нацелена на выработку у студентов творческого подхода к решению конкретных задач и сознанного применения различных методов анализа. В ходе выполнения курсовой работы студент должен применить на практике все методы анализа, с которыми он был ознакомлен в ходе аудиторных занятий.

ЧАСТЬ I. ИСТОРИЧЕСКИЕ И МЕТОДОЛОГИЧЕСКИЕ ОСНОВАНИЯ КОНТЕНТ-АНАЛИЗА

1.1. Из истории метода

В советской социологической литературе происхождение контент-анализа связывалось с именами У. Томаса и Ф. Знанецкого, однако ныне многие отечественные исследователи отмечают, что он возник сто и более лет тому назад. Первый упоминаемый в литературе опыт использования метода, очень близкого к этому (прикладная цель которого выглядит очень узнаваемой) Г.Г. Почепцов[1] относит к XIII в., когда в Швеции был осуществлен анализ сборника из 90 церковных гимнов, прошедших государственную цензуру и приобретших большую популярность, но обвиненных в несоответствии религиозным догматам. Наличие или отсутствие такого соответствия и определялось подсчетом в текстах этих гимнов религиозных символов и сравнения их с другими религиозными текстами, в том числе тех, которые считались еретическими. Частота использования определённых заранее собранных слов и тем позволяла судить о том, насколько корректен текст с точки зрения официального учения церкви.

Важно отметить, что простой подсчёт частотности употребления какого-либо слова давал сравнительно мало материала для точного и глубокого анализа проблемы. Установление семантических связей между отдельными единицами контент-анализа позволяет получить более полную картину.

В конце XIX – начале XX вв. в США появились первые контент-аналитические исследования текстов массовой информации. Их мотивация выглядит удивительно знакомой: авторы задавались целью продемонстрировать прискорбное пожелтение тогдашней нью-йоркской прессы. На рубеже XIX и XX веков развитие средств массовой коммуникации, увеличение количества информационных каналов и потоков и, как следствие, их дезориентирующее влияние на человека потребовали метода систематизации материала, его обобщения. Сам термин контент-анализ (content-analysis) впервые был использован в США журналистами Д.Уипкинсом, А.Тенни, Д.Спиидом, Б.Мэттью. Принципы методики также были частично описаны французским журналистом Ж.Кайзером.

Контент-анализ как сформировавшийся метод исследований изучения массовых коммуникаций первоначально был количественно-ориентированным. Впервые он был использован Максом Вебером в 1910 году для анализа освещаемости прессой политических акций в Германии. Позднее, в 1937 году метод контент-анализа был использован в США в

1 Г.Г. Почепцов – доктор филологических наук, профессор, заведующий кафедрой информационной политики Украинской академии государственного управления, автор книг в области паблик рилейшнз, имиджелогии, теории коммуникации и семиотики.

исследовании инаугурационных речей американских президентов, в рамках которого были изучены наиболее общие категории, отражающие национальные, исторические, фундаментальные и оценочные аспекты.

Чтобы получить материалы для своей книги о судьбе польских крестьян, эмигрировавших в США, ***У. Томас и Ф. Знанецкий***[2] провели колоссальную работу по сбору личной документации. Одним из путей решения этой задачи была публикация, объявленная в газете с просьбой к полякам, приехавшим в США, присылать свои жизнеописания и письма родственников по определенному адресу за незначительную плату - 10 центов за материал.

Этот метод сбора материала, точнее методологические позиции авторов, были раскритикованы спустя почти 20 лет американским социологом ***Блумергом***. Он отметил, что эти материалы носили лишь иллюстративный характер и никоим образом не могли быть использованы в качестве доказательства конкретной точки зрения. После этого экспертная комиссия Национального совета по социальным исследованиям США создала специальный комитет, на котором анализировались проблемы, связанные со степенью искажения материала при передаче мысли и при записи. В этой связи ставился вопрос о том, в какой мере само оформление того или иного личного документа – заявления или дневника – соответствует реальным намерениям этого автора и действительному положению дел.

Опыт первой мировой войны сформировал большую группу серьезных исследователей в области пропаганды, и паблик рилейшнз. Тогда в США был создан комитет под руководством ***Джорджа Криля***, который занимался составлением пропагандистских материалов.

Во время второй мировой войны, в США и Великобритании контент-анализ использовался государственными структурами в военных целях и в целях исследования направлений пропагандистской деятельности. В это же время в Великобритании сотрудники радио ВВС анализировали пропагандистские материалы нацистов и составляли прогнозы по поводу ведения ими внешней и внутренней политики. Один из самых замечательных примеров использования контент-анализа принадлежит британским аналитикам, верно предсказавшим время запуска крылатых ракет «ФАУ-1» и баллистических ракет «ФАУ-2» Германией против Великобритании.

2 Томас Уильям Айзек (1863—1947), американский социолог. Профессор в Чикаго (1895—1918) и Нью-Йорке (1923—28). Сочинения: «Пол и общество» (1907) и др. Знанецкий Флориан Витольд (1882—1958), социолог. Родился в Польше. Основал Польский социологический институт (1921). С 1941 живет в США. Один из авторов теории социального действия. Основная работа обоих социологов — «Польский крестьянин в Европе и Америке» (1918—20).

В исследования пропаганды значительный вклад внес ***Гарольд Лассуэлл***.[3] В 1927 г. вышла его докторская диссертация под названием «Техники пропаганды в мировой войне». Эта книга была качественной (с точки зрения методологии), в ней оценивались техники пропаганды двух сторон военных действий. В частности он произвёл анализ содержания газеты «истинный американец» и привёл аргументированные доказательства того, что она поддерживает фашизм, после чего публикация газеты была запрещена. При этом Лассвелла критикуют за некорректное соотнесение качественных и количественных методов, не позволяющее провести верификацию результатов.

Г. Лассуэл сформулировал три основные функции коммуникации в обществе:

1. Наблюдение над окружающим миром: эта роль масс-медиа позволяет индивиду видеть гораздо больше, чтобы узнавать о событиях во всем мире.
2. Корреляция ответа общества на события в окружающем мире: масс-медиа рассказывает индивидууму как интерпретировать происходящие события.
3. Передача культурного наследия, например: дети изучают жизнь других людей, что такое хорошо и что такое плохо, чем они отличаются от других людей.

Широко известна формула Лассуэлла из пяти вопросов: «Кто и что говорит, по какому каналу, кому и с какими эффектами?», дающая простой и четкий формат описания коммуникации.

Накопленный опыт лёг в основу создания книги, написанной ***Б. Берелсоном*** в начале 1950-х годов XX века «контент-анализ в коммуникационных исследованиях». Она до сих пор считается фундаментальным трудом, описывающим наиболее общие положения этой молодой методики исследований. После её появления метод приобрёл большую популярность и стал широко использоваться и совершенствоваться в самых разных сферах. Так, например, появилась методика связанности символов ***Ч. Осгуда***[4], позволявшая определить коррелирующие между собой части содержания текста. Европейские исследователи опирались главным образом на опыт американских специалистов в области контент-анализа.

В начале 1960-х гг. Г. Лассуэлл осуществил попытку политологического анализа СМИ, исходя из учета формальных критериев. Он ввел в научный оборот абстрактную единицу – слово. Целью работы Лассуэлла было получение собственно социологического результата на нетипичном для социологии материале – текстах печатных изданий. Иссле-

3 Лассуэлл (Лассвелл) (Harold Dwight Lasswell, 1902-1978) – американский политолог, представитель бихевиористского подхода в политической науке и один из основателей чикагской школы социологии.

4 Чарльз Осгуд (Charles Osgood 1916 – 1991 гг.) — американский психолог, разработчик методики семантического дифференциала.

дователь проделал огромную работу, но, поскольку в методике Лассуэлла качественные оценки не были адекватно соотнесены с количественными, результаты его трудов с трудом поддавались верификации.

В этот же период ***Ж. Кайзер*** разработал новую методику статистического анализа периодических изданий, в основе которой лежал подход к тестовому массиву, как информационной системе. Тем самым Кайзер сформулировал теоретическую базу последующего распространения социологических методов в сферы изучения всех нарративных источников, включая эпиграфический и эпистолярный материал. В работе Ж. Кайзера акцентировалось внимание на внешней форме организации материала: его расположении, оглавлении, оформлении и т.д. Кайзер разработал целый комплекс исследовательских процедур, обеспечивающих полную формализацию, как единичного газетного номера, так и совокупности однотипных периодических изданий. Тем самым Ж.Кайзер сформулировал систему, позволяющую фиксировать развитие тенденций в публикациях СМИ.

Дальнейшее развитие кайзеровское направление методологии контент-анализа получило в работах ***Э. Морэн***, которая ввела в научный оборот термин *единица информации* – семантический блок, содержание которого отвечает на вопрос: «О чем говорится?» Последнее обстоятельство сделало возможным изучение любых форм организации текстового материала, причем, как на терминологическом уровне, так и на уровне фразы, абзаца, статьи и даже целых книг. Тем самым, Э. Морэн разрушила критерий однородности, применявшийся ранее при статистической обработке нарративов. Взамен, она предложила идеологию семантических групп, которые, по ее мнению, должны учитываться по тематическому признаку. Кроме того, Э. Морэн разработала концепцию тона материала, который определялся социометрически: положительная информация, отрицательная, нейтральная.

Следующим этапом в развитии метода в области проведении исследований, имеющих дело с большими объёмами текста, стало использование ресурсов ЭВМ. Так в 1974 году в Италии на конференции, посвящённой проблемам контент-анализа, было представлено несколько проектов, реализуемых посредством машинной обработки данных. Они заключались в анализе заголовков статей опубликованных в большом количестве газет и сравнении степени внимания в них к региональным, общегосударственным и международным проблемам; в сравнении интереса американских и европейских СМИ к тенденциям развития «общего рынка» и т.д. На этом собрании Германия выступала с проектом создания словаря, который мог бы быть использован в проведении контент-анализа текстов.

В СССР метод контент-анализа стал использоваться с конца 1960-х годов. Например, это исследования А.В. Баранова, направленные на изучение степени обращения к

субъективным интересам читателей в газете «Известия»; исследования Б.А. Грушина по изучению информированности читателей ряда СМИ о существующих проблемах.

Наиболее широкое распространение контент-анализ получил в теории массовой коммуникации, политологии и социологии. Этим отчасти объясняется тот факт, что иногда этот термин используется как обобщающий для всех методов систематического и претендующего на объективность анализа политических текстов и текстов, циркулирующих в каналах массовой коммуникации. Однако такое расширительное понимание контент-анализа неправомерно, поскольку существует ряд исследовательских методов – либо специально разработанных для анализа политических текстов (например, метод когнитивного картирования), либо применимых и применяемых для этой цели (например, метод семантического дифференциала или различные подходы, предполагающие изучение структуры текста и механизмов его воздействия), – которые не могут быть сведены к стандартному контент-анализу даже при максимально широком его понимании.

Становление и распространение приобретающей всё большую популярность всемирной глобальной сети Интернет также дало много ресурсов для развития метода контент-анализа и усилило потребность в нём. Сегодня контент-анализ широко используется именно в этой коммуникативной среде, и развитие программного обеспечения, способного в той или иной мере автоматизировать процесс проведения метода, остаётся приоритетной задачей для расширения диапазона его применения. Его разработкой занимаются главным образом в США, Великобритании и Германии. К сожалению, уникальные отечественные разработки программного обеспечения практически неизвестны зарубежным исследователям, использующим метод контент-анализа. Подробнее о зарубежных компьютерных пакетах, применяемых для проведения контент-анализа можно узнать из Приложения №1.

1.2. Назначение, область применения и особенности контент-анализа

Виды документов. Прежде чем приступить к рассмотрению способов анализа документов, необходимо дать определение и классификацию документов. *Документом в социологии называется специально созданный человеком предмет, предназначенный для передачи или хранения информации.*

По способу фиксирования информации различают *рукописные* и *печатные документы; записи на кино-* или *фотопленке, на магнитной ленте*. Сегодня, в связи с широким распространением и универсализацией электронных средств хранения, передачи и обработки информации, решающее значение приобретает классификация документов на цифровые (машинно-читаемые) и аналоговые (не читаемые с помощью компьютера).

С точки зрения целевого назначения, выделяют *материалы, которые были провоцированы самим исследователем* (к примеру биография эмигранта в работе Томаса и Знанецкого — в числе использованных документов была уникальная автобиография одного из крестьян, написанная по просьбе исследователей и составившая около 300 страниц). Эти документы называют *целевыми*. Но социолог имеет дело и с материалами, составленными независимо от него, ради каких-то других целей, т.е. с *наличными документами*. Обычно эти материалы называют собственно документальной информацией в социологическом исследовании.

По степени персонификации документы делятся на *личные* и *безличные*. К личным относят карточки индивидуального учета (например, библиотечные формуляры или анкеты и бланки, заверенные подписью), характеристики и рекомендательные письма, выданные данному лицу, письма, дневники, мемуарные записи. Безличные документы — это статистические или событийные архивы, данные прессы, протоколы собраний.

В зависимости от статуса документального источника выделим документы *официальные* и *неофициальные*. К первым относятся правительственные материалы, постановления, заявления, коммюнике, стенограммы официальных заседаний, деловая корреспонденция, протоколы судебных органов и прокуратуры, финансовая отчетность. Неофициальные документы — это многие личные материалы, упомянутые выше, а также составленные частными гражданами безличные документы (например, статистические обобщения, выполненные другими исследователями на основе собственных наблюдений). Особую группу документов образуют многочисленные материалы средств массовой информации: газет, журналов, радио, телевидения, кино, видеоматериалы.

По источнику информации документы разделяют на *первичные* и *вторичные*. Первичные составляются на базе прямого наблюдения или опроса, на основе непосредственной регистрации совершающихся событий. Вторичные представляют обработку, обобщение или описание, сделанное на основе данных первичных источников.

Помимо этого, можно, конечно классифицировать документы по их прямому содержанию, например литературные данные, исторические и научные архивы, архивы социологических исследований, видеохроники общественных событий.

Определения контент-анализа. Существует множество определений контент-анализа, но при этом большинство из них едва ли полно отражает его сущность. Приведём наиболее часто употребляемые определения контент-анализа.

- Это методика объективного качественного и систематического изучения содержания средств коммуникации (Д. Джери, Дж. Джери).

- Это систематическая числовая обработка, оценка и интерпретация формы и содержания информационного источника (Д. Мангейм, Р. Рич).
- Это качественно-количественный метод изучения документов, который характеризуется объективностью выводов и строгостью процедуры и состоит в квантификационной обработке текста с дальнейшей интерпретацией результатов (В. Иванов).
- Это исследовательская техника для получения результатов путем анализа содержания текста о состоянии и свойствах социальной действительности (Е. Таршис).
- Контент-анализ состоит в нахождении в тексте определенных содержательных понятий (единиц анализа), выявлении частоты их встречаемости и соотношения с содержанием всего документа (Б. Краснов).

Наиболее компактное формальное определение контент-анализа звучит так: «Любая систематическая редукция потока текста (или других символов) к стандартном набору статистически обрабатываемых символов, отражающих присутствие, интенсивность или частоту характеристик, значимых для социальной науки».[5]

Эти определения дают фрагментарное представление о методе и не учитывают новых возможностей многомерного статистического анализа. Все эти определения могут быть сгруппированы следующим образом:

- статистическая семантика;
- техника для объективного количественного анализа содержания коммуникации;
- техника для разработки обобщений при помощи объективного и систематического установления характеристик сообщений.

Как нам представляется, одним из наиболее адекватных является определение контент-анализа, разработанное психологами. КОНТЕНТ-АНАЛИЗ (англ. content – содержание) – метод выявления и оценки специфических характеристик текстов и других носителей информации (видеозаписей, теле- и радиопередач, интервью, ответов на открытые вопросы и т.д.), при котором в соответствии с целями исследования выделяются определенные смысловые единицы содержания и формы информации. Затем производится систематический замер частоты и объема упоминаний этих единиц в определенной совокупности текстов или другой информации. Контент-анализ дает возможность выявлять отдельные психологические характеристики коммуникатора, аудитории, сообщения и их взаимосвязи. В отличие от элементарного содержательного анализа, контент-анализ, как научный метод, используется для получения информации, отвечающей некоторым критериям качества (объективность, надежность и валидность). Заметную роль в повышении

5 Определение дано Шапиро и Маркоффом (Shapiro and Markoff) в 1997; цит. по: Popping, Roel. Computer-assisted Text Analysis. SAGE Publications: London, 2000, p. 7.

качества контент-анализа играет возможность использования методов многомерного статистического анализа данных. Особенно широко используется факторный анализ, способствующий выявлению скрытых факторов, определяющих содержание текстов. Такое определение несколько громоздко и, по сути, представляет собой описание исследовательской техники, тем не менее оно позволит нам отойти от представлений о контент-анализе как простом пересчете слов в текстах.

Специфика метода. Специфика анализа текстов как метода раскрывается через пары понятий, описывающих основные контрасты метода. Дедукция или индукция: от общего к частному или от частностей к общим закономерностям? Количественная или качественная стратегия: количественная стратегия предполагает более формальный подход и применение статистических методов, тогда как качественная опирается на способность человека понимать и интерпретировать смыслы.

Денотация и коннотация: денотация и коннотация связаны с социальным значением слов, а не с грамматическими правилами их употребления. Денотация – это фиксированное отношение слова к объектам, которые оно описывает (одно слово может иметь несколько денотаций, например, в языке разных социальных групп). Коннотация – это контекстно-зависимое значение слова или ценностная (оценочная) нагрузка. Примеры: окно – отверстие в стене, форточка – часть окна, но также окно компьютерной программы, окна, форточки – сленговое обозначение ОС Windows (вторичная денотация). Слово социализм или демократия будет иметь разные коннотации для молодого и старшего поколения, как в смысле социального опыта, так и в смысле оценки.

Открытое или скрытое значение, смысл слов – описание и интерпретация: различение открытого, непосредственно выраженного смысла и скрытого смысла, значения. Данное противопоставление напоминает различие открытого и скрытого смысла слов, но относится к текстам в целом. Описательный, дескриптивный анализ концентрируется на таких вопросах, как насколько часто и каким образом слово встречается в тексте, тогда как интерпретационный анализ задается вопросами значения слова и причин его употребления в том или ином контексте.

Область применения. Важной особенностью этого метода является систематизация большого по объёму тематически связанного, но не структурированного массива (чаще всего текстового). Предварительная систематизация такого материала позволяет сократить время на его обработку. В этой связи существенную важность имеет грамотный подбор источников получения информации, например печатных и электронных СМИ нужных тематических групп, ориентированных на определённые целевые аудитории. Роль

и функции их варьируются в зависимости от особенностей освещения экономико-политических аспектов общественной жизни, идеологической, религиозной и многих других её составляющих, по типу социализирующей и образовательной деятельности, по методам воздействия на целевую аудиторию, по степени объективности публикуемой информации и т.д. Для исследователя крайне важно чётко идентифицировать позиции медиа при отборе массива, в противном случае противоречивость, размытость результатов могут помешать в полной мере, убедительно подтвердить либо опровергнуть гипотезы исследования.

Сферы социологических исследований коммуникации, в которых может применяться анализ текстов:

- Анализ содержания коммуникации;
- Анализ формы коммуникации;
- Анализ производителей коммуникации;
- Анализ аудитории;
- Анализ эффектов коммуникации.

Три типа гипотез, которые могут быть протестированы с помощью анализа текстов:

1. гипотезы относительно частоты встречаемости тех или иных терминов, понятий;
2. гипотезы о связи понятий в тексте, отдельных частях текста или совокупностях текстов;
3. гипотезы, касающиеся соотношения между текстуально-аналитическим исследованием и другими видами исследований; гипотезы такого типа используются для сравнения результатов исследований, проведенных с помощью различных методов или для установления связей между текстуальными и не-текстуальными явлениями (например, для сравнения высказываний и реальных действий людей).

Ограничения анализа текстов как метода:

- для количественного анализа необходимо статистически значимое количество текстуальной информации, он не предназначен для анализа уникальных текстов;
- анализируемые тексты должны поддаваться формализации, поэтому данный метод лишь ограниченно пригоден для анализа художественной литературы и совсем не пригоден для анализа поэзии;
- качественный анализ позволяет глубже понять текст, но он требует значительного количества времени и усилий; таким образом, традиционный качественный анализ малопригоден для исследования больших объемов текста. Последнее ограничение ныне снимается посредством создания программных средств, осуществляющих лексический анализ текстов. В последние годы предпринимаются по-

пытки и семантического машинного анализа вербальной информации;

- главным ограничением является то обстоятельство, что текст менее сложен, чем индивидуальное или общественное сознание, которыми он порожден; текст является упрощенным, редуцированным отражением социальной реальности.

Метод занимает особое место среди других в силу своей эффективности при анализе больших информационных массивов. Чаще всего он используется при анализе текста и заключается либо в подсчёте наиболее часто встречающихся в нём слов, словосочетаний, самостоятельных тем, выраженных, например, целостными абзацами, и других лексических единиц, либо единицами контент-анализа выступают такие величины как протяжённость текста, численность строк, абзацев, колонок, страниц. Метод также применяется и при изучении видео и аудио материала и единицами анализа становятся графическая составляющая, сопровождающая тексты, метраж аудио и видео плёнки с материалами, интересующими исследователя, объём эфирного времени, время суток, в которое материал транслируется аудитории. С помощью этого метода можно изучать такие материалы как, например, статьи в СМИ, речи политиков, партийные программы, программы общественных движений, видеоматериалы массовых мероприятий, съездов и митингов, нормативно-правовые акты, рекламные сообщения, произведения художественной литературы, исторические тексты, письма и многое другое. Обязательным условием проведения контент-анализа является фиксация материала на материальном носителе. Только при его соблюдении возможно использование этого метода.

Часто результаты контент-анализа дополняются использованием других методов. Интересен он также и тем, что не требует больших материальных затрат, несложен в использовании, не подразумевает ощутимых технических и других трудностей при использовании специализированного компьютерного программного обеспечения. Полевой этап исследования более прост, чем при использовании многих других методов. Так, проведение простого (хотя и неглубокого) контент-анализа доступно даже при использовании базовых средств Microsoft Office или его аналогов.

ЧАСТЬ II. МЕТОДОЛОГИЯ, МЕТОДИКА И ТЕХНИКА КОНТЕНТ-АНАЛИЗА

2.1. Основные методологические категории метода

Контент-анализ как метод предоставляет исследователю богатые и разнообразные возможности, но требует тщательного формирования исследовательской стратегии путем

выбора из нескольких альтернатив. Рассмотрим эти альтернативы.

Основа контент-анализа – это подсчет встречаемости некоторых компонентов в анализируемом информационном массиве, дополняемый выявлением статистических взаимосвязей и анализом структурных связей между ними, а также снабжением их теми или иными количественными или качественными характеристиками. Отсюда понятно, что главная предпосылка контент-анализа – это выяснение того, что считать; иными словами, определение единиц текста.

Единицы текста. *Единица – это отдельная группа слов, рассматриваемая как целое.* Выделяется несколько типов единиц.

Единицы анализа – это единицы, составляющие основу анализа, единицы, которые исследователь стремится охарактеризовать. Пример: слово, газетная статья.

Единицы выборки – части наблюдаемой реальности или потока текста, которые рассматриваются как независимые друг от друга. Они имеют ясно различимые границы, им могут быть присвоены уникальные номера и они могут включаться в выборку с заранее известной вероятностью.

Единицы кодирования (также единицы записи или единицы текста) – это отдельные сегменты текста, помещаемые в ту или иную категорию. Для каждой единицы кодирования исследователь принимает решение, имеет ли она те или иные атрибуты, которые интересуют его в данном исследовании, относятся ли они к теме исследования. Пример: идея превосходства мужчин над женщинами (идея, формирующая категорию) может быть выражена в таких единицах кодирования, как слово, смысл слова, предложение, тема, абзац, текст целиком.

Единицы контекста – это та совокупность текстов, которую необходимо принять в расчет, характеризуя единицу кодирования. Они формируют контекст, который определяет значение, смысл единиц кодирования, в том случае, если этот смысл контекстно-зависим. Например, в статье, посвященной финансовым вопросам, слово долг будет иметь другое значение, чем в тексте, посвященном религиозным вопросам. При анализе текстов без применения компьютера контекст обычно легко распознаваем. В компьютерном анализе контекст, как правило, определяется через анализ слов, окружающих в тексте единицу кодирования.

Единицы счета – это те единицы, с помощью которых квантифицируются атрибуты текста. Они совпадают с единицами кодирования, если исследователь заинтересован в подсчете количества слов или других элементов текста. Другими словами, единицы счета – это именно то, что подсчитывается в процессе исследования, то, к чему относятся числа в матрице данных. Примеры: 5 слов были идентифицированы как относящиеся к агрессии

(попадающие в данную категорию). В матрицу ставится число 5 – в данном случае единица кодирования совпадает с единицей счета. Пример несовпадения этих единиц: анализ пространства на страницах газеты, отданного под освещение определенной темы. Статья, идентифицированная как относящаяся к теме – это единица кодирования, а число квадратных сантиметров (в которых измерена площадь статьи и полученный результат занесен в матрицу) – единица счета.

Физические единицы имеют отдельную физическую форму (например, отдельный номер газеты).

Синтаксические единицы – те, которые являются естественными для грамматики соответствующего средства коммуникации (например, слово во фразе или отдельная новость во фразе выпуска новостей). *Единицы референции* – те, которые описывают разными словами один и тот же объект (например, «глава государства», «президент», «Путин», в определенном контексте – просто «он»). *Пропозиционные единицы* – это части сложных предложений, имеющие собственную структуру, описания конкретных положений дел (ситуаций). Такие единицы используются для того, чтобы избежать сложности естественного языка. Например, фраза «Агрессивный вор угрожает полицейскому» распадается на два простых предложения «Вор агрессивен» и «Вор угрожает полицейскому».

Единицы различного рода могут пересекаться и включать друг друга. Например, при анализе книг первая единица анализа – это книга, вторая – главы в книгах, третья – параграфы или абзацы. В случае если параграф – наименьшая из единиц, на которые исследователь разбил текст, он также служит и единицей кодирования. Однако можно продолжить делить тест дальше вплоть до предложений или грамматических частей предложений. В таком случае единицей выборки может стать абзац. Каждая единица, которая больше, чем составляющие ее части, может служить единицей контекста: фраза для слова, глава для параграфа и т.д.

Концептуальные категории. *Концепт* – это единица смысла, отдельная идея. Концептуальные категории – это агрегации единиц текста, основанные на общей идее, релевантной для теоретической основы исследования. Иными словами, категории – результат операционализации идей с помощью слов и фраз. Концепты могут быть образованы дедуктивно (на основе теории) или индуктивно (на основе исследуемых текстов).

Количественный или качественный подход. Количественный контент-анализ в первую очередь интересуется частотой появления в тексте определенных характеристик (переменных) содержания. Качественный контент-анализ позволяет делать выводы даже на основе единственного присутствия или отсутствия определенной характеристики содержа-

ния.

Различие двух подходов довольно легко проиллюстрировать примерами. В 1950-е годы западные аналитики на основе количественного анализа статей газеты "Правда" обнаружили резкое снижение числа ссылок на Сталина. Отсюда они сделали вывод, что последователи Сталина стремятся дистанцироваться от него. С другой стороны, качественный аналитик мог бы сделать аналогичный вывод на основе единственного факта, что в публичной речи одного из партийных функционеров, посвященной победе СССР в Великой Отечественной войне, Сталин вообще не был упомянут. Прежде такое было бы немыслимо.

Качественный контент-анализ не слишком высоко оценивается позитивистски ориентированными исследователями. На западе ему отдают предпочтение исследователи, придерживающиеся феминистских, а также критических или интерпретативных подходов. Сторонники количественного подхода также иногда включают качественный контент-анализ в свой методологический арсенал с целью усилить надежность количественных исследований в исследовании содержания текста. В дискуссии о качественном или количественном контент-анализе существует и объединяющая точка зрения, которая представляется наиболее продуктивной. Ее защитники[6] утверждают, что должно использоваться некоторое сочетание количественного и качественного анализа.

Тематический или семантический подход. Основной отличительной особенностью *тематического* подхода является то, что выводы делаются на основании данных о встречаемости концептов, понятий в тексте. Основное допущение подхода состоит в том, что существует связь между наличием в тексте тех или иных тем и интересом к этим темам у автора текста. Сделав это допущение, исследователь может ставить вопрос об уровне интереса, проявляемого тем или иным коммуникатором к той или иной теме. Можно также сравнивать интерес к теме у разных коммуникаторов или в различные периоды времени. Применимость тематического анализа определяется следующими основными критериями:

1. можно сформулировать четкие и однозначные правила кодирования;
2. единицей анализа выступает слово или устойчивое словосочетание, и выводы будут делаться на основе частот встречаемости слов или словосочетаний в тексте;
3. в анализ включено большое количество концептов.

Основное различие между тематической и *семантической* стратегиями – в ответе на вопрос «что считать?». В тематической стратегии кодируются концепты, семантическая предполагает подсчет и кодирование отношений между концептами. Преимуществом семантического анализа по сравнению с тематическим является то, что он позволяет сохранить больше изначальной сложности, больше оригинальных смыслов исследуемого

6 Berg, B.L. Qualitative research methods for the social sciences / Bruce L. Berg. Boston, 1995.

текста.

Семантика [<гр. semanticos обозначающий] — смысловая сторона языка, отдельных слов и частей слова. *Семантический* [<гр. semanticos обозначающий] — смысловой, относящийся к значению слова (в отличие от звуковой и формальной стороны слова).

Семантический анализ в значительно большей мере, чем тематический, опирается на свойства естественного языка. Поэтому анализ (особенно компьютерный) в значительной степени специфичен для каждого языка. Каждый блок текста, попавший в выборку, суммируется при помощи нескольких кодов, взаимосвязанных в соответствии с общим пониманием предмета изучения.

Рассмотрим пример семантической грамматики, состоящей из четырех компонентов:

- Агент (agency) – инициатор какого-либо действия, активности;
- Позиция (position) – позиция относительно действий агента, их оценка;
- Действие (action) – анализируемая деятельность;
- Объект (object) – цель, предмет деятельности агента.

Например, фрагмент текста, имеющий смысл «политики слишком много спорят», будет закодирован следующим образом: политики (агент) не должны (позиция) спорить (действие) с политиками (объект).

Методологические основания кодирования текста. *Кодирование* – это процесс систематической трансформации и агрегации исходных данных в категории, которые позволяют точно описать характеристики текста, релевантные для исследования. Основная проблема кодирования – неоднозначность текста. Слова могут иметь различные коннотации. Способ избежать проблем неоднозначности – выработка детальных правил кодирования и принятие мер к избежанию неоднозначности, основанной на контексте. Кодирование может производиться как вручную, так и с использованием компьютера. Каждый способ имеет свои преимущества и недостатки. Люди более способны к совершению осмысленного выбора из нескольких значений или вариантов смысла, но они работают медленно и иногда склонны присваивать словам значения, не предусмотренные исследователем. Компьютерное кодирование значительно быстрее, но хуже справляется с неоднозначностью текста.

В рамках инструментального подхода набор категорий создается на основе теории, которая есть у исследователя еще до начала исследования. В таком случае, целью исследования может быть подтверждение или опровержение теории. Важно, что в данном случае значения, которые не релевантны по отношению к теории, а также варианты значений, возмож-

но подразумеваемые автором текста, не принимаются во внимание.

Выделяется два базовых подхода к кодированию: кодирование открытого, эксплицитно выраженного содержания текста (манифестное или инструментальное) и кодирование скрытых, неявно выраженных смыслов, которые, по мнению исследователя, подразумевались создателем текста (латентное, репрезентационное).

Манифестное кодирование. Кодирование содержания текста, лежащего на поверхности, называется манифестным (открытым, явным). Исследователь подсчитывает количество появлений фразы или слова (например, красный) в письменном тексте; или количество определенных действий (например, поцелуев или ударов), представленных на фотографиях или в видеосцене.

Система кодирования включает перечень терминов или действий, помещенных в тексте. Именно таким перечнем является, например, словарь. Исследователь может использовать компьютерную программу (например, Lekta) для поиска слов или фраз и, таким образом, переложить на компьютер всю работу по кодированию. Для этого исследователю нужно изучить компьютерную программу, составить список соответствующих слов или фраз и затем представить текст в форме, которая может быть пригодна для компьютера.

Манифестное кодирование весьма надежно, поскольку фраза или слово может либо наличествовать, либо отсутствовать. К сожалению, манифестное кодирование не принимает в расчет коннотации слов или фраз. Одно и то же слово может иметь различные значения в зависимости от контекста. Возможность каждого слова выступать во множестве значений ограничивает валидность манифестного кодирования.

Латентное кодирование. Исследователь, который использует латентное (скрытое) кодирование, ищет скрытые, имплицитные значения содержания текста. Например, он прочитывает весь абзац целиком и решает, присутствует ли в его содержании эротика или же это романтический жанр. Применяемая им система кодирования должна следовать общим правилам, устанавливающим принципы интерпретации текста и определяющим, имеют ли место те или иные темы или жанры.

По сравнению с манифестным латентное кодирование менее надежно. Оно зависит от степени владения исследователем языком и общепринятыми значениями. Здесь очень важной становится позиция автора по отношению к исследуемой проблеме. Повысить надежность могут тренинг, практика и описание правил, но остаются трудности, связанные с правомерностью идентификации тем, жанров и т.д. Латентное кодирование может быть валиднее манифестного, поскольку люди передают значение множеством неявных способов, прежде всего, в зависимости от контекста, а не от тех или иных слов.

С технической точки зрения, исследователь может разработать систему кодов для

латентных смыслов, содержащихся в тексте. Затем в конце каждого фрагмента после текста вставляются кодовые обозначения. Например, крайне негативный, критический настрой автора текста, отраженный в отдельном фрагменте, может быть обозначен кодом «НГТ». Степень негативности может быть передана неоднократностью кода НГТ. Коды затем учитываются при обработке текста как отдельные фильтры.

Исследователь может использовать и манифестное, и латентное кодирование. Если в использовании двух подходов нет противоречий, результат будет сильнее; если же манифестное и латентное кодирование используются несогласованно, исследователю может потребоваться перепроверить операциональные и теоретические дефиниции.

Итогом проведения контент-анализа как самостоятельного или дополняющего метода является аналитическая записка, интерпретирующая полученный материал, классифицированный по семантическим группам. Она включает в себя либо только общие выводы, вытекающие из проведённого анализа, либо также подробную информацию, описывающую изучаемые массивы.

Наряду с очевидным достоинством метода контент-анализа – возможностью его применения при анализе больших массивов неструктурированного текста – существуют и недостатки. Чаще всего выделяют высокую вероятность субъективной интерпретации полученных данных, субъективизм при отборе исходных данных, категорий. Обычно проблема субъективного трактования полученных данных решается коллективным участием в работе нескольких исследователей.

В процессе работы над контент-анализом может вызвать сомнение выбор объёма массива, достаточного для определения зависимости между категориями и индикаторами, транслирующими семантическую нагруженность конкретной части массива. Следует помнить, что контент-анализ неразумно и неэффективно использовать при изучении уникальных несхожих между собой документов, для исследования которых необходимо получить всестороннее и полное их описание, не допуская игнорирования тех или иных уникальных, а потому не встречающихся постоянно в массиве элементов. Это также справедливо и для анализа весьма сложных документов и документов, в которых явно недостаточно материала для проведения контент-анализа, результаты которого не будут репрезентативны.

2.2. Организация исследования

Отбор источников и построение выборки. Как уже было сказано ранее, этап определения фильтров поиска полезных в исследовании источников информации и отсеи-

вание ненужных, несвязанных с темой, либо не имеющих весомых связей с ней, является очень важным и не должен рассматриваться как формальный, механический и малозначимый этап работы. Ошибки при выборе типологических групп источников могут отрицательно сказаться на всёх этапах последующего исследования, включая получение необъективных его результатов. В связи с этим прежде всего, необходимо определить круг источников потенциально полезных для исследования, содержащих в себе материалы по заданной теме. Далее важно установить дополнительные рамки отбора материала: определить тип источника (телевидение, пресса, рекламные материалы, радио и др.). Затем нужно определить вид сообщения (публицистические статьи в электронном либо в печатном СМИ, информационные заметки, рекламные плакаты) роль участника коммуникации (отправитель или получатель сообщения). Определяются минимальные и максимальные границы объёма сообщений, их протяжённости, частота, время, место и средство трансляции сообщений целевой аудитории. Существуют и другие критерии отбора сообщений, и их количество и выбор варьируется в зависимости от поставленных задач исследования.

Далее следует этап определения объёма выборочной совокупности. В случае ограниченного количества материала по заданной теме, выборочная совокупность может быть эквивалентна генеральной. Это актуально, например, при предварительном проведении интервьюирования на заданную узкую тему, при котором весь массив текстов будет использоваться для анализа. Классическая трактовка метода контент-анализа подразумевает возможность сокращения выборочной совокупности сообщений при их схожести и однородности в соответствии с вышеописанными критериями. Такая редукция допустима, если объём генеральной совокупности очень велик. Выборка при исследовании больших совокупностей данных случайная и производится так же в соответствии с заданными вышеуказанными критериями. Безусловно, необходимо рассчитать её объём так, чтобы она оставалась репрезентативной, важно определить допустимую ошибку выборки. Техническое задание исследования должно содержать такие критерии сбора материала, регламентируя этот процесс и не давая нежелательному (ненужному или деструктивному для исследования) тексту проникнуть в массив. Для определения конгруэнтности особенностей преподнесения информации в СМИ, выбранных исследователем для проведения контент-анализа, и конкретных сообщений, потенциально попадающих в выборку, может быть произведён эксперимент. Его целью будет определение степени точности эмпирической интерпретации категорий исследования. Стоит добавить, что зачастую объём выборочной совокупности объясняется исследователями, исходя из понятий здравого смысла, доступности материала, скорости анализа материала в сжатые сроки, а не расчетом допустимой ошибки выборки, не достаточностью массива для сохранения его репрезентативности.

Для того чтобы применение контент-анализа было успешным, источник должен отвечать определенным требованиям. При выборе источника, прежде всего, нужно определить, в какой мере его содержание соответствует поставленной задаче. Необходимо также изучить все существующие источники по данной проблеме и, если понадобиться, выявить оптимальный размер репрезентативной случайной выборки. При построении выборки необходимо учитывать уровень исследуемой проблемы, цели и задачи исследования, ресурсоемкость (затраты труда, времени и средств) построения выборки и последующего проведения исследования на ней.

Пример построения выборки. Например, перед исследователем стоит задача узнать, как изображаются женщины и представители меньшинств в американских еженедельных журналах. В качестве единицы анализа избирается статья. Генеральная совокупность (популяция) включает все статьи, опубликованные в Time, Newsweek, U.S. News & World Report между 1969 и 1989 гг. Сначала нужно проверить, издавались ли названные журналы в указанные годы и определить, что понимается под статьей. Например, являются ли статьями обзоры кинофильмов? Можно ли определить минимальный размер текста (например, текст, состоящий из двух предложений), позволяющий квалифицировать его как статью? Если статья состоит из нескольких частей (и печатается в нескольких номерах), следует ли рассматривать эти части как отдельные статьи, или же как одну? Исследование указанных трех журналов показывает, что в среднем каждый номер содержит 45 статей. В год издавалось 52 еженедельных номера. Учитывая 20 лет определенных временных рамок, генеральная совокупность включает приблизительно 140000 статей (3 х 45 х 52 х 20 = 140400). Рамочные параметры для выборки задаются перечнем всех этих статей. Затем нужно принять решение об объеме и виде выборки. Допустим, что исходя из размеров бюджета и времени выборка ограничивается 1400 статьями. Таким образом, пропорция выборки составляет 1%. Необходимо также избрать вид выборки. Систематизированная выборка не подходит, поскольку журнальные издания выходят в свет циклично (интервал между выходом каждого из 52 номеров на протяжении каждого года – всегда неделя). Все номера важны для исследования, поэтому используется стратифицированная выборка. Стратификация проводится по журналам: 1400 / 3 = 467. Выборка стратифицируется также и по годам. Результат - примерно 23 статьи из каждого журнала за год. Наконец, составляется случайная выборка с использованием таблицы случайных чисел, чтобы отобрать 23 номера для 23 выбранных статей из каждого журнала за каждый год.

Организационные моменты проведения контент-анализа на этапе сбора информации. Традиционные методы работы с текстами подразумевают, что организаторы

работы весьма тщательно должны работать с бригадой кодировщиков, ибо именно они выполняют весь объем технической работы от которой зависит качество всего исследования. Контент-анализ часто включает кодирование очень большого круга источников информации. Исследовательский проект может потребовать просмотра содержания нескольких десятков книг, сотен часов телевизионных передач или тысяч газетных статей. В этой работе часто используется помощь ассистентов, предварительно обученных правилам ведения записей и методике кодирования. Кодировщики должны владеть моделями системы кодирования и консультироваться по всем вопросам неоднозначного характера. Исследователь обязан фиксировать все решения, которые он принимает относительно того, как трактовать возникшую при кодировании новую ситуацию.

1. Определение круга кодировщиков. В соответствии с поставленными проблемами, сроками, ресурсами рассчитывается число работников, привлекаемых к кодированию, обработке материала. Содержание в штате исследовательской лаборатории постоянных кодировщиков малоэффективно. Целесообразнее привлекать временных исполнителей. Это могут быть работники библиографических отделов, в случае анализа статей, ведь именно они разносят карточки по различным направлениям, работники службы отдела кадров, если ведется анализ аттестационных характеристик, работники канцелярий и отделов по связям с общественностью.

2. Организация кодирования. Инструкция по кодированию должна быть написана языком, доступным кодировщикам. Инструктаж должен быть проведен как в устной, так и, что особенно важно, в письменной форме. Это позволит устранить элементы недопонимания между методологами, аналитиками и исполнителями. Важны и моральные стимулы: технические работники были в курсе целей, задач, важность и назначение исследования. Поэтому в ходе беседы перед кодировщиками должны быть раскрыты основные задачи и гипотезы исследования. Необходимо наладить тесный контакт с кодировщиками и разработчиками инструментария по вопросам толкования различных моментов текста.

3. Мотивация труда кодировщиков. Здесь особенно важны материальные стимулы, их достаточность и справедливость распределения средств. В связи с этим могут использоваться различные варианты оплаты: по затратам времени; сдельная, за каждую найденную статью; пропорции к обработанному материалу. Могут быть введены поощрения в виде премий за скорость обработки информации; качество обработки; творческий подход к делу.

4. Контроль результатов кодирования. Необходимо осуществление выборочного контроля. Из бланков первичного кодирования исследователь отбирает несколько бланков, отличающихся от среднестатистического объема заполнения, например тем, что:

а) в клеточках преобладают нули; б) все клеточки заполнены.

Как правило, наименования, газеты, даты выхода и наименования статьи заносятся в бланк полностью, что значительно облегчает процедуру контроля.

Исследователь, который пользуется помощью нескольких кодировщиков, должен всегда проверять однозначность кодирования. Для этого он просит кодировщика закодировать текст самостоятельно и затем сопоставить полученные результаты с уже имеющимися по всему тексту. Таким образом замеряется надежность воспроизведения информации, полученной другими кодировщиками, что является типом эквивалентной надежности со статистическим коэффициентом, который передает степень согласованности действий кодировщиков. Этот коэффициент приводится в отчете по результатам контент-анализа.

По прошествии известного времени (например, трех месяцев) исследователь также проверяет, насколько устойчива надежность взаимодействия, для чего каждый кодировщик заново самостоятельно кодирует текст, который он уже кодировал ранее. На основании полученных результатов исследователь делает вывод о том, сохраняется ли стабильность кодирования. Например, шесть часов телевизионных эпизодов кодировались в апреле, а затем подверглись новой кодировке в июле, при этом кодировщик не имел возможности пользоваться полученными ранее результатами. Любое значительное отклонение обязывает переобучить кодировщика и повторно провести кодирование текста.

Специфика работы с данными на электронных носителях. Если анализируемые документы имеют электронные копии, процесс анализа значительно упрощается, но необходимо не забывать о разнице между возможностями человека и машины. Так, например, с применением компьютера более уместен анализ документов управления, где не нужны эмоциональные оценки и большее внимание может быть уделено анализу лексики.

Особое внимание здесь следует уделить установлению однозначного синонимического отношения. Например, дать программе команду считать синонимами или индикаторами одного и того же явления в рамках поставленной задачи слова, например, нормативы и тарифы.

Все определяется постановкой задачи; если, например, анализируется стиль управления государственным имуществом, то в рамках гипотезы могут быть рассмотрены два направления – рыночное и директивное, где рыночному соответствуют термины: цена спроса, цена предложения, конъюнктура; а директивному – нормативы отчислений, тарифная база. Если же исследование текста идет с позиций анализа состояния рынка, то термины спрос и предложение не могут быть использованы в качестве единой смысловой единицы.

Процедура проведения контент-анализа. Подробнее остановимся на одном из подходов к анализу документов и рассмотрим его практическую реализацию. Далее речь пойдет о контент-анализе, сочетающем качественную и количественную стратегии (в том числе, применение методики факторного анализа). Единицей анализа является фильтр – семантическая цепочка, состоящая из некоторого количества лексем.

Лексема [<гр. lexis слово, выражение, оборот речи] — *лингв.* единица словаря языка; в одну лексему объединяются разные парадигматические формы одного слова и разные смысловые варианты слова, зависящие от контекста, в котором оно употребляется.

Описывая процедуру контент-анализа, можно выделить несколько этапов:

1. Разработка программы исследования (цели, задачи, гипотезы). Этот этап работы определяет срезы содержания. На этом этапе, как правило, формулируется т.н. эмпирическая теория исследования. То есть, в ходе подготовки к проведению контент-анализа, ученый систематизирует гипотезы, существующие в контексте данной проблематики и отсеивает те из них, которые не поддаются верификации на данных информационного массива.
2. Построение выборки документов на основе определения общей совокупности, какие документы являются носителями необходимой информации.

- Определение круга и объема документов, являющихся носителями необходимой информации (наименование, периодичность выхода, период, тиражи).
- Построение выборки: какие документы по каким критериям будут привлечены для анализа.
- Анализ правильности построения выборочной совокупности.

3. Моделирование содержательного плана текста.

- Классификация социальных ситуаций, соответствующих рассматриваемому кругу проблем.
- Определение набора единиц анализа.
- Проверка надежности методики.

4. Кодирование единиц анализа.
5. Проведение непосредственного анализа-расчета информации – сбор информации.
6. Анализ результатов.
7. Оформление полученных результатов.

8. Написание отчета.
9. Презентация результатов.

0100090000030202000002008a01000000008a01000026060f000a03574d464301000000000000100b67b
0000000001000000e802000000000000e8020000010000006c0000000000000040000005a000000e400
0000000000000000000005b3c0000e718000020454d4600000100e80200000e000000200000000000000
0000000000000000622400001e350000c5000000200100000000000000000000000001a0203005064
0400160000000c000000180000000a000000100000000000000000000000090000001000000861c0000
c50b0000250000000c0000000e000080120000000c000000010000005200000070010000010000038fff
fff0000000000000000000000009001000000000cc0440001254006900 6d006500730020004e00650077
00200052006f006d0061006e000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000000000000000000
0000000000000000000000000000000000000047169001cc000202060305040502030 4873a0020000
0000000000000000000000ff0100000000000540069006d006500730020000000650077002000 52006f00
6d0061006e00000039ffca3078011400080400000000000d824510c0000000090ef0f0d12b5023090ef0f
0d4c6eaf30a8ef0f0d6476000800000000250000000c0000000100000018000000 0c00000000000002540
0000054000000000000000040000005a000000e400000001000000ba5d0740a28c074000000000b800000
0010000004c0000000400000000000000000000000841c0000c50b0000500000002000040 05b000000460
00000280000001c000000474449430200000 0ffffffffffffffff871c0000c60b000000000004600000014000
0008000000474449430300000025000000 0c0000000e0000800e000000140000000000000100000001
4000000040000000301080005000000 0b020000000050000000c021f01b702040000002e0118001c000
000fb02020001000000000bc02000000cc0102022253797374656d000000000000000000000000000000
0000000000000000000000000040000002d0100000400000002010100 1c000000fb02edff00000000000090
01000000cc0440001254696d6573204e657720526f6d616e000000000000000000000000000000000004
0000002d0101000500000009020000000 20d000000320a120000000100040000000000b7021f01205109
00040000002d010000030000000000

Рис. 1. Сетевая схема организации контент-анализа

Построение выборки и моделирование содержательного плана текста являются параллельными этапами исследования, взаимообуславливающими друг друга. Следует помнить, что проблемы, возникающие при кодировании, нередко ведут к пересмотру моделей выборки и текста. Анализ результатов и их оформление также могут идти параллельно. Например, при появлении статистических материалов принимается решение об оформлении их в виде графиков или диаграмм и эти графические материалы становятся объектом анализа; оформление статистических данных в виде таблиц приводит в иной форме представления и описания данных.

Следующим этапом проведения контент-анализа является составление словаря. Словарь часто называют классификатором контент-анализа, разработанной категориальной сеткой или таблицей контент-анализа, представляющей собой совокупность систематизированных и субординированных категорий и единиц. Он строится на основе создан-

ной системы категорий контент-анализа. Категориями являются генерализированные ключевые понятия, отражающие цель и задачи исследования. Категориальный аппарат и подчинённые ему единицы счёта, введённые в словарь в соответствии с созданной классификацией, должны идентифицировать общую тематику исследования и его частные особенности, то есть охватывать её полностью и максимально точно. Необходимо избегать крайностей при определении категорий контент-анализа. Так, при включении слишком крупных и размытых категорий, исследователь рискует получить тривиальные результаты, отражающие только общую суть вопросов. При введении слишком узких категорий есть вероятность получить большое количество малозначимой информации, которую крайне трудно будет в дальнейшем интегрировать и обобщить, для того, чтобы дать комплексную, но ёмкую оценку исследуемой проблеме. Категории должны максимально полно охватывать исследуемую тему, быть взаимоисключающими, не позволяющими включить одни и те же единицы одновременно в несколько категорий. Они должны обладать надёжностью и трактоваться единым образом. От корректного выбора категорий во многом зависят результаты всей работы, и поэтому исследователей уже давно интересует вопрос автоматизации выбора категорий. Решение этого вопроса позволило бы существенно экономить время проведения контент-анализа и также получать более достоверные и объективные результаты. Стоит оговориться, что автоматизированное создание системы категорий возможно только при работе с большими массивами.

При работе с текстом, в категориальную сетку заносится выбранный массив слов, словосочетаний, лексем (форм слов, имеющих сходное значение) в соответствии с поставленными задачами, определяются ключевые единицы, имеющие чётко идентифицируемую семантику, максимально точно соответствующие выделенным категориям контент-анализа, и их синонимический ряд. Важно учесть всю совокупность вариантов единиц, отражающих широту категории, не игнорируя кажущиеся малозначительными, но так же близкие по значению единицы. Часто в словарь вводятся единицы полярные по значению и оценивающим свойствам (например, антонимы), характеризующие своеобразные символьно-знаковые поля, в рамках которых и существуют эти оценочные единицы. Такое позитивное и негативное маркирование важно на последующих этапах контент-анализа при расчёте и изучении корреляций единиц. Серьёзным препятствием при определении позиции и дальнейшей оценки корреляции единицы того или иного оценочного символьно-знакового поля является различное отношение представителей целевых групп к одной и той же единице в конкретной изучаемой коммуникативной ситуации. Эта особенность исследуемой аудитории ставит дополнительные трудоёмкие задачи по исключению из категориальной матрицы единиц, трактуемых целевыми группами различно, способных

привести к получению некорректных результатов исследования. При невозможности исключения таких единиц из словаря определяющую роль при работе с ними играет исследование контекста конкретных элементов выборочной совокупности текстов, что позволяет верно трактовать значение использованной единицы.

На основе построенной таблицы создаётся так называемая кодировальная матрица, служащая инструментом квантификации заданных единиц в исследуемом массиве, о которой пойдёт речь в следующем параграфе.

Как уже было отмечено выше, единицами контент-анализа являются наиболее часто встречающиеся в тексте слова, словосочетания, предложения, абзацы, строки, колонки, физическая протяжённость и площадь текста, его доля в общем изучаемом массиве. Квантификации могут быть подвергнуты также и нетекстовые объекты, такие как аудио или видео плёнка, длительность трансляции по радио или телевидению.

Условно можно разделить словари на два вида: частотные и семантические. Первые подразумевают выделение единиц контент-анализа на основе частоты их использования по отношению к суммарному количеству слов в массиве и к другим единицам потенциально подходящим для включения в словарь. Второй вид представляет собой включение в словарь категорий и единиц на основе заранее проработанных текстов, максимально точно описывающих предмет исследования, а потому уже содержащих большинство единиц будущего словаря. Чаще всего в словарь включаются имена существительные, отглагольные имена, реже глаголы, прилагательные и наречия, совсем редко частицы и союзы. При этом важность выбора видов частей речи включённых в словарь варьируется в зависимости от исследуемого массива, поставленных задач, интуиции исследователя.

Стоит отдельно отметить важную особенность проведения контент-анализа художественных текстов. В них основной единицей счёта выступают не лингвистические единицы (предложения, словосочетания, слова), а смысловые единицы. Они могут содержаться, например, в одном словосочетании, в предложении, а могут находиться в одном абзаце, что существенно усложняет процесс поиска данных. Смысловые единицы помимо своей неструктурированности, могут быть имплицитны и неидентифицируемы в рамках поиска синтаксических единиц, а потому упущены. Существует также мнение о том, что смысловые нагрузки художественного произведения не могут быть соотнесены с нетекстовой действительностью, то есть быть подвержены кодированию и квантификации, а в дальнейшем – качественной обработке – интерпретации, в силу чего производить контент-анализ художественных произведений малоэффективно.

Квантификация и интерпретация результатов проведения контент-анали-

за. По завершению подготовительного периода следует этап работы с единицами подсчёта, выбранными в соответствии с установленной системой категорий, опирающейся в свою очередь на цель, задачи и гипотезы исследования. Здесь исследователь прибегает к помощи таких инструментов как регистрационная карточка или кодировальная матрица, бланк контент-анализа также называемый протоколом итогов контент-анализа. На основе систематизированного и дифференцированного материала исследователь пишет работу – записку по результатам контент-анализа, опираясь, главным образом, на протокол итогов, полученный в ходе полной дистрибуции категорий текстового массива. Определение тенденций и особенностей функционирования социальной реальности и является итогом проведения квантификации материала и контент-анализа в целом.

Сама процедура подсчёта (квантификации) близка стандартным действиям классифицирования по взаимоисключающим темам. Оперирование данными производится с помощью таблиц, математических формул, шкалирования, выстраивания данных в определённом заранее заданном порядке, специализированных компьютерных программ и т.д. Интерпретация, полученного числового материала и его дальнейшая тематическая градация, построение искомых моделей социальной действительности производится в соответствии с установленным изначально категориальным аппаратом, задачами, целями и гипотезами исследования.

Важно отметить, что иногда контент-анализ используется для определения лёгкости чтения конкретного набора текстов. Единицей счёта здесь является слово, имеющее любое значение. Основу квантификации в этом случае составляет длина слов, количество слов в предложении. На основе таких данных высчитывается общий индекс читабельности текста. Безусловно, такие характеристики, как жанр текста, язык, форма шрифта и другие визуальные и лингвистические особенности текста в немалой мере влияют на степень лёгкости восприятия текста. Контент-анализ и здесь не претендует на получение абсолютно точной информации. Такой вид контент-анализа позволяет также узнать приблизительный уровень образования, требуемый для понимания анализируемого текста. С этой целью используется формула Фреча. Применяют её для изучения англоязычных текстов.

Процесс квантификации чаще всего производится при использовании так называемых простых частот, подразумевающих поиск единиц счёта в одном текстовом массиве. Этот подход неприменим в случае сравнения текстовых массивов. Для сравнения используются относительные частоты, отражающие количество упоминаний единицы счёта на заданный фиксированный по объёму массив (например, на 1000 слов или 1000 страниц текста).

При анализе небольших массивов текста чаще всего единицей счёта является сло-

во, входящее в созданную систему категориального аппарата анализа. Но при исследовании больших массивов иногда допускают некоторую редукцию значимости количества слова в рамках заданной по объёму части текста. Так, абзац, в котором искомое слово упоминается 1 раз, будет приравнен к абзацу, в котором оно использовано многократно.

Регламентирует работу исследователя специализированная инструкция кодировщика, призванная определять то, каким образом будет собираться и регистрироваться (кодироваться) информация. Другими словами, эти система норм и правил, в частности устанавливающая определённые рамки, за которые нельзя выходить исследователю при работе над массивом. В ней приводятся конкретные примеры кодирования, алгоритмы работы со спорными случаями, характеристика категорий и единиц анализа. Серьёзные трудности могут возникнуть при квантификации и обработке данных массива, состоящего из художественных произведений, в том случае, если работа над категориальным аппаратом была проведена недостаточно скрупулезно. Интерпретация смысла при работе с такими массивами заключается в идентификации смысловых единиц. Поиск их, как уже было сказано, затруднён, а семантическая нагрузка может также сильно варьироваться и качественно и количественно, что в ещё большей мере усложняет кодирование и может привести к получению необъективных результатов. При этом в большинстве случаев контент-анализ доверяет регистрации лингвистических единиц, оперируя предположением о соответствии в большинстве случаев смысла отрывка текста семантике включённых в него единиц счёта. Такое формальное отношение к художественному тексту допустимо в меньшей степени, чем к другим его видам. Для решения этого вопроса в анализ вводят дополнительную единицу – тему. Это позволяет редуцировать вероятность несоответствия слова искомому в рамках конкретного этапа исследования значению. Другим вариантом преодоления такой трудности является использование «мнения арбитров» – то есть кодировщиков, классифицирующих контекст, в котором была использована единица счёта. Безусловно, и сам исследователь может им являться.

В ходе квантификации также производится анализ взаимодействия (корреляций) единиц счёта и далее интерпретация этих корреляций. Средством их определения может служить включение в контент-анализ других видов статистиче6ского анализа, например, факторного. Такое сочетание методов позволяет производить глубокий, разнонаправленный и точный контент-анализ. На примере функционирования этой программы легко убедиться в том, что контент-анализ это качественно-количественный метод, способный идентифицировать не только эксплицитные характеристики текста, определить его тематику, оценочную составляющую и т.д., но и проследить имплицитные сюжетные линии,

которыми изобилует массив, скрытые для читателя или исследователя не вооружённого таким инструментарием.

2.3. Процедура проведения контент-анализа в пакете Lekta

Пакет Lekta – лексико-семантический текстовый анализатор – был создан с целью проведения контент-анализа больших текстовых массивов. Помимо удобного интерфейса программы, позволяющего решать все базовые задачи метода, она способна производить факторный анализ лексем, выделенных при первичной обработке массива. Это делает возможным идентификацию лексически и тематически коррелирующих текстовых фрагментов, основных и частных латентных и более ярких сюжетных линий, лежащих в основе изучаемого текста, что даёт возможность сделать контент-анализ более глубоким и многомерным. Работа в пакете Лекта подчиняется классическим правилам и канонам контент-анализа, описанных в предыдущих параграфах. В проведении контент-анализа с помощью программы можно выделить несколько наиболее существенных этапов и несколько второстепенных, но предпочтительных.

Набор текстового массива производится из текстов одного жанра и стиля, зачастую важно также не выходить за определённые установленные временные рамки создания анализируемых сочинений. Сами тексты помещаются в документы формата txt. Возможно размещение всех текстов в одном документе. В дальнейшем средствами пакета он будет разбит на фрагменты в соответствии с установленными критериями.

Создание единого реестра текстов. Если количество текстов невелико рекомендуется создать реестр материалов, вошедших в массив в формате xls или аналогичных. В такой реестр имеет смысл включить по возможности как можно больше данных о материалах. Например, если исследователь работает с публицистическими статьями, размещёнными в сети Интернет, важно зафиксировать в реестре оригинальное название статьи, закодированное название статьи, дату её написания, URL адрес, тему в соответствие с выбранной градацией (если массив текстов включает несколько очевидных тематических групп), фамилию и имя автора, название издания, и т.д. В дальнейшем при необходимости всегда можно обратиться к такому реестру.

Кодирование исходных данных текстов. Удобным и ванным инструментом при работе в программе Лекта является кодирование названий текстов. В таком коде можно коротко отобразить те особенности материала, на основе которых можно идентифицировать конкретную статью во всём массиве. Такие данные в качестве примера были приведены выше при описании единого реестра материалов контент-анализа. Так, например, если исследователю требуется закодировать статью, опубликованную 7 августа 2008 года, вы-

шедшую в Российском СМИ (допустим, что в массиве также используются материалы зарубежных русскоязычных изданий) и описывающую международный военный конфликт на Северном Кавказе, он может закодировать документ следующим образом: 7Cau070808, где: 7 – международный телефонный код России (статья была опубликована в российском издании); Cau – сокращение от Caucasus (Кавказ) эта составляющая кода описывает общую тему исследуемого вопроса; 070808 – дата публикации статьи. Если в этот день в Российских СМИ, включённых в выборку, было опубликовано несколько статей, касающихся изучаемой проблемы, то и их можно упорядочить, пронумеровав. Например: 7Cau070808_01, 7Cau070808_02, 7Cau070808_03 и т.д.

Деление текстов на фрагменты. Непосредственно через интерфейс пакета ЛЕКТА тексты, включённые в массив можно поделить на приблизительно равные части. Диапазон, в рамках которого производится разбивка текстов, вводится исследователем самостоятельно. Существует альтернатива такому типу разбития текстов. В пакете предусмотрена возможность делить их на части по определённой группе знаков, находящихся в тексте и являющейся границей разделения документа. Отдельной строкой вводится совокупность значков, которые наверняка не встречаются в тексте, например, @@ZZZ. Такой принцип деления хорош при необходимости сохранить целостность небольших по объёму текстов, например, постов на Интернет-форуме, коротких писем, глубинных интервью. Это особенно важно при некоторой латентности высказываемых мыслей, слабо поддающихся лингвистической идентификации. Важно подчеркнуть, что независимо от типа деления, тексты должны быть близки друг другу по своему объёму. В случае дальнейшей факторизации фрагментов, имеющих принципиально разную протяжённость, фрагменты, большие по размеру, станут препятствием при классификации и идентификации сравнительно небольших текстов, что может пагубно повлиять на результаты всего исследования.

Создание словаря контент-анализа. Ранее в работе уже был приведён материал, касающийся работы со словарём. Применительно к пакету ЛЕКТА, важно отметить удобную возможность программы производить просмотр словосочетаний, в которые входит конкретная лексема. Это особенно важно при необходимости включить в словарь полисемантическую лексему и отсеять все ненужные её значения. При запросе словосочетаний произвести такую верификацию несложно. Пакет даёт возможность осуществить сортировку слов по следующим критериям: лексема (по алфавиту), длина (по длине слов) и частоте встречаемости в массиве. Такая полифункциональность делает его применимым для решения различных задач составления словаря.

Проведение факторного анализа массива. Средства пакета ЛЕКТА позволяют произвести факторный анализ переменных (лексем), включённых в созданный исследователем словарь, в массиве изучаемого текста, что позволяет получить информацию о корреляциях групп слов в массиве. Как уже было сказано ранее, этот метод используется не только для классификации, интеграции тематик, связанных между собой в одни группы, но и для получения материала, выраженного в массиве неявно, имплицитно, что делает такой симбиотический союз методик особенно привлекательным.

Уточнение словаря. Так как выбор слов, включаемых в словарь, не может быть сделан безупречно при первом отборе, не все переменные органично группируются в результате факторизации. В этой связи следующим шагом при использовании пакета ЛЕКТА является повторная работа со словарём, заключающаяся в удалении малоинформативных (в рамках исследуемого массива) слов и добавлении слов, часто встречающихся вместе с уже включёнными в словарь. Такая работа производится при использовании полученных по итогам факторизации матриц собственных значений факторов и факторных нагрузок. После корректирования словаря вновь производится факторизация. Если переменные органично сгруппировались и позволяют дать чёткую характеристику каждому из факторов, то работу можно приступать к следующему этапу работы в пакете. Если же этого не произошло, необходимо повторить работу по удалению и добавлению слов в словарь.

Работа с темами. Необходимо дать краткую характеристику каждому из факторов, полученных в результате обработки массива в ЛЕКТА. После этого нужно сгруппировать все факторы по 3-м – 5-ти тематическим группам, предварительно определив их и дав и им названия, а внутри каждой группы определить удобный и логичный порядок описания каждого из факторов. Это поможет получить стройный и удобный для восприятия аналитический отчёт по проделанной работе. Результатом такой работы аналитический отчет или аналитическая записка по исследованию, статья для прессы, научная публикация, информационная записка.

Подробная инструкция к проведению контент-анализа с помощью пакета ЛЕКТА доступна для просмотра на http://nisoc.ru/.

ЛИТЕРАТУРА

1. Ньюман Л. Неопросные методы исследования // Социологические исследования, 1998, №6.
2. Психолингвистическая экспертная система Ваал(R) Руководство пользователя // http://www.vaal.ru.
3. Федотова Л.Н. Анализ содержания – социологический метод изучения средств

массовой коммуникации / Федотова Л.Н. // М., 2001.

4. Шалак, В. Компьютерный контент-анализ текстов как метод экономической разведки (полный вариант статьи) / Владимир Шалак // http://www.vaal.ru.
5. Шалак, В. Технология прогноза (подробности) / Владимир Шалак // http://www.vaal.ru.
6. Ядов В.А. Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности / В. А. Ядов // М.: «Добросвет», 1998.
7. Berg, B.L. Qualitative research methods for the social sciences / Bruce L. Berg. Boston, 1995.
8. Csigo, Peter The Discourse of Consolidation in Hungary / Peter Csigo, Balazs Vedres // Доклад на 5-й европейской конференции по социальным сетям, 27-31 мая 1998, Испания http://www.bke.hu/~socnet/Disccons/Disccons.htm.
9. Krebs, Valdis E. Uncloaking Terrorist Networks/ Valdis E. Krebs// First Monday, volume 7, number 4 (April 2002), http://firstmonday.org/issues/issue7_4/krebs/index.html.
10. Popping, Roel. Computer-assisted Text Analysis / Roel Popping // London: SAGE Publications, 2000.

Приложение I

Зарубежные компьютерные пакеты контент-анализа

TABARI (KEDS) – программа для автоматизированного кодирования данных политических событий. Она использует встроенный анализатор для идентификации слов используемых для проведения контент-анализа. При работе программа обращается к встроенным и загружаемым словарям. Данные могут быть использованы для работы в других специализированных программах, таких как SPSS и SAS.

JFreq создает матрицы частот употребления слов в массиве, используется для проведения контент-анализа и работает с большинством языков мира. Не работает с японским, китайским и тайским языками из за принципиально отличной от большинства языков лингвистической системы этих языков. Программа позволяет исключить из массива нечитаемые символы и знаки не входящие в алфавитную базу данных. Работает на любой операционной системе.

Concordance – программа, используемая для проведения контент-анализа электронных документов, В ней можно создавать списки связанных единиц счёта, индексов, слов, при работе над электронным текстом. Позволяет обрабатывать большие массивы. Даёт возможность просматривать корреляции между словами, входящими в словарь контент-анализа. Результаты работы легко разместить в Интернет с помощью встроенных инструментов программы.

HyperRESEARCH позволяет кодировать, находить и декодировать текстовые, аудио и видео материалы. Позволяет проводить анализ таких форматов данных.

LEXIMANCER – мультиязычное программное обеспечение, производящее контент-анализ больших объёмов текста, позволяющее совмещать в массиве тексты разных жанров и стилей, включая диалектические и другие нетрадиционные формы языка.

PROTAN – комплекс 30-ти программ, интегрированных в один блок, позволяющих проводить контент-анализ массивов текста с помощью встроенных словарей и идентифицировать сюжетные линии, определяя корреляции между словами словаря посредством проведения факторного анализа. Обладает большим количеством других функций.

TextQuest – описать (нужен ответ от Harald Klein)

TEXTPACK – кодирует тексты на основе созданных пользователем словарей. Производит сравнение 2-х документов, сравнивая их словарное наполнение, обнаруживает схожие отрывки внутри документов. Данные легко импортируются в такие пакеты, как SPSS или SAS.

QDA Miner является средством качественного анализа текстовых данных, аннотирования, получения и просмотра кодированных данных. Программа позволяет работать большим числом документов, содержащих как текст, так и числовые данные. QDA Miner также предоставляет широкий спектр поисковых средств для выявления корреляций кодированных данных.

WordStat – модуль анализа текста, предназначенный специально для обработки материалов, таких как журнальные статьи, литературные произведения, интервью. Как и другие

аналогичные программы позволяет создавать категориальный аппарат и словарь контент-анализа. Дальнейший анализ может быть произведён с помощью создания и расчета перекрестных таблиц, а также KWIS-метода. Пакет позволяет работать и с более сложными методами статистического анализа, такими как кластеризация и многомерное шкалирование. Созданные категориальные аппараты и словари схемы могут быть применены к другим текстовым массивам в дальнейшем.

SALT – программное обеспечение, анализирующее содержимое текстового массива. Поддерживает работу со всеми языками. Определяет среднюю длину предложения, количество искомых слов, общее количество слов. Может создавать алфавитный список слов, кодировать текстовый массив в соответствии с определёнными исследователем кодами. Работает только с операционной системой Windows.

MonoConc – производит поиск единиц текстового анализа, определяет корреляции между ними в массиве.

TROPES – производит хронологически-ролевой качественный анализ текста. Также позволяет получать общую информацию по частотности использования тех или иных единиц счёта.

Qualrus является инструментом проведения качественного анализа данных, кодирующий элементы массива для дальнейшей обработки. Qualrus может быть использован для проведения полного спектра качественных исследований, в том числе в культурологическом анализе, интерпретации методов, семиотике, истории, брекетинге, эмпирическом анализе, анализе рассказов и произведений других жанров.

CAMEO – система, созданная для кодирования и аналитики политических коммуникаций. Включает в себя 20 главных событийных категорий и 200 субкатегорий, обширную базу для кодирования имён политиков в тексте.

AnnoTape – это программное обеспечение для записи и анализа аудио, видео, графических и текстовых данных, предназначенное для качественных исследований, маркетинга, журналистики средств массовой информации, архивных служб. Производит запись звуковых файлов – интервью, бесед, радиопередач напрямую на жесткий диск компьютера. Позволяет сохранять до ста часов звука вместе с текстовыми данными в одной интегрированной базе данных. Производит анализ данных, аннотирование и индексирование оригинальных звуковых и текстовых файлов. Эффективно разбивает на фрагменты массивы аудио-данных