

*Г.В. КАНЬГИН*

## **ДАННЫЕ В СОЦИОЛОГИЧЕСКОМ ИССЛЕДОВАНИИ**

«Данные» являются одним из самых широко используемых понятий современной науки. В статье представлено представление о данных рассматривается в связи с развитием специализированных программных средств для гуманитарных приложений. Здесь можно выделить программные продукты двух классов: пакеты программ для анализа качественных данных (АКД) и системы компьютерного ассистирования интервьюированию (КАИ). Главным ресурсом развития программных средств для социолога являются высокие технологии. В составе пакетов АКД высокие технологии помогают исследователю при анализе плохо структурированной информации, а в КАИ системах способствуют получению надежных результатов анкетирования.

Адаптация современных информационных технологий (различные среды программирования, Интернет-решения, новые форматы данных и т. п.) к задачам вычислительного ассистирования для социолога сама по себе является трудоемким процессом; разработчики таких компьютерных средств не всегда осознают важность модельного выражения собственно гуманитарных проблем.

В данной работе мы преследуем две основные цели. Во-первых, показать специалистам по компьютерным технологиям необходимость учитывать методологические принципы социальных наук при создании для них программных средств. Во-вторых, развить структурную модель данных, которая позволяет конструктивно выразить проблему неоднозначности понятий, характерную для любого гуманитарного проекта. Предлагаемая модель данных рассматривается как проблемно-ориентированный ресурс разработки программных средств для социологов; в частности, речь пойдет о пакетах АКД и КАИ систем.

### **Компьютерное ассистирование в социологическом исследовании**

*Анализ качественных данных.* На «монографическом» сайте [1], посвященном программным средствам АКД, приводятся описания десятков систем. Среди них такие широко известные, как Atlas.ti, BEST, HyperResearch, NUD\*IST, NVivo, WinMAX и др.

---

**Каныгин Геннадий Викторович** — кандидат экономических наук, старший научный сотрудник Социологического института РАН.

**Адрес:** 196233 Санкт-Петербург, пр. Космонавтов, 52–5–20. **Электронная почта:** genaka@hotmail.com **Веб-сайт:** www.genaka.spb.ru

Статья написана на основе работ, поддержанных Российским фондом фундаментальных исследований (грант № 03–06–80273).

Рассмотрим, какие исходные идеи закладывают разработчики пакетов АКД в основу своих программных продуктов. В качестве примеров воспользуемся приводимыми на сайтах описаниями двух популярных программ анализа качественных данных.

NVivo является высокоразвитым программным инструментом, предоставляющим широкие возможности для анализа качественных данных, в том числе относительно плохо структурированных, не поддающихся адекватному переводу в числовую форму. «Исследователь, обращающийся к таким данным, как правило, нуждается в новом понимании изучаемого явления. При этом он либо исходит из формулировок самих респондентов, либо анализирует свои собственные полевые наблюдения, либо черпает факты из документов» [2].

ATLAS.ti предназначен для качественного анализа больших массивов текстовых, графических, аудио и видео данных. ATLAS.ti предоставляет разнообразный набор средств, позволяющих решать задачи по систематическому анализу «неструктурированных» данных, то есть исследовательских материалов, которые не могут быть удовлетворительно изучены с помощью формальных или статистических подходов [3].

Таким образом, пакет АКД позволяет обрабатывать исследовательские материалы «неколичественной» природы — качественные данные, или неструктурированную информацию.

Как отличить этот вид материалов от данных, которые поддаются обработке, например, с помощью традиционных статистических техник? Для того чтобы конструктивно выразить эти различия, представление о качественных данных обычно переопределяется через понятия, заимствованные из современной информатики. Типичным является сведение понятия качественной информации к типам «нечислового» поля, например, «мемо», «форматированное мемо», «ссылка», «изображение» и т. п. При этом и само понятие поля, и его различные спецификации считаются известными из соответствующих областей информатики. Указанное переопределение представлений о качественных данных страдает одним заметным недостатком. Хотя понятие качественных данных становится конструктивным за счет обращения к компьютерно реализованным конструкциям, оказывается, что специфика «качественных материалов» выражается в уже существующих информационных терминах. Подобная метафора прикладного программного обеспечения является принципиально неудовлетворительной: прикладные программы должны развиваться исходя из требований своей предметной области.

**Компьютерное ассистирование интервьюированию.** Примерами популярных систем такого рода служат: Blaise, BellView, MaCATI, CASES, Tric-Trac, Sphinx и др. Эти системы предназначены

для анкетных опросов различного вида: телефонного, почтового, через Интернет, самоопроса респондентов и т. п.

На сайтах, посвященных этим системам, приводятся сведения об их отличительных чертах. Например, специалисты, развивающие Blaise, подчеркивают наличие в этой системе мощного командного языка и интегрированных программных инструментов, удовлетворяющих любым исследовательским запросам при проведении анкетирования [4]. Создатели BellView обращают особое внимание на графический интерфейс анкеты [5]. Разработчики Sphinx предлагают развитые функции для работы с текстом при анкетном опросе (версия Lexica) [6]. Достоинства этих программных продуктов несомненны, и мы не будем на них останавливаться. Однако хотелось бы отметить, что в случае КАИ систем, как и в случае пакетов АКД, разработчики программных средств для социологических исследований стремятся скорее к заимствованию информационных достижений, чем к выражению гуманитарной специфики своих программных продуктов.

Применительно к КАИ системам в сравнении с АКД такая безусловная информационная ориентация выглядит еще более странной. Для анкетного опроса существуют методологические разработки, которые могут быть положены в основу программной архитектуры КАИ систем. Мы имеем в виду, в первую очередь, двухкомпонентную модель анкетного вопроса, предложенную С. Садменом и Н. Брэдберном [7].

Фактическое моделирование сугубо проблемной стороны социологических методов позволяет повысить надежность результатов социологического исследования не только за счет методических процедур, развивающих предложения указанных авторов и получивших название когнитивных методов анализа опросного инструмента [8], но и за счет проблемно-ориентированного компьютерного ассистирования. Чтобы перенести идеи когнитивного анализа в область, связанную с разработкой программного обеспечения социологических опросов и анализом эмпирического материала нечисловой природы, мы предлагаем собственное понимание того, чем являются данные в социологическом исследовании. Вводимый нами принцип разделительной трактовки данных есть средство структурирования предметной области компьютерного ассистирования, предназначенного для гуманитарных проектов.

### **Структура данных**

**Традиционная трактовка данных.** В различных программных средствах, например, системах управления базами данных (СУБД) и статистических пакетах, широко применяемых для социологических приложений, обычная процедура создания нового файла данных выглядит следующим образом. Сначала пользователь заполняет таблицу

описания полей структуры (табл. 1), а затем программа по этой таблице создает файл данных (табл. 2).

Таблица 1

Описание полей структуры данных		
ИМЯ ПОЛЯ	ТИП ПОЛЯ	РАЗМЕР ПОЛЯ
НОМЕР	ЦЕЛОЕ	
ИМЯ РЕСПОНДЕНТА	СТРОКА	25
ПОЛ	ЦЕЛОЕ	
ЗАРПЛАТА	ВЕЩЕСТВЕННОЕ	

Таблица 2

Файл данных			
НОМЕР	ИМЯ РЕСПОНДЕНТА	ПОЛ	ЗАРПЛАТА
1	ИВАНОВ	1	4500
2	ПЕТРОВ	1	6800
3	СИДОРОВА	2	5500

Что означают текстовые метки, выделенные жирным шрифтом в шапках обеих таблиц? Для табл. 1 это понятия, однозначно определенные в некоторой программной среде и описывающие типы полей данных, накапливаемых в табл. 2. Кроме того, указанные текстовые метки табл. 2 являются значениями поля ИМЯ ПОЛЯ табл. 1. В классическом понимании данных табл. 1 является определением для табл. 2.

Рассмотрим теперь заголовки (шапки) этих таблиц отдельно от строк значений, игнорируя структурную квазиреляционную связь между колонкой ИМЯ ПОЛЯ и заголовком табл. 2. Естественно считать, что в обеих таблицах понятия, вынесенные в заголовки, обладают свойством однозначности. Благодаря этому свойству в колонках таблицы появляются экземпляры этого понятия, наследующие однозначность от понятия, формирующего заголовок столбца.

Теперь обратим внимание на контекст определения данных, имеющий место в прикладном исследовании применительно к табл. 2. Прежде чем сформировать метки, например, в том виде, как они фигурируют выше (ИМЯ РЕСПОНДЕНТА, ПОЛ и др.), исследователь выполняет ряд действий. Хотя мы не можем предложить соответствующую модель этих действий, их смысл станет понятным, если перечислить некоторые варианты меток, которые могли бы появиться при создании рассматриваемого заголовка. Например, вместо указанного имени первого поля можно ожидать такие словосочетания, как: НОМЕР ПО ПОРЯДКУ, УНИКАЛЬНЫЙ НОМЕР, ИДЕНТИФИКАТОР РЕСПОНДЕНТА и т. п. Аналогично для имени второго поля — ФИО, ФАМИЛИЯ, РЕСПОНДЕНТ и т. д. Иными словами, введение в заголовок той или иной метки является некоторым когнитивным действием. Его цель — ориентировать актора, воспринимающего метку в исследовательских ситуациях (коммуникация при сборе данных, составление документов, обработка результатов и т. п.), на правильное

понимание тех ассоциаций, которые имеет в виду автор, предложивший метки. Такие действия отчетливо проявляются при разработке сугубо социологических представлений (ИМЯ РЕСПОНДЕНТА, ПОЛ, ЗАРПЛАТА и т. п.) и поэтому пояснялись нами на примере табл. 2. Однако мы относим подобные ментальные операции и к формулированию меток, принадлежащих заголовку табл. 1. Для того чтобы сформулировать заголовок в виде ИМЯ ПОЛЯ, ТИП ПОЛЯ, РАЗМЕР ПОЛЯ, необходимо ответить на такие вопросы, как: почему не НОМЕР ПОЛЯ; зачем РАЗМЕР ПОЛЯ, если есть ТИП ПОЛЯ и т. п., — иными словами, выполнить «социологическую» работу по прояснению ассоциаций, стоящих за указанными метками заголовка.

Обычно считается, что операции по определению данных в форме табл. 1 отделены от действий по работе с данными в форме табл. 2. Для подтверждения этого тезиса можно вспомнить, как появляется файл данных, например, в статистическом пакете SPSS™: сначала форма описания полей, а затем — операции с данными. Однако если принять во внимание отмеченные нами когнитивные действия, то можно посчитать, что определение данных — это формирование заголовков в каждой из таблиц, а работа с данными — это манипулирование со значениями строк в таблицах. Такой взгляд отличается от классического определения, при котором одни данные (в форме табл. 1) являются определением для других данных (в форме табл. 2).

**Разделительная трактовка данных.** Попробуем обобщить приведенные соображения по определению данных с помощью понятия поименования. Будем называть поименованием *формально-когнитивную конструкцию*, связывающую однозначно идентифицируемую компоненту (метку) и когнитивные ассоциации. Назовем определением понятия ряд поименований, которые осуществляет исследователь с целью перехода от неясных ассоциаций к исследовательским терминам. Введем нотацию для процедуры определения понятия в виде формулы (1):

$$F_s^{(i)}, F_s^{(i-1)}, \dots, F_s^{(1)}, F^{(0)} \quad (1)$$

Рассмотрим, что подразумевается под обозначениями, входящими в состав выражения (1). Термин «поименование» предполагает, что за каждой из меток  $F_s^{(i)}$  стоят связанные с ней ассоциации, то есть когнитивные конструкции, понимаемые в русле теории метафоры<sup>1</sup>. В составе определения каждое поименование  $F_s^{(i)}$  имеет однозначную метку, которую следует отличать от стоящих за ней ассоциаций. Метка — результат поименования. Под меткой мы понимаем любое конструктивное выражение ассоциаций автора, то есть конструкции,

<sup>1</sup> О понятии ассоциации в контексте нашего подхода см.: [9].

безусловно воспроизводимые на всем множестве акторов. Например, алфавитно-цифровая последовательность (текст), простой рисунок, общеизвестный жест и т. д. Для сохранения конструктивного представления о метке мы будем считать ее компьютерно моделируемой в виде различных информационных объектов — графических файлов, текстовых переменных, компонентов среды программирования и т. д.

Ключевой вопрос — как представить обращение к ассоциациям с помощью метки — решается в традиционно социологической манере. Формальная компонента поименования  $F_s^{(i)}$  предполагает наличие носителя ассоциаций и позволяет ему обращаться к своим ментальным образам при помощи обозначения. Нижний индекс  $s$ , реферирующий к слову *subject*, подчеркивает, что такое обращение является авторским, то есть существует только для того, кто предложил метку<sup>2</sup>. Любой другой актор, использующий любую из меток  $F_s^{(i)}$ , обращается к другому набору ассоциаций. В рамках процедуры определения для актора не существует иного способа выяснить, какие ассоциации стоят за любой из меток  $F_s^{(i)}$ , кроме получения от ее автора новых поименований.

Индекс  $i$  указывает, что в рамках одной процедуры определения может быть предложено несколько меток, за каждой из которых стоит свой ментальный образ. Мы используем обычное понимание индекса, при котором  $i$  — это натуральное число. Это означает, что, с одной стороны, обозначения  $F_s^{(i)}$  различаются между собой, с другой, — что они образуют некоторую последовательность, идентифицируемую с помощью значений индекса:  $i, i-1, \dots, 0$ . Наличие такой последовательности не означает существование меры, с помощью которой введен порядок поименований от  $F_s^{(i)}$  к  $F_s^{(i-1)}$ . Эта последовательность является отражением шагов сведения неясных ассоциаций автора к однозначному понятию. Такое сведение только именуется стадией, через которые проходит автор при определении конкретного понятия. Итак,  $F_s^{(i)}$  и  $F_s^{(i-1)}$  — это два разных поименования.

Результат применения поименований, или процедуры определения, всегда мыслится как однозначное понятие. Однако весь процесс определения предполагает существование не гипотетической направленности — от нечетких ассоциаций к ясной разделяемой формулировке, — а цели, которая как раз и обладает требуемым свойством однозначности. Эта цель может быть достигнута с помощью любого социологического метода. Такие возможности социологического метода объясняются не его научностью, а наличием автора, который

<sup>2</sup> Существует и второй план этой нотации. *Subject* — это также предмет исследования, то есть введение индекса  $s$  символизирует то обстоятельство, что ассоциации обуславливаются для их носителя рамками некоторой предметной области.

декларирует «однозначную понимаемость» своих формулировок. Однако формулирование критериев того, состоялся ли переход от авторских образов к разделяемым терминам, находится в ведении все того же автора, то есть является авторским произволом. Поэтому номинальное различие, проводимое с помощью индекса  $i$ , и однозначно воспринимаемые понятия  $F^{(0)}$  существуют только в рамках конкретной процедуры определения. Такая упорядоченность целиком зависит от авторских соображений, и для нее не существует выражения в общем виде. Поэтому нельзя считать поименования номинальной переменной<sup>3</sup>.

В процедуре определения всегда существует метка, относительно которой автор полагает, что стоящие за ней ассоциации одинаковы для всех реципиентов определения. В нашей нотации этим свойством обладает метка  $F^{(0)}$ . Такие метки мы обозначаем без использования нижнего индекса  $s$ . Поименование имеет автора, то есть носителя языка, который предложил данную метку в качестве выражения своих ассоциаций, возникающих при рассмотрении того или иного фрагмента предметной области. Поименование имеет совокупность акторов (ареал поименования), то есть лиц, использующих поименование, например, как элемент социологического метода, при распространении результатов исследования или разработке методик.

Таким образом, нотация  $F_s^{(i)}$  имеет два принципиальных аспекта. Во-первых,  $F_s^{(i)}$  является однозначно воспринимаемым изображением. Во-вторых, поименование — это аналог отношения неоднозначности, которое существует в рамках любого социологического метода между исследователем и актором метода, например, социологом, составившим анкету и предложившим поименования, и респондентом, который оказывается их реципиентом. И социолог, и респондент — акторы поименования.

Будучи меткой, выразимой в «количественных» понятиях, например, таких, как *алфавитно-числовая последовательность*, в отличие от *подразумеваемых ассоциаций*,  $F_s^{(i)}$  может быть сформулирована в обычных формальных терминах, то есть представлена как переменная, вектор, матрица и т. п. Эти понятия однозначны для всех использующих их акторов. Поэтому  $F_s^{(i)}$  может быть изображено, например, как множество, состоящее из однозначно понимаемых элементов, то есть  $F_s^{(i)} = \{F_s^{(ik)}\}$ , где  $k=1 \dots N$ . Однако каждый элемент такого изображения  $F_s^{(ik)}$ , как и  $F_s^{(i)}$ , трактуется в нашем подходе как поименование, то есть любой элемент  $F_s^{(ik)}$  обладает теми же свойствами<sup>4</sup>, что и  $F_s^{(i)}$ . Поэтому в общей формулировке нашего подхода мы

<sup>3</sup> Далее мы показываем на примерах конкретных терминов, как выглядит подобная упорядоченность в прикладном исследовании.

<sup>4</sup> Конструктивное выражение этих свойств мы приводим ниже в пунктах «Однозначность» и «Конструктивность» последнего параграфа.

используем нотацию  $F_s^{(i)}$ , а разложение  $\{F_s^{(ik)}\}$  применяем в примерах для наглядности.

Последовательность поименований (1), приводящая от  $F_s^{(i)}$  к  $F^{(0)}$ , в силу своей зависимости от ассоциаций, связываемых автором поименования с каждой из меток  $F_s^{(i)}$ , также является субъектно-зависимой. Другими словами, хотя процедура определения приводит в своих рамках к понятиям, декларируемым как однозначные, не существует критериев, которые позволяли бы указать последовательность поименований, необходимых для такого приведения.

Используя компьютерную метафору, можно сказать, что метка в составе поименования является обращением к процедуре, выполняющей алгоритм поименования, а ассоциации, возникающие при когнитивном действии, — это сама реализация процедуры, обычно называемая ее телом. В терминах поименования классическое понятие данных имеет вид (2):

$$F^{(0)} \Rightarrow G^{(0)}, \quad (2)$$

где через  $F^{(0)}$ ,  $G^{(0)}$  обозначены различные между собой однозначные понятия, то есть поименования, имеющие один и тот же набор ассоциаций на всем ареале акторов. Знаком « $\Rightarrow$ » передана квазиреляционная связь между столбцом экземпляров понятия  $F^{(0)}$  и метками однозначно воспринимаемого поименования  $G^{(0)}$ , упомянутая в вышеприведенном примере определения данных. Именно в силу этого квазиреляционного отношения  $F^{(0)}$ ,  $G^{(0)}$  являются единым понятием данных, причем  $F^{(0)}$  считается определением для  $G^{(0)}$ .

В нашем подходе каждая из частей представления (2) имеет вид (1), поэтому классическое представление о данных должно быть переписано в виде (3):

$$F_s^{(i)}, F_s^{(i-1)}, \dots, F_s^{(1)}, F^{(0)} \Rightarrow G_s^{(i)}, G_s^{(i-1)}, \dots, G_s^{(1)}, G^{(0)}, \quad (3)$$

где  $F_s^{(i)}$ ,  $G_s^{(i)}$  — поименования, описанные выше при разъяснении формулы (1), а  $F^{(0)}$ ,  $G^{(0)}$  — данные в классическом понимании.

Если придерживаться традиционного взгляда на данные, то в нашей нотации следует рассматривать как единое целое каждую из процедур определения, то есть  $F_s^{(i)}$ ,  $F_s^{(i-1)}$ , ...,  $F_s^{(1)}$ ,  $F^{(0)}$  имеет одно структурное выражение, а  $G_s^{(i)}$ ,  $G_s^{(i-1)}$ , ...,  $G_s^{(1)}$ ,  $G^{(0)}$  — другое. Однако нам кажется предпочтительной другая структуризация представления о данных. Стремясь к выделению когнитивных свойств, проявляющихся в процедуре определения данных, мы считаем, что для любого из поименований  $F_s^{(i)}$ ,  $F_s^{(i-1)}$ , ...,  $F_s^{(1)}$  и  $G_s^{(i)}$ ,  $G_s^{(i-1)}$ , ...,  $G_s^{(1)}$ , присутствующих в соотношении (1), существует одна и та же структурная модель, которая отличается от структурной модели, описывающей каждую из компонент  $F^{(0)}$ ,  $G^{(0)}$ .



Иными словами, в данных всегда существует метафорическая составляющая  $F_s$  (4) и однозначно понимаемая  $F$  (5):

$$F_s = F_s^{(i)}, F_s^{(i-1)}, \dots, F_s^{(1)} \quad (4)$$

$$F = F^{(0)} \quad (5)$$

Метафорическая составляющая выражает для всех акторов данных процесс поименования, осуществляемого как обозначение авторских ассоциаций. Однозначно понимаемая или количественная составляющая представляет собой значения или экземпляры однозначных понятий.

### Графическая интерпретация

У одних акторов социологических определений — одни ассоциации с вводимыми терминами, у других — другие, но такие ассоциации существуют для каждого актора. Признав это явление, попытаемся дать ему графическую интерпретацию. Изобразим табл. 1 исходного примера в виде рисунка:

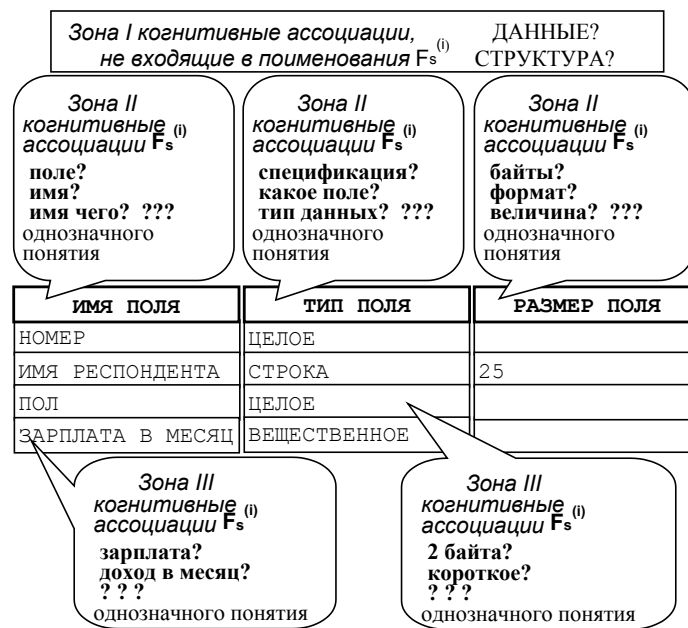


Рис. Условности определения данных

С помощью рисунка мы хотим показать, что определение данных — это не только перечисление экземпляров однозначных понятий, но и процесс порождения, развития и отсеивания когнитивных образов, которые предшествуют финальным формулировкам. Для каждого понятия могут быть перечислены связанные с ним поименования, то есть ассоциации, помеченные однозначно воспринимаемой

меткой. Совокупность таких поименований мы рассматриваем как процедуру определения понятия. Для простоты изложения мы представляем метки поименований в составе рисунка как тексты со значком вопроса: «поле?», «короткое?» и т. п. Однако метка может быть изображением, аудиофайлом, гиперссылкой и т. п., то есть любым понятием, имеющим конструктивную, например, компьютерную реализацию. Поместив изображение из трех знаков вопроса, мы хотим подчеркнуть, что состав поименований в общем случае неизвестен и подчинен лишь произволу автора.

Поименования сведены в три группы. В группе I находятся общие ассоциации, возникающие в связи с понятием данных. В группе II показаны поименования, приводящие к однозначным понятиям, которые используются для определения таблицы социологических данных. В группе III собраны ассоциации, возникающие в связи со значениями данных.

На рисунке наглядно проявляется особенность нашего взгляда на данные — любое однозначное понятие имеет набор предшествующих поименований. Не столь важно, где расположены эти однозначные понятия — в заголовке или в строках таблицы. Эти различия вторичны по сравнению с их общим свойством: однозначность этих понятий определяется только авторским произволом. Этот произвол выражается в том, что автор формулировок считает их одинаково понятными для всех использующих их лиц. При этом автор не предлагает никакого конструктивного способа проверки этого предположения.  $F^{(0)}$  — это те представления, которые в формальных системах называются первопонятиями, а в рамках конкретного социологического определения оказываются выражением исследовательской гипотезы об однозначности.

С традиционной точки зрения на данные, заголовок столбца отличается от значения, фигурирующего в нем, тем, что заголовок требует дополнительных описаний свойств столбца, а значения данных такого описания не требуют, так как черпают его из свойств самого заголовка. С нашей точки зрения, если автор захочет трактовать значения в строках не как однозначные понятия, а как поименования, то и строки следует признать неоднозначными. В этом случае приходится учитывать дополнительные ассоциации, например, те, которые показаны в зонах III. В классическом случае этого не происходит в силу соглашения, которое присутствует в самом понятии данных. Графически эти соглашения могут быть истолкованы как переход последовательности поименований, показанных метками в зонах II, в однозначные понятия. Например, совокупность поименований (спецификация? какое поле? ... тип данных?) составляют определение ТИП ПОЛЯ.

Соглашения, отмечаемые в традиционном представлении о данных, сформулируем в виде двух правил. Во-первых, значение в таблице



имя ответа. В-третьих, код градации — это целое число, а мнемоническое имя — это последовательность букв.

По существу же эта схема означает, что, отвечая на вопрос о поле респондента, интервьюер выражает понятие ПОЛ, имеющее неоднозначно трактуемые ассоциации, в реализациях однозначно понимаемых понятий ЦЕЛОЕ и СТРОКА. Такое утверждение может показаться нелепым в силу очевидных различий между ассоциациями, помеченных словами, например, ЖЕНСКИЙ и СТРОКА. Однако акторы никогда не обмениваются собственно ассоциациями: между акторами осуществляется только передача меток. Поэтому при конструктивных реализациях социологических методов, например, с помощью компьютерных средств, следует ориентироваться не на моделирование ассоциаций, скрытых в когнитивном аппарате исследователя, а на выражение действий с метками, которые осуществляет актор. Причем такая позиция будет являться не упрощением, на которое нас вынуждает недостаточное развитие компьютерных средств, а точной моделью того, как решается проблема неоднозначности понятий с помощью современных социологических методов. Действительно, при анкетировании исследователь с помощью вариантов ответа 1 — МУЖСКОЙ, 2 — ЖЕНСКИЙ закладывает структуру, состоящую из двух полей ЦЕЛОЕ<sup>5</sup> и СТРОКА<sup>6</sup>, и предполагает, что респондент останется при ответе в рамках их значений.

В чем состоят правила сведения понятий к их однозначным трактовкам? Во-первых, наличие в описании термина ПОЛ двух понятий КОД ГРАДАЦИИ и МНЕМОИМЯ ГРАДАЦИИ, а не трех или одного является авторской условностью, которая не меняет смысла получения однозначности. И в случаях, скажем, добавления третьего поименования — развернутого описания ТЕКСТ — или сохранения только одного описания — КОД ГРАДАЦИИ — исходный концепт ПОЛ подвергается аналогичным когнитивным преобразованиям. Смысл этих преобразований состоит в направленном сужении круга ассоциаций, стоящих за этой трехсимвольной меткой.

Во-вторых, то, что КОД ГРАДАЦИИ приведен к термину ЦЕЛОЕ, а МНЕМОИМЯ ГРАДАЦИИ к термину СТРОКА, также является условностью, которая не меняет основного требования процесса определения — однозначности финального понятия. Необходимость подобных редукций не отменяется путем назначения других однозначных терминов, например, КОРОТКОЕ ЦЕЛОЕ или СТРОКА 25.

<sup>5</sup> Это поле, где хранится код, раскрывающий исследовательское понятие ПОЛ.

<sup>6</sup> Это поле сохраняет короткую последовательность символов, описывающую то же исследовательское понятие.

Также условностью (в рамках процедуры определения понятия ПОЛ) является декларация свойства однозначности для поименования ЦЕЛОЕ. Вполне возможен вариант соглашения, при котором исходная процедура определения примет вид:

$F_s^{(2)} = \text{ПОЛ}$ ,  $F_s^{(1)} = \text{ГРАДАЦИИ ОТВЕТА}$ ,  $F^{(0)} = \{\text{КОД ГРАДАЦИИ, МНЕМОИМЯ РЕСПОНДЕНТА}\}$ . Такое обозначение будет соответствовать ситуации, при которой между всеми акторами определения ПОЛ существует общее понимание того, что такое КОД ГРАДАЦИИ и МНЕМОИМЯ РЕСПОНДЕНТА. Исходный вариант нотации описывает случай, при котором автор анкеты считает, что КОД ГРАДАЦИИ и МНЕМОИМЯ РЕСПОНДЕНТА нуждаются (для подразумеваемого круга акторов) в дальнейших уточнениях. Такие уточнения как раз и выполняются с помощью введения поименований {ЦЕЛОЕ, СТРОКА}, которые оказываются финальными терминами.

Конечно, считая поименование {ЦЕЛОЕ, СТРОКА} однозначным или безусловно понятным, мы тоже формулируем лишь одну из условностей процесса определения данных. Однако мы опираемся на то обстоятельство, что целью нашего исследования является анализ проблем компьютерного ассистирования, а понятия ЦЕЛОЕ и СТРОКА имеют конструктивную реализацию в различных системах программирования, поэтому не обладают ненаблюдаемой когнитивной составляющей, то есть не зависят от актора. Условность в данном случае возникает из-за того, что далеко не все акторы социологических методов представляют себе эти конструктивные реализации.

В-третьих, однозначность ответов актора — это одинаковость структур значений, которые служат для выражения изучаемого социального феномена (в примере — ПОЛ). Значения — это реализации, которые создают акторы в ситуации применения метода, в данном случае — анкетирования.

Иными словами, финальные описания в примере — ЦЕЛОЕ и СТРОКА — имеют статус однозначных не в силу того, что они охватывают все возможные ассоциации, возникающие в случае, если трактовать значение в данных как поименование, то есть метка плюс когнитивная ассоциация. Их статус однозначности возникает в связи с тем, что они предполагают однозначно понимаемую структуру, в терминах которой должен быть сформулирован «ответ».

Условность определения данных как раз и состоит в том, что по авторскому произволу из поименований возникают значения, то есть метки, связываемые всеми акторами с однозначно продуцируемыми ассоциациями. В качестве наблюдаемого признака однозначности таких ассоциаций возникает повторяемость меток или значений. Таким образом, возвращаясь к исходному примеру классического определения данных, в наших терминах можно сказать, что значения

столбца ИМЯ ПОЛЯ табл. 1 и заголовки полей табл. 2 — это однозначные понятия.

Вместе с тем, значения — это всегда экземпляры понятия, сформулированного как однозначное. А заголовок — это первое однозначное понятие в ряду поименований. Это не бросающееся в глаза различие в нашем подходе является более важным, чем свойства, возникающие в результате самой однозначности. Добиваясь однозначности понятий, посредством которых действует респондент в ситуациях опроса, социолог заведомо получает немоделируемость этих действий. Это означает следующее: респондент, конечно, укажет вариант ответа, то есть значение данных, но ментальный механизм, который позволит ему выполнить такое указание, окажется вне рамок схемы определения, приводящей к возможности самого указания.

Финальные понятия никоим образом не претендуют на моделирование когнитивного аппарата. Их цель прямо противоположная — избавиться от когнитивной основы социологических понятий, сделав конструктивным терминологический произвол. Таким образом, социологические определения не более чем мнемоническая последовательность, которая, разумеется, может быть полезной в прикладном исследовании, но никак не может считаться «научной», «обоснованной», «валидной» и т. п.

**Когнитивная природа принципа.** Когнитивные подходы, получившие развитие в связи с проблематикой анкетного метода, используют представление о когнитивной трудности — трудности, ощущаемой респондентом в том случае, если формулировка, входящая в текст вопроса или ответа, репрезентирует для него сюжет, изучаемый с помощью опроса. В работе Д.М. Рогозина [8] обстоятельно рассмотрены различные приемы, предлагаемые в современной методической литературе для фиксации трудностей такого рода.

В когнитивных подходах исследователь обращается к актору для улучшения словесного описания изучаемого фрагмента. За этим обращением стоит инструментальное использование актора, отличающее когнитивный подход от традиционных позитивистских схем. Например, от представлений о латентности, которые конструируют числовые оценки скрытых от непосредственных измерительных методик параметров на основе статистических гипотез исследователя. С помощью этих гипотез исследователь считает свои вербальные конструкции одинаково интерпретируемыми каждым из респондентов, то есть навязывает актору свою терминологию описания изучаемого фрагмента. В когнитивном подходе предполагается наличие внутреннего знания актора, на которое не распространяются модельные представления исследователя. В дополнение к нему всегда существует конструктивное или наблюдаемое выражение этого внутреннего знания — документы, тексты, формулировки и т. п. Актор используется

как инструмент, обладающий обеими составляющими когнитивного подхода, то есть когнитивным знанием и его текстовым (или любым другим конструктивным) выражением. Причем у актора существует только один способ выразить внутреннее знание — сформулировать наблюдаемую составляющую. Актор, соотносящий указанные компоненты друг с другом, играет роль исследовательского инструмента для выявления когнитивных трудностей.

Когнитивные подходы всегда предлагают лишь косвенные оценки конструктивных характеристик (меток, формулировок и т. п.), так как между такой характеристикой и порождающей ее когнитивной основой всегда находится актор. Причем оказывается, что это — «второй» актор. Например, в процессе анкетирования социолог (первый актор, или автор) использует респондента (второй актор) для проверки того, насколько удачны формулировки, выражающие авторское когнитивное знание.

Можно сказать, что отправной пункт современных когнитивных подходов — это идея полиграфа: исследователь не знает, почему «волнуется» респондент, то есть наблюдаемые характеристики не позволяют предсказать факт «волнения» респондента. Но факт «волнения» является индикатором когнитивных трудностей, которые подлежат устранению. Чем меньше значение «индекса волнений», тем лучше составлен опросный инструмент. Сами методы оказываются специфицированными по способу построения такого индекса. Например, метод артикулированного суждения [10] предполагает дополнительные вопросы респонденту, направленные на выявление мотивов его ответа; наличие интерпретационных затруднений при ответах фиксируется исследователем, то есть от актора требуется их осознание. При использовании метода временных характеристик [11], построенного на замерах продолжительности отдельных фаз коммуникации при интервьюировании, от респондента не требуется осмысления своего опыта ответов на предложенные вопросы.

Косвенный характер когнитивных методов, связанный с немоделируемостью ментальных механизмов, порождает естественное желание если не устранить, то ослабить эту немоделируемость. В этом случае естественной выглядит попытка описать ментальный аппарат актора. Одной из наиболее известных разработок такого рода является четырехшаговая схема ответа на анкетный вопрос, предложенная П. Туранжо [12]. Стратегический результат подобных подходов состоит в имитации ментальных механизмов акторов. Получение такого результата в будущем станет означать отказ от инструментального использования актора и тем самым устранение терминологического произвола социологических определений, о котором шла речь выше. Однако до тех пор, пока не удалось построить модели ментальных механизмов, при решении проблем вычислительного ассистирования

в социологии предпочтительным выглядит использование сугубо социологического описания социальной реальности в виде когнитивных схем.

Мы придерживаемся мнения, что когнитивные методы недостаточно последовательно освобождаются от позитивистской ориентации. Это проявляется в том, что второй актор рассматривается как инструмент измерения (пусть косвенного, пусть измерения чего-то ненаблюдаемого, скрытого внутри человеческого сознания). В результате такой измерительной подоплеки так же, как в случае измерительных схем, где сосуществуют факторный анализ, автоматическая классификация, метод детерминаций и другие техники, в когнитивных подходах развиваются различные приемы выявления ментальных затруднений акторов на формулировках вопросов.

Исходная неточность когнитивных взглядов состоит в отношении к вербальной деятельности гуманитария в социологическом исследовании как к научной работе. Более перспективным для решения проблем моделирования в социологии представляется следующий тезис. Когнитивные техники так же, как и классическое измерение, например, латентный анализ, не являются в рамках социологии научным методом. Когнитивность метода возникает не в связи с тем, что автор использует актора для проверки своих научных взглядов. Идея когнитивности предопределена тем, что социологические термины являются выражением не согласованного мнения научного сообщества, а произвола автора. Моделирование когнитивности не следует строить по естественнонаучным образцам, скорректированным за счет привлечения второго актора. Моделирование когнитивности — это создание средств для конструктивного выражения произвола при определении понятий первым актором, то есть исследователем.

Для компьютерного ассистирования это означает, что все средства, моделирующие авторский произвол при определении понятий, должны быть отданы автору этих понятий. На практике все терминологические построения социолога сводятся к данным, поэтому, другими словами, компьютерное ассистирование в социологии — это информационные модели терминологического произвола, осуществляемого социологом при создании баз данных.

**Конструктивность.** Один из главных принципов нашего подхода заключается в утверждении о фундаментальности терминологического произвола, который характеризует социологическое исследование. Под произволом социолога или произвольностью терминов в социологии мы понимаем присвоение гуманитарием себе права решать, до какой степени следует разъяснять используемые предметные понятия. Это утверждение, на первый взгляд, выглядит странным, если принять во внимание, например, упомянутые выше когнитивные



методы. Действительно, разве не использует исследователь респондента как стороннего наблюдателя с целью улучшить формулировки вопросов, то есть разъяснить предлагаемую терминологию?

Само суждение о произвольности исследовательских терминов в социологии может быть истолковано как оценка качества прикладных исследований. В этой связи тезис о произвольности терминов может быть подвергнут критике на различных основаниях. Особенно убедительной такая критика выглядит с позиций социолога, занимающегося какой-либо устоявшейся предметной областью. При ее идентификации обычно в названии на первом месте находится слово «социология»: социология семейных отношений, социология девиантного поведения, социология морали, социология общественных движений и т. д. Критика может исходить из «очевидного факта» существования «семейных отношений», «девиантного поведения», «морали», «общественных движений» и т. д. Такое существование зафиксировано в разговорной практике, в научных текстах, в сообщениях средств массовой информации и т. п. Поэтому возможно заключение, что говорить о произволе социолога в использовании терминов, описывающих такие предметные области, по меньшей мере, странно.

Для нас терминологический произвол является не столько оценочным высказыванием, сколько наблюдаемым фактом, который зафиксирован не только нами. Например, С.В. Чесноков в автобиографической публикации [13] красноречиво свидетельствует о собственных наблюдениях такого рода. Описываемый произвол отчетливо проявляется как раз в том случае, когда возникает необходимость выходить за границы предметных областей, чтобы, например, согласовать результаты исследований, полученные в каждой из них. В этом случае обнаруживается, что одни и те же по написанию термины в разных исследованиях имеют разные смыслы, то есть связаны с различными ассоциациями. Находясь в границах своей предметной области, исследователь волен игнорировать такого рода разночтения. Однако как только речь заходит о разработке компьютерного ассистирования не для отдельной области, а для социологического исследования как специфического способа описания действительности, проблема терминологических расхождений негативно влияет на успех разработки. Сосуществование различных предметных терминологий в рамках социологического исследования оказывается основной трудностью при разработке средств его компьютерного ассистирования.

Наш подход строится как конструктивная реализация идеи произвольности социологических понятий. В этих условиях естественным выглядит стремление выражать в компьютерных моделях не столько конструктивные термины предметных областей, сколько механизмы авторского произвола, приводящие к указанному сосуществованию.

Это проявляется в том, что мы предлагаем средства, которые моделируют механизм произвола, но не реализуют его. Само понятие произвола предполагает наличие субъекта, без которого термин теряет смысл. В связи с этим терминологический произвол социологического исследования не может существовать без своего автора — социолога.

На традиционный вопрос, который мы неоднократно слышали от коллег — практикующих социологов — теперь можно дать новый, вполне конструктивный ответ. Смысл этого вопроса понятен, например, из следующей формулировки: необходимо измерить уровень бедности, как это сделать с помощью компьютерного ассистирования? Наш ответ таков: мы принципиально не можем указать таких средств, так как подобное указание будет прямым нарушением авторского права социолога на развитие собственных представлений, в данном случае, о бедности. Взамен мы предлагаем программные средства, более адекватные природе социологического исследования, чем навязывание той или иной формальной схемы, закодированной, например, в первопонятиях признакового пространства, эмпирического множества, метода агрегирования данных и т. п. Эти средства, во-первых, позволяют определять представления, например, о бедности, самому социологу в рамках его текущего исследования; во-вторых, внедрить в это определение, в случае необходимости, понятия, которые развиты в предметных областях, не относящихся к области выполняемого исследовательского проекта.

Предлагаемые средства реализованы в виде программных конструкций среды объектно-ориентированного программирования. В силу этой реализации наши модели сами могут быть усовершенствованы другими авторами не в результате дискуссий, а путем видоизменения программного кода, то есть абсолютно конструктивно.

Отметим два других проявления конструктивности предлагаемых решений. Во-первых, наша претензия на разработку понятийного аппарата, воссоздающего механизм авторского произвола в конструировании представлений социологического исследования, реализуется как программная модель ментального объекта, который мы назвали поименованием. Эта модель описана в терминах объектно-ориентированного программирования и представляет собой набор согласованных классов, в том числе класса поименований. В составе этих классов, названных нами базовыми, конструктивно определены свойства и операции, которые передают наш подход. Важным частным случаем разработки представлений о поименовании как конструктивном способе объединения в социологическом исследовании понятий различных предметных областей является сопряжение терминологий, с одной стороны, социологических изысканий (анкетирование, когнитивные методы и т. п.) и, с другой — собственно компьютерного ассистирования (подключение алгоритмов, определение

данных, построение индексов и др.). Во-вторых, структурирование представления о данных на основе понятия поименования в случае программной реализации приводит к оригинальной машине данных. Предлагаемое структурирование служит совершенствованию программных средств различных видов, в том числе: СУБД, статистических пакетов, пакетов АКД, систем КАИ и др. В нашем подходе машина данных реализована в составе базовых классов.

### **Заключение**

В основе работы социолога лежит аппарат поименований, который не является естественнонаучным инструментом описания реальности, но адекватен современному состоянию методологии социальных наук. Предложенный в статье принцип разделительной трактовки данных позволяет описать поименование как специфический когнитивный инструмент, открывающий новые возможности при разработке архитектуры компьютерного ассистирования, предназначенного для социологических приложений.

Поименование как специфически социологический способ моделирования предполагает развитие соответствующей теории. В качестве конструктивного способа такого развития мы видим программную реализацию предлагаемых взглядов. В таком случае вербальные конструкции, оформляемые в соответствии с традициями в виде научных текстов, оказываются вторичными документами, имеющими однозначно понимаемую первооснову.

Вместе с тем, прикладное программирование для любой предметной области может стать эффективным только при выполнении двух условий: во-первых, учета в своей архитектуре специфики самой предметной области; во-вторых, возможности использования наработок из других предметных областей. Развитие представлений о поименовании имеет целью выполнение обоих этих условий. Такое развитие мы предполагаем осуществить путем создания на основе концепции поименования специального языка программирования, с помощью которого возможна разработка как социологической предметной области (социально-культурные изменения, общественные движения, социальное неравенство и др.), так и ее вычислительной поддержки (специализированные СУБД, КАИ, АКД и т. п.).

### **Термины или группы синонимов**

метка; изображение; однозначно идентифицируемая конструкция,  $F_s^{(i)}$ ,  
наблюдаемая составляющая понятия;  
ассоциация; образ; ненаблюдаемая составляющая понятия; когнитивное  
знание;  
когнитивный; ментальный;  
ментальные операции; когнитивные действия: генерация метки, соотнесение метки и ассоциации;

однозначное понятие; финальный термин;  $F^{(0)}$ ;  
 поименование; соединение метки и ассоциации;  $F_s^{(i)}$ ; ментальный объект;  
 определение понятия; процедура определения; процесс определения;  
 совокупность поименований;  
 актор; носитель когнитивных ассоциаций поименования;  
 реципиент поименования; актор, но не автор;  
 ареал поименования; совокупность акторов поименования; ареал состо-  
 ит из автора и реципиентов;  
 автор, один из акторов; актор, предложивший метку.

### Аббревиатуры

КАИ — компьютерное ассистирование интервьюированию

АКД — анализ качественных данных

СУБД — система управления базами данных

ООП — объектно-ориентированное программирование

SPSS — Statistical package for social sciences

### ЛИТЕРАТУРА

1. <http://caqdas.soc.surrey.ac.uk/> (по состоянию на 11.11.03).
2. NVivo (1999): Demonstration software for qualitative data analysis. Version 1.1. 1999. September (CD-ROM).
3. <http://www.atlasti.de/intro.shtml> (по состоянию на 11.11.03).
4. <http://www.westat.com/blaise/capabilities.html> (по состоянию на 11.11.03).
5. [http://www.pulsetrain.com/solutions/application/visual\\_qls.htm](http://www.pulsetrain.com/solutions/application/visual_qls.htm) (по состоянию на 11.11.03).
6. <http://www.scolari.co.uk/> (по состоянию на 11.11.03).
7. *Sudman S., Bradburn N.M.* Asking questions: A practical guide to questionnaire design. San Francisco: Jossey-Bass Publisher, 1992.
8. *Рогозин Д.М.* Когнитивный анализ опросного инструмента. М.: Институт Фонда «Общественное мнение», 2002.
9. *Каныгин Г.В.* Базовые структуры анкетного метода // Социологический журнал. 1999. № 1/2. С. 143–155.
10. *DeMaio T.J., Rothgeb J.M.* Cognitive interviewing techniques: In the lab and in the field // Answering questions: Methodology for determining cognitive and communicative processes in survey research / Ed. by N. Schwarz, S. Sudman. San Francisco: Jossey-Bass Publisher, 1996. P. 177-196.
11. *Bassili J.N.* The how and why of response latency measurement in telephone surveys // Answering questions: Methodology for determining cognitive and communicative processes in survey research / Ed. by N. Schwarz, S. Sudman. San Francisco: Jossey-Bass Publisher, 1996. P. 319-346.
12. *Tourangeau R.* Cognitive sciences and survey methods // Cognitive aspects of survey methodology: Building a bridge between disciplines / Ed. by T.B. Jabine, M.L. Straf, J.M. Tanur, et al. Washington, DC: National Academy of Science Press, 1984.
13. *Чесноков С.В.* «Мне интересен человек как человек...» // Социологический журнал. 2001. № 2. С. 63–122.