УДК 658.012.56

МОДЕЛИРОВАНИЕ МОРФОЛОГИИ ЕСТЕСТВЕННОГО ФЛЕКТИВНОГО ЯЗЫКА

Т.П. Любченко

Украинский языково-информационный фонд НАН украины, г. Киев, Украина, ltp@i.com.ua

Построена формальная модель словоизменения флективного языка. При моделировании процессов словоизменения использован формализм мультимножеств. Изложенная модель составляет концептуальную основу для компьютерного моделирования и реализации функции парадигматических отношений для определенного класса флективных языков. Практическая реализация модели выполнена на материале русского языка.

ЯЗЫК ЕСТЕСТВЕННЫЙ ФЛЕКТИВНОГО СТРОЯ, СЛОВОИЗМЕНЕНИЕ, ПАРАДИГМАТИЧЕСКИЙ ТИП, ГРАММАТИЧЕСКИЙ КЛАСС, ПАРАДИГМАТИЧЕСКИЙ КЛАСС, ГРАММАТИЧЕСКОЕ ЗНАЧЕНИЕ, СЛОВОИЗМЕНИТЕЛЬНАЯ ПАРАДИГМА, ОТНОШЕНИЕ ПАРАДИГМАТИЗАЦИИ, ОПЕРАТОР ПАРАДИГМАТИЗАЦИИ, ГРАММАТИЧЕСКАЯ ЛЕКСИКОГРАФИЧЕСКАЯ СИСТЕМА

Введение

Неотъемлемой составляющей автоматизированной системы обработки флективного естественного языка является подсистема морфологического анализа (морфологический анализатор), в результате работы которой для каждой текстовой словоформы определяется ее начальная форма, а также необходимые значения грамматических параметров, которые идентифицируют грамматический статус словоформы в тексте. Эффективность систем обработки в целом определяется, прежде всего, результатом работы морфологического анализатора. Например аннотирование текстов лингвистических корпусов без использования морфологического анализатора вообще немыслимо.

Существуют разные схемы построения морфологических анализаторов. Среди большого количества работ, посвященных этой теме, можно выделить три основных подхода: подход, основывающийся на использовании больших словарей (так называемая четкая морфология) [1, 4, 5, 7, 8, 11], подход на основе правил (нечеткая морфология) [14, 15] и вероятностный подход [16]. Первый из них позволяет получать точные результаты морфологического разбора (но лишь при условии, если анализируемое слово имеется в словаре). Второй позволяет строить гипотезы относительно возможного грамматического статуса слов, «незнакомых» системе. Третий подход эффективен для языков с четким порядком слов в предложении. Сочетаемость первого и второго подходов разрешает получать точный и полный морфологический анализ [14, 18]. Таким образом, морфологические анализаторы целесообразно основывать на использовании больших морфологических или грамматических словарей, что обеспечивает морфологический анализ с высокой точностью (см. например, [1], [4], [8]).

В Украинском языково-информационном фонде НАН Украины (УЯИФ) алгоритмы морфологического анализа ориентированы на использование грамматических словарей [1, 11]. Электронные грамматические словари разработаны для нескольких языков (украинского, русского, английского, немецкого, испанского, французского и турецкого). Словари разрабатывались на основе теории лексикографических систем [2, 3]. Они ориентированы на письменные варианты языков и предназначены, прежде всего, для использования их в системах автоматической обработки текстов. Помимо этого, предусмотрено предоставление пользователю возможности доступа к словарю как к справочно-информационной системе (поиск слов, предоставление информации относительно словоизменения определенных реестровых единиц, их грамматической параметризации и тому подобное).

При создании грамматической лексикографической системы флективного языка определяющим является наличие формализованной модели словоизменения лексики, которая, в свою очередь, означает построение определенной классификации (то есть разбиения лексического состава по определенным параметрам на классы слов, внутри которых словоизменение происходит согласно единому правилу). Заметим, что модели и действующие автоматизированные системы словоизменения для русского [4–8, 14–19] и украинского языков [12, 13] построены уже довольно давно. Они изложены, в частности, и в наших работах [1, 3, 20, 21]. Тем не менее, корректной формальной модели словоизменения нам до сих пор не попадалось. Подчеркнем, что речь идет именно о формальной модели, которая требует для своего формулирования применения стандартных математических объектов, например, теоретико-множественных. В то же время, применение теоретико-множественных объектов, так сказать, в лоб натыкается на определенные формальные преграды. Ведь первым шагом при формулировании модели есть определение некоего «множества слов», что на самом деле есть и первым шагом математического «грехопадения», поскольку, как правило, совокупности рассматриваемых слов, строго говоря, не являются множествами – они содержат повторяющиеся элементы, что является следствием грамматической омонимии. Поэтому мы сочли необходимым публикацию работы, в которой наведен математический порядок и соблюдены необходимые условия корректности.

1. Построение формальной модели

Построение формальной модели словоизменения флективного языка требует установления и формализации лингвистических критериев, согласно которым лексический состав языка разбивается на определенные непересекающиеся совокупности, внутри каждой из которых словоизменение происходит в соответствии с одинаковыми правилами. Совокупности слов с такими свойствами назовем словоизменительными (или парадигматическими) классами. При формализации процессов словоизменения мы будем использовать формализмом мультимножеств[1].

Моделирование распределения языкового лексикона на парадигматические классы осуществляется в несколько этапов. На первом определяется понятие парадигматического типа; при этом принципиальную роль играют понятия грамматической категории, грамматического значения и грамматической формы [9].

Пусть L – некоторый фиксированный язык (для примера будем рассматривать русский);

A – множество[2] слов языка L, которое является порождающим множеством для мультимножества (ММ) W, представляющего лексикон рассматриваемого языка;

Ω – множество грамматических значений (значения грамматических категорий таких как род, число, падеж, залог, наклонение, время, лицо);

$\Omega(T_i)$ – множество грамматических значений, которые отвечают типу T_i.

Рассмотрим множество грамматических значений детальнее.

Элементами множества грамматических значений (значений грамматических категорий) русского языка $\Omega = \{\Omega^1, \Omega^2, \Omega^3, \Omega^4, \Omega^5, \Omega^6, \Omega^7, \Omega^8\}$ являются следующие подмножества:

$\Omega^1 = \{\omega_1^1, \omega_2^1, \omega_3^1\}$ ≡ {*м.р., ж.р., с.р.*} – множество значений грамматической категории *род*;

$\Omega^2 = \{\omega_1^2, \omega_2^2\}$ ≡ {*единственное число*, *множественное число*} – множество значений грамматической категории *число*;

$\Omega^3 = \{\omega_1^3, \omega_2^3, \omega_3^3, \omega_4^3, \omega_5^3, \omega_6^3\}$ ≡ {*именительный, родительный, дательный, винительный, творительный, предложный*} – множество значений грамматической категории *падеж*;

$\Omega^4 = \{\omega_1^4, \omega_2^4\}$ ≡ {*действительный, страдательный*} – множество значений грамматической категории *залог*;

$\Omega^5 = \{\omega_1^5, \omega_2^5, \omega_3^5\}$ ≡ {*изъявительное, сослагательное, повелительное*} – множество значений грамматической категории *наклонение*;

$\Omega^6 = \{\omega_1^6, \omega_2^6, \omega_3^6\}$ ≡ {*настоящее, прошлое, будущее*} – множество значений грамматической категории *время*;

$\Omega^7 = \{\omega_1^7, \omega_2^7, \omega_3^7\}$ ≡ {1 *лицо*, 2 *лицо*, 3 *лицо*} – множество значений грамматической категории *лицо.*

$\Omega^8 = \{\omega_1^8, \omega_2^8, \omega_3^8\}$ ≡ {*совершенный вид, несовершенный вид, совершенный и несовершенный вид*} – множество значений грамматической категории *вид.* Или в обозначениях А. Зализняка [10]: $\Omega^8 = \{\omega_1^8, \omega_2^8, \omega_3^8\}$ = {*св, нсв, св-нсв*}.

Множество грамматических значений может быть представлено в виде графа (рис. 1).

Определенные комбинации определенных грамматических значений (элементов множеств Ω^i) характеризуют словоизменительную форму слова. Например словоизменительные формы лексемы «*дирижёр*» (которая является существительным) характеризуются комплексом грамматических значений, состоящим из пар элементов множеств значений грамматических категорий *число* и *падеж*: $\{\langle\omega_i^2, \omega_j^3\rangle\}$, $i = 1, 2$; $j = 1,2,...,6$. Каждая пара значений ⟨*число, падеж*⟩ отвечает определенной форме лексемы из ее словоизменительной парадигмы: $\langle\omega_1^2, \omega_1^3\rangle$ – *дирижёр*, $\langle\omega_1^2, \omega_2^3\rangle$ – *дирижёра*, ..., $\langle\omega_1^2, \omega_5^3\rangle$ – *дирижёром*, ..., $\langle\omega_2^2, \omega_6^3\rangle$ – *дирижёрах.* Грамматические формы глагола характеризуются другим комплексом грамматических значений, который состоит из кортежей (из n-*ок*, где n может принимать значение от 1 до 6) элементов множеств грамматических значений категорий *залог*, *наклонение*, *время*, *лицо*, *число*, *род.* Например словоизменительная парадигма глагола «*произойти*» состоит из 15 словоформ[3], которые определяются таким комплексом грамматических значений:

$$\{\omega_1^V, \omega_2^V, \omega_3^V, ..., \omega_{15}^V\} = \{\langle\omega_i^4, \omega_j^5, \omega_j^6, \omega_i^2, \omega_j^7, \omega_j^1\rangle_k\},$$

$$i = 1,2; j = 1,2,3; k = 1,2,...,15.$$

[1] Мультимножество или множество с повторяющимися элементами как новое математическое понятие, насколько нам известно, впервые встречалось еще в работах Д. Кнута [22, С. 498]. В последние годы появилась серия работ А.Б. Петровского [23-29], посвященных развитию теории мультимножеств и проблемам применения данной теории для принятия решений в условиях нечеткой исходной информации, в кластерном анализе многопризнаковых объектов, объектов с противоречивыми свойствами, в сетях Петри и тому подобное.

[2] Множество (в данном случае совокупность слов языка), состоящее из элементов (слов), среди которых нет повторяющихся. Это множество порождает мультимножество слов языка (или множество с повторяемыми элементами).

[3] Потенциально глагольная парадигма в общем случае может состоять из большего количества форм – в нашей модели максимально возможное количество разных форм – 45 (см. [1, 21]).

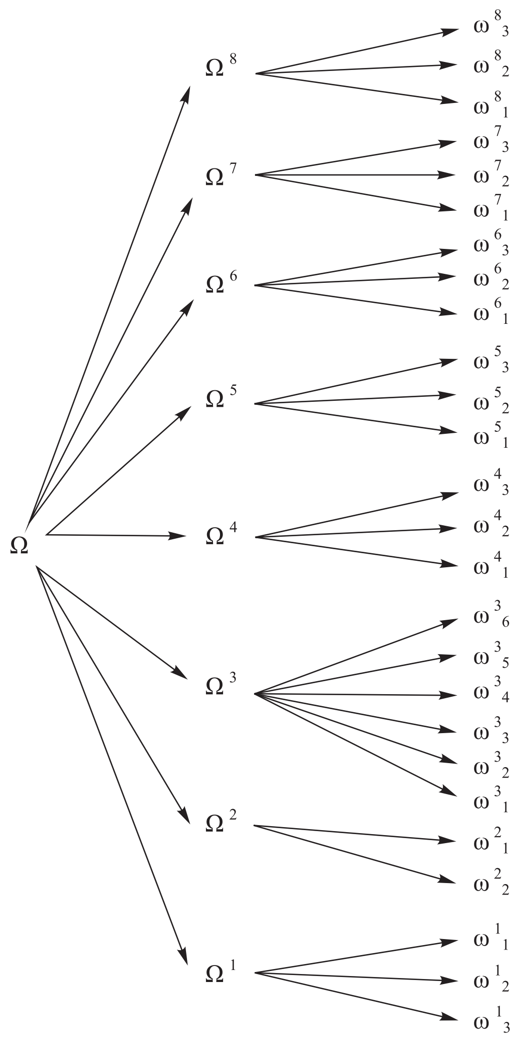

Рис. 1. Множество грамматических значений

Причем значение категории рода не определено для $k=2, 3, 4, 5, 6, 7, 12, 13, 14$, значение категории числа не определено для $k=14$, значение категорий лица и наклонения не определены для $k=14, 15$, а для инфинитивной формы не определенными являются все грамматические значения.

Каждой из *n-ок* грамматических значений отвечает определенная парадигматическая форма рассматриваемой лексемы: ω_1^V – *произойти*, $\omega_2^V = \langle\omega_1^4,\omega_1^5,\omega_3^6,\omega_1^2,\omega_1^7\rangle$ – *произойду*, ..., $\omega_8^V =$ $= \langle\omega_1^4,\omega_1^5,\omega_2^6,\omega_1^2,\omega_{1,2,3}^7,\omega_2^1\rangle$ – *произошла*, $\omega_{11}^V =$ $= \langle\omega_1^4,\omega_1^5,\omega_2^6,\omega_1^2,\omega_{1,2,3}^7,\omega_{1,2,3}^1\rangle$ – *произошли*, ..., $\omega_{15}^V =$ $= \langle\omega_1^4,\omega_2^6,\omega_1^2\rangle$ – *происшедший*.[4]

Как видим, словоизменительная парадигма слов из разных классов определяется своим индивидуальным набором словоизменительных категорий.

В соответствии со словоизменительными категориями, определяющими словоизменительную парадигму конкретных слов, вводятся парадигматические типы (ПТ). Пусть T_i ($i=1, 2,..., N$) – парадигматические типы, N – количество парадигматических типов; $W(T_i)$ – слова языка L, имеющие парадигматический тип T_i. Так, что $W=\bigcup_{i=1}^{N} W(T_i)$. Заметим, что $W(T_i)$, как, собственно, и весь лексический состав языка, является мультимножеством (ММ), или множеством с повторяющимися элементами. То есть, на самом деле, $W(T_i)=\{(n_{T_i}(w_j)\bullet w_j) \mid w_j\in A, n_{T_i}(w_j)\in Z^+\}$, где $n_{T_i}(w_j)$ называется функцией числа экземпляров элемента w_j в ММ $W(T_i)$ и определяет, сколько раз элемент $w_j\in A$ входит в ММ $W(T_i)$; символ $\bullet$ обозначает кратность вхождения элемента в соответствующую компоненту ММ. Элементы w_j мультимножества $W(T_i)$, для которых $n_{T_i}(w_j)>1$, это слова, тождественные графемно, но имеющие определенные различия в своих грамматических и семантических свойствах.

Подробное описание парадигматических типов русского языка представлено в наших публикациях [1, с. 128-156, с. 218-223], [21]. Для изменяемых слов русского языка определяется четыре парадигматических типа: субстантивный, адъективный, глагольный и тип количественных числительных, а для неизменяемых слов — так называемый «нулевой парадигматический тип».

Каждый парадигматический тип характеризуется определенным набором грамматических значений.

Субстантивный парадигматический тип характеризуется комплексом значений грамматических категорий *число* (Ω^2) и *падеж* (Ω^3):

$$\Omega(T^S)\subseteq\Omega^2\times\Omega^3=\{\langle\omega_i^2,\omega_j^3\rangle\},$$
$$\omega_i^2\in\Omega^2, i=1,2;\ \omega_j^3\in\Omega^3, j=1,2,...,6 \quad (1)$$

Категория рода не включается в число дифференцирующих словоизменительных категорий.

Субстантивный парадигматический тип может рассматриваться как характеризующийся мультимножеством, полученным (порожденным) из множеств Ω^2 и Ω^3 значений грамматических категорий число и падеж:

$$\Omega(T^S)\equiv\Omega^S$$

$$=\{2\bullet\omega^2, 6\bullet\omega^3 \mid \omega^2=\{\omega_1^2,\omega_2^2\}, \omega^3=\{\omega_1^3,\omega_2^3,...,\omega_6^3\}\}=$$

$$=\{6\bullet\omega_1^2, 6\bullet\omega_2^2, 2\bullet\omega_1^3, 2\bullet\omega_2^3, 2\bullet\omega_3^3, 2\bullet\omega_4^3, 2\bullet\omega_5^3, 2\bullet\omega_6^3\}.$$

Графическое изображение ММ грамматических значений субстантивного парадигматического типа продемонстрировано на рис. 2.

[4] Нижние индексы в записи $\omega^7_{1,2,3}$ означают, что у соответствующей парадигматической словоформы грамматическая категория лицо может принимать какое-либо из трёх возможных значений – первое, второе или третье лицо (я, ты, она — пришла; я, ты, он — пришёл; я, ты, оно — пришло; мы, вы, они — пришли). В записи $\omega^1_{1,2,3}$ нижние индексы означают возможность какого-либо (одного или комбинации нескольких) из значений грамматической категории род.

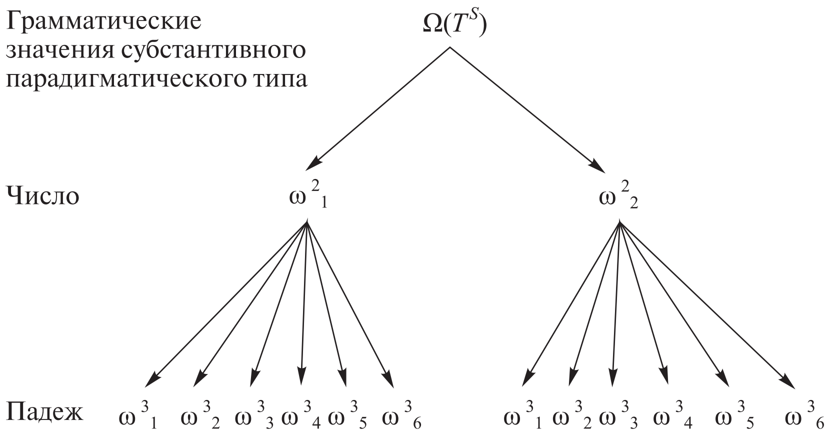

Рис. 2. Отношение между грамматическими значениями субстантивного парадигматического типа

По субстантивному парадигматическому типу происходит словоизменение у существительных и местоимений-существительных.

Адъективный парадигматический тип характеризуется комплексом грамматических категорий *род* (Ω^1), *число* (Ω^2) и *падеж* (Ω^2):

$$\Omega(T^A) \subseteq \Omega^1 \times \Omega^2 \times \Omega^3 = \{\langle \omega^1_i, \omega^2_j, \omega^3_k \rangle\} ,$$

где $\omega^1_i \in \Omega^1, i = 1,2,3; \omega^2_j \in \Omega^2, j = 1,2;$

$$\omega^3_k \in \Omega^3, k = 1,2,...6 . \qquad (2)$$

Графическое представление отношений между грамматическими значениями адъективного парадигматического типа изображено на рис. 3.

По адъективному парадигматическому типу происходит словоизменение у прилагательных, причастий, местоимений-прилагательных и порядковых числительных.

Грамматические значения глагольного парадигматического типа определяются значениями грамматических категорий *залог* (Ω^4), *наклонение* (Ω^5), *время* (Ω^6), *лицо* (Ω^7), *число* (Ω^2), *род* (Ω^1):

$$\Omega(T^V) \subseteq \Omega^4 \times \Omega^5 \times \Omega^6 \times \Omega^7 \times \Omega^2 \times \Omega^1 . \qquad (3)$$

Графическое представление отношений между грамматическими значениями глагольного парадигматического типа дано на рис. 4–6.

Парадигматический тип количественных числительных характеризуется грамматическими значениями категории *падеж* (Ω^3):

$$\Omega(T^C) \subseteq \Omega^3 . \qquad (4)$$

Графическое представление отношений между грамматическими значениями парадигматического типа количественных числительных изображено на рис. 7.

Обозначим P_j , ($j = 1,2,..., p$) – лексико-грамматические классы, где p – количество лексико-грамматических классов для данного языка; $W(P_j)$ – слова языка L , принадлежащие лексико-грамматическому классу P_j .

По признаку принадлежности к определенной части речи и по дополнительным признакам, являющимся классифицирующими (не словоизменительными) в пределах определенной части речи, ММ слов W распределяем на подмультимножества, которые будем называть *грамматическими классами* (обозначим их P_j), следующим образом.

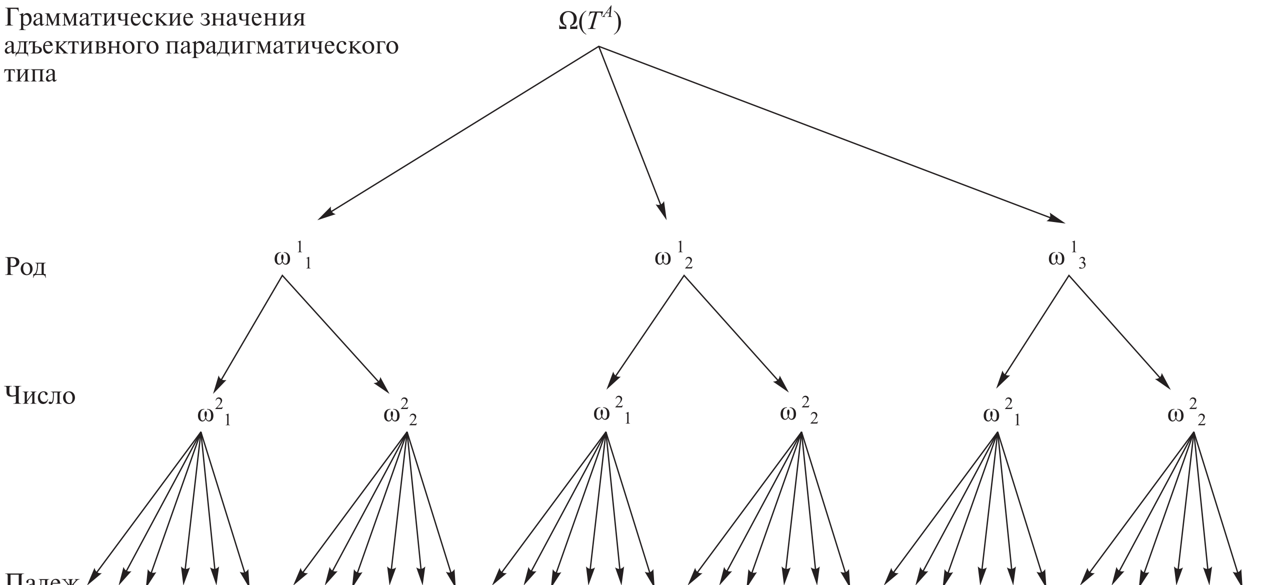

Рис. 3. Отношение между грамматическими значениями адъективного парадигматического типа

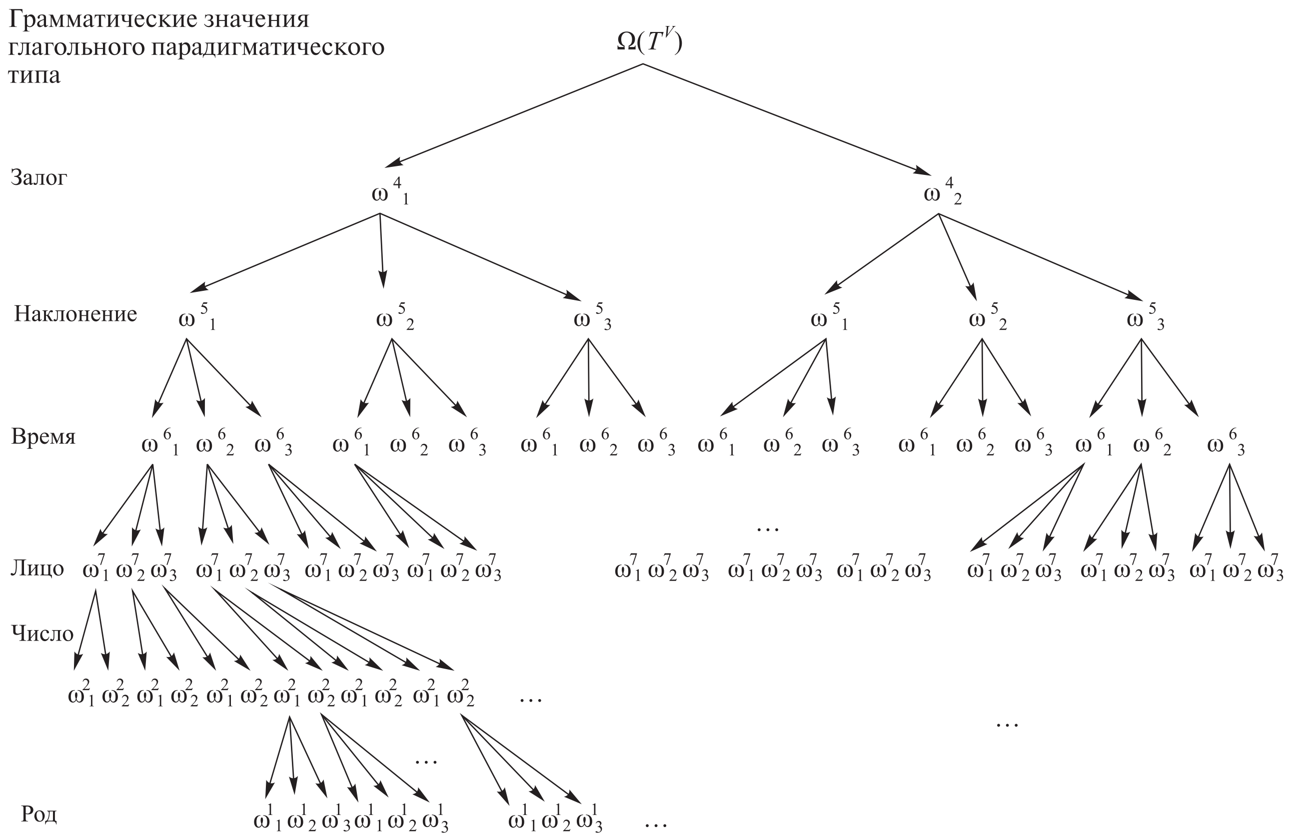

Рис. 4. Отношение между грамматическими значениями глагольного парадигматического типа

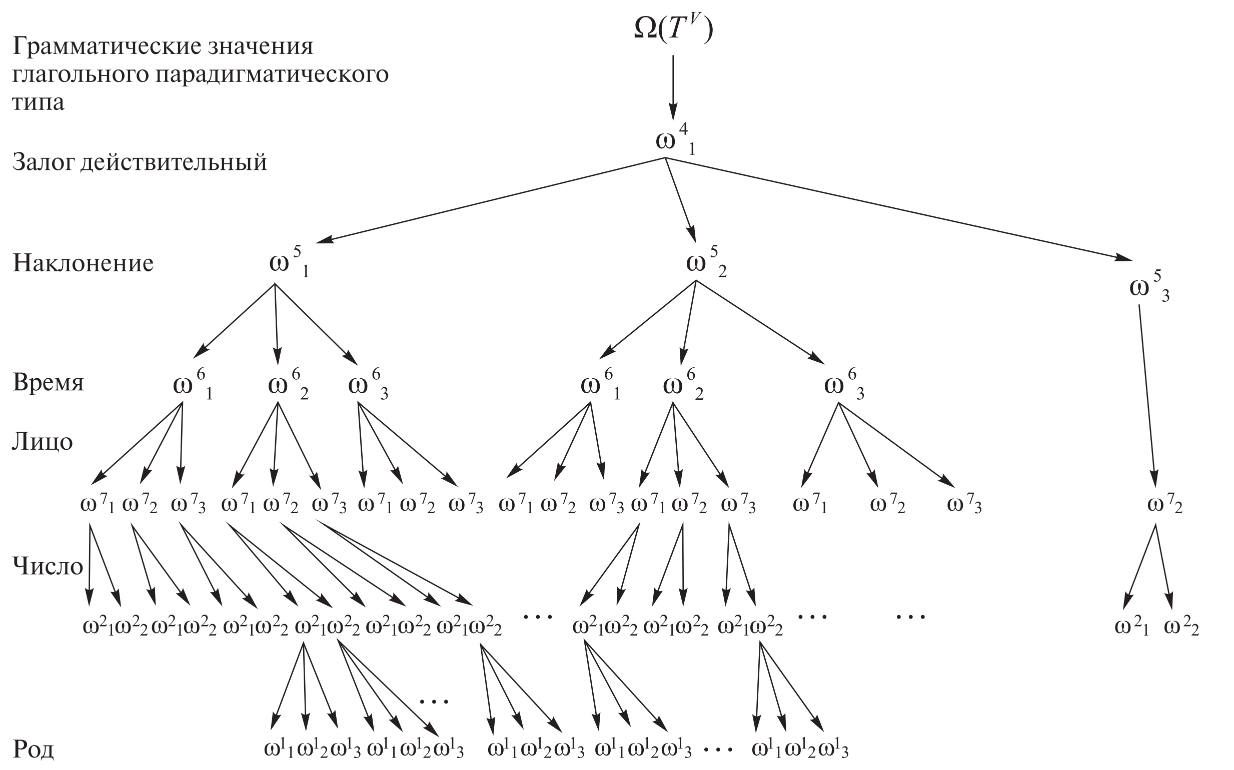

Рис. 5. Отношение между грамматическими значениями глагольного парадигматического типа (действительный залог)

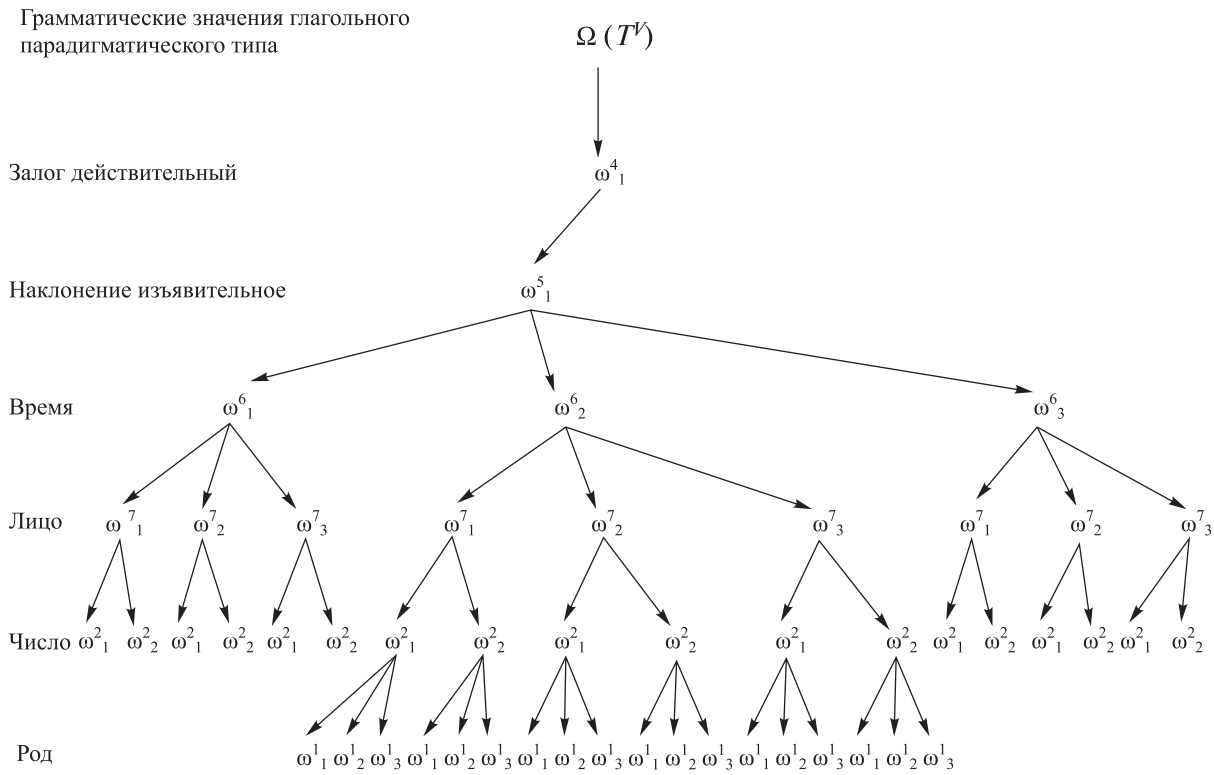

Рис. 6. Отношение между грамматическими значениями глагольного парадигматического типа (действительный залог, изъявительное наклонение)

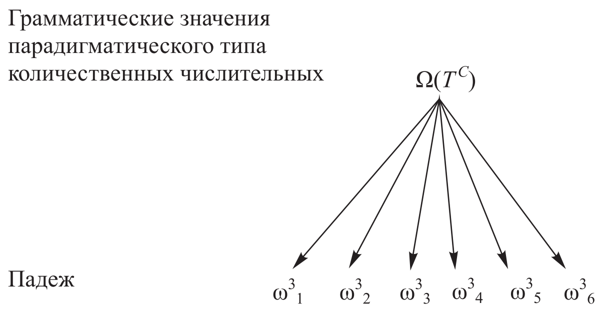

Рис. 7. Отношение между грамматическими значениями парадигматического типа количественных числительных

Существительные по значению грамматической категории «*род*» (которая в пределах этой части речи является классифицирующим признаком) [10] распределяются на 3 грамматических класса: существительные мужского рода, существительные женского и существительные среднего рода; множественные существительные (существительные pluralia tantum) образуют отдельный грамматический класс. Таким образом, существительные составляют 4 грамматических класса, которые в последующем изложении будем обозначать соответственно P_1, P_2, P_3 и P_4.

Глаголы по значению грамматической категории «*вид*» (которая рассматривается нами как классифицирующая) [10] распределяются на такие три грамматических класса: глаголы совершенного вида (св), глаголы несовершенного вида (нсв) и двувидовые глаголы (св-нсв).

Местоимения разделяются на два грамматических класса: местоимения-прилагательные и местоимения-существительные .

Грамматический класс адъективов составляют прилагательные и порядковые числительные (то есть этот грамматический класс образован путем объединения).

Все другие слова из ММ W слов языка отнесены к своим грамматическим классам по признаку принадлежности к конкретной части речи (то есть в этом случае понятие «грамматический класс» совпадает с понятием «часть речи»).

Таким образом, в русском языке нами выделены грамматические классы, которые мы представляем в табл. 1.

$$W = \bigcup_{j=1}^{19} W(P_j)\,.$$

Определенный парадигматический тип может быть присущ нескольким грамматическим классам, внутри каждого из которых выделяются парадигматические классы.

Схема отношений между парадигматическими типами, грамматическими классами и парадигматическими классами представлена на рис. 8.

Таблица 1

Грамматические классы русского языка

№	Грамматический класс	Символ ГК
1	Существительные мужского рода	P_1
2	Существительные женского рода	P_2
3	Существительные среднего рода	P_3
4	Множественные существительные	P_4
5	Адъективы (прилагательные+порядковые числительные)	P_5
6	Глаголы совершенного вида	P_6
7	Глаголы несовершенного вида	P_7
8	Двувидовые глаголы	P_8
9	Причастия	P_9
10	Местоимения (местоимения-существительные)	P_{10}
11	Местоимения-прилагательные	P_{11}
12	Числительные количественные	P_{12}
13	Наречия	P_{13}
14	Междометия	P_{14}
15	Союзы	P_{15}
16	Частицы	P_{16}
17	Предлоги	P_{17}
18	Предикативы	P_{18}
19	Сокращения	P_{19}

Как видно из схемы, согласно принципу распределения:

$$W(T_i) = \bigcup_{j=1}^{p_i} W(P_j), \qquad (5)$$

где $P_j \subseteq T_i$, p_i – количество грамматических классов, в которых словоизменение имеет парадигматический тип T_i.

$$W(P_j) = \bigcup_{k=1}^{n_j} W(\Pi_k), \qquad (6)$$

где $\Pi_k \subseteq P_j \subseteq T_i$; n_j – количество парадигматических классов грамматического класса P_j, в которых словоизменение имеет парадигматический тип T_i.

В соответствии с принципами, изложенными в наших предыдущих работах [1, 11, 20, 21], внутри грамматических классов выделяются парадигматические классы, строятся отношения парадигматизации и оператор парадигматизации, а каждая лексема *x* приобретает представление в виде комбинации ее квазиосновы и квазифлексии:

$$x = c(x) * f(x), \qquad (7)$$

где $c(x)$ – часть лексемы *x*, которая в процессе словоизменения остается неизменной (квазиоснова); $f(x)$ – ее изменяемая составляющая (квазифлексия); * – конкатенация, а словоизменительная парадигма представляется в виде:

$$\pi(x) = c(x) * \{f_j(x)\}, \qquad (8)$$

где $f_j(x)$; $j = 0,1,2,...,n(T_i)$ – квазифлексии в соответствующих грамматических формах. Например лексема *x* =‘*лёд*’ представляется в виде *x* =‘*л*’ * ‘*ёд*’, где ‘*л*’ – квазиоснова, а ‘*ёд*’ квазифлексия начальной формы лексемы. Ее словоизменительная парадигма:

$\pi(x) = c(x) * \{f_i(x)\}\}$ = ‘*л*’*{‘*ёд*’; ‘*ьда*’, ‘*ьду*’; ‘*ьду*’; ‘*ёд*’; ‘*ьдом*’; ‘*ьду*’,‘*ьде*’; ‘*ьды*’; ‘*ьдов*’; ‘*ьдам*’; ‘*ьды*’; ‘*ьдами*’; ‘*ьдах*’}.[5]

Отношения парадигматизации π_i строятся на декартовых произведениях $P_i \times P_i$, i = 1,2,...,19, (то есть построение оператора парадигматизации выполняется на мультимножестве лексем каждого грамматического класса отдельно) . Как было указано выше, в нашей системе имеется 19 грамматических классов, а, следовательно, соответственно столько же отношений парадигматизации π_i вида:

$$\forall x^1, x^2 \in P_i \;\; x^1 \pi_i x^2 : \; x^1 = c(x^1) * f^k, \quad x^2 = c(x^2) * f^k, \; f^k \in [F]^k, \qquad (9)$$

где $[F]^k$ состоит из квазифлексий слов, имеющих во всех своих грамматических значениях $\omega_1, \omega_2, ..., \omega_{n(T_i)}$ (парадигматический тип T_i) одинаковые изменяемые составляющие. Обозначим $C = \{n(c_i) \bullet c_i\}$ – ММ квазиоснов, $i = \overline{1, n_{co}}$ – индекс на обозначение уникального номера квазиосновы, n_{co} – количество квазиоснов в словаре; $F = \{n(f_i) \bullet f_i\}$ – ММ квазифлексий, $j = \overline{1, n_f}$ – индекс на обозначение уникального номера квазифлексии, n_f – количество квазифлексий в словаре.

Отношение парадигматизации является отношением эквивалентности, поскольку оно, очевидно, рефлексивно, симметрично и транзитивно. Фактор-множество P_i / π_i составляет множество парадигматических классов грамматического класса P_i (парадигматического типа T_i): $P_i = \bigcup_{j=1}^{n} \Pi_j$. К одному парадигматическому классу относятся слова, имеющие одинаковые наборы квазифлексий для всех грамматических форм, а отличаются один от другого лишь неизменяемой составляющей $c(x)$.

В каждом из грамматических классов (парадигматических типов T_i) построим разбиение на парадигматические классы и рассмотрим все возможные отношения между элементами указанных выше множеств Ω, T, P, C, F, Π .

Отношение между грамматическими значениями определенных парадигматических типов описаны выше (1) – (4). Далее приведем другие отношения.

[5] Квазифлексии в разных граматических формах записаны через точку с запятой; если для граматической формы существует несколько вариантов словоформ, то соответствующие квазифлексии поданы через запятую.

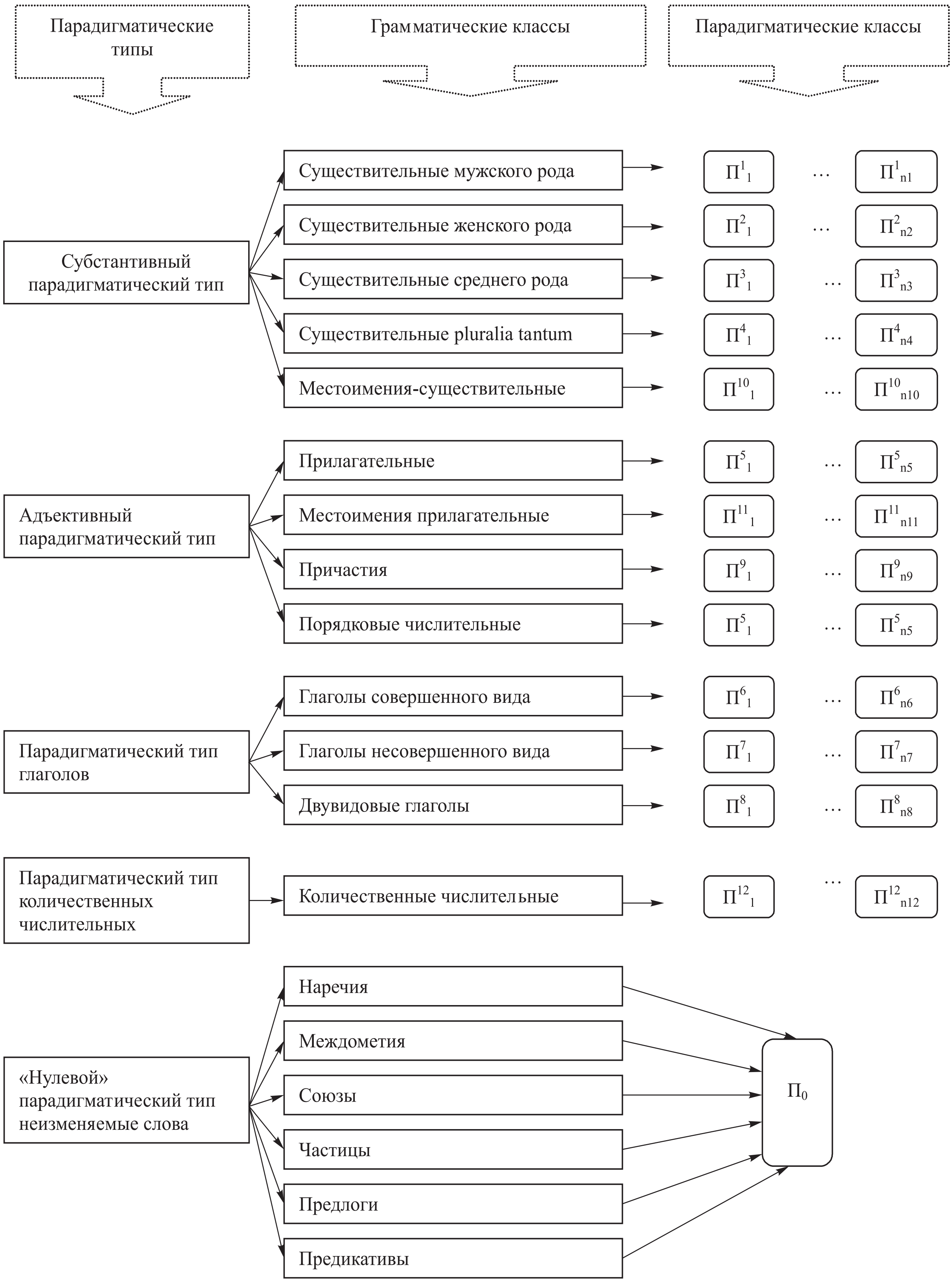

Рис. 8. Схема отношений между парадигматическими типами, грамматическими классами и парадигматическими классами

Отношение «парадигматический тип»:

$$T \subseteq T^S \cup T^A \cup T^C \cup T^0 \quad (10)$$

Отношение «грамматическое значение»:

$$\Omega(T) \subseteq \Omega(T^S) \cup \Omega(T^A) \cup \Omega(T^C)\,. \quad (11)$$

Отношение «парадигматический тип – грамматический класс»:

$$R_{TP} \subseteq T \times P : \{\langle T_i, P_j \rangle\}\,, \quad (12)$$

где $i = 1,2,3,4$; $j = 1,2, ..., 19$;

$$T_i = \{T_0 = T^0, T_1 = T^S, T^2 = T^A, T_3 = T^V, T_4 = T^C\}\,.$$

Соответствие между парадигматическими типами и грамматическими классами хорошо видно на рис. 8.

$$R_{TP} \subseteq T \times P: \ \{\langle T_1,P_1\rangle, \langle T_1,P_2\rangle, \langle T_1,P_3\rangle, \langle T_1,P_4\rangle, \langle T_1,P_{10}\rangle, \langle T_2,P_5\rangle, \langle T_2,P_9\rangle, \langle T_2,P_{11}\rangle, \langle T_3,P_6\rangle, \langle T_3,P_7\rangle, \langle T_3,P_8\rangle, \langle T_4,P_{12}\rangle, \langle T_0,P_{13}\rangle, \langle T_0,P_{14}\rangle, \langle T_0,P_{15}\rangle, \langle T_0,P_{16}\rangle, \langle T_0,P_{19}\rangle\} \quad (13)$$

Отношение «грамматический класс – грамматическое значение»:

$$R_{P\Omega(T)} \subseteq P \times \Omega(T) : \{\langle P_j, \Omega(T_i)\rangle\}\,, \quad (14)$$

где j=1,2, ..., 19; $T_0 = T^0$; $T_1 = T^S$; $T^2 = T^A$; $T_3 = T^V$; $T_4 = T^C$.

$$R_{P\Omega(T)} \subseteq P \times \Omega(T): \ \{\langle P_1,\Omega(T_1)\rangle, \langle P_2,\Omega(T_1)\rangle, \langle P_3,\Omega(T_1)\rangle, \langle P_4,\Omega(T_1)\rangle, \langle P_{10},\Omega(T_1)\rangle, \langle P_5,\Omega(T_2)\rangle, \langle P_9,\Omega(T_2)\rangle, \langle P_{11},\Omega(T_2)\rangle, \langle P_6,\Omega(T_3)\rangle, \langle P_7,\Omega(T_3)\rangle, \langle P_8,\Omega(T_3)\rangle, \langle P_{12},\Omega(T_4)\rangle, \langle P_{13},\Omega(T_0)\rangle, \langle P_{14},\Omega(T_0)\rangle, \langle P_{15},\Omega(T_0)\rangle, \langle P_{16},\Omega(T_0)\rangle, \langle P_{17},\Omega(T_0)\rangle, \langle P_{18},\Omega(T_0)\rangle, \langle P_{19},\Omega(T_0)\rangle\}\,. \quad (15)$$

Отношение «лексема – грамматический класс»:

$$R_{WP} \subset W \times P\,. \quad (16)$$

Например *'стол'* $\in P_1$, *'вишня'* $\in P_2$, *'слово'* $\in P_3$, *'ножницы'* $\in P_4$, *'горячий'* $\in P_5$, *'выучить'* $\in P_6$, *'учить'* $\in P_7$, *'автомтизировать'* $\in P_8$, ...

Отношение «лексема – грамматическое значение»:

$$R_{W\Omega(T)} \subset W \times \Omega(T)\,. \quad (17)$$

Например для лексемы *'слово'*, принадлежащего к классу существительных среднего рода (*'слово'* $\in P_3$) и характеризуется набором грамматических значений $\Omega(T_1)$, отношение $R_{W\Omega(T)}$ имеет вид:

$$\forall w \in P_3 : R_{W\Omega(T)} \subset [w] \times \Omega(T_1) = [w] \times \Omega_2 \times \Omega_3 :$$

{⟨ *'слово'*, ⟨ ед., им. ⟩ ⟩, ⟨ *'слова'*, ⟨ ед.ч., род. ⟩ ⟩, ⟨ *'слову'*, ⟨ ед.ч., дат. ⟩ ⟩, ⟨ *'слово'*, ⟨ ед.ч., вин. ⟩ ⟩, ⟨ *'словом'*, ⟨ ед.ч., твор. ⟩ ⟩, ⟨ *'слове'*, ⟨ ед.ч., предл. ⟩ ⟩, ⟨ *'слова'*, ⟨ мн.ч., им. ⟩ ⟩, ⟨ *'слов'*, ⟨ мн.ч., род. ⟩ ⟩, ⟨ *'словам'*, ⟨ мн.ч., дат. ⟩ ⟩, ⟨ *'слова'*, ⟨ мн.ч., вин. ⟩ ⟩, ⟨ *'словами'*, ⟨ мн.ч., твор. ⟩ ⟩, ⟨ *'словах'*, ⟨ мн.ч., предл. ⟩ ⟩}.

Проекция отношения $R_{W\Omega(T)}$ по первому компоненту $\pi_W(R_{W\Omega(T)})$ есть мультимножество (в словоизменительной парадигме имеются словоформы, которые графемно совпадают, но имеют различающиеся грамматические характеристики. Одинаковая форма слова в именительном и винительном падежах единственного числа – *'слово'*, а форма *'слова'* в парадигме фигурирует трижды – в родительном падеже единственного числа, в именительном и винительном падежах множественного числа:

$\pi_W(R_{W\Omega(T)})$ ={2·(*'слово'*), (*'слова'*), (*'слову'*), (*'словом'*), (*'слове'*), 3·(*'слова'*), (*'слов'*), (*'словам'*), (*'словами'*), (*'словах'*)}).

Отношение «лексема – грамматический класс – грамматическое значение»:

$$R_{WP\Omega(T)} \subset W \times P \times \Omega(T)\,. \quad (18)$$

Учитывая формулу (7), согласно которой лексема представляется в виде комбинации неизменяемой и изменяемой составляющих (квазиосновы и квазифлексии), рассмотрим отношение «квазиоснова – квазифлексия – грамматический класс – грамматическое значение»:

$$R_{CFP\Omega(T)} \subset C \times F \times P \times \Omega(T)\,. \quad (19)$$

Распишем его детальнее:

$$\left.\begin{array}{l} \langle c^{j_{Pi}}, f_0^{\,j_{Pi}}, P_i, {\omega_0}^S\rangle \\ \langle c^{j_{Pi}}, f_1^{\,j_{Pi}}, P_i, {\omega_1}^S\rangle \\ \ldots \\ \langle c^{j_{Pi}}, {f_{n(T^S)}}^{j_{Pi}}, P_i, {\omega_{n(T^S)}}^S\rangle \end{array}\right\} \ i = 1,2,3,4,10; \ j_{P_i} = \overline{1, n_S},$$

$$n_S = \sum_{i=1}^{4} n(P_i) + n(P_{10}), \quad (20)$$

где $n(P_i)$ – количество слов грамматического класса P_i; n_S – количество существительных в реестре; $n(T^S)$ – количество словоизменительных форм в парадигме субстантивного типа. Общепринятым для русского языка: $n(T^S)$ = 12;[6] ${\omega_m}^S = \langle {\omega^2}_i, {\omega^3}_j\rangle \in \Omega(T^S)$, ${\omega^2}_i \in \Omega^2$, $i = 1,2$; ${\omega^3}_j \in \Omega^3$, $j = 1,2,...,6$;

$$\left.\begin{array}{l} \langle c^{j_{Pi}}, f_0^{\,j_{Pi}}, P_i, {\omega_0}^A\rangle \\ \langle c^{j_{Pi}}, f_1^{\,j_{Pi}}, P_i, {\omega_1}^A\rangle \\ \ldots \\ \langle c^{j_{Pi}}, {f_{n(T^A)}}^{j_{Pi}}, P_i, {\omega_{n(T^A)}}^A\rangle \end{array}\right\}, \quad (21)$$

$i = 5,9,11$.

$j_{P_i} = \overline{1, n_A}, n_A = n(P_5) + n(P_9) + n(P_{11})\,,$

где n_A – количество адъективов (прилагательные + порядковые числительные + причастия + местоиме-

6 Имеются и иные мнения относительно количества словоизменительных форм в парадигме существительного, а именно: 14. См., например, [27; 28].

ния-прилагательные) в реестре; $n(T^A)$ – количество словоизменительных форм в парадигме адъективного типа: $n(T^A) = 28$; $\omega_m{}^A = \langle \omega_l^1, \omega_i^2, \omega_j^3 \rangle \in \Omega(T^A)$; $\omega^1{}_l \in \Omega^1, l = 1,2,3; \omega^2{}_i \in \Omega^2, i = 1,2;\quad \omega^3{}_j \in \Omega^3, j = 1,2,...6$; $m = \overline{1,28}$.

$$\left.\begin{array}{l}\langle c^{j_{Pi}}, f_0{}^{j_{Pi}}, P_i, \omega_0{}^V \rangle \\ \langle c^{j_{Pi}}, f_1{}^{j_{Pi}}, P_i, \omega_1{}^V \rangle \\ \ldots \\ \langle c^{j_{Pi}}, f_{n(T^V)}{}^{j_{Pi}}, P_i, \omega_{n(T^V)}{}^V \rangle\end{array}\right\}$$

$$i = 6,7,8.$$

$$j_{P_i} = \overline{1, n_V},\ n_V = \sum_{i=6}^{8} n(P_i), \tag{22}$$

где n_V – количество глаголов в реестре; $n(T^V)$ – количество словоизменительных форм в парадигме глагольного типа: $n(T^V) = 49$; $\omega_m^V = \langle \omega_i^4, \omega_j^5, \omega_j^7, \omega_j^2, \omega_j^1 \rangle \in \Omega(T^V)$, $m = \overline{1,49}$; $i = 1, 2$; $j = 1, 2, 3$; $\omega_i^4 \in \Omega^4$, $\omega_j^5 \in \Omega^5$, $\omega_j^6 \in \Omega^6$, $\omega_j^7 \in \Omega^7$, $\omega_i^2 \in \Omega^2$, $\omega_j^1 \in \Omega^1$.

$$\left.\begin{array}{l}\langle c^{j_{P_i}}, f_0{}^{j_{P_i}}, P_i, \omega_0{}^C \rangle \\ \langle c^{j_{Pi}}, f_1{}^{j_{Pi}}, P_i, \omega_1{}^C \rangle \\ \ldots \\ \langle c^{j_{Pi}}, f_{n(T^C)}{}^{j_{Pi}}, P_i, \omega_{n(T^C)}{}^C \rangle\end{array}\right\}$$

$$i = 12.$$

$$j_{P_i} = \overline{1, n_C},\ n_C = n(P_{12}), \tag{23}$$

где n_C – количество числительных в реестре; $n(T^C)$ – количество словоизменительных форм в парадигме количественного числительного: $n(T^C) = 6$; $\omega_m{}^C = \langle \omega_j^3 \rangle \in \Omega(T^C)$, $m = \overline{1, n(T^C)}$.

Построим селекции $R_{CFP\Omega(T)}$ по условию $P = P_i$ ($i = 1,2,...12$)[7] $\sigma_{P=P_i}(R_{CFP\Omega(T)})$.

$$\bigcup_{i=1}^{4} \sigma_{P=Pi}(R_{CFP\Omega(T)}) \cup \sigma_{P=P_{10}}(R_{CFP\Omega(T)}) = R_{CFP\Omega^S} \tag{24}$$

задает отношение «квазиоснова – квазифлексия – грамматический класс – грамматическое значение» для лексем парадигматического типа T^S.

Проекцией отношения $R_{CFP\Omega^S}$ по второму столбцу (по атрибуту F) является мультимножество квазифлексий F(T^S) для лексем парадигматического типа T^S:

$$\pi_F(R_{CFP\Omega^S}) = \{\{f_0{}^{j_{Pi}}\}, \{f_1{}^{j_{Pi}}\}, ..., \{f_{n(T^S)}{}^{j_{Pi}}\}\},\ j_{P_i} = \overline{1, n_S}. \tag{25}$$

$$F_0{}^{T^S} = \{f_0{}^{j_{Pi}}\},\ F_1{}^{T^S} = \{f_1{}^{j_{Pi}}\}, ..., F_{n(T^N)}{}^{T^S} = \{f_{n(T^S)}{}^{j_{Pi}}\}. \tag{26}$$

Построим на $F_i^{T^S}, i = \overline{0, n(T^S)}$ отношение

$$R_{F(T^S)} \subseteq F_0{}^{T^S} \times F_1^{T^S} \times ... \times F_{n(T^S)}{}^{T^S}. \tag{27}$$

Отношение $R_{F(T^S)}$ по построению является отношением эквивалентности a, следовательно, разбивает множество $F^{T^S} \subset F$ на классы. Кортежи $\{f_0{}^j, f_1^j, ..., f_{n(T^S)}{}^j\}$ задают классы $\Pi_j(P, T^S)$. Нижний индекс при f отвечает номеру грамматического значения в парадигме, верхний индекс – номер парадигматического класса. Например кортеж квазифлексий $\{\langle$'*ь*'$,\omega_1^S\rangle$, $\langle$'*я*'$,\omega_2^S\rangle$, $\langle$'*ю*'$,\omega_3^S\rangle$, $\langle$'*я*'$,\omega_4^S\rangle$, $\langle$'*ем*'$,\omega_5^S\rangle$, $\langle$'*е*'$,\omega_6^S\rangle$, $\langle$'*ья*'$,\omega_7^S\rangle$, $\langle$'*ей*'$,\omega_8^S\rangle$, $\langle$'*ьям*'$,\omega_9^S\rangle$, $\langle$'*ей*'$,\omega_{10}^S\rangle$, $\langle$'*ьями*'$,\omega_{11}^S\rangle$, $\langle$'*ьях*'$,\omega_{12}^S\rangle\}$ задает парадигматический класс существительных, представителем которого есть которого слово '*князь*'.

Аналогично,

$$\sigma_{P=P_5}(R_{CFP\Omega(T)}) \cup \sigma_{P=P_9}(R_{CFP\Omega(T)}) \cup$$
$$\cup\, \sigma_{P=P_{11}}(R_{CFP\Omega(T)}) = R_{CFP\Omega^A} \tag{28}$$

задает отношение «квазиоснова – квазифлексия – грамматический класс – грамматическое значение» для лексем парадигматического типа T^A.

Проекция отношения $R_{CFP\Omega^A}$ по второму столбцу (по атрибуту F) является множеством квазифлексий F(T^A) для лексем парадигматического типа T^A:

$$\pi_F(R_{CFP\Omega^A})) = \{\{f_0{}^{j_{Pi}}\}, \{f_1{}^{j_{Pi}}\}, ..., \{f_{n(T^A)}{}^{j_{Pi}}\}\},$$
$$j_{P_i} = \overline{1, n_A}, \tag{29}$$

где $F_0{}^{T^A} = \{f_0{}^{j_{Pi}}\}$, $F_1{}^{T^A} = \{f_1^{j_{Pi}}\}$, ..., $F_{n(T^A)}{}^{T^A} =$

$$= \{f_{n(T^A)}{}^{j_{Pi}}\}. \tag{30}$$

Отношение

$$R_{F(T^A)} \subseteq F_0{}^{T^A} \times F_1^{T^A} \times ... \times F_{n(T^A)}{}^{T^A}, \tag{31}$$

построенное на декартовом произведении $F_i^{T^A}, i = \overline{0, n(T^A)}$, разбивает множество $F^{T^A} \subset F$ на классы эквивалентности $\Pi_j(P, T^A)$.

Кортежи $\{f_0{}^j, f_1^j, ..., f_{n(T^A)}{}^j\}$ задают классы $\Pi_j(P, T^A)$. Например кортеж $\{\langle$'*лый*'$,\omega_1^A\rangle$, $\langle$'*лого*'$,\omega_2^A\rangle$, $\langle$'*лому*'$,\omega_3^A\rangle$, $\langle$'*лый*'$,\omega_4^A\rangle$, $\langle$'*лым*'$,\omega_5^A\rangle$, $\langle$'*лом*'$,\omega_6^A\rangle$, $\langle$'*лая*'$,\omega_7^A\rangle$, $\langle$'*лой*'$,\omega_8^A\rangle$, $\langle$'*лой*'$,\omega_9^A\rangle$, $\langle$'*лую*'$,\omega_{10}^A\rangle$, $\langle$'*лой*'$,\omega_{11}^A\rangle$, $\langle$'*лой*'$,\omega_{12}^A\rangle$, $\langle$'*лое*'$,\omega_{13}^A\rangle$, $\langle$'*лого*'$,\omega_{14}^A\rangle$, $\langle$'*лому*'$,\omega_{15}^A\rangle$, $\langle$'*лое*'$,\omega_{16}^A\rangle$, $\langle$'*лым*'$,\omega_{17}^A\rangle$, $\langle$'*лом*'$,\omega_{18}^A\rangle$, $\langle$'*лые*'$,\omega_{19}^A\rangle$, $\langle$'*лых*'$,\omega_{20}^A\rangle$, $\langle$'*лым*'$,\omega_{21}^A\rangle$, $\langle$'*лые*'$,\omega_{22}^A\rangle$, $\langle$'*лыми*'$,\omega_{23}^A\rangle$, $\langle$'*лых*'$,\omega_{24}^A\rangle$, $\langle$'*ел*'$,\omega_{25}^A\rangle$, $\langle$'*ла*'$,\omega_{26}^A\rangle$, $\langle$'*ло*'$,\omega_{27}^A\rangle$, $\langle$'*лы*'$,\omega_{28}^A\rangle\}$ задает парадигматический класс адъективов, в состав которого входит прилагательное '*светлый*'.

$$\bigcup_{i=6}^{8} \sigma_{P=Pi}(R_{CFP\Omega(T)}) = R_{CFP\Omega^V} \tag{32}$$

[7] Рассматриваем лишь словоизменительные классы

задает отношение «квазиоснова – квазифлексия – грамматический класс – грамматическое значение» для лексем грамматического типа T^V.

Проекция отношения $R_{CFP\Omega^V}$ по второму столбцу (по атрибуту F):

$$\pi_F(R_{CFP\Omega^V}) = \{\{f_0^{j_{Pi}}\}, \{f_1^{j_{Pi}}\}, ..., \{f_{n(T^V)}{}^{j_{Pi}}\}\}, \; j_{P_i} = \overline{1,n_V}, \tag{33}$$

где $$F_0^{T^V} = \{f_0^{j_{Pi}}\}, \; F_1^{T^V} = \{f_1^{j_{Pi}}\}, ..., F_{n(T^V)}{}^{T^V} = \{f_{n(T^V)}{}^{j_{Pi}}\}. \tag{34}$$

На декартовом произведении элементов $F_i^{T^V}$, $i = \overline{0, n(T^V)}$ строим отношение эквивалентности

$$R_{F(T^V)} \subseteq F_0^{T^V} \times F_1^{T^V} \times ... \times F_{n(T^V)}{}^{T^V}, \tag{35}$$

которое разбивает множество $F^{T^V} \subset F$ на классы эквивалентности $\Pi_j(P,T^V)$.

Кортежи $\{f_0{}^j, f_1{}^j, ..., f_{n(T^V)}{}^j\}$ задают классы $\Pi_j(P,T^V)$. Например парадигматический класс, представителем которого является глагол совершенного вида '*произойти*', задается кортежем

$$\{\langle flex, i\rangle | \, i = 0,...,n(T^V)\} = \{\langle\text{'}\textit{зойти}\text{'}, \omega_1^V\rangle, \langle\text{'}\textit{зойду}\text{'}, \omega_2^V\rangle, \langle\text{'}\textit{зойдёшь}\text{'}, \omega_3^V\rangle, \langle\text{'}\textit{зойдёт}\text{'}, \omega_4^V\rangle, \langle\text{'}\textit{зойдём}\text{'}, \omega_5^V\rangle, \langle\text{'}\textit{зойдёте}\text{'}, \omega_6^V\rangle, \langle\text{'}\textit{зойдут}\text{'}, \omega_7^V\rangle, \langle\text{'}\textit{зошёл}\text{'}, \omega_8^V\rangle, \langle\text{'}\textit{зошла}\text{'}, \omega_9^V\rangle, \langle\text{'}\textit{зошло}\text{'}, \omega_{10}^V\rangle, \langle\text{'}\textit{зошли}\text{'}, \omega_{11}^V\rangle, \langle\text{'}\textit{зойди}\text{'}, \omega_{12}^V\rangle, \langle\text{'}\textit{зойдите}\text{'}, \omega_{13}^V\rangle, \langle\text{'}\textit{зойдя}\text{'}, \omega_{14}^V\rangle, \langle\text{'}\textit{зошедший}\text{'}, \omega_{15}^V\rangle, \langle\text{'}\textit{сшедший}\text{'}, \omega_{15}^V\rangle\},$$

где ω_i^V ($i = 1,2,...,15$) – комплексы грамматических значений $\langle$ *залог, наклонение, время, число, лицо, род* $\rangle$, которые для данного примера парадигматического класса имеют такое содержание:

ω_1^V – $\langle$ неопределенная форма/ инфинитив>;

ω_2^V – $\langle$ действительный залог, изъявительное наклонение, будущее время, единственное число, первое лицо $\rangle$;

ω_3^V – $\langle$ действительный залог, изъявительное наклонение, будущее время, единственное число, второе лицо $\rangle$;

ω_4^V – $\langle$ действительный залог, изъявительное наклонение, будущее время, единственное число, третье лицо $\rangle$;

ω_5^V – $\langle$ действительный залог, изъявительное наклонение, будущее время, множественное число, первое лицо $\rangle$;

ω_6^V – $\langle$ действительный залог, изъявительное наклонение, будущее время, множественное число, второе лицо $\rangle$;

ω_7^V – $\langle$ действительный залог, изъявительное наклонение, будущее время, множественное число, третье лицо $\rangle$;

ω_8^V – $\langle$ действительный залог, изъявительное наклонение, прошедшее время, единственное число, третье лицо, мужской род $\rangle$;

ω_9^V – $\langle$ действительный залог, изъявительное наклонение, прошедшее время, единственное число, третье лицо, женский род $\rangle$;

ω_{10}^V – $\langle$ действительный залог, изъявительное наклонение, прошедшее время, единственное число, третье лицо, средний род $\rangle$;

ω_{11}^V – $\langle$ действительный залог, изъявительное наклонение, прошедшее время, множественное число, третье лицо>;

ω_{12}^V – $\langle$ действительный залог, повелительное наклонение, будущее время, единственное число, второе лицо $\rangle$;

ω_{13}^V – $\langle$ действительный залог, повелительное наклонение, будущее время, множественное число, второе лицо $\rangle$;

ω_{14}^V – $\langle$ действительный залог, прошедшее время $\rangle$;

ω_{15}^V – $\langle$ действительный залог, прошедшее время, единственное число $\rangle$.

$$\sigma_{P=P_{12}}(R_{CFP\Omega(T)})) = R_{CFP\Omega^C}. \tag{36}$$

$$\pi_F(R_{CFP\Omega^C}) = \{\{f_0^{j_{P_i}}\}, \{f_1^{j_{P_i}}\}, ..., \{f_{n(T^C)}{}^{j_{Pi}}\}\}, \; j_{P_i} = \overline{1,n_C}, \tag{37}$$

где $$F_0^{T^C} = \{f_0^{j_{Pi}}\}, F_1^{T^C} = \{f_1^{j_{Pi}}\}, ..., F_{n(T^C)}{}^{T^C} = \{f_{n(T^C)}{}^{j_{Pi}}\}. \tag{38}$$

Отношение

$$R_{F(T^C)} \subseteq F_0^{T^C} \times F_1^{T^C} \times ... \times F_{n(T^C)}{}^{T^C} \tag{39}$$

разбивает множество $F^{T^C} \subset F$ на классы эквивалентности $\Pi_j(P,T^C)$. Кортежи $\{f_0{}^j, f_1{}^j, ..., f_{n(T^C)}{}^j\}$ задают классы $\Pi_j(P,T^C)$.

Отношение «квазифлексия – парадигматический класс»

$$R_{F\Pi} \subseteq R_{F(T)} \times \mathrm{N}, \text{ где } \mathrm{N} = \{1,2,3,...\}. \tag{40}$$

Объединение отношения $R_{CFP\Omega(T)}$ и $R_{F\Pi}$ по атрибуту F:

$$\pi_{C, R_{F\Pi}.F, P, \Omega(T), \Pi}(\sigma_{R_{CFP\Omega(T)}.F = R_{F\Pi}.F}(R_{CFP\Omega(T)} \times R_{F\Pi})) \tag{41}$$

дает разбиение множества $W(P,T) \subset W_P \subset W_T$ на классы (парадигматические классы).

3. Оператор парадигматизации

Для автоматического построения полной парадигмы по исходной (начальной) форме x_0 определяется оператор парадигматизации:

$$\begin{aligned} \mathrm{H}: x_0 \to [x] = c(x) * \{f_0(x), f_1(x), ..., f_n(x)\} \equiv \\ \equiv \{c(x) * f_0(x), c(x) * f_1(x), ..., c(x) * f_n(x)\}, \end{aligned} \tag{42}$$

действие которого определяется отношением $\pi(x^1, x^2)$.

Оператор полной парадигматизации (который действует на ММ лексем W) определяется по формуле

$$\mathrm{H} = \sum_{i=1}^{N} \mathrm{H}_i \cdot \delta(x; T_i), \tag{43}$$

где

$$\delta(x; T_i) = \begin{cases} 1, & x \in W(T_i), \\ 0, & x \notin W(T_i), \end{cases} \tag{44}$$

N – количество парадигматических типов, H_i – оператор парадигматизации, действующий на ММ лексем соответствующего парадигматического типа $W(T_i)$ [8]:

$$\mathrm{H}_i : x_0 \to [x] \; \forall x \in W(T_i). \tag{45}$$

Поскольку имеет место (2), справедлива формула:

$$\mathrm{H}_i = \sum_{j=1}^{p} \mathrm{H}_i^j \cdot \delta(x; P_j), \tag{46}$$

где

$$\delta(x; P_j) = \begin{cases} 1, & x \in W(P_j), \\ 0, & x \notin W(P_j), \end{cases} \tag{47}$$

$i = 1,2,..., N$; p – количество грамматических классов в множестве $W(T_i)$.

Оператор H_i^j действует на множестве лексем в пределах грамматического класса P_j парадигматического типа T_i:

$$\mathrm{H}_i^j : x_0 \to [x] \quad \forall x \in W(P_j) \subseteq W(T_i) \tag{48}$$

В свою очередь, поскольку P_j является объединением парадигматических классов ($P_j = \bigcup_{k=1}^{n} \Pi_k$), можно записать, что

$$\forall x \in W(\Pi_k) \subseteq W(P_j) \subseteq W(T_i),$$

$$\mathrm{H}_i^{j^k} : x_0 \to c(x) * [F]_{ij}^k \quad \mathrm{H}_i^j = \sum_{k=1}^{n_j} \mathrm{H}_i^{j^k} \cdot \delta(x, \Pi_k), \tag{49}$$

где $\mathrm{H}_i^{j^k}$ – оператор парадигматизации, который действует в пределах парадигматического класса Π_k; $i = 1,2,..., N$; $j = 1,2,..., n_j$; n_j –количество парадигматических классов в ММ $W(P_j) \subseteq W(T_i)$; $[F]_i^k$ – множество наборов квазифлексий слов грамматического класса P_j парадигматического типа T_i; функция

$$\delta(x; \Pi_k) = \begin{cases} 1, & x \in W(\Pi_k), \\ 0, & x \notin W(\Pi_k). \end{cases} \tag{50}$$

Таким образом, для каждого из парадигматических типов $W(T_i)$ оператор парадигматизации определяется независимо.

[8] на ММ лексем каждого из парадигматических типов действует свой оператор парадигматизации. Прежде всего потому, что каждый из парадигматических типов характеризуется своим комплексом значений грамматических категорий, который оличен для каждого ПТ.

Оператор H отображает лексему x на ее полную парадигму $[x]$. Его реализация для лексики соответствующего флективного языка осуществляется при помощи словаря квазифлексий и набора алгоритмов построения полных словоизменительных парадигм. С помощью парадигматического словаря любой лексеме приписывается ее словоизменительный тип, далее по алгоритму морфологического анализа (с использованием набора алгоритмов построения полных словоизменительных парадигм) осуществляется грамматическая идентификация лексемы x. После этого лексема приобретает представление (4).

Алгоритмическая реализация оператора H^{-1} осуществляет процесс лемматизации, то есть сведение произвольной словоформы к ее исходной канонической форме.

Изложенная морфологическая модель составляет концептуальную основу для компьютерного моделирования и реализации функции парадигматических отношений для определенного класса флективных языков. Согласно изложенной концепции строение парадигматической лексикографической системы для флективного языка показано на рис. 9.

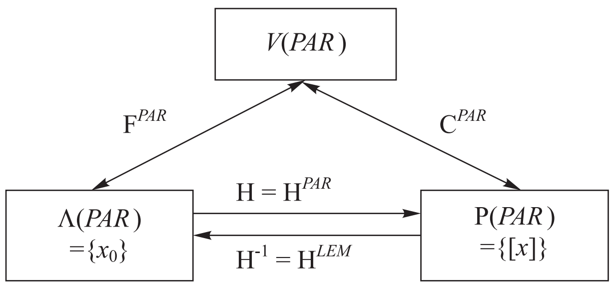

Рис. 9. Структура парадигматической Л-Системы

В приведенной схеме элементы имеют такую интерпретацию:

$V(PAR)$ – множество словарных статей грамматической Л-Системы;

$\Lambda(PAR) = F^{PAR} V(PAR) = \{x_0\}$;

$\mathrm{P}(PAR) = C^{PAR} V(PAR) = \{[x]\}$;

$\mathrm{H} = \mathrm{H}^{PAR}$ – оператор парадигматизации:

$$\mathrm{H} = \mathrm{H}_{x_0}^{PAR} = [x];$$

$\mathrm{H}^{-1} = \mathrm{H}^{LEM}$ – оператор лемматизации:

$\mathrm{H}^{LEM} \circ \chi(x) = x_0$, где $\chi(x)$ – любой элемент парадигмы $[x]$.

Выводы

На основе описанного формализма нами была построена грамматическая лексикографическая система для русского языка, в основу которой положена классификация, базирующаяся на выделении в лексическом составе языка парадигматических (словоизменительных) типов и подтипов, грамматических классов. Внутри грамматических классов выделялись парадигматические классы, каждый из которых включает лексемы с одинаковыми правилами образования определенных словоизменительных форм. Построен и программно реализован оператор парадигматизации, в результате чего каждому слову был поставлен в соответствие его парадигматический класс. Это позволяет автоматически получать полную словоизменительную парадигму любого слова. Реализация и верификация описанной модели морфологии русского языка выполнена на объеме лексики около 170 тыс. единиц. Действующая модель системы представлена в свободном доступе в Интернете на сайте Украинского лингвистического портала (www.ulif.org.ua, Словарь русского словоизменения).

Список литературы: 1. *Корпусна* лінгвістика: Монографія / В.А. Широков, О.В. Бугаков, Т.О. Грязнухіна, Т.П. Любченко, О.Г. Рабулець, О.О. Сидоренко, Н.М. Сидорчук, И.В. Шевченко, О.О. Шипнівська, К.М. Якименко; Український мовно-інформаційний фонд НАН України. – К.: Довіра, 2005. – 472 с. 2. *Широков В.А.* Інформаційна теорія лексикографічних систем. – К.: Довіра, 1998. – 331 с. 3. *Широков В.А.* Феноменологія лексикографічних систем. – К.: Наукова думка, 2004. – 327 с. 4. *Белоногов Г.Г.,Зеленков Ю.Г.* Алгоритм морфологического анализа русских слов // Вопросы информационной теории и практики. М.: ВИНИТИ, 1985, № 53. 5. *Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л.*и др. Лингвистическое обеспечение системе автоматического перевода Этап-2. – М.: Наука, 1989. – 296 с. 6. *Бидер И.Г.*, *Большаков И.А., Еськова Н.А.* Формальная модель русской морфологии. – ИРЯ АН СССР, Проблемная группа по экспериментальной и прикладной лингвистике. – Вып. 112. – М.: 1978. – I – 60 с., II – 59 с. 7. *Хорошилов А.А.* Автоматическая нормализация слов в системах обработки научно-технической информации: Автореф. дис. ... канд. техн. наук. – М., 1987. – 18 с. 8. *Гельбух А.Ф.*, *Сидоров Г.О.* К вопросу об автоматическом морфологическом анализе флективных языков // Компьютерная лингвистика и интеллектуальные технологии: Тр. междунар. конф. “Диалог’2005” – http://www.dialog-21.ru/Archive/2005/Gelbukh%20Sidorov/GelbukhA.htm 9. *Лингвистический* энциклопедический словарь /Гл. ред. В.Н. Ярцева. М.: Советская энциклопедия, 1990. – 688 с. 10. *Зализняк А.А.* Грамматический словарь русского языка: Словоизменение. – М.: Русский язык, 1978. – 878 с. (4-е изд., испр. и доп.: 2003). 11. *Грязнухина Т.А.*, *Любченко Т.П., Рабулец А.Г.* Электронная версия грамматического словаря русского языка (А.А. Зализняк) как инструмент автоматического морфологического анализа русского текста // Докл. науч. конф. «Корпусная лингвистика и лингвистические базы данных». – Санкт-Петербург, март 2002. – С. 63-70. 12. *Шевченко І.В.* Алгоритмічна словозмінна класифікація української лексики // Мовознавство. – 1996. – № 4-5. – С. 40-44. 13. *Шевченко І.В.* Моделі та алгоритмічно-програмне забезпечення лексикографічних систем: Автореф. дис. ... канд. техн. наук. – К., 2000. – С. 20. 14. *Гарант-парк-интернет*. Технологии анализа и поиска текстовой информации / URL: http://research.metric.ru 15. *Ножов И.М.* Процессор автоматизированного морфологического анализа без словаря. Деревья и корреляция // Диалог’2000. Труды конференции – Протвино, 2000. – Т.2. – С. 284-290. 16. *SRILM* – The SRI Language Modeling Toolkit / URL: http://www.speech.sri.com/projects/srilm 17. *Селезнев К.* Обработка текстов на естественном языке // Журнал “Открытые системы”, #12, 2003 ч // Изд. “Открытые системы” / URL: http://www.osp.ru/os/2003/12/048.htm . 18. *Сокирко А. В.* Морфологические модули на сайте www.aot.ru // Диалог-2005. 19. *Коваленко А.* Вероятностный морфологический анализатор (стеммер) // Системный администратор.– №1. – Октябрь, 2002. / URL: http://linguist.nm.ru/ling/ 20. *Любченко Т.П.* Програмно-технологічні аспекти створення граматичних лексикографічних систем // Проблеми програмування. – 2007. – № 3. – С. 61-75. 21. *Любченко Т.П.* Морфологічна модель словозміни флективної мови та електронний граматичний словник //Біоніка інтелекту: наук.-техн. журнал. – 2006. – №1 (64). – С. 72-77. 22. *Кнут Д.Э.* Искусство программирования. – М.: Мир, 1977.– т.2.: Получисленные алгоритмы. 23. *Петровский А.Б.* Метрические пространства мультимножеств // Доклады Академии наук. – 1995. – Т.344, № 2. – С. 175-177. 24. *Петровский А.Б.* Комбинаторика мультимножеств // Доклады Академии наук. — 2000. – Т.370, №6. – С. 750-753. 25. *Петровский А.Б.* Пространства множеств и мультимножеств. – М.: Едиториал УРСС, 2003. 26. *Петровский А.Б.* Многокритериальное принятие решений по противоречивым данным: подход теории мультимножеств // Информационные технологии и вычислительные системы. – 2004. – №2. – С. 56-66. 27. *Коваль С.А.* К вопросу о числе падежей русского существительного (Решения для компьютерной лингвистики) // http://dialog-21.ru/Archive/2004/Koval.htm 28. *Успенский В.А.* К определению падежа по А.Н. Колмогорову // Бюллетень Объединения по проблемам машинного перевода. – № 5. – М.: [I МГПИИЯ], 1957. – С. 11-18.

Поступила в редколлегию 24.03.2008