

## **Дискурсивный контент-анализ. Исходные методические позиции эмпирической практики<sup>1</sup>**

В. Г. Погорецкий, Н. С. Смирнова

Прогресс информационных технологий создает основу для роста, расширения сферы и интенсивности вербальной коммуникации. К числу средств обработки информации, выраженной на естественном языке, относится контент-анализ. Универсальность и многофункциональность метода определяют его ценность для анализа текстов. В статье рассмотрены аспекты развития метода в связи с его междисциплинарным характером, проблемой категориальной эволюции, дискурсивной логической природой текстового сообщения.

## **Зависимость инструментальных ресурсов контент-анализа от ресурсов информационных технологий**

Контент-анализ (К-А) как метод эмпирического социологического исследования нуждается как в общей оценке его ресурсов и возможностей, так и в постоянной переоценке его инструментального потенциала в связи со вновь открывающимися перспективами. Будучи универсальным средством анализа текстов, К-А при решении каждой конкретной социологической задачи должен осознавать свой статус как средства эмпирического измерения, но в то же время в силу своей универсальности и пластичности вызывает такой поток ассоциаций, что попытки определить конкретный инструментарий, а в последнее время возникают все новые технические средства, ориентированные на контент-аналитические решения проблем социальной коммуникации, выливаются в создание различных списков возможностей. Пространные или лаконичные, они совпадают или пересекаются, при этом оставляя впечатление достаточно произвольных

---

<sup>1</sup> Работа выполнена при финансовой поддержке РГНФ (проект № 02–03–18007а).

перечней функций метода в целом или в связи с его конкретной технической реализацией.

Мы говорим о «технической реализации» метода постольку, поскольку современный К-А осуществляется в среде компьютерной техники информационных технологий, и содержательные идеи, которые закладываются в конкретное воплощение метода К-А как средства анализа, неизбежно облекаются в форму компьютерных программ, те же, в свою очередь, требуют программного обеспечения для их реализации, могут быть автономны, могут быть совместимы или несовместимы с подобными им, технически реализующими тождественные, сходные или альтернативные содержательные идеологии метода К-А.

Контент-анализ сочетается с целым рядом подходов, ассимилирует различные принципы из различных сфер научной проблематики, поэтому попытки определить совокупность его функций и возможностей в связи с конкретной разработкой влекут за собой такой шлейф заслуживающих описания обстоятельств, что непосредственно методика анализа данного инструментального средства утопает во множестве, впрочем, вполне оправданных, инструктивных указаний.

Социологический контент-анализ традиционно относится к обширной сфере исследования документов или к более узко определяемой области анализа документации в рамках фиксированных границ социального объекта конкретного эмпирического исследования, традиционно определяя зону своего непосредственного применения как количественного формализованного анализа текстов [1, с. 318]. Учитывая историю развития метода, когда на первоначальных стадиях измерения степени значимости материала документа для субъекта исследования оценивалась, например, по площади на газетной полосе, посвященной ему, против подобной трактовки не может быть возражений. Соответственно, все те методологические принципы, подходы, методические описания общей организации общей организации исследования документов в полной мере могут быть отнесены и к К-А как одному из типов более широкого класса социологических исследований. При описании контент-аналитических методик нет необходимости описывать всю совокупность процедур осуществления эмпирического исследования, достаточно ясного представления принадлежности метода к области, относительно которой корпус методологической и методической литературы создается.

Естественно, специфика каждого конкретного метода анализа документов определенного типа объектов привнесит свои характерные приемы, принципы, стратегии. Это очевидно, как очевидно и то, что эти специфические, сформированные для специальных и даже индивидуальных задач методы могут оказывать влияние и на общеметодологическое пространство данного класса исследований. Исследователь сознает такую

практику и, решая непосредственно поставленную задачу, имеет в виду перспективу более широкой значимости применяемой им техники и инструментария, чем только в пределах данного конкретного социологического исследования, что и находит отражение в его методических и методологических рефлексиях собственного методического опыта.

Но данная деятельность развивается в пределах области заданной исходной типологии методов социологических исследований. Она же эволюционирует, и происходит это под влиянием техники и инструментария конкретных прецедентов эмпирической практики. В этом процессе К-А занимает особое место. Опуская эволюции метода в процессе его исторического становления, можно сказать, что современный К-А не существует вне информационных технологий, компьютерной техники, технических средств передачи, хранения и обработки информации. Технологии в сфере электронных носителей информации демонстрируют уникальную способность к увеличению ресурсов, форм и способов обработки информации. И соответственные возможности для роста потенциальных возможностей, эволюции всей системы анализа текстовой социальной коммуникации приобретает и К-А. Так, например, в результате развития сугубо технических возможностей электронных технологий последнее время становится возможным непосредственный К-А фонетической информации. Это только характерная иллюстрация, тогда как непосредственно рост ресурсов увеличивает доступные для анализа объемы первичной информации, возможности архивирования информации, диапазон «исторической глубины» анализа социальной проблемы. Параллельное же и развивающееся опережающими темпами программное обеспечение не только способствует развитию изощренности анализа, но расширяет круг заинтересованных в К-А социологов, сокращает дистанцию между специалистами по прикладному программированию и социологами-эмпириками, расширяет возможности их непосредственного участия в обработке и анализе данных. Очевидно, что по мере развития компьютерных технологий расширяются и возможности постановки и решения теоретических проблем развития метода. Во всяком случае, когда выдвигаются новые проблемы, предлагаются новые подходы и решения, и опираются они на последние (из доступных) достижений этих технологий.

## **Междисциплинарные контент-аналитические исследования как источник развития метода**

Главным фактором эволюции К-А является то, что его объектом является текст. Если текст мы рассматриваем как объект социологического анализа, как документ, как социальный факт информационной коммуникации,

то текст как факт культуры рассматривается в качестве объекта множественном других дисциплин, и проблематика социологического К-А не может изолироваться от их интересов и проблем, тем более она не может изолироваться от задач использования К-А за пределами собственно социологических исследований. В связи с ними К-А развивается как специальный метод, как социологическое направление, как профессиональная специализация исследований и как явление социологической субкультуры.

Эпистолярный жанр «Новой Элоизы» не слишком удобен для цитирования, но здесь мы находим характерное обобщенное выражение доктрины социальной интеллектуальности Руссо: «Отдавшись своей склонности к наблюдениям... я почувствовал... что, если ограничиться одним лишь созерцанием, ничего не увидишь, — надо действовать, и тогда увидишь, как люди действуют. Я стал актером, чтобы сделаться зрителем» [2, с. 423]. Данный фрагмент показывает, что созерцание, наблюдение, действие здесь связываются воедино («наблюдение» в данном случае, в отличие от созерцания, предполагает рациональное осмысление). Текст как средство социальной коммуникации, как универсальное средство фиксации социального опыта в качестве объекта ряда научных дисциплин становится предметом различных типов абстрагирования. Каждая, исследующая его дисциплина, находит в нем содержание собственной проблематики, но его общая природа, внутреннее единство как дискурсивной формы рационального сознания создает среду необычайно тесной связанности, смысловой тождественности результатов, получаемых исследователями, представляющими разные научные традиции.

Общность проблематики, постоянно создает наложение зон интересов, зон научной компетентности, побуждает обращаться к одним и тем же фундаментальным положениям языкознания, лингвистики, социолингвистики, психолингвистики, модальных и интенциональных логик. Одним примером можно проиллюстрировать это положение дел. Логический квадрат — фигура формальной логики находит применение в общих методах эмпирической социологии, является также непременным атрибутом К-А. В семиотике логический квадрат трансформируется в «семиотический квадрат», превращаясь из наглядной схемы отношений суждений, имеющих одинаковый материал, но различную форму, в метаструктуру порождения членов категории, соотносимую с формальными структурами шестиугольника З. Бланше, группами Клейна и Пиаже [3, с. 502]. Однако, почему-то никто не ссылается на критическое отношение Канта к подобным конструкциям, недвусмысленно выражаемое им в «Критике чистого разума» [4, с. 105], которое можно принимать или не принимать, но которое указывает на отношение данного предмета и к сфере метафизики рациональности. Кстати сказать, К-А, активно использующий термин «категория», может с пользой заимствовать из семиотики понятие «члена

категории», что позволяет более дифференцировано описывать внутриметодические проблемы К-А. Налицо тесное переплетение сфер проблематики, относящихся к разным дисциплинам.

Контент-анализ не сводим к формальной технике, к совокупности или пакету фиксированных процедур, он органично сочетает подходы количественных и качественных методов социологии. Попытки определить ему роль анализа подобного некоему пакету статистического анализа данных содержит признаки технократического мышления, основывающего свой оптимизм на иллюзии исчерпывающих возможностей формальных методов. Он может успешно развиваться только как область междисциплинарной проблематики. Поскольку текст является универсальным носителем смысла, то круг объектов К-А не может быть замкнут, а становясь достоянием анализа различных научных дисциплин, текст создает основания и для создания собственно присущих данному направлению аспектов научного метода. Если нужно произвести подсчет, сколько раз в древнеиндийском эпосе «Махабхарате» упоминаются имена богов Митры и Варуны, то это задача К-А является классической не только потому, что сводится к элементарному подсчету терминов, их распределению в тексте, но и потому, что эти божества — антиподы, образуют оппозицию, и анализ их распределения и соотношения является задачей содержательного анализа текста. Это задача филологическая, но в методическом плане она оказывается совершенно совпадающей с целым рядом социологических задач. Но когда в той же филологии К-А решает задачу обозначения имен божеств в «Илиаде» в зависимости от того, употребляются их имена непосредственно в авторском тексте или в прямом обращении к божеству одного из героев, то тогда задача становится сугубо специфической, особенность постановки которой диктуется характером филологической проблемы. Но решение ее методом К-А обогащает общий арсенал его ресурсных возможностей, который так или иначе может либо непосредственно, либо в результате некоторого переосмысления быть использован при анализе других объектов, в том числе и в социологии.

В задачах более высокого порядка поиска смысла и значения временных эпистемологических контекстов, задачах интенциональных логик в сфере информатики и искусственного интеллекта на новом уровне воспроизводятся все те же исходные элементы категориальных структур, что и в К-А. Более того, задача формальной интерпретации текста как осмысливаемого сообщения ставится непосредственно и в самом К-А [5]. И. А. Герасимова, останавливаясь на наличии двух путей логического анализа языка «концептуального» (собственно логического) — от формальной структуры абстрактных моделей к эмпирическим контекстам и «лингвистического» — от эмпирических контекстов естественного языка к формальным моделям, констатирует необходимость совмещения обоих

подходов [6, с. 79], что в очередной раз иллюстрирует наличие глубокого взаимопроникновения предметного и методологического аспектов различных дисциплин при решении прикладных и теоретических проблем анализа текста. Свой вклад в развитие метода может внести и социологическая практика использования К-А, представляя методологам результаты анализа больших объемов стереотипных по содержанию текстов, в свою очередь воспринимая и используя эффективные средства логического и лингвистического анализа.

Восприятие К-А как чисто технического средства, как ресурса исключительно компьютерного формального анализа содержит опасность рутинизации его в качестве социологического метода именно в силу его эффективности, возможности предоставления по результатам формальной обработки массивов текстов достаточного материала для аналитической работы. Опасность рутинизации состоит в том, что набор процедур обработки и порядок их использования остаются неизменными, тогда как рассмотрение метода через призму междисциплинарного взаимодействия содержит перспективу актуализации множества потенциальных возможностей.

## **Необходимость терминологических эволюций**

Доказывая свою эффективность, К-А, находится в процессе постоянных трансформаций, его круг целей и объектов постоянно расширяется. В ходе его развития расширяется арсенал его инструментария, формируются альтернативные подходы. Как метод эмпирического исследования он пополняет свой багаж с решением проблем каждого нового класса реальности, к которому он оказывается приложим как средство анализа. Это очевидно. Но при этом оказывается, что терминологический аппарат, который используется для описания процедур метода, развивается, перестает быть отчетливым, приобретая даже черты противоречивости. С развитием метода должна развиваться и его терминология, причем не только новые средства и факты должны определяться новыми понятиями, но и переосмысливаться и переопределяться сложившиеся ранее, в том числе и исходные понятия. Как метод количественного измерения К-А требует точных определений, и если общую ревизию терминологии, в принципе, производить можно, но процесс это длительный, требует проведения широких дискуссий, обсуждения разнообразного эмпирического материала, то уточнение основных операциональных понятий, применительно к каждому случаю использования инструментария становится необходимым требованием проведения эмпирического исследования.

Так в К-А понятия текста и объекта связываются далеко не однозначно. Если макросоциологический анализ требует рассмотрения некоторой

совокупности текстов, характеризующих эволюцию социального объекта на протяжении длительного исторического периода, то в условиях, когда социально-экономических, статистических показателей может быть недостаточно, а это достаточно обычная ситуация при изучении прошлых эпох, такой анализ текстов становится ведущим методом описания. На основании проведенного исследования формируются социокультурные индикаторы, которые и характеризуют тенденции изменения объекта.

Собственно, объект не исследуется, исследуются тексты как его артефакты. Индикаторы — характеристики объекта. К-А здесь выступает в качестве одной из многочисленных форм вторичного анализа [7]. Если К-А исследует некоторую совокупность документов, регламентирующих деятельности организаций, то актуальными становятся не процессы функционирования этих организаций, а то, насколько исчерпывающе, адекватно исследуемая документация эти процессы отражает. Здесь, собственно, объектом становится текст. Анализ же текстов прессы, излюбленного материала К-А, может фокусироваться на фактах, описываемых в публикациях, на позициях социальной самоидентификации авторов, на психологических характеристиках аудитории, для которой текст предназначен. В любом случае объектом анализа становится многоплановая реальность, и вскрыть ее можно, лишь отчетливо представляя отношения объект — текст в каждом конкретном случае. Самоочевидное, формальное как бы уже ритуальное употребление этих понятий более невозможно.

Аналогично ситуация складывается в отношении понятия индикатора. При традиционном взгляде на К-А как одного из методов социального измерения предполагается, что есть социальное поле, есть инструментарий. Если социальные факты, подлежащие исследованию, не могут быть исчерпывающе описаны, решение об их наличии, их качестве, состоянии принимается на основании индикаторов, о которых достоверно известно, что эти индикаторы представляют собой эмпирически однозначно фиксируемые признаки, которые коррелируют с изучаемой реальностью социальных фактов. Установление наличия индикатора свидетельствует о наличии искомого социального факта. Несколько индикаторов свидетельствуют о наличии факта с более высокой степенью достоверности, измерения, проведенные несколькими независимыми исследователями, дают большую уверенность в объективности результата. Совпадение результатов, произведенных разными методами, в еще большей мере подтверждает объективность полученного результата. Такого рода абстрактная модель отвечает позитивистскому идеалу социологического измерения.

До некоторой степени и К-А может отвечать такого рода требованиям. Так, в условном примере контент-анализа открытого вопроса, приводимого в работе [8, с. 424]: «Какие блюда к праздничному столу вы предпочитаете?» Из списка ответов: бифштекс, отбивная, котлета, борщ, антрекот,

винегрет, соленые грибы, икра, мороженное, рассольник — любое блюдо из предложенного списка можно со стопроцентной уверенностью отнести к одному из трех заданных классов: горячие блюда, закуски, десерт. Подобное положение возникает и в ситуации реального исследования, когда на основании текстов газетных публикаций по каждой из них судят, поддерживает автор кандидата Дж. Буша или Дж. Керри на пост президента США.

В ситуациях более проблемных, когда выявление и оценка категорий, в соответствии с которыми исследователь делает свои выводы об объекте, зависит от даваемых этим категориям определений в тексте на основании разнородного лексического материала, смысл и значение отдельных слов в котором меняется в зависимости от контекста, ряд определений не текстуально, но с известной степенью уверенности может квалифицироваться как тождественных. Данные определения выступают как индикаторы значений категорий по отношению к непосредственным характеристикам текста, на основании которых исследователь, эксперт судит о социальных фактах. Все это свидетельствует, как в каждом случае разработки методики для конкретных исследовательских программ конкретных классов объектов исследования терминология, операционализируемые понятия должны переосмысливаться и переопределяться заново.

Социологический метод К-А устанавливает, регистрирует, измеряет социальные факты на основании констатации тождества элементов текста-объекта с концептуально определенными и заданными эталонами элементов текста в своей совокупности в фиксированном виде отражающими цели, доктрину и гипотезы автора исследования. Отождествление может быть непосредственным, т. е. слово (или совокупность слов) текста-объекта совпадает со словом из перечня эталонов-категорий, определенных исследователем, или опосредованным, когда слова текста-объекта относятся к некоторому вербальному классу, охватываемому эталоном-категорией. В этом случае отождествление элемента исследуемого текста и эталона происходит на основании явным образом заданных правил и процедур. Степень формализации правил и процедур может быть различной, в зависимости от того насколько факт отождествления элемента текста с эталоном отделяется от констатации непосредственного совпадения, а также формализованных нормативов идентификации и осуществляется на основании неформализованных доводов и представлений. Вследствие увеличения неформализованных факторов идентификации метод приобретает характеристики качественного, феноменологического анализа. Необходимость интерпретации контекста для констатации принадлежности элемента текста к той или иной категории-эталону создает поводы для скептицизма в отношении объективности метода. Все эти обстоятельства разработчик инструментария и методики проведения конкретных контент-аналитических исследований имеет в виду в ходе своей деятельности.

Установленные в результате анализа показатели позволяют сравнивать совокупности социальных фактов, рассматривать временные ряды социальных фактов (time series), осуществлять кросс-групповое сопоставление (cross-section). Последний способ оценки тенденций измерения социальных явлений базируется на отождествлении групп социальных явлений, находящихся на разных стадиях развития, с гипотетически имеющим место явлением, которое в различные моменты своего развития соответствует параметрам групп реальных явлений, с которыми оно отождествляется. Таким образом, возникает возможность для экстраполяции тенденций развития событий, для которых еще нет достаточной базы измерения их эволюции во времени. В силу того что тексты существуют по поводу любых социальных фактов, и круг объектов К-А не ограничен.

В принципе мы в праве заявить, что поскольку любой текст есть продукт сознания человека, постольку он как факт культуры и человеческой деятельности может быть объектом К-А и становится объектом К-А как документ этой культуры, деятельности, сознания и социальной коммуникации. Такая трактовка текста как собственно социологического объекта (наряду с объектами других социальных дисциплин) вполне правомерна, но она предельно расширена, и вместе с расширенной трактовкой социологического объекта предельно расширяется и трактовка социологической проблемы. Как направление социального научного мышления социология действительно универсальна. Рассматривая же социологию как сферу продуцирования научных прикладных массовых результатов и К-А как одного из методов получения таких результатов, мы должны сфокусировать представление об объекте К-А на определенных видах текстов, которые в наибольшей мере соответствуют массовым задачам массовых измерений социальной реальности, т. е. являются подобными измерениям, производимым в социологии другими методами, что, конечно, не предполагает внесения ограничений на развитие универсальных подходов и получения, соответственно, уникальных результатов.

## **Тематически общие тексты как объект конкретного социологического исследования**

В распоряжении социолога могут оказаться текстовые документы разных типов. В связи с хрестоматийным примером исследования В. Томаса и Ф. Знанецкого «Польский крестьянин в Америке», использовавшим 754 письма польских эмигрантов, оказывается не лишним замечание Сенеки о значении письма как социального, исторического документа: «Имени Аттика не дают погибнуть письма Цицерона. Тут не помогло бы ни то, что зятем его был Агриппа, ни то, что внучка его была

замужем за Тиберием и Цезарь Друз приходился ему правнуком: среди столь громких имен об Аттике и помину бы не было, если бы Цицерон не связал его имя со своим» [9, с. 45]. Письма являются чрезвычайно разнообразным, чрезвычайно информативным источником фактов для социолога. Любая историческая эпоха от древности до современности, любая социальная субкультура, сохранившая предметы письменной культуры, обладает ценнейшими свидетельствами социальной жизни.

То обстоятельство, что за решение задачи формирования социальных индикаторов за длительные исторические отрезки времени при отсутствии данных социальной и социально-экономической статистики можно браться за счет обращения к личным письмам имеет под собой реальную почву [10]. Огромное число исторических и общественных деятелей оставляет помимо основных трудов и эпистолярное наследие, ограничимся единственной ссылкой на опубликованные письма П. Л. Капицы [11]. Ту же роль выполняют судебные речи выдающихся ораторов древности, нового времени и современности — жанр, берущий начало от классических образцов античности, Демосфена, Цицерона, Апулея, к которому примыкает и жанр контroversии, разрабатывавшийся в трудах античных историков. Последний завоевал столь широкую популярность, что в современном кинематографе существует целое направление, в рамках которого интрига строится на рациональности дискурса противостоящих сторон. В более широком плане это жанр публичного выступления знаковой фигуры социального, социально-политического процесса. Подобного рода тексты становятся предметом К-А. Анализ инаугурационных речей американских президентов, обращений английской королевы к парламенту, речей германского кайзера стали своего рода эталоном аналитических возможностей метода.

Применение К-А к анализу открытых вопросов анкет характеризует непосредственное использование метода в процессе традиционного социологического исследования, является некоторым прообразом комплексного анализа, когда он включен в качестве составляющей в саму структуру совокупности действий по реализации исследовательских программ. Использование К-А в комплексе с другими социологическими методиками является очень важным начинанием и имеет широкие перспективы, но как область непосредственного применения метода ограничивает его возможности достаточно узкими рамками.

Использование прессы рассматривается в социологической литературе в разнообразных аспектах. Если оценивать ее с точки зрения К-А, то это индустрия производства текстов, если воспользоваться метафорой Маклюэна, то его определение производства печатной продукции как «Галактики Гуттенберга» в наибольшей степени относится именно к прессе. Пресса воспроизводит социальную структуру социума, адресуется ко всем

социальным группам, отражает события, социальные движения, мотивационные и поведенческие образцы, аксиологические нормы. Массовое и элитарное сознание представляется в прессе в самых разнообразных аспектах.

Для К-А это необозримое поле деятельности. Материалы прессы как объект представляют тексты семантически не нормализованные, т. е. смысловые единицы здесь выражены на естественном языке, не ограниченном шаблонами, как в деловых документах типа должностных инструкций, направленных на то, чтобы исключить неоднозначность истолкования, избежать противоречий, представить последовательно исчерпывающие перечни позиций предписываемых обязанностей. Таким образом, тексты публикаций — это естественных язык.

В то время как конкретные публикации представляют читательской аудитории единичные события, эти события являются представителями массовых классов событий. Сообщения строятся по принципам обращения к структурам массового сознания, т. е. тексты прессы обладают всеми параметрами социологического объекта, они позволяют по совокупности параметров идентифицировать объект, позволяют на основе многократных измерений совокупности однородных объектов выявлять объективные характеристики как собственно социальных объектов, ставших предметом публикаций так и самих текстов публикаций с особенностями их задач, целевых ориентаций социальных предпочтений авторов и других факторов издательской мотивации. Конечно, жанровое разнообразие публикаций чрезвычайно велико, но для отработки техники дискурсивного контент-анализа можно выделить сферы тематического единства, в рамках которых будут формироваться объекты исследования и которые будут характеризоваться однородностью предмета описания, размеров текста, быть идентифицируемыми по объективным параметрам объекта эмпирического социологического измерения. Отрабатываемая таким образом техника измерения метода К-А будет содержать необходимые базовые элементы, которые при анализе объектов других типов будут модифицироваться, но основа оставаться стабильной или меняться в ходе эволюции с относительно медленной динамикой.

## **Вариативность идентифицируемых референтов категорий**

Так как К-А базируется на принципе фиксации, измерения и количественной оценке текстовых единиц, квалифицируемых как тождественные, на основании таких массовых измерений и делаются выводы о социальной реальности. Вопросы о единицах регистрации и их тождестве становятся для К-А ключевыми, в зависимости от их решения формируются

те или иные способы реализации метода. Собственно говоря, в зависимости от глубины понимания сложности этих вопросов и развивается метод. Процедура регистрации будет элементарной, если нужно зафиксировать слово, которое и будет носителем необходимого значения, являющегося предметом анализа.

Предположим, исследуются досуговые занятия и нужно оценить, какое место по значимости в публикациях о легкой атлетике как оздоровительной форме физической культуры занимает бег. Определенную картину даст сопоставление частоты употребления слова «бег» со словами «метания», «прыжки» и т. п. Проблем регистрации для данной задачи не возникает. Так же как ее не возникает, если нужно оценить сложность предложений, составляющих текст, по числу запятых в их составе, т. е. некоторым образом охарактеризовать готовность аудитории к восприятию текста того или иного уровня сложности. Фиксация числа предложений и числа запятых в каждом из них представляет чисто формальную задачу. Именно возможность формализации делает реальным применение компьютерной техники для нужд К-А. Правда, в связи с этим возникает проблема корректности формализации регистрации элементов текста. Фиксировать запятые в исследовательских целях — задача маловероятная, скорее такая информация может служить некоторой предварительной характеристикой исследуемого текста, но на ней возможность непосредственной автоматической исчерпывающей регистрации элементов кончается. Уже в приведенном примере регистрации слова «бег» возникают некоторые проблемы. «Бег» — склоняемое существительное, неизменяемая часть, основа, совпадает с исходной словоформой, соответствующим образом организовав программу поиска, можно найти данное слово, в каком бы числе и падеже оно не было употреблено в тексте. Но бег, несмотря на его конкретность, все-таки обобщающее понятие. Слово «пробежка» относится к логическому классу понятия «бег», но оно не будет зафиксировано, потому что в основе произошло изменение «г» на «ж». Подобного рода другие особенности языка, типа неразличимости в тексте двух разных слов: «за́мок» и «замо́к»; «о дне» (сосуда) и «о дне» (недели); «лень» (как существительное) и «лень» (как предикатив), возможно, будут встречаться, но число их в общей массе окажется незначительным, погрешность измерения, связанная с возникающими ошибками идентификации будет несущественна, тем более что они вполне поддаются устранению в процессе коррекции текста. Надо отметить, что подобный уровень обработки текста допускает высокую степень автоматизации со всеми вытекающими отсюда последствиями: большие объемы текстов, быстрота обработки, текущий, практически, немедленный мониторинг поступающих сообщений, сокращение затрат на ручную обработку, устранение ошибок, связанных с человеческим фактором.

Это, отчасти, и явилось источником оптимизма в отношении перспектив решения проблем К-А путем автоматизации в 80-х гг. XX века. Естественно, удовлетворительное решение задач на простом уровне побуждает к постановке более сложных. Проблемы возникают, стоит только перейти от регистрации конкретных понятий к понятиям абстрактным. Понятия «благополучие», «безопасность», «солидарность», «конфронтация» имеют широкий спектр значений, а следовательно, и определенное количество синонимов. Словарь синонимов русского языка определяет синонимы как «слова, обозначающее одно и то же понятие, следовательно, тождественные, или предельно близкие по значению» [12, с. 1]. «Предельно близкие» — критерий довольно расплывчатый, лингвисты расходятся в его трактовке. Сюда же примыкает и проблема паронимов [13]. Технически в К-А нужно знать, фиксировать ли два различно написанных слова в качестве одного и того же понятия? При утвердительном ответе, это значит, что содержащиеся в словаре эталоны эти два различных по написанию слова должны быть зафиксированы как принадлежащие к одной категории. Но в пределах смыслового контекста одного текста разные слова могут использоваться как тождественные, тогда как в другом, в силу различий их функционального диапазона, трактоваться как различные понятия. Поэтому создание контекстно-зависимых словарей-эталонов достаточно трудоемко, а эффективность их остается проблематичной. Каждый раз, приступая к новому эмпирическому контент-исследованию, такие словари необходимо корректировать, естественно, общий массив словарей-эталонов при этом пополняется. Понятие контекстно-зависимых синонимов существенно шире просто синонимов. Механически словари-эталоны не могут пополняться, возникает опасность слишком широких трактовок и даже противоречий, точность измерения падает. Склонность перевести К-А в режим исключительно автоматизированного процесса приводит к иллюзиям по поводу его эффективности. Функция неформализованного анализа на ранних стадиях исследования, при регистрации вновь становится актуальной. Организация целесообразного взаимодействия формализованных и неформализованных процедур в процессе реализации метода осознается как необходимая.

## **Индикаторы категорий в логической структуре дискурса**

До сих пор нами рассматривалась относительно простая ситуация, когда единичное слово в тексте является смысловой единицей, фиксируется как единица регистрации, и в последующем, после обработки, в анализе фигурирует в качестве категории К-А. Это вполне реальная ситуация, но необходимость использования К-А в широком диапазоне требует

модификации методических средств. Так, понятия «Соединенные Штаты Америки» и «США» или «жилищно-коммунальное хозяйство» и «ЖКХ» обозначают в первом и втором случаях один и тот же объект. Это разные семантические средства выражения одного понятия, их смысл тождественен, но синонимами их не назовешь, поскольку в двух случаях понятие конкретного объекта выражено комплексом слов. Если аббревиатуры США и ЖКХ носят устойчивый характер и однозначно воспринимаются как слово, носитель смысла, то понятие «уголовный кодекс» воспринимается как целостное понятие исключительно в сочетании существительного и прилагательного. То же самое с армейским термином «живая сила», отличие только в том, что понятие «уголовный кодекс» содержит специфицированное определение, тогда как ни «живая», ни «сила» такой специфической определенности не содержит, и понятие приобретает значение характерной определенности только в сочетании обоих слов. Это разнообразие, неоднозначность исходного материала должна быть так или иначе отражена в методике анализа.

Если для фиксации в качестве категории устойчивой аббревиатуры ЖКХ достаточно ввести ее в эталонный словарь, то процедура поиска в тексте слова «пенсия» предполагает модифицированный алгоритм в качестве эталона поиска следует ввести основу слова «пенс\*»), чтобы фиксировать все словоформы слова «пенсия». Мало вероятно, но выборку может попасть название английской денежной единицы — пенс, в целом при таком способе поиска ошибки включения в выборку не соответствующих искомой категории слов вполне вероятно. Ошибки устраняются в ходе редактирования полученного результата. Для подобных целей в программное обеспечение включается внутренний редактор [8, с. 425; 14, с. 31]. Вообще действует общий принцип: все, чего не может обеспечить текущая версия программы в автоматическом режиме, переводится в режим внутреннего редактирования и подготавливается оператором. Так если программа не может содержать в качестве эталонов группы слов, например, уголовный кодекс, то в качестве индикатора может использоваться одно слово уголовный или слово кодекс, соответственно попавшие в выборку сочетания уголовный процесс или гражданский кодекс исключаются при редактировании. Процедура редактирования неизбежно становится более трудоемкой, если мы ищем категорию живая сила.

Аналогично посредством редактирования решаются проблемы замены в тексте по стилистическим соображениям термина, соответствующим местоимением. Дискурсивный К-А ключевым объектом анализа считает фразу. То есть предложение, содержащее одно суждение. Поэтому предложение «Ипотека реформируется, и предоставление ссуд упрощается» преобразуется в два самостоятельных предложения: «Ипотека реформируется», и «Предоставление ссуд упрощается». Конкретно в нашей программе

конец фразы определяется по наличию точки, поэтому фраза «Обвинение предъявлено по статье n, ч. m.» будет записана как два предложения (второе будет состоять из одного символа m). После редактирования этот дефект будет устранен. Собственно говоря, объем и стереотипность задач внутреннего редактирования являются основным источником совершенствования версий, используемых в К-А программ.

Единодушные в признании того обстоятельства, что понятия выражаются не одним словом, а некоторой их совокупностью нуждается в детализации. Во-первых, очевидно, что понятие может выражаться одним словом, и тогда единица регистрации может определяться безошибочно, регистрация будет проводиться автоматизированным образом, а результат будет исчерпывающим. Но далее, выражение отдельных понятий сочетаниями слов действительно становится проблемой для определения регистрируемых единиц и для автоматической их фиксации в частности. В простейших случаях, как показано выше, понятие может выражаться устойчивым сочетанием двух слов, имеющих значение термина, из которых, по крайней мере, одно в своем изолированном самостоятельном значении имеет непосредственное отношение к предметной тематике исследуемого текста. Это усложняет задачу автоматизированной регистрации элементов текста, но возникающие при этом трудности вполне преодолимы. И, как уже говорилось, те проблемы, которые не решаются в автоматическом режиме переносятся на уровень редактирования и решаются в комбинированном режиме человеко-машинной процедуры регистрации. То же самое, когда понятие определяется сочетанием двух слов в совокупности составляющих специфический термин, но независимые и изолированные принадлежащие к общепотребительной лексике, усложняет задачу не принципиально.

Более сложные проблемы возникают, когда понятия, которые должно фиксировать в соответствии с моделью содержательного анализа текста, фиксируются совокупностью слов, имеющих значение консолидированного комплекса только в пределах контекста данного анализируемого текста. Если проблема синонимических вариаций выражения понятий возникает уже на стадии «одно слово — одно понятие», существует на стадии «одно понятие — терминологическое сочетание слов», то на стадии «понятие — контекстно зависимое сочетание слов» проблема синонимического выражения понятий существенно увеличивается. Семантическая вариабельность выражения смысла сообщения резко возрастает, трактовки совокупностей слов как синонимических комплексов тождественных фиксируемым понятиям становятся проблематичными. Процедуры идентификации понятий переходят на качественный уровень, что, в целом, характерно для современных тенденций в социологии [15].

Сочетание качественного и количественного подходов характерно для К-А. По мере усложнения задач, феноменологический аспект

контент-аналитических исследований возрастает. Качественная составляющая увеличивается тогда, когда фиксируемая категория представляет собой не конкретное слово, или сочетание слов, составляющих термин, а некоторый обобщенный класс, к которому с равным основанием могут принадлежать несколько понятий и число их заранее не известно. Аналогично, когда фиксируемая категория описывается совокупностью признаков, причем для уверенной констатации присутствия категории достаточно не всех, а только нескольких признаков из совокупности и существует некоторое пороговое значение, при недостижении которого бесспорная уверенность сменяется проблематичностью.

До сих пор мы рассматривали ситуацию констатации факта присутствия в тексте отдельного изолированного понятия, которое задано исследователем в качестве регистрируемой категории. Реальное исследование предполагает регистрацию некоторой совокупности структурно связанных категорий, перечень которых определяется содержанием исследовательского задания. Но прежде содержательных аспектов формирования перечня категорий необходимо определиться с формальными аспектами структуры. Если существует понятие, то существует и его альтернатива. Для К-А важно фиксировать не только факт присутствия данной категории в тексте, но и явно выраженный факт ее отрицания. Для исследователя факт отрицания будет иметь позитивное значение. Анализируя суждения общей логики, это отмечает еще Кант [4, с. 82], следовательно, формальная структура отношений контрарности и субконтрарности актуальна и для К-А. В наглядной форме она демонстрируется в виде логического квадрата, и эта формальная структура активно используется в К-А, так же как в целях классификации суждений, фиксирующих исследуемые категории, могут использоваться формальные структуры типа шестигульника отношений модальности и количества суждений [16, с. 62, 67]. Факт альтернативы в регистрации категории для К-А имеет принципиальное значение как социологического метода, поскольку в ней фиксируется социальный выбор или решение об оценке. Нет однозначного представления, фиксировать термин со значением утверждения и со значением отрицания как одну категорию или как различные.

Мы акцентируем внимание на логических отношениях дискурса постольку, поскольку источником исходных данных в К-А является текст, естественная грамматически оформленная речь, но в процессе анализа элементы текста как объекта смысловой регистрации квалифицируются как элементы логической структуры. Основная проблема практического К-А состоит в том, чтобы квалификация элементов текста, представляющего совокупность предложений как элементов логических суждений, происходила с достаточной степенью точности. На это направляются многочисленные усилия разработчиков методик К-А. Это область, где сопрягаются качественный и количественный подходы.

Проблемы неформальной качественной идентификации возникают уже на начальных стадиях. Для иллюстрации воспользуемся примером Канта [4, с. 131]. Понятие «человек молодой» и «человек старый» являются контрарными. Если они избраны в качестве идентификационных категорий, то некое действующее лицо по тексту будет отнесено к одной из них. Кант заменяет определение «старый» на определение «немолодой», т. е. «старый». Эти выражения эквивалентны. Если теперь и мы заменим «молодой» на «не старый», то эквивалентность сохранится и здесь. Стирается, однако, ощущение границы между понятиями. Очевидно, ряд описаний в тексте будут находиться в некоей промежуточной зоне между двумя полюсами, где понадобится некоторое количество доводов на качественном уровне, чтобы квалифицировать отдельный случай упоминания действующего лица в тексте как относящегося к одному из двух альтернативных классов. В практике К-А в подобных случаях можно прибегнуть к дискуссии, или к параллельному кодированию, с последующей математической обработкой разных версий.

Обратим внимание на то, что проблема неопределенности классифицирования возникла на стадии непосредственной совместимости логического суждения и предложения, тогда как подлинные проблемы категориальной квалификации элементов текста начинаются, когда идеальная формальная структура логических суждений выражена семантическими средствами естественного языка. Предложений, смысловая форма которых непосредственно совпадет с логической формой суждения и к тому очевидной для восприятия не так уж и много. Так, например, у Канта: «Явления не есть вещи сами по себе» [4, с. 138]. Предложение в чистом виде представляет собой общеотрицательное логическое суждение. Или, там же: «Осознание самого себя далеко еще не есть познание самого себя» [4, с. 114]. Высказывание представлено в грамматической форме очень близкой к форме (кстати, весьма редкой) частноотрицательного суждения, если признать, что семантическое выражение «далеко еще не есть» эквивалентно логической формуле «некоторые... не есть». Операции с элементами естественного языка вызывают необходимость трансформации формальных логических структур в структуры семантические. Как упоминалось выше, в семиотике логический квадрат трансформируется в квадрат семиотический. Авторы «Объяснительного словаря теории языка» подчеркивают, что методический подход, применяемый к конкретным лингвистическим объектам, отличается от чисто логического или математического, которые «не зависят от семантической составляющей» [3, с. 501, 502]. Семиотический квадрат демонстрирует зависимость формальных логических структур от смысловых содержательных аспектов конкретного текста. Дальнейшее развертывание структур категориальной идентификации обнаруживает еще более тесную связь логического и содержательного, а отнесение

регистрируемой единицы к одной из альтернативных категорий делается еще более проблематичным. Показателен пример противоречивости оценок текста у Плиния Младшего. Своему адресату он пишет: «Мне показалось, ты в моих сочинениях напыщенным счел то, что я считаю возвышенным, дерзким — то, что я считал смелым, перегруженным — то, что я считаю полным» [17, с. 169]. Многообразие однопорядковых категорий порождает структуры, организуемые по принципу разделительных суждений, где количество альтернатив непосредственно задается содержанием текста. Их количество, в принципе, может быть бесконечно. В социальной прогностике это превращается в серьезную проблему «разбегающихся сценариев». На данном основании базируется тезис о недопустимости смешения нормативного и поискового типа сценариев [18, с. 145].

Структуры категориальной идентификации элементов текста расширяются за счет введения градаций интенсивности выражения качеств в однородных элементах, дальнейшее развитие структур происходит за счет сугубо предметных аспектов анализируемых текстов, логики содержательного описания событий и проблем. Категории, фиксирующие содержательные аспекты процессов, могут быть бесконечно разнообразны, следуя содержанию описываемого материала, но среди них значимое место занимают категории, фиксирующие стадии, состояния процессов, позиции участников социального действия, целерациональные и ценностно-рациональные оценки. Здесь возможны коллизии несовпадения моделей логики реального процесса, логики описания процесса автором текста, или авторов некоторого множества однородных текстов, и логики анализа текста исследователем. Возникновение альтернативных моделей категориального отражения и регистрации вполне возможно. В связи с этим может возникнуть проблема избыточности или недостаточности первичной информации в текстах для наполнения категориальной модели регистрируемыми данными.

Огромное разнообразие текстовых объектов К-А требует и чрезвычайно пластичных средств анализа. Программы компьютерного К-А должны строиться с учетом значения в условиях контекста, порождения контекстом практики словоупотребления, дальних ассоциаций формирования значения в условиях контекста.

## Литература

1. Рабочая книга социолога. М.: Наука, 1976; 5-е изд.: М.: Книжный дом «ЛИБРОКОМ»/URSS, 2009.
2. *Руссо Ж.-Ж.* Юлия, или Новая Элоиза. Письма двух любовников, живущих в маленьком городке у подножия Альп. Собраны и изданы Ж.-Ж. Руссо. 1761 // Руссо Ж.-Ж. Избр. соч. Т. 2. М.: Гос. изд-во худож. лит., 1961.
3. *Греймас А. Ж., Курте Ж.* Семиотика. Объяснительный словарь теории языка // Семиотика. М.: Радуга, 1983.

4. *Кант И.* Критика чистого разума. М.: Мысль, 1994.
5. *Weber R. P.* Basic content analysis. Beverly Hills; London; New Delhi, 1988.
6. *Герасимова И. А.* Формальная грамматика и интенциональная логика. М.: Ин-т филологии РАН, 2000.
7. *Погорецкий В. Г.* Дифференциация доходов населения — эволюционирующий вторичный анализ // Системные исследования. Методологические проблемы. Ежегодник 2001. М.: URSS, 2003.
8. *Ядов В. А., Беляев В. Я., Марарица В. Ф., Оконешников П. Н.* Общая социология. Теория и прикладные исследования. СПб.: Социум, 2001.
9. *Сенека.* Нравственные письма к Луцилию. Кемерово: Кемеровское книжное изд-во, 1986.
10. *Weber R. Ph., Namenwirth J. Zvi.* Dynamics of Culture. London; Sydney; Wellington, 1993.
11. *Каница П. Л.* Письма о науке. 1930–1980. М.: Московский рабочий, 1989.
12. Словарь синонимов. Л.: Наука, 1975.
13. Словарь паронимов современного русского языка. М.: Русский язык, 1994.
14. *Чураков А. Н.* Компьютерный контент-анализ. М.: Ин-т социологии РАН, 1996.
15. *Семенова В. В.* Качественные методы в социологии // Ядов В. А. (в сотрудничестве с Семеновой В. В.) Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности. М.: Добросвет, 1998.
16. Формальная логика. Л.: Изд-во Ленингр. ун-та, 1977.
17. Письма Плиния Младшего. М.: Наука, 1982.
18. *Бестужев-Лада И. В.* Поисковое социальное прогнозирование: перспективные проблемы общества. М.: Наука, 1984.

