

МОДЕЛИРОВАНИЕ ПРОФЕССИОНАЛЬНЫХ ЗНАНИЙ ЛЕКСИКОГРАФА

Аннотация

В работе предложен метод моделирования профессиональных знаний лексикографа. В рамках разработки метода предложена математическая модель лексикографического произведения. Процесс лексикографирования представлен в виде структуры целевых установок реализации замысла создания лексикографического произведения.

Значительные достижения последнего десятилетия в области информационных и коммуникационных технологий позволяют решать задачи, которые до настоящего времени считались слабоструктурированными и слабо формализуемыми. Одной из таких задач является создание интеллектуальных систем, обеспечивающих лексическую поддержку образовательных процессов высших учебных заведений. Важным элементом такой системы должна стать специальная база знаний, построенная на основе моделей знаний профессиональных лексикографов.

Целью настоящей статьи является разработка метода моделирования профессиональных знаний лексикографа.

Для достижения поставленной цели необходимо сформулировать исходные посылки, которые лягут в основу моделирования профессиональных знаний лексикографа.

Определим, какими знаниями должен обладать лексикограф для построения словарей, при этом будем использовать последние достижения в области теории и практики лексикографии [1].

Во-первых, знаниями принципов лексикографирования, к которым относят:

1. Преемственность лексикографических произведений (ЛП).
2. Прагматизм при создании ЛП.
3. Эвристика решений в процессе создания ЛП.
4. Нормативность в отборе и подаче лексики.
5. Теоретическая и практическая многоплановость ЛП.

Во-вторых, знаниями универсальных составляющих словаря любого типа (лексикографические универсалии), к которым относят:

1. Словник (левая часть словаря).
2. Заголовочная единица.
3. Толкование, дефиниция, переводной эквивалент и т.д. (правая часть словаря).
4. Системность языкового материала, описываемого в словаре.
5. Нормативность – допустимая не нормативность языковых единиц.
6. Определенный порядок подачи словарного материала.

В-третьих, знаниями о параметризации (характеристиках однопараметрических и многопараметрических словарей), а также структурах словарей (макро- и микроструктурах).

В-четвертых, знаниями одной или нескольких типологий словарей, т.е. знаниями классификационных признаков, по которым словари можно отнести к тому или иному классу (типу). В дальнейших исследованиях будем использовать типологию, предложенную в работе [1].

В-пятых, знаниями правил морфологии, орфографии, синтаксиса, а также правилами орфоэпии.

В основу модельного представления профессиональных знаний лексикографа положим как логические, так и эвристические методы, которые подробно изложены в работах [2, 3].

Опираясь на вербальное описание профессиональных знаний лексикографов, и используя когнитивный метод постановки задач принятия решений, приведенный в работе [4], представим процесс лексикографирования в виде обобщенной структуры целевых установок, которая приведена на рис.1. Здесь структура целевых установок изображена графом древовидной формы. В теории принятия решений такой граф называется деревом решений.

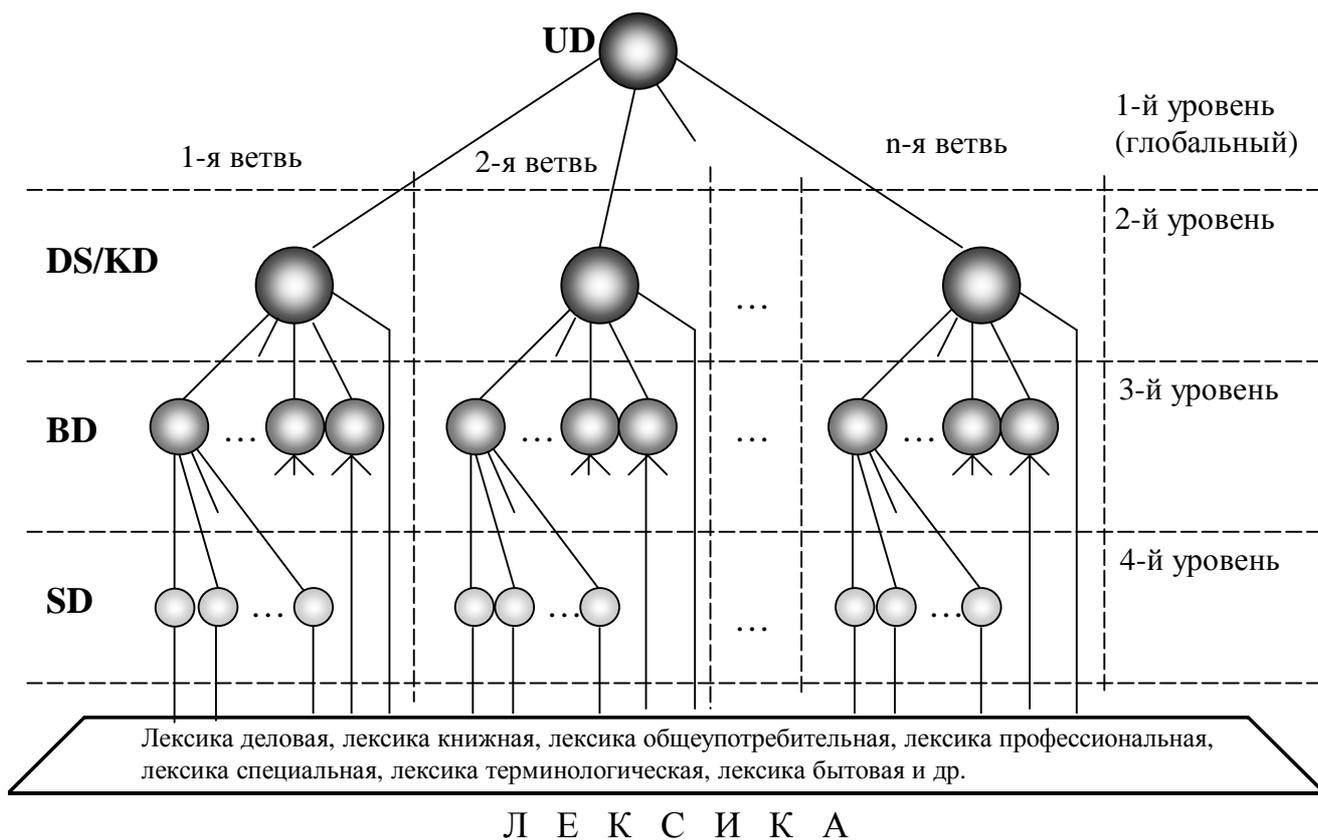


Рис.1. Обобщенная структура целевых установок лексикографирования

По сути приведенная на рисунке структура целевых установок отражает замысел создания того или иного ЛП, которая ориентируется на конкретную группу пользователей. Здесь выделены четыре уровня иерархии дерева решений и n его

ветвей. Каждая из вершин дерева решений может быть конкретизирована, детализирована и представлена самостоятельной иерархической структурой.

1-й уровень иерархии содержит одну единственную цель: «Удовлетворение лексикографических потребностей всех категорий пользователей». Последствием достижения данной цели является универсальный словарь (universal dictionary, UD), при создании которого он обладал бы совокупностью свойств всех известных ЛП и, кроме того, содержал полный словарный состав языка. В работе [1] отмечается, что создание универсального словаря практически неразрешимая задача, следовательно, рассматриваемая цель является абстрактной.

2-й уровень иерархии дерева решений содержит множество целей: «Удовлетворение лексикографических потребностей больших групп пользователей с широким кругом интересов». Последствиями достижения этого множества целей являются словарные системы или комплексные словари (dictionary system/ **komplex dictionary**, DS/KD).

3-й уровень иерархии дерева решений содержит множества целей: «Удовлетворение лексикографических потребностей больших групп пользователей с узким кругом интересов». Последствиями достижения настоящего множества целей являются лексикографические произведения для различных отраслей экономики государства (branch dictionary, BD), например, словари, справочники, отраслевые стандарты и т.д.

4-й уровень иерархии дерева решений содержит множество целей: «Удовлетворение лексикографических потребностей малых групп пользователей с узким кругом интересов». Последствиями достижения данного множества целей являются специализированные словарно-справочные средства (special dictionary, SD).

Ветви в дереве решений соответствуют множеству вариантов решений по созданию различных видов словарей, например, толковых, идеографических, орфографических, фразеологических, переводных, обратных и т.д.

Корнями дерева решений в данном случае является лексический материал, размещенный в ЛП различного вида, что является основой лексикографирования, т.е. первым принципом лексикографирования – преемственности лексикографических произведений. Что касается второго принципа лексикографирования - прагматизма при создании ЛП -, то он выражается собственно в формулировках множества целевых установок, которые направлены на удовлетворение лексикографических потребностей конкретных групп с конкретными интересами.

Значительный и разнотипный исходный (лексический) материал, а также множество правил его обработки при создании ЛП обуславливает творческий характер деятельности лексикографа, который предполагает принятие им эвристических решений с целью сокращения вариантов анализа лексического материала. Этот факт отражен на дереве решений значительной его разветвленностью и возможностью поиска кратчайших путей к достижению поставленной цели.

Принцип нормативности в отборе и подачи лексики некоторым образом связан с предыдущим принципом лексикографирования и опосредованно отражен начальными условиями принятия решений, которые в дереве решений являются корнями. Другими словами, нормативность и подача лексики при создании слова-

ря зависит от лексики, которую использует лексикограф, а также вида создаваемого словаря.

Принцип теоретической и практической многоплановости ЛП в дереве решений отражен множеством ветвей, каждая из которых соответствует варианту достижения целей построения того или иного вида лексикографического произведения.

Проанализируем цепь рассматриваемых категорий: профессиональные знания \rightarrow цели профессиональной деятельности \rightarrow последствия достижения целей профессиональной деятельности лексикографа, которая иллюстрируется рис.2. В верхней части рисунка показано, как на основе профессиональных знаний достигаются (или не достигаются) цели профессиональной деятельности, которые в случае их достижения образуют последствия (конкретные лексикографические произведения) достижения этих целей.

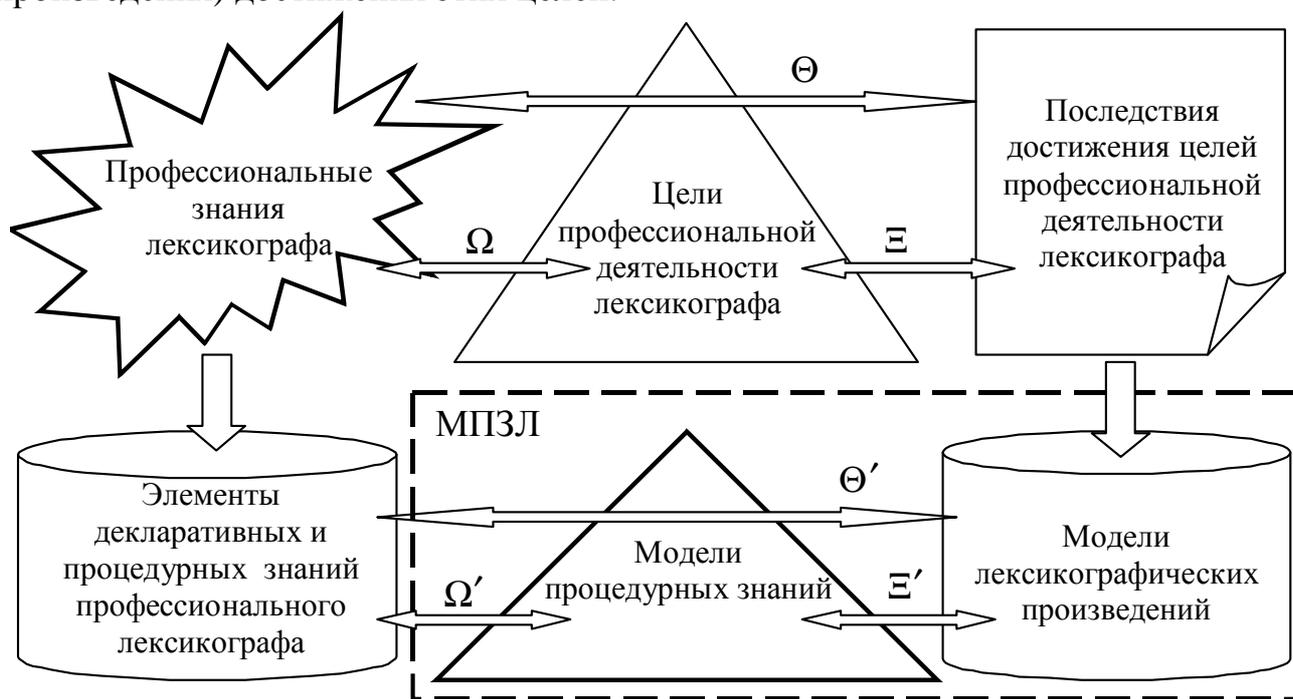


Рис.2. Обобщенная схема моделирования профессиональных знаний лексикографа

Здесь: Θ – множество отношение между элементами профессиональных знаний лексикографа, например, лексического материала, расположенного в определенном порядке, и элементами конечного продукта его профессиональной деятельности, т.е. лексическим материалом конкретного лексикографического произведения; Ω – множество отношений между элементами профессиональных знаний, например, лексикографическими универсалиями и структурными единицами целевых установок создания того или иного ЛП; Ξ – множество отношений, связывающих элементы структуры целевых установок с характеристиками и параметрами конкретного ЛП.

В нижней части рис. 2 показано, что профессиональные знания лексикографа можно декомпозировать на две составные части – декларативную, которая может быть выражена только на вербальном уровне и процедурную (знания конкретных правил манипулирования морфологическим и лексическим материалом). Здесь

множества отношений Θ' , Ω' и Ξ' являются аналогом множества отношений Θ , Ω и Ξ полученных путем их корреспонденции (переноса) из реальных процессов создания ЛП в процессы их моделирования. Понятие термина «корреспонденция отношений» используется в работе [5] и трактуется как операция над отношениями и обозначается $\Theta \xrightarrow{\text{Kor}} \Theta'$, $\Omega \xrightarrow{\text{Kor}} \Omega'$ и $\Xi \xrightarrow{\text{Kor}} \Xi'$.

Из вышесказанного следует, что формальное представление ЛП играет важную роль при создании модели профессиональных знаний лексикографа (МПЗЛ). Поэтому воспользуемся идеями и методами корпусной лингвистики [6, 7] для представления макро- и микроструктур лексикографических произведений. Покажем на примере создание математической модели ЛП общенаучного назначения [8].

Будем считать лексикографическое произведение [8] корпусом текстов, в котором выделим основные его элементы, как это показано на рис. 3.

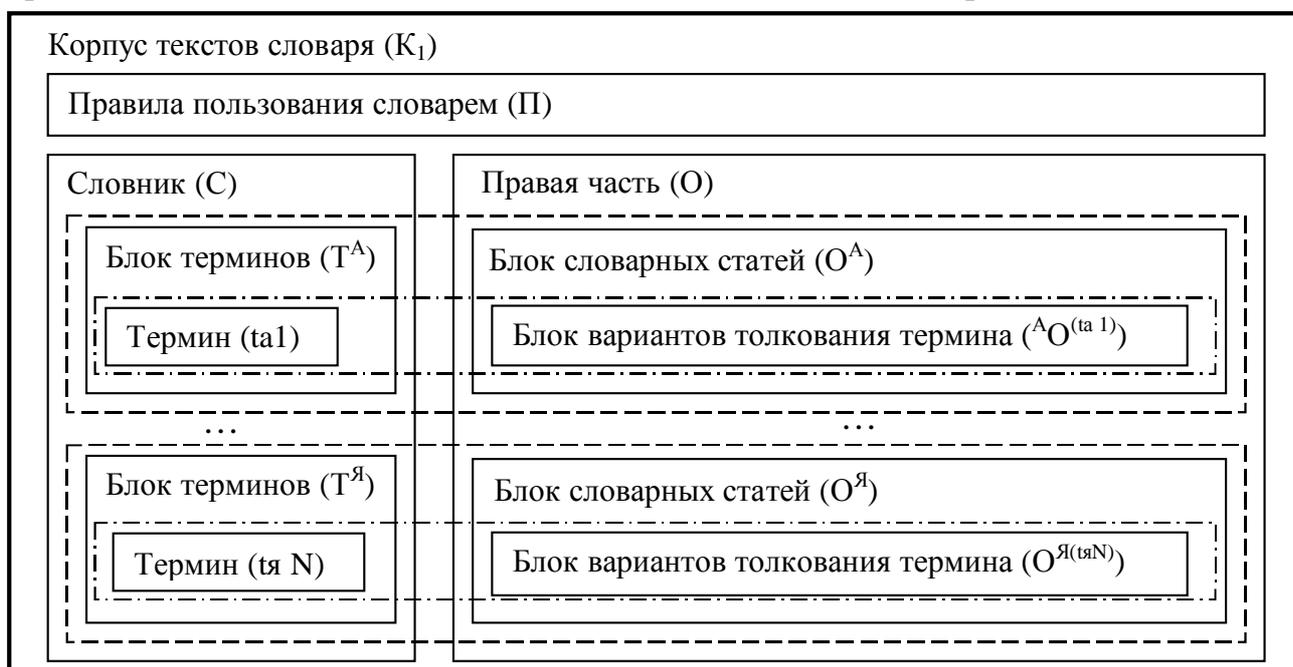


Рис.3. Структура корпуса текстов толкового словаря Ожегова С.И.

Из анализа структуры словаря и рис. 3 видно, что на его элементах заданы несколько видов отношений. Это отношения строгого порядка (блоки терминов и собственно термины расположены в алфавитном порядке), обозначим их символом ($>$), и включения (\subset), например, между корпусом текстов и словарем, между словарем и блоком терминов и т.д. Кроме того, между словарем (С) и правой частью словаря (О), а также их элементами существуют бинарные отношения.

Обозначим:

K_1 - корпус текстов образующий толковый словарь;

$\Pi = \{ \pi_i \}, i = \overline{1, n}$ - множество текстов образующих правила пользования словарем, элементами которого могут быть список принятых сокращений (памет), характеристика употребления слов и другие;

$C = \{ T^A, \dots, T^Я \}$ - левая часть словаря, которая представляет собой множество блоков терминов, расположенных в алфавитном порядке от А до Я;

$T^A = \{ta_1, \dots, ta_\alpha\}$ - блок терминов, у которого литеры начинаются с буквы А, α - количество терминов, помещенных в словарь с первой литерой А, остальные блоки терминов (от Б до Я) имеют аналогичный состав;

$O = \{O^A, \dots, O^Я\}$ - правая часть словаря, которая состоит из подмножеств блоков словарных статей $O^A, \dots, O^Я$, расположенных также в алфавитном порядке;

$O^A = \{A O^{(ta_1)}, \dots, A O^{ta_\beta}\}$ - блок интерпретаций, термины которых начинаются с литеры А, где β - количество словарных статей этого блока или кардинал подмножества O^A ;

$A O^{ta_1} = \{a_o^{ta_1.1}, \dots, a_o^{ta_1.4}\}$ - подмножество вариантов толкования термина (в этом случае левый верхний индекс показывает, что в блоке словарных статей O^A словарная статья соответствующая термину ta_1 содержит четыре варианта толкования).

Подвергнем более детальному анализу отношения, в которых находятся элементы корпуса текстов.

Свяжем элементы корпуса текстов отношением включения

$$\begin{aligned} \{П, С, О\} \subset K_1; \{T^A, \dots, T^Я\} \subset C; \{ta_1, \dots, ta_\alpha\} \subset T^A, \dots, \{тя_1, \dots, тя_\delta\} \subset T^Я; \\ \{O^A, \dots, O^Я\} \subset O; \{A O^{ta_1}, \dots, A O^{ta_\beta}\} \subset O^A, \dots, \{Я O^{тя_1}, \dots, Я O^{тя_\gamma}\} \subset O^Я; \\ \{a_o^{ta_1.1}, \dots, a_o^{ta_1.\phi}\} \subset A O^{ta_1}, \dots, \{я_o^{тя_1.1}, \dots, я_o^{тя_1.\epsilon}\} \subset Я O^{тя_N}. \end{aligned}$$

Отношения строгого порядка между подмножествами С и О и их элементами зададим в следующем виде:

$$\begin{aligned} T^A > T^Б > \dots > T^Я; ta_1 > ta_2 > \dots > тб_1 > тб_2 > \dots > тя_1 > \dots > тя_\delta; \\ O^A > O^Б > \dots > O^Я; A O^{ta_1} > \dots > A O^{ta_\beta} > Б O^{тб_1} > Б O^{тб_2} > \dots > Я O^{тя_1} > \dots > Я O^{тя_\gamma}; \\ a_o^{ta_1.1} > \dots > a_o^{ta_1.\phi}; \dots; я_o^{тя_1.1} > \dots > я_o^{тя_1.\epsilon}. \end{aligned}$$

Учитывая, что между подмножествами С и О, а также их элементами существуют некоторые бинарные отношения, можно записать следующие соответствия

$$\begin{aligned} E \subseteq C \times O; H^A \subseteq T^A \times O^A, \dots, H^Я \subseteq T^Я \times O^Я; \\ G^A \subseteq ta_1 \times A O^{ta_1}, \dots, G^Я \subseteq тя_\delta \times Я O^{тя_\delta}, \end{aligned}$$

где $(C, O) \in E$ - множество пар, в данном случае одна (словник и правая часть словаря), $(T^A, O^A) \in H^A, \dots, (T^Я, O^Я) \in H^Я$ - множество пар блоков терминов и их словарных статей, $(ta_1, A O^{ta_1}) \in G^A, \dots, (тя_\delta, Я O^{тя_\delta}) \in G^Я$ - множество пар терминов и блоков вариантов его толкования.

В случае, когда одному термину соответствует несколько толкований можно записать, например, для термина начинающегося с литеры А, $D^A \subseteq ta_1 \times (a_o^{ta_1.1}, \dots, a_o^{ta_1.\phi})$, где множество пар $(ta_1, a_o^{ta_1.1}) \in D^A, \dots, (ta_1, a_o^{ta_1.\phi}) \in D^A$ термина и его толкований.

В обобщенном виде, отождествляя корпус текстов с моделью общенаучного словаря $K_1 \equiv M^{об}$ и не учитывая соответствие термину нескольких определений, запишем

$$M^{об} = \left\langle \Pi, T^A, \dots, T^Я, O^A, \dots, O^Я, ta_1, \dots, тя_\delta, A_{O^{ta_1}}, \dots, Я_{O^{тя_\delta}}, c, > \right\rangle. \quad (1)$$

По аналогии можно построить модели других видов ЛП.

Процедурную составляющую МПЗЛ будем представлять системой правил, к которой отнесем:

1. Правила отбора ЛП для создания системы словарей или комплексного словаря.
2. Правила создания словарей конкретных типов.
3. Правила интеграции свойств лексикографических произведений различных типов.
4. Морфологические и орфографические правила языка, на котором разрабатывается лексикографическое произведение.
5. Синтаксические правила и правила пунктуации.
6. Орфоэпические правила.

Правила отбора ЛП, в частности, для обеспечения лексической поддержки образовательных процессов в высшей школе, в настоящее время не разработаны. Поэтому, на наш взгляд, необходимо разработать специальную методику экспертного оценивания ЛП с целью формирования системы словарей, отражающей лексику учебных дисциплин, которые изучаются в рамках конкретного учебного плана по специальности конкретного вуза. Разработка такой методики выходит за рамки настоящей работы.

Для формального представления остальных перечисленных правил воспользуемся методами формального описания текстовой информации в полнотекстовых базах данных, которые используются в современных автоматизированных библиотечных комплексах [9, 10]. Здесь для описания структур текстовой информации используются ситуационно-текстовые предикаты.

Покажем на отдельных примерах возможность описания вышеперечисленных правил на основе логики исчисления предикатов.

В обобщенном виде типологию ЛП можно представить в следующем виде:

$P(l_i, x_j)$ - предикат « l_i имеет признак x_j », где l_i - лексикографическое произведение i -го вида, x_j - признак, по которому классифицируется то или иное ЛП.

Лексикографические произведения могут быть следующих видов: энциклопедические, толковые, терминологические, этимологические, ономастические, учебные, лингвострановедческие, словари иностранных слов и т.п.

Признаки, которые предлагается использовать в работе [1] для классификации словарей, будем использовать, как предметные переменные, имеющие следующие значения:

- x_1 - признак одноязычного словаря;
- x_2 - признак переводного словаря;
- x_3 - признак использования лексики «без ограничений»;

x_4 - признак использования определенных лексических пластов в словарях;
 x_5 - признак большого объема лексики в словаре;
 x_6 - признак краткого словаря;
 x_7 - признак лексического минимума, помещенного в словарь;
 x_8 - признак электронного носителя лексической информации;
 x_9 - признак бумажного носителя лексической информации;
 x_{10} - признак функционально-отраслевого словаря;
 x_{11} - признак функционально-языкового словаря;
 x_{12} - признак функционально-образного словаря;
 x_{13} - признак алфавитного порядка расположения заголовочных слов в слова-

ре;

x_{14} - признак квазипорядка между семантическими полями заголовочных слов словаря;

x_{15} - признак алфавитного порядка заключительных букв в заголовочных словах в словаре;

x_{16} - признак алфавитного порядка имен и фамилий;

x_{17} - признак алфавитного порядка названий государств;

x_{18} - признак использования в словаре нормативной лексики;

x_{19} - признак использования в словаре ненормативной лексики;

x_{20} - признак компиляции различных видов словарей;

x_{21} - признак дидактических свойств словаря.

Естественно предположить, что приведенный список признаков не полный и при необходимости может быть дополнен новыми признаками.

Исходя из перечисленных признаков, запишем типологию словарей.

$P_1(l_i, x_1) \vee P_1(l_i, x_2) = 1$ - типы словарей, зависящих от количества описываемых языков;

$P_2(l_i, x_3) \vee P_2(l_i, x_4) = 1$ - типы словарей, классифицирующиеся по охвату лексики;

$P_3(l_i, x_5) \vee P_3(l_i, x_6) \vee P_3(l_i, x_7) = 1$ - типы словарей, классифицирующиеся по объему лексического материала;

$P_4(l_i, x_8) \vee P_4(l_i, x_9) = 1$ - типы словарей, классифицирующиеся по носителю лексической информации;

$P_5(l_i, x_{10}) \vee P_5(l_i, x_{11}) \vee P_5(l_i, x_{12}) = 1$ - типы словарей, зависящие от функциональной направленности;

$P_6(l_i, x_{13}) \vee P_6(l_i, x_{14}) \vee P_6(l_i, x_{15}) = 1$ - типы словарей, зависящих от порядка подачи лексического материала;

$P_7(l_i, x_{16}) \vee P_7(l_i, x_{17}) \vee P_7(l_i, x_{18}) \vee P(l_i, x_{19}) = 1$ - типы словарей, отражающих культурологический аспект языка;

$P_8(l_i, x_{20}) = 1$ - смешанные, или комплексные словари;

$P_9(l_i, x_{21}) = 1$ - учебные словари.

Из приведенных выше выражений можно создавать более сложные формулы, которые будут описывать типы и виды словарей с различными признаками. Например, $P_1(l_i, x_1) \wedge P_2(l_i, x_4) \wedge P_3(l_i, x_6) \wedge P_4(l_i, x_9) \wedge P_6(l_i, x_{13}) = 1$. Подставляя в это выражение предметную переменную l_i - «толковый словарь» и соответствующие признаки получим высказывание «Толковый словарь является одноязычным, описывающим определенный пласт лексики, кратким по объему на электронном носителе, заголовочные слова которого расположены в алфавитном порядке».

Такое построение формул отражает суть интеграции свойств словарей различного вида, которую будем называть лексикографической интеграцией первого рода.

Под лексикографической интеграцией второго рода будем понимать интегральное описание лексем интегрального словаря, т.е. описание лексем во всех вариантах ее использования в языке. Другими словами, описание лексем морфологическими, орфографическими, семантическими, синтаксическими и орфоэпическими средствами языка.

Например, указания в словарной статье заголовочного слова части речи, к которой оно принадлежит, ударений, синонимов, омонимов, гиперонимов, фразеологических зон, иллюстраций и т.п.

Формально множество заголовочных слов словаря обозначим $Z = \{z_v\}$, $v = \overline{1, n}$, где n – кардинальное число этого множества. Каждому элементу этого множества соответствует некоторое множество атрибутов, определяющих степень параметризации словаря и в конечном итоге его свойства в целом.

Логику лексикографической интеграции второго уровня можно задать, используя аналогию интеграции приведенной выше.

Обозначим $G(z_v, h_m)$ - предикат « z_v имеет параметр h_m », где z_v - заголовочное слово словарной статьи, h_m - параметр словаря.

Имея полный список заголовочных слов, а также словарных параметров, можно по аналогии записать множество формул, на языке логики предикатов. При этом множество заголовочных слов и словарных параметров будут являться предметными переменными в этих формулах.

Из вышесказанного видно, что для представления декларативной составляющей знаний лексикографа можно воспользоваться известными фреймовыми представлениями знаний. Это предположение основывается на том, что приведенная на рис.3 структура модели словаря имеет высокую степень подобия со структурой фреймовых представлений [3]. Основным бинарным элементом в структуре словаря является «заголовочная единица – словарная статья», а во фреймовой системе представления знаний «имя слота – значение слота». Поэтому большинство ЛП как гомоморфное отображение знаний лексикографа можно представить в виде фреймовой системы.

Естественно предположить, что сложность структуры фреймовой системы будет различной в зависимости от учета при моделировании количества параметров словаря. Так, например, слоты фреймов толкового одноязычного словаря [8] будут иметь вид, приведенный на рис. 4.



Рис. 4. Фрагмент структуры фрейма толкового одноязычного словаря

При создании фреймовой системы для многопараметрического словаря ее структура усложняется и принимает вид, фрагмент которого приведен на рис.5. Здесь во фреймовую структуру вводятся дополнительные поля «Атрибуты» и «Служебное поле», которые предназначены для управления выводом лексической информации. В поле «Атрибуты» помещаются необходимые словарные параметры, а в «Служебное поле» - присоединенные процедуры, при помощи которых можно вызвать специальные программы обработки текстовой информации, например, программы для разрешения синонимической неопределенности при обращении к слоту, заголовочная единица которого имеет несколько толкований.

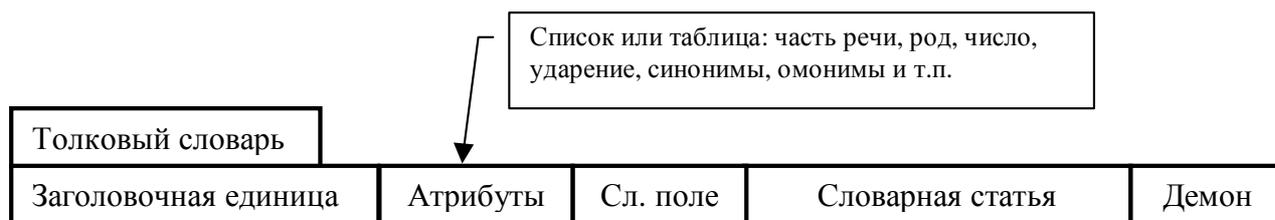


Рис.5. Фрагмент структуры фрейма многопараметрического словаря

Процедурные знания лексикографа будем формально представлять продукционными моделями «ЕСЛИ условие ..., ТО действие», опираясь при этом на логику лексикографической интеграции второго рода.

Продукционные модели в данном случае должны отражать правила пользования ЛП, которые помещаются в специально предназначенном разделе словаря.

Например: «ЕСЛИ заголовочная единица является существительным, ТО она подается в именительном падеже единственного числа», «ЕСЛИ заголовочная единица является прилагательным, ТО она подается в именительном падеже мужского рода в полной форме», «ЕСЛИ заголовочная единица является глаголом с окончаниями –ся, -сь, ТО он (его постфиксы) произносится мягко: *училс^бя, учи-лас^б*» и т.д.

Представление ЛП фреймовыми системами ставит вопрос об организации более крупных лексических образований - систем словарей (DS/KD), имеющих различную топологию и степень параметризации для решения задач лексической поддержки образовательных процессов в вузах. С целью формализации знаний лексикографа о таких системах воспользуемся еще одним эвристическим методом представления знаний, который позволяет представлять знания в виде семантической сети.

Для уменьшения размерности решаемой задачи будем считать, что семантическая сеть содержит три вершины, как это показано на рис.6. Каждая из вершин представляет собой систему фреймов ЛП.

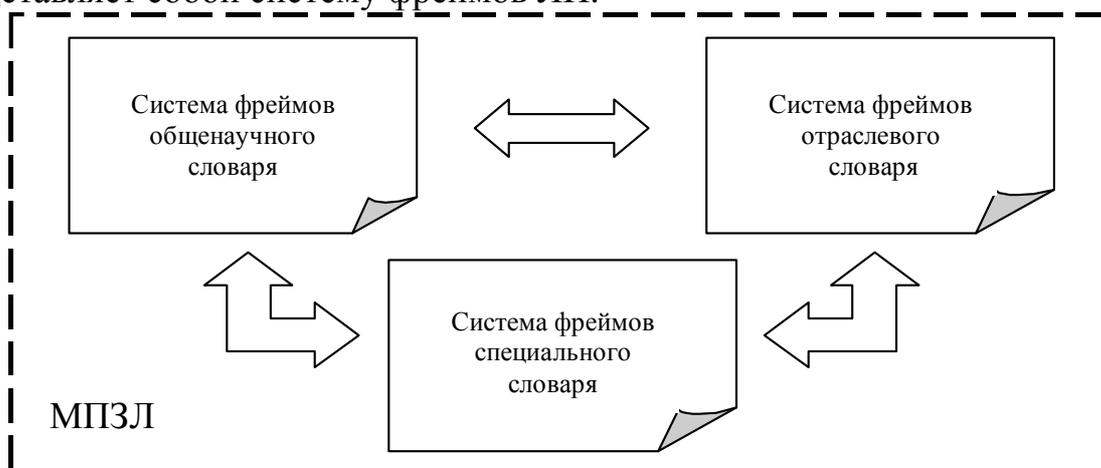


Рис. 6. Обобщенная схема семантической сети фреймовых моделей ЛП

На рис. 6 двойными стрелками показаны связи между фреймовыми системами, которые можно организовать при помощи продукционных правил на основе логики лексикографической интеграции первого рода.

Примерами в этом случае могут стать следующие правила: «ЕСЛИ общенаучный, отраслевой и специальный являются толковыми словарями, ТО их лексика соответствует интегральному словарю со свойствами этих словарей»; «ЕСЛИ лексикографическое произведение является общенаучным толковым словарем, ТО оно обеспечивает лексикографическую поддержку всех учебных дисциплин вуза»; «ЕСЛИ лексикографическое произведение является многоязычным словарем, ТО оно осуществляет лексикографическую поддержку учебных дисциплин, связанных с переводом иностранных слов» и т.п.

Подводя итог наших умозаключений, сделаем следующие выводы.

Моделирование знаний профессиональных лексикографов позволило по-новому взглянуть на задачу интеграции лексики с целью лексической поддержки образовательных процессов в вузе.

Использование в процессе моделирования как логических, так и эвристических модельных представлений является отличительной особенностью представления профессиональных знаний лексикографов. В известной литературе имеются лишь описания процессов и явлений в лексикографии на основе теоретико-множественного и категорного языков только логическими методами.

Разработка соответствующих эвристик процесса создания словарей позволит вплотную подойти к решению проблемы построения лингвистических баз знаний, которые обеспечат лексическую поддержку всех участников образовательных процессов в вузе на основе методов искусственного интеллекта.

ЛИТЕРАТУРА

1. Дубичинский В.В. Теоретическая и практическая лексикография. Вена -Харьков. 1998 - 160 с.
2. Осуга С., Саэки Ю. Приобретение знаний. - М.: Мир, 1990. - 304 с.

3. Представление и использование знаний / Под ред. Х.Уэно, М.Исидзука. - М.: Мир, 1989. - 220 с.
4. Ярушек В.Е., Прохоров В.П., Судаков Б.Н., Мишин А.В. Теоретические основы автоматизации процессов выработки решений в системах управления. - Харьков, ХВУ. 1993 – 442 с.
5. Шрейдер Ю.А. Равенство, сходство, порядок. - М.: Наука. 1971. - 254 с.
6. Баранов А.Н. Введение в прикладную лингвистику: учебное пособие. - М.: Эдиториал УРСС, 2001. - 360 с.
7. Марчук Ю.Н. Основы компьютерной лингвистики. Учебное пособие. – М.: Изд-во МПУ «Народный учитель», 1999. - 221 с.
8. Ожегов С.И. Словарь русского языка: Ок 57 000 слов / Под ред. Н.Ю. Шведовой. – 16-е изд., испр. – М.: Рус. яз., 1984. – 797 с.
9. Шабанов-Кушнарченко Ю.П., Шаронова Н.В. Компараторная индентификация лингвистических объектов. - К.: ИСДО, 1993. - 116 с.
10. Хайрова Н.Ф. Компараторная идентификация документов в полнотекстовой базе данных // АСУ и приборы автоматики. - 1999. - № 109. С. 67 – 76.
11. Тлумачний словник офіцера протиповітряної оборони з оперативно-тактичної підготовки / За ред. Шмакова О.М. – Харків, ХВУ. 2000. – 122 с.