

ФОРМАЛИЗАЦИЯ ЛЕКСИКОГРАФИЧЕСКИХ ПРОИЗВЕДЕНИЙ

Современные достижения в компьютерной лексикографии открывают большие возможности по созданию системы лексикографической поддержки учебного процесса в вузах. Такая система может являться составной частью информационной подсистемы, обеспечивающей процессы выработки педагогических решений в современных обучающих системах (вузах), в которых обучение осуществляется на основе интегрированного интеллекта, т.е. с использованием, как естественного интеллекта преподавателей, так и их моделей профессиональных знаний [1, 2].

В настоящее время создано большое количество словарей как в традиционном исполнении в виде лексикографических произведений на бумажной основе, так и различных электронных словарно-справочных средств, которые обладают отдельными функциями переработки текстовой информации, например, тезаурусы, конкордансы, частотные словари и т.д.

Словарно-справочные средства вуза составляют лексический фонд подготовки специалистов по всем специальностям и специализациям вуза. Коротко назовем его лексическим фондом (ЛФ) вуза. На рис.1 иллюстрируется соответствие словарно-справочных средств вуза дисциплинам, которые в учебных планах образуют блоки: гуманитарных и социально-экономических дисциплин; фундаментальных и профессионально-ориентированных дисциплин, а также блок специальных учебных дисциплин.

Практика показывает, что в настоящее время отсутствуют методы и методики целенаправленного использования в процессе образовательной деятельности лексических фондов вузов, тем более методы и методики создания лингвистических баз данных учебного назначения, с помощью которых осуществлялась бы целенаправленная лексикографическая поддержка образовательных процессов вуза.

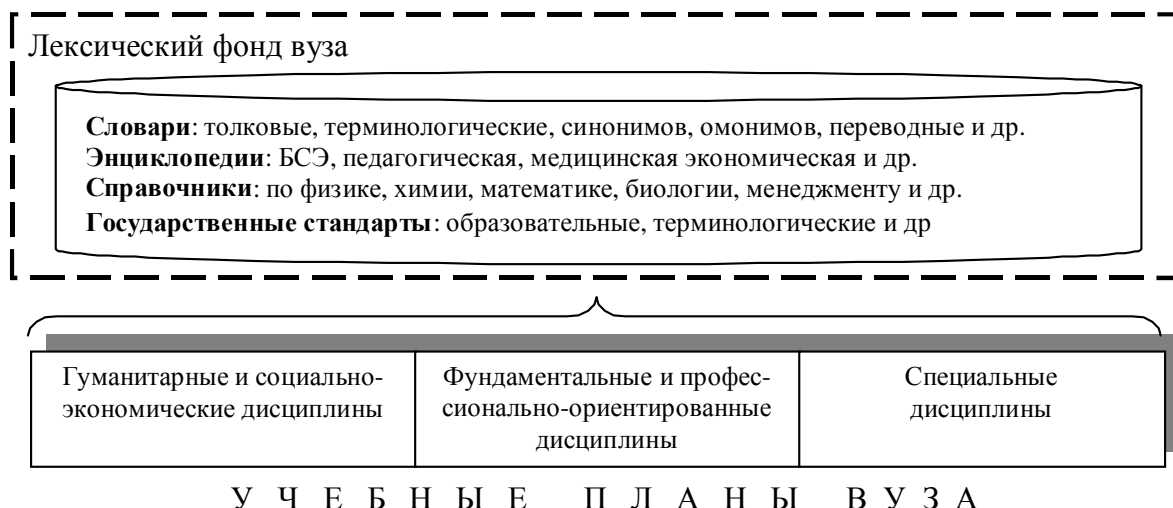


Рис.1. Схема существующего лексикографического обеспечения вуза

Целью настоящей работы является разработка моделей лексикографических произведений для дальнейшего построения на этой основе лингвистических баз знаний учебного назначения.

Подойдем к исследованию вопроса использования словарно-справочных средств в образовательных процессах высшей школы с общенаучных позиций и определим их целевое назначение с точки зрения философии, естественных, общественных и технических наук.

С философской точки зрения словарь создается как средство для познания истины человеком.

С точки зрения естественных наук словарь – это средство для познания и систематизации знаний человеком о явлениях и процессах в живой природе.

С точки зрения общественных наук словарь – средство для познания и систематизации знаний человеком о явлениях и процессах в обществе.

С точки зрения технических наук словарь является средством для познания и систематизации знаний человеком о явлениях и процессах протекающих в технике.

Кроме того, настоящие исследования будут опираться на последние достижения прикладной лингвистики, в частности компьютерной и корпусной лингвистики [3, 4]. Под углом зрения социолингвистики будем рассматривать вопросы, связанные с интернациональным составом участников учебного процесса.

Отметим тот факт, что любой учебный план подготовки специалиста в вузе содержит дисциплины в той или иной степени, обеспечивающие изучение перечисленных выше процессов и явлений. Кроме того, в лексикографии различают литературные и общенаучные словари и энциклопедии, например лексикографические произведения [5, 6], словарно-справочные средства, обеспечивающие лексикографическую поддержку в конкретных областях (отраслях) знаний, например [7, 8], а также специальные справочные средства [9, 10].

Поставим в соответствие классифицированные таким образом словарно-справочные средства учебному плану конкретной специальности, а именно, блокам: гуманитарных и социально-экономических дисциплин; фундаментальных и профессионально-ориентированных дисциплин, а также блоку специальных учебных дисциплин. Это соответствие схематично показано на рис. 2.

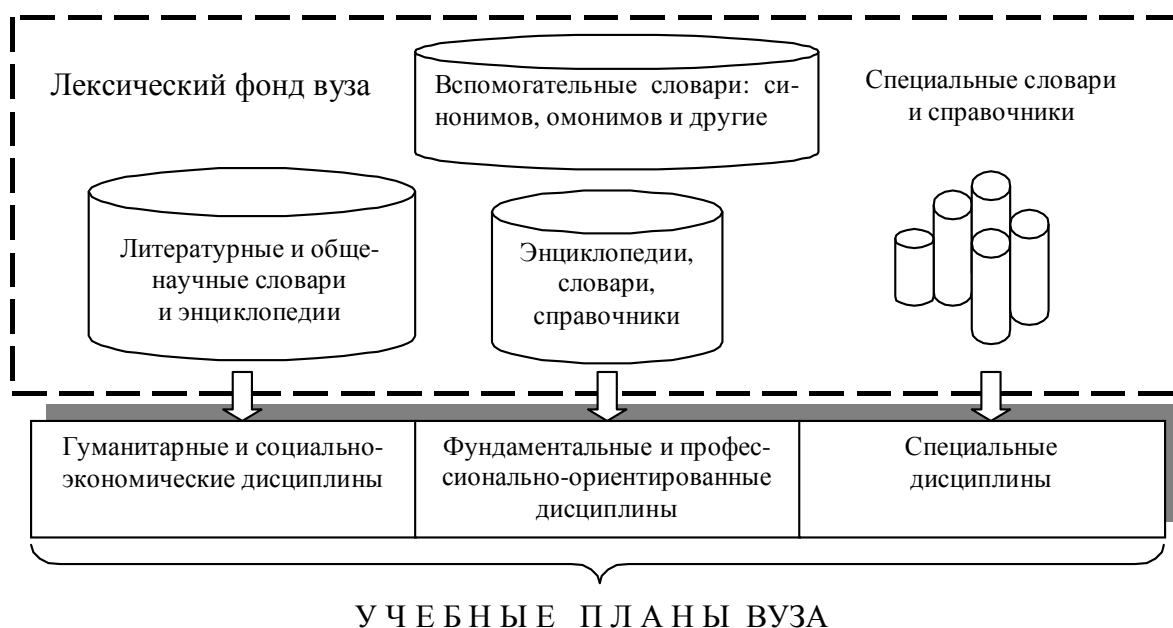


Рис. 2. Иллюстрация классификации словарно-справочных средств

К сожалению, в настоящее время отсутствуют, какие либо нормативные документы, которые бы ставили в строгое соответствие набор словарно-справочных средств той или иной специальности вуза или тому или иному блоку дисциплин учебных планов.

На рис.2 кроме приведенных выше словарно-справочных средств показаны вспомогательные лексикографические средства в виде словарей синонимов, омонимов и других, которые являются неотъемлемой составляющей для организации

процесса интеграции словарно-справочных средств и их целенаправленного использования всеми участниками учебного процесса.

Возникает задача на основе интеллектуальных информационных технологий создать интегрированные объектно-ориентированные словарно-справочные средства (ИОО ССС) с целью повышения эффективности образовательных процессов в вузе за счет оперативности использования лексического фонда вуза в учебном процессе, а также появления новых возможностей его использования.

Термин «объектно-ориентированных» введен потому, что словарно - справочные средства должны ориентировать участников образовательных процессов на использование некоторой совокупности лексикографических данных для решения учебно-воспитательных задач. Под лексикографическими данными будем понимать не только информацию справочного типа. Лексикографическими данными может быть информация о состоянии и пополнении лексического фонда вуза, его частотная характеристика и другая информация, обеспечивающая принятие решений направленных на совершенствование учебного процесса в вузе.

Зададим требования к ИОО ССС вуза.

Из названия «интегрированные объектно-ориентированные словарно - справочные средства» следуют основные требования, к которым отнесем следующие.

1. Способность из выбранных словарно-справочных средств формировать интегрированные словари различных типов и различного объема, например, тезаурусы, частотные словари, тематические (по дисциплинам, разделам, учебным темам, конкретному занятию) и др.

2. Обеспечение оперативного доступа к лексическому фонду вуза участников образовательных процессов.

3. Обеспечение орфографической и грамматической проверки и корректировки текста.

4. Обеспечение перевода терминов и определений несколькими иностранными языками.

5. Возможность построения «деревьев» понятий в рамках конкретных учебных планов и учебных дисциплин для обеспечения дидактических функций словарно-справочных средств.

6. Создание персональных словарей-минимумов.

7. Осуществлять исследование профессиональных языков научно - педагогических работников вуза и пополнять их лексику новыми лексическими единицами.

8. Осуществлять исследование языка обучающихся с целью определения метрики для оценивания близости лексики обучающихся и преподавателей.

Естественно полагать, что построению ИОО ССС должен предшествовать подготовительный этап, на котором осуществляется экспертиза словарно - справочных средств с целью отбора лексикографических произведений и установления возможности их использования при обучении по конкретным специальностям вуза. Отобранные в результате экспертизы лексикографические произведения должны быть подвергнуты тщательному анализу и на основе методов корпусной лингвистики построены их математические модели.

Покажем на примере создание математической модели лексикографического произведения общенаучного назначения [5].

Будем считать лексикографическое произведение [5] корпусом текстов, в котором выделим основные его элементы, как это показано на рис. 3.

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 словником, между словником и блоком терминов и т.д. Кроме того, между словником (С) и правой частью словаря (О), а также их элементами существуют бинарные отношения.

Обозначим:

K_1 - корпус текстов образующий толковый словарь;

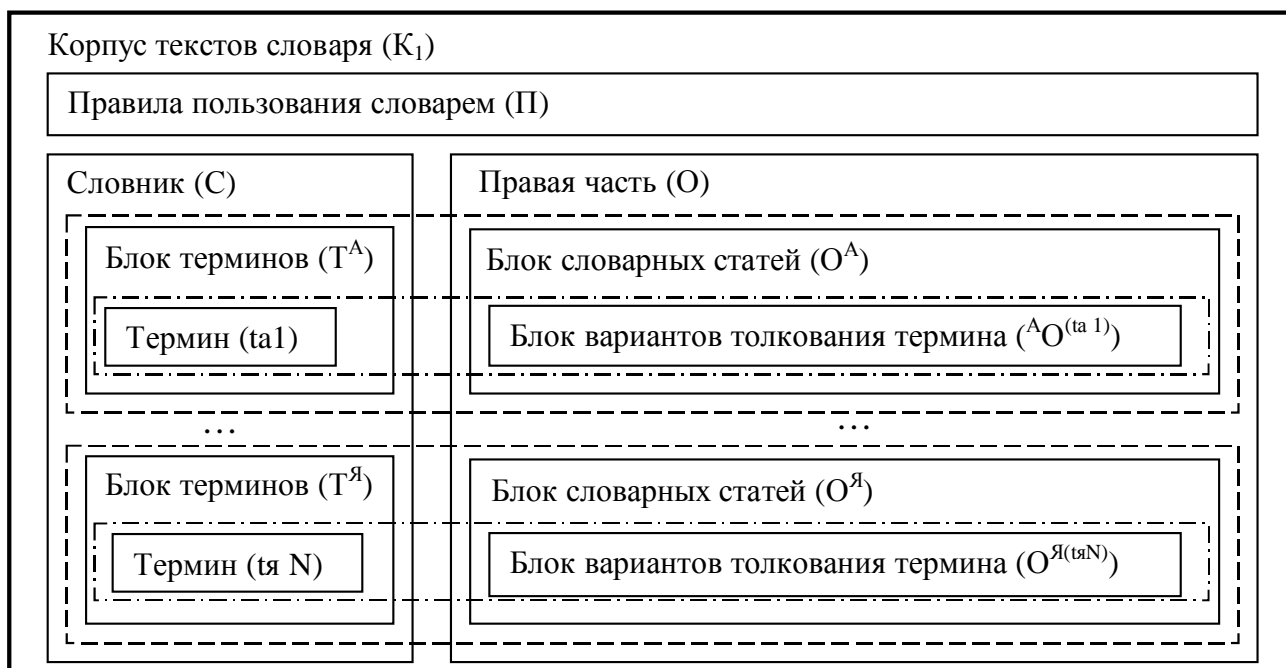


Рис.3. Структура корпуса текстов толкового словаря Ожегова С.И.

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

$C = \{ T^A, \dots, T^Я \}$ - левая часть словаря 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 словарных статей, 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

$$\begin{aligned} \{ П, С, О \} \subset K_1; \{ T^A, \dots, T^Я \} \subset C; \{ ta\ 1, \dots, t\alpha \} \subset T^A, \dots, \{ т\alpha\ 1, \dots, т\delta \} \subset T^Я; \\ \{ O^A, \dots, O^Я \} \subset O; \{ A_{O^{ta\ 1}}, \dots, A_{O^{ta\ \beta}} \} \subset O^A, \dots, \{ Я_{O^{т\alpha\ 1}}, \dots, Я_{O^{т\alpha\ \gamma}} \} \subset O^Я; \\ \{ a_{O^{ta\ 1.1}}, \dots, a_{O^{ta\ 1.\phi}} \} \subset A_{O^{ta\ 1}}, \dots, \{ я_{O^{т\alpha\ 1.1}}, \dots, я_{O^{т\alpha\ 1.\epsilon}} \} \subset Я_{O^{т\alpha\ N}}. \end{aligned}$$

Отношения строгого порядка между подмножествами C и O и их элементами зададим в следующем виде:

$$\begin{aligned} T^A > T^B > \dots > T^Я; ta\ 1 > ta\ 2 > \dots > t\beta\ 1 > t\beta\ 2 > \dots > т\alpha\ 1 > \dots > т\delta; \\ O^A > O^B > \dots > O^Я; A_{O^{ta\ 1}} > \dots > A_{O^{ta\ \beta}} > B_{O^{t\beta\ 1}} > B_{O^{t\beta\ 2}} > \dots > Я_{O^{т\alpha\ 1}} > \dots > Я_{O^{т\alpha\ \gamma}}; \\ a_{O^{ta\ 1.1}} > \dots > a_{O^{ta\ 1.\phi}}; \dots; я_{O^{т\alpha\ 1.1}} > \dots > я_{O^{т\alpha\ 1.\epsilon}}. \end{aligned}$$

Учитывая, что между подмножествами C и O , а также их элементами существуют некоторые бинарные отношения, то можно записать следующие соответствия

$$\begin{aligned} E \subseteq C \times O; H^A \subseteq T^A \times O^A, \dots, H^Я \subseteq T^Я \times O^Я; \\ G^A \subseteq ta\ 1 \times A_{O^{ta\ 1}}, \dots, G^Я \subseteq т\alpha\ \delta \times Я_{O^{т\alpha\ \delta}}, \end{aligned}$$

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, $(ta\ 1, A_{O^{ta\ 1}}) \in G^A, \dots, (т\alpha\ \delta, Я_{O^{т\alpha\ \delta}}) \in G^Я$ - множество пар терминов и блоков вариантов его толкования.

В случае, когда одному термину соответствует несколько толкований можно записать, например для термина начинающегося с литеры A , $D^A \subseteq ta\ 1 \times (a_{O^{ta\ 1.1}}, \dots, a_{O^{ta\ 1.\phi}})$, 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

$$M^{об} = \langle \Pi, T^A, \dots, T^Я, O^A, \dots, O^Я, ta_1, \dots, тя_\delta, {}^A O^{та1}, \dots, O^{тя\delta}, <, > \rangle.$$

В таком компактном виде модель общенаучного словаря может быть использована для построения модели состояний интегрированного словаря.

По аналогии разработаем математическую модель специализированного словаря [9], корпус текстов которого показан на рис. 4.

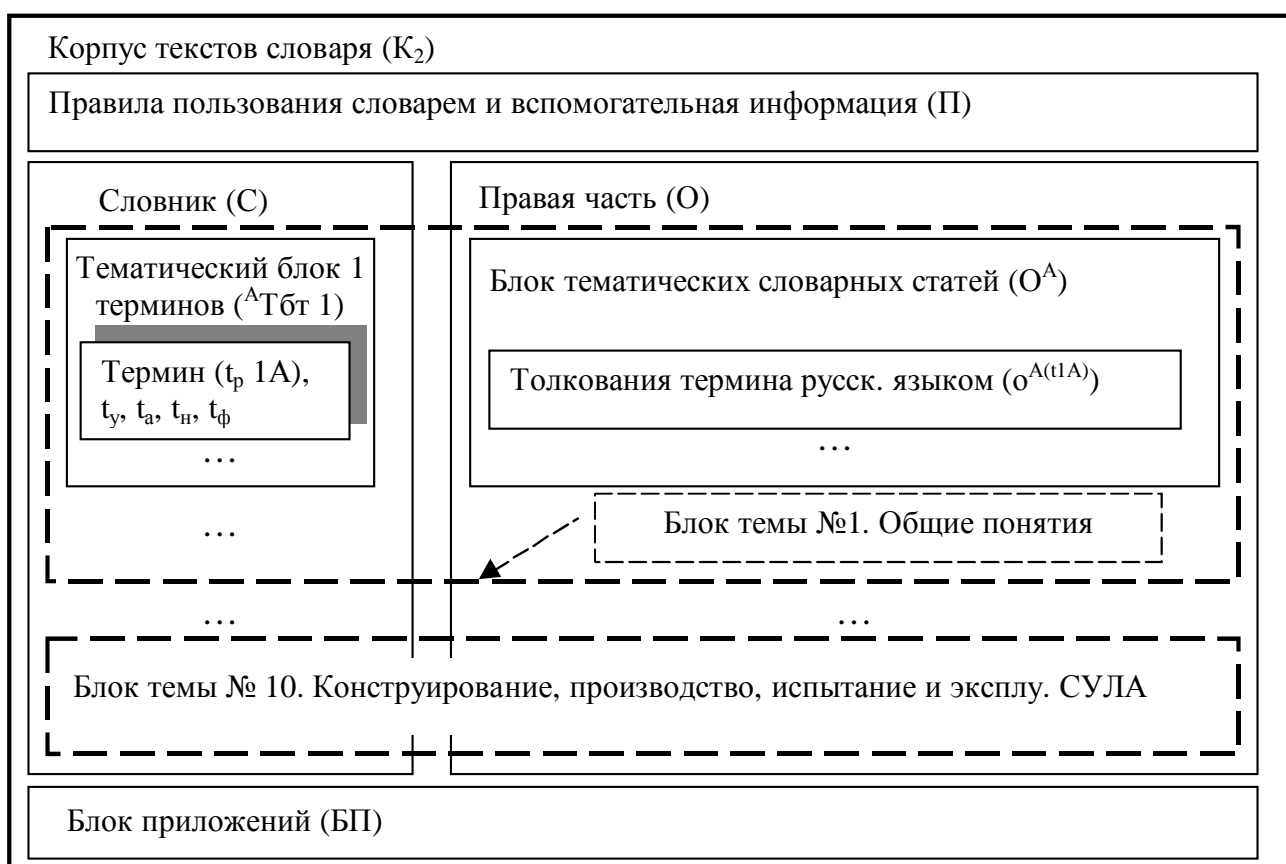


Рис. 4. Структура корпуса текстов специализированного словаря

Из рисунка видно, что структура словаря имеет отличия от структуры словаря приведенного на рис.3. Здесь имеются тематические блоки, термины которых упорядочены по алфавиту. Кроме того, терминам, записанным русским языком в левой части словаря, ставятся в соответствие термины в этой же части словаря, записанные на украинском (t_y), английском (t_a), немецком (t_n) и французском (t_f) языках.

По аналогии с формальным представлением общеобразовательного словаря запишем $K_2 = \{ П, С, О, БП \}$ - корпус текста специализированного словаря, состоящий из четырех подмножеств, которые имеют названия приведенные на рис.4. Отличительной особенностью данного корпуса является то, что он имеет тематические блоки терминов $\{ \overset{A}{Я}Т\bar{б}т\ i \}, i = \overline{1,10}$, где его элементы расположены в алфавитном порядке в пределах каждого блока. Кроме того, словник С содержит тематические блоки терминов, которые также расположенных в строгом порядке.

Каждый термин имеет аналог, записанный на четырех языках $t_{\triangleright} = \{ t_y, t_a, t_n, t_\phi \}$, где индекс \triangleright обозначает, что разноязычные термины между собой находятся в отношении строгого порядка $t_y > t_a > t_n > t_\phi$. Эту структурную единицу корпуса текстов будем называть блоком многоязычных терминов, и задавать простым их перечислением. Тогда справедлива обобщенная запись

$$\overset{A}{Я}Т\bar{б}т\ i = \{ (t_p\ 1A, t_{\triangleright}), \dots, (t_p\ NJ, t_{\triangleright}) \}, i = \overline{1,10},$$

где t_p - термин, записанный русским языком.

Факт соответствия элементов подмножества С подмножеству О зададим следующими формулами:

$U \subseteq C \times O, (C, O) \in U$ - пара устанавливающая соответствие левой части словаря С его правую часть О;

$R \subseteq \overset{A}{Я}Т\bar{б}т\ i \times \overset{A}{Я}О, (\overset{A}{Я}Т\bar{б}т\ i, \overset{A}{Я}О) \in R$ - множество пар устанавливающих соответствие между тематическими блоками терминов и тематическими блоками словарных статей;

$Q_k^i \subseteq (t_p\ 1A, t_{\triangleright}) \times o^{1A}$ - множество пар терминов с соответствующими им словарными статьями, где $k = \overline{1, d}$, d - количество терминов в i -ом тематическом блоке.

Отметим еще одну особенность данного словаря. В отдельных его словарных статьях имеются рисунки, схемы, чертежи, фотографии и т.д. Они представляют собой дидактический материал для создания у пользователей определенного образа, который соответствует некоторому явлению, процессу или устройству. Та-

кие дидактические элементы будем называть гештальтами и считать их специфическими лексическими единицами текста, которые связаны с текстом словарной статьи отношением включения. Кроме того, текст, который поясняет суть гештальта, также включен в основной текст словарной статьи. Поясним сказанное рис. 5.

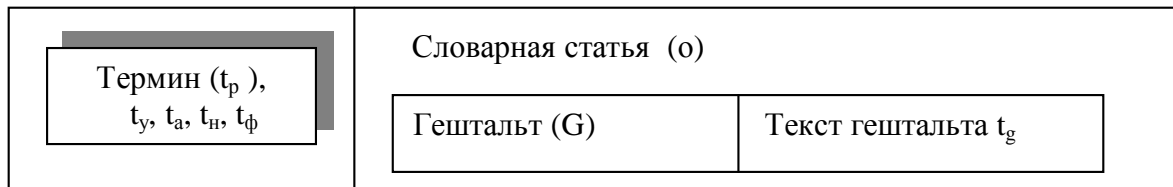


Рис.5. Структура текста словарной статьи с гештальтом

Тогда текст словарной статьи, содержащий гештальт, можно представить следующими соотношениями $L \subseteq G \times t_g, (G, t_g) \in o$, где L – пара гештальта и поясняющего его суть текста.

Последним элементом корпуса K_2 является блок приложений (БП), который представляет собой предметные указатели для терминов записанных различными языками. Формально запишем $БП = \{ \Omega_p, \Omega_y, \Omega_a, \Omega_n, \Omega_f \}$, где, например, $\Omega_p = \{ \omega_{a,б,...,я} \}$ - подмножество терминов записанных русским языком и расположенных в приложении в алфавитном порядке.

Обобщая полученный формализм, модель специализированного словаря, окончательно запишем следующую формулу

$$M^{сп} = \langle \Pi, \overset{A}{\underset{Я}{T}} \overset{A}{\underset{Я}{b}} t \overset{A}{\underset{Я}{i}}, \overset{A}{\underset{Я}{O}} \overset{A}{\underset{Я}{i}}, БП \rangle.$$

Предложенные выше формальные представления словарей различного типа базируются на методе моделирования письменных источников информации приведенном в работе [11], который является составной частью технологии формализации, разработанной в работе [12].

Такое представление словарей позволяет манипулировать их составными частями и создавать словарно-справочные средства любой конфигурации в зависимости от их целевого назначения.

Следующим шагом в создании ИОО ССС является формализация знаний лексикографов, которые являются «носителями» знаний о методах и способах по-

строения словарей различного типа. Однако эта задача выходит за рамки данной работы, и здесь рассматриваться не будет.

Таким образом, на основе анализа лексикографического обеспечения вуза поставлена задача создания интегрированных объектно-ориентированных словарно-справочных средств вуза. Для этого предлагается провести экспертизу лексикографических произведений и выявить словарно-справочные средства, которые в наибольшей степени отражали понятийную основу учебных дисциплин (блоков учебных дисциплин) конкретных учебных планов.

Электронные варианты словарей и справочников должны иметь специальную разметку в соответствии с предложенным методом представления словарей корпусами текстов. Совокупность, таким образом размеченных словарей, представляют собой специализированную лингвистическую базу данных, в которой размещаются данные в виде текстовой информации различного объема. Для манипулирования этими данными необходимо разработать эвристические правила (модели), которые бы смогли реализовать требования к ИОО ССС сформулированные в данной работе.

ЛИТЕРАТУРА

1. Метешкин К. Интеллектуальные информационные технологии в организации учебного процесса // Новый коллегіум, 2002, №1, С. 24-28.
2. Метешкин К.А., Шаронова Н.В. Использование гибридного интеллекта в учебном процессе высших учебных заведений // "Alma mater" ("Вестник высшей школы"). - 2001. - №11. С. 10 - 15.
3. Баранов А.Н. Введение в прикладную лингвистику: учебное пособие. - М.: Эдиториал УРСС, 2001. - 360 с.
4. Марчук Ю.Н. Основы компьютерной лингвистики. Учебное пособие. – М.: Изд-во МПУ «Народный учитель», 1999. - 221 с.
5. 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.
6. Большая Советская Энциклопедия. – М.: Советская энциклопедия, 1978.

7. Энциклопедический словарь бизнесмена: менеджмент, маркетинг, информатика / под общ. ред. М.И. Молдованова. Київ, «Техніка», 1993. – 854 с.
8. Педагогическая энциклопедия. В 4-х томах. – М.: Советская энциклопедия, 1968.
9. Словарь терминов по системам управления летательных аппаратов (СУЛА) / А.С. Кулик, А.Г. Гордон, В.Н. Картунов, В.Ф. Симонов, Ю.Н. Соколов. - Харьков: Нац. аэрокосмический ун-т. "ХАИ", 2001. - 224 с.
10. Русско-английский специализированный словарь, англо-русский специализированный словарь. Программа "Циклон" / Под ред. С.Н. Конюхова. Днепропетровск. ГКБ "Южное", 1999. - 335 с.
11. Метешкин К.А. Теоретические основы построения интеллектуальных систем управления учебным процессом в вузе: Монография. - Харьков: Экограф, 2000. - 278 с.
12. Белова Л.А., Метешкин К.А. Уваров О.В. Логико-математические основы управления учебными процессами вузов. - Харьков: Восточно-региональный центр гуманитарно-образовательных инициатив, 2001. - 272 с.