

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ
ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 10

Москва 2012

ИНФОРМАЦИОННЫЙ АНАЛИЗ

УДК 001.102 – 047.44

С.Г. Климова, М.А. Михеенкова

Формальные средства ситуационного анализа: опыт применения*

Описаны проблемы представления социологических данных и знаний для ситуационного анализа трудовых отношений. Предлагается последовательное изложение технологии поиска эмпирических закономерностей средствами ДСМ-метода автоматического порождения гипотез на материалах исследования, проведенного авторским коллективом на двух предприятиях. Делаются выводы о возможностях и границах применимости метода.

Ключевые слова: интеллектуальный анализ данных, формализованный качественный анализ социологических данных, ДСМ-метод автоматического порождения гипотез, трудовые отношения, трудовые права, анализ зависимостей, прогноз поведения, ситуационный анализ

ВВЕДЕНИЕ

Возможности интеллектуального анализа данных в области обнаружения новых знаний и извлечения за-

кономерностей особенно востребованы в областях с обширными массивами исходных фактов, нуждающихся в структурировании, упорядочении и систематизации. Это отвечает положению в современных эмпирических социологических исследованиях, во многом ориентированных на получение не количественных (качественных) данных, которые соответствуют номинальному или порядковому уровню измерений. Развитие формальных инструментов преоб-

* Исследование ведется в рамках проекта 10-06-00033 «Методология и методы ситуационного анализа на предприятии», финансируемого РФФИ. Руководитель проекта – С. Г. Климова. Участники проекта: Н.В. Авдошина, В.Ю. Бочаров, И.А. Климов, М.А. Михеенкова, Б.Г. Тукумцев, В.К. Финн, В.А. Ядов.

разования таких данных в знания означает интеллектуализацию соответствующих исследовательских эвристик (и, соответственно, адекватное представление данных и знаний) с последующей реализацией в современных компьютерных системах.

Возникновение «понимающей» социологии М. Вебера, «...стремящейся, истолковывая, понять социальное действие и тем самым каузально объяснить его процесс и воздействие» [1, с. 453], привело к появлению и развитию методов исследования, концентрирующихся на социальных взаимодействиях, личностном повседневном опыте действующего субъекта. Характерной особенностью таких методов, получивших название «качественных», стал отказ от использования формальных средств – в противоположность «количественным» методам, использующим статистические методы для анализа числовых данных, характеризующих обезличенного субъекта. В этом источник фундаментальной слабости качественного анализа – прямой зависимости обобщений и выводов от способностей, здравого смысла и воображения исследователя и непрозрачности (уникальности и невоспроизводимости) исследовательских технологий. В этих условиях статус социолога как беспристрастного исследователя, продуцирующего новое знание, оказывается под вопросом.

Казалось бы, широкая доступность компьютерной техники неизбежно должна была привести к объективизации качественных исследований. Но развитие пакетов CAQDAS (Computer-Assisted Qualitative Data Analysis Software) изначально пошло по пути усовершенствования обработки, структурирования и управления данными. И хотя представители современной генерации таких инструментов выглядят достаточно впечатляющими [2], проблемы нестатистического анализа социологических данных, в основном, остаются за рамками их возможностей.

Большие ожидания в социологии связывались с развитием «кибернетических» методов data mining (DM), в частности, искусственных нейронных сетей, эволюционного программирования, поиском ассоциативных правил, деревьев решений и, наконец, систем обработки экспертных знаний. Теперь уже статистические пакеты, являющиеся основным инструментом обработки социологических данных, в качестве обязательных составляющих содержат приложения DM (SPSS Clementine 9.0, STATISTICA Data Mining и т.п.). Эти средства позволяют одновременно анализировать неоднородные и неполные данные («нерепрезентативные» с точки зрения статистики), причём, нечислового характера, учитывают нелинейные связи¹. Однако, как убедительно показано в известной работе [4], DM – применение конкретных алгоритмов для извлечения моделей (образцов) – составляет лишь один из шагов knowledge discovery (KD) – процесса извлечения полезных знаний из данных. Для полноты KD – интеллектуального анализа данных – необходимы также отбор и представление

данных и знаний, объяснение и интерпретация результатов анализа.

Таким образом, собственно интеллектуальным анализ данных оказывается при использовании компьютерных систем, содержащих средства извлечения знаний из баз фактов (БФ), автоматического порождения гипотез и объяснения имеющихся фактов на основании порожденных гипотез, а также способных осуществлять дедуктивный вывод из исходных и полученных знаний (баз знаний – БЗ) [5]. Современная социологическая наука, задача которой – не просто восприятие и описание непосредственно эмпирических явлений, но и объяснение каузальных связей в этих явлениях, крайне заинтересована в использовании таких инструментов. Вот почему формализация аналитических процедур, прежде всего – процедур выявления причинных зависимостей из анализа имеющихся данных, индуктивных стратегий создания теорий и связанных с этим когнитивных процессов абдукции, – рассматривается действующими социологами как подлинный вызов времени [6, 7].

ФОРМАЛЬНОЕ ПРЕДСТАВЛЕНИЕ ДАННЫХ

С точки зрения эпистемологического содержания качественный анализ может рассматриваться как «восходящая» стратегия «доказательного и последовательного» построения теории на основе эмпирических фактов, т.е. получения нового знания из эмпирического материала. Напомним, что подобным образом (с использованием неформализованного индуктивного вывода) формулируется основная задача в одной из наиболее структурированных и разработанных методологий качественного анализа – обоснованной теории (grounded theory, см. [8]). Стало быть, интеллектуальный анализ социологических данных (ИАСД) – построение теории на основе эмпирических фактов средствами формализованных познавательных процедур в компьютерных системах высокого уровня – есть не что иное, как формализованный качественный анализ социологических данных (ФКАСД). Такой анализ предполагает более высокий, чем это принято в качественном анализе, уровень формализации изучаемых психосоциальных явлений и нуждается в первичном структурировании данных и знаний и формировании системы отношений для них. Следующим шагом оказывается выбор адекватных предложенному представлению формальных средств анализа. В совокупности это означает создание формального языка с дескриптивной и аргументативной функциями [9] для представления данных и знаний (с возможностью определения их сходства) и формализации рассуждений (и выдвижения гипотез), соответственно².

Фундаментальным принципом качественного анализа данных является принцип: «сходство фактов влечет наличие (отсутствие) изучаемого эффекта и его повторяемость». Конкретизируя его для реализации идеи причинности, можно сказать, что в рамках

¹ К сожалению, развитие арсенала современных методов анализа социологических данных в нашей стране пока заметно отстаёт от общемировых тенденций [3]

² Очевидно, что главной слабостью как пакетов CAQDAS, так и используемых абстрактных методов DM, является не просто отсутствие такого языка, но даже представления о необходимости его создания.

качественного анализа исследуется тип каузальности «структура – эффект» (а не «явление – явление»). Гипотетические причины представляются в виде сходств фактов, имеющих определенную структуру, т.е. «сходство» в этой схеме является нестатистическим и может анализироваться в рамках логико-алгебраического подхода (на чём основан качественный сравнительный анализ – Qualitative Comparative Analysis, QCA [10]) и средств интеллектуального анализа данных, включающих формализацию индуктивных процедур (которые реализованы в ДСМ-методе автоматического порождения гипотез [11])³.

ДСМ-метод автоматического порождения гипотез (ДСМ-метод или ДСМ-метод АПГ) позволяет реализовать сформулированную в [5] программу интеллектуального анализа применительно к социологическим данным. Интеллектуальный анализ социологических данных (ИАСД) и реализующая его интеллектуальная система (в частности, интеллектуальная система типа ДСМ, ИС-ДСМ) предполагают возможность решения ряда проблем, связанных с представлениями классической «понимающей социологии» М. Вебера и составляющих круг задач ФКАСД. К ним относятся исследование индивидуального поведения, порождение детерминант поведения и типологизация социума на их основе, анализ и прогнозирование мнений респондентов как варианта поведения; существенным при решении этих задач является выяснение влияния ситуации на поведение индивидуума.

Из описанного принципа структурного сходства как источника детерминаций вытекает потребность первичной (до решения собственно задач ФКАСД) структуризации данных и знаний, т.е. предварительной алгебраической формализации сходства объектов и их свойств. Основной принцип представления знаний в ИС-ДСМ, предназначенных для анализа социологических данных и возможного прогнозирования социальных действий или мнений (решения задач ФКАСД), – «постулат поведения» Р1. Согласно этому постулату, используемые параметры описания индивида выражают его социальный характер, черты личности и нетривиальные биографические данные, отражающие историю становления личности (или, по крайней мере, адаптацию в социальной среде). Кроме того, в соответствии с исследованиями в социальной и когнитивной психологии, детерминация социального поведения субъекта напрямую связана также и со средовыми факторами [14]. Соответственно, «постулат ситуационизма» Р3 предполагает включение в анализ ситуационных параметров, позволяющих учесть контекст реализации поведения. Наконец, анализ рационального поведения предполагает возможность формализации мнений субъекта (в том числе, с учётом его социальных установок). Логическая систематизация знаний о субъекте, его поведении и мнениях означает выделение исходных отношений и установление зависимостей между ними на основе формализованных рассуждений.

Специальный класс таких рассуждений – ДСМ-рассуждения – представлен в ДСМ-методе. Метод реализует синтез познавательных процедур – эмпирической *индукции* (формальных расширений и уточнений индуктивных методов английского философа и логика Д.С. Милля, в честь которого и назван метод), структурной аналогии (перенос гипотез о причинах на случаи с неопределённым исходом) и *абдуктивного рассуждения* Ч.С. Пирса как средства принятия гипотез на основе объяснения начальных данных. Синтез этих познавательных процедур является операциональным определением отношения «причина – следствие» (а не просто условного высказывания «если p , то q ») относительно данных и знаний. Средством формализации ДСМ-рассуждений, реализующих синтез познавательных процедур, являются бесконечнозначные логики степеней правдоподобия порождаемых гипотез (1-го порядка для конечных моделей и слабой логики предикатов 2-го порядка для бесконечных моделей [15]).

Исследование каузальности типа «структура объекта – эффект», составляющее основу ДСМ-метода, укладывается в рамки перечисленных выше задач ФКАСД, направленных на изучение поведения Y субъекта X , находящегося в ситуации S и обладающего мнением φ (отображающим субъективный мир личности – см. [5]). Для представления знаний и формализации ДСМ-рассуждений формулируется ДСМ-язык [11, с. 10-48], в котором для представления субъектов поведения используются индивидуальные переменные $X, Z, V, \dots$ 1-го сорта (быть может, с нижними индексами) и константы $C, C_1, C_2, \dots$, являющиеся значениями переменных для объектов и подобъектов X, Z, V и т.д. Свойства объектов (например, эффекты поведения субъектов) представляются индивидуальными переменными 2-го сорта $Y, U, W, \dots$ (быть может, с нижними индексами) и константами $Q, Q_1, Q_2, \dots$, а также A, A_1, A_2 .

Пусть даны конечные множества $U^{(i)}, i = 1, 2, 3$, $U^{(1)} = \{d_1, \dots, d_{n_1}\}$ – множество характеристик социальных субъектов, $U^{(2)} = \{a_1, \dots, a_{n_2}\}$ – множество эффектов их поведения (действий и установок). Определим на них две булевы алгебры $\mathcal{R}_i = \{2^{U^{(i)}}, \emptyset, U^{(i)}, \text{—}, \cap, \cup\}$, $i = 1, 2$. $\mathcal{R}_1$ – алгебра объектов, $\mathcal{R}_2$ – алгебра свойств. Тогда объекты (субъекты поведения) $X \in 2^{U^{(1)}}$, множества свойств (эффекты поведения) $Y \in 2^{U^{(2)}}$.

Атомарные формулы ДСМ-языка применительно к задачам ФКАСД интерпретируются следующим образом. В исходном состоянии база фактов (БФ) содержит (+)- и (-)-факты наличия и отсутствия изучаемого эффекта, соответственно, а также неопределённые (τ)-факты, представляющие предикат $X \Rightarrow_1 Y$. Предикат означает, что «субъект X обладает/не обладает эффектом поведения Y », где X – структурированное описание субъекта, Y – переменная для представления действий и установок (в общем случае – и мнений). Производные предикаты $V \Rightarrow_2 W$ и $W \Leftarrow_3 V$ означают, что «подмножество характеристик V есть причина эффекта поведения W » и «эффект поведения W есть следствие подмножества характеристик V » (т.е. переменная V принимает значения на множестве возможных характеристик субъ-

³ Теоретическое сравнение этих подходов проведено в [12]; опыт практического их применения в одном исследовании изложен в [13].

ектов данного социума). $V \Rightarrow_2 W$ и $W_3 \Leftarrow V$ представляют собой гипотезы о причинах наличия/отсутствия изучаемых эффектов, порожденные на основании индуктивного анализа предиката $X \Rightarrow_1 Y$. Расширение и уточнение представленного в начальном состоянии БФ отношения $\Rightarrow_1^*$ (которому соответствует предикат $\Rightarrow_1$) – предсказание поведения субъектов – осуществляется с помощью правил вывода по аналогии, которые используют порождённые гипотезы $V \Rightarrow_2 W$ и $W_3 \Leftarrow V$. Цикл «индукция – аналогия» повторяется до стабилизации множества гипотез. ДСМ-рассуждение завершается применением абдукции (формализованной посредством критерия достаточного основания принятия гипотез) – процедуры объяснения исходного множества фактов. При необходимости (если порожденных гипотез недостаточно для объяснения) БФ может быть пополнена новыми примерами, и ДСМ-рассуждение выполняется заново.

Такое представление БФ является базовым для ДСМ-метода и допускает различные модификации с учётом потребностей предметной области. Так, введение в рассмотрение ситуационных параметров из множества $U^{(3)} = \{s_1, \dots, s_n\}$ позволяет включить в анализ поведения влияние средовых факторов. Тогда $\mathcal{B}_3 = \{2^{U^{(3)}}, \emptyset, U^{(3)}, -, \cap, \cup\}$ – алгебра ситуаций (внешних обстоятельств), параметры ситуации представлены переменными 3-го сорта $S, S_1, \dots, S_n, \dots$ и константами $\bar{S}, \bar{S}_1, \dots, \bar{S}_n, \dots$ ($S, S_1, \dots, S_n, \dots \in 2^{U^{(3)}}$).

Расширим ДСМ-язык введением термина $\bar{X}$, где $\bar{X}$ – полный объект $\bar{X} = \langle X, S, [\varphi] \rangle$, X – субъект поведения, S – контекст (ситуация) поведения, φ – мнение субъекта (о ситуации, о возможном поведении)⁴. Отделение мнения от действий и установок обусловлено специфическим характером этого вида поведения, для которого в [17] предложено формальное представление, опирающееся на аргументированное принятие темы опроса T^* посредством раскрывающих её содержание утверждений $p_1, \dots, p_n$, каркаса P темы T^* . Множество ответов на вопросы $?p_i$ – «какова оценка v утверждения p_i ?» – представляет мнение $[\varphi]$ респондента. В стандартном варианте ДСМ-метода используются 4 типа истинностных значений v для представления фактов: 1 – «фактическая истина», -1 – «фактическая ложь», 0 – «фактическое противоречие», τ – «фактическая неопределённость». В общем случае m -значного ($m \geq 2$) закрытого опроса респонденту предлагается m вариантов ответа (см., например, [12]).

Общая задача изучения явления, понимаемого как отношение «объект – эффект», соответствует двум классам задач, которым отвечают два типа ДСМ-рассуждений – прямой и обратный [17, 18].

(1) Пусть в исходном состоянии БФ представлена предикатами $\bar{X} \Rightarrow_1 Y$ ($\langle X, S, [\varphi] \rangle \Rightarrow_1 Y$), где Y – эффект (действие или установка к действию). Тогда применение прямого ДСМ-рассуждения («от причины – к следствию») порождает гипотезы о причинах вида $\bar{V} \Rightarrow_2 Y$, где $\bar{V}$ представляет сходство объектов $\bar{X}$, $\bar{V} = \bar{X}_1 \cap \dots \cap \bar{X}_k$, $\bar{V} = \langle V, S', [\psi] \rangle$, где $V = X_1 \cap \dots \cap X_k$, $S' \subseteq S_1 \cap \dots \cap S_k$, $[\psi] \subseteq [\varphi_1] \cap \dots \cap [\varphi_k]$.

(2) База фактов для анализа такого вида поведения, как мнение, представлена предикатами $\langle X, S \rangle \Rightarrow_1 [\varphi]$, T , где T – одна из возможных тем опроса (в случае проведения опроса по одной теме представление упрощается до $\langle X, S \rangle \Rightarrow_1 [\varphi]$). В этом случае применяется обратный метод и порождаются гипотезы вида $W_3 \Leftarrow \langle V, S' \rangle$, $W/[\psi]$, $V = X_1 \cap \dots \cap X_k$, $S' \subseteq S_1 \cap \dots \cap S_k$, $[\psi] = [\varphi_1] \cap \dots \cap [\varphi_k]$. Решение задачи выявления детерминаций мнений с учётом ситуации проведения опроса подробно описано в [19].

В n ситуациях r полных объектов могут быть представлены таблицей:

	$\bar{X}_1$		$\bar{X}_j$		$\bar{X}_r$
S_1	$\langle X_1, S_1, [\varphi_1^{(1)}] \rangle$	...	$\langle X_j, S_1, [\varphi_j^{(1)}] \rangle$	...	$\langle X_r, S_1, [\varphi_r^{(1)}] \rangle$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
S_i	$\langle X_1, S_i, [\varphi_1^{(i)}] \rangle$	...	$\langle X_j, S_i, [\varphi_j^{(i)}] \rangle$	...	$\langle X_r, S_i, [\varphi_r^{(i)}] \rangle$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
S_n	$\langle X_1, S_n, [\varphi_1^{(n)}] \rangle$	...	$\langle X_j, S_n, [\varphi_j^{(n)}] \rangle$	...	$\langle X_r, S_n, [\varphi_r^{(n)}] \rangle$

Представленному в таблице соответствию $\langle X_j, S_i \rangle \Rightarrow_1 [\varphi_j^{(i)}]$ отвечает описанный в [20] предикат $P(X, S, Y)$ ($X_j/X, S_i/S, [\varphi_j^{(i)}]/Y$) – «субъект X в ситуации S демонстрирует эффекты поведения (мнения) Y ». Предикат причинности принимает вид $R(\langle V, S' \rangle, W)$ – «пара $\langle$ подмножество характеристик субъекта V и характеристики ситуации $S'\rangle$ есть причина множества эффектов поведения W ». В случае анализа мнений в различных ситуациях для индуктивного порождения причин $\bar{R}^s(\langle V, S' \rangle, W)$ или $\bar{R}^p(\langle W, S' \rangle, V)$ используются предикаты ${}_2\bar{M}_{a,n}^\sigma(V, W, S')$ или ${}_2\bar{M}_{a,n}^\sigma(V, W, S')$ (подробно описанные в [19]), соответственно (или их усиления; $\sigma = +, -$), входящие в посылки индуктивных правил правдоподобного вывода I-го рода. Основой эмпирических зависимостей в этих предикатах является сходство не только субъектов поведения и их мнений, но и ситуации проявления этих мнений. В дальнейшем полученные гипотезы о причинах используются для предсказания возможного поведения и/или мнений с помощью вывода по аналогии – правил правдоподобного вывода II-го рода, а также для абдуктивного объяснения исходной БФ с помощью аксиом каузальной полноты.

Рассмотрим, каковы возможности и результаты анализа средствами ДСМ-метода представленной подобным образом БФ на примере решения конкретной социологической задачи.

ПОСТАНОВКА ИССЛЕДОВАТЕЛЬСКОЙ ЗАДАЧИ

Адекватность использования того или иного метода определяется особенностями изучаемого объекта и спецификой исследовательской задачи. Наш объект – современные трудовые отношения на российских предприятиях – весьма сложен для социологического анализа.

⁴ О введении кортежных термов см. [16].

Во-первых, из-за того, что они в значительной мере определяются внешними обстоятельствами. Например: местоположение предприятия, ситуация на рынках сбыта, институциональные условия (действующие правила и практики). В них прямо или опосредованно включены многие субъекты: собственники, менеджмент, власти разных уровней, конкуренты, наёмные работники (непосредственно либо опосредованно, в лице своих представительных органов). Это значит, что ограничение исследовательских задач, с одной стороны, практически неизбежно, а с другой – такое ограничение ведет (осознанно либо неосознанно) к неполной адекватности исходных концептуальных построений и результатов. Так, объяснение изучаемых процессов установками участников, «менталитетом» может определяться тем, что эти установки выявить проще, чем другие компоненты проблемы, ставшей стимулом исследования.

Во-вторых, многие предприятия сейчас переживают период резких организационных, экономических, технологических перемен, которые требуют оперативных данных о ситуации. Необходимость оперативного реагирования приводит социологов к подчас вынужденному отказу от создания моделей изучаемых процессов для их интерпретации и прогноза развития в пользу «широкого нарратива» – регистрации реальных событий, описания конфликтов, кризисов, или, напротив, каких-либо свершений. Такая исследовательская стратегия вольно или невольно ведет к вынужденному отказу от обобщений, от выводов о закономерностях в пользу «описания случая». Тем самым социологи предоставляют читателю самому судить о характере связей между описываемыми феноменами и их устойчивости, что лишает исследование научной объективности и полноценности.

В-третьих, на современных предприятиях весьма затруднительно использовать методы, основанные на статистическом анализе данных из-за, как правило, малых объемов выборки и, соответственно, невозможности анализировать характеристики групп, выделенных по тому или иному признаку (квалификация, возраст, подразделение, и пр.).

В-четвертых. Хотя имеющиеся в распоряжении социологов методы многомерного анализа данных позволяют выделять сочетания признаков, определяющие возможные значения зависимых характеристик, учёт контекста рассматриваемого явления остаётся за рамками их возможностей. Вместе с тем, известно, что значения признаков всегда контекстны. Самый простой пример – оценка «достаточности» заработка в зависимости от места проживания и состава семьи. Зачастую именно влияние контекста определяет множественность причин проявления рассматриваемых эффектов, однако эта множественность не выявляется применением указанных методов.

Эти обстоятельства обусловили выбор исследовательским коллективом ситуационного анализа в качестве концепции исследования социальных проблем предприятий [21]. В качестве формального инструмента ситуационного анализа был выбран ДСМ-метод автоматического порождения гипотез,

возможности которого, как уже говорилось выше, позволяют рассматривать его как метод формализованного качественного анализа социологических данных (ФКАСД).

НЕФОРМАЛЬНОЕ ОПИСАНИЕ И ОБЩАЯ ИДЕОЛОГИЯ МЕТОДА

ДСМ-метод – логико-комбинаторный (нестатистический) метод, использующий средства математической логики и предназначенный для анализа *зависимостей* между сочетанием признаков и искомым эффектом. Искомым эффектом в социологии может считаться категоризация и описание сообщества по заданным критериям. Это могут быть признаки, имеющие объективированное содержание (например, квалификация или состояние здоровья); какое-то поведение или поведенческая установка; мнения или оценки людей; предположения относительно возможного развития ситуации (поведения участников). Установление зависимости – это поиск ответа на вопрос: «Какие *сочетания признаков* порождают появление искомого эффекта, и при каких условиях?». Например, семья может считаться бедной не потому, что в ней много иждивенцев, или невысок заработок работающих, или высоки цены на местных продовольственных рынках, или высоки расходы на лечение и образование детей, а потому, что имеют место одновременно все перечисленные обстоятельства (либо их различные комбинации).

Такую зависимость мы называем причинной обусловленностью тогда, когда обнаруживается отношение, при котором сходство респондентов по некоторому сочетанию признаков порождает повторяющийся эффект и при этом отсутствуют контрпримеры (альтернативное поведение или поведенческая готовность при том же сочетании детерминирующих признаков). Это условие означает, что какое-то сочетание признаков считается влияющим на искомое поведение (в нашем случае – готовность отстаивать трудовые права) только тогда, когда отсутствует то же самое сочетание признаков, описывающих респондентов с альтернативным поведением (не будут отстаивать трудовые права). В ДСМ-методе такое требование обеспечивается самой структурой правил правдоподобного вывода, в котором решающие предикаты разных знаков (например, ${}^s\tilde{M}_{a,n}^+(V,W,S')$ и ${}^s\tilde{M}_{a,n}^-(V,W,S')$ или ${}^p\tilde{M}_{a,n}^+(V,W,S')$ и ${}^p\tilde{M}_{a,n}^-(V,W,S')$) являются взаимно фальсифицирующими. Однако заранее неизвестно, во-первых, каково число таких сочетаний, а во-вторых, какие именно характеристики каждого из трех блоков (респондент, его установки, ситуация) войдут в то или иное детерминирующее сочетание.

Сразу следует сказать, что причинная обусловленность в данном контексте – это не причина-универсальный закон, а причина-объяснение события, произошедшего в данной ситуации. Причинную обусловленность мы понимаем как предрасположенность к неким действиям, которая реализуется при отсутствии противодействующих влияний (как внутренних – личностных, так и внешних – ситуационных).

Содержательно задача формулировалась как выяснение вопросов: кто, какие работники и в какой ситуации готовы отстаивать свои трудовые права и кто, в какой ситуации решает отказаться от их защиты. Эмпирической базой исследования стали результаты опросов работников двух предприятий г. Самары. В число детерминирующих признаков вошли характеристики работников двух предприятий по объективным параметрам; их мнения и оценки работы в разных аспектах; некоторые параметры ситуации на рынке труда г. Самары. В качестве детерминируемого признака взят утвердительный либо отрицательный ответ на вопрос: *«Если в будущем Ваши трудовые права окажутся нарушенными, Вы будете или не будете добиваться справедливости?»*.

Исследовательская проблема, для решения которой были использованы возможности ДСМ-метода, может быть представлена следующим образом. Имеется множество разнородных признаков, бесспорно влияющих, как показывает наш исследовательский опыт, на искомое поведение (готовность отстаивать трудовые права), причем влияющих по-разному в зависимости от их сочетания. Например, уже на этапе выбора объекта было ясно, что признак «зарплата» для установки на готовность защищать свои трудовые права значим не сам по себе, а в сочетании с другими признаками, такими как «доход», «удовлетворенность социальным пакетом», «возможность найти другую работу» и пр. Помимо содержательных соображений, неприменимость традиционных статистических процедур обусловлена практическими трудностями. Предприятия, на которых проводилось исследование, были недостаточно велики для того, чтобы можно было опросить необходимое для традиционного статистического анализа число работников, и тем более для того, чтобы сформировать частные группировки по подразделениям или другим значимым для исследования признакам.

Отказ же от «социологического нарратива» – при том, что решение поставленной задачи может рассматриваться как традиционное для качественного анализа построение теории на основе эмпирических фактов – объясняется стремлением к объективизации процесса построения такой теории. ДСМ-метод, являющийся инструментом для анализа *нечисловых* отношений и установления причин эффектов (готовности к поведению определенного типа) на основе этого анализа и располагающий для этого формальными средствами, как нельзя лучше отвечает потребности такой объективизации – автоматического получения нового знания из эмпирического материала.

Возможность применения этого метода в социологических исследованиях опирается на представление о социальных системах как системах с детерминированными (когда факты являются зависимыми и не образуют системы случайных событий) многофакторными разнонаправленными ($\pm$)-влияниями (при этом возможные случайные возмущения не включаются в анализ)⁵. Наиболее существенные и устойчи-

вые влияния рассматриваются как причины тех или иных социальных явлений. Такое представление о «причине» вызывается скрытым комплексным характером причинно-следственных отношений в рассматриваемой области.

Особенности применения ДСМ-метода автоматического порождения гипотез как формального аппарата для качественного анализа в социологии описаны в ряде работ авторов (см., например, [22]). Успешное использование метода – получение полезных результатов – требует его адаптации для социологических задач. В частности, необходимы компетентные решения социолога-исследователя, основанные на предварительных обоснованных допущениях об обстоятельствах (признаках), которые могут находиться в причинной связи с изучаемым явлением (эффектом).

Напомним, что основу правил индуктивного вывода в ДСМ-методе составляют формальные уточнения и расширения индуктивных «методов опытного исследования» Д.С. Милля, идеи которого имеют долгую историю применения в «нравственных» науках – науках о человеке и обществе. Для современных исследователей также характерно стремление к использованию потенциала этих идей в социологии. Так, Н. Смелзер высказывает убеждение, что методы Милля обеспечивают общее направление сравнительных исследований, индуктивных по своей природе, а конкретно методы сходства и различия составляют ядро экспериментальных исследований [23]. О возможности/необходимости использования индукции Милля в сравнительных исследованиях говорят также отечественные специалисты [24, с. 349, 367], причём индукции иногда отводится роль инструмента для выявления необходимой причинной связи (в противопоставлении возможной статистической) [25, с. 149].

Важную роль идей Милля в качественных социологических исследованиях признают авторы широко распространённого на Западе метода, предназначенного для работы с социологическими данными – упомянутого выше «качественного сравнительного анализа» (QCA). QCA опирается на конъюнктивно-дизъюнктивное представление о причинах явлений: существуют комбинации причин, приводящие к изучаемому эффекту; различные комбинации могут приводить к одному результату; действие комбинаций причин контекстно-зависимо. В результате порождаются различные каузальные модели, релевантные существующему эмпирическому материалу.

Слабостью метода QCA сами авторы считают нерешенность проблемы внутренней и внешней валидности (т.е. грамотного выделения факторов, определяющих вариации интересующего исследователя признака, и возможности экстраполяции результатов за рамки конкретного опыта). Существуют подходы к формированию исходной выборки для использования QCA. Так, для максимизации внутренней валидности предлагается формировать исходную выборку из как можно более близких друг другу примеров (например, имеющих наибольшее количество общих характеристик), которая разделяется на две в зависимости от наличия или отсутствия рассматриваемого эффек-

⁵ «(+)-влияния» – те, которые способствуют возникновению или развитию события, «(-)-влияния» – те, которые этому препятствуют.

та. Далее неформально применяются методы сходства и различия и порождаются гипотезы о причинах. Для внешней валидности отбираются как можно более далёкие друг от друга примеры, на которых тестируются полученные гипотезы.

Современная формализация методов сходства и различия в ДСМ-методе автоматического порождения гипотез была использована нами для анализа материалов исследования на двух предприятиях (2011 г., Самара)⁶ с некоторыми уточнениями. В ситуации реального полевого исследования, а тем более на таком сложном объекте, как производственное предприятие, крайне трудно найти даже два таких предприятия, которые имели бы «максимально сходные» характеристики. Дело в том, что те или иные общие характеристики (например, отраслевая принадлежность или расположенность в одном и том же городе) точно так же, как и различия, не могут исчерпывающим образом объяснять искомый эффект. Эффект всегда объясняется сочетанием причин.

ОСНОВНЫЕ ПРОЦЕДУРНЫЕ РЕШЕНИЯ

Анализ социологических данных с помощью ДСМ-метода требует знания о системе отношений между некоторым набором признаков, описывающих субъект, и поведением этого субъекта. Это – знание о закономерностях поведения исследуемого социума, полученное ранее. Например, считается установленным, что сейчас наиболее сильным стимулом трудовой активности является заработок. Но удовлетворительность заработка определяется многими контекстными обстоятельствами. Значимость некоторых очевидна (размер и состав семьи), а других – нет (с какими профессиональными группами идет сравнение по критерию удовлетворительности заработка).

Ещё одна необходимая предпосылка анализа социологических данных с помощью ДСМ-метода – требование более строгого, чем это принято в традиционном качественном анализе, уровня формализации данных об изучаемых явлениях. Для анализа средствами ДСМ-метода в соответствии с предложенным представлением данных необходимо их структурирование и категоризация на: 1) характеристики респондентов по каким-то объективированным (социально-демографическим) критериям; 2) мнения и оценки; 3) поведенческие установки, 4) параметры ситуации.

В исследованиях *мнений* какого-то сообщества результатом, удовлетворяющим заказчика, считается анализ мнений и оценок по какому-то значимому вопросу. Например, оценка работы руководителя, одобрение или неодобрение каких-то принятых решений. Выводы о мнениях и оценках, в зависимости от задач исследования, могут быть результатом работы социолога, а могут включаться в исследование как детерминирующий фактор. Например, в нашем случае предполагалось, что мнения и оценки по поводу некоторых параметров трудовой ситуации будут определять готовность отстаивать трудовые права.

Когда заказчиков исследований интересуют *поведенческие установки* какого-то сообщества, запрос на исследование формулируется как ожидание выводов о связи характеристик респондента, его установок, мнений, оценок с готовностью к поведению определенного типа. Например, в политической или управленческой практике заказчик задаёт исследователю такие вопросы: «Кто намерен увольняться (бастовать), а кто готов терпеть дальше?»; «Кто примет предполагаемую управленческую новацию, а кто – нет?»; «Кто уедет, а кто – останется?»; и т.п.

Ситуационный анализ в нашем случае предполагает включение признаков, характеризующих *объективную ситуацию* [26]. Это оказалось довольно сложной задачей, потому что изначально нет уверенности, какие признаки окажутся работающими, а какие – нет. Абстрактно говоря, все характеристики поселения (включая характеристики локального рынка труда) значимы для работников. Однако «вес» этой значимости разный в зависимости от конкретного поселения. Например, бесспорно значимы экологические характеристики города в принятии решения о том, чтобы уволиться из-за необходимости переселиться в более экологически благополучное место. Однако в нашем случае (в г. Самаре) исследователи (которых можно считать экспертами) решили, что вопросы экологии для характеристики трудового поведения самарцев не значимы, но значимы параметры заработка и доходов.

Следующий шаг – это анализ множества социологических признаков, полученных в исследовании, для включения в анализ с помощью ДСМ-метода. Они должны быть: а) предположительно значимыми для искомого эффекта; б) независимыми друг от друга⁷; в) предположительно влияющими на поведение людей в определенной ситуации.

Первичная проверка структуризации данных включала также анализ того, насколько симметрично представлены исходные данные – включены ли в БФ примеры наличия исследуемого эффекта и его отсутствия. Это требование, называемое требованием «квазисимметричности» исследуемого мира, является одним из существенных условий применимости ДСМ-метода. Собственно, это требование к шкалам – одно из базовых в социологии. Однако его нарушения (осознанные или нет) встречаются довольно часто. Здесь ДСМ-метод становится инструментом контроля методической строгости. В нашем случае предполагалось, что в изучаемом массиве эмпирических данных отношение «субъект $\Rightarrow$ поведение» представлено позитивными (наличие поведенческой готовности: «буду добиваться справедливости в случае нарушения трудовых прав»), негативными («не буду добиваться справедливости в случае нарушения трудовых прав») и неопределёнными (вариант ответа: «трудно сказать») примерами. Детерминирующие признаки также были проверены на соответствие требованию квазисимметричности и из анализа были исключены те, которые не удовлетворяли или не вполне удовлетворяли ему.

⁶ В разработке инструментария, полевого исследовании и обработке данных участвовали: Б.Г. Тукумцев (социологический институт, С-Пб), В.Ю. Бочаров, Н.В. Авдошина, Ю.В. Васькина (СамГУ).

⁷ Это требование, существенное для многих статистических процедур, не является обязательным для ДСМ-метода.

Мы предполагали, что субъекты, выбирающие ту или иную поведенческую стратегию, будут обладать близкими, похожими признаками (статусные характеристики, мнения, оценки). Кроме того, признаки, описывающие параметры ситуации, существенные для выбора стратегии поведения, также предполагались сходными.

После обработки данных анкетного опроса, на этапе, предваряющем работу с процедурами формализованного качественного анализа социологических данных, обнаружилось, что два обследованных предприятия значительно различаются по оценкам работниками ситуации с трудовыми правами.

Ситуацию с трудовыми правами неблагополучной находят более 4/5 опрошенных работников (83%) на предприятии «А»; на предприятии «Б» – примерно 2/3 (64%). Полагают, что права соблюдаются, 7% работников предприятия «А» и 22% работников предприятия «Б». Оценивая ситуацию с трудовыми правами за последние полгода, сказали, что их трудовые права в этот период нарушались, 63% работников предприятия «А» и лишь 23% работников предприятия «Б». Вместе с тем, на предприятии «А» лишь четверть работников предпринимали те или иные действия, чтобы восстановить справедливость, а на предприятии «Б» – почти половина (48%). При том, что воспринимаемая ситуация с соблюдением трудовых прав различается весьма заметно, почти нет разницы в распределении мнений о готовности в будущем добиваться справедливости, если трудовые права будут нарушены. Заявили о своей готовности «во всех случаях» отстаивать нарушенные права 16% работников предприятия «А» и 19% работников предприятия «Б». «Только в серьезных случаях» будут отстаивать свои права 48% первых и 51% – вторых. Сказали, что откажутся от борьбы за свои права, 18% и 12% работников, соответственно.

Таким образом, мы имеем дело с ситуацией, когда для интерпретации результатов опроса нужна дополнительная информация. В самих данных нет объяснения тому факту, что те работники, чьи права нарушаются чаще, готовы бороться за них реже. Искомый признак (эффект) оказывается обусловленным множеством обстоятельств. Обычно исследователи в таких случаях мобилизуют запасы своей эрудиции и неформализованных данных о предприятии, чтобы объяснить полученный парадокс. В нашем случае можно сказать следующее. На предприятии «А» доминирующая установка работников – «жаловаться бесполезно», потому что руководство объяснило работникам, что завод находится в крайне тяжелом экономическом положении, что заказов нет, что в будущем, возможно, ситуация изменится. Но это объяснение – лишь одна из версий, услышанная в неформальной беседе. А их было много, в том числе и «менталитетных». Для того чтобы избежать произвола в выборе интерпретативной схемы, мы попытались систематизировать объективные данные о ситуации на обоих предприятиях.

Оба предприятия относятся к одной отрасли (машиностроение), но к разным подотраслям; предъявляют достаточно высокие требования к квалификации работников, характерные для машиностроительной отрасли в

целом; имеют сходные квалификационные характеристики персонала и остро нуждаются в квалифицированных работниках (как инженерах, так и рабочих). На обоих предприятиях существует профсоюз «советского» типа; заключен коллективный договор; есть, хотя и скромный, «социальный пакет».

Вместе с тем, сильно различается материальное положение работников занятых на этих двух предприятиях. Месячный душевой доход работников предприятия «А» примерно на две тыс. руб. меньше, чем работников предприятия «Б». Есть разница и в характеристиках персонала. Так, работники со стажем более 10 лет заметно больше на предприятии «А», чем на предприятии «Б» (85% и 57%, соответственно). Мнения и оценки тоже различны по некоторым параметрам. Например, доля тех, кто в целом удовлетворен трудом, на предприятии «А» – 47%; на предприятии «Б» – 67%. Эти данные позволяют предположить, что более пожилым работникам предприятия «А» просто некуда деться, и они мирятся с малыми заработками и их регулярными задержками. Однако данные анкетного опроса и здесь неоднозначны. Тех, кто хотели бы сменить место работы, на предприятии «А» на треть больше, чем на предприятии «Б» (32% и 21% соответственно). Но перспективы альтернативного трудоустройства и те, и другие оценивают примерно одинаково. Сказали, что им трудно было бы найти другую работу с лучшими условиями 55% – первых и 54% – вторых. Итак, мы имеем дело с исследовательской ситуацией, когда нужны дополнительные возможности для интерпретации данных. Такие возможности обеспечивает анализ средствами ДСМ-метода автоматического порождения гипотез.

Чтобы выяснить, какие сочетания характеристик респондентов сопутствуют двум искомым поведенческим признакам: готовности либо неготовности добиваться справедливости в случае нарушения трудовых прав, мы проанализировали данные исследования с тем, чтобы выбрать те из них, которые, по нашим предположениям, будут сопутствовать искомой характеристике (влиять на неё). Как уже отмечалось выше, это были данные трех типов: характеристики ситуации (S); характеристики респондентов (X) и признаки, описывающие мнения и установки опрошенных (элементы каркаса темы P).

В первую группу признаков, характеризующих ситуацию, вошли: s_1 – индекс «соответствие заработной платы работника и среднемесячной заработной платы по предприятиям обрабатывающих производств г. Самары»; s_2 – индекс «соответствие реальной заработной платы работника и среднемесячной заработной платы в соответствующем сегменте производств»; s_3 – индекс «соотношение среднедушевого дохода респондента и личных доходов по г. Самаре в среднем»; s_4 – предприятие (А или Б).

Во вторую группу признаков, описывающих респондента, вошли: d_1 – категория работников (ИТР или рабочие определенной квалификации); d_2 – стаж работы на предприятии; d_3 – потенциальная текучесть; d_4 – вовлеченность в работу предприятия; d_5 – степень инновативности трудового поведения; d_6 – предрасположенность к изменениям в работе; d_7 – возраст.

В третью группу признаков, описывающих мнения и оценки респондентов, включены: p_1 – самооценки уровня жизни работников; p_2 – коэффициент желаемой заработной платы (кратный существующей); p_3 – удовлетворенность работой; p_4 – оценка возможности найти другую работу; p_5 – оценка изменения материального положения семьи по сравнению с прошлым годом; p_6 – самооценки информированности о своих трудовых правах; p_7 – мнение о соблюдении администрацией прав наёмных работников; p_8 – случаи нарушения трудовых прав и действия в этих случаях (нарушались или нет трудовые права; предпринимал ли что-то человек, если такие нарушения были; удалось или нет добиться успеха, если что-то предпринимал); p_9 – мнения о соответствии работы квалификации; p_{10} – оценки социального пакета на предприятии по сравнению с другими, известными респонденту, предприятиями; p_{11} – информированность и мнение о соблюдении коллективного договора (знают его содержание и полагают, что администрация его соблюдает; знают содержание коллективного договора и полагают, что администрация его не соблюдает; не знают содержание коллективного договора).

Из этого перечня признаков видно, что некоторые из них «простые», т.е. получены непосредственно из какого-то источника и в таком виде включены в анализ, а некоторые – сложные, т.е. превращенные в индексы. Некоторые индексы или интегральные показатели были разработаны участниками проекта ранее и использовались ими для мониторингового анализа ситуации на предприятии⁸. Другие создавались непосредственно для анализа средствами ДСМ-метода.

В нашей работе использовались средства интеллектуальной системы типа ДСМ (ИС-ДСМ), разработанные аспиранткой РГГУ А.Ю. Волковой [27]. Кратко алгоритмический процесс ДСМ-рассуждения описывается следующим образом. Закодированные данные просматриваются шаг за шагом, сравниваются, находятся группы респондентов, которые имеют одинаковые сочетания значений переменных, описывающих социально-демографические характеристики, биографические данные, мнения, признаки, характеризующие ситуацию этих конкретных людей. Выделяются группы кодов с аналогичным содержанием. Сочетание признаков, которые сопутствуют искомой характеристике (готовность либо неготовность добиваться справедливости в случае нарушения трудовых прав), интерпретируется как причина этой характеристики. При этом предполагалось изначально (и было подтверждено результатами эксперимента), что одно и то же явление (искомая характеристика) может порождаться разными причинами

(т.е. искомой характеристике сопутствует не одно, а несколько сочетаний признаков). Такое сочетание признаков мы называем гипотезой. Процесс поиска новых гипотез заканчивается тогда, когда исчерпаны все возможные сочетания признаков, сопутствующих искомой характеристике. Гипотезы о причинах исследуемого феномена порождаются автоматически, но из того материала, который предложил исследователь. Поэтому, помимо предположений о возможных детерминантах, мы можем оценить валидность используемых данных.

Комбинаций признаков (гипотез) может быть много; один и тот же признак может входить в разные комбинации. Функция социолога-аналитика здесь – оценить, насколько содержательны полученные комбинации и существенны их различия. Если различия существенны, это дает возможность получения обобщающих признаков (генерализаций) и формирования типологических единиц. Мы исходили из того, что если в изучаемой совокупности совпадают характеристики у нескольких людей, то это может свидетельствовать о возможном наличии типа. Предполагалось также, что количество типов ограничено.

Существенное облегчение работы социолога для типологизации на основе выявленных детерминант (причинных комбинаций) обеспечивается визуализацией результатов в системе по «кустовому» принципу. «Ядро» куста образуют максимальные пересечения описаний респондентов (как правило – несколько характеристик). Соответственно, дополнительные к «ядерным» «периферийные» признаки возникают при пересечении подмножеств респондентов, входящих в максимальное пересечение. Это даёт возможность охарактеризовать нюансы позиции субъектов, относящихся к одному и тому же типу поведения, т.е. предложить уточнение типологизации.

СОДЕРЖАТЕЛЬНЫЕ РЕЗУЛЬТАТЫ И ИХ ИНТЕРПРЕТАЦИЯ

В нашем исследовании сформировано два **результатирующих типологических** признака, описывающих готовность либо неготовность добиваться справедливости в случае нарушения трудовых прав: 1 – не будут добиваться справедливости; 2 – будут добиваться справедливости. На первом этапе были проанализированы данные по каждому предприятию в отдельности, чтобы выяснить, есть ли разница в типах работников на этих предприятиях. Затем данные анализировались на слитом массиве, что позволило включить параметр принадлежности к предприятию как особый признак, характеризующий ситуацию, и сделать выводы об устойчивости данных.

Первая группа – *не будут добиваться справедливости.*

На предприятии «А» был выделен один тип работников, не желающих отстаивать свои трудовые права, добиваться справедливости. Этот тип, метафорически названный «пассивный бедняк», включает два подтипа.

Первый подтип – пожилые инженеры. Сочетания ядерных признаков, описывающих эту группу, позволяют сказать, что это инженеры, у которых боль-

⁸ Формулы расчета приверженности работе на предприятии и вовлеченности в деятельность предприятия разработана участниками проекта Б.Г. Тукумцевым, В.Ю. Бочаровым и Н.В. Авдошиной как интегральный показатель, в котором учитывается ряд исходных показателей. Для приверженности работе – потенциальная текучесть и готовность порекомендовать работу на предприятии знакомым. Для вовлеченности в деятельность предприятия – информированность о задачах и проблемах предприятия, личная значимость этих проблем, готовность проявлять инициативу.

шой стаж работы, высокий уровень квалификации; низкие заработки (получают меньше, чем те, кто работает на предприятиях такого же типа в городе – от 0,8 до 0,6). Они не склонны менять место работы и не уверены, что у них есть возможность найти работу с лучшими условиями.

Дополнительные (периферийные) признаки – это те, которые встречаются не у всех, а лишь у некоторых представителей этого типа. Среди тех, кто не будут добиваться справедливости, есть и квалифицированные инженеры, и те, у кого сравнительно низкая квалификация. В некоторых сочетаниях встречается такой параметр, как уровень запросов относительно желаемой заработной платы. Как правило, они демонстрируют довольно скромные запросы. При том, что их заработки крайне низки, они хотят, чтобы их зарплата выросла примерно на треть или наполовину. Они не склонны к инновативному поведению, но иногда заявляют о своей готовности его принять, если это предложат другие инициаторы (администрация). Некоторые из них склонны думать, что администрация не соблюдает права наёмных работников.

Второй подтип – пожилые рабочие. От инженеров они отличаются в некоторых ядерных и некоторых периферийных признаках. Явное отличие в ядерных признаках: 1 – они готовы сменить место работы, если представится случай, и 2 – у них нет мнения по поводу того, соблюдает или нет администрация права наёмных работников. Самое заметное отличие в периферийных признаках этого подтипа рабочих в том, что они не готовы принимать новации в работе. В остальном они выглядят столь же бедными и безынициативными. Некоторые сказали, что не знают своих прав.

Таким образом, на предприятии «А» были выделены два подтипа работников, не склонных добиваться справедливости в случае нарушения их трудовых прав (пассивных бедняков). Общее у них – пожилой возраст, средний или выше среднего уровень квалификации, бедность. Различие в том, что представители первого типа убеждены в том, что хотя администрация и нарушает трудовые права, но нужно с этим мириться, потому что нет альтернатив занятости. Представители второго типа не рассматривают свою трудовую ситуацию в категории соблюдения либо несоблюдения трудовых прав. Они оценивают её, учитывая лишь возможность альтернативной занятости: если есть куда уйти, нет смысла отстаивать свои трудовые права.

На предприятии «Б», помимо таких же, как и на предприятии «А», двух подтипов работников, составляющих тип *пассивных бедняков*, выделился второй тип – *довольные работой*. Они убеждены в том, что бороться за свои права не нужно, потому что их никто не нарушает. Их отличие от первого типа в том, что они демонстрируют разные признаки включенности в дела предприятия. Они готовы принимать инновации и сами заниматься разными формами новаторства; они знают содержание коллективного договора; довольны содержанием своей работы и социальным пакетом; не намерены искать другую работу. Такие установки демонстрируют и рабочие, и инженеры. По уровню материального положения (зарботкам и доходу) они не отличаются от «пассивных

бедняков». Но есть отличия по возрасту: среди них встречаются и сравнительно молодые люди. Похоже, что именно они, несмотря на стесненные материальные обстоятельства, надеются на то, что им удастся сделать карьеру на предприятии в будущем.

Вторая группа – это те, кто будут добиваться справедливости, если столкнутся с нарушением трудовых прав.

На предприятии «А» основные (ядерные) характеристики тех, кто будут добиваться справедливости, если трудовые права будут нарушены, это: доход от 0,3 до 0,5 от среднего по г. Самаре; обеспеченность на крайне низком уровне – от одного до трех прожиточных минимумов и ниже. Дополнительные (периферийные) признаки: считают, что администрация не соблюдает либо «не всегда соблюдает» коллективный договор; что социальный пакет на предприятии хуже, чем на других; что их права нарушались; стабильные. В некоторых сочетаниях встретились молодые работники – 36-45 лет.

Главное отличие тех, кто будут добиваться справедливости, от тех, кто отказывается отстаивать трудовые права в том, что не оказался значимым признак профессионального статуса: рабочий или инженер. Но те, кто намерены добиваться справедливости, во-первых, беднее; во-вторых, не готовы искать другую работу; в-третьих, убеждены в том, что администрация нарушает их трудовые права и условия коллективного договора.

Таким образом, те, кто говорят, что будут добиваться справедливости, на предприятии «А» – это, если исключить нюансы, «отчаявшиеся бедняки». Других признаков, которые могли бы составить тип, нам не встретилось.

На предприятии «Б», помимо типа «отчаявшихся бедняков», у которых примерно те же типобразующие признаки, нам встретилось ещё два подтипа, которые различаются доминирующими ориентациями в труде: ориентированные на заработок и ориентированные на карьеру.

Основная характеристика первых – *ориентированные на заработок*. Они не склонны менять место работы (стабильные); высокая квалификация; стаж от 5 до 10 лет – не самый большой, но достаточный для того, чтобы освоиться на предприятии. Более высокие запросы относительно роста зарплаты, чем у тех, кто не намерен бороться за свои права (от 0,8 до 1), но при этом слабый интерес к новациям на работе и слабые проявления инициативы. Слабое знание трудовых прав и слабый интерес к ним. Готовы сменить место работы, если будут выгодные предложения.

Вторые – *ориентированные на карьеру*. У них те же характеристики, что и у первого подтипа (т.е. они сравнительно молоды и бедны), но они демонстрируют большую включенность в дела предприятия; знание содержания коллективного договора и своих прав. Вместе с тем, они полагают, что администрация не соблюдает или не всегда соблюдает условия коллективного договора и трудовые права работников. Работу менять они не намерены.

Анализ сочетаний признаков на слитом массиве показал наличие тех же типов, но в группе намеренных добиваться справедливости в некоторых сочета-

ниях встретился признак принадлежности к предприятию (А), что может свидетельствовать о большей социальной напряженности на этом предприятии.

Выделенные типы – это номинальные группы людей, объединенные общими искомыми признаками. Сочетание признаков, составляющих типы, позволяет понимать проблемы этих людей и принимать соответствующие управленческие решения по отношению к каждой из выделенных групп применительно к конкретной ситуации. В частности, тот факт, что о несоблюдении трудовых прав и готовности их отстаивать говорят не только (и даже не столько) самые обездоленные, но, прежде всего, сравнительно молодые и ориентированные на карьеру на данном предприятии (Б) работники, свидетельствует о слабом внимании администрации к обеспечению возможностей для карьерного роста, кроме того, неготовность бороться за свои трудовые права, добиваться справедливости вовсе не означает, что с этим на предприятии всё в порядке. Массовыми являются ситуации, когда работники не выступают в защиту своих трудовых прав не потому, что их никто не нарушает, а потому, что они боятся, не верят в то, что удастся добиться справедливости. Но пассивность вовсе не означает готовности демонстрировать трудовой энтузиазм. Скорее напротив. Пассивность в отстаивании своих трудовых прав означает и пассивность на работе.

МЕТОДИЧЕСКИЕ ВЫВОДЫ

1. Использованный вариант формализованного анализа социологических данных предъявляет высокие требования к языку представления данных, при соблюдении которых метод способен порождать причинно-следственные зависимости. В противном случае результаты его работы можно рассматривать как «информацию к размышлению» для социолога, требующую коррекции, а не готовые выводы. Соответственно, формирование типов – это лишь отчасти результат использования математического аппарата, а в значительной мере – результат размышлений социолога. Поэтому, если в результате формализованного анализа получаются типы, мало различающиеся между собой, это означает либо то, что исходные показатели слишком дробные (например, разные степени бедности, которые, как оказалось, не влияют на искомый результат), либо то, что они не слишком значимы для исследуемого эффекта. Такие признаки образуют «шумы» в виде сочетаний признаков, не прибавляющих ничего к уже имеющимся типам. В нашем случае таким признаком оказались информированность о наличии коллективного договора и мнение о его соблюдении администрацией.

Дальнейший анализ допускает возможность агрегирования исходных данных, сведения их в комплексы (генерализации) для уменьшения количества гипотез. Такие сочетания уже предъявлены в данных, и, тем самым, процесс формирования генерализаций становится объективированным и прозрачным. Например, тот факт, что в качестве детерминант для описания типа обнаружились признаки, описывающие в совокупности адаптированность респондента к условиям работы на предприятии, позволяет сфор-

мировать из нескольких исходных признаков (освоился на работе, знает свои права, знает задачи предприятия, знает содержание коллективного договора) комплексную характеристику (генерализацию): «адаптированность – неадаптированность». Здесь можно усмотреть некоторые аналогии с порождением категорий в обоснованной теории, для которой существенной проблемой является субъективное представление кодов. В предложенном подходе эта процедура оказывается формализованной.

Разумеется, выделение именно тех факторов, которые определяют явление (внутренняя валидность качественного эксперимента) – непростая задача, от квалифицированного решения которой зависит результат формализованного анализа. В нашем случае она обеспечивается выполнением так называемого критерия достаточного основания гипотез, реализующего абдуктивное объяснение имеющихся фактов. Абдуктивное объяснение позволяет говорить о принятии порождённых в результате индуктивного вывода гипотез с некоторой степенью правдоподобия. Следовательно, полученные выводы могут быть перенесены на случаи, отличные от участвовавших в эксперименте, также с некоторой степенью правдоподобия. Отметим, что получение устойчивых гипотез о детерминантах изучаемого явления на множестве расширяющихся БФ (что предусмотрено процедурами ДСМ-метода) позволит говорить о повышении степеней правдоподобия таких гипотез – вплоть до придания им статуса эмпирических закономерностей [28].

2. С помощью ДСМ-метода можно решать и более простые, по сравнению с описанным примером, задачи. Например, можно ответить на вопрос: «Какие из частных признаков формируют общую, интегральную характеристику?». Так, показатель удовлетворенности трудом в целом, вообще говоря, связан с показателями удовлетворенности разными аспектами трудовой ситуации. Ответ на вопрос об их связи может быть получен и статистическими методами, однако для этого должны выполняться условия их применимости. В нашем случае (малые выборки) возможности ДСМ-метода оказываются практически незаменимыми. Метод показал свою применимость как исследовательская стратегия, обеспечивающая преимущества при исследовании малых и средних массивов данных (~ до 100 случаев).

3. ДСМ-метод позволяет учесть в анализе и прогнозе разные уровни детерминаций (ситуационный, установочный, ценностный, нормативный) и оценить вклад каждого уровня в изучаемое поведение. При этом, в отличие от некоторых статистических методов (например, факторного анализа), не требуется, чтобы признаки, включенные в анализ, имели один и тот же формат, но нужно, чтобы они изначально обладали существенными дискриминирующими свойствами (например, не были слишком дробными). Если у исследователя нет предположений относительно «шага» возможных различий (например, размер или другие характеристики дохода, позволяющие отнести респондента к «бедным», «средним» или «состоятельным»), то ДСМ-метод способен предложить такие решения. Например, в нашем случае респон-

денты, имеющие доход от 0,3 до 0,5 среднего по г. Самаре, и те, у кого доход от 0,6 до 0,8 от среднего по г. Самаре, мало отличались друг от друга, что обусловило необходимость менее дробных делений.

4. Социология, как правило, имеет дело со статистическим, а не с реальным объектом. ДСМ-метод формирует группы реальных респондентов, обладающих некоторым набором признаков, отличающих их от другой, противоположной по какому-то признаку группы, и, соответственно, может предложить описания этих групп.

ЗАКЛЮЧЕНИЕ

Предлагаемые средства представления и последующего интеллектуального анализа социологических данных представляют собой приближение к формализации исследовательских эвристик «анализ данных – предсказание – объяснение», характерных для качественного анализа социологических данных. Работа в этом направлении является ответом специалистов в области искусственного интеллекта на вызовы потребностей современной социологической науки. Однако достижение интересных результатов возможно лишь при взаимодействии глубокого (гуманистического) понимания сущности изучаемых социальных проблем и профессионального применения исследовательских процедур, что требует от социолога серьезной предварительной работы. Существующие традиции сбора и подготовки социологических данных, ориентированные на применение статистических методов анализа, нуждаются в серьезной трансформации. С другой стороны, именно потребности формализованного качественного анализа социологических данных (ФКАСД) оказываются полезным стимулом для развития инструментов компьютерной поддержки такого анализа средствами искусственного интеллекта.

СПИСОК ЛИТЕРАТУРЫ

1. Вебер М. Избранное: протестантская этика и дух капитализма. – М.: РОССПЭН, 2006. – 656 с.
2. Lewins A., Silver C. Using Software in Qualitative Research: A Step by Step Guide. – London: Sage Publications, 2007. – 304 p.
3. Давыдов А.А. Меморандум IV Конференции памяти А.О. Крыштановского «Современные проблемы формирования методного арсенала социолога». – 2010. – URL: http://www.ssa-rss.ru/index.php?page_id=273
4. Fayyad U. M., Piatetsky-Shapiro G., Smyth P. From Data Mining To Knowledge Discovery in Databases // AI Magazine. – 1996. – Vol. 17, № 3. – P. 37-54.
5. Арский Ю. М., Финн В. К. Принципы конструирования интеллектуальных систем // Информационные технологии и вычислительные системы. – 2008. – № 4. – С. 4-36.
6. Fielding N. G. Automating the ineffable: Qualitative software and the meaning of qualitative research // Qualitative research in action / ed. Tim May. – London: Sage Publication, 2003. – P. 161-178.
7. Gobo G. The Renaissance of Qualitative Methods // FQS Forum: Qualitative Social Research. – 2005. – Vol. 6, № 3. – Art. 42.
8. Страусс А., Корбин Дж. Основы качественного исследования. Обоснованная теория. Процедуры и техники. – М.: КомКнига, 2007. – 256 с.
9. Поппер К. Эволюционная эпистемология // Эволюционная эпистемология и логика социальных наук. – М.: Эдиториал УРСС, 2000. – С. 57-74.
10. Configurational Comparative Methods. Qualitative Comparative Analysis (QCA) and related techniques (Applied Social Research Methods) / eds. B. Rihoux, C. C. Ragin. – Thousand Oaks; London: Sage Publications, 2009. – 240 p.
11. Автоматическое порождение гипотез в интеллектуальных системах / сост. Е.С. Панкратова, В.К. Финн; под общ. ред. В.К. Финна. – М.: Книжный Дом «Либроком», 2009. – 528 с.
12. Михеенкова М.А. О логических средствах интеллектуального анализа социологических данных // Искусственный интеллект и принятие решений. – 2010. – № 1. – С. 20-32.
13. Кученкова А.В. Логико-комбинаторные методы анализа социологических данных: эвристический потенциал и методическая специфика: автореф. дис. ... кандидата социологических наук. – М., 2012. – 25 с.
14. Росс Л., Нисбетт Р. Человек и ситуация (уроки социальной психологии). – М.: Аспект Пресс, 2000. – 429 с.
15. Виноградов Д.В. Формализация правдоподобных рассуждений в логике предикатов // НТИ. Сер. 2. – 2000. – № 11. – С. 17-20.
16. Скворцов Д.П., Финн В.К. Замечания об одном расширении языка многосортной логики предикатов // НТИ. Сер. 2. – 1981. – № 8. – С. 25-26; Finn V.K., Skvortsov D.P. A Note on Extension of Multisort Predicate Logic Language // Automatic Documentation and Mathematical Linguistics. – 1981. – Vol. 15. – No 4. – P. 89-91.
17. Гусакова С.М., Михеенкова М.А., Финн В.К. О логических средствах анализа мнений // НТИ. Сер. 2. – 2001. – № 5. – С. 4 – 24.
18. Финн В.К. Индуктивные методы Д.С. Милля в системах искусственного интеллекта. Часть I // Искусственный интеллект и принятие решений. – 2010. – № 3. – С. 3-21.
19. Финн В.К., Михеенкова М.А. О логических средствах концептуализации анализа мнений // НТИ. Сер. 2. – 2002. – № 6. – С. 4-24.
20. Финн В.К., Михеенкова М.А. О ситуационном расширении ДСМ-метода автоматического порождения гипотез // НТИ. Сер. 2. – 2000. – № 11. – С. 20-30; Finn V.K., Mikheyenkova M.A. Situation Extension of the JSM Automatic Hypotheses Generation Method // Automatic Documentation and Mathematical Linguistics. – 2000. – Vol. 34. – No 6. – P. 11-26.

21. Ситуационный анализ социально-трудовой сферы на предприятии: концептуальные основания // Социологические методы в современной исследовательской практике: сб. статей, посвященный памяти первого декана факультета социологии НИУ ВШЭ А.О. Крыштановского / отв. ред. и вступит. ст. О.А. Оберемко; НИУ ВШЭ, ИС РАН, РОС. – М.: НИУ ВШЭ. – 2011. – С. 106 – 112.
22. Климова С.Г., Михеенкова М.А., Панкратов Д.В. ДСМ-метод как метод выявления детерминант социального поведения // Автоматическое порождение гипотез в интеллектуальных системах. - М.: URSS, 2009. – С. 410-427.
23. Смелзер Н. Дж. О компаративном анализе, междисциплинарности и интернационализации в социологии // Социологические исследования. – 2011. – № 11. – С. 3-11.
24. Ядов В. А. Стратегия социологического исследования. – М.: Добросвет, 2003. – 596 с.
25. Толстова Ю. Н. Математико-статистические модели в социологии. – М.: Издательский дом ГУ-ВШЭ, 2008. – 243 с.
26. Климова С.Г. Региональный контекст проблемной ситуации на предприятии: опыт анализа // Социологические исследования. – 2011. – № 11. – С. 25-34.
27. Волкова А.Ю. Алгоритмизация процедур ДСМ-метода автоматического порождения гипотез // НТИ. Сер. 2. – 2011. – № 5. – С. 6-12.
28. Финн В.К. Об определении эмпирических закономерностей посредством ДСМ-метода автоматического порождения гипотез // Искусственный интеллект и принятие решений. – 2010. – № 4. – С. 41-48.

Материал поступил в редакцию 10.07.12.

Сведения об авторах

КЛИМОВА Светлана Гавриловна – кандидат философских наук, ведущий научный сотрудник Института социологии РАН

E-mail: sgklimova@mail.ru

МИХЕЕНКОВА Мария Анатольевна – кандидат технических наук, старший научный сотрудник ВИНТИ РАН

E-mail: mmikh@viniti.ru

Методы генерации словарей

*Предлагается классификация лексикографических ресурсов, необходимых для поддержки функционирования систем автоматического анализа текста. Выделяются и описываются четыре вида словарей: терминологические, терминологико-статистические, тезаурусы, онтологии. Методы генерации словарей классифицируются на статические и динамические; линейные и ступенчатые. Методы взвешивания терминов делятся на интертекстуальные и интратекстуальные. Подробно рассматриваются особенности алгоритма TF*IDF. Описываются две методики генерации словарей.*

Ключевые слова: лексикографические ресурсы, классификация, методы генерации словарей, формула TF*IDF

Функционирование современных систем автоматической обработки текста невозможно без лексикографической поддержки, в качестве которой выступают словари различных типов. Как мы полагаем, лексикографические источники можно разделить на четыре основных вида: терминологические словари, терминологико-статистические словари, тезаурусы, онтологии.

Терминологические словари представляют собой списки терминов - единиц текста. В качестве терминов выступают ключевые слова и/или словосочетания, выделяемые по некоторым параметрам и отражающие специфику данного текста/текстов и/или его основное содержание. Основным параметром для распознавания терминов-ключевых слов служат их весовые коэффициенты. В связи с этим возникает проблема разработки адекватных алгоритмов взвешивания терминов. Существенной особенностью этого вида

словарей является возможность их динамической генерации: словарь не закладывается заранее в базу данных, а создаётся на лету, в ответ на запрос пользователя. На рис.1. в левой части окна – словарь, автоматически генерируемый программой Event Tracking Summarizer (ETS) в процессе реферирования.

Остальные виды словарей предварительно записываются в базу данных, т.е. создаются статически, причём достаточно часто вручную, экспертами.

Терминологико-статистические словари содержат статистические данные о распределении единиц текста. К таким данным могут относиться частотность, распределение по файлам, ранг. В качестве примера можно привести лемматизированные и нелемматизированные списки слов Британского национального корпуса (БНК), составленные А.Килгарифом (табл. 1) [1].

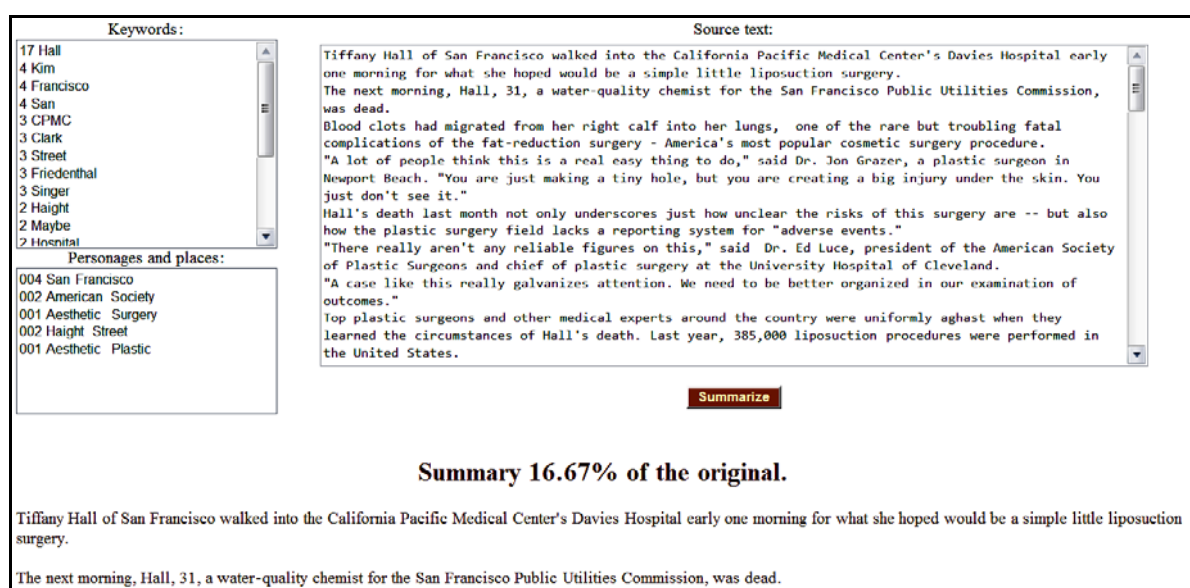


Рис. 1. Интерфейс программы ETS¹

¹ Интерфейс онлайн-версии, доступной на http://csl.khsu.ru/ets_web/default.aspx

Терминологический-статистический словарь БНК А.Килгарифа (выдержка)

Лемматизированный список				Нелемматизированный список			
Ранг	Частотность	Слово	Тег части речи	Частотность	Слово	Тег части речи	Количество файлов
2333	3703	bomb	n	22147	book	nn1	2336
4960	1202	bomb	v	47	book	vvb	28
252	37675	book	n	2321	frequent	aj0	1094
3623	1894	book	v	39	frequent	vvb	37

Статистические данные, содержащиеся в словарях этого типа, имеют существенное значение для определения вероятностных характеристик, которые необходимы при разработке ряда систем автоматического анализа текста. Например, вероятностные параметры можно учитывать при разработке стохастических теггеров частей речи, причём можно игнорировать случаи маловероятного использования токенов с определёнными тегами. Как показано в табл. 1, частотность использования *frequent* как прилагательного почти в 60 раз выше его использования в качестве глагола, в вероятностных величинах разница составит $0,002321 - 0,000039 = 0,002282$ (при 1000000 слов в корпусе). В данном случае вполне возможно игнорировать глагольные формы, и всем токенам *frequent* приписывать тег прилагательного, поскольку вероятность ошибки крайне мала. То же самое относится и к использованию *book* в качестве существительного и личного глагола (*vvb*). Однако вряд ли это возможно по отношению к глагольным и именным формам *bomb*, поскольку разница их вероятностных значений не так существенна.

Тезаурусы предоставляют информацию о терминах, связанных с данным термином структурно-семантическими связями: синонимическими, антонимическими, гипоним-гиперонимическими. Наиболее широко известным тезаурусом для английского языка является *Wordnet*, разработанный в Принстонском университете США и распространяющийся с открытым исходным кодом, локализованным для различных языков программирования [2]. По проблемам применения этого тезауруса в целях автоматического анализа текста даже проводятся международные конференции, что свидетельствует об актуальности разработок словарей этого типа.

Основным понятием, лежащим в основе архитектуры *Wordnet*, является понятие синонимических рядов (*synsets*) – группы семантически связанных терминов, распределённых по частям речи, которые различаются в зависимости от степени смысловой близости. Смысловая близость определяется расстоянием от исходного (текущего) слова. Если взять в качестве исходного, например, слово *courage*, то синонимический ряд первого уровня составят слова, через которые непосредственно толкуется данное слово: *courageousness*, *bravery*, *braveness*. К синонимическому ряду второго уровня относятся слова, выделяемые стрелкой в нижней строке (рис. 2). В данном случае синонимический ряд второго уровня включает только одно слово: *spirit*. К синонимиче-

скому ряду третьего уровня будут относиться синонимы *spirit*: *character*, *fiber*, *fibres*. К синонимическому ряду четвёртого уровня относятся синонимы слов, находящихся на третьем уровне и т.д. Таким образом, создаётся типичная гипертекстовая структура с непрерывными переходами от одного кластера синонимов к другому.

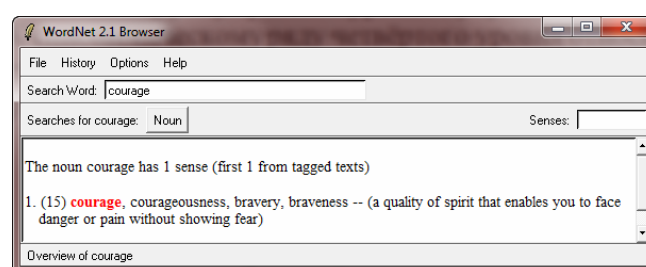


Рис.2. Интерфейс тезауруса WordNet

Очевидно, что на каждом уровне уменьшается степень смысловой близости с исходным словом; в сущности на втором и третьем уровнях располагаются не синонимы, а его гиперонимы. Группа экспертов во главе с Т.Педерсеном разработали программу *WordNet Similarity*, которая позволяет вычислять коэффициент смысловой близости между словами по различным методикам [3]. Для *courage* и *bravery* коэффициент смысловой близости, подсчитанный по расстоянию между синсетам (*path length*) равен 1, а для *courage* и *character* (в значении "совокупность свойств личности") – 0,3333.²

Тезаурусы применяются в информационно-поисковых системах и системах автоматического реферирования, системах автоматической классификации и категоризации текстов, системах интеллектуального анализа текста. В ИПС применение тезаурусов является эффективным средством расширения поисковых запросов. По данным, которые приводятся в [4], в ИПС *Яндекс* средняя длина поискового запроса в 1997 г. составляла 1,2 слова, а в 2011 г. – 2,8 слова. В [5] приводятся следующие данные о распределении поисковых запросов разной длины в ИПС *Яндекс* и *Google* для русского сектора Интернета. Запросы, состоящие из одного слова, составляют 16,999% от общего количества запросов в *Яндексе* и 17,17 – в *Google*; состоящие из двух слов – 28,076% и 27,784%; состоящие из трёх слов – 24,261% и

² Данные получены с использованием веб-интерфейса <http://talisker.d.umn.edu/cgi-bin/similarity/similarity.cgi>

24,226%; состоящие из четырёх слов - 14,966% и 15,123%; состоящие из пяти слов - 7,82% и 8,054%, соответственно. Очевидно, что если запрос состоит из одного - двух слов, то достаточно сложно определить релевантные документы, анализируя распределение терминов запроса в текстах, находящихся в базе данных информационно-поисковой системы. Поэтому к терминам поискового запроса автоматически, с помощью словаря-тезауруса, добавляются их синонимы и/или другие семантически соотносимые термины, что позволяет повысить эффективность поиска.

Под онтологиями понимаются сложно структурированные словари, моделирующие структуру определённой предметной области на основе функциональных отношений между компонентами. Сложная структура проявляется в многоуровневой иерархии, причём на определённых уровнях компоненты онтологии - понятия и категории - соотносятся с конкретными терминами – инстанциями [6]. Онтологии классифицируются на формальные и лингвистические [7]. Специфика лингвистической онтологии состоит в том, что она связана с грамматикой, правила которой позволяют распознавать компоненты онтологии и отношения между ними в тексте. В [8] описывается шестиуровневая лингвистическая онтология, разработанная для поддержки системы автоматического анализа мнений пользователей о коммерческих продуктах. На верхнем уровне онтологии находятся категории семантических и синтаксических терминов, которые, соответственно, либо выражают положительную или отрицательную оценку, либо изменяют её интенсивность. Онтология связана с линейной грамматикой, благодаря которой имена продуктов, указанные в поисковом запросе, соотносятся с компонентами онтологии.

Формальные онтологии никак не связаны с анализом текста и содержат термины и соотносящиеся с ними количественные и числовые данные. К ним могут относиться собственные имена, обозначающие, например, названия продукции; количественные данные о покупках и продажах; данные о возрасте, профессиональном статусе, поле и т.д. Такая информация представляется в табличном формате баз данных и обрабатывается средствами СУБД. Онтологии такого типа содержат на верхнем уровне иерархии две основных категории: *покупатели* и *продукты*. Категория покупателей делится на подкатегории в зависимости от типов личных данных, а категория продуктов – в зависимости от ассортимента продукции. В качестве инстанций выступают названия конкретных продуктов. Применение формальных онтологий позволяет выявить имплицитные зависимости между покупкой товаров определённого ассортимента и такими данными о покупателях, как наличие в собственности дома, марка принадлежащего автомобиля, возраст, профессия, доход, расстояние между домом и магазином. В результате может быть принято решение об изменении ассортимента товаров, которое повлияет на увеличение количества продаж. Основной проблемой, возникающей при разработке формальных онтологий, является сбор личных данных пользователей. Обычно пользователь не готов тратить большое количество времени на за-

полнение анкет и, тем более, предоставлять личные данные. Поэтому крупные производители и торговые сети стимулируют пользователей, предоставляя за заполнение анкет скидки на товары, бонусы, или просто переводя деньги на счета пользователей и покупателей. В последнее время в отдельную профессию выделилось написание обзоров товаров и услуг [9].

Методы создания словарей можно, как мы полагаем, разделить на три основных вида: динамические, статические, гибридные. К динамическим относятся методы генерации словарей на лету, по определённому алгоритму, в результате запроса пользователя в процессе функционирования системы. Таким динамическим способом создаются словари в информационно-поисковых системах, системах автоматического реферирования текста универсального типа. Существенным требованием в данном случае является быстроедействие, поскольку пользователь не готов долго ждать результатов работы системы. Некоторые ИПС выводят данные о количестве времени, затраченного на обработку запроса. К статическим относятся методы, предусматривающие создание и запись словаря в базу данных до начала функционирования системы. Часто такие словари создаются и редактируются вручную экспертами, и к процессу составления словаря не предъявляются требования по быстродействию. Экспертные словари используются в системах, предназначенных для обработки текстов, относящихся к определённой предметной области. В случае применения гибридных методов часть словаря закладывается в систему заранее, а другая часть генерируется динамически. В качестве примера можно привести предложенную нами ранее методику создания словаря для автоматического анализа научного текста [10]. Вначале динамически создаётся часть словаря, включающая знаменательные слова из заглавия текста, далее к этим словам добавляются семантически соотносимые термины из заранее интегрированной лексической базы данных типа *WordNet*.

В зависимости от темпа создания словаря мы предлагаем разделить методы генерации на линейные и ступенчатые. Линейные методы предусматривают создание словаря в один темп, на одной стадии, одном этапе, а ступенчатые включают два или более этапов. На первом этапе выделяются наиболее значимые термины, составляющие ядерную часть словаря, его первый эшелон; на втором этапе ядерная часть расширяется, к ней добавляются другие термины, связанные с терминами в ядерной части, например синонимы, составляющие второй эшелон. На следующем этапе возможно расширение второго эшелона и образование третьего эшелона и т.д.

Применение как линейных, так и ступенчатых методов генерации словарей предусматривает распознавание наиболее значимых терминов (*salient terms*), выражающих основное содержание данного текста. Обычно такое распознавание проводится на основе различных алгоритмов и программ взвешивания, на выходе у которых - список терминов, ранжированных по весовым коэффициентам, отражающих значимость каждого термина. Термины с наиболее высокими коэффициентами отбираются в словарь, репрезентирующий содержание текста. Как мы пола-

Вычисление коэффициентов по формуле $TF \cdot IDF$

Термин	Частотность в документе D_i (tf_{ij})	N	n	w_{ij}
information	4	4	3	1,66
artificial	12	4	1	24
the	15	4	4	0
intelligence	16	4	2	16
computer	5	4	3	2,075

гаем, алгоритмы взвешивания можно разделить на два основных вида: интертекстуальные и интратекстуальные. Интертекстуальные алгоритмы предполагают сопоставление распределения терминов в данном тексте с их распределением в другом тексте/текстах. Интратекстуальные алгоритмы выполняются только с учётом внутренней структуры данного текста. Наиболее простым является взвешивание терминов на основе их частотностей. В текстовом редакторе *MS Word 2003* сервисная функция "Автореферат" выполнялась именно на основе подсчёта сырых частотностей по следующему алгоритму: 1) подсчет частотности слов в данном тексте (служебные слова игнорируются); 2) приписывание весовых коэффициентов каждому слову в зависимости от его частотности; 3) определение весовых коэффициентов предложений. Весовой коэффициент каждого предложения определяется соотношением между суммой весовых коэффициентов входящих в него слов и общим количеством слов в предложении; 4) ранжирование предложений в зависимости от весовых коэффициентов; 5) отбор предложений в реферат в соответствии с установленным размером реферата. По умолчанию размер реферата определялся в 25 процентов от оригинала, однако пользователь может выбрать размер, указав количество предложений, либо количество слов, либо процент компрессии [11]. В более поздних версиях *MS Word* (2007, 2010) функция автоматического реферирования была удалена, по-видимому, из-за низкой эффективности. В разработанной нами методике симметричного взвешивания весовые коэффициенты приписывались предложениям на основе рекуррентности и симметричности [10].

Существенным недостатком взвешивания на основе сырых частотностей (*raw counts*) является зависимость от размера текста: чем больше текст, тем больше частотность слов. Поэтому такой тип взвешивания применяется достаточно редко и только в интратекстуальных методиках взвешивания, а наиболее простой вариант устранить зависимость от размера текста – вычисление вероятностных величин, которые получаются делением частотности данного термина на общее количество терминов в тексте. Особую важность устранение зависимости весовых коэффициентов от размера текста имеет для интертекстуальных методов взвешивания.

Наиболее распространённым интертекстуальным методом взвешивания считается метод $TF \cdot IDF$ (*term frequency * inverse document frequency*), который легко алгоритмизируется на основе формулы:

$$w_{ij} = tf_{ij} * \log_2 \frac{N}{n}, \quad (1)$$

где

w_{ij} = вес термина t_j в документе d_i ,

tf_{ij} = частотность термина t_j в документе d_i ,

N = общее количество документов в корпусе,

n = количество документов в корпусе, в которых

термин t_j встречается хотя бы один раз [12].

Применение формулы можно проиллюстрировать на примере некоторого документа D_i , статистические данные о котором приведены в табл. 2.

Наибольший коэффициент получает термин, имеющий высокую частотность в данном документе и наиболее редко встречающийся в корпусе текстов. Вместе с тем, наименьшие коэффициенты получают термины, встречающиеся в большем количестве текстов, а термины, которые есть во всех текстах корпуса, получают нулевые коэффициенты (т.е. при $N=n$). К последним относятся стоп-слова: артикли, предлоги, местоимения, союзы. Таким образом, алгоритм позволяет как фильтровать стоп-слова (без применения специальных списков), так и определять наиболее значимые термины в данном тексте. К его недостаткам можно отнести то, что термины, достаточно частотные в данном тексте, но не встречающиеся в корпусе ($n=0$), также получают нулевые коэффициенты (возникает известная ошибка деления на ноль). Однако в ряде случаев именно эти термины могут быть значимыми для содержания текста. Например, в научной работе автор может использовать неологизмы для обозначения вновь открытых объектов, а в научно-фантастическом произведении – для описания своего представления о будущем. Поэтому предлагается применять модифицированный вариант формулы со знаменателем в виде $n+1$. В этом случае авторские неологизмы получают достаточно высокие весовые коэффициенты, зависящие от их частотности в данном тексте. Ещё одной проблемой является количественный и жанрово-стилистический состав корпуса текстов (величина N в формуле). Количество текстов в корпусе задаёт величину N , значение которой непосредственно влияет на результат вычислений. К настоящему времени не сформулированы однозначные требования к репрезентативности корпуса, что ведёт к произвольности определения этой величины. Не вполне определены и требования к жанрово-стилистическому составу корпуса. Если, например, анализируемый текст художественный, то следует ли его сопоставлять с корпусом художественных текстов? Или в корпус нужно включить научные и/или газетные тексты?

Мы можем предложить следующие варианты решения данных проблем, исходя из общепринятой жанрово-стилевой классификации текстов на художественный, газетные, научные. Контрастивное сопоставление имеет смысл в том случае, если анализируемый текст с большой долей вероятности содержит неологизмы. К таким текстам могут относиться научные статьи и произведения жанров науч-

ной фантастики и фэнтези. Возможно и двухступенчатое взвешивание: вначале анализируемый текст сопоставляется с текстами того же жанра, затем с текстами противоположного жанра. В случае применения модифицированного варианта формулы, сопоставление научного текста с корпусом художественных текстов поможет выявить лексику, характеризующую специфику научного жанра в целом, а сопоставление с корпусом научных текстов – определить терминологию, специфичную для данного текста. Для газетных текстов целесообразно сопоставление с текстами того же жанра, поскольку газетные сообщения могут соотноситься с разными предметными областями (медициной, техникой, образованием, политикой). Проблема репрезентативности и количественного состава может быть решена включением в корпус всех текстов данного жанра, которые используются в национальных корпусах. Для английского языка это – Британский национальный и Американский национальный корпуса, оффлайн-версии которых распространяются платно. Как известно, национальные корпуса составляются экспертами, и включаемые в них тексты полностью соответствуют критериям репрезентативности.

В табл. 3 приводятся 10 слов с наиболее высокими коэффициентами, полученными в результате сопоставления научного текста³ с корпусом художественных и научных текстов. Художественные тексты были взяты из каталога *Fiction* Американского национального корпуса, а научные – из каталога *Biomed* того же корпуса. Количество текстов и в том и в другом корпусе (N) – 105. Коэффициенты подсчитывались по классической формуле $Tf*Idf$ (Формула 1) и в соответствии с модифицированным вариантом (Формула 2), согласно которому дополнительный весовой коэффициент I , приписывался только словам, которые встречаются в анализируемом тексте, но отсутствуют в корпусе.

$$w_{ij} = tf_{ij} * \log_2 \frac{N}{\{n+1 | n=0\}}. \quad (2)$$

Из табл. 3 видно, что распределение слов по коэффициентам существенно различается. Наиболее адекватные результаты получились при сопоставлении с художественным текстом на основе модифицированной формулы. В список попали слова, отражающие содержание статьи: *collocate*, *concordance*, *computational*, которые отсутствуют в списке, составленном на основе применения классической формулы. Существенно хуже результаты на основе сопоставления с корпусом научных текстов, поскольку в списке оказались общенаучные и общелитературные слова *Language*, *traffic*, *English*. Очевидно, что следует учитывать предметную область, с которой соотносится анализируемый текст – в данном случае это корпусная лингвистика. Сопоставление с текстами по биологии и медицине не вполне корректно. В любом случае, результаты, представленные в табл. 3, в явном виде демонстрируют зависимость вычислений по формуле $Tf*Idf$ от состава корпусов.

Данные, приводимые в табл. 3, были получены с помощью *TF*IDF Ranker* – программы, специально разработанной нами для вычисления коэффициентов по двум вариантам формулы. Программа распространяется бесплатно на сайте нашей лаборатории.⁴

Мы можем предложить следующую методику генерации словаря с использованием *TF*IDF Ranker*:

1) используя *TF*IDF Ranker*, сопоставьте анализируемый текст с корпусом текстов, относящихся к той же предметной области/жанру. Скопируйте результаты и составьте список 1;

2) сопоставьте анализируемый текст с корпусом текстов, относящихся к другой предметной области/жанру. Скопируйте результаты и составьте список 2;

3) из верхней части каждого списка отберите n слов с наиболее высокими коэффициентами. В качестве порогового уровня можно применить среднеарифметическое значение. В новые списки (1-1 и 2-1) войдут все термины, вес которых превышает данное значение;

4) произведите пересечение списков 1-1 и 2-1 и составьте новый список (список 3), который включает термины, встречающиеся в обоих списках;

5) нормализуйте коэффициенты терминов в списке 3. Для этого можно применить вероятностные величины, которые получаются делением частотности данного термина в анализируемом тексте на общее количество терминов в данном тексте.

Мы также можем предложить интратекстуальную методику генерации словаря, которая предусматривает применение разработанной нами ранее интерпретации закона Брэдфорда в терминах геометрической прогрессии (Y -интерпретации) [13] на основе формулы

$$S_n = J_2(q^n - 1)/(q - 1), \quad (3)$$

где $n = 3$ – постоянная величина, равная количеству зон по закону Брэдфорда; q – знаменатель прогрессии, равный константе Брэдфорда. Для анализа текстовых данных можно установить значение $q=3$; J_2 – вторая зона, характеризующаяся наибольшей степенью рассеяния информации; S_n – сумма членов геометрической прогрессии:

1) получите список терминов данного текста, ранжированных по частотности. Для этого можно применить существующее бесплатное программное обеспечение, например конкорданс AntConc [14];

2) суммируйте частотности всех терминов. Будет получена величина S_n в формуле (2);

3) определите числовое значение зоны J_2 . Для этого преобразуйте формулу (2) в уравнение;

$$J_2 = S_n / (q^n - 1) / (q - 1) = S_n / (3^3 - 1) / (3 - 1) = S_n / (27 - 1) / 2 = S_n / 13 \quad (4)$$

4) определите числовое значение зон J_1 и J_0 , что выполняется последовательным умножением на коэффициент q . Вначале на этот коэффициент умножается полученное ранее значение зоны J_2 , в результате чего получается значение зоны J_1 , затем значение зоны J_1 также умножается на этот коэффициент и находится значение зоны J_0 ;

³ Danielsson P. What constitutes a unit of analysis in language? – URL: http://www.linguistik-online.de/31_07/danielsson_a.html

⁴ URL: http://vetsky.narod2.ru/catalog/tfidf_ranker

**Результаты сопоставления научного текста с корпусом
художественных и научных текстов**

Художественный текст				Научный текст			
Слово	Формула 1	Слово	Формула 2	Слово	Формула 1	Слово	Формула 2
jam	0,02638	jam	0,02638	word	0,013297	jam	0,030996
Corpus	0,009955	units	0,022224	meaning	0,007718	word	0,013297
Unit	0,007553	Corpus	0,009955	Language	0,007487	Corpus	0,011697
multi	0,006887	Unit	0,007553	units	0,006853	meaning	0,007718
node	0,005264	collocate	0,007018	words	0,006481	Language	0,007487
frequency	0,004679	multi	0,006887	traffic	0,006433	collocate	0,007018
meaning	0,004152	concordance	0,005848	English	0,005264	units	0,006853
word	0,003974	computational	0,005264	multi	0,005103	words	0,006481
identified	0,003956	node	0,005264	Unit	0,004805	traffic	0,006433
frequent	0,003696	frequency	0,004679	concordance	0,003597	English	0,005264

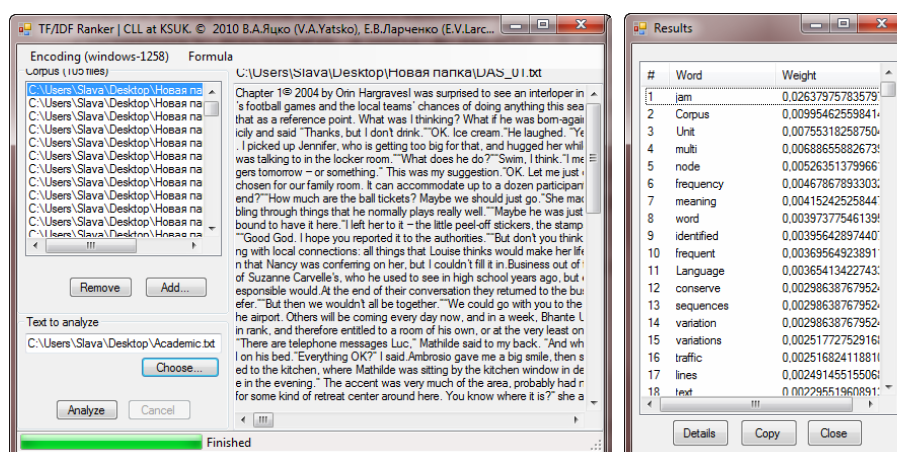


Рис.3. Программа *TF*IDF Ranker*

5) определите состав каждой из зон, т.е. какие именно термины входят в каждую из зон. Для этого с помощью табличного процессора вычислите диапазон, дающий сумму числового ряда, соответствующего той или иной зоне, меняя количество членов числового ряда.

Термины, входящие в зону J_0 , – служебные слова, часто встречающиеся во всех текстах. В информационно-поисковых системах такие термины обычно игнорируются, фильтруются. Вместе с тем анализ распределения служебных слов часто используется с целью автоматической классификации текстов [15]. Термины, входящие в зону J_1 , представляют основное содержание текста. Их распределение по частотности уникально характеризует данный текст, позволяя отличить его от других текстов. Распределение терминов в этой зоне может быть использовано в целях информационного поиска и автоматического реферирования, а также и автоматической классификации. В последнем случае вычисления следует проводить не на основе сырых частотностей, а на основе

нормализованных величин. Термины, входящие в зону J_2 , – редко встречающиеся слова и/или словосочетания. Их характеризует низкая частотность как в данном тексте, так и в других текстах.

В настоящей статье мы предложили классификацию лексикографических ресурсов, необходимых для поддержки систем автоматического анализа текста, на терминологические словари, терминологико-статистические словари, тезаурусы, онтологии, а также классификацию методов генерации словарей на статические и динамические; линейные и ступенчатые. Методы взвешивания терминов, используемые в процессе генерации словаря, мы предлагаем классифицировать на интертекстуальные и интратекстуальные.

Одной из проблем, которая возникает в процессе создания словарей, является проблема определения пороговых уровней, применяемых для отбора терминов в словарь. Наиболее типичное решение этой проблемы – применение среднеарифметических коэффициентов: в словарь включаются все термины, вес

которых превышает или равен среднеарифметическому значению. Другое решение – применение разработанной нами интерпретации закона Брэдфорда (Y-интерпретации). Эта интерпретация предоставляет адекватные критерии для выделения подмножества элементов в каком-либо множестве и может быть применена по отношению не только к лингвистическим данным, но и к другим предметным областям. С её помощью можно выделить, например, из множества регионов страны наиболее отсталые и, напротив, наиболее развитые регионы; из множества студентов (абитуриентов, учащихся, кандидатов на какую либо должность) – группы лучших и отстающих студентов. Основные требования к применению Y-интерпретации: наличие имени объекта и его количественной характеристики, причём количество объектов должно быть не меньше четырёх.

Существенная проблема при применении интертекстуальных методов взвешивания - обеспечение репрезентативности корпуса. Нами было предложено использовать национальные корпуса, поскольку при их создании соблюдается требование репрезентативности. Однако такие корпуса не могут включать тексты, относящиеся ко всем предметным областям. В Американском национальном корпусе научные тексты представлены текстами по биологии и медицине, соответственно вряд ли их можно использовать при анализе текста, относящегося к другой предметной области.

Решение этих проблем имеет значение для повышения эффективности функционирования всех систем автоматического анализа текста.

СПИСОК ЛИТЕРАТУРЫ

1. Kilgarriff A. BNC database and word frequency lists. - 1998. – URL: <http://www.kilgarriff.co.uk/bnc-readme.html>
2. About WordNet. = 2012. – URL: <http://wordnet.princeton.edu>
3. Pedersen T. WordNet::Similarity. -2008. – URL: <http://wn-similarity.sourceforge.net>
4. Зуёнок Т. Что должен знать о Яндексе каждый. Часть 1. - 2012. – URL: - <http://www.thepromo.ru/articleview.php?id=51>
5. Николаев В.Г. Узкоспециализированные поисковые интернет-системы. – 2010. - URL: – <http://www.in-internet.narod.ru/teor/poisk3.html>

6. Pretorius A. J Ontologies - introduction and overview. 2004. – URL: http://www.starlab.vub.ac.be/teaching/Ontologies_Intr_Overv.pdf
7. Добров Б.В., Лукашевич Н.В. Лингвистическая онтология по естественным наукам и технологиям для приложений в сфере информационного поиска // Web journal of formal, computational and cognitive linguistics. - 2006. - URL: <http://fccl.ksu.ru/fcclpap.htm>
8. Яцко В.А., Стариков М.С. Опыт разработки онтологии для автоматического анализа мнений пользователей о коммерческих продуктах // НТИ. Сер.2 - 2011. -№ 7. - с.9-14.
9. SurveysPaid – get paid taking surveys at home. – 2012. - URL: <http://www.surveypaid.com>
10. Яцко В.А Симметричное реферирование: теоретические основы и методика // НТИ. Сер.2. - 2002. - № 5. – С. 18-28.
11. Gore K. What less can we say? Computers have the answer. – 2006. - URL: http://www.slate.com/articles/news_and_politics/slates_10th_anniversary/2006/06/cogito_auto_sum.html
12. Salton G., Buckley C. Term-weighting approaches in automatic text retrieval. // Information Processing & Management. – 1988. – Vol.24, N 5. - P. 513–523.
13. Яцко В.А. Интерпретация закона Брэдфорда в терминах геометрической прогрессии // НТИ. Сер. 2 - 2012 - № 4. - С. 19-23.
14. Laurence Anthony's software. 2011. – URL: <http://www.antlab.sci.waseda.ac.jp/software.html>
15. Debole F., Sebastiani F. Supervised term weighting for automated text categorization // Proceedings of the 2003 ACM Symposium on Applied Computing (SAC). – Melbourne (FL, USA), 2003. – P. 784-788.

Материал поступил в редакцию 05.07.12.

Сведения об авторе

ЯЦКО Вячеслав Александрович – доктор филологических наук, профессор Хакасского государственного университета им. Н. Ф. Катанова, г. Абакан

E-mail: viatcheslav-yatsko@rambler.ru

Заимствования в современной украинской документоведческой терминосистеме

Исследуются источники и механизмы заимствования лексем украинской документоведческой терминологией. Делается вывод, что заимствование слов украинской терминосистемой документоведения началось на ранних исторических этапах. Этот процесс был обусловлен экстра- и интралингвальными факторами, а именно: развитием политических, экономических и культурных связей между народами; разной степенью достижений того или иного народа в определенной отрасли, необходимостью уточнения или детализации соответствующих понятий с помощью иностранных слов и т.д.

Термины документоведения заимствовались как непосредственно из языков, так и опосредованно – через язык-посредник. Источниками заимствований были греческий, латынь, английский, французский, немецкий, реже – голландский, итальянский, польский языки.

Ключевые слова: документоведческая терминология, заимствование, прямое заимствование, косвенное заимствование

Научно-техническое и общественное развитие обуславливают появление новых изобретений, открытий, а вместе с ними – понятий. Это объективный процесс, потому что формирование терминологических единиц требует учета их основных функций, в частности тех, которые призваны четко фиксировать научные понятия, а через понятие – передавать результаты познавательной деятельности человека в сжатых, лаконичных выражениях [1, с. 142]. В современном языкознании в связи с необходимостью упорядочения национального научно-технического языка все большее значение приобретает вопрос анализа целесообразности попадания в его лексический состав значительного количества языковых заимствований.

В научных работах последних лет иноязычные слова трактуют как нормальную функцию лингвистического образа жизни (В. М. Лейчик, Г. С. Онуфриенко), которая приобретает особую перспективность и вследствие этого заслуживает серьезного внимания исследователей. Заимствованные слова являются органической составной частью почти каждого современного языка: «Практически ни один природный язык при создании терминов не может обойтись собственными ресурсами, поэтому терминологии языков многих отраслей знания свои лексические средства заимствуют отовсюду, что делает их состав слишком пестрым, сочетая в себе национальные и интернациональные элементы» [2, с. 226].

В таких условиях очень важно найти «золотую середину» между немотивированными заимствован-

ными номинациями научных понятий и узконациональной замкнутостью в развитии терминологии. О сложности лавирования при употреблении заимствованной терминологии говорил в свое время И. Я. Франко: с одной стороны «ученый, излагая нам достижения науки, должен пользоваться языком, и то не каким-то абстрактным, а тем обычным, созданным исторически, привычным для нас», а с другой – должен заботиться о максимальной точности и однозначности термина: «чем более подробно, доказательна наука, тем сильнее ученый должен бороться с поэтикой» [3, с. 46].

Проблема языковых заимствований в последние два с половиной века занимает важное место в лингвистике. Она рассматривается в различных аспектах: И.И. Огиенко поднимает вопрос лексических заимствований из иностранных языков, исследуя историю их проникновения в украинский язык [4]; А.А. Потебня на материале немецкого и украинского языков доказал, что заимствования попадают в язык из языка-источника как непосредственно, так и через язык-посредник [5, с. 262]; Ю.А. Жлуктенко исследует механизмы заимствования, проблемы языковых контактов, в частности украинско-английских, и разграничивает понятия «речевой контакт» и «взаимодействие языков» [6, с. 12]; в исследованиях С.В. Семчинского рассматриваются пути обогащения лексики, изучаются теоретические аспекты взаимодействия и сближения языков на лексико-семантическом уровне, их морфологическая и семантическая структура [7]; источникам заимствований посвящены ра-

боты Т.К. Черторижской (1979), структуре терминологической лексики – А.В. Крыжановской (1987), Т.И. Панько (1994), Л.А. Симоненко (1987); из работ последних лет, посвященных вопросам проникновения и ассимиляции иноязычной лексики, можно отметить исследования Б.Н. Ажнюка (1998), Г.В. Зимовец (1997), В.П. Симонок (2002), И.Н. Кочан (2006). Исследователи стремятся выработать единое определение для понятия «заимствование», ведь ученые – представители разных направлений в языкознании – неединогласны во взглядах на механизм и классификацию заимствованных слов, а отсюда – и на природу этого явления. Так, Ю.А. Жлуктенко отрицает понятие «заимствование», которое, по его мнению, «в значительной мере является условным» [8, с. 42]; Л.П. Крысин под заимствованием понимает «процесс перемещения различных элементов из одного языка в другой» [9, с. 18]; С.В. Семчинский рассматривает заимствования как форму интерференции языков [7]. Учитывая различные взгляды лингвистов на проблему заимствований, в своем исследовании мы руководствуемся мнением, что терминологические заимствования являются одним из путей пополнения терминосистем разных языков, имея в виду, что такие лексические единицы характеризуются однозначностью, отсутствием эмоционально-экспрессивных коннотаций: если «слово родного языка должно еще стать термином», то слово иноязычное – «готовый термин; оно является знаком больше, чем слово родного языка» [10, с. 89].

В русском языкознании лингвистическому описанию потенциала и структурно-семантической специфики терминов и понятий документоведения в английском языке посвящены исследования И.Ю. Булгаковой. Она указывала, что «сейчас существует серьезная необходимость в систематизации свойств, признаков и характеристик англоязычных терминологических единиц документоведения, так как указанная терминология представляет лингвистический инструмент научного познания» [11, с. 4-5]. Согласимся, что проблема научной систематизации и классификации терминологии документоведения – весьма актуальна.

Процесс вхождения в украинскую терминологию документоведения заимствованных терминов еще не был предметом отдельного исследования, поэтому попытаемся проанализировать заимствования в системе украинской документоведческой терминологии в двух аспектах: социолингвистическом, при котором наблюдаются процессы вхождения заимствований в язык в связи с развитием истории и культуры народонаселения, и в лексико-семантическом, при котором исследуются отдельные группы заимствованных слов.

Социолингвистический аспект предполагает, в частности, исследование механизмов вхождения в национальный язык определенных групп заимствованной лексики в результате общественно-экономических и политических контактов Украины с другими государствами на разных исторических этапах. Вопрос контактов, несмотря на его якобы очевидную простоту, сложный, хотя бы вследствие разнообразия и многоаспектности этнических контактов украинцев, их предков с раз-

ными народами в различных географических направлениях – восток, запад, юг, север.

Вопрос хронологизации заимствований также остается в центре внимания исследователей, поскольку углубляет понимание многопланового в своих проявлениях процесса контактирования и взаимодействия языков. Не менее сложным для выяснения в каждом конкретном случае остается вопрос о путях проникновения заимствованных слов на украиноязычную почву, ведь каждое слово имеет свою, индивидуальную историю вхождения в язык-реципиент. Эта история не ограничивается лишь указанием на время заимствования. Важным остается установление путей заимствования, поскольку такие слова могли проникать в украинский язык непосредственно из языка-продуцента или через посредничество других языков. «Определение пути заимствования всегда оставляет место для дискуссий и уточнений, поскольку проявление новых источников может изменить оценки по наличию и характеру посредничества при заимствовании конкретной языковой единицы. Приходится идти от противного: если в возможных языках-посредниках отсутствуют следы соответствующего влияния ранее, чем в украинском, тогда можно утверждать о факте непосредственного заимствования» [12, с. 107].

Процесс заимствования иностранных терминов может происходить различными путями: непосредственно (прямые заимствования), через посредничество других языков (косвенные заимствования).

К прямым заимствованиям относим терминологические единицы, которые проникли непосредственно в документоведческую терминосистему из того или иного иностранного языка, например: укр. *декрет* ← лат. *decretum* – указ, постановление: один из видов нормативно-правовых актов государственной власти [13, с. 206]; укр. *статут* ← среднелат. *statutum* – устанавливаю, решаю: сборник основных принципов, положений о чем-либо [13, с. 643].

К косвенным терминологическим заимствованиям относятся слова, которые были заимствованы другими языками и уже из него вошли в терминосистему украинского языка. Так, в частности, в сфере документоведения употребляются такие термины: *архів* (от греч. *archaios* – древний → греч. *archeion* – административное учреждение, ведомство → лат. *archivum*) – 1) учреждение (или отдел в учреждении), которое сохраняет документальные материалы, 2) письма, рукописи, фотографии и т.д., касающиеся деятельности какой-либо организации или лица [13, с. 74]; *бібліометрія* – (от греч. *Biblion* – книга и *metron* – мера → англ. *Bibliometrics*) – научная дисциплина, которая изучает закономерности отражения в научной литературе процессов развития определенных отраслей знания, проблемы авторства и т.п. [13, с. 103]. К косвенным относятся также общенаучные термины типа: *метод*, *норматив* и т.д. Значительная часть этих терминов является интернационализмами. Под термином *лексические интернационализмы* понимаем «языковые единицы лексического уровня, употребляемые во многих (не менее трех)

неродственных языках, сохраняющие вместе с тем близость или общность своей семантики и фонетикоморфологического строения» [2, с. 160]. Сравним: укр. *документ*, англ. *document*, нем. *Dokument*, франц. *document*, испан. *documento*, литовск. *dokumentas*; укр. *реквизит*, датск. *rekvisitter*, голл. *rekwisieten*, нем. *Requisiten*. Распространение интернационализмов современные языковеды склонны оценивать как положительное явление, поскольку они являются «одинаково чужими и, одновременно, родными для любого языка» [14, с. 113].

Украинская документоведческая терминология (УДТ), исходные элементы которой зарождаются вместе с первыми попытками человека зафиксировать важную для него информацию, имеет давнюю историю. На разных этапах развития УДТ в ее составе начинали функционировать термины, заимствованные из других языков. Чужие лексемы пополняли терминосистему документоведения с неодинаковой интенсивностью в разные периоды ее становления, причем на каждом этапе отчетливо проявлялось влияние какого-то одного языка. Влияние каждого из языков определялось прежде экстралингвальными факторами: если на начальных этапах контакта с тем или иным языком наблюдается определенная хронологическая последовательность, то позже дифференциация происходит не по хронологии, а по сферам функционирования, представляющим этнические предпочтения в какой-либо области.

Лексика латинского и греческого языков стала пополнять научную терминологию многих европейских языков из-за того, что эти языки имели статус международных в течение достаточно длительного времени. Разумеется, греко-латинские заимствования широко употребляются в украинской научно-технической терминологии вообще и в документоведческой терминологии как в ее составляющей, в частности. Языками-продуцентами стали также французский, немецкий, голландский, итальянский, английский, испанский, турецкий, арабский, польский языки.

ЗАИМСТВОВАНИЯ ИЗ ГРЕЧЕСКОГО ЯЗЫКА

Одной из экстралингвистических причин распространения грецизмов было большое влияние византийской церкви на украинскую, которая находилась в канонической юрисдикции цареградского патриарха: «слова из греческого языка попадали в древнерусский язык двумя путями: посредством переводов древнегреческих произведений и вследствие контактов населения Киевской Руси с населением греческих колоний на Северном Причерноморье» [2, с. 160]. Приняв крещение от Византии и пригласив иерархов греческого происхождения, Киевская Русь «приняла» ряд слов из греческого языка. Одним из древнейших греческих заимствований было слово *библиотека*, которым в Греции называли сосуд для хранения книг [15, с. 186]. Как свидетельствует исследователь лексики русского языка XVII в. в материалах к курсу «История русского литературного языка» Е.М. Иссерлин, этот термин впер-

вые был использован в переводе Библии 1499 г. [16]. Среди документоведческих терминов греческое происхождение имеют названия дисциплин, отраслей научного знания (*библиография*, *библиология*), названия видов изданий, носителей информации (*апокриф*, *антология*, *энциклопедия*, *хрестоматия*, *каталог*, *диск*), структурно-композиционных частей документов (*пролог*, *эпilog*). В более поздние времена, особенно на протяжении XIX–XX вв. – в период технологического перевооружения всех отраслей научного знания – для создания новых номинаций стали активно использоваться компоненты древнегреческого происхождения. На современном этапе в УДТ функционирует большое количество терминов, в состав которых входят интернациональные терминологические элементы греческого происхождения, среди которых наиболее часто такие: *-граф*, *поли-*, *гидро-*, *микро-*, *моно-*, *макро-*, например: *библиограф*, *полівидання*, *гідрограф*, *мікроформа*, *макрозгорання* и многие другие. Количество таких терминологических единиц постоянно возрастает в разных терминологических системах, что и обуславливает актуальность исследований таких лексем [17].

Интернациональные терминологические элементы имеют определенную специализацию. Так, важнейшим универсальным показателем отношения к книге был терминологический элемент *библио-* (гр. *biblion* – книга). Соединяясь с иноязычными и автохтонными основами, он образует термины для обозначения научных дисциплин, направлений исследований (*библиометрия*, *библиотекознание*, *библиология*, *библиография*, *библиографознание*), специализированных процессов (*библиографование*), лиц по профессии, выполняемой ими деятельности (*библиотекар*, *библиограф*, *библиографознавец*, *библиофил*, *библиоман*), учреждений, организаций (*библиотека*, *библиотечная система*, *библиотека-филиал*). Эти существительные имеют свойство образовывать производные прилагательные: *библиотечный*, *библиографический*, *библиометрический* и под.

Как указывает М.В. Комова, «создание новых терминов с помощью терминологического элемента *библио-* продолжалось до 1960-х гг. В это время появились термины-синонимы *библиотека-автомобиль* и *библиобус* [15, с. 245]. На современном этапе есть тенденция заимствования определенных терминологических элементов без перевода. В исследуемой терминологической системе эта тенденция тоже имеет проявление, о чем свидетельствует лексема *буккроссинг* (от *book* – книга, *crossing* – переход, переправлять). *Буккроссинг* – это перемещение книги в основном между незнакомыми людьми. Идею буккроссинга выдвинул специалист по интернет-технологиям Рон Хорнбекер в мае 2001 г. Именно он предложил оставлять книги в общественных местах и отслеживать их перемещения через сайт в Интернете. Суть этого явления заключается в том, что, прочитав книгу, вы должны ее оставить в общественном месте для того, чтобы книгу нашел кто-то другой. Буккроссинг – это и хобби, и общественное движение, а также неофициальный

книжный клуб. Наибольшее распространение буккроссинг получил в Италии, где власти Флоренции решили подарить буккроссингу четыре тысячи книг. Буккроссинг – это движение людей, которые воплощают в жизнь идею создать мировую общественную библиотеку. Буккроссинг имеет несколько разновидностей, а если точнее, то несколько способов передачи (путешествия) книг. Например, букрей – так называют пересылку книг по почте от одного участника к следующему, книга практически никогда не возвращается к владельцу. Безопасные книжные полки или шкафы – еще одна разновидность буккроссинга, который заключается в том, что, например, в кафе выделяют место под книжную полку. На этой полке можно как оставлять книги, так и забирать. Букринг – это пересылка книги по почте, но, в отличие от букрея, в этом случае книга должна вернуться к первому владельцу.

В Украине движение буккроссинга пока развито очень слабо. Есть попытки его распространения в больших городах: Киеве, Харькове, Днепрпетровске.

Итак, большинство заимствований из греческого языка – это интернационализмы, усвоенные украинским языком. Они традиционно используются в УДТ, несмотря на попытки создать собственные дублетные названия.

ЗАИМСТВОВАНИЯ ИЗ ЛАТИНСКОГО ЯЗЫКА

Латинский язык, который в средневековье играл ведущую роль в общественной и научной жизни всей Европы, стал основным источником пополнения лексики многих национальных языков. Заимствованию слов из латыни способствовало «интенсивное изучение латинского языка, переводы латинских трудов по многим отраслям науки и техники, а также художе-

ственных произведений» [2, с. 161]. Латинское происхождение имеет немало древних названий государственно-административных документов: *декларация* (declaration) – свидетельство, заявление, постановление, определение, резюме, условие; *контroversия* (controversia) – заявление; *реляция* (relatio) – рапорт, донесение; *интpомісія* (intromissio) – акт об официальном введении в право на владение недвижимым имуществом, внесенным в актовую книгу; *метрика* (matricula) – сборник древних правительственных актов, официальный перечень, список, ведомость; *пакт* (pactum) – соглашение; *мембран*, *мембрана* (membrana) – документ, бланк, заготовленный для написания документа, пергаментная хартия; *трактат* (tractatus) – договор между государствами; *універсал* (universale) – правительственный распорядительный акт, адресованный представителям королевской или выборной сословной администрации. Ряд названий, которые стали использоваться позже, тоже латинского происхождения: названия видов изданий (*кодекс*, *календар*, *патент*, *проспект*, *факсиміле*), названия каталогизационных понятий (*індекс*, *рубрика*, *специфікація*, *фасцикула*), названия элементов оформления книги (*екслібрис*, *орнамент*).

И хотя латинский давно не употребляется самостоятельно как язык науки, он продолжает служить источником для создания новых терминов, которые приспособляются к нормам языка-реципиента графически и фонетически. В составе современной украинской документоведческой терминологии функционирует немало терминов, образованных с помощью префиксов и терминологических элементов латинского происхождения, среди которых выделим наиболее распространенные:

Префикс, термино-элемент	Украинский эквивалент	Значение	Примеры
ад-	до-, в-, при-	‘приближения, пребывание у чего-нибудь’	адаптація, адсорбція
де- дез-	з-, від-, зне-	‘отмена, устранение, удаление, нехватка чего-либо, уничтожение или искажение чего-либо’	декодування, декаталогізація, децентралізація, дезінформація
екс-	ви-, від-	‘выход за пределы чего-нибудь; из’	експлікація, експозиція, екслібрис
ин-	у-	‘проникновение в что-нибудь’	інноваційний
интра-	у-, усередині	‘внутренний’	інтразональний
инфра-	під-, нижче, поряд	‘подчинение’	інфраструктура
кон- ко-	спів- разом	‘объединение, общность, совместимость’	конструкція
контр-	проти-	‘противоположность места размещения предмета’	контртитул
квази-	ніби, майже	‘мнимый, ложный’	квазіпублікований
ре-	від- знову	‘повторное, обратное действие’; ‘противоположное действие или противодействие’	реорганізація, рекаталогізація, реконструкція, репродукція, рекласифікація
суб-	під-, нижче, до, біля	‘размещение под чем-нибудь’; ‘подчиненность’	суборенда

Отметим, что в некоторых случаях латинский язык становится посредником при заимствовании греческих слов (в таких случаях процесс заимствования имеет такой схематический вид: украинский язык ← латинский язык ← греческий язык), в некоторых – украинский язык заимствует из латинского слова через некоторые западноевропейские языки, воздействуя на фонетическое и грамматическое оформление слов в языке-реципиенте. Так, из французского языка были заимствованы такие лексемы латинского происхождения: *масив, мемуари, фронтиспис, резерв, фонд, експлікація, режим, ресурс, баланс*; из английского – *репортаж*; из немецкого – *раритет, формуляр, штамп*; из польского – *реєстр, лист, предмет, таблиця*. Иногда бывает трудно определить конкретный язык, из которого заимствован термин. Так, латинизм *фонд* был примерно одновременно заимствован французским (фран. *fond* < лат. *Fundus*) и английским (англ. *fund* < лат. *Fundus*) языками. Поэтому «Современный словарь иностранных слов» [13] указывает на романо-германское посредничество при заимствовании этой лексемы.

ЗАИМСТВОВАНИЯ ИЗ НЕМЕЦКОГО ЯЗЫКА

Заимствования из немецкого языка восходят еще ко временам великого переселения народов во время прохождения готов с севера на запад. Так, «Лексикон латинский» Е. Славинецкого свидетельствует о роли немецкого как языка-продуцента: *друкарня, друкар, палітурник*. Среди поздних заимствований – *абзац, кант, каптал, курсив, штемпель, штрих, фоліант, форзац, шрифт*.

Немецкий язык был посредником при заимствованиях из латинского и некоторых европейских языков: *раритет* (нем. ← лат.), *формат* (нем. ← фр. ← лат.), *формуляр* (нем. ← лат.), *цитата* (нем. ← лат.), *шмуцтитул* (нем. ← лат.), *плакат* (нем. ← фр.), *штамп* (нем. ← итал.).

ЗАИМСТВОВАНИЯ ИЗ ФРАНЦУЗСКОГО ЯЗЫКА

В состав исследуемой терминосистемы вошли заимствованные из французского языка названия видов изданий, документов (*афіша, буклет, бланк, блокнот, паспорт, альбом*), названия элементов печатных изданий (*макет, этикетка, віньєтка*), названия лиц по роду выполняемой работы, деятельности (*адресат, адресант, абонент, хронікер*) и другие термины: *репертуар, жанр, стаж, фасет, платформа*.

Термин *сценарій*, чтобы попасть в украинский язык, «пережил» ряд «переселений»: франц. *scenario* ← итал. *scenariò* ← лат. *scaena* – сцена.

ЗАИМСТВОВАНИЯ ИЗ АНГЛИЙСКОГО ЯЗЫКА

Сейчас нет, кажется, ни одной терминосистемы, в которой отсутствовали бы англицизмы. Первые из них появились в УДТ в конце XIX в. Среди древнейших англицизмов отметим такие: *пакет, карта, фильм*. Новую волну заимствований из английского языка наблюдаем в течение последних десятилетий в связи с резким, «скачкообразным» развитием техники, внедрением компьютерных технологий и закреп-

лением за английским языком статуса средства международного общения. В XX–XXI вв. заимствовались термины информационных технологий (*комп'ютер, принтер, процесор, байт, біт, дисплей, інтернет, файл*), названия документов (*бестселер, стандарт, слайд, препринт*).

ЗАИМСТВОВАНИЯ ИЗ ДРУГИХ ЯЗЫКОВ

В состав терминологии документоведения вошли в разное время термины из итальянского языка – *газета* (13, с. 147), *лібрето* (13, с. 412), *трафарет* (13, с. 684), из польского – *реєстр* (13, с. 586), из голландского – *лоція* (13, с. 422), из тюркского – *сургуч, ярлик*.

Отдельно следует отметить роль посредничества русского языка в процессе заимствования иноязычных терминов. Начиная со второй половины XVIII в., происходит «постепенное, но последовательное ограничение функций украинского литературного языка и вытеснение его из многих сфер общественной жизни» [18, с. 684]. В настоящее время вследствие увеличения влияния русского языка в терминосистеме различных отраслей знания начали интенсивно проникать русизмы (*гражданка, опечатка*). Этот процесс вызвал нивелирование выразительных национальных черт УДТ, активизацию в терминообразовании русских словообразовательных моделей и копирования. Среди терминов документоведения до недавнего времени встречалось немало лексических русизмов (*полоса, багаточисленний, згідно методики предметизації*) и терминов, образованных по русским словообразовательным моделям (отглагольные существительные с суффиксом -к(а): *підгонка, вичитка, розробка, відмітка*; названия исполнителей действия с суффиксом чик/-щик: *мірщик (обмірщик), креслярщик, газетчик, наборщик*).

Современный этап развития терминоведения характерен тем, что основным направлением работы в этой области является практический аспект совершенствования и кодификации терминологии. Процесс нормирования терминосистемы документоведения предполагает систематизацию, стандартизацию, кодификацию и унификацию – комплекс мероприятий, направленных на то, чтобы отраслевая терминология могла адекватно отражать систему специальных понятий, в том числе и номинированных заимствованными словами. Заимствование – это естественный процесс, и никакая терминология не может без него обойтись. Однако следует помнить, что чрезмерные заимствования из любого языка негативно влияют на терминообразование. Заимствования должны быть обоснованы и мотивированы. Выбирая между собственным и заимствованным терминам, следует отдавать предпочтение собственному, но учитывать, насколько термин точно передает обозначаемое понятие, соответствует нормам украинского литературного языка и традициям терминопотребления.

Итак, заимствование слов украинской терминосистемой документоведения началось на ранних исторических этапах. Этот процесс был обуслов-

лен екстра- и интралингвальними факторами, а именно: розвитком політичних, економічних і культурних зв'язів між народами; різною ступеню досягнень того чи іншого народу в одній галузі; необхідністю уточнення або деталізації відповідних понять з допомогою іноземних слів і т.д.

Терміни документознавства заїмалися як не-посередньо з мов, так і опосередковано – через мов-посередників. Джерелами заїмивань були грецький, латинь, англійський, французький, німецький; режі – голландський, італійський, польський мови.

СПИСОК ЛІТЕРАТУРИ

1. Ктитарова Н. Формування і розвиток металургійної термінології в Україні // Проблеми української термінології. Вісник Нац. ун-ту «Львівська політехніка». – № 453. – Львів, 2002. – С. 142 – 147.
2. Національні та інтернаціональні компоненти в сучасних терміносистемах / відп. ред. Л. О. Симоненко. – Київ : Наукова думка, 1993. – 237 с.
3. Франко І. Із секретів поетичної творчості // Твори в 50-ти тт. – Т. 31. – Київ : Наукова думка, 1981. – 595 с.
4. Огієнко І. І. Іноземні елементи в російській мові: Історія проникнення слів в російську мову. – Київ : Тип. Бондаренко і Гнездовського, 1915. – 136 с.
5. Потебня А. А. Из записок по русской грамматике: в 2 т. – М. : Учпедгиз, 1958.
6. Жлуктенко Ю. А. Языковые контакты. Проблемы интерлингвистики. – Киев : Изд-во Киевского ун-та, 1966. – 135 с.
7. Семчинський С. В. Семантична інтерференція мов. – Київ : Вища школа, 1974. – 256 с.
8. Жлуктенко Ю. А. Лингвистические аспекты двуязычия. – Київ : Высш. шк., 1974. – 176 с.
9. Крысин Л. П. Иноязычные слова в современном русском языке. – М. : Наука, 1968. – 208 с.
10. Кутина Л. Л. Языковые процессы, возникающие при становлении научных терминологических систем // Лингвистические проблемы научно-

технической терминологии / под ред. С. Г. Бархударова. – М. : Наука, 1970. – С. 82–101.

11. Булгакова И. Ю. Терминологические аспекты развития документоведения в английском языке с позиции социолингвистики : автореф. дисс. ... канд. филол. наук. / Омский государственный технический университет. – Омск, 2011. – 23 с.
12. Гриценко С. П. Лексичні латинізми в українських текстах кінця XVI – XVII століття : монографія. – Київ : Видавничо-поліграфічний центр «Київський університет», 2011. – 367 с.
13. Сучасний словник іноземних слів: близько 20 тис. слів і словосполучень / укл.: О. І. Скопченко, Т. В. Цимбалюк. – Київ : Довіра, 2006. – 789 с.
14. Д'яков А. С., Кияк Т. Р., Куделько З. Б. Основи терміноутворення: Семантичні та соціолінгвістичні аспекти. – Київ : Вид. дім «КМ Academia», 2000. – 218 с.
15. Комова М. Українська документознавча термінологія: шляхи творення та функційні особливості. – Львів : Вид-во Львівської політехніки, 2011. – 316 с.
16. Иссерлин Е. М. Лексика русского литературного языка XVII века: материалы к курсу «История русского литературного языка». – М., 1961. – 80 с.
17. Кочан І. М. Динаміка і кодифікація термінів з міжнародними компонентами у сучасній українській мові. – Львів : Вид. центр ЛНУ ім. І. Франка, 2004. – 519 с.
18. Українська мова: Енциклопедія. – 2-ге вид., випр. і доп. – Київ : Вид-во «Укр. енциклопедія» ім. М. П. Бажана, 2004. – 824 с.

Матеріал надійшов в редакцію 03.07.12.

Сведения об авторе

ТУР Оксана Николаевна - кандидат филологических наук, доцент кафедры украиноведения Кременчугского национального университета им. Михаила Остроградского (Украина)

E-mail: oktur@ukr.net

Национально и культурно обусловленная детерминированность локативной семантики в языковых картинах мира

Язык рассматривается в качестве составной части культуры. Выявляются некоторые особенности национально-ментального характера в использовании лексики с локативной семантикой, находящие отражение в языковой картине мира. Делается вывод о причинах данного явления.

Ключевые слова: пространство, локативность, культура, язык, лексика, семантика.

Язык – это национально специфический компонент культуры, знак принадлежности его носителей к определенному социуму. «План содержания языка (включая содержание лексических единиц и грамматикосинтаксические категории) варьируется от языка к языку, максимально соотносясь с национальной культурой. ...Во-первых, ... язык как идеальная, объективно существующая структура подчиняет себе, организует восприятие мира его носителями. А во-вторых, ... язык – система чистых значимостей – образует собственный мир, как бы наклеенный на мир действительный» [1, с. 75]. «Язык не существует вне культуры, т.е. вне социально унаследованной совокупности практических навыков и идей, характеризующих наш образ жизни», – отмечает Э. Сепир [2, с. 211].

На современном этапе развития науки слово должно рассматриваться не только как лексическая единица номинативного плана, но и как отражение национальной картины мира носителей языка. «Речь уже идёт ... о необходимости более глубокого и тщательного изучения мира (не языка, а мира) носителей языка, их культуры в широком этнографическом смысле слова, их образа жизни, национального характера, менталитета и т.п., потому что реальное употребление слов в речи, реальное речепроизводство в значительной степени определяется знанием социальной и культурной жизни говорящего на данном языке речевого коллектива», – считает С.Г. Тер-Минасова [3, с. 29].

Лексика, в том числе с локативным значением, часто обладает скрытой «культуроносностью». Каждое слово имеет значение, т.е. соотносится с реальным предметом или явлением окружающего нас мира. Но эти предметы, явления, а также представления о них неэквивалентны, т.е. неодинаковы, в разных культурах.

В связи с этим С.Г. Тер-Минасова приводит интересный пример: «...если рассмотреть те речевые ситуации, в которых *дом* и *house* совпадают по семантике, а значит, должны быть эквивалентны и легко переводимы, необходимо учитывать разницу культур на уровне если не реальных предметов, то представ-

лений и понятий о них. Иначе говоря, понятие, выражаемое словом *дом*, и то, что стоит за английским словом *house*, – это разные вещи, определяемые разными культурами.

Для того чтобы понять и, соответственно, правильно перевести английское предложение *That morning she had a headache and stayed upstairs*, нужно знать, что представляет собой английское слово *house*. Буквальным эквивалентом английских слов этого предложения будут русские слова: *в то утро она имела головную боль и осталась наверху*. Правильный перевод, передающий смысл предложения, – *В то утро у нее болела голова, и она не вышла к завтраку*.

Дело в том, что в традиционном английском доме наверху только спальни, а гостиная, столовая, кухня – на первом этаже. Поэтому понятия *upstairs* (вверх, поднявшись по лестнице) и *downstairs* (внизу, спустившись по лестнице) подразумевают образ жизни и устройство дома, т.е. все то, что обозначается словом *house* и что в ряде моментов существенно отличается от русского слова-понятия *дом*. И то, и другое – и *house*, и *дом* – складывалось веками под влиянием образа жизни, климата, географических условий и еще самых различных факторов» [3, с. 57].

Проведенный нами лингвистический эксперимент с использованием метода свободных дефиниций также доказывает различие представлений об одном и том же предмете у представителей различных культур. В ходе эксперимента было опрошено по 35 российских и иностранных респондентов, соответственно, на предмет толкования с функционально-семантических позиций слов *дача*, *ферма*, *таунхаус*, *пентхаус*. Цель – выявить различия в понимании локативов *на даче*, *на ферме*, *в таунхаусе*, *в пентхаусе* у представителей разных лингвокультур.

Все российские респонденты продемонстрировали одинаковое понимание лексемы *дача*, предложив для неё близкие дефиниции. Например: 1. *Загородное строение помимо основного городского жилья, чаще всего имеет небольшой участок земли. Используется в тёплое время года в качестве места отдыха и для*

выращивания овощей и фруктов. 2. Место, где люди могут жить, отдыхать, сажать цветы или обрабатывать огород. Это второй дом для хозяина. 3. Собственность владельца, небольшой участок за городом, на который человек имеет право, подтверждённое документом. Данный участок владелец может использовать по своему усмотрению, чаще всего на нём выращивают какие-либо культуры. 4. Место за пределами города, которое находится в частной собственности владельца. Функции: выращивание овощей и фруктов, отдых. 5. Загородный участок с жилым помещением. Служит для занятий сельским хозяйством, отдыха. Используется для личных нужд. И тому подобное.

Иностранные респонденты (англоязычные студенты 2-3 курсов из Индии и стран Африки) дали слову дача различные толкования, вплоть до абсурдных: 1. Старый маленький дом в деревне, где люди живут на каникулах (11 чел.). 2. Дом, где можно отдохнуть (4 чел.). 3. Место в деревне (1 чел.). 4. Дом, который находится на пляже (1 чел.). 5. Дом в лесу (1 чел.). 6. Крыло дома (1 чел.). 7. Место, где люди живут летом, когда возможно вырастить фрукты или овощи; на даче есть красивый дом и сад рядом с ним (2 чел.). 8. Место, где родились дедушка и бабушка; мы туда ездим отдыхать (1 чел.). 9. Сад (1 чел.). 10. Место, где родились мама и папа (1 чел.). 11. В нашей стране этого нет, потому что природа везде одинаковая (1 чел.). 12. Страна (7 чел.). 13. Красивое место в Индии, где люди живут, учатся и работают (1 чел.). Два человека не смогли дать определение слову дача.

Представители русской лингвокультуры проявили единодушие, по всей видимости, благодаря распространённости данной реалии в России конца XX в. – начала XXI в. и устойчивости представлений о роли этого объекта действительности в жизни соотечественников. Инокультурные респонденты разошлись во мнениях, и, даже если дефиниции близки друг к другу (см. дефиниции 1, 2, 3), они выявляют отсутствие понятийной эквивалентности русского слова дача и английских *summer cottage*, *bungalow*. Такое явление, как владение загородным участком земли с домом, во-первых, не имеет, по-видимому, широкого распространения на родине иностранцев (об этом свидетельствует отказ объяснить слово, значительное количество абсурдных толкований: см. дефиниции 12,13), во-вторых, даже в случае наличия этот объект используется, как правило, исключительно для отдыха, в отличие от российской дачи новейшего времени.

Подобная ситуация сложилась при толковании лексемы ферма.

Российские респонденты дали сходные толкования этому слову. Ср.: 1. Место, используемое для разведения домашнего скота, чаще всего находится в пригороде. 2. Место для выращивания скота, получения от него различных продуктов. 3. Большое хозяйство, где разводят животных разных пород, выращивают овощи и фрукты. 4. Дом с плантацией, загоном для скота. Там получают различные продукты: овощи, мясо, молоко. Находится за городом, в селах, на полях. 5. Место для разведения животных. 6. Место выращивания культурных растений, разведения жи-

вотных ради прибыли. 7. Место, где человек, занимающийся сельским хозяйством, выращивает овощи, фрукты или разводит животных на продажу. 8. Место, где разводят крупный рогатый скот. 9. Участок земли с домом, загонами для домашних животных и, возможно, огородом. Используется для постоянного проживания фермера и его семьи. И другие подобные, свидетельствующие об устойчивости представления о данном объекте в русском языковом сознании.

Студенты из Индии и стран Африки разошлись во мнениях по поводу реалии, обозначаемой словом ферма: 1. Сад (15 чел.). 2. Красивое место для гостей (3 чел.). 3. Дом далеко от города (1 чел.). 4. Место в Индии в деревне (4 чел.). 5. Место, где мы отдыхаем (1 чел.). 6. Место, где люди работают (5 чел.). 7. Место, где выращивают урожай (6 чел.).

Таким образом, иностранные респонденты, в отличие от российских, не единодушны в своих представлениях о ферме, из чего можно сделать вывод о нераспространённости данного объекта в Индии и африканских странах.

Что касается слов *таунхаус* и *пентхаус*, то при их толковании и российские, и иностранные респонденты продемонстрировали в большинстве своем далекое от реальности или весьма неточное представление о предметах действительности, ими обозначаемых.

Ср. дефиниции российских студентов для слова *таунхаус*: 1. Место для проживания в городе (1 чел.). 2. Дом в большом городе, служит для проживания людей (2 чел.). 3. Частный дом в центре города (1 чел.). 4. Дом в городе или просторная квартира (1 чел.). 5. То же самое, что *пентхаус*, но в черте города (1 чел.). 6. Дом в городе для отдыха (1 чел.). 7. Место отдыха, находящееся рядом с местом работы (1 чел.). 8. Загородный домик (1 чел.). 9. То же, что *пентхаус* (1 чел.). 10. Дом в сельской местности (1 чел.). 11. Огромные строения для проживания людей, используются для нахождения в них гостиниц; может быть, там проживают главы государств при выезде в чужую страну (1 чел.). 12. Городской дом (4 чел.). 13. Место жительства горожан (1 чел.). 14. Дом, расположенный в зеленой зоне города (1 чел.). 15. Место на крыше большого многоквартирного дома (1 чел.). 16. Городская квартира (1 чел.). 17. Частный дом в городе (1 чел.). 18. Человек не смог растолковать это слово.

Ср. дефиниции иностранных студентов для слова *таунхаус*: 1. Дом, где живет много людей (2 чел.). 2. Дом, где живет много людей (1 чел.). 3. Маленький дом (2 чел.). 4. Высокий большой дом, в 2 или 3 этажа (3 чел.). 5. Дом в деревне (3 чел.). 6. Дом в городе (2 чел.). 7. Как дача: место, где разводят животных, делают шашлык (1 чел.). 8. Ферма (1 чел.). 9. Большой дом, в котором живут богатые люди (2 чел.). 10. Большой дом, где отдыхают (2 чел.). 11. Дом в районе, где находятся банки (2 чел.). 12. Место для гостей (1 чел.). 13. Верхняя часть в доме (3 чел.). 14. Место в городе, где люди работают (1 чел.). 15. Место, где отмечают праздники (1 чел.). 16. Офис (3 чел.). 17. Большой дом в городе, где живет одна семья (2 чел.). 18. Место в столице (1 чел.).

чел.). 19. Место для важного дела (1 чел.). 20. Место для проведения операций (1 чел.).

Пентхаус в анкетах российских респондентов также представлен множеством различных вариантов, в большинстве своем далеких от истины: 1. Что-то типа дачи (1 чел.). 2. Облагороженная квартира или дом для проживания богатых людей (1 чел.). 3. Помещение, занимающее целый этаж, используется для отдыха, проживания (1 чел.). 4. Вип-номер в отеле (2 чел.). 5. Большой дом или квартира (1 чел.). 6. Квартира в многоэтажке (1 чел.). 7. Дом с улучшенной планировкой (1 чел.). 8. Здание нового поколения (1 чел.). 9. Чердак или элитное жилье на самом последнем этаже, используется для жилья, отдыха (1 чел.). 10. Высокое здание, где на самом верху находится помещение с двумя этажами, т.е. жилая квартира (1 чел.). 11. Жилье (1 чел.). 12. Дом с комфортными условиями проживания (1 чел.). 13. Место для отдыха в пригородном доме (1 чел.). 14. Здание для общественных целей (1 чел.). 15. Большой дом европейского типа (1 чел.). 16. Дом большой площади с многочисленными комнатами (1 чел.). 17. Большой жилой дом за чертой города, принадлежащий состоятельным людям (1 чел.). 18. Загородный дом, в котором люди отдыхают в свободное время (1 чел.). 19. Английский вариант дачи (1 чел.). 20. Верхний «люксовый» этаж в домах, как правило, с удобствами частного дома (бассейн, палисадник) (1 чел.). 21. Вид квартир в современных новостройках с какими-либо преимуществами и отличиями от обычной квартиры (1 чел.). 22. Престижное жилье (1 чел.). 23. Дорогое большое строение старинного типа, используемое для проживания обеспеченных людей (1 чел.). 24. Место, где живут обеспеченные люди (1 чел.). 25. Место проживания высокой комфортности (1 чел.). 26. Жилая площадь с хорошим расположением (1 чел.). 27. Загородный дом, дача (1 чел.). 28. Квартира, занимающая несколько этажей, большая по размеру, красиво меблированная и хорошо отремонтированная (1 чел.). 6 человек не смогли дать определение этому слову.

Иностранные респонденты предложили также большое количество дефиниций к слову *пентхаус*: 1. Маленькое здание (1 чел.). 2. Высокое здание для работы (6 чел.). 3. Квартира (1 чел.). 4. Место для каникул (1 чел.). 5. Место для отдыха (5 чел.). 6. Место для занятий искусством (1 чел.). 7. Дом рядом с морем (1 чел.). 8. Большой старый дом, который построен 20 лет назад (2 чел.). 9. Место в городе, где люди работают (3 чел.). 10. Дом в городе, где есть оборудование (2 чел.). 11. Лес, где живут животные (1 чел.). 12. Большой дом (2 чел.). 13. Дом для гостей (3 чел.). 14. Сад (2 чел.). 4 человека отказались дать толкования слову *пентхаус*.

Таким образом, о лексемах *таунхаус* и *пентхаус* можно сделать следующие выводы: реалии, ими обозначаемые, не нашли четкого отражения в языковом сознании как российских, так и иностранных респондентов. Дефиниции к этим словам, составленные опрошенными, в подавляющем большинстве случаев далеки от истинного положения вещей. Ср. дефиниции из Википедии: *Пентхаус* (англ. penthouse) – особняк на крыше небоскрёба или отделённая пло-

щадь на верхнем этаже здания. Подобная структура может содержать верхнюю часть шахты лифта, оборудование для кондиционирования воздуха или лестницы, которые ведут на крышу. Она также может использоваться для проживания или как подсобное помещение. Обычно пентхаус находится сзади вертикальной лицевой стороны здания, таким образом предусмотрены открытые площадки или террасы с одной или более сторон. Однако в современной практике архитекторы и агенты по аренде квартир и домов относят термин «пентхаус» к верхнему этажу любого здания, независимо от расположения по отношению к фасаду. ... в настоящее время слово «пентхаус» обозначает роскошные и дорогие апартаменты, откуда открывается панорамный вид... [4]. *Таунхаус* (англ. townhouse) – комплекс малоэтажных комфортабельных домов, совмещённых друг с другом боковыми стенами. Каждый из таких домов имеет свой вход, иногда гараж и небольшой палисадник. Существует русский термин для обозначения этого типа жилья — блокированный дом [5].

Нераспространенность этих реалий на родине опрошенных привела к отсутствию правильного представления о данных объектах в их языковом сознании (табл. 1 и 2).

Таблица 1

Уровень наличия/отсутствия каких-либо представлений о реалиях: дача, ферма, таунхаус, пентхаус

Лексемы	Количество толкований, %	
	Российские респонденты	Иностранные респонденты
Дача	100%	94%
Ферма	100%	100%
Таунхаус	54%	100%
Пентхаус	83%	88%

Таблица 2

Уровень адекватности представлений о реалиях: дача, ферма, таунхаус, пентхаус

Лексемы	Количество правильных толкований, %		Количество неправильных толкований, %	
	Российские респонденты	Иностранные респонденты	Российские респонденты	Иностранные респонденты
Дача	100%	51%	—	49%
Ферма	100 %	31%	—	69%
Таунхаус	26 %	11%	74%	89%
Пентхаус	6%	—	94%	100%

Таким образом, локативы *на даче, на ферме, в таунхаусе, в пентхаусе* по-разному понимаются представителями различных лингвокультур, и расхождение во мнениях в подавляющем большинстве случаев довольно велико. Степень адекватности восприятия, т.е. правильного представления об объекте, обозначенном каждым

из данных локативов, напрямую зависит от его распространенности / нераспространенности в родной стране респондентов. Налицо факт национально и культурно обусловленной детерминированности локативной семантики в различных языковых картинах мира.

СПИСОК ЛИТЕРАТУРЫ

1. Антипов Г.А., Донских О.А., Марковина И.Ю., Сорокин Ю.А. Текст как явление культуры. – Новосибирск, 1989.
2. Сепир Э. Коммуникация // Избранные труды по языкознанию и культурологии. – М., 1993.
3. Тер-Минасова С.Г. Язык и межкультурная коммуникация: учебное пособие. – М., 2000.

4. Википедия. – URL: <http://ru.wikipedia.org/wiki/%CF%E5%ED%F2%F5%E0%F3%F1>
5. Википедия. – URL: <http://ru.wikipedia.org/wiki/%D2%E0%F3%ED%F5%E0%F3%F1>

Материал поступил в редакцию 05.07.12.

Сведения об авторе

ФЕДОСЕЕВА Лариса Николаевна - кандидат филологических наук, старший преподаватель кафедры латинского и русского языков Рязанского государственного медицинского университета им. И.П. Павлова

E-mail: ln-fedoseewa@yandex.ru

Информационная специфика дефиниций

Обосновывается необходимость адекватного описания юридических терминов. Рассматриваются основные требования к терминологическим дефинициям на материале французского, английского и русского языков.

Ключевые слова: термин, дефиниция, право, научный стиль, референция,

В последнее время намечается особое внимание к проблеме взаимосвязи лингвистики и правоведения. Научные исследования по юриспруденции начинаются с обращения к значению основных терминов и их этимологии. Право фиксируется в форме словесных знаков. Связь между правом и словом отразилась, прежде всего, в терминологии, которая стремится к точности передачи содержания. Научный стиль, стиль правовых определений реализуется прежде всего в письменной форме речи. Полагаем, что обращение к языку закона, к дефинициям юриспруденции является весьма актуальным: юридическая терминология считается одной из наиболее значимых отраслевых терминосистем. Правовые понятия и нормы выражаются только посредством языка. Следует заметить, что терминология права широко используется не только в профессиональной среде, но и в других сферах языковой коммуникации.

В центре нашего внимания находятся дефиниции основных терминов правоведения. В соответствии с определением лингвистического энциклопедического словаря, «термин (от лат. *Terminus* - предел, граница) – слово или словосочетание, обозначающее понятие специальной области знания или деятельности» [1, с. 508]. К термину предъявляются такие требования, как однозначность, определенность, ясность, непротиворечивость и т.д. Термин – слово или словосочетание естественного языка, в терминологиях происходят те же лексико-семантические процессы, что и в лексике общелитературного языка. Это полисемия, омонимия, синонимия, что вызывает определенные проблемы. А.А. Реформатский писал: «Один и тот же термин может входить в разные терминологии данного языка, что представляет собой межнаучную терминологическую омонимию, например: реакция 1) в химии, 2) в физиологии 3) в политике; редукция 1) в философии, 2) в юриспруденции, 3) в фонетике» [2, с. 62]. Среди терминов много заимствованных слов, например из французского языка: *арбитраж, унитаризм, федерация*. Важно отметить, что французский язык стал проводником латинских заимствований в английском языке, например: *justice, accuse, cause, jury, prison*. Поскольку многие термины являются заимствованиями, то при переходе слова из одного языка в другой его форма и значение могут изменяться. Интернационализмы могут оказаться тем, что называют ложными друзьями переводчика. Так, французское *société anonyme* переводится как «акционерное общество». Многочисленны

примеры заимствований из английского языка: *импичмент, киллер, консалтинг, ракет, секьюрити*.

«Юридические термины – элемент юридической техники, словесные обозначения государственно-правовых понятий, с помощью которых выражается и закрепляется содержание нормативно-правовых предписаний государства» [3]. Термин может быть выражен одним словом или словосочетанием. Чаще всего термин – это имя существительное или именное словосочетание. Имена – это тот класс слов, который предназначен для выполнения референциальной функции языка. Все определения референции представляют ее как функцию, реализуя которую языковые знаки отсылают к экстралингвистической действительности.

Особый характер имен существительных в плане выполнения референциальной функции языка отмечается уже Дж. Миллем. Им проводится первое фундаментальное деление имен на общие и единичные. Рассуждая по поводу семантического различия между общими и единичными терминами, О. Куайн отмечает, что «единичный терм указывает или имеет целью указать только на один объект, сколь бы он ни был сложен или расплывчат, в то время как общий терм истинен для каждого или для каждой группы из любого количества объектов» [4, с.59]. Эта дихотомия может рассматриваться как основополагающая для разработки теории референции. Возможности языковых выражений обозначать внелингвистическую реальность, т.е. обладать референцией, неоднородны. Эта неоднородность обусловлена разнообразием объектов, свойств и отношений, которые они называют.

При анализе корпуса французских и английских юридических терминов можно отметить, что большую часть представляют абстрактные имена, в которых преобладают сигнификативные компоненты. Так, в английском языке это *abatement, abduction, hearing*, представляющие собой отглагольные существительные. Называя процессы, действия, состояния в сочетании с дефинициями они не имеют конкретной референции, но обладают способностью характеристики конкретных ситуаций. Их особенностью является сходство с соответствующими глаголами и прилагательными, которое объясняется общностью концептов, обозначаемых ими; это, однако, не исключает их различия, которое проявляется в особенностях грамматической категоризации.

Выделяется небольшая группа имен, имеющих метаязыковой характер: *fact, case, thing* etc. Во фран-

пузском языке соответствующее имя *fait (m)*, часто входит в состав словосочетаний: *fait juridique, faits justificatifs, faits du process*.

В состав правовой терминологии включены также и конкретные имена классов, обладающие такими семантическими признаками, как конкретность, одушевленность, исчисляемость, лицо. В последнем случае это обозначение лица по его отношению к различного рода деятельности – *avocat (m), demandeur (m), notaire (m)* (франц.).

Термин стремится к однозначности, т. е. он должен быть своего рода «жестким десигнатором», который указывает на один и тот же объект в различных возможных ситуациях (возможных мирах). Реальность, которую отражает язык, может быть действительной, вымышленной или концептуальной. Референтность именной группы зависит и от типа высказывания. Существует логическое разделение высказываний на аналитические и синтетические. Особенностью высказываний первого типа является то, что они истинны в любом из возможных миров, для любого говорящего. Таковы обобщающие высказывания. Понятийную базу высказывания составляют концепты, обозначающие объекты, свойства, действия. Они могут быть представлены как субстанция, квалификация, действие, например, «право», «правовой», «управлять». Термины указывают на предмет в его основных, а не случайных свойствах

Термины существуют в рамках определённой научной системы, они входят также в конкретную лексическую систему языка. В терминологии определённой науки и техники за термином закреплёно точное логическое определение – дефиниция. Определение – это «формулировка, раскрывающая содержание, сущность чего-л., характеризующая основные черты чего-л.» [5]. Его необходимость обусловлена также тем, что он часто сохраняет свою многозначность [6]. Русский термин «определение» говорит о некотором ограничении, которое, очевидно, является необходимостью при исследовании.

Дефиниции включают не все существенные признаки предметов, а лишь те из них, которые необходимы для того, чтобы отличить определяемый предмет от других, подобных ему предметов. Очевидно, что дефиниция должна обладать таким свойством, как универсальность.

Полагаем, что сочетание термина и его дефиниции представляет собой предложение тождества. Предложения сигнификативного тождества характерны для научных дефиниций, например:

Accused - The person that is charged with a crime and has to go to criminal court. -

Обвиняемый – Лицо, которому предъявлено обвинение в совершении преступления и которое должно явиться в суд по криминальным делам [7].

При выработке дефиниции термина в работах по терминологии исходят из того, что дефиниция должна содержать признаки, каждый из которых необходим, а все вместе достаточны, чтобы подчеркнуть родство одного понятия с другими, с одной стороны, и выявить его специфику, с другой.

Проблема юридических дефиниций является одной из наиболее актуальных и сложных в современной лингвистике. В настоящей статье мы стремились показать многоаспектность разрабатываемой нами проблематики и неоднозначность возможных решений.

СПИСОК ЛИТЕРАТУРЫ

1. Лингвистический энциклопедический словарь / под ред. В.И. Ярцевой. – М.: Большая российская энциклопедия, 1990. – 709 с.
2. Реформатский А.А. Введение в языковедение. – М.: Аспект Пресс, 1996. – 536 с.
3. Большой юридический словарь. – URL: <http://www.info-law.ru/dic/1/symb> (visited 30 May 2012).
4. Куайн У.В.О. Слово и объект // Новое в зарубежной лингвистике. – М.: Прогресс, 1986. – Вып. 18. – С. 5-98.
5. Словарь русского языка: в 4-х т. / под ред. А.П. Евгеньевой. – 3-е изд., испр. и доп. – М.: Русский язык, 1988. – Т. 2. – 356 с.
6. Хайруллин В.И. Перевод неоднозначных терминов // Лингвострановедение: методы анализа, технология обучения. – Ч. 1. – М.: МГИМО (У), 2011. – с. 156-160.
7. English/Russian Legal Glossary. Юридический глоссарий. Translated by Yana Berrier, Esq Superior Court of California, County of Sacramento, 2005. – URL: www.saccourt.com (visited 11 May 2012).

Материал поступил в редакцию 13.06.12.

Сведения об авторе

ВАСИЛЬЧЕНКО Кирилл Алексеевич - аспирант кафедры делового иностранного языка и перевода Института права Башкирского государственного университета, г. Уфа

E-mail: kyril_korol@mail.ru