

ИНФОРМАТИВНО-ЦЕЛЕВОЙ АНАЛИЗ ДОКУМЕНТОВ

Под информативным анализом текстов в широком значении обычно понимается вся совокупность разновидностей анализа содержания любых документальных источников. Сюда относятся применяемый в социологии и социальной психологии контент- анализ, используемые в архивоведении и информатике методы аналитической обработки и «свертывания» содержания первоисточников для создания массивов вторичных документов (обзоров, рефератов, аннотаций и пр.). К этой совокупности принадлежат все виды индексирования (кодирования) содержащейся в текстах информации с помощью ограниченного количества так называемых ключевых слов — основы поисковых дескрипторных словарей, а также матричное представление содержащейся в источниках фактической информации. Наконец, сюда относятся и методы специального (целевого) преобразования первоисточников ранее уже использованной информации в так называемые банки данных, пригодные для хранения на электронных носителях, используемые в ходе вторичной обработки. Этот перечень можно продолжать.

При всем многообразии этих методов их объединяет одна общая и, пожалуй, наиболее существенная характеристика: они призваны минимизировать неизбежный элемент субъективности при изучении текстовых материалов и поставить анализ текстов на возможно более объективную научную основу. Отсюда тенденция к разработке обоснованных критериев как для отбора источников, подлежащих информационному анализу, так и для формализованного представления содержащейся в них информации — неременного условия последующей статистической обработки. Качественные содержательные выводы информационного обследования соответствующей текстовой выборки опираются именно на статистические показатели.

В этой теме будет рассмотрена новая разновидность информационного анализа текстов, основанная на

семиосоциопсихологической концепции социальной коммуникации. Эта концепция акцентирует внимание на семиотических механизмах общения, формирующих действия порождения текстов и их интерпретации. Целью текстовой деятельности является реализация определенного содержательного замысла общения, а ее продуктами и объектами — тексты: научные и технические, литературно-художественные и публицистические, правовые, политические и др. Особенностью текстов является их способность сохранять в себе образы замыслов вне зависимости от той семиотической системы, в которой они порождены: книги, ноты, картины, чертежи и т. д.

Трактовка текста как единицы общения позволяет определить его как иерархию коммуникативно-познавательных программ, которые объединены в целостную содержательно-смысловую структуру авторским замыслом (коммуникативным намерением). Это объединение функционально, т.е. ориентировано на достижение конкретной цели общения. На этой основе возникает трактовка взаимопонимания, а, следовательно, и диалога как смыслового контакта, который может возникнуть лишь в том случае, если партнеры по общению адекватно интерпретируют коммуникативные намерения, отраженные в содержании порождаемых текстов. Анализу процессов установления смысловых контактов в ходе социальной коммуникации и служит метод информативно-целевого анализа текстов, один из вариантов которого будет описан ниже.

Метод информационно-целевого анализа текстов средств массовой коммуникации был разработан Т.М. Дридзе¹ в рамках семиосоциопсихологического подхода.

Информативно-целевой анализ выявляет содержательно-смысловую структуру текста и соотносит ее с замыслом общения, что позволяет обнаружить возможные отклонения в интерпретации текста со стороны других участников общения, т.е. оценить успешность коммуникации. Анализируя тот или иной текст как воплощение замысла автора, исследователь пытается ответить на следующие вопросы:

- Почему и для чего в тексте вообще что-то говорится? Зачем публикуется текст?
 - О чем говорится в тексте (каков объект описания)?
 - Что говорится (какие сведения представлены)?
 - Как говорится (с использованием каких языковых средств)?

ИНФОРМАТИВНОСТИ ТЕКСТА

При информативно-целевом анализе тексту приписывается некоторая мера информативности, которая является ключевым для понимания этого вида анализа документов.

Информативность текста может рассматриваться с двух позиций:

1. Под информативностью понимается способность текста донести замысел, основные коммуникативные намерения автора до получателя информации. Отсутствие общего замысла означает, что автор не справился с задачей создания текста как единицы общения, не определил цель коммуникации.
2. Под информативностью понимается способность текста быть источником каких-либо сведений, т.е. текст рассматривается как совокупность единиц информации.

¹ См.: Дридзе Т.М. Информативно-целевой анализ содержания текстовых источников // Методы сбора информации в социологических исследованиях. М.;1990. Кн. 2.

Выделяют информативность первого рода, или *первичную информативность*, характеризующую способность конкретного текста передать замысел автора; а также информативность второго рода, или *вторичную информативность*, характеризующую потенциальную способность конкретного текста служить источником некоторых сведений, которые ищет читатель, не задумываясь о целях порождения автором этого текста. Как правило, читатели находят много полезного, интересного в тексте «поверх» или вопреки коммуникативным намерениям автора. Это означает, что первичная и вторичная информативности чаще всего не совпадают друг с другом, создавая «смысловые ножницы». В процессе коммуникации вторичная информативность весьма часто не совпадает с первичной, а будучи оторванной от нее, приводит к искажению смысловой нагрузки, заложенной в тексте.

Когда истолкование замысла получателем информации соответствует замыслу автора, то интерпретацию текста можно считать адекватной. Элементы текста, способствующие адекватному истолкованию, обуславливают ту или иную меру информативности. Таким образом, информативность может быть измерена, но не абсолютным количеством информации в тексте, а той ее долей, которая используется для истолкования замысла автора.

Социолог, приступая к анализу текстов, исходя из своих исследовательских задач, тоже рассчитывает на вторичную информативность: преобладающее большинство документов, циркулирующих в обществе, не создается для социолога, для целей научного анализа, у их авторов — свои собственные коммуникативные замыслы, зачастую далекие от интересов социологии.

В этом случае информативно-целевой анализ больших текстовых разноплановых массивов ориентирован на изучение содержания с целью выявления, сопоставления, систематизации и сведения в системы данных о социальных явлениях или процессах. Здесь конкретный текст признается

информативным в той мере, в какой в его содержании обнаруживаются сведения, соответствующие целям исследования.

В этом своем варианте информативно-целевой анализ соединяет в себе элементы дедуктивного анализа (контент-анализ), индуктивного, где исследователь идет от текста, и информационного, когда извлеченные из текста индикаторы смысловых единиц «ставятся в жесткие координаты существующих между ними семантических отношений»².

Для оценки первичной информативности текста исследователь анализирует макро- и микроструктуры текста, где макроструктура — это иерархия разнопорядковых смысловых блоков, а микроструктура — набор внутритекстовых связей между опорными смысловыми узлами текста (фактами), образующими логико-фактологическую цепочку, на основании которой рассчитывается гипотетический коэффициент информативности. По величине этого коэффициента судят об уровне первичной информативности текста (об этом см.далее).

Информативность характеризует не абсолютное количество информации в тексте, не общую его насыщенность информацией в соответствии с коммуникативным намерением автора, но лишь ту ее долю, которая станет достоянием реципиентов, будет освоена множеством предполагаемых интерпретаторов.

Выявить подлинную информативность текста можно только в эксперименте, однако предварительно можно вывести гипотетический коэффициент информативности того или иного текста.

Введенное понятие первичной информативности текста показывает цель (отражающую коммуникативное намерение) отправителя. Мы сочтем текст лишенным первичной информативности, если он оказался информативным с точки зрения каких-то весьма конкретных целей его потребления реципиентом, но побочных для автора данного текста. Читатель

² ¹Дридзе Т.М. Указ.соч.. С. 102.

не всегда стремится (да и не всегда умеет) извлечь из текста основное коммуникативное намерение его автора.

Просматривая и время от времени читая тексты, исходя из так или иначе понимаемых своих собственных коммуникативно-познавательных потребностей, читатель нередко извлекает из сообщений множество разнообразной информации, которую считает для себя полезной и которую стремится найти. В этом последнем случае мы имеем дело со вторичной информативностью текста. Она является следствием общей информационной насыщенности текста, которая позволяет извлекать из него самые разнообразные сведения, побочные относительно коммуникативного намерения автора. Такие сведения могут возникнуть и как ассоциации или выводы, совершенно не предусмотренные коммуникативным намерением автора, в том числе обусловленные неумением общаться, отсутствием установки на диалог с партнером.

Следует отметить особые случаи неадекватной интерпретации текстов, когда исследователь обращается к их анализу со специальными научными целями, отнюдь не совпадающими с коммуникативными целями автора. Именно для решения таких задач создаются конкретные методики информативно-целевого анализа, предназначенные для обследования больших текстовых массивов с целью извлечения информации социологического характера.

Организовать текст так, чтобы коммуникативное намерение адекватно интерпретировалось адресатом, значит сделать его информативным в первичном смысле этого слова и в этом же смысле эффективным. Обеспечивая тексту структурно-смысловую целостность, являющуюся одной из наиболее существенных его характеристик, основная концепция текста (коммуникативное намерение автора) может и не иметь в нем словесно выраженной формулировки. Тем не менее она всегда присутствует в виде определенной мотивации и в случае необходимости может быть словесно сформулирована. Отсутствие цементирующего текст замысла

означает, что коммуникатор не справился с задачей создания текста как единицы общения или в лучшем случае запрограммировал определенную нечеткость воплощения своей коммуникативно-познавательной цели.

АЛГОРИТМ ИНФОРМАТИВНО-ЦЕЛЕВОГО АНАЛИЗА

Прежде чем анализировать текст, следует прочесть его полностью от начала до конца, как минимум, дважды, вдумавшись в его содержание и характер (назначение). Первоначально следует освоить методику на небольших текстах. Впоследствии, когда будет выработан навык адекватного общения, необходимость перечитывать текст отпадет сама собой. Она сохранится лишь в случае препарирования текста для обучения или в ходе эксперимента, когда требуется оценить меру адекватности интерпретации конкретного текста его читателями.

Следует спросить себя:

1) Представлена ли в тексте и чем именно та проблемная ситуация, к разрешению которой стремился (стремится) автор теми или иными средствами? Имеется ли в тексте анализ, или текст содержит лишь описание, констатацию, оценку этой ситуации?

2) Можно ли ответить на вопрос — ради чего, для чего вообще был написан (порожден) этот текст? Чем скорее всего мотивировал бы автор написание именно этого, а не какого-либо иного текста? Что именно хотелось бы ему вообще сказать этим текстом? Что именно хотелось бы ему сказать этим текстом в первую очередь? Как бы скорее всего он сформулировал свой замысел, свое коммуникативное намерение или ближайшую содержательную цель и в рамках какой более широкой сверхзадачи?

При повторном чтении текста (оно может быть на первых порах и трех- и четырехкратным) следует сосредоточить внимание на поисках непосредственной содержательной цели сообщения, реализация которой входит в коммуникативное намерение автора, в его замысел (т.е. выявить предикацию первого порядка) и выписать ее из текста. Следует обратить

внимание на то, что цель сообщения может и не иметь в тексте словесно выраженной формулировки. В этом случае ее надо сформулировать своими словами и записать. Остальное содержание текста или все его содержание, если в тексте не удалось обнаружить словесно выраженной предикации первого порядка, принимается за общее содержание текста.

Внутри общего содержания текста нужно затем выделить следующие элементы:

1. Основные тезисы-аргументы — некоторые предложения, являющиеся опорными для подкрепления цели сообщения; тезисы могут констатировать какие-либо положения рассматриваемой ситуации, выводить постулаты или описывать установки.
2. Разъяснения (развертывание) тезиса — предложения, содержащие комментарии (например, оценку преимуществ и недостатков) к выдвинутому тезису.
3. Описание или анализ проблемной ситуации, из которой вытекает основной тезис.

Элементы содержания текста могут иметь второстепенное значение для развертывания цели сообщения. К второстепенным элементам текста относятся иллюстрации к тезисам, пояснения тезиса или оценке ситуации, а также элементы фона (описания событий, являющихся фоном для развития рассматриваемой ситуации).

Обобщая вышесказанное, можно определить возможности метода следующим образом. Информативно-целевой анализ позволяет:

- оценить воплощение цели коммуникации, заложенной в тексте;
- определить структуру текста;
- выяснить возможность интерпретаций содержания текста;
- выяснить адекватность восприятия текста;
- осуществить «сжатие» текста для удобства анализа или работы с ним.

Пример: Предикация 1-го порядка: «Голосуйте за кандидата Х» Предикация

2-го порядка: «О его честности говорит бескомпромиссная борьба с ворами на его родном заводе». Предикация 3-го порядка: «Кандидат Х вышел из простой семьи, поэтому он хорошо знает интересы простого народа, а также честен, смел и не замечен в порочащих связях». Предикация 4-го порядка: «В наше нелегкое время непрекращающихся кризисов России нужен президент, который выведет страну из тупика».

Таким образом, информативно-целевой анализ состоит из двух блоков процедур: первый блок связан с определением иерархической структуры элементов текста, во втором выявляется избыточность задействованных автором символов (слов) для передачи своей идеи. Реализация поставленных задач анализа требует от исследователя хорошего знания текста, поэтому перед началом работы следует детально ознакомиться с содержанием, прочитав текст несколько раз.

Последовательность процедур информативно-целевого анализа может быть следующей³:

Первый блок - определение иерархической структуры элементов текста (рис.1)

1. Ознакомление с содержанием текста.
2. Выделение из текста предложения (словосочетания) или даже абзаца, отражающего цель сообщения, его замысел.

Определить цель сообщения помогают такие вопросы:

- Ради чего автором написан данный текст?
- Что именно хотел сказать автор?

Цель сообщения может не иметь четкой формулировки; в этом случае ее необходимо сформулировать своими словами. Оставшаяся часть текста — это общее содержание текста, в рамках которого будут выделяться смысловые блоки (элементы текста).

3. Выделение основных элементов в рамках общего содержания:

³ См.: Андреев В. Г., Маслова О. М. Методы сбора информации в социологических исследованиях. Кн. 2. Организационно-методические проблемы опроса. Анализ документов. Наблюдение. Эксперимент, М., 1991.

Элементы первого уровня представляют собой тезисы-аргументы.

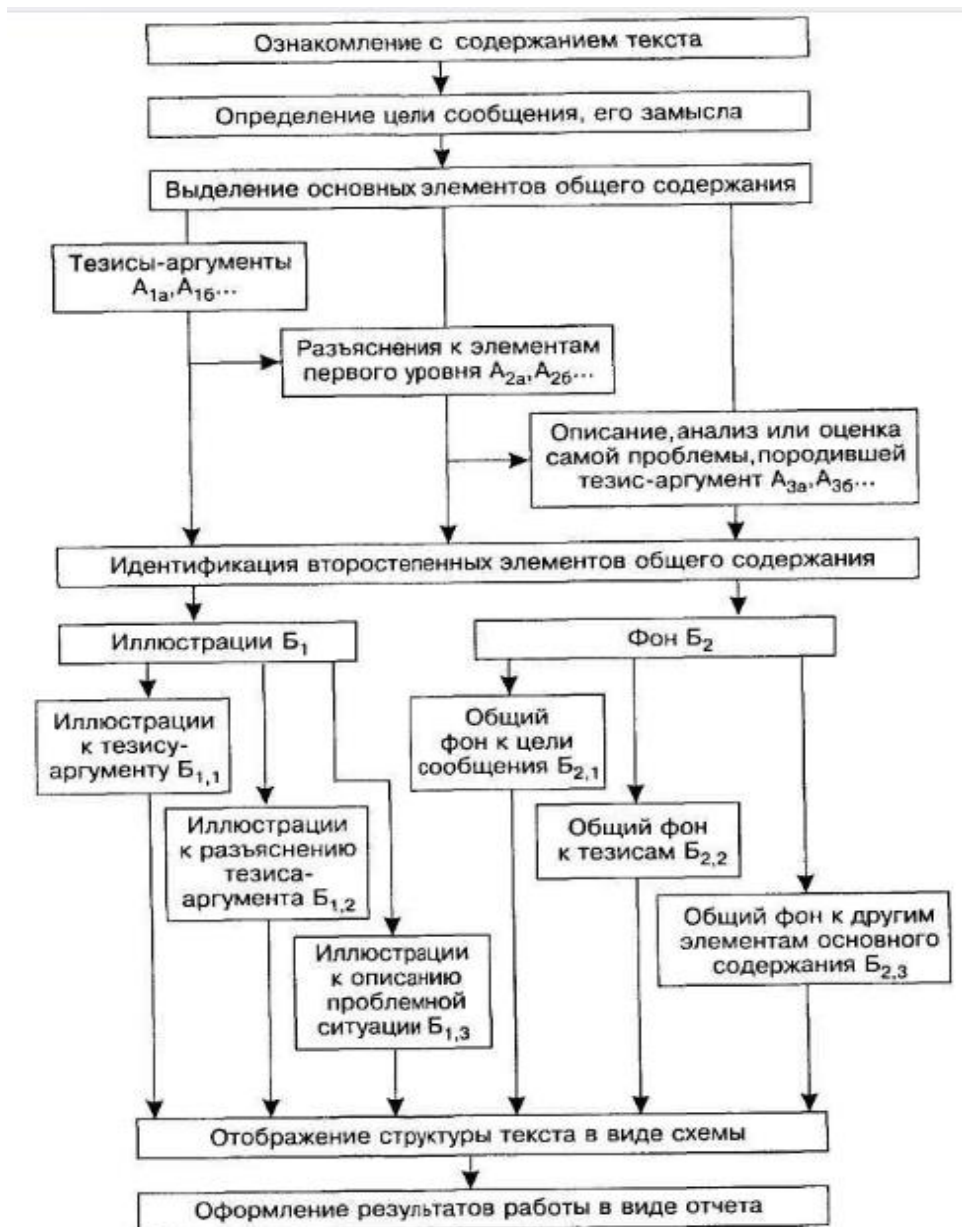


Рис. 1. Определение иерархической структуры элементов текста

(Источник: Андреенков В. Г., Маслова О. М. Указ.соч.)

Тезисы-аргументы — это элементы текста, которые являются опорными для цели сообщения, подкрепляющими ее развитие. Тезис-аргумент может выступать в виде констатирующего высказывания, в виде постулата или установки. Для составления структуры текста эти элементы обозначаются A_1 ; при наличии нескольких тезисов используются индексы A_{1a} , $A_{1б}$ и т. д.

Элементы второго уровня представляют собой разъяснения к элементам первого уровня. Для составления структуры текста эти элементы

обозначаются A_2 ; при наличии нескольких элементов разъяснения используются индексы A_{2a} , $A_{2б}$ и т. д.

Элементы третьего уровня представляют собой описание, анализ или оценку самой проблемы, породившей тезис-аргумент. Для составления структуры текста эти элементы обозначаются A_3 .

4. Выделение второстепенных элементов в рамках общего содержания. Второстепенные элементы содержания подразделяются на три группы иллюстраций. В зависимости от иерархии элементов основного содержания выделяют *иллюстрации к тезису-аргументу* (для составления структуры текста эти элементы обозначаются $B_{1,1}$), *иллюстрации к разъяснению тезиса-аргумента* (для составления структуры текста эти элементы обозначаются $B_{1,2}$), *иллюстрации к описанию проблемной ситуации* (для составления структуры текста эти элементы обозначаются $B_{1,3}$).

Второй блок выявляет избыточность задействованных автором символов (слов) для передачи своей идеи (рис. 2).

Дополнительные сведения, имеющие косвенное отношение (фон) к цели сообщения при составлении структуры текста обозначаются как B_2 . В зависимости от иерархии элементов основного содержания выделяют общий фон к цели сообщения, общий фон к тезисам и другим элементам основного содержания (для составления структуры текста эти элементы обозначаются как $B_{2,2}$, $B_{2,3}$ и т. д.).

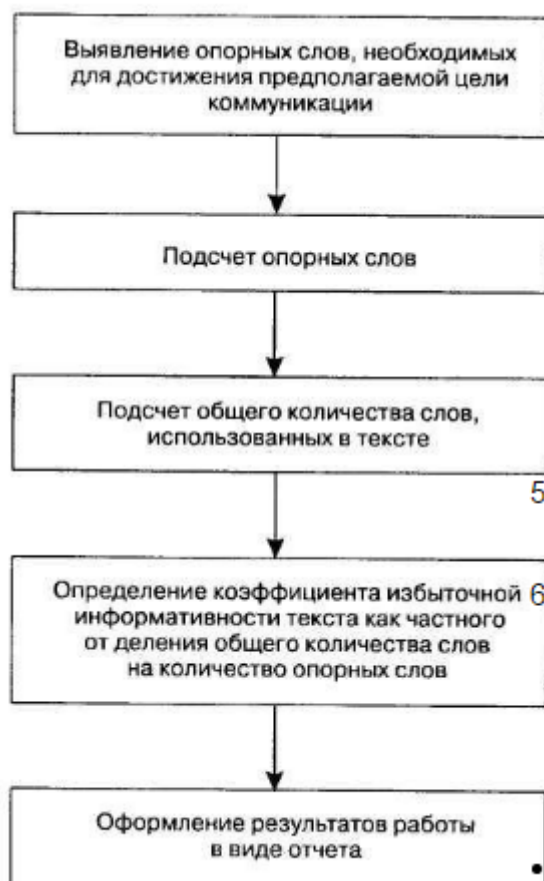


Рис.2 Определение избыточной информативности

(Источник: Андреенков В. Г., Маслова О. М. Указ.соч.)

Указанные элементы текста могут быть выделены при наличии таковых в тексте. Их отсутствие желательно фиксировать, что полезно при сравнении структуры различных текстов.

5. Схематическое изображение выделенных смысловых цепочек структуры текста (рис. 3).

6. Определение коэффициента избыточной информативности текста. Для этого необходимо:

- выделить слова, являющиеся необходимыми для достижения цели коммуникации, заложенной в тексте. Как правило, слова являются опорными для изложения содержания и связаны между собой в единую смысловую цепочку;
- пронумеровать выделенные слова;
- подсчитать общее количество слов, использованных в

тексте;

- определить коэффициент избыточной информативности текста как частное от деления общего количества слов на выделенные опорные слова. Чем ближе полученный в итоге коэффициент информативности (избыточности) текста к единице, тем менее он избыточен по отношению к авторскому замыслу. Таким образом, за идеальное значение гипотетического коэффициента условно принята единица, т. о. такая ситуация, когда все синтаксемы текста включаются в состав логико-фактологической цепочки, ориентированной на цель сообщения.

7. Оформление результатов работы в виде отчета.

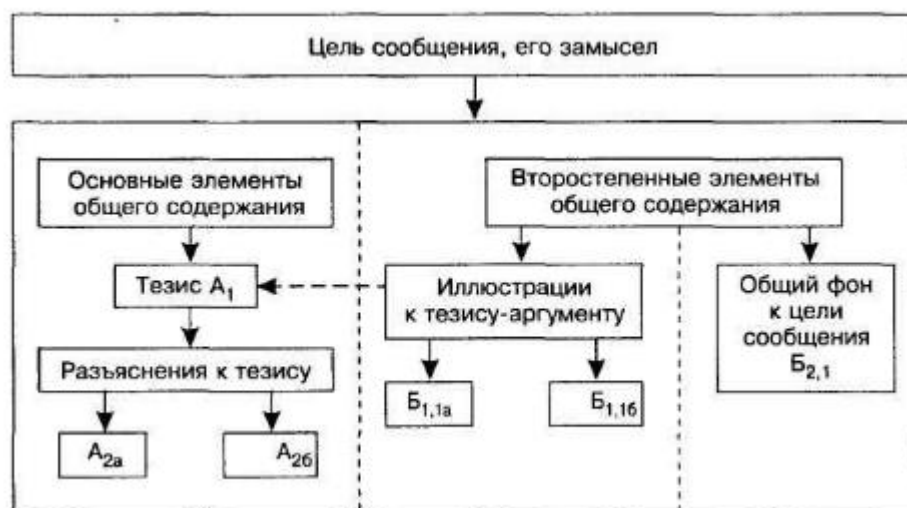


Рис. 3. Схематическое изображение структуры текста

Информативно-целевой анализ как метод изучения содержания документальных источников ориентирован на свойство их вторичной информативности. Это свойство допускает многофункциональную интерпретацию содержания текста. Обращаясь к тексту, пользователь может извлекать не только ту информацию, которая отражает коммуникативное намерение автора, но и другую, отвечающую познавательным целям самого пользователя. Социолог, использующий в своих исследованиях массовые документальные источники для получения эмпирической социологической

информации, постоянно имеет дело именно с вторичной информативностью документов.

Большинство документов, к которым обращается социолог, первоначально создавались их авторами совсем не с теми целями, которые преследуются в социологических исследованиях. Информативно-целевой анализ представляется единственно возможным средством изучения больших выборок разнопредметных и разнотипных текстов, описывающих самые разноплановые подсистемы общественных отношений, с целью выявления, сопоставления, систематизации и сведения в системы данных о сложных социальных объектах.

В этом случае на первый план выступает вторичная информативность текстов, а в ходе обработки текстового массива выявляются и систематизируются только те структурные единицы текстов, которые позволяют выявить и систематизировать интересующие исследователя факты. Иными словами, каждый конкретный текст признается информативным для исследования лишь в той мере, в какой в его содержании обнаруживаются сведения, соответствующие целям предпринимаемого исследования.

Особое значение приобретают гипотезы о характеристиках феномена, реконструируемого из содержания текстов. На основании таких гипотез формируются индикаторы изучаемых характеристик феномена, а также единицы наблюдения и регистрации в содержании анализируемых текстов. В своем варианте информативно-целевой анализ текстов соединяет в себе элементы, свойственные различным методам информационного анализа — дедуктивному (контент-анализу), когда исследователь движется к тексту от априорно сконструированных им аналитических категорий, и индуктивному, при котором исследователь идет от текста, фиксируя использованные в нем термины и терминологические конструкции (в данном случае те, которые соответствуют задачам исследования). Рассмотрим пример реализации информационно-целевого подхода.

Допустим, что предполагается реконструировать из текстов представления о нормативных характеристиках некоего образа жизни. В этом случае необходимо, во-первых, определить исходный теоретический принцип его рассмотрения и, во-вторых, сообразуясь с этим принципом, избрать категории для анализа текстов. Если в качестве исходного избирается, например, теоретический принцип деятельности, который в наибольшей степени отвечает сущностным характеристикам образа жизни, то необходимо извлечь из текстов данные, характеризующие как саму деятельность, так и условия ее протекания.

Исследователь вправе ограничить задачи любого конкретного исследования. Допустим, что ограничением для рассмотрения деятельности будет служить представление о сфере деятельности, а ограничением для рассмотрения условий деятельности — принятая в обществе система норм, имеющая немаловажное значение для реализации деятельности. Тогда на основе информативно-целевого анализа текстов, в той или иной мере содержащих сведения об избранной для изучения сфере образа жизни, можно, с одной стороны, получить целостное представление о содержании и структуре нормативного сознания (перечни действующих общих и конкретных правовых, морально-этических, эстетических и прочих норм и общепринятых ценностей) как одном из условий деятельности социальных субъектов.

С другой стороны, могут быть получены сведения о распространенных и рекомендуемых моделях деятельности и взаимодействия: перечни функций и признаков, присущих разным социальным субъектам, общих и частных целей деятельности, а также перечня самих действий, операций, способов и инструментов деятельности, применяемых в тех или иных ситуациях взаимодействия.

Обращаясь к разнопредметным источникам, можно восстановить из текстов целостные представления не только о сложных социальных феноменах, но и о сложных феноменах естественного характера.

Необходимым предварительным условием как в том, так и в другом случае является наличие у исследователя четкого представления о цели проводимого исследования и гипотезы относительно общих принципов структурирования явлений и процессов, сущностные характеристики которых предполагается реконструировать из текстов.

При этом важно понимать, что одним из результатов такого анализа окажется весьма существенная корректировка исходных принципов структурирования. Иными словами, предлагаемая методика дает возможность моделировать сложные феномены, основываясь на больших массивах текстов междисциплинарного характера.⁴

Преимущества метода:

- Присутствует определённый уровень формализации.
- Этот метод позволяет измерить силу информированности текста.
- Техника хороша для работы с результатами опросного метода

Недостатки метода:

- Невозможно работать с большими текстовыми массивами.
- значительная степень субъективности в полученных результатах.

⁴ В. Г. Андреенков, О. М. Маслова. Методы сбора информации в социологических исследованиях. Кн. 2. Организационно-методические проблемы опроса. Анализ документов. Наблюдение. Эксперимент - М.: Наука, — 224 с. 1991.