**Методика анализа содержания текстов**

**Федотова Л.Н.**

**Программа исследования как жанр**

Первоначально социологическое исследование существует как замысел исследователя. И если мы здесь проведем аналогию со словами Рене Клера, который в свое время воскликнул: «Фильм готов! Осталось только снять его!», то это не будет большой натяжкой. Именно на этапе замысла, который формально выливается у социолога в своеобразный документ, называемый программой исследования, автор представляет совершенно отчетливо все остальные этапы.

По В.А. Ядову, программа - это изложение основных задач, методологических предпосылок и гипотез исследования с указанием правил процедуры и логической последовательности операций по проверке гипотез [1].

Хотя на этих страницах исполнитель исследования чаще упоминается в единственном числе, делается это ради простоты изложения. На самом деле технология добычи социологического знания требует довольно разнообразных профессиональных умений большого числа исполнителей. Еще и поэтому исследование на этапе своего замысла должно иметь организационный план своего осуществления, представление о финансовых затратах, о профессиональных ресурсах, о сроках исследования, о форме демонстрации (выдачи) результатов заказчику и т.п.

Прежде всего в программе формулируется социальная проблема, к изучению которой приступает исследователь. Это может быть новое социальное явление, распространенность и характеристики которого неизвестны обществу и поэтому являются самодостаточной задачей для социолога; это может быть и уже известное явление, но оставшеся до поры до времени вне поля зрения социологов - ситуация, при общей неразвитости социологических исследований в нашем обществе, довольно частая; это может быть и уже изученное социологами явление, но нуждающееся в повторном исследовании в связи с явными изменениями обстановки в стране и т.п.

В свое время американские исследователи считали вполне заслуживающим внимание социолога-исследователя текстов такие задачи:

Какие изменения претерпели в СССР первомайские пропагандистские призывы за время существования советской власти?

Как отражается склад характера писателя на том, что он пишет?

Каковы важнейшие тенденции в использовании научной литературы химиками и физиками за последнее время?

Как человеческие ценности, отображенные в американских пьесах, отличаются от ценностей, фигурирующих в немецких пьесах того же периода? Как моральные принципы, содержащиеся в американской бойскаутской литературе, отличаются от тех понятий о морали, которые мы встречаем в литературе «гитлерюгенда» (или в повести Гайдара «Тимур и его команда» - добавим мы: Л.Ф.)?

Как можно проверить коммуникацию, заподозренную в том, что она преследует подрывные цели, как достоверно выяснить, имеет ли она пропагандистское содержание или нет?

Какое сравнение можно провести между газетами и радиостанциями в отношении освещения ими сенсационных убийств?

Что делает ту или иную статью или произведение писателя «читабельными»?

Каковы сходные и отличительные черты политических символов, которые предлагаются вниманию людей в отдельных государствах?

Что происходит с хорошей книгой, когда на ее основании делается кинокартина? [2]

Приведем примеры таких вопросов, «привязанных» к нашим сегодняшним реалиям: каково содержание рекламы продуктов отечественного производителя? Чем эта реклама отличается от импортной? К каким ценностям личности она апеллирует, не является ли она дисфункциональной, не следует ли изменить ее характер? Таким вопросам несть числа.

В любом случае исследователь так составляет программу, чтобы в результате ответить на поставленный вопрос. Прежде всего, он должен отразить степень изученности проблемы до него. Эта стадия является началом любого исследования. Таким образом, анализ предшествующей литературы, содержащей попытки коснуться аналогичной проблематики, становятся не просто данью профессиональной этике, - хотя и это важно, - но и демонстрирует фундаментальность подготовки исследователя. Помимо этого данный этап имеет сугубо прагматическую ценность. Исследователь получает возможность поставить перед собой ранее нерешенные задачи, получить возможность для сопоставления своих результатов с ранее полученными и т.п.. Только так, в результате преемственности научного знания, его углубления общество получает знание о самом себе, что составляет главную ценность социологии.

Но - любой заданный исследователем самому себе вопрос, это лишь поверхностный слой проблемы. Действительно, каково тематическое распределение материалов в газете - это еще не проблема, это лишь материал для решения проблемы, а каким ему быть? Это распределение должно быть для молодежной газеты одним, другим для издания, считающегося в общественном мнении качественным, третьим для городского радио и т.д.

Теоретик может написать книгу о том, каким должно быть это распределение для всех этих каналов, ссылаясь на то, что все они потребляются одной аудиторией, или наоборот, потому что они потребляются разной аудиторией, поскольку все они вместе образуют иерархию каналов и потому каждый должен «обслуживать» свой участок, что они должны при этом учитывать иерархию тематических интересов аудитории...

Но при всем этом нужно знать реальное распределение материалов по темам, чтобы это было информацией к действию: к изменению, к переструктурированию тем, к усекновению одних и появлению других. Сейчас даже не важно, кто будет осуществлять это изменение: сам ли коммуникатор, или ему укажет на необходимость этого издатель, или это останется «гласом вопиющего» исследователя. Только так замкнется, так сказать, технологический цикл социологического исследования.

Каково распределение в газете «Истинный американец» (см. раздел 2) суждений о Гитлере и проч. - это не проблема; это лишь материал для решения проблемы: а каким ему быть в газете, издаваемой в стране, находящейся в состоянии войны с Гитлером? Симптоматично, что это исследование было буквальной информацией к действию - для суда, который обвинил газету в профашистской ориентации. Исследование, скажем тут мы, было, как никакое другое, «действенным».

Каков набор «моделей поведения» для домохозяек в американских «мыльных операх» - это не проблема; вопрос в том, каким ему быть, если исходить из более гуманистических теорий и концепций личности, чем рожденных в лоне общества потребления?

Хотелось бы предостеречь от узкого понимания роли контент-аналитика, как умещающейся только в рамках поставщика этой самой «добротной» информации для принятия решения. Именно степень «добротности» этой информации зависит от понимания аналитиком всего цикла исследования.

Я бы даже высказала одну кажущуюся парадоксальной мысль - неважно, присутствует ли весь этот цикл, прописывается ли он автром в отчете - а чем иным являются, например, книги, по которым мы знакомимся с историей контента, как не отчетом авторов о проделанном исследовании? Важно, чтобы он - этот цикл - осознавался в начале исследования и особенно - если это исследование осуществляется начинающим исследователем. Тут нелишне напомнить, что социологическое исследование - это всегда «точка зрения»: в данном случае на выбор определенных характеристик текста (по степени важности?), на интерпретацию «что такое хорошо» и «что такое плохо», на выбор определенных выводов для практических рекомендаций (по степени настоятельности?). Вспомним здесь и фигуру заказчика исследования. Из истории архитектуры известно, что там многое зависит не только от гениальности мастера, но и от гениальности заказчика.

Золотое правило, действующее во всех исследованиях средств массовой коммуникации, говорящее, что необходимо ясно осознавать, в чем состоит проблема, прежде чем исследовать ее, справедливо и для метода анализа содержания. Этот метод подразумевает большую кропотливую работу для того, чтобы решить проблему, Но не менее большая работа нужна, чтобы нащупать эту проблему.

В практике исследований текстов методом анализа содержания некоторым аналогом проблемы в конкретном исследовании выступает категория анализа. Это емкое, но вместе с тем лаконичное выражение проблемы. Категория анализа не является рабочей, непосредственно данной нам в тексте; эта характеристика скорее принадлежит всей совокупности текстов. Так, в нашем иллюстративном материале мы разбирали понятие «расовой дискриминации». Мы показали на примерах, что реальность дает нам самые разнообразные проявления расовой дискриминации, причем на этих проявлениях не всегда есть этикетка с надписью «расовая дискриминация». Авторы этого конкретного исследования работали с категорией анализа «квалификация героев рассказов разной расовой принадлежности» и пришли к выводу, что в текстах имела место расовая дискриминация. То есть - наблюдая свокупность текстов под углом зрения «квалификация героев разной расовой принадлежности» автор выходит на проблему «расовой дискриминации».

Или возьмем случай, когда в качестве категории анализа выступает «оперативность». Для ряда источников информации, подытожив результаты, мы бы вышли на проблему недостаточной оперативности его материалов. Этот пример как нельзя лучше показывает, что, введя только понятие «категория анализа», мы остаемся на очень высоком уровне абстракции. В самом деле - оперативность бывает разная. Какая? Материалы могут быть с «сегодня» и с «не-егодня» случившимся поводом. Можно предложить другой вариант дробности: так, в анализе содержания местных газет, радио и телевидения как источников информации, осуществленных в рамках проекта «Общественное мнение» [3], использовались следующие градации оперативности:

Сообщение о событии дается в день совершения последнего

Сообщение о событии дается на следующий день после совершения последнего

Сообщение о событии дается через два дня после свершения

Сообщение о событии дается через три-семь дней

Сообщение касается события, произошедшего более семи дней назад или, как указывается в сообщении «недавно», «за последние дни» и т.д.

Сообщение касается события, дата свершения которого не указывается

Сообщение касается вневременных проблем

Видно, что анализировалась именно местная пресса тридцать лет назад - если бы мы исследовали сегодняшние информационные телевизионные выпуски, мы должны были бы предусмотреть репортажи с места события «минута в минуту» и т.д. А поскольку это были местные издания, где существовал такой способ донесения информации до читателя, как перепечатка из центральных газет, появление здесь градации с оперативностью трех-четырехдневной давности было вполне оправданным.

Когда же мы стали анализировать информационную телевизионную программу «Время» [4]- средство (и канал) заведомо более оперативное - мы пошли по пути дальнейшей дифференциации событийного повода материалов, идущих с грифом «сегодня». Это могли быть информационные сюжеты о событиях, которые можно было предсказать за много десятилетий до этого дня: так называемые календарные, юбилейные или знаменательные даты - «сегодня исполнилось 200 лет со дня рождения А.С. Пушкина». С другой стороны, это могут быть сюжеты типа «сегодня проснулся вулкан X » - событие, обладающее для аудитории максимумом неожиданности, непредвиденности, непредсказуемости. Потребовались и более «оперативные» градации для телевизионной программы общенационального масштаба.

Если проследить за нашей логикой описания оперативности, то она состояла в том, что мы описали ее с помощью разных типов, разных случаев оперативности. По сути дела, взамен однородного понятия оперативности мы создали группировку, классификацию разных типов оперативности. Как в жизни - для каких-то случаев достаточно сказать о спортсмене, что он хороший стрелок, а для каких-то - в спорте - надо уточнить, что он стрелок 3-го, 2-го, 1-го разряда, мастер спорта и т.п.

Обратимся с этой точки зрения к географии события, а значит, и материала о нем. Известно, что в нашей стране больше тысячи городов, десятки тысяч сел. Мы не можем в анализе состояния дел с «географической сбалансированностью», - если бы это было категорией анализа в нашем будущем исследовании, - отражаемой источником информации об объекте, говорить о представительности каждого населенного пункта: точек на карте слишком много. Сбалансированность такого рода может и должна подразумевать пропорциональную представительность достаточно крупных регионов. Каких? Существует множество типологий такого рода, исследователь вправе выбрать любую. Страна разделена на области, края, республики - можно говорить о представленности таких понятий; существует деление на экономико-географические районы: Урал, Западная Сибирь и т.д. - можно на этом построить анализ; существует типология поселений: село, малый город, средний город, крупный город - можно говорить о представленности таких типов поселений.

Мы, по сути дела, сделали следующий шаг в построении, создании программы исследования, который состоит в конструировании категориальной сетки, или, если выразиться на языке терминов, в классификации, группировке семантического пространства категории анализа. Если говорить образно, то речь идет о создании своеобразного сита, через которое будет просеяно реальное содержание исследуемого информационного источника, где роль ячеек будут выполнять типы, разные модификации категории анализа.

Обращаясь к конкретным текстам газет, радио и телевидения, исследователь оперирует рабочей единицей анализа. В нашем пример из исследования о расовой дискриминации таковой будет «персонаж» - фигура белого рабочего или темнокожего с определенной степенью квалификации.

Исследование оперативности материалов имеет в качестве единицы анализа отдельный материал, поскольку мы связываем оперативность с оперативным поводом целого материала.

Когда речь идет о географии информационного материала, единицей анализа может быть отдельный материал. Но вот большие газетные, теле- и радиоматериалы содержат (могут содержать) информацию о разных населенных пунктах, о разных областях и т.п. Значит, надо предусмотреть объем информации для анализа, который будет мельчайшей единицей текста, где может появиться упоминание искомого нами географического типа. Прежде всего это может быть слово, равное названию пункта (края, города, области, региона); это может быть абзац, посвященный нужной нам «географической точке»; это могут быть более пространные текстовые фрагменты, содержащие разговор о нужной нам точке.

Вот здесь пора вспомнить, что анализ содержания выясняет объем внимания к какой-либо проблеме в рамках отдельного материала, газетного номера, радио- или теледня, недели, месяца и любых других временных отрезков в деятельности анализируемого канала [5].

Оперируя словом как единицей анализа, мы подсчитываем, сколько слов приходится на тот или иной тип в избранной нами типологии, подразумевая, что чем больше слов посвящено какому-либо типу, тем больше внимания коммуникатор уделяет именно этому типу.

Оперируя абзацем как единицей анализа, мы подсчитываем, сколько абзацев посвящено тому или иному типу в избранной нами типологии, подразумевая, что чем больше абзацев, тем больше внимания уделяется тому или иному типу... При этом мы игнорируем разницу в размерах абзаца, рассматривая их равноправно, расцениваем, что несмотря на разницу в размерах, каждый из них прочитывается как бы на «едином дыхании», а с точки зрения автора текста - выражает законченную мысль.

Оперируя фрагментом текста как единицей анализа, мы должны измерить, какой фрагмент текста - крупный или небольшой - соответствует определенному типу поселения, Значит, появляется нужда в пространственных - если речь идет о газете, или временных - если речь идет о радио- или телепередачах, мерах. Можем измерять эти куски строками, площадью или минутами, секундами. То есть здесь мы выходим на новые, по сравнению со счетом по нарастающей, соответствующей количеству фиксируемых нами признаков, единицы измерения.

Таким образом, если сгруппировать все возможные методики оперирования с текстом, с учетом разных единиц анализа и разных единиц измерения, то грубо их можно разделить на две группы:

1. Единицей анализа выступает признак материала, который характеризует его в целом: это может быть оперативность - фиксация (или нефиксация) временной вехи события, которое выступает событийным поводом появления материала в потоке информации; это может быть функциональная принадлежность материала: так, радио- и телевизионные материалы традиционно делятся на информационные, развлекательные, учебные, утилитарные (реклама и справки).

Группировка может быть другой [6], но важно, чтобы каждый материал относился к той или иной группе по своей доминирующей функции.

В таких случаях признак должен описываться конечным числом случаев, или, иначе говоря, исследователь должен предусмотреть все возможные модификации признака, чтобы каждый материал был зачислен к той или иной модификации. Есть какие-то материалы, в которых бессмысленно искать оперативный повод, например художественные. Значит, исследователь предусматривает графу «материалы, не классифицируемые по этому признаку» и относит туда эти материалы [7]. Есть материалы, для которых трудно вычленить главную функцию - значит, исследователь предусматривает графу «трудно классифицируемые случаи» и относит туда этот материал. В совокупности все анализируемые материалы распадаются на столько групп, сколько модификаций признака предусмотрел исследователь. Подсчет количества материалов в каждой группе, принадлежащей одной модификации, происходит двумя способами, которые мы тут подытоживаем:

а - подсчитывается само количество этих материалов, выражаемое натуральным рядом чисел (от одного до n ...);

б - подсчитывается сумма площадей (в газете) или эфирного времени (на радио и телевидении) материалов, приходящихся на каждую группу.

2. Единицей анализа выступает отдельный фрагмент текста: слово, персонаж, суждение, абзац, или нетвердо фиксируемый отрезок текста, совпадающий с определенной темой. Во всех случаях речь идет о довольно ограниченном наборе слов, персонажей, суждений, который интересует исследователя.

Их появление в тексте фиксируется уже значительно большим числом способов:

а - возможность подсчета количества упоминаемых в тексте слов, персонажей, суждений, тем и т.п. сводится к таким вариантам (разберем эти варианты для самого простого случая, когда единицей анализа является слово):

подсчитывается общее число упоминаний каждого слова;

подсчитывается количество предложений, абзацев, которые могут содержать искомые слова;

подсчитывается число материалов, в которых встретилось хотя бы одно искомое слово.

б - подсчитывается сумма площадей (в газете) или эфирного времени (на радио и телевидении) отрезков текста, содержащих нужное слово:

тут мельчайшей частицей выступают законченные, «оформленные» единицы площади (абзац, страница микрофонной папки с радио- и телематериалами) с нужным словом, сумма которых и подсчитывается;

суммируются площади материалов в целом, если они содержат хотя бы одно искомое слово.

Здесь следует ввести понятие единицы контекста. Обсудим его на примере подсчета слов, соответствующих категориальной сетке исследователя. Слово живет в рамках предложения; более сложной пространственной единицы, состоящей из нескольких предложений, если все они обеспечивают законченное суждение; абзаца и, наконец, материала в целом. Если объектом исследования служит небольшой текстовой материал, имеет смысл подсчитывать все количество упоминаний слова. Но иногда при больших объемах исследователи устанавливают единицы контекста, и тогда употребление хотя бы одного искомого слова в рамках абзаца считается равноценным его неоднократному употреблению в этом же абзаце. Более того, упоминание слова в рамках материала иногда фиксируется как равнозначное неоднократному его употреблению в рамках этого же материала.

Так, в исследовании языка политики школой Г. Лассвелла, когда кодировщики оперировали совокупностью анализируемых текстов в 20000 редакционных статей и 416 символами, отсутствие или присутствие которых они фиксировали в тексте, тактика кодировки состояла в том, что достаточно было встретить хотя бы одно упоминание - независимо от того, сколько на самом деле их было в тексте: ведь, как объяснялось в инструкции, вариации числа появления могли зависеть от индивидуального стиля автора [8].

Как правило, перед тем, как принять решение, какой стратегии подсчета придерживаться, проводится пробный подсчет вариантов - не искажает ли выводов выбор той или иной стратегии.

Чтобы проиллюстрировать механизм существования обозначенных нами категорий, единиц анализа, измерения и контекста, разберем с этой точки зрения некоторые исследования, которые мы ввели в пласт наших рассуждений о контент-анализе и которые, таким образом, уже знакомы читателю.

1 пример. В исследовании Г. Лассвелла « World Attention Survey », который мы подробно разбирали с точки зрения его фактуры, категорией анализа были символы политики, язык политики (поименованные политические деятели, названия идеологий, политических движений и партий, властных институтов); единицей анализа было слово ; единицей счета было упоминание отдельного слова, количество их измерялось натуральным рядом чисел в абсолютном выражении или процентами в относительном выражении; единицей контекста - отдельный абзац в газетном материале.

В исследовании телемостов СССР - США мы использовали в качестве единицы анализа местоимения «мы», чтобы выйти на категорию «степень объединительных тенденций у участников телемостов». Как представляется, таблица 3.1 чрезвычайно показательна для оценки таких тенденций.

Таблица 3.1

Употребление местоимений участниками диалога «Ленинград-Сиетл» (в % к общему числу употребления «мы» каждым субъектом высказывания)

|  |  |  |
| --- | --- | --- |
| Употребление местоимения «мы»  | Советская сторона  | Американская сторона  |
| ведущий  | аудитория  | ведущий  | Аудитория  |
| Американский и советский народ в совокупности  | 8  | 4  | 3  | 7  |
| Участники телемоста в совокупности  | 8  | 5  | 5  | 9  |
| Или только американский, или только советский народ  | 29  | 39  | 26  | 33  |
| Или только американские, или только советские участники телемоста  | 16  | 5  | 26  | 1  |

Укажем, что единицей счета было тут каждое упоминание местоимения, а единицей контекста весь телемост в целом (понятно, что для такого скрупулезного анализа исследователь должен располагать стенограммой состоявшегося общения).

2 пример. Исследование Г. Лассвелла газеты «Истинный американец»: категорией анализа было отношение газеты к властным структурам трех стран (США, Германии, Японии); единицей анализа было суждение относительно правительства, лидеров или политики этих стран; единицей счета было упоминание этих суждений, количество их измерялось натуральным рядом чисел в абсолютном выражении; единицей контекста - отдельное предложение, в рамках которого могло появиться или одно, или несколько интересующих исследователя суждений.

3 пример. В исследовании телемоста «Ленинград-Сиетл» между американской и советской аудиториями мы получили по одной из задач исследования такие результаты (табл. 3.2):

Таблица 3.2

Характеристика задаваемых в ходе телемоста вопросов (в % к числу вопросов, заданных сторонами)

|  |  |  |
| --- | --- | --- |
| Характеристика вопросов  | Советская сторона  | Американская сторона  |
| Вопросу предшествовала негативная информация о стране-адресате  | 18  | 32  |
| Вопросу не предшествовала негативная информация  | 82  | 68  |

Здесь категорией анализа была оценочная нагрузка задаваемых в ходе обмена мнениями вопросов; единицей анализа - фрагмент текста, совпадающий с вопросом, заданным в ходе дискуссии, и вводкой к этому вопросу; единицей счета - натуральный ряд чисел как в абсолютном, так и относительном выражении; единицей контекста - каждый вопрос [9].

4 пример. В исследовании "Общественное мнение" в процедуре, связанной с анализом содержания местных газет, радио и телевидения, в одной из задач мы брали в качестве категории анализа сферу социальной действительности; единицей анализа был фрагмент текста, совпадающий с темой при описании каждой из этих сфер; единицей счета - пространственные (метрические) куски газетных колонок, совпадающие с этими темами, измеренные в условных единицах строкомером; единицей контекста - весь материал.

Как правило, перечень категорий и единиц анализа, способы оперирования с анализируемым текстом представлены в особом документе - инструкции кодировщика. Основная ее часть составляет перечень категорий и их типологических (или лексических) модификаций.

Эта часть обычно называется «кодом». Если продолжить пример с такой категорией анализа как «оперативность», то наш гипотетический код выглядел бы следующим образом:

Оперативность

1 - информация со словом «сегодня», в том числе:

календарные даты

непредсказуемые события

остальные случаи

информация о ранее случившихся событиях

информация без указания времени события.

Или возьмем категорию «география» (следующую типологию выбираем из-за краткости).

География

1 - крупный город

2 - средний город

3 - малый город

4 - поселок городского типа.

Осталось только снабдить этот фрагмент собственно инструкцией, например, указать, города с какой численностью населения относить к крупным, средним и малым, указать, какие справочники могут служить для уточнения данных о поселках городского типа.

Надо сказать, что реальные исследования редко ограничиваются одной-двумя категориями анализа.

Так, исследование местных средств массовой информации и пропаганды в проекте «Общественное мнение», которое мы уже представляли, имело следующие категории анализа: функциональная принад лежность материала, тип автора, жанр, оперативность, модальность, локальность, тема, элементы содержания, содержащие суждения о соотношении социалистической и капиталистической систем по признакам экономики, демократии, идеологии, условий развития личности [10].

Исследование информационной телевизионной программы Центрального телевидения «Время» (1984 - 19 87 гг.) оперировало такими категориями: тематика, суждения о роли СССР и США в борьбе за мир, авторство таких суждений, суждения об актуальных народно-хозяйственных проблемах СССР, география, степень критичности материалов, род занятий лиц, выступающих в программе, и лиц, упоминаемых участниками программы, жанр.

Теперь понятно, почему кодировщик должен располагать документами, где бы были перечислены все модификации характеристик, а также оговорены все правила анализа. Все это трудно запомнить, но, кроме того, такой текст нужен для решения спорных вопросов. По крайней мере нулевая отметка будет записана в этом исходном материале. И еще одно соображение, вес которого трудно переоценить - по правилам строгости анализа автор не может выступать в роли кодировщика. Слишком большой объем информации он содержит в уме и иногда принимает решения, исходя из этого обстоятельства. Трудно будет доказать критикам работы, выдерживался ли одинаковый под ход к разным текстам, не было ли ситуаций, когда «три пишем - два в уме». Автор должен снабдить кодировщика, а иногда и двух, «маршрутом и компасом» анализа, чтобы они, работая по одной инструкции, пришли к одинаковому выводу.

Кроме того, это действительно те документы, объясняющие, как получены те результаты, которые явились итогом анализа текста. Демонстрация инструментария считается одним из первых требований к корректному социологическому исследованию. Дело читателя исследования соглашаться или не соглашаться с авторской интерпретацией результатов, но он как минимум должен знать, как эти результаты получены.

Если представить себе конкретную ситуацию, когда кодировщик приступает к анализу выбранного массива газет или садится к экрану телевизора, чтобы отслеживать в тексте определенные характеристики, оговоренные кодом и инструкцией, то нужно предусмотреть еще один документ, где кодировщик будет фиксировать свои наблюдения (о возможностях применения на этой стадии компьютера мы еще будем говорить - соображения необходимости контроля за итогами кодировки иногда перевешивают и заставляют работать с реальными документами, зафиксировавшими итоги кодировки).

Эта фиксация может происходить одним из трех способов:

1-й из них может быть уподоблен работе с библиографическими карточками: на чистый бланк заносятся названия характеристик или числовая нумерация их по коду, эти бланки накапливаются и затем по мере необходимости суммируется частота появления той или иной характеристики.

2-й способ состоит в том, что предварительно заготовляется массив бланков с отпечатанными на них названиями характеристик (по типу «листка по учету кадров»). Кодировщику остается в нужном месте вписать обнаруженную им характеристику.

3-й способ - самый экономный, но возможный только для решения задач, когда исследователя интересуют лишь упоминания характеристик, но не объемы текста, где эти характеристики встречаются, - когда карточка представляет собой сплошной ряд чисел (совпадающий с общим числом кодов), а кодировщик обводит кружком лишь те коды, которые соответствуют встреченным в тексте характеристикам. Подсчет частоты появления каждой характеристики в целом по массиву будет итогом первичного анализа выбранной совокупности текстов.

Кодировщик выступает в исследовании непосредственным исполнителем, который работает по инструкции, составленной автором-исследователем.

По социологическим нормам принято, чтобы один и тот же массив газет и журналов (или текстов радио- и телепередач) обрабатывался двумя кодировщиками. Это повышает надежность полученной информации. Когда анализ не очень сложен, речь может идти об установлении чисто механических ошибок. Иногда же при сложности отнесения того или иного куска текста к той или иной единице анализа, могут возникать различные интерпретации текста и тогда ошибки возникают из-за недостаточной ясности инструкции кодировщику. Одно же из требований к любой методике анализа содержания состоит в том, чтобы она была составлена так ясно, чтобы разные исследователи, работающие по одной методике, пришли к одному и тому же выводу.

Когда в исследовании кодировщики имеют дело с идентификацией оценки коммуникатора того или иного факта, то случаи несовпадений бывают не такими уж редкими. Так, в исследовании 1995 г. российской прессы на предмет освещения ею лидеров и фракций Государственной Думы [11] процент совпадения решений кодировщиков по количеству материалов и абзацев с интересующими исследователя субъектами достигал 96 - 94%, в то время как для единиц, требующих идентификации текстового отношения к предмету внимания, совпадение достигалось лишь в 88% случаев.

В предыдущих разделах мы оперировали примерами из исследований прессы, радио и телевидения практически как равноправными. Это нуждается в некоторых пояснениях. Если представить информацию, идущую по каналам СМК, как передачу познавательной и эмоциональной структуры, то надо сказать, что контент-анализ имеет дело в основном с познавательной структурой именно потому, что она более «предметна», «объектна», следовательно, можно утверждать, что она и для анализа более доступна.

При том что теоретики радио и телевидения охотно отмечают как несомненно влияющие на эффективность восприятия такие специфические выразительные средства, как звук (шумовые эффекты) на радио, крупный план, ракурс и раскадровка на телевидении практики анализа содержания не включают (или редко включают) измерение этих вещей в свои программы.

Зависимость восприятия этих переменных может быть выяснена из экспериментов с приемником информации на психологическом уровне. Именно поэтому анализ содержания как метод, отмечая в каждом сообщении диалектическую связь чувственного и рационального, идет за рациональным, за мыслью.

Обратимся в этой связи к проблеме видеоряда, который помимо того что он состоит из планов, кадров и ракурсов воспроизводит на экране объективную реальность. Как общая методическая платформа для возможностей анализа этой реальности нашим методом может быть представлена такая практическая установка. Можно анализировать то, что поддается однозначной словесной интерпретации. Так, наличие на экране изображения какой-либо социальной реальности практически без трудностей поддается классификации в плане традиционного тематического членения.

Иногда такой анализ может существенно обогатить выводы исследователя. Например, в анализе телевизионной рекламы на трех ведущих коммерческих телесетях США, осуществленной в Анненбергской школе коммуникаций (Филадельфия, США), получено много информации - специалисту по рекламе покажутся значимыми и такие сведения о ней: в 63% случаев предлагаемый товар не показывается в телеролике, о нем только говорят, в остальных 33% он демонстрируется; статус пользователя, который мы можем увидеть на экране, в 95% случаев - это обычные люди, остальные - знаменитости; мужчины - 46%, женщины - 26%, вместе - 29%; возраст пользователя - ребенок (43%), молодой (13%), средний возраст (23%), старший (1%), смешанные ситуации (20%) [12].

Видеоряд в телевизионной рекламе вообще живет особой жизнью. Иногда даже «вещный» мир кино может восприниматься как реклама. И в этом ничего парадоксального нет. Героя в кино окружает современная ему жизнь. И почему бы в этой жизни не доминировать тем предметам, которые важны для рекламодателя? Действительно, коль скоро киноискусство воспроизводит «вещный» мир, было бы странно, если бы рекламодатели упустили такую блестящую возможность показать в этом мире именно ту «вещь», которую производят они.

Есть множество примеров, которые позволяют говорить об этом как о распространенной практике. Сидней Поллак, американский режиссер, автор фильма «Фирма» откровенно говорил Сергею Шолохову в его программе «Тихий дом», что появление в его фильме модели автомобиля «Линкольн» (если я не путаю марку) как свершения всех мечтаний героя не случайно. Между режиссером и автофирмой был заключен контракт, что по выходе фильм будет рекламироваться на всех автосалонах, где будет присутствовать эта модель.

Недавно газеты сообщили такую новость: оказывается, корпорация «Браун энд Уильямсон тобакко», выпускающая сигареты « Cool », платила актерам за то, что они курили на экране продукцию их фирмы. Клинт Иствуд и Пол Ньюмен получили машины, Шон О'Коннери - ювелирные украшения, Сильвестр Сталлоне - гонорар. Как сообщает журнал « Priier », в период с 1979 по 1983 гг. компания потратила около миллиона долларов на «подарки» кинозвездам-курильщикам. Сигареты этой марки фигурируют в таких фильмах, как «Рэмбо», «Рокки-4», «Внезапная коллизия» и «Лихорадка тела». Сталлоне курил эту марку в пяти фильмах, про Иствуда подробностей не сообщается, ведь он в образе ковбоя предпочитает сигары, но где он изменил им в пользу сигарет, неясно. По ходу дела выяснилось, что любимая марка Джеймса Бонда - это все та же любимица домохозяек. Вот этого англичане никак не могли простить своему национальному агенту 007 - Шону О'Коннери.

В 1990 г. американская киноиндустрия приняла даже решение не рекламировать больше табачную продукцию на киноэкранах. Не стали бы теперь из старых фильмов вырезать эти кадры, как сделали у нас в 1985 году, когда мы боролись по постановлению Партии и Правительства с алкоголизмом!

Еще один пример из моего собственного досье. Как-то по телевидению шел фильм-ужастик «Серебряная пуля». Там по ходу дела детишки борются с вампиром, которого, как утверждает местная легенда, можно убить только серебряной пулей. Сказано - сделано. Не пожалев фамильного серебра, брат с сестрой у местного же оружейника отливают такую пулю. И вот брат в роковую минуту мчится на мотороллере, то ли от вампира, то ли на встречу с ним. И что можно было прочитать на этом мотороллере - ну конечно « Silver Bullet »! Я с одной стороны чувствовала себя обманутой, а с другой восхитилась - надо же, целый фильм сочинили, и все во славу своей мотомодели!

Ясно, что такая форма телепродукции, как сериалы, не могла остаться в стороне от таких процессов по определению: во-первых, сериалы воспроизводят действительность максимально приближенным к ней способом (кино, например, в огромной степени отличается от других видов искусства именно по этому признаку), во-вторых, они собирают огромную аудиторию, по своему составу являющуюся заманчивой мишенью для рекламодателя (женская часть населения, на которой лежат преимущественно покупки для дома), в-третьих, к рекламе в телесериалах (до, после, и внутри) отношение более доброжелательное («сериальность» этого жанра, специальная драматургия, которая предусматривает места врезки телерекламы), чем появление в кинофильме.

Несколько лет назад по российским телеканалам прошел сериал «Моя вторая мама». По крайней мере в двух сериях главным действующим лицом был... океанский туристический лайнер с туром по Карибскому бассейну. Он доминировал над всеми сюжетными хитросплетениями и любовными интригами. Его было много. Герои были включены в самые разнообразные способы проведения времени на корабле: кино, танцы, бар, поездки на острова на лодках, загорание и демонстрация вечерних туалетов. Нужно ли говорить, что сам лайнер с огромной надписью его названия по борту всегда был, так сказать, в кадре, причем крупным планом. Я не удивлюсь, если окажется, что у себя на родине выход этого сериала способствовал резкому увеличению спроса на услуги, предлагаемые этим круизом.

Это обычная практика для телесериалов. В телесериале «Жестокий ангел» супруги-молодожены едут в свадебное путешествие, и супруг, в прошлом житель Португалии, по сюжету совсем не обремененный излишними знаниями, проводит настоящую экскурсию по столице, с наименованиями стилей архитектуры, датировкой строительства и т.д. Туристическая фирма, организующая поездки из Латинской Америки в Европу, несомненно, была в числе спонсоров фильма.

Конечно, такого рода примеры могли быть отслежены скорее традиционным путем - с помощью неспешного многолетнего наблюдения, с помощью собирания досье... Но все же и контент-анализ мог бы быть тут активным помощником. Было бы желание финансировать такой проект.

Здесь возникает проблема интерпретации видеоряда, когда заходит речь об оценках. Идея определить качество изображаемого объекта на оценочных весах «хорошо», «плохо» влечет за собой проблемы, которые не всегда просто решаются. Действительно, оценка - это свойство предмета, где на правах соавтора активно выступает сам оценивающий. Недаром существуют специальные психологические тесты, когда изображения одинаковых человеческих лиц предъявляются на оценку испытуемым (в данном случае - тестируемым) и они дают этим «героям», вернее, их портретам такие взаимоисключающие друг друга характеристики, что для психолога эти характеристики становятся предметом для выводов относительно самих испытуемых. Мы говорили уже о случаях, когда для выработки решения об оценке изучаемой характеристики приходится прибегать к «методу судей» и в контент-анализе.

Конечно, у изобразительного ряда есть свои возможности выржать чувства, вызывать оценки, достаточно однозначные у всей аудитории. Вспомним хрестоматийный киноведческий пример, когда изображение актера Мозжухина в кинокадре с тарелкой супа и то же изображение актера с гробом интерпретируется аудиторией по-разному: в одном случае актер воспринимается как комик, в другом - как трагик. Но для анализа содержания это слишком тонкие материи.

Чаще всего изобразительный ряд в СМК - на телевидении - живет по другим законам. Там он чаще всего (и тут могут быть проведены параллели между фото в газете) активно интерпретируется. Телекоммуникатор (в любой ипостаси - диктора, ведущего, репортера и т.п.) как раз старается снабдить визуальный ряд оценками, своими качествами. И вот уже в этом может быть обнаружено поле деятельности и для аналитика содержания. Не оказывается ли изображение и словесный комментарий к нему в конфликтном положении?

Хочется привести пример одного исследования влияния подписей под фотографиями на восприятие этих фотографий (аналогия с видеорядом и словесным комментарием к нему полная). В результате исследования автор получил такой набор возможностей сочетания характера подписи и самой фотографии:

1. Подпись значительно влияет на интерпретацию снимка, она может сформировать интерпретацию аудитории, коренным образом отличную от явного содержания снимка.

2. Именно содержание подписи может повлиять на то, чтобы интерпретация снимка произошла в нужном автору подписи направлении.

3. Иногда подпись вызывает реакцию реципиента, прямо противоположную той, которую вызывала своим содержанием фотография; если подпись предполагает в содержании фото смысл, несовместимый с его истинным содержанием, то она, как правило, игнорируется и интерпретация зрителя ориентируется только на снимок [13].

Проблема эта чрезвычайно интересная и разработка ее применительно к телевидению кажется весьма плодотворной.

Вообще контент-анализ телевизионных материалов помимо обозначенных проблем имеет и организационные проблемы, связанные с мимолетностью существования этой продукции в эфире. Как правило, если исследование оперирует большим числом переменных, не обойтись без предварительной консервации материала и неоднократного его прокручивания. Иногда исследование осуществляется непосредственно с эфира. Тогда непременным условием становится просмотр материала несколькими кодировщиками, обязательное протоколирование (хотя бы схематичное) эфира, и все равно количество характеристик, которые могут быть уловлены, уменьшается на порядки...

Таким образом, схематически процесс использования контент-анализа в конкретном эмпирическом социологическом исследовании можно представить так: формулировка задач исследования, определяющих выбор единиц наблюдения и разработку категорий анализа; операционализация последних, осуществляемая в инструкции кодировщику; перевод анализируемого текста в совокупность единиц анализа, подвергаемую счету; представление частотного распределения переменных; статистическая обработка и интерпретация результатов. Часть этого пути мы уже прошли.

**Машинный способ анализа текста**

Далее мы коснемся возможностей машинной обработки материалов способом контент-анализа. Из всех возможностей контент-анализа, которые суммированы в группах 1 и 2, ЭВМ можно поручить задачи анализа текстов, когда в качестве единицы анализа выступает слово. Когда мы говорим, что речь идет о достаточно ограниченном наборе слов, мы имеем в виду ограниченность с точки зрения человеческого мышления. Память же компьютера оперирует списком, состоящим из тысяч слов. Когда мы проверяем орфографию напечатанного текста, происходит операция сличения вашего текста с набором слов, уже введенных в память компьютера. И человеку, выступающему тут в качестве «машиниста», кажется, что число таких слов, введенных в память, беспредельно. Тем не менее оно «конечно», и машинный анализ текста возможен, когда в память ЭВМ заложены эти тысячи слов.

И тогда машина при обработке текстов ничем иным не занимается, как подсчитывает число материалов, где встречаются те слова, которые интересуют исследователя.

Такие операции находят все большее применение и предложениями - провести такой анализ - уже пестрят рекламные странички профессиональных журналов. Дело тут за малым: предусмотреть, например, возможности синонимического ряда понятий...

Более сложные случаи, когда речь идет о генерализации вывода, не обеспеченного симметричным рядом известных слов, пока неподвластны компьютеру, и на этот счет иллюзий пока питать не следует. Другое дело, что компьютер может выступать как машинописный лист фиксации результатов, к которым приходит кодировщик текста. Но и тут существует соображение контроля за этим выводом. Думается, что возможности предъявления такого анализа для контроля пока сдерживают применение компьютера в этом качестве.

Но вспомним, как все это начиналось. Группа ученых Гарвардского университета в США (рук. Ф. Стоун) в 1961 г. начала разработку принципов подобной машинной обработки, проведя серию анализов самых различных текстов: газет, сочинений членов малых групп, программных речей, произнесенных кандидатами на президентский пост от демократической и республиканской партий США, личных документов (писем, дневников, автобиографий) [14].

В рамках этого исследования было разработано несколько программ со следующими общими требованиями: компьютер отыскивает в тексте лингвистические аналоги тем категориям, которые интересуют и оговорены исследователем; подсчитывает частоту употребления этих категорий и их сочетания; выводит результаты в таблицы распределений; группирует текст в зависимости от употребляющихся там категорий и т.д.

Эти операции компьютер осуществлял только после того, как в его память были заложены языковые знаки (слова, идиомы, фразы), которые в своих группировках могут быть выражены на понятийном языке социальной теории исследователя и каждая из которых представляет из себя определенную переменную величину в его гипотезах. Для этого исследователь составляет словарь своего исследования: понятийному слову в качестве его расшифровки даны «носители» его смысла в бытующем в реальности языке.

Построение такого словаря аналогично построению «тезауруса» — языка определенной области человеческого знания, когда ключевым словам из этой области знаний соответствует синонимический ряд общеупотребительных слов. Таким образом, словарная статья в «тезаурусе» - это более крупный семантический блок, чем словарная статья обычного толкового словаря.

Обозреваемое нами комплексное исследование включает следующие самостоятельные анализы.

Исследование Д. Данфи связано с изучением малых групп, оно основано на анализе вербального поведения членов малых групп. Целью этого изучения было выяснение ролевых различий в группах, связей, отношения к фигуре формального наставника - все это во временном развитии. Материалом для анализа были сочинения испытуемых групп. Основными категориями были самоотождествление, величина группы, служащей для испытуемых референтной, действия и эмоции, институциональный контекст и т.д.

Группа исследований была связана с политическими текстами, в частности, анализ М. Смитом, Ф. Стоуном и Е. Гленн программных речей, произнесенных кандидатами на президентский пост от демократической и республиканской партии США в течение 36 лет (от Смита и Гувера до Джонсона и Голдуотера).

Исследователи Дж. Нейменвирс и Т. Брейер провели анализ статей, опубликованных в газетах The Times (Англия), Le Monde (Франция), Frankfurtur Allgieine Zeitung (тогда еще Западная Германия), The New York Times (США), на предмет того, как, по их мнению, обстоит дело с политической интеграцией западных стран в связи с существованием экономических и военных союзов между ними.

Процесс межгосударственной интеграции понимался как изменение следующих четырех элементов: характер принятия решений, касающихся судеб различных стран; обмен между странами почтой, студентами; торговля, иммиграция; мнения масс и элиты по поводу перспектив на этот счет. Показателями изменений в ориентации анализируемых престижных газет были выбраны внимание к национальным и региональным символам и степень тождественности в ориентации газет всех четырех стран.

Даниэл М. Огилви построил специальный словарь для проверки гипотезы о существовании так называемого «комплекса Икара», психологического явления, которое отмечалось по наблюдениям психологов за рядом своих пациентов. Огилви решил подвергнуть испытанию эту гипотезу, проанализировав текстовые материалы, созданные в рамках разных культур. В качестве такого материала были выбраны сказки - 626 сказок из 44 примитивных сообществ. Контент-анализ этих сказок был предварен созданием так называемого «Икарианского словаря», охватывающего 74 категории и 2500 слов, их наполняющих. Для разработки системы категорий использовались разные источники: работы исследователей, описывающих комплекс Икара, истории жизни людей, явно обладавших этим комплексом, тексту наблюдений за ними психотерапевтов и т.д.

Поскольку изначально комплекс Икара определялся как направленность на «полет, падение, огонь, воду, бессмертие и нарциссизм»,, гипотезу о существовании комплекса на уровне анализируемых сказок можно было, по мнению автора, считать доказанной, если эти темы окажутся «связанными» в конкретных единицах фольклора, или на определенном этапе окажется возможным по одним темам предсказать появление в анализируемом тексте смежных тем. Итог анализа состоял в расчете корреляций между 74 категориями и факторном анализе конечной матрицы.

С самых первых попыток использования электронно-счетных устройств для работы с текстом становились ясными преимущества и недостатки обработки текстового материала с помощью машин - они обеспечивали адекватность анализа огромных текстовых материалов, но требовали огромных усилий по составлению программы - собственно «словника», с учетом всех синонимических вариантов понятий, которые надо будет отыскивать в этом море пропускаемых через машину слов. Не случайно, что многие из тех анализов, которые были проделаны в Гарвардском университете, использовали категориальный аппарат ранее проделанных исследований и «обкатывали» его на компьютерах.

Для начала же авторы системы «Дженерал Инкуайерер» создали отперфорированные тексты с общим количеством слов, равным 6 миллионам. Из этих слов была произведена выборка в 511 тыс. слов. Тексты, попавшие в выборку, были просеяны через процедуру поиска ключевого слова. В результате осуществления этой процедуры исследователи получили распечатку всех случаев употребления интересующих их слов в выбранной совокупности текстов. На реализацию процедуры ушло шесть часов машинного времени, а полученная распечатка была толщиной в несколько десятков сантиметров. Одним из самых неожиданных результатов процедуры было то, что для определения смысла ключевого слова оказалось вполне достаточно нескольких слов, расположенных по обе стороны ключевого слова. То есть для ряда задач не нужно было иметь значения всего предложения целиком, включая анализ его синтаксической структуры. Но оставались еще местоимения, идиомы и т.д.

Хотя первые словники создавались исследователями под конкретные задачи и главным тут была идентификация ключевых слов, сами авторы отлично понимали перспективы развития компьютерного дела - их размышления о том, что наборная клавиатура, связанная с компьютером, будет обычной принадлежностью школьного класса и делового офиса учителя, бухгалтера, психотерапевта, библиотекаря (ясно видно, что эти размышления относились к дореволюционной эпохе - до революции, когда был изобретен персональный компьютер), звучат сегодня как предвидение. Но для того чтобы выполнять все эти ожидаемые операции, нужно было, чтобы компьютер мог различать не только слова, но и смысл более пространных языковых единиц. Для этого нужно было решение проблемы синтаксического анализа. К сегодняшнему дню частично такие проблемы решены, коль скоро мы доверяем своему компьютеру - при наличии в нем специфических программ - проверку ошибок набранного текста.

Отметим существенный момент в разработке методологии анализа текста с помощью машины (что помимо всего прочего означало движение в направлении общения человек-машина): как только исследователи вышли на изучение диалога, они осознали, что эта проблема выводит их за пределы узкоспециальных «ведомственных» словарей...

С тех пор машинный метод обработки текстовых массивов все больше находит себе применение. При чем мы имеем в данном случае не те чисто прикладные случаи, когда машина помогает отыскать нужную вам научную литературу по ключевым словам, которые такая литература имеет заранее в виде своеобразной паспортички (или патент изобретения, если вы хотите проверить, не изобрел ли кто уже предлагаемый вами велосипед); или практику, по которой работает Международный междисциплинарный Индекс публикаций ( Citation In dex ) - индекс представляет из себя распечатку статей из 7000 журналов, издаваемых во всем мире, по лицам, упоминаемым там. Каждое упоминание лица классифицируется по источнику: является ли оно автором статьи, упоминается ли в ходе дискуссии, появляется ли его имя в рецензии, в библиографии и т.д. Более детальный вариант этого индекса по социальной проблематике содержит роспись всех статей по ключевым словам.

Так, выпуск 1984 г. содержит данные о 1000000 статей из этих журналов, а поскольку форма его выхода - алфавитный список упоминаемых фамилий - то это составляет 10,5 млн. ссылок. К примеру, в течение нескольких минут мы можем (а этот Индекс есть в Государственной российской библиотеке) определить, какие статьи по этой проблематике напечатаны во всем мире в период, например, мая - августа 1985 г. Оказалось, что за это время было опубликовано 7 статей (все в американских журналах): три посвящены исследованиям речи лиц с психопатологическими изменениями, статья У. Миллса по содержанию китайской пропаганды, статья К. Уинника «Контент-анализ журналов с сексуальной тематикой». Все упоминания снабжены указанием названия журнала, тома, номера и страниц.

Уже через десять лет после этого пионерского проекта Ф. Стоуна и его коллег практически повсеместно стали осуществляться проекты контент-аналитических исследований с применением ЭВМ. На состоявшемся в 1974 г. в Италии рабочем совещании по проблемам контент-анализа было представлено несколько таких проектов, в частности проект международного исследования газетных заголовков с задачами определения внимания различных газет к местным, общенациональным и международным событиям, сравнения внимания американских и европейских газет к проблемам «общего рынка», сравнения освещения гражданской войны в Нигерии газетами разных стран и др. Германия была представлена на этой конференции проектом по созданию специализированного словаря для целей анализа содержания текстов [15].

Как ни странным это покажется на первый взгляд, именно пример с машинной обработкой текстов иллюстрирует очень важную для понимания сущности контент-анализа мысль. И в другом месте учебника эта мысль не прозвучала бы так явственно. Анализ содержания как метод не обладает магическими качествами - вы не получите из него больше, чем вложили в него. Если нечто значительное, важное, необычное не предусмотрено процедурой, то оно не появится в результате анализа, каким бы сложным и кропотливым он ни был.

Говоря о машинной обработке текстов, мы должны уточнить, что сейчас обсуждали случаи, когда компьютер оперирует непосредственно текстом. Гораздо более часты случаи, когда ЭВМ используется, грубо говоря, как арифмометр, оперируя с введенными в него кодировочными карточками, которые уже содержат итоги наблюдения за текстом, осуществленного исследователем. Сейчас это уже распространенная практика.

**Выбор единиц наблюдения**

Даже только начинающие овладевать методом анализа содержания исследователи начинают с вопроса: сколько текстов надо исследовать, чтобы исследование считалось корректным? Мне даже кажется, что они не до конца осознают важность этого вопроса - а важность в ответе, которым должно начинаться каждое исследование, где автор заявляет: «Я изучил такое-то количество текстов. Этого достаточно, чтобы сделать выводы о таком-то периоде деятельности источника». Просто начинающего исследователя берет легкая оторопь от осознания сложности, тщательности, пунктуальности, трудоемкости метода - подчеркиваем, что эти характеристики должны быть присущи методу всегда - и он задает этот вопрос, чтобы получить логичный ответ, насколько минимальной должна быть выборка изучаемых текстов, чтобы по возможности максимальным был период, на который мы распространяем свои выводы.

Вопрос этот вполне логичен. Методология многих наук основана на том, что целое описывается по его части.

Вспомним, что, например, такая отрасль социологии, как зондирование общественного мнения, смогла встать на индустриальные рельсы не раньше, чем для институтов, специализирующихся на таком зондировании, был решен вопрос с научной выборкой. Стали известны законы отбора отдельных индивидуумов, чтобы сравнительно неболшое их число - в случае, например, с американским институтом общественного мнения Дж. Гэллапа, это 1500 - 2000 американцев - позволило распространять данные опроса на все население США. Говоря научным языком, с разработкой научной теории выборки была решена проблема, как добиться «надежных обобщений при интенсивном изучении относительно небольшого числа случаев». Реально выборка представляла возможность экономично, при большой скорости проведения работ, профессионально провести опрос, который зачастую дает более верную информацию, чем сплошное обследование.

Кстати, вспомним, что на заре исследований общественного мнения именно выборка подвергалась поистине уничтожающей критике оппонентов изучения общественного мнения. Обыденное сознание никак не могло согласиться с тем, что достаточно опросить всего 1500 человек, чтобы знать, что думают по определенному поводу 150 млн. человек.

Главное, что мы должны себе представить, это то, что существуют различные типы выборок и что далеко не все из них сможет применить на практике контент-аналитик.

Не последняя тут проблема - доступность материала. Хотя библиотеки многих стран уже давно озаботились проблемой сохранности продукции СМК для потомства, пробелов здесь множество. Как правило, хранятся только самые престижные газеты [16]. Особенно остро стоит эта проблема по отношению к продукции радио и телевидения. Это может быть естественным ограничителем при общей разработанности технологии выборок.

Так, существует большой класс выборок репрезентативных, т.е. таких, которые претендуют на то, чтобы быть представительными для более широкой совокупности. Этот класс можно разделить на две неравные группы по способу отбора единиц уже для конкретного исследования. Большая часть - это так называемые квотные выборки (или многочисленные модификации их: районированные, стратифицированные, многоступенчатые, послойные и т.п.). При составлении их нужно обладать предварительной информацией обо всей совокупности единиц, из которой производится отбор. В случае, например, с конкретной страной нужно знать распределение по количеству населения в каждом регионе, причем надо знать распределение этого населения по полу, возрасту, образованию или другой характеристике, которая, по мнению исследователя, влияет на то конкретное поведение, которое он изучает - например, потребление конкретных средств массовой коммуникации.

Понятно, что для аналитика содержания такая возможность отпадает. Всякий раз исследователь начинает с «белого листа». Ему как раз и надо знать распределение характеристик текста - это не те достаточно постоянные характеристики населения, как пол, возраст, и прочие, которые нужны при определении выборки в исследованиях аудитории. Действительно, они для определенного региона могут считаться постоянными. Для больших совокупностей людей меняются мало, а если меняются, то эти изменения тут же становятся известными соответствующим статистическим органам, поскольку это их работа, которая заканчивается тем, что они сразу же информируют социум об этих изменениях.

Существуют и другие выборки в классе репрезентативных - это вероятностные, или случайные выборки. Их построение основывается на том факте, что если для каждой единицы генеральной совокупности, например, населения всей страны, будет выдерживаться равновероятная возможность попасть в число отобранных для исследования, т.е. конкретных людей будут отбирать случайным образом, то такая выборка будет отвечать высшему критерию представительности - она будет репрезентативна для всего населения страны.

Мы повели разговор о репрезентативной выборке, а именно о ней идет речь, когда встает вопрос, в какой мере выводы исследования по ряду случаев приложимы (характерны, репрезентативны) ко всему явлению в целом, идя на поводу у неискушенного исследователя-аналитика. Репрезентативная выборка представляет лишь один из двух типов выборок, которыми практически (в принципе) исчерпываются все случаи выборок в социальных исследованиях. Другой класс выборок - типологическая выборка - в ходе исследования говорит нам, что наблюдаемое нами в обществе явление есть, оно при этом имеет определенные, зафиксированные в ходе нашего исследования характеристики (хотя может этими характеристиками и не исчерпываться).

Говоря о репрезентативных выборках, мы остановились на том, что один из подвидов их - и только он один - применим в контент-аналитических исследованиях.

Но здесь вероятностная, или случайная выборка - а это как раз тот самый подвид - приобретает некоторые особенности, с которыми специалист не может не считаться.

В самом деле - в случае с исследованиями общественного мнения мы имеем дело с объектом, протяженным в пространстве, скажем, с населением страны. Как ни сложно оперировать с таким объектом (или, как говорят социологи, с генеральной совокупностью исходных единиц), социологи, установив определенную ступенчатость отбора для разных характеристик, обеспечивают для каждого человека, входящего в эту генеральную совокупность, искомую равновероятную возможность попасть в выборку.

В случае с исследованиями содержания прессы мы имеем дело с объектом, протяженным во времени. Что тут считать генеральной совокупностью? Всю совокупность текстов с первого дня существования газеты? Десятилетие? Пятилетие? Год? Другими словами - если мы возьмем эмпирическую выборку Гэллапа в 1500 человек за образец и будем строить свою выборку текстов в 1500 единиц (пока отвлечемся от того, что считать тут единицей: это с людьми все ясно, а с текстом, как говорится, возможны варианты - это могут быть отдельные дни/номера газеты целиком или отдельные материалы), то реально мы должны промерить все тексты гигантским циркулем, «шаг» которого будут составлять эти 1500 единиц, от сегодняшнего дня в прошлое. Сделаем пять этих гигантских шагов (или двадцать пять), а потом пойдем в обратном порядке - возьмем в отмеренном пространстве каждую пятую единицу (или двадцать пятую) и получим идеальную случайную выборку.

Но - где остановиться циркулю? В пятой точке или двадцать пятой? Это всегда произвольное решение исследователя. Иначе говоря — определение границ генеральной совокупности, из которой будет произведена выборка - это авторитарное решение исследователя и никто ему этого подсказать не сможет. На его выбор должно влиять одно тактическое соображение - его выборка репрезентативна для всего выбранного временного интервала, но не для отдельного периода внутри этого интервала. Другими словами, если для исследования выбраны 1500 текстов газеты с определенным интервалом внутри периода 1963 - 1966 гг., то результаты будут относиться ко всему этому периоду, но не к маю месяцу 1964 г. (точно так же выбранные случайным образом 1500 опрошенных по России демонстрируют мнение, репрезентативное в целом для России, но не для Красноярского края - такая задача потребует увеличения выборки, при чем все на тех же началах отбора уже по Красноярскому краю).

Каковы должны быть эти временные границы, повторяем, это воля исследователя. Ответ на этот вопрос теснее всего связан с программой исследования, т.е. с его задачами. Ясно, что логика определения этих временных границ должна быть одной для выяснения воздействия текстов на аудиторию, процесса, который по самой своей сути является долговременным, и другой - для снятия картины деятельности источника, которая может быть и одномоментной; одной - когда само исследование замышляется для того, чтобы внести коррективы в сегодняшнюю деятельность источника, но она может быть совершенно другой в случае с историко-сравнительными исследованиями.

Во всех случаях, когда речь идет о долговременном процессе, повышается необходимость репрезентативных выборок со случайным отбором единиц наблюдения. Вспомним исследование, связанное с изучением языка идеологий в мировой прессе (см. с. данной работы). Для анализа брался каждый номер престижных газет нескольких стран, вышедший первого и пятнадцатого числа каждого месяца.

Случайность этого отбора уравнивала источники с точки зрения частоты употребления в них политических символов. Поясним свою мысль - определенная газета могла продемонстрировать первого числа какого-то месяца в данном пятилетии необычную даже для нее «густоту», частоту обращения к политической тематике, но эта же газета пятнадцатого числа другого месяца (в другом пятилетии) по причинам того, давно ушедшего пятилетия, абсолютно не касалась политики: она вышла под лозунгом «день спорта». Но в пределах одного-двух-трех десятилетий, на которые распространялся анализ, любая другая газета могла быть точно в таком же положении в другие первое и пятнадцатое числа.

В этом смысле справедливо замечание, что большой временной отрезок для отбора и сам механизм случайного отбора (а это мог быть каждый второй и шестнадцатый номер газеты и т.п.) уравняли источники и в этом смысле сделали надежными выводы исследования: они были характерны для всего анализируемого периода деятельности источника.

Такой случайный отбор уравнял газеты еще в одном отношении. Известно, что в газетах разных стран существенно разнятся номера в зависимости от дня недельного цикла. Так, например, американские ежедневные газеты имеют различную толщину в разные дни недели. Это зависит в основном от количества рекламы, размещаемой на страницах газет, а она в свою очередь отражает динамику привычек покупателей в течение недели. И тогда выпуски газеты в четверг имеют одну толщину, а субботние - другую. Когда газеты отбираются на большом временном отрезке случайным образом, возрастает вероятность того, что в выборку попадут номера газет разных дней недели.

Когда исследователя интересует современная ему деятельность источника и он хочет ограничиться заведомо небольшим временным периодом для анализа, он должен специально учесть этот фактор. Так, известно, что каждая газета имеет сменные тематические рубрики для каждого дня недели, телевизионные каналы приурочивают особые развлекательные передачи к концу недели и т.д.

Чтобы отразить в выборке недельную цикличность, столь характерную для деятельности СМК, у аналитика содержания есть уже опробованный метод «конструирования» недели (авторы этого метода - Роберт Джоунс и Рой Картер). По этому методу в годичном выпуске газет отмечаются все выпуски в понедельник, вторник и т.д. Затем берется каждый десятый выпуск из числа выпусков в понедельник, каждый десятый выпуск из выпусков во вторник и т.п. Сконструированная таким образом неделя будет репрезентировать весь год, выбранный для анализа. Надо сказать, что этот вывод не голословен: исследователи, рекомендовавшие такой подход к отбору единиц для анализа, в ходе своих сравнительных исследований показывали, что точно такие же результаты получаются, если анализировать весь материал за год сплошь.

Интересный вариант выборки с использованием такой недели содержит исследование Ч. Окигбо (отделение массовых коммуникаций, Университет Нигерии).

После отбора четырех газет - объекта исследований, репрезентирующих разные формы собственности, разные политические направления: частное предприятие, наиболее элитарную газету, государственную собственность/контроль, газету - собственность управления штата -формы, которые, по мысли автора, в значительной мере определяют отбор новостей (механизм отбора новостей был предметом исследования), автор использовал следующий механизм построения выборки. На первом этапе из годичной подшивки каждой газеты за 1986 г. (52 недели) методом случайного отбора была взята одна сплошная неделя для каждой газеты - это была неделя с понедельника 10 марта по понедельник 17 марта 1986 г. Причем, поскольку воскресные номера газет в этой стране скорее напоминают еженедельник (большой объем, множество редакционных материалов и аналитических статей), они были изъяты из анализа. К этим семи номерам была прибавлена сконструированная неделя (случайным образом был отобран понедельник из всех понедельников и т.д.; в итоге в выборку попал понедельник 14 июля, вторник 9 сентября, среда 14 мая, четверг 28 августа, пятница 7 февраля, суббота 26 июля и понедельник 20 октября). Таким образом, исследование было осуществлено на 14 номерах анализируемых газет [17].

Но начинающий аналитик содержания, как и его оппоненты, должен отчетливо сознавать, что в таком случае выводы его исследования относятся к этому обозначенному как основание для выборки году, но не к каждой неделе в отдельности.

И все-таки вопрос «сколько» остается. И здесь мы опять обращаемся к изначальным задачам исследования: если мы анализируем разнородный материал, например, совокупный теле- или радиодень с их разнообразием рубрик или весь номер газеты, где есть информационные и очерковые материалы, редакционные статьи и теоретические «подвалы», справочные материалы и уголок юмора, надо выбирать больший объем; если в этой совокупности мы имеем дело с отдельной передачей или жанровым куском газеты как с объектом анализа, мы ограничиваемся меньшим объемом.

И здесь самое место «закрыть» еще один вопрос: о единицах, которые мы выбираем - дни (выпуски) или отдельные материалы. Согласимся, что разница существенная - проанализировать 1500 газетных выпусков или 1500 отдельных материалов. Исходя из тактики случайного отбора, это должны быть отдельные материалы: только они образуют совокупность, уменьшение которой в тысячу или в десять тысяч раз (как в опросах общественного мнения) делает посильным, а значит, и осуществимым анализ текстовой продукции.

Но тактика вероятностного случайного отбора требует, чтобы исходные единицы были, образно говоря, хорошо перемешаны [18]: единицы должны иметь равновероятную возможность попасть в выборку, это обязательное условие такого отбора. Теоретики вероятностного метода приводят в качестве примера в таких случаях урну, в которой перемешаны разноцветные шары или лотерейные билеты.

Если обсуждать с этой точки зрения текстовую продукцию СМК, то мы убеждаемся, что она существует в виде устоявшихся многоступенчатых структур: каждая полоса газеты имеет сложившуюся структуру информации (по тематике, по локальности, по расположению официальных материалов и справочных документов). Даже в пределах более простого случая, например одной передачи на телевидении, сложились определенные модели сосуществования отдельных материалов - как самый красноречивый пример такого рода - программа «Время».

В таком случае отбор отдельных материалов может привести к значительным систематическим искажениям. Как правило, контент-аналитики, оперируя сравнительно небольшим искомым числом отдельных материалов (сопоставимым с вышеуказанными 1500 - 500 единицами), случайно отбирают их в пределах одного теледня или одного выпуска газеты. Примеры со сконструированной неделей, о которой мы говорили, или с отбором каждого первого и пятнадцатого номера каждой газеты на протяжении десятилетий в исследовании языка политической пропаганды могут служить тут иллюстрациями.

Все вышеизложенные соображения о сложности обоснования выборки в контент-аналитических исследованиях объясняют, почему на практике за всю историю применения этого метода исследователи - авторы одноразовых, эпизодических «замеров», «зондажей» содержания СМК объясняют свою выборку на уровне здравого смысла: доступностью единиц наблюдения, возможностью изучить данный объем в короткий срок и т.д.

Практически за этим стоит то, что аналитики всякий раз имеют - за редкими исключениями - дело с типологическими выборками: они скрупулезно констатируют, что в деятельности источника есть в достаточно короткий, точно ими фиксируемый отрезок времени. Социологи знают, что все социальные исследования чрезвычайно чувствительны к фактору времени. Тем не менее, поскольку - как мы можем судить из ретроспективного анализа применения контент-анализа в мировой социологии - границы этого отрезка времени подвижны, мы можем сформулировать несколько методических правил, которые позволят начинающим аналитикам содержания определиться со своей выборкой.

Прежде всего исследователь должен учитывать реальную периодичность, цикличность, ритмичность в деятельности анализируемых СМК или отдельных газет, радио- и телеканалов. Ясно, например, что при анализе передачи, которая выходит в эфир один раз в месяц, и которая выходит несколько раз в день (информационные выпуски), мы должны будем предусмотреть - если нас интересует их сравнение - такой промежуток времени, чтобы обе эти передачи были представлены в нем равновесомо.

Если мы сравниваем ежедневную общенациональную газету с районной, которая имеет другую - более редкую - периодичность, соответственно, номеров для анализа ежедневной газеты мы можем брать меньше. Поможет принять тут правильную тактику такое методологическое объяснение, что чем чаще воспроизводятся в деятельности источника какие-то характеристики (а комплекс определенных характеристик программы «Время» воспроизводится, например, ежедневно), тем меньше единиц для наблюдения за этим источником нужно выбрать по сравнению с другим источником, характеристики которого воспроизводятся реже.

Если мы интересуемся какой-либо одной характеристикой в деятельности источника, то мы при определении выборки руководствуемся следующими соображениями. Рассмотрим их на примере программы «Время», на характеристике «Обоснование коммуникатором выбора факта». Допустим, что мы на основании пробного пилотажного исследования, предварительного знакомства с этой телевизионной информационной программой убедились, что существуют такие виды этого обоснования:

1 - сюжеты с лексической маркировкой времени события;

2 - сюжеты с лексической маркировкой качественных признаков события и факта;

3 - сюжеты без лексического обоснования выбора факта.

Начиная отслеживать материал (точка отсчета, начало отсмотра произвольно задается самим исследователем в зависимости от его интереса к определенному временному периоду), мы фиксируем все случаи появления разных видов «обоснования выбора факта». Как только появилась наиболее редко встречающаяся характеристика, мы можем закончить наше микроисследование. Мы получили результаты классического типологического исследования: мы получили количественное распределение массива информации по одной интересующей нас характеристике. Оказалось, что нам понадобилось проанализировать для этого 6 - 7 телевыпусков программы «Время» (табл. 3.3).

Таблица 3.3

Распределение сюжетов программы «Время» по обоснованию выбора факта за май - июнь 1985 г. (в абс. цифрах)

|  |  |
| --- | --- |
| Сюжеты  | Количество сюжетов  |
| А. С лексической маркировкой вр6мени события  | 98  |
| Событие произошло сегодня  | 71  |
| Событие произошло вчера или ранее этого срока  | 3  |
| Юбилейные даты со словом «сегодня»  | 16  |
| Событие состоится в будущем  | 8  |
| Б. С лексической маркировкой качественных признаков  | 22  |
| Новизна  | 16  |
| Лучшие образцы  | 3  |
| Сложности, трудности  | 2  |
| Важность, ценность  | 1  |
| В. Без лексического обоснования выбора факта  | 41  |

Если бы нас интересовала частотность появления в эфире областей, краев, республик России, можно было бы заведомо утверждать, что в эфир пройдет гораздо больше выпусков «Времени», пока в поле внимания коммуникатора, а значит, и исследователя, попадет наименее часто упоминаемая территориальная единица.

С учетом такого поведения реальных характеристик содержания исследователь, включающий в одну программу исследования до десятка признаков, берет достаточно большой отрезок времени, чтобы подстраховать себя. В случае с ежедневной газетой, информационной программой и т.п. это, как правило, два-три месяца.

Если исследователя заинтересует какое-либо конкретное событие, происходящее сегодня в мире, и информационная политика нескольких источников по освещению этого события, то потребуется сплошное наблюдение за всеми источниками в сроки, максимально совпадающие с самим событием. Если мы вспомним исследование газеты «Истинный американец», то пропуск какого-то номера за исследуемый период дал бы великолепную возможность адвокату подследственных оспорить выводы экспертов. Или вообразим, что мы исследуем освещение ведущими телеканалами России кризис в Косово. Ясно, что исчерпывающий ответ на вопрос, как освещали этот конфликт ОРТ, РТР, НТВ и т.д., мы получим, если в поле зрения попадет весь временной отрезок этого конфликта.

Следует предусмотреть чисто организационные моменты. Сегодня нашелся заказчик, который заказывает вам исследование информационного источника, и как всегда в таких случаях, результаты ему требуются «вчера». Сколько вам понадобится времени, чтобы разработать программу исследования, осуществить собственно анализ, написать комментарий и сдать отчет заказчику? Тогда проблема срока наблюдения за объектом приобретает совсем другой характер. И ничто тут не заменит опыта исследователя, а также его знания всех элементов, из которых состоит исследование. К следующему такому элементу мы и переходим.

**Полевой этап**

Каждое социологическое исследование венчается результатом. Но ему предшествует большая работа, собственно социологическое «поле» - сбор информации, обработка, и только после этого - написание отчета. Чаще всего именно последняя стадия становится известной широкой научной общественности и лишь отдельные фрагменты предшествующей работы попадают в поле ее зрения. Методические работы, воспроизводящие и объясняющие инструментарий исследования, его средства и способы добывания научного знания рассчитаны, как правило, на более узкую аудиторию профессионалов, издаются малыми тиражами, быстро становятся библиографической редкостью.

Вот почему возникает необходимость в пределах этого пособия обсудить некоторые проблемы профессиональной «кухни» контент-аналитика. Это не заменит специальной работы для каждого отдельного исследователя, когда он окажется на этой стадии - скорее, это будет акцент на самом важном для этого этапа.

Формулировка задач, которые поставил перед собой исследователь, тесно увязывается в самом исследовании с теми способами получения социологической информации, которые мыслятся самим автором как наиболее адекватные. Образно говоря, социологический инструментарий - это та узда, которой сдерживается воображение социолога. Ставя перед собой определенную задачу, он должен примерить ее на способы решения.

Далее в рассуждениях автора должен появиться сам объект исследования: исследователь описывает этот объект - информационный источник в определенных границах его деятельности, обосновывая отбор единиц наблюдения. Описание выборки мы осуществили в предыдущем параграфе этого раздела, но не исключено, что на реальном исследовании стоит прибегнуть к консультации математика-специалиста по теории выборки. Следующий шаг - составление кода и инструкции к нему. Эти документы разрабатываются в тесном взаимодействии со специалистом, который будет осуществлять обработку полученной информации на ЭВМ. Чем в большей степени построение документа, на котором будет фиксироваться исходная информация, будет соответствовать правилам ввода информации в компьютер, а также правилам ее дальнейшей агрегации, тем меньше затрат вам потребуется на этой стадии.

После составления кода требуется специальная стадия - проверка кода на соотносимость результатов ( reliability ). Как правило, эта стадия состоит в том, что исследователь предлагает для анализа по данному коду один и тот же текст нескольким кодировщикам и проверяет, однообразно ли они используют код и насколько сходятся их результаты [19]. Другой вариант проверки - повторно дать закодировать этот же текст нескольким кодировщикам по истечении определенного промежутка времени. Цель в данном случае та же: посмотреть на идентичность анализа, так сказать, «во времени и в пространстве».

Все необходимые поправки и изменения, которые еще можно будет предусмотреть в коде и в инструкции к нему, следует делать имено на этой стадии. Как только начнется само исследование, то есть работа большого коллектива кодировщиков над текстами, вошедшими в выборку, поправки будут означать сбой в работе. Инструкции уже напечатаны, кодировщики проинструктированы - работу эту придется начинать заново.

Пока же «снимаются» все вопросы кодировщиков, происходит коллективное обсуждение реальной кодировки, оттачиваются примеры, которые должна содержать инструкция [20]. Так, в исследовании российской прессы на предмет ее отношения к лидерам и фракциям Госдумы, на которое мы уже ссылались, инструкция содержала примеры конкретных текстов из реальных газет, содержащих разные оценки: «Пока же преимущество в искусстве ведения судебного спора явно на стороне президентско-правительственной команды. Во-первых, она хорошо подобрана и в ней четко распределены роли. Впереди официальные представители - Сергей Шахрай, Юрий Батурин и Олег Кутафин. Все трое - юристы со степенями, хорошо освоившие приемы публичного спора. И психологическая уравновешенность у них тоже на высоте». («Авторов Чеченской республики пока не нашли» // Известия, 13.07.95) - позитивная оценка С. Шахрая, как лидера определенной фракции ;

«Кроме Егора Гайдара, предстоящие выборы могут принести большое разочарование и Жириновскому и его ЛДПР. Некогда мощнейшей партии сегодня симпатизирует лишь 6,6 процента голосов. Очевидно, имиджу Жириновского не помог пролитый в сражении с Б. Немцовым апельсиновый сок. Да и в целом скандальностью наших людей теперь удивить крайне трудно, хотя тяга к ярким личностям остается». («Аркадий Вольский встретился...» // Комсомольская правда, 01.07.95) - негативная оценка Жириновского и его фракции;

«Иван Петрович Рыбкин символизирует интересную тенденцию в российской политике. Его опыт, предшествовавший посту председателя Госдумы, оказался уникальным. Он хорошо знал законы аппарата, обрел незаурядные качества публичного политика в оппозционном состоянии. Наконец, это человек природного ума. Рыбкин внешне открыт и опрощен, но внутренне у него есть та российская хитринка, когда "сказка ложь, да в ней намек". Полтора года в логове Госдумы Иван Петрович осуществлял психотерапевтические функции и занял уникальную позицию. Сегодня он представляет интересы многих социальных групп в кремлевско-президентском властном режиме. Но надо же было такому случиться, чтобы Президент Ельцин двумя словами всю эту изнурительную работу опрокинул». (Интервью Г. Бурбулиса «Какофония новой власти» //Аргументы и факты. 1995. № 28) - сбалансированная оценка И. Рыбкина и негативная -Госдумы.

Мы специально привели эти примеры - установки руководителя исследования здесь недостаточно и некоторые кодировщики оценили второй пример как сбалансированную оценку В. Жириновского, а третий пример как позитивную оценку И. Рыбкина. Недаром в итоге оказалось, что степень совпадения мнений кодировщиков в той части иследования, которая касалась фиксации оценок, равнялась 88% (76%, когда речь шла о сбалансированных оценках, 89%, когда речь шла о позитивных оценках, 91% - нейтральная подача, 92% - негативные оценки).

Чтобы представить себе детализацию проблем в тексте инструкции, приведем начало инструкции к этому исследованию: «Кодировщик читает всю газету целиком, чтобы не пропустить упоминания интересующих исследование Субъектов/Объектов анализа, которые перечислены на карточке для кодировки. Кодировка начинается с того, что выписывается дата и номер газетного выпуска и число материалов с упоминаниями. После того как на отдельном листке выписаны заголовки всех материалов, где встретились эти упоминания, - чтобы не упустить ни одного из них, - приступаем к анализу каждого из этих материалов. Для этого требуется более скрупулезное, внимательное, чем на первой стадии, чтение каждого из этих материалов». И далее подробно объясняется каждая характеристика, даются примеры, объясняются способы фиксации результатов (если обработка результатов ручная) и т.д.

Если пробный анализ показывает недостаточную степень совпадения (американские авторы допускают в некоторых случаях степень совпадения в 60%), следует выяснить причины - это могут быть как механические ошибки (пропустил, не отметил и т.д.), так и содержательные расхождения - закодировано неправильно. Поэтому пилотный анализ должен предусмотреть проверку каждого материала - кто из кодировщиков допустил ошибки.

Более сложная процедура оценки работы с кодом - это проверка результатов на их обоснованность ( validity ) [21]. Общая идеология, связанная с проблемой обоснованности в социологических методах (в специальной литературе можно встретить и языковую кальку «валидность»), выражается в том, что исследователь предоставляет проверяемые доказательства, что именно, какую характеристику он измеряет.

Хрестоматийным примером в таких случаях является ссылка на такой измерительный прибор, как часы, измеряющие хронологическое время. Исправные, не останавливающиеся часы - надежный инструмент регистрации времени. Если же мы не знаем, по какому часовому поясу выверены эти часы, то их точные показания для нас бесполезны: мы не знаем, какое время, что именно они измеряют. Часы как инструмент не обоснованы [22].

Результаты анализа будут считаться обоснованными, если методика анализа замеряет в массиве текста то, что предполагалось замерять. Американский специалист в области контент-анализа У. Даниельсон приводит на этот счет такой пример. Допустим, вы фиксируете заметки, которые вы определяете как информацию о преступлениях, но действительно ли эта информация помещается для того, чтобы повествовать о преступлениях? Исследователь, отвечая на вопрос, должен обосновать свое определение преступления и такой заметки. Ему придется сказать, что в эту группу он относит происшествия, которые расцениваются юридическими властями, или общественностью, или, наконец, прессой как криминалистические по природе, а не как шалости подростков или результат паники. Иногда приходится прибегать к соображениям здравого смысла по принципу «широко известно...». Главный принцип - чтобы не было большого разрыва между категориальной сеткой и эмпирической действительностью.

Так, в ходе многочисленных исследований прессы мы убеждались, что при определении оценок, отношения автора текста к описываемым им событиям недостаточно оперирования только качественными прилагательными «хороший», «плохой». Про одного героя он напишет, что тот вытащил ребенка из проруби, или помог старушке перейти улицу (был у нас в отдаленные времена такой стереотип «хорошего» пионера - он обязательно поможет старушке перейти улицу), а про другого (скажем, депутата В. Жириновского), что тот в политической дискуссии облил своего коллегу апельсиновым соком или просто покажет это. А потом еще раз покажет... Средство не из слабых.

Нельзя было игнорировать эту мощную оценочную палитру. Поэтому в ходе многочисленных исследований, которые я цитирую и в которых выступаю как автор методики и интерпретации результатов, используется в инструкции кодировщикам такое требование-объяснение: «Оценка формируется как с помощью "сильных" оценочных эпитетов, обнаруженных в тексте, так и самим содержанием ситуаций, в связи с которыми появляется в тексте интересующий нас субъект или группа: это могут быть как положительные ситуации, с которыми традиционно, в массовом сознании ассоциируется "благо ", так и ситуации, с которыми ассоциируется "зло "».

Именно здесь мы должны упомянуть характеристику нашего метода, которая нами еще не привлекалась для его описания. В наборе характеристик метода мы находим такое ограничение, что анализ работает только с «открытым» ( manifest ) содержанием [23].

Обычно в качестве примера этой характеристики используют как объяснение на уровне здравого смысла, так и в сугубо научной парадигме. Начнем с первого. Фраза, произнесенная женщиной в адрес своей соперницы: «Какая чудная шляпка на вас!» - может быть расцеена мужчиной так, как будто о шляпе говорилось, что она выше всяческих похвал, красивая и элегантная. Женщина же скорее заподозрит, что при этом в шляпе хотели подчеркнуть, что она «дешевая, немодная, аляповатая, и скорее бы подошла юной девушке». И хотя аргументация черпается тут из области психологии, контент-анализ не может работать на значениях и смыслах второго, скрытого плана, он будет оперировать содержанием, что называется, «на поверхности».

Действительно, если представить множество жанров на протяженности по мере того, насколько адекватно они воспринимаются потенциальной аудиторией, одинаково понимаются, то информационная заметка из СМК окажется на одном конце, а заумная мистическая поэма - на другом (поскольку вариабельность прочтения ее разными людьми возрастает многократно). Анализ содержания работает как раз с теми текстами, где вероятность одинакового понимания смысла очень высока. И это не случайно. Содержание СМК потому (в том числе) и является массовым, что оно рассчитано на максимально возможное одинаковое понимание. В этом его природа. И анализ содержания как методика его изучения «снимает» именно этот уровень значений. Все остальное для него - от лукавого.

Именно поэтому и нужны тесты на обоснованность кода - что понимается разными кодировщиками под значением текста, и что говорит нам об алгоритме «декодирования» этого значения предлагаемая для исследования инструкция.

Один из тестов на обоснованность результатов, который, судя по специальной литературе, использовался неоднократно, состоит в том, что исследователь, исходя из того, что содержание меняется определенным образом, предсказывает, что измерение другого периода покажет точно такое же соотношение характеристик.

Проблематика, которую мы обсудили, позволяет нам нагрузить некоторые принципиальные характеристики анализа содержания, которые мы приводили ранее, такими обертонами, которые без дополнительной информации на их счет остались бы непонятыми. Объективность анализа состоит также и в том, что категории, которыми оперируют в исследовании, определены так четко, что различные исследователи, анализируя одно и то же содержание с помощью этих определений, получают один и тот же результат. Кто-то может не согласиться с тем определением оценки, которую мы только что привели, и будет использовать свою метрику - брать только качественные прилагательные как фактор отношения. Результаты будут разные, но ведь и методика была разной. Если вы принимаете мою методику и работаете с ней, мы закодируем текст совершенно одинаково, а значит, получим на выходе одинаковый результат.

После обстоятельного разговора о выборе единиц наблюдения станет понятным, что систематичность анализа состоит в том, что этот выбор должен основываться на формальных, беспристрастных признаках, другими словами, исследователь не может выбрать для анализа только те части текста, которые подтверждают его гипотезу, и отвергать другие. Это требование позволяет избежать аргументированной подтасовки фактов.

Перечисление этапов исследования выводит нас еще на одну очень важную характеристику контент-анализа, как и любого другого социологического исследования - научная строгость исследования подразумевает обязательное соблюдение этих этапов с набором требований, предъявляемых каждому из них.

Полевая стадия исследования, как никакая другая, связана с такими характеристиками исследования, а значит, и его программы, как его организационные параметры - этапы исследования, сроки исполнения, способы обработки, задействованные в исследовании исполнители, финансовые затраты. Зависимость содержательных характеристик и финансовых затрат огромна. Финансирование социологического исследования никогда не бывает абстрактным - чаще всего средства адресуются конкретным проектам. В таких случаях само обоснование исследования - некоторая гарантия финансирования работы. Более того - в последнее время многие проекты защищают свое право на существование на конкурсной основе. Программа исследования тут вполне способна показать товар лицом.

[1] Ядов В.А. Социологическое исследование. М., 1998.

[2] Danielson W. Content Analysis in Communication Research // R. Nafsiger and D. White (eds). Introduction to Mass Communication Research. Lousiana , 1963.

[3] Имеется в виду генеральный проект «Функционирование общественного мнения в условиях города и деятельность государственных и общественных институтов», реализованный в Институте конкретных социальных исследований АН СССР в 1967-74 гг. (руководитель проекта доктор философских наук Б.А. Грущин). См.: Массовая информация в советском промышленном городе. Опыт комплексного социологического исследования / Под общей ред. Б.А. Грушина, Л.А. Оникова. М., 1980; см. также Федотова Л.Н. Анализ содержания местных газет, радио и телевидения как источников информации. Автореф. дис. на соискание ученой степени канд. Филолог. наук. М., 1969. В настоящее время данные этого исследования можно запросить по электронному адресу butaev @ expert . extech . msk . su (база данных INSYS ).

[4] См.: Федотова Л.Н., Красавченко И.А. О критериях отбора фактов для информационной телевизионной программы // Вести. Моск. ун-та. Сер. 10. Журналистика. 1986. №6

[5] Оказывается, в свое время И. Кант размышлял над тем, как измерять информацию. Он считал, что ее нужно мерить по тому времени, которое нужно для усвоения ее адресатом. Этот пример приводил в своих лекциях Ю.А. Левада.

[6] Классификация телепередач ЮНЕСКО (1983 г.) имеет такую типологию: информативные передачи, образовательные программы, культурные программы, религиозные, программы для детей, развлекательные, реклама. По этой классификации, которая в инструкции хорошо детализирована, ЮНЕСКО осуществило международное исследование.

[7] Следует иметь в виду, что если в ходе анализа окажется, что эта «свалочная» яма чересчур переполнена, то надо признать неудачной саму классификацию, причина также может быть в том, что вы недостаточно ясно описали признаки каждой группы в этой классификации и кодировщик затрудняется принять решение - ему легче отнести это к «трудным» случаям. В любом случае обнаружить причину поможет стадия пилотного (или, как говорят некоторые социологи, пилотажного) анализа, которая для этого и существует.

[8] См .: Lerner D., I. Pool, Lasswell H. Analysis of Political Ideologies // Public Opinion Quarterly. 1952. Vol.15. No.4. P.720.

[9] См.: Федотова Л.Н. Массовая информация… С. 31-56.

[10] 47 пятниц, выпуск 5: Контентный анализ - методика и организация / Редактор-составитель и автор предисловия В.В. Сазонов. Информационный бюллетень ИКСИ АН СССР. 1969. №15. С.187.

[11] См.: Федотова Л.Н. Массовая информация… С. 57-94.

[12] Wolfenson W. Consumerism and Consumption: A Message Systi Analysis of Television Commercials. University of Pennsylvania . 1980.

[13] Kerrick .I. The Influence of Captions on Picture Interpretation // Journalism Quarterly. 1955. Yol. 32. No.2.

[14] Stone Ph., D. Dunphy, M. Smith, D. Ogilvie. The General Inquirer: A Computer Approach to Content Analysis. Cambridge , 1966.

[15] См.: Федотова Л.Н. Социологи анализируют прессу // Вести. Моск. ун-та. Сер.11, Журналистика. 1976. №6. С. 82-85. В моем архиве давно хранится вырезка из газеты «Советская культура» (21.05.85), где рассказывается, как через компьютер был пропущен роман классика итальянской литературы Алессандро Мандзони «Обрученные». Оказалось, что роман состоит из 8950 слов, некоторые из которых повторяются неоднократно - общий итог 223000 слов. Филологи также узнали, что самыми употребляемыми подлежащими были «дом» и «слово», прилагательными - «большой» и «добрый», наречиями - «хорошо» и «мало». Главная героиня, деревенская девушка Лючия, чаще всего произносит слово «сердце».

[16] Известно, что с началом перестройки в России произошел настоящий «бум» прессы. Тогда известный российский социолог А.Н. Алексеев (Санкт-Петербург) создал специальный фонд российской прессы периода начала перестройки. Исследователь, который поставит своей задачей провести контент-анализ прессы этого периода, вспомнит об этом фонде.

[17] Доклад был представлен на Международном конгрессе социологов по проблемам, массовых коммуникаций (круглый стол по контент-анализу). Барселона, Испания, 24-29 июля 1988 г.

[18] При отборе опрашиваемых случайным методом социологи часто пользуются списком избирателей, где алфавитный принцип списка существенно «перемешивает» население. См.: Капелюш Я.С., Сазонов В.В., Федотова Л.Н. Учреждения культуры в небольшом городе и население. М., 1985. С. 10.

[19] В профессиональной литературе термин « reliability » переводится также как «надежность» и «устойчивость». Обсуждение общих проблем соотносимости /надежности см., например. Anastasi A. Psychological Testing. New York : McMillan, 1961; Garrett H. Statistics in Psychology and Education. New York : Logmans, 1955; Guilford J. Psychometric methods. New York : McGraw-Hill, 1954.. обсуждение специального применения методик в контент-анализе см.: Janis I ., R . Fadner and M . Janowit z. The Reliability of a Content Analysis Technique // Public Opinion Quarterly. 1943. Vol.7. No.2. P.243.

[20] См . об этом : Lasswell H. and associates. The Politically Significant Content of the Press: Coding Procedures // Journalism Quarterly. March 1942. Vol.19. No. l. P.12-21.

[21] Janis I. The Probli of Validating Content Analysis // Lasswell H., N. Leites and associates. Language ot'Polilics: Studies in Quantitative Siantics. Revised Edition. Cambridge , Mass. : The M. I. Т . Press, 1965.

[22] Горбунов Н. Обоснованность, надежность и переносимость вопросников // Труды по социологии. Выпуск 2. Тарту, Тартуский Государственный университет. 1972. С .195-221. См . также Alien R. Personality Assessment Procedures. New York : Harper, 1958; Cattell R. and Warburton F. Objective Personality and Motivation Tests. Chicago : University Illinois Press, 1967.

[23] Лассвелл употреблял этот термин по контрасту со «скрытым, подразумеваемым» ( latent ) содержанием и определял его так: «Мы расцениваем содержание как открытое, если слова интерпретируются так, как они понимаются большинством носителей данного языка». См .: Lasswell H. The Uses of Content Analysis Data in Studying Social Change // Science and Culture. 1967. Vol. 33. No. 4. P. 163.