

ПРОБЛЕМЫ ИДЕОГРАФИЧЕСКОЙ РЕПРЕЗЕНТАЦИИ СЕМАНТИКИ СВОБОДНЫХ ВЫСКАЗЫВАНИЙ ТАКСОНОМИЧЕСКОГО ТИПА

© 2013 А.Н.Краснов¹, А.П.Журавлёв², Е.А.Слоева¹

¹Самарский государственный медицинский университет

²Самарский государственный технический университет

Статья поступила в редакцию 26.03.2013

В данной статье проводится анализ, выявляющий некоторые проблемы, возникающие при репрезентации семантической структуры предметной области при помощи идеографического метода, разработанного Ю.Н.Карауловым. Статья показывает, что для полного и адекватного отображения семантической классической идеографический метод Караулова должен быть усовершенствован.

Ключевые слова: метаязык, лексико-семантическая группа, предметная область.

Введение. Проблема идеографической репрезентации семантики свободных высказываний таксономического типа в языкознании связана с задачами компьютерного распознавания предложений и небольших текстов, относящихся к конкретным предметным областям. Разработка данной проблемы привела к появлению работ, посвящённых как языковому моделированию, так и проблеме разработки общей методики такого моделирования¹.

Известно, что необходимыми условиями для адекватного протекания процесса коммуникации является сходство языковых картин мира коммуникантов². Иными словами, в процессе кодирования и декодирования посылаемых и получаемых сообщений коммуниканты должны опираться на сходные системы значений. Эти системы значений Н.И.Жинкин обозначил как «универсальный предметный код» (УПК)³.

Проблематика УПК в особенности актуальна для коммуникации в профессиональных облас-

тях, где функционируют устойчивые терминосистемы и определяемые научной парадигмой языковые картины предметных областей. Действительно, специфичность профессионального знания вкупе с однозначной формулировкой его терминов требуют наличия единой системы значений, поскольку эти значения должны одинаково восприниматься всеми, кто имеет отношение к данной предметной области. В свете этого весьма актуальным направлением представляется создание единой семантической модели конкретной предметной области. Модель такого рода должна строиться на основе лексической базы и описывать предметную область с точки зрения её структуры – виды составляющих её элементов и объединяющих их связей.

Предложения и тексты таксономического типа (все «А» делятся на «а₁», «а₂», ... «а_n»), имеют достаточно стандартную семантическую конфигурацию и широко распространены в различных науках. Они являются удобным языковым материалом для наглядной репрезентации их семантики. Для этих целей в лингвистике используется идеографический метод Ю.Н.Караулова. В своём классическом виде данный метод представляет семантику предметной области в форме ориентированного плоского древовидного графа. Вместе с тем, развитие современной лингвистики и смежных наук, опирающихся на её разработки, выявило некоторые слабые стороны классического идеографического метода.

Во-первых, на практике широко распространено деление одного понятия одновременно на различные группы (т.е. по нескольким критериям). Например, понятие «человек / люди» можно делить по полу, возрасту, образовательному цензу, национальности и т.д. При этом появляются антонимические группы, а сам граф перестаёт быть плоским и превращается в квазипространственный.

¹Краснов Александр Никодимович, кандидат медицинских наук, профессор, заведующий кафедрой педагогики, психологии и психолингвистики.

E-mail: ankrasnov@mail.ru

Журавлёв Александр Павлович, ассистент кафедры иностранных языков, ст.лаборант лаборатории электронных учебных пособий кафедры педагогики, психологии и психолингвистики.

E-mail: palych32@rambler.ru

Слоева Елена Анатольевна, кандидат филологических наук, доцент кафедры педагогики, психологии и психолингвистики. E-mail: Elena-sloeva@efindex.ru

¹ Караулов Ю.Н. Общая и русская идеография. – М.: 1976; *Он же.* Лингвистическое конструирование и тезаурус литературного языка. – М.: 1981; *Кубрякова Е.С.* Язык и знание. – М.: 2004; *Лосев А.Ф.* Введение в общую теорию языковых моделей. – М.: 1968; *Савицкий В.М.* Английская фразеология: проблемы моделирования. – Самара: 1993.

² Знаков В.В. Понимание в познании и общении. – Самара: 2000. – С. 19.

³ Жинкин Н.И. Речь как проводник информации. – М.: 1982; *Он же.* Язык – речь – творчество. М.: 1998.

Во-вторых, появление критериев, по которым выделяются группы, с точки зрения языкознания означает появление на графе мета-языковых структур, не предусмотренных в классическом идеографическом методе.

В-третьих, различные значения, входящие по различным критериям в разные группы, могут обозначаться одними и теми же словами. Например: *открытая* и *закрытая* дверь (по положению в пространстве); *открытый* и *закрытый* человек (по характеру); *открытая* и *закрытая* научная тема (по уровню секретности) и т.д. Это порождает полисемию слов и требует для распознавания их значений опоры на лексическое окружение. Но в самом графе это требует не объединения различных значений, а наоборот, их чёткого различения.

В-четвёртых, одно и то же значение может одновременно входить в две или более семантических группы. Например, элемент *костный мозг* будет одновременно и частью кости, и частью кроветворной, и частью иммунной систем человека. При этом граф, отображающий такие семантические связи, превращается из плоского или даже квазипространственного в сетевую структуру. Это было показано нами ранее на примере анализа анатомии таза человека⁴.

Наконец, в-пятых, семантика выделенных групп может иметь как континуальный (непрерывный), так и дискретный (чётко разделённый) характер. Примером первого случая может быть лексика деления человека по возрасту (от младенца до старика) или силы ветра на шкале Бофорта (от штиля до урагана). В этом случае мы сталкиваемся не просто с группами, а со шкалами. В этих случаях деление шкал носит достаточно условный характер, а семантика становится нечёткой. Понятно, что и идеографическое отображение таких шкал становится также достаточно условным. Во втором случае семантика слов чётко отделена друг от друга (например, названия драгоценных камней – алмаз, сапфир, рубин и т.п.). С точки зрения идеографии только такая семантика может быть репрезентирована классическим древовидным графом Ю.Н.Караулова.

Таким образом, на сегодняшний день в языкознании возникла необходимость в коррекции классического идеографического метода. Актуальность данной задачи стимулируется развитием компьютерной лингвистики и систем автоматического распознавания естественного языка. Если в процессе коммуникации в режиме «че-

ловек-человек» процессы кодирования и декодирования сообщений носят естественный характер и едва ли осознаются коммуникантами, то в режиме «человек-машина» необходимо делать поправку на особенности второго коммуниканта – компьютера. Это означает, что упомянутые выше когнитивные процессы необходимо описать и алгоритмизировать, а саму языковую картину предметной области – переработать в соответствующую базу знаний.

В программировании одним из методов описания структуры баз знаний является теория графов, что и делает сходными эти две задачи: репрезентации семантики свободных высказываний с помощью идеографического метода Ю.Н.Караулова и описания структуры компьютерной базы знаний с помощью теории графов. Но поскольку теория графов носит более развитой характер, чем идеографический метод Ю.Н.Караулова, то перед нами встаёт вопрос: что для лингвистики необходимо заимствовать из теории графов, чтобы дополнить и развить классический идеографический метод и сделать его более адекватным для задач языкового описания семантики предметных областей.

Терминология. *Граф* – базовое понятие; включает множество вершин и множество рёбер, являющееся подмножеством декартова квадрата множества вершин (т.е. каждое ребро соединяет ровно две вершины). *Вершина графа* – базовое понятие: точка, где могут сходиться / выходить рёбра дуги графа. *Дерево (древовидный граф)* – связный граф, не содержащий циклов. *Корень графа* – выбранная вершина дерева. *Лес* – ориентированные древовидные графы без циклов. Компонентами связности леса являются деревья. *Лист дерева* – вершина дерева с единственным ребром или входящей дугой. Обычно это нижние (терминальные) вершины на древовидном графе. *Ребро (дуга) графа* – базовое понятие: способ графического отображения смысловой связи двух соседних вершин.

Цели и задачи. Целью нашего исследования стал сравнительный анализ идеографического метода Ю.Н.Караулова и теории графов, используемых для репрезентации семантики свободных высказываний таксономического типа, относящихся к заданной предметной области.

Задачами исследования стали: а) анализ таксономических высказываний с точки зрения соотношения в них элементов метаязыка и языка-объекта; б) семантический анализ структуры составного термина, относящегося к данной предметной области; в) определение возможностей для репрезентации выявленной семантики с помощью идеографического метода и выявление возникающих при этом проблем.

⁴ *Краснов А.Н.* Психолого-педагогические основы технологии открытого тестирования. – Самара: 2005; *Краснов А.Н., Словца Е.А.* Языковая репрезентация концептуальной модели анатомического объекта. LAP LAMBERT Academic Publishing GmbH & Co., 2011.

Материал и метод. Материалом нашего исследования являются 343 ответа студентов 5-го курса лечебного факультета на вопрос о классификации повреждений таза, а также 17 учебников и научных монографий, относящихся к данной предметной области (травматология). В общей сложности объём материала составил 360 высказываний таксономического типа.

В ходе исследования использованы следующие методы: 1) метод экспертных оценок имеющегося речевого материала; 2) метод системного анализа; 3) тезаурусный метод; 4) метод когнитивной графики (графическое представление семантики предметной области с помощью идеографического метода и теории графов); 5) сравнительный метод; 6) статистический метод.

Результаты работы. 1. Анализ таксономических высказываний с точки зрения соотношения в них элементов метаязыка и языка-объекта. Для решения первой и третьей задач рассмотрим несколько высказываний студентов с точки зрения соотношения в них элементов метаязыка и языка-объекта, а также с точки зрения возможности их идеографической репрезентации: А) *Все переломы таза делятся*

на открытые, закрытые, со смещением, без смещения, с повреждением внутренних органов, без повреждения внутренних органов. Б) *Все переломы таза делятся на: 1. Открытые, закрытые; 2. Со смещением и без смещения; 3. С повреждением внутренних органов, без повреждения внутренних органов.* В) *Все переломы таза делятся: 1) по наличию раны; 2) по наличию смещения; 3) по локализации повреждений.*

В первом предложении основная семантика представлена семантически однородными признаковыми элементами. Они являются либо прилагательными, либо предложно-падежными конструктами и относятся к языку-объекту. Видно, что студент воспринимает картину повреждений таза слитно, не выделяя явных оппозиционных групп. Семантический граф этого предложения может быть легко представлен в рамках классического идеографического метода Ю.Н.Караулова, где корнем графа является тема предложения (Т – «Все переломы таза»), а его листьями – члены ремы предложения (Р₁ – Р₆) (рис. 1).

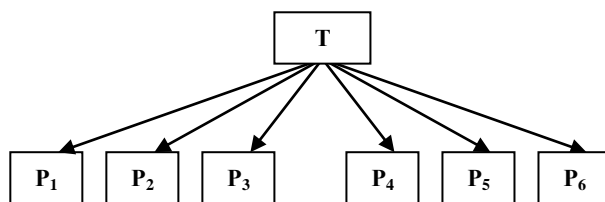


Рис. 1. Граф ответа «А»

Во втором предложении рема разделена на три группы, которые представлены семантически оппозиционными парами. Эти пары обозначены цифрами, которые могут быть легко заменены словами. Очевидно, что в данном высказывании цифровые обозначения «1», «2» и «3» могут рассматриваться как элементы метаязыка.

Это усложняет графическую репрезентацию данного высказывания за счёт появления вершин и связей, которые относятся к метаязыковому слою (рис. 2). Классический идеографический метод Ю.Н.Караулова не предусматривает репрезентации метаязыковых элементов.

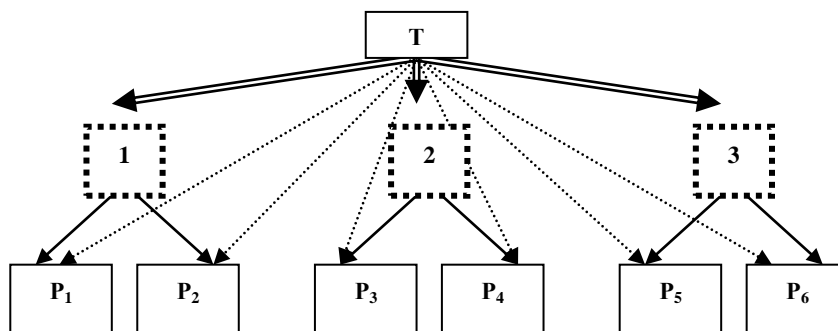


Рис. 2. Граф ответа «Б»

Третий ответ представлен только критериями (К) выделения видов переломов таза. Элементы языка-объекта в данном случае отсутствуют, они лишь подразумеваются. Идеографический метод в его классическом варианте здесь уже не применим. Сравнение графов №1 и №3 будет

некорректным (рис. 3). Даже при их внешнем сходстве они имеют лишь один общий элемент – корень Т. Это лучший из трёх ответов: студент показал не только знание классификации, но и понимание принципов её построения.

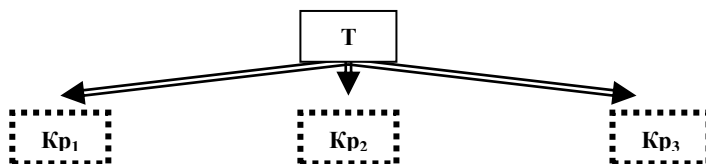


Рис. 3. Граф ответа «В»

Таким образом, для того, чтобы классический идеографический метод Ю.Н.Караулова мог использоваться для репрезентации, а тем более для сравнения различных высказываний таксономического типа, он должен быть расширен и дополнен: а) возможностями для репрезентации метаязыковых элементов высказываний; б) построением обобщённого графа, описывающего семантику всех слоёв метаязыка и языка-объекта данной предметной области как некоторого универсального предметного кода этой области.

2. Семантический анализ структуры составного термина, относящегося к предметной области «Повреждения таза». Для решения второй задачи проанализируем семантическую структуру составного термина, описывающего семантику данной предметной области. Проведённый анализ показал, что в основе описания семантики классификации повреждений таза лежат составные термины, структуру которых можно выразить абстрактной формулой: $\{P\} \Leftarrow \{S\} \rightarrow \{A\}$, где знаком « \Leftarrow » обозначено согласование, а « \rightarrow » – управление между элементами лексико-семантических групп, образующих предметную область. Примером таких составных терминов могут быть выражения *открытый перелом таза, закрытый отрыв правого крыла таза, передний вывих копчика* и т.д. В реальной речи врачей эти составные термины могут быть фрагментами классификации повреждений таза, готовыми диагнозами, а также фрагментами сложных предложений.

Главными словами этих выражений являются слова, обозначающие виды повреждений: *переломы, отрывы, вывихи* и т.д. Обозначим их как S_1, S_2, \dots, S_M . Т.о., мы получаем субъектную лексико-семантическую группу (ЛСГ $\{S\}$). Вторую группу слов и словосочетаний составляют анатомические термины *таз, кости таза, копчик, правое крыло таза* и т.д. Это объектная ЛСГ; обозначим её элементы как A_1, A_2, \dots, A_N . Наконец, третью, признаковую ЛСГ составляют прилагательные вроде *открытый, закрытый, передний* и т.д. Обозначим их как P_1, P_2, \dots, P_L .

Т.о., общую семантическую структуру предметной области «Повреждения таза» можно представить в виде т.н. леса, т.е. графа, состоящего из нескольких плоских подграфов-деревьев. В данном случае лес образован графами трёх ЛСГ – признаковой ($\{P\}$), субъектной ($\{S\}$) и объектной ($\{A\}$). Каждый граф состоит из определённого числа элементов двух видов – элементов метаязыка и языка-объекта. При этом вся собранная нами лексика, называющая признаки, виды повреждений и анатомические объекты, относится к языку-объекту и размещается в листьях графа (овальные элементы). Лексику, описывающую критерии выделения семантических оппозиций, следует отнести к метаязыку. На графе метаязыковые элементы представлены прямоугольными вершинами.



Рис. 4. Принципиальная схема семантической структуры предметной области «Повреждения таза»

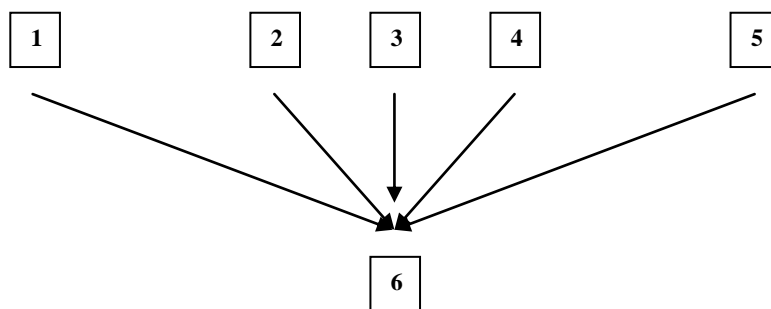


Рис. 5. Схема вхождения одного и того же элемента в разные подгруппы

В количественном отношении ЛСГ {S} представлена 24 терминами ($L = 24$). ЛСГ {A} состоит из минимум 1600 лексических единиц – это предметная терминология, общемедицинская терминология и общая (бытовая) лексика ($M \approx 1600$). При этом было выявлено наличие номинативных терминологических лакун, т.е. отсутствие терминологических наименований для реальных анатомических объектов⁵. Признаком лексико-семантическую группу (ЛСГ {P}) формируют около ста терминов-прилагательных ($N \approx 100$). Т.о., общее количество комбинаций элементов этих трёх ЛСГ по самой минимальной оценке должно составить около четырёх миллионов ($100 * 24 * 1600 = 3\,840\,000$).

Анализ семантики ЛСГ {A} позволил выявить тот факт, что одна и та же лексическая единица или словосочетание может входить в несколько семантических подгрупп одновременно. Например, словосочетание *верхняя ветвь правой лонной кости* (рис. 5, №6) является элементом деления концепта *правая лонная кость* (рис. 5, №1). Одновременно с этим данное словосочетание является элементом деления концептов «*правая тазовая кость*» (рис. 5, №2), «*правая половина таза*» (рис. 5, №3), «*переднее полукольцо таза*» (рис. 5, №4) и «*правая колонна таза*» (рис. 5, №5). Т.о., словосочетание *верхняя ветвь правой лонной кости* одновременно входит в пять семантических подгрупп. Идеографически данная семантика будет репрезентироваться уже не исходящими (как в классическом методе Ю.Н.Караулова) из неё стрелками, а входящими в неё (рис 5). Это приводит к переходу от классических древовидных структур к сетевым, а сам граф становится не плоским, а квазипространственным.

Ещё одна проблема идеографической репрезентации семантики была выявлена при анализе

ЛСГ {S} (виды повреждений). Среди лексических единиц, обозначающих виды повреждений (*переломы, вывихи, растяжения, отрывы и т.д.*) был выявлен термин *переломовывих*⁶. Очевидно, что он занимает промежуточное положение между строгими определениями *перелом* и *вывих*. Следует отметить, что сочетание двух и более видов других повреждений встречается весьма часто (классическим литературным примером может стать сочетание огнестрельных ранений ног с их последующим отморожением у известного лётчика А.Маресьева). Однако кроме термина «*переломовывих*», нами не зафиксировано лексикой, обозначающей такой род сочетаний.

При идеографической репрезентации это приведёт к появлению значительного количества значений, имеющих онтологическое содержание, но не имеющих терминологического обозначения, а только дескриптивное. На нашем материале (при 22 дочерних вершинах относительно корня «Повреждения» и одной переходной «Переломовывихи» в ЛСГ {S}) это явление при семантическом моделировании может привести к появлению 462 ($22 * 21 = 462$) словосочетаний. Иными словами, при идеографической репрезентации семантики данной ЛСГ за вершинами, обозначенными 22 зафиксированными терминами, имеется на порядок большее количество «пустых», или невербализованных с точки зрения репрезентирующей их лексикой концептов.

Обсуждение полученных результатов. Проведённый нами анализ семантической структуры терминосистемы «Повреждения таза» с применением для её визуализации идеографического метода Ю.Н.Караулова показал, что для адекватной идеографической репрезентации материала такого рода необходимо в первую очередь различать язык-объект и описывающий его метаязык. На графе это приводит к появлению различных типов вершин, а также связы-

⁵ Краснов А.Н. Психолого-педагогические основы технологии открытого тестирования....; Краснов А.Н., Слоева Е.А. Языковая репрезентация концептуальной модели

⁶ Краснов А.Ф., Мирошниченко В.Ф., Котельников Г.П. Травматология. М.: 1995. – С. 132.

вающих их отношений, характер которых необходимо учитывать.

Отдельного комментария требует методологическая сторона работы. Выявленное нами количественное ограничение (порядка 4 млн. теоретически возможных сочетаний), делает традиционный метод филологических исследований (сбор лексической базы, описание парадигматических, гипо- и гиперонимических отношений между её элементами, построение концептуальной модели) – изначально непродуктивным, поскольку для того, чтобы полученная в рамках данной предметной области модель была более или менее достоверной, филологу необходимо описать хотя бы 1 – 1,5 миллиона терминов. Из-за такого количественного ограничения данная задача перестаёт быть решаемой традиционными языковыми методами.

В этом отношении в методологическом плане более продуктивным представляется иной путь: от эпистемологии предметной области через конечные задачи обработки речевого материала к способам языкового выражения установленной предметной семантики. Иными словами, это путь от предметной экстралингвистики и её семантического моделирования к описанию способов номинации (т.е. после получения ограниченного материала сразу переходить к моделированию семантики). Минусом такого пути является организационная сторона работы: необходимо создание устойчивых коллективов из специалистов-предметников, филологов и программистов, работающих длительное время на конкретном направлении.

Выводы. Таким образом, отвечая на главный вопрос – об определении возможностей репрезентации семантики с помощью идеографического метода и возникающих при этом проблем – мы можем сказать следующее: 1) Классический идеографический метод визуальной (идеографической) репрезентации семантических отношений в рамках предметной области, предложенный Ю.Н.Карауловым, нуждается в усовершенствовании. 2) Потребность в совершенствовании метода Караулова обусловлена, помимо прочего, необходимостью учитывать не только элементы языка-объекта (как у Ю.Н.Караулова), но и элементы метаязыка, что делает необходимым различать на графе как характер вершин, так и характер связывающих их отношений. 3) При построении графа составного термина, лежащего в основе описания семантики данной предметной области, мы переходим от одного «дерева» (как у Ю.Н.Караулова) к «лесу», представленному в данном конкретном случае сразу тремя деревьями. Это связано с выявленной нами трёхмерной семантической структурой пространства полного составного термина, и, соответственно, с трёхмерностью концептуальной семантической модели данной классификации, что отражается на общей структуре графа. 4) Наконец, мы должны констатировать возможность появления при моделировании семантики не только древовидных (как у Ю.Н.Караулова), но и сетевых структур.

PROBLEMS OF IDEOGRAPHIC REPRESENTATION OF TAXONOMIC-TYPE EXPRESSIONS SEMANTICS

© 2013 A.N.Krasnov¹, A.P.Zhuravlyov², E.A.Sloeva^{1*}

¹Samara State Medical University

²Samara State Technical University

This article presents the analysis that reveals certain problems that occur in the process of representation of semantic structure in the subject matter with the use of ideographic method developed by Y.Karaulov. The article shows that Karaulov's classic ideographic method needs to be modified in order to represent the semantic structure more properly.

Keywords: metalanguage, lexical-semantic group, subject matter.

^{*} Aleksandr Nikodimovich Krasnov, Ph.D in Medicine, Professor, Head of the pedagogy, psychology and psycholinguistics department. E-mail: ankrasnov@mail.ru

Aleksandr Pavlovich Zhuravlyov, teaching assistant of the foreign languages department (Samara State Technical University), senior laboratory assistant of the electronic educational aid laboratory of the pedagogy, psychology and psycholinguistics department (Samara State Medical University). E-mail: palych32@rambler.ru

Elena Anatolevna Sloeva, PhD in Philology, associate professor of the pedagogy, psychology and psycholinguistics department. E-mail: Elena-sloeva@efndex.ru