РОССИЙСКАЯ АКАДЕМИЯ НАУК

ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ
(ВИНИТИ)

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г. **№ 4** **Москва 2013**

ИНФОРМАЦИОННЫЙ АНАЛИЗ

УДК 519.216.3:004.738.5

К.Е. Бауман, А.Н. Корнетова, В.А. Топинский, Д.А. Хакимова

Оптимизация прогноза вероятности посещения контекстной рекламы в поисковой системе «Яндекс»

Обсуждается проблема построения оценки вероятности клика пользователя по рекламному объявлению, показанному на странице поисковой выдачи. Методы, предложенные в работе, повысили качество прогноза (в терминах как метрики правдоподобия, так и основных показателей системы). Также были рассмотрены случаи показов рекламных объявлений, когда история их показов достаточно мала (т.е. таких объявлений, которые можно считать новыми). Предложенный в работе вид формулы прогноза учитывает дисперсионность и большие риски показа нового рекламного объявления.

Ключевые слова: *оценка вероятности клика, метрика качества прогноза, off-line-предсказание, планирование эксперимента*

ВВЕДЕНИЕ

«Яндекс» – крупнейшая российская поисковая интернет-система. На сегодняшний день «Яндекс» в ответ на запрос пользователя на странице результатов поиска отображает два различных вида информации: результаты поиска информации для ответа на запрос и рекламные объявления, некоторым образом отобранные для данного запроса.

Основной частью предоставляемой информации – в виде списка результатов первого вида, является поисковая выдача. Поисковая система формирует список результатов поиска с целью максимально точно ответить на конкретный запрос. Исходя из этого, до показа пользователю результатов поиска они ранжируются, основываясь на мере их релевантности запросу.

Вторым типом информации являются рекламные объявления, которые отображаются справа от результатов поиска (так называемые «показы справа») и над результатами поиска («спец-размещение»). Для всех крупных поисковых интернет-систем доход, полученный с показов поисковой рекламы, составляет важную, а зачастую определяющую часть от общего дохода компании. Исходя из этого критерия,

формирование рекламной выдачи должно происходить с учётом нескольких факторов: соответствие рекламного объявления запросу пользователя и ожидаемый доход при условии показа этого объявления.

В настоящее время задача выбора рекламного объявления для показа по конкретному запросу решается с помощью нескольких механизмов отбора. Каждому объявлению рекламодатель приписывает набор ключевых фраз. В зависимости от выбранного рекламодателем механизма подбора, рекламное объявление может быть отобрано для возможного показа в следующих случаях: точного совпадения хотя бы одной ключевой фразы и запроса, частичного совпадения (когда какая-то из ключевых фраз является частью запроса) или семантического соответствия ключевых фраз запросу (broad match, близкие по смыслу или тематике).

Статья авторов посвящена решению второй подзадачи выбора рекламных объявлений – учёту ожидаемого дохода от показа.

Сейчас большинство поисковых систем, в том числе и Яндекс, используют механизм продаж рекламы на поиске, при котором рекламодатель платит за клик пользователя по его объявлению, другими словами, рекламодатель платит лишь за реальные переходы пользователей с поисковых страниц на рекламируемые сайты. При этом ставки рекламодатели выставляют по обобщённому аукциону второй цены [2].

Количество объявлений, которые могут быть показаны пользователю на странице результатов поиска, ограничено. Как правило, это не более трех объявлений над поисковой выдачей и не более четырех-пяти объявлений в блоке справа. Система отбирает наиболее эффективные рекламные объявления среди уже прошедших отбор.

На практике этот отбор реализуется выбором объявлений с наибольшими значениями характеристики CPM (cost-per-mille, стоимость рекламы за тысячу показов). Название этой характеристики объясняется историей развития интернет-рекламы, когда всё начиналось с продаж показов рекламных объявлений. Но в наши дни рекламодатели торгуются за клики, поэтому CPM теперь вычисляется как:

$$CPM = BID \cdot CTR,$$

где BID – это ставка рекламодателя за клик по рекламе, а CTR *(click-through rate)* – вероятность клика по данной рекламе.

Более подробный обзор рекламных технологий в интернете можно найти в [3].

Таким образом, задача эффективного отбора рекламных объявлений сводится к задаче построения оценки вероятности клика по данной рекламе. Эта задача - одна из наиболее важных задач современных поисковых интернет-систем, а качество ее решения имеет прямое влияние на меру удовлетворения пользователей поисковой системы, на эффективность контекстной рекламы для рекламодателей и на прибыль самой поисковой системы.

В данной работе обсуждаются проблемы, возникающие при решении задачи построения оценки CTR рекламного объявления.

1. OFF-LINE-ОПТИМИЗАЦИЯ

Рассмотрим метрики, с помощью которых мы бы могли сравнить друг с другом разные прогнозы вероятности клика. В литературе по прикладным вопросам теории вероятности и математической статистики широко используются такие классические метрики, как логарифм правдоподобия, средний квадрат ошибки или средняя абсолютная ошибка, меры корреляции, площадь под ROC-кривой (см. [5]) и т.д.

На практике прогноз или оценка CTR является лишь частью большой и сложной системы показов рекламы в интернете. Наиболее интересным вопросом является то как новый прогноз CTR скажется на таких общих показателях системы, как средний CTR по системе в целом или по отдельным частям трафика, суммарная прибыль компании за день (за месяц), количество трафика, выраженного в кликах и т.д.

Наилучший, но довольно сложный и дорогой в реализации способ – это провести on-line-эксперимент на части поисковых запросов, однако такие эксперименты могут требовать много времени и могут стать причиной возможных потерь компании на экспериментальном трафике (как прибыли, так и кликов). Поэтому мы разработали методики off-line-сравнения двух различных прогнозов вероятности клика по рекламному объявлению.

1.1. Метод ROC-подобных кривых

Опишем суть построения графиков, необходимых для сравнения прогнозов.

В качестве входных данных мы берем имеющиеся записи из базы данных о прошлых показах рекламы. Используя прогноз CTR, мы приписываем каждой такой записи ряд производных характеристик: CTR, CPM и, ожидаемое списание за клик, cost.

Далее, упорядочив все события по значению характеристики CPM, мы строим следующую кривую, например, для анализа числа кликов: текущее значение абсциссы мы рассматриваем как CPM порог, тогда значение ординаты – это количество кликов в тех событиях, чьи соответствующие значения CPM для объявления превысили данный порог. Таким же образом мы строим графики для прибыли, CTR (см. рис. 1).

Такие кривые позволяют нам сравнить различные прогнозы. Если кривая нового прогноза всегда доминирует над кривой текущего прогноза, то очевидно, что новый прогноз предпочтительней текущего. Это можно представить, как если бы мы откидывали или запрещали к показу объявления с меньшим значением CPM (что в реальности практически так и организовано), то, согласно первому прогнозу, были бы показаны объявления с большей оценкой ожидаемых кликов или денег (в зависимости от того, на каких кривых было видно явное доминирование).

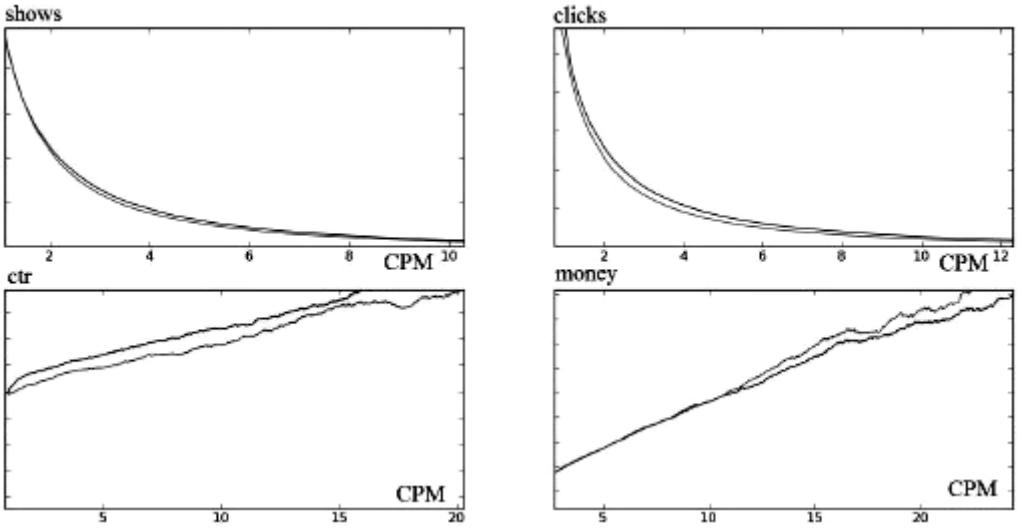

Рис. 1: Метод *ROC* -подобных кривых: *shows* – количество показов, *clicks* – количество кликов, *ctr* – средний *CTR* по системе, *money* – количество денег.

Такой подход обладает одной сложностью. Зачастую две кривые могут пересекаться на внутренней части диапазона *CPM* -порогов. Но очевидно, что на концах области определения эти кривые всегда имеют одинаковые значения. Разрешить ситуацию с внутренними пересечениями, а следовательно, с возможным чередованием доминирования, могут дополнительные ограничения. Например, поисковая система может иметь ограничение на долю генерируемых поисковых страниц, на которых показывается реклама. В этом случае соответствующий *CPM* - порог может быть легко вычислен, и прогнозы можно будет сравнивать в соответствующих точках.

1.2. Моделирование системы показов

Второй способ сравнения двух прогнозов в терминах ожидаемых кликов, денег – это построить off-line-модель всей системы показов рекламы.

С этой целью мы делаем копию всех текущих характеристик системы: всю информацию о имеющихся объявлениях, их ставках, бюджетах и т.п. Далее мы связываем эти копии с текущей версией реализации программного комплекса системы показов и таким образом получаем локальную копию системы показов рекламы с актуальными данными.

Для получения основных характеристик системы мы берем случайное множество запросов (возможно с повторениями, с целью сохранить пропорции естественного трафика) и подаем его на вход локальной копии системы. Выходом этой системы на конкретный запрос является множество всех объявлений, прошедших процедуру отбора. Для каждого объявления для данного запроса мы можем посчитать интересующие нас версии прогноза *CTR* и завершить формирование списков объявлений для показа, на выходе мы получаем столько наборов рекламных объявлений для каждого запроса пользователя, сколько различных прогнозов мы хотим исследовать.

После этого по конкретной версии рекламной выдачи мы можем посчитать ожидаемые количества денег, кликов и т.д. Для подсчета математических ожиданий в качестве истинного распределения вероятностей, мы используем версию прогноза с наилучшими показателями метрик качества (правдоподобие, линейная корреляция и пр.).

Опишем эти вычисления детальнее.

Пусть $H(Q,CTR)$ обозначает множество наборов рекламных объявлений, прошедших отбор на запросы из множества Q с использованием прогноза $CTR(\cdot)$. Каждое $h \in H(Q,CTR)$ есть множество объявлений, отобранных для показа на конкретный запрос. Нижний индекс i будет использоваться нами для ссылки на конкретное объявление из всех отобранных на показ. Пусть $CTR_{new}(\cdot)$ – новый прогноз вероятности клика (исследуемый), а $CTR_{base}(\cdot)$ – текущий прогноз.

Тогда мы можем посчитать ожидаемые изменения в основных показателях системы согласно следующим формулам.

$$\Delta Покрытие = \frac{\left|H(Q,CTR_{new})\right|}{\left|H(Q,CTR_{base})\right|} - 1,$$

$$\Delta Трафик = \frac{\sum_{h \in H(Q,CTR_{new})} \sum_{i \in h} CTR_{new}(i)}{\sum_{h \in H(Q,CTR_{base})} \sum_{i \in h} CTR_{new}(i)} - 1,$$

$$\Delta Прибыль = \frac{\sum_{h \in H(Q,CTR_{new})} \sum_{i \in h} \cos t_i \cdot CTR_{new}(i)}{\sum_{h \in H(Q,CTR_{base})} \sum_{i \in h} \cos t_i \cdot CTR_{new}(i)} - 1.$$

Здесь под Покрытием и Трафиком понимаем долю запросов из общего множества, по которым было показано хотя бы одно рекламное объявление, и количество кликов по этим объявлениям, соответственно. $CTR_{new}(i)$, $CTR_{base}(i)$, $\cos t_i$ - есть предсказанные вероятности клика новым и текущим прогнозами, и возможный платеж за клик, посчитанные для конкретного объявления i.

Мы сравниваем друг с другом производные характеристики работы алгоритма, основанного на разных версиях прогноза и, возможно, с разными специально подобранными гипер-параметрами модели. Для того чтобы вычисленные характеристики были сравнимы, мы должны при их вычислении использовать одно и то же распределение вероятностей, считая его оценкой истинного неизвестного нам распределения. В случае Покрытия мы не используем умножения на вероятности, так как в вычислении этой метрики мы не зависим от кликов – единственного источника случайности в данной модели.

Какое конкретное распределение выбрать? Мы предлагаем выбирать то, которое отвечает наилучшим показателям по метрикам качества прогноза, как наилучшую аппроксимацию неизвестного истинного распределения (в нашем случае, новый прогноз достоин исследования, только если он лучше по следующим метрикам качества: правдоподобие, средняя ошибка и прочее).

Отдельно стоит отметить, что этот метод с помощью полного восстановления правил отбора показов рекламы позволяет нам произвести оптимизацию по свободным гиперпараметрам системы. Таким образом, сравнение двух прогнозов производится в условиях того, что соответствующие рекламные выдачи были сформированы с использованием индивидуальных оптимальных значениях этих свободных параметров системы.

2. ПОСТРОЕНИЕ ОЦЕНОК И ПРОБЛЕМА РЕГУЛЯРИЗАЦИИ

2.1. Используемые данные и текущая оценка CTR

Текущая система показов рекламных объявлений в Яндексе устроена так, что про каждое объявление, отобранное для данного показа на данный запрос, мы располагаем рядом следующих статистик: количество показов и соответствующее этим показам количество кликов по различным частям поискового трафика.

Здесь и далее пусть $clicks_b$ - количество кликов по данному объявлению b, а $shows_b$ – соответствующее количество показов данного объявления, за которое были набраны эти клики. Предположим, что мы зафиксировали конкретный срез, например, мы собираем такую статистику для пары: объявление и купленная фраза, по которой происходил показ. Через Λ мы обозначаем множество объявлений, для которых мы храним данные статистики.

Событие, что по показанному объявлению произойдет клик, мы моделируем через бернуллиевскую случайную величину, принимающее значения 0 (неклик) и 1 (клик), где вероятность клика есть $p = CTR_b$.

На момент данного исследования в Яндексе уже была реализована некоторая оценка вероятности клика, основанная на множестве таких кликовых статистик. Этот прогноз обладал рядом некоторых свойств: на достаточно больших объемах показов прогноз сходился к отношению клики/показы, а в условиях отсутствия достаточной статистики был использован некоторый статический априорный прогноз (достаточно простая кусочно-постоянная функция).

Наша цель в этом исследовании – построить на тех же данных новый прогноз, обладающий такими же свойствами, использование которого приносило бы лучшие результаты по основным метрикам качества работы системы.

2.2. Оценка CTR и минимизация MSE

Рассмотрим вопрос, как на основе статистики по некоторому срезу можно построить новый прогноз CTR.

Предположим, что мы обладаем некоторой дополнительной информацией о множестве $C = \{CTR_b \mid b \in \Lambda\}$. Например, эта информация сводится к $\overline{CTR}_b$, оценке среднего значения имеющихся CTR_b. Тогда мы постараемся построить такой прогноз, который бы, в зависимости от количества показов в статистике, сходился бы от этой дополнительной информации к актуальному отношению кликов к показам. Более формально, мы ищем прогноз CTR в виде линейной комбинации:

$$\hat{CTR}(shows) = \alpha(shows) \cdot \overline{CTR} + \beta(shows) \cdot \frac{clicks}{shows},$$

где $\overline{CTR} \approx mean_{b \in \Lambda}(CTR_b)$.

Для нахождения неизвестных функций $\alpha(shows)$, $\beta(shows)$ мы можем выбрать в качестве целевого функционала для минимизации математическое ожидание квадратов ошибок $Q(shows)$.

$$Q(shows) = E \sum_{b \in \Lambda} (CTR_b - CTR_b(shows))^2 .$$

Тогда легко находим, что оптимальные значения этих функций имеют вид:

$$\alpha(shows) + \beta(shows) = 1, \;\; \alpha(shows) = \frac{S_0}{S_0 + shows},$$

где $$S_0 \approx \frac{mean_{b \in \Lambda}(CTR_b \cdot (1 - CTR_b))}{\mathrm{var}_{b \in \Lambda}(CTR_b)} .$$

Поэтому вид итогового прогноза можно записать как

$$\hat{CTR}_b = \alpha(shows) \cdot \overline{CTR} + \beta(shows) \cdot \frac{clicks}{shows} = \frac{clicks + C_0}{shows + S_0},$$

где $$C_0 = \overline{CTR} \cdot S_0 . \qquad (1)$$

Ниже, на рис. 2 мы представили сравнение данного вида прогноза с текущим.

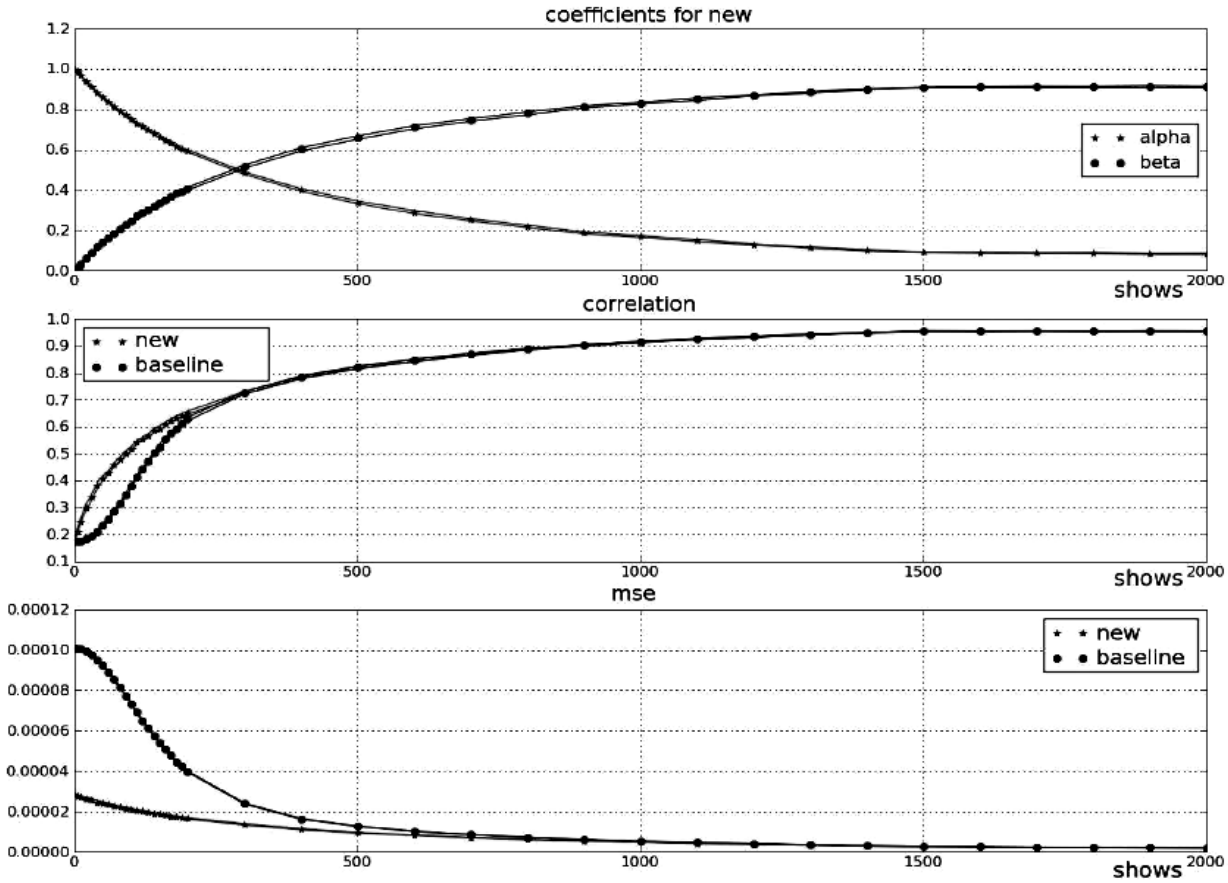

Рис. 2. Верхний график показывает поведение функций α , β ; средний – представляет корреляцию двух прогнозов с кликами; нижний – средний квадрат ошибок.

2.3. Оценка CTR и максимизация правдоподобия

Другой широко известный метод для получения прогноза этого же вида - это Байесовский подход с предполагаемым априорным распределением. Пусть $P_{prioir}(CTR)$ – априорное распределение истинных CTR для объявлений в множестве Λ. Тогда апостериорную вероятность для CTR_b легко вычислить как

$$P_{posterior}(CTR_b) = P_{prior}(CTR_b) \cdot CTR_b^{clicks_b} \cdot (1 - CTR_b)^{shows_b - clicks_b}$$

.

Если предположить, что априорное распределение $P_{prioir}(CTR)$ принадлежало к семейству Бета-распределений (т.е. $P_{prioir}(CTR) = \frac{CTR^a (1-CTR)^d}{B(a+1, d+1)}$), то тогда соответствующее апостериорное распределение, как известно, будет тоже Бета-распределением с параметрами $(a + clicks_b, d + shows_b - clicks_b)$. Поэтому апостериорное математическое ожидание вероятности клика по рекламному объявлению b будет:

$$\hat{CTR}_b = \frac{a + clicks_b}{a + clicks_b + d + shows_b - clicks_b} = \\ = \frac{clicks_b + a}{shows_b + a + d} = \frac{clicks + C_1}{shows + S_1}$$

.

Таким образом, при разных начальных предположениях мы дважды пришли к одному и тому же виду оценки. В данном случае, у нас остались неизвестными две константы, которые при желании можно получить, построив оценку априорного распределения. Но вместо этого, мы предложим другую эвристику, пришедшую к нам из рассуждений в пункте 2.2.

2.4. Практическое использование предложенной оценки

Как было показано, довольно распространённой формой для оценки вероятности клика является регуляризированное отношения количества кликов к количеству соответствующих показов с регуляризацией. Запишем это отношение так, как оно было получено в пункте 2.2.

$$\hat{CTR}_b = \frac{clicks + D_0 \cdot \tilde{CTR}}{shows +} D_0 ,$$

где $\tilde{CTR}$ - некоторая дополнительная информация.

Теперь допустим, что в качестве такой дополнительной информации $\tilde{CTR}$ мы используем некоторый "статический прогноз" (т.е. прогноз, вычисляющийся без прямого использования статистик количества кликов и показов; например, это может быть логистическая регрессия, построенная на множестве тек-

стовых признаков, информация про конкретного пользователя и пр.).

Предположим, что этот статический прогноз имеет некоторую позитивную корреляцию R с соответствующей случайной величиной $click = \sum_{b \in \Lambda} click_b \cdot show_b$, где $show_b = 1_{[b-был-показан]}$, $click_b = 1_{[b-был-кликнут]}$. Тогда, если предположить, что данный прогноз обладает достаточно малой дисперсией: $\Sigma < R^2 \cdot \mathrm{var}[clicks]$, то нетрудно показать, что $D_0 > S_0$, при условии $\alpha(shows) + \beta(shows) = 1$.

Данный факт имеет естественную интерпретацию. Если мы построили статический прогноз CTR лучшего качества, чем оценка CTR среднего истинных вероятностей, тогда итоговый прогноз будет дольше "доверять" нашему начальному приближению. То есть сходимость к отношению кликов к показам будет более медленной.

3. МИНИМИЗАЦИЯ РИСКОВ

При практическом использовании любого прогноза вероятности клика в текущей системе показа объявлений мы можем столкнуться с нежелательными эффектами, рисками. Очевидно, что прогноз при большем объеме статистики обладает меньшей дисперсией. Поэтому для новых объявлений, например, мы сталкиваемся с высоким разбросом прогноза CTR, но в силу алгоритма отбора объявления для показа, мы получаем большие риски того, что минимальная цена попадания в интересующий блок на странице результата поисковой системы может оказаться неоправданно высокой. Подобного рода эффекты являются нежелательными, поэтому мы должны уменьшить вероятности появления такого рода ситуаций.

С этой целью мы провели ряд симуляций, в результате чего для заданного истинного значения CTR объявления была вычислена вероятность того, что минимальная цена входа в соответствующие блоки возросла в k раз по сравнению с ценами при текущем прогнозе.

Симуляция подтвердила, что новый прогноз обладал большими рисками для объявлений с малой статистикой, поэтому может возникнуть нежелательное препятствие для участия объявления в показах – высокая начальная цена.

Предложенный вид прогноза был модифицирован таким образом, чтобы его дисперсия при малых значениях показов совпадала с той дисперсией, которая устанавливается при большом количестве показов.

Пусть число n_0 – минимальное число показов, для которого мы считаем дисперсию прогноза приемлема. Для каждого количества показов n в статистике считаем ожидаемое количество кликов M_n и стандартное отклонение std_n для этого количества кликов. Тогда желание избежать высокой дисперсионности прогноза, сохранив его несмещенность, можно выразить с помощью следующих ограничений:

1) $(std_n)_{new} = std_{n_0}$ для каждого $n < n_0$.

2) $(M_n)_{new} = M_{n_0}$ для каждого $n < n_0$.

Чтобы удовлетворить этим ограничениям достаточно произвести следующее линейное преобразование:

$$CTR^*_{new} = \frac{std_{n_0}}{std_n} \cdot CTR_{new} + \left(1 - \frac{std_{n_0}}{std_n}\right) \cdot M_n, \forall n < n_0 .$$

После такого преобразования CTR^*_{new} обладает заданным уровнем дисперсии (Рис.3).

Этим мы сократили риски рекламодателей иметь завышенные цены за клик, вызванные высокой волатильностью начального прогноза.

3.1. Эксперимент

На ранее отобранном тестовом материале была проверена новая формула (1) и получен прирост в CTR на 1.9%. Мы посчитали, что прирост такого объёма достаточен для проведения on-line-эксперимента.

По on-line-эксперименту прирост в среднем CTR по системе составил +1.53%. Рис. 4 (слева) показывает, что в области применения нового прогноза произошло ожидаемое улучшение значения CTR. На рис. 4 (справа) видно, что новый прогноз генерировал не меньшую прибыль, чем текущий.

Теперь рассмотрим еще одну важную характеристику системы, *CPC (cost-per-click)*:

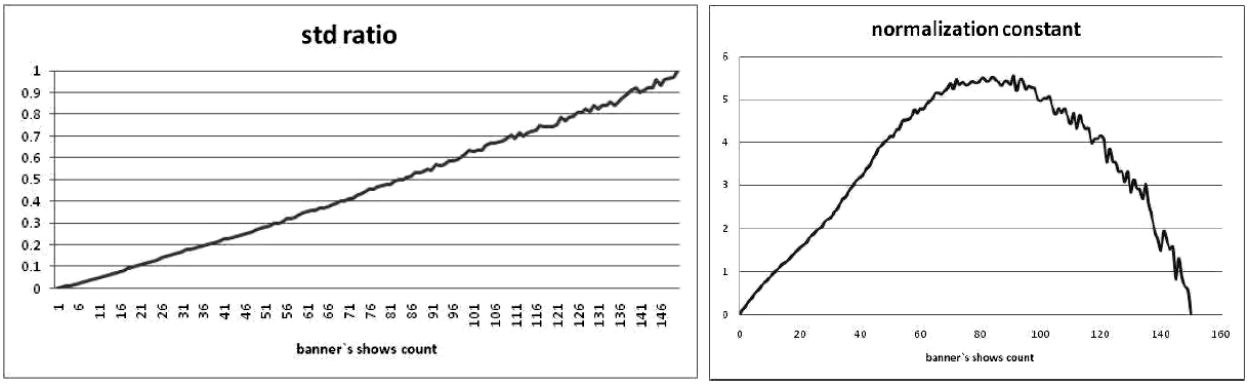

Рис.3. Слева – график отношения $\frac{std_{n_0}}{std_n}$; справа – параметр сдвига $\left(1 - \frac{std_{n_0}}{std_n}\right) \cdot M_n$

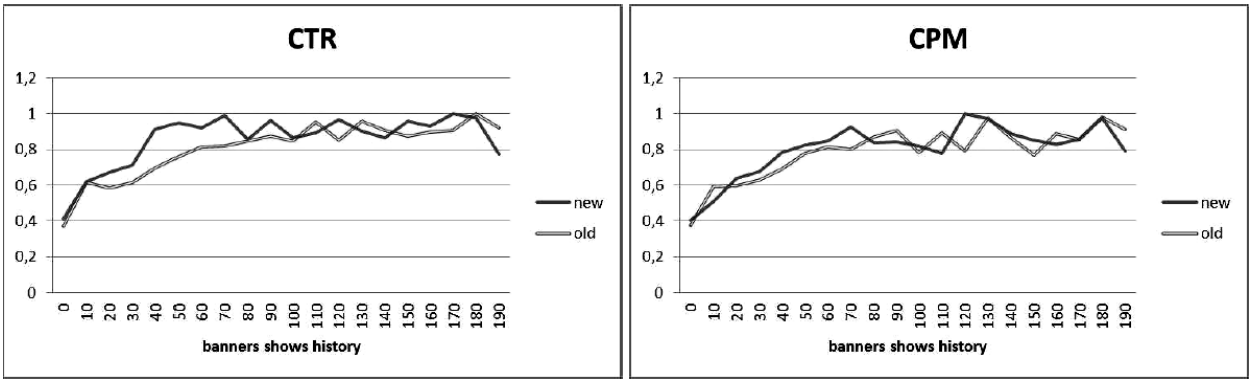

Рис.4. Слева - *CTR* ; справа - *CPM* (в зависимости от количества показов в статистике объявления)

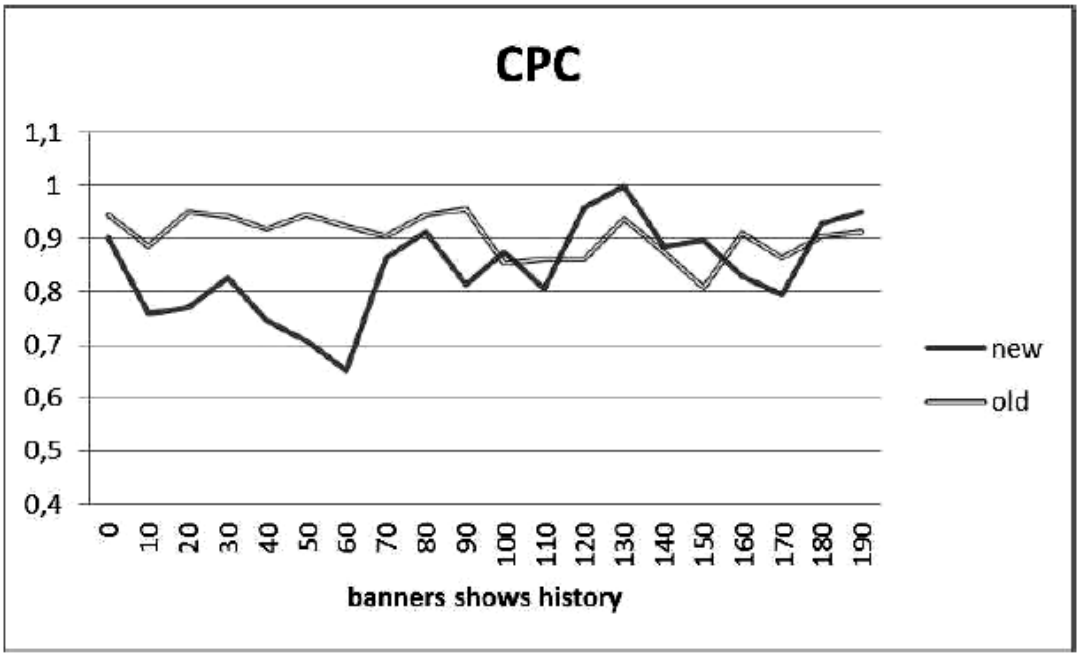

Рис.5. Динамика изменения средней цены за клик *CPC*

Из графика на рис.5 видно, что для объявлений с маленькой историей продукт стал более дешевым и, следовательно, более привлекательным. Таким образом мы увеличиваем приток клиентов и даём им более выгодные условия для размещения объявлений. Так же подобного рода эффекты могут приводить к положительным последствиям в долгосрочной перспективе: реклама на поиске становится более эффективной для рекламодателей, что стимулирует увеличение существующих бюджетов или привлечению новых.

ЗАКЛЮЧЕНИЕ

В настоящей работе, мы обсудили ряд проблем, возникающих при попытке оптимизации текущей системы показов рекламных объявлений в Яндексе.

Прежде всего было отмечено, что для любой оптимизации необходима возможность вычислять оптимизируемые показатели в новых точках. При этом запускать с этой целью каждый раз новый on-line-эксперимент не представляется возможным в силу естественной дороговизны данного подхода.

В силу недостоверности статистик для новых объявлений для прогноза вероятности клика для таких ситуаций использовался альтернативный "статический" прогноз. Как следствие возникает проблема оптимального перехода от такого начального прогноза к основному.

Свойства прогноза зачастую сильно меняются в зависимости от объема доступных статистик, в связи с этим для объявлений, находящихся в начале своего "жизненного пути" в рамках их рекламных кампаний, могут возникать нежелательные, отрицательные эффекты. В качестве примера мы привели проблему недопустимо высокой волатильности прогноза.

Были предложены методы по решению возникающих проблем и представлен результат работы построенного прогноза на реальном трафике.

Были описаны два способа для сравнения друг с другом качества работы алгоритма с разными версиями прогноза в терминах итоговых характеристик системы. В связи с чем было предложено проводить оптимизацию всех параметров off-line и лишь затем проводить итоговый on-line-экспе-

римент, как верификацию полученных результатов данной аппроксимации.

Мы описали два различных подхода для решения проблемы оптимального перехода от начального прогноза к основному. Предложенный вариант кажется достаточно убедительным в силу того, что для его построения мы использовали форму решения, которая оказалась общей и для подхода минимизации квадрата ошибок, и для байесовского подхода с достаточно широким классом априорных вероятностей для данной задачи

Задача уменьшения волатильности прогноза была обсуждена как задача редукции нежелательных эффектов. Мы описали методику вычисления необходимых показателей прогноза и предложили линейное преобразование, разрешающее эту проблему.

Итогом проделанной работы являются результаты on-line-эксперимента, проведенного для сравнения показателей системы показа рекламы, полученных с использованием предложенной модификации прогноза и текущей версии. Точность off-line-прогноза улучшения естественных показателей системы оказалось приемлемой, в связи с чем данный подход мы считаем успешным.

СПИСОК ЛИТЕРАТУРЫ

1. Ashkan A., Clarke C.L.A., Agichtein E., Guo Q. Estimating Ad Clickthrough Rate through Query Intent Analysis // WI-IAT '09. Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology. - 2009 - № 1. - P. 222-229.
2. Edelman B., Ostrovsky M., Schwarz M. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords // American Economic Review. – 2007. - № 97. – P. 242-259.
3. Broder A., Josifovski V. Introduction to Computational Advertising. – 2011. – URL: http://www.stanford.edu/class/msande239/
4. Dembczynski K., Kotlowski W., Weiss D. Predicting ads' click-through rate with decision rules //Workshop on Targeting and Ranking in Online Advertising. – 2008. – T. 2008.
5. Fawcett T. ROC Graphs: Notes and Practical Considerations for Researchers // In HP Labs Tech Report, No. HPL-2003-4.
6. Graepel T., Candela J.Q., Borchert T., Herbrich R. Web-Scale Bayesian Click-Through Rate Prediction for Sponsored Search Advertising in Microsoft's Bing Search Engine // ICMLOmnipress. – 2010. – P. 13-20.

Материал поступил в редакцию 19.09.12.

Сведения об авторах

БАУМАН Константин Евгеньевич - аспирант Математического института им. В. А. Стеклова РАН, Москва
E-mail: kbauman@yandex-team.ru

КОРНЕТОВА Анна Николаевна - аспирант Национального Исследовательского Университета - Высшая Школа Экономики, Москва
E-mail: kornetovaanna@yandex.ru.

ТОПИНСКИЙ Валерий Александрович - аспирант Национального Исследовательского Университета - Высшая Школа Экономики
E-mail: topinsky@gmail.ru

ХАКИМОВА Диляра Азатовна - аспирант Национального Исследовательского Университета - Высшая Школа Экономики
E-mail: d.a.khakimova@gmail.com

ИНФОРМАЦИОННЫЕ ЯЗЫКИ

УДК 811.93’36

В.В. Грибова, А.С. Клещев, Д.А. Крылов

Контекстно-свободные грамматики искусственных языков*

Представлен математический аппарат для описания контекстно-свободных грамматик искусственных языков: определен синтаксис языка для их описания, логическая и порождающие семантики; содержатся примеры, демонстрирующие применимость предложенного аппарата.

Ключевые слова: *искусственный язык, контекстно-свободная грамматика, контекстно-зависимая грамматика, логическая семантика, порождающая семантика*

ВВЕДЕНИЕ

В настоящей работе под искусственными языками будут пониматься языки, специально создаваемые людьми для некоторых целей, в отличие от естественных языков, о происхождении которых мы можем судить лишь гипотетически. Будем считать, что для искусственного языка точно известно множество смыслов, которые на нем можно выразить, и множество текстов, эти смыслы выражающих. Примерами искусственных языков могут служить нотная запись, языки, применяемые в научных исследованиях (языки математических и химических формул, языки математической логики), языки обработки информации (языки программирования, операционных систем, описания данных, запросов к базам данных) и др. С развитием искусственного интеллекта их дополнили языки представления знаний и онтологий.

Принятый в настоящее время подход к определению искусственных языков восходит к идеям Н. Хомского: сначала определяется синтаксис языка с помощью порождающих грамматик, а затем иными средствами определяется семантика языка в терминах его грамматики [1]. Этот подход сталкивается с определенными трудностями: из всех классов грамматик, введенных Н. Хомским, только контекстно-свободные грамматики оказались удобными для практического использования, в то время как большинство практически полезных искусственных языков являются контекстно-зависимыми; формализмы, используемые для определения семантики (в том числе контекстных условий), обычно имеют иную природу, чем порождающие грамматики.

Целью данной работы является описание математического аппарата для определения контекстно-свободных искусственных языков, допускающего в рамках одной модели описание и множества смыслов, и множества текстов, выражающих эти смыслы, а также их синтаксиса, логической и порождающей семантик.

ОСНОВНЫЕ ПОНЯТИЯ

Определение искусственного языка и его математической модели введем по аналогии с определением формального языка и его математической модели. Напомним, что формальный язык определяется как множество текстов, которые могут быть написаны на этом языке [2]. Тексты формального языка не имеют смысла, в то время как искусственный язык создается для того, чтобы передавать определенный смысл. Поэтому искусственный язык определим как множество смыслов, которые могут быть выражены на этом языке.

Математической моделью текста является конечная строка, состоящая из знаков, принадлежащих заданному конечному набору. Таким образом, математической моделью (экстенсионалом) формального языка является некоторое (бесконечное) множество таких строк. Математической моделью смысла (некоторой информации) будем считать, как это принято в области искусственного интеллекта, конечную семантическую сеть, т.е. ориентированный размеченный связный граф без циклов, содержащий конечное число вершин. Способы представления смыслов (информации) семантическими сетями определяются правилами, специфическими для каждого искусственного языка; их рассмотрение выходит за рамки данной работы. Таким образом, математической моделью (экстенсионалом) искусственного языка будем считать некоторое (бесконечное) множество конечных семантических сетей.

Грамматикой, порождающей некоторый формальный язык, является исчисление, порождающее экстенсионал этого формального языка. Грамматикой, порождающей некоторый искусственный язык, бу-

* Работа выполнена при финансовой поддержке ДВО РАН в рамках Программы ОНИТ РАН (проект 12-I-ОНИТ-04) и РФФИ (проект 12-07-00179).

дем считать исчисление, порождающее экстенсионал этого искусственного языка. Грамматика искусственного языка может быть дополнена набором правил (проекцией), дающих возможность перейти от смысла (информации) к тексту (выражающему эту информацию), т.е. позволяющих по семантической сети построить соответствующую ей конечную строку из заданного конечного набора знаков. Здесь можно увидеть поверхностную аналогию между предлагаемым подходом к определению искусственных языков и теорией «смысл <–> текст» И.А. Мельчука для моделирования естественных языков [3].

Для представления грамматик искусственных языков требуется некоторый метаязык. Поскольку такие грамматики должны определять (бесконечные) подмножества множества семантических сетей, в качестве метаязыка естественно взять логический язык. В этом случае грамматика искусственного языка есть логическая формула этого метаязыка, истинная для тех и только тех семантических сетей, которые принадлежат этому искусственному языку. Поскольку логическая семантика метаязыка (способ вычисления значения истинности его формул) будет определена относительно семантических сетей, т.е. объектов, имеющих структуру, естественно ввести структуру и в синтаксис метаязыка – формулами метаязыка будут определенным образом размеченные графы (вообще говоря, с циклами), в каждом из которых выделена одна вершина, называемая начальной вершиной этой формулы. Наряду с логической семантикой метаязыка будет определена и его порождающая семантика. Начальным состоянием порождающего процесса является начальная вершина (корень) порождаемой сети, которая объявляется активной; с каждой вершиной порождаемой сети связывается формула метаязыка; на очередном шаге порождающего процесса из множества активных вершин выбирается одна, из которой выполняется очередной шаг порождения в соответствии с формулой, связанной с этой вершиной (при этом вершина, из которой выполнен шаг порождения, перестает быть активной, но могут возникнуть новые активные вершины); порождающий процесс заканчивается, когда в порождаемой сети нет ни одной активной вершины. Логическая и порождающая семантики метаязыка связаны следующим естественным условием: по каждой формуле метаязыка может быть порождено множество тех и только тех семантических сетей, относительно которых эта формула истинна.

ИЕРАРХИЧЕСКИЕ СЕМАНТИЧЕСКИЕ СЕТИ

Мы рассматриваем иерархические семантические сети, определяемые следующим образом. Иерархическая семантическая сеть есть связный ориентированный граф без циклов, в котором каждая дуга имеет метку, а вершины могут быть одного из двух типов – простые и структурные; одна из вершин сети, в которую не входит ни одна дуга, выделяется и называется ее корнем. Каждая простая терминальная вершина сети имеет в качестве метки константу некоторого сорта. Корень сети имеет две метки-термина, из которых первая метка есть метка искусственного языка (класса), а вторая – индивидуальная метка сети. Каждая структурная вершина сети является контейнером, содержащим упорядоченное конечное множество иерархических семантических сетей с одной и той же меткой класса и попарно различными индивидуальными метками. Далее вместо термина «иерархическая семантическая сеть» для краткости будет использоваться термин «сеть».

КОНТЕКСТНО-СВОБОДНЫЕ ГРАММАТИКИ ИСКУССТВЕННЫХ ЯЗЫКОВ

Контекстно-свободная грамматика искусственного языка есть формула без переменных метаязыка, начальная вершина которой имеет метку – название этого искусственного языка.

Синтаксис формул без переменных. Формула без переменных метаязыка может относиться к одной из следующих групп (рис. 1): простая формула без переменных; простая кванторная формула без переменных; унарная формула без переменных; пропозициональная формула без переменных; структурная кванторная формула без переменных.

Чтобы читателю было легче различать группы, к которым относятся те или иные формулы, каждой группе формул будет сопоставлен свой графический образ.

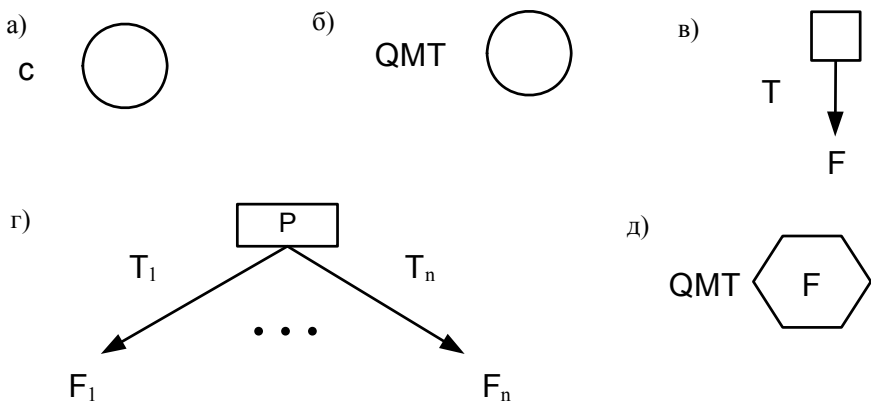

Рис. 1. Группы формул без переменных: а) простая, б) простая кванторная, в) унарная, г) пропозициональная, д) структурная кванторная

Простая формула без переменных есть граф, состоящий из единственной вершины с меткой **c** (рис. 1а). Эту единственную вершину будем называть начальной вершиной простой формулы без переменных. Метка с является константой.

Простая кванторная формула без переменных есть граф, состоящий из единственной вершины с меткой, имеющей вид QMT, где Q – знак квантора, M – описание (конечного или бесконечного) множества, а T – термин (рис. 1б). Знаком квантора может быть: ∀ (для всех), ∃ (существует), ∃2 (существует не менее двух), ∃? (существует, но не для всех), ∃! (существует и единственен), ∃[] (существует подынтервал). Множество кванторов является расширяемым.

Описание конечного множества может иметь вид $\{c_1, \ldots, c_n\}$, где $c_1, \ldots, c_n$ – попарно различные константы, либо быть целым конечным интервалом; в этом случае знаком квантора могут быть ∀, ∃, ∃2, ∃? или ∃!. Описание бесконечного множества может быть названием сорта, неименованным множеством «*» или вещественным конечным интервалом; в этом случае знаком квантора может быть ∃, ∃2, ∃!. Если знак квантора ∃[], то описанием множества может быть только конечный целый или вещественный интервал.

Названием сорта может быть: "строка", "целый", "вещественный", "целый интервал", "вещественный интервал", "дата–время". Множество сортов является расширяемым.

Унарная формула без переменных есть граф, состоящий из начальной вершины и дуги с меткой T (термином), выходящей из начальной вершины и входящей в начальную вершину некоторой формулы без переменных F (рис. 1в).

Пропозициональная формула без переменных есть граф, состоящий из начальной вершины с пропозициональной меткой P и выходящих из нее *n* дуг (не менее двух), каждая из которых имеет метку T_i (термин; i – от 1 до *n*; все эти метки должны быть попарно различны) и входит в начальную вершину некоторой формулы без переменных F_i (рис. 1г). Пропозициональными метками P являются: & – конъюнкция, ∨ – дизъюнкция, | – исключающее "или". Множество пропозициональных меток является расширяемым. Любая из дуг, выходящая из начальной вершины с пропозициональной меткой &, может иметь метку факультативности "[]".

Структурная кванторная формула без переменных состоит из начальной вершины с меткой, имеющей вид QMT, где Q – знак квантора, M – описание (конечного или бесконечного) множества, а T – термин, и формулы без переменных F, начальная вершина которой изображается внутри начальной вершины структурной кванторной формулы без переменных (рис. 1д). Описание конечного множества может иметь вид $\{c_1, \ldots, c_n\}$, где $c_1, \ldots, c_n$ – попарно различные константы, либо быть целым конечным интервалом; в этом случае знаком квантора может быть ∀, ∃, ∃2, ∃?. Описание бесконечного множества может быть названием сорта, неименованным множеством «*» или вещественным конечным интервалом; в этом случае знаком квантора может быть ∃, ∃2.

Логическая семантика формул без переменных. Формула без переменных F метаязыка истинна относительно сети N, если F истинна на корне R сети N, а метка входной вершины формулы без переменных F совпадает с меткой класса корня R сети N.

Простая формула без переменных F (рис. 2) истинна на вершине V сети N, если V является терминальной и простой, а ее меткой является константа **c**.

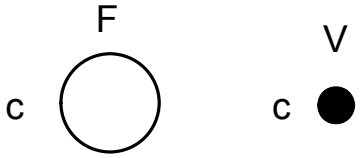

Рис. 2. Семантика простой формулы без переменных

Если в ***простой кванторной формуле без переменных*** (рис. 1б) описание конечного множества имеет вид $\{c_1, \ldots, c_n\}$, где $c_1, \ldots, c_n$ – попарно различные константы, то конечное множество состоит из элементов $c_1, \ldots, c_n$; если же описание конечного множества есть конечный целый интервал, то конечное множество состоит из всех целых констант, принадлежащих этому интервалу. Если описание бесконечного множества есть название сорта, то бесконечное множество состоит из констант, представляющих все значения этого сорта; если описание бесконечного множества есть «*», то бесконечное множество состоит из всех констант специального вида. Если описание бесконечного множества есть конечный вещественный интервал, то бесконечное множество состоит из всех вещественных констант, принадлежащих этому интервалу.

Если метка входной вершины ***простой кванторной формулы без переменных*** F (рис. 3а) имеет вид ∀ FS T, где FS – описание конечного множества $\{d_1, \ldots, d_n\}$, а T – термин, то F истинна на вершине V сети N, если V является структурной и терминальной вершиной, содержащей *n* сетей, каждая из которых состоит из одного корня с меткой класса – T и индивидуальными метками – элементами множества $\{d_1, \ldots, d_n\}$.

Если метка входной вершины ***простой кванторной формулы без переменных*** F (рис. 3б) имеет вид ∃ M T, где M – описание (конечного или бесконечного) множества, а T – термин, то F истинна на вершине V сети N, если V является структурной и терминальной вершиной, содержащей непустое множество сетей, каждая из которых состоит из одного корня с меткой класса – T и индивидуальной меткой – одним из элементов множества (индивидуальные метки попарно различны).

Если метка входной вершины ***простой кванторной формулы без переменных*** F (рис. 4а) имеет вид ∃2 M T, где M – описание (конечного или бесконечного) множества, а T – термин, то F истинна на вершине V сети N, если V является структурной и терминальной

вершиной, содержащей не менее двух сетей, каждая из которых состоит из одного корня с меткой класса – T и индивидуальной меткой – одним из элементов множества (индивидуальные метки попарно различны).

Если метка входной вершины ***простой кванторной формулы без переменных*** F (рис. 4б) имеет вид ∃? FS T, где FS – описание конечного множества $\{d_1, \dots, d_n\}$, а T – термин, то F истинна на вершине V сети N, если V является структурной и терминальной вершиной, содержащей менее чем *n* сетей, каждая из которых состоит из одного корня с меткой класса – T и попарно различными индивидуальными метками – элементами множества $\{d_1, \dots, d_n\}$.

Если метка входной вершины ***простой кванторной формулы без переменных*** F (рис. 5а) имеет вид ∃! M T, где M – описание (конечного или бесконечного) множества, а T – термин, то F истинна на вершине V сети N, если V является простой и терминальной вершиной, а ее меткой является один из элементов множества.

Если метка входной вершины ***простой кванторной формулы без переменных*** F (рис. 5б) имеет вид ∃[] I T, где I – конечный целый (вещественный) интервал, а T – термин, то F истинна на вершине V сети N, если V является простой и терминальной вершиной, а ее метка – это целый (вещественный) интервал, являющийся подынтервалом интервала I.

Унарная формула без переменных F (рис. 6) истинна на вершине V сети N, если V является простой вершиной, из нее выходит в точности одна дуга с меткой T и эта дуга входит в вершину V_1, на которой истинна формула без переменных F_1.

Рис. 3. Семантика простой кванторной формулы без переменных: а) с квантором ∀ и б) с квантором ∃

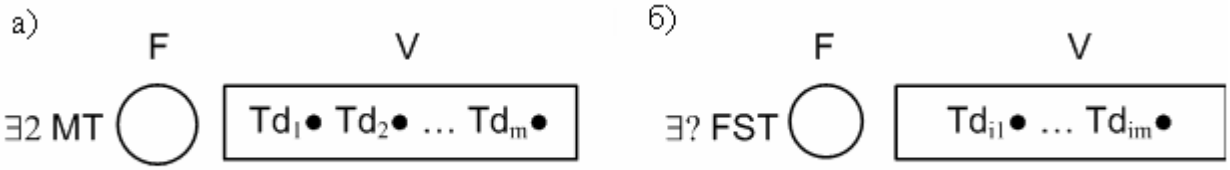

Рис. 4. Семантика простой кванторной формулы без переменных: а) с квантором ∃2 и б) с квантором ∃?

Рис. 5. Семантика простой кванторной формулы без переменных: а) с квантором ∃! и б) с квантором ∃[]

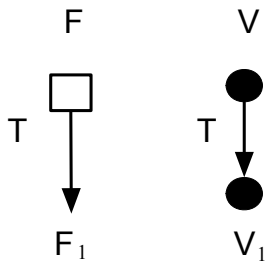

Рис. 6. Семантика унарной формулы

Если пропозициональной меткой начальной вершины W ***пропозициональной формулы без переменных*** F (рис. 7) является &, а некоторые дуги, выходящие из W, имеют метки [], то F истинна на вершине V сети N, если V является простой вершиной, для каждой дуги с меткой T_i, выходящей из V, существует дуга с такой же меткой T_i, выходящая из W, для каждой дуги с меткой T_i, выходящей из W и не имеющей метки [], существует дуга с такой же меткой T_i, выходящая из V, а каждая формула без переменных F_i истинна на вершине, в которую входит дуга с меткой T_i, выходящая из вершины V.

Если пропозициональной меткой начальной вершины W ***пропозициональной формулы без переменных*** F (рис. 8) является $\vee$, то F истинна на вершине V сети N, если V является простой вершиной, существует подмножество дуг, выходящих из W, между которым и множеством дуг, выходящих из V, существует взаимно-однозначное соответствие, при котором метки T_i соответствующих дуг совпадают, а формулы без переменных Fi, в начальные вершины которых входят дуги, выходящие из W, истинны на вершинах, в которые входят соответствующие дуги с метками T_i, выходящие из V.

Если пропозициональной меткой начальной вершины W ***пропозициональной формулы без переменных*** F (рис. 9) является |, то F истинна на вершине V сети N, если V является простой вершиной, из нее выходит единственная дуга, существует дуга, выходящая из W, такая, что метки T_i этих дуг совпадают, а формула без переменных F_i, в начальную вершину которой входит дуга с меткой T_i, выходящая из W, истинна на вершине, в которую входит дуга с меткой T_i, выходящая из V.

Если метка входной вершины ***структурной кванторной формулы без переменных*** F (рис. 10а) имеет вид $\forall$ FS T, где FS – описание конечного множества $\{d_1, \dots, d_n\}$, а T – термин, то F истинна на вершине V сети N, если V является структурной вершиной, содержащей *n* сетей, каждая из которых имеет метку класса – T и индивидуальные метки – попарно различные элементы множества $\{d_1, \dots, d_n\}$, и на корне каждой из которых истинна формула без переменных F_1.

Если метка входной вершины ***структурной кванторной формулы без переменных*** F (рис. 10б) имеет вид $\exists$ M T, где M – описание (конечного или бесконечного) множества, а T – термин, то F истинна на вершине V сети N, если V является структурной вершиной, содержащей непустое множество сетей, каждая из которых имеет метку класса – T и индивидуальные метки – попарно различные элементы множества, и на корне каждой из которых истинна формула без переменных F_1.

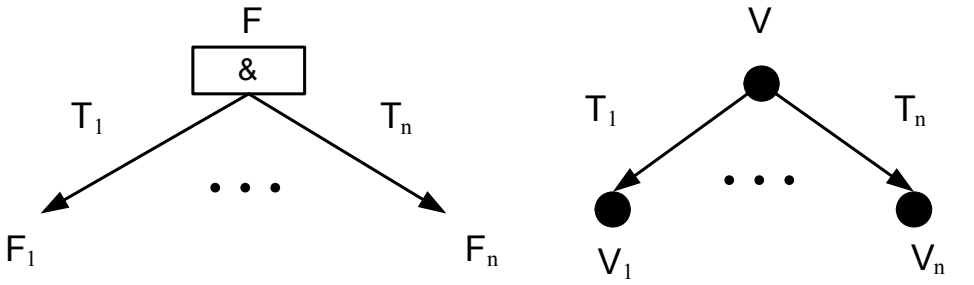

Рис. 7. Семантика конъюнкции без переменных

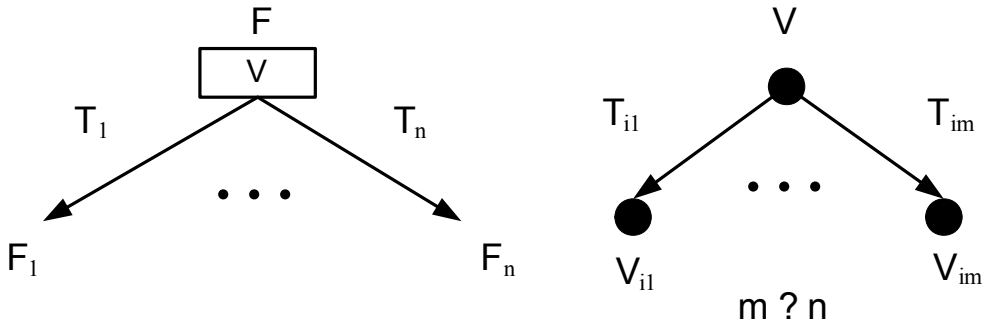

Рис. 8. Семантика дизъюнкции без переменных

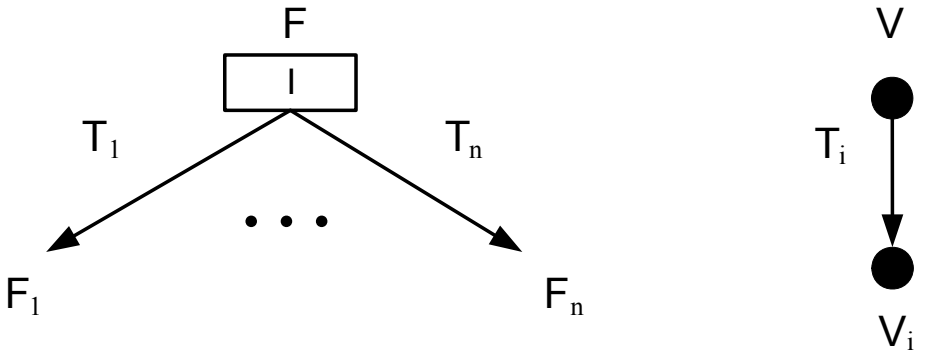

Рис. 9. Семантика исключающего ИЛИ без переменных

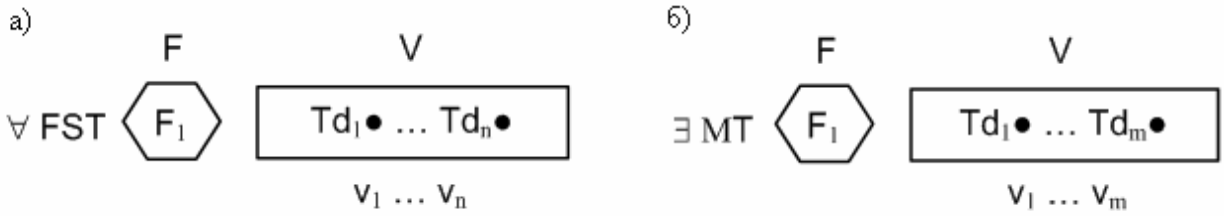

Рис. 10. Семантика структурной кванторной формулы без переменных F: а) с квантором ∀ и б) с квантором ∃.

Если метка входной вершины ***структурной кванторной формулы без переменных*** F (рис. 11а) имеет вид ∃2 M T, где M – описание (конечного или бесконечного) множества, а T – термин, то F истинна на вершине V сети N, если V является структурной вершиной, содержащей не менее двух сетей, каждая из которых имеет метку класса – T и индивидуальные метки – попарно различные элементы множества, и на корне каждой из которых истинна формула без переменных F_1.

Если метка входной вершины ***структурной кванторной формулы без переменных*** F (рис. 11б) имеет вид ∃? FS T, где FS – описание конечного множества $\{d_1, \dots, d_n\}$, а T – термин, то F истинна на вершине V сети N, если V является структурной вершиной, содержащей менее чем n сетей, каждая из которых имеет метку класса – T и индивидуальные метки – попарно различные элементы множества $\{d_1, \dots, d_n\}$, и на корне каждой из которых истинна формула без переменных F_1.

Порождающая семантика формул без переменных. Сеть N порождается по формуле без переменных F в ходе следующего порождающего процесса.

На первом шаге порождается корень сети N, метка класса которого совпадает с меткой начальной вершины формулы без переменных F, а индивидуальная метка порождается. Корень сети N становится ее активной вершиной, а начальная вершина формулы без переменных F ставится ей в соответствие.

Если активной вершине V сети N соответствует начальная вершина ***простой формулы без переменных*** F, то вершина V становится простой и терминальной с меткой **c** (рис. 2). После этого вершина V перестает быть активной.

Если активной вершине V сети N соответствует начальная вершина ***простой кванторной формулы без переменных*** F (рис. 3а), метка которой имеет вид ∀ FS T, где FS – описание конечного множества $\{d_1, \dots, d_n\}$, а T – термин, то вершина V становится структурной и терминальной вершиной, в которой, как в контейнере, порождается *n* сетей, каждая из которых состоит из одного корня с меткой класса – T и индивидуальными метками – элементами множества $\{d_1, \dots, d_n\}$.

Если активной вершине V сети N соответствует начальная вершина ***простой кванторной формулы без переменных*** F (рис. 3б), метка которой имеет вид ∃ M T, где M – описание (конечного или бесконечного) множества, а T – термин, то вершина V становится структурной и терминальной вершиной, в которой, как в контейнере, порождается непустое конечное множество сетей, каждая из которых состоит из одного корня с меткой класса – T и индивидуальной меткой – одним из элементов множества (индивидуальные метки попарно различны).

Если активной вершине V сети N соответствует начальная вершина ***простой кванторной формулы без переменных*** F (рис. 4а), метка которой имеет вид ∃2 M T, где M – описание (конечного или бесконечного) множества, а T – термин, то вершина V становится структурной и терминальной вершиной, в которой, как в контейнере, порождается не менее двух сетей, каждая из которых состоит из одного корня с меткой класса – T и индивидуальной меткой – одним из элементов множества (индивидуальные метки попарно различны).

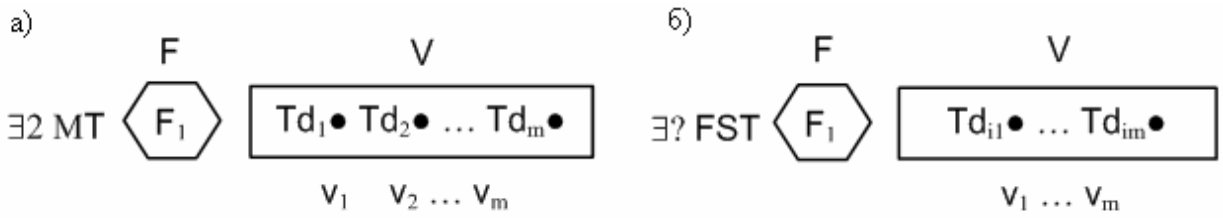

Рис. 11. Семантика структурной кванторной формулы без переменных F:
а) с квантором ∃2 и б) с квантором ∃?

Если активной вершине V сети N соответствует начальная вершина ***простой кванторной формулы без переменных*** F (рис. 4б), метка которой имеет вид ∃? FS T, где FS – описание конечного множества $\{d_1, \dots, d_n\}$, а T – термин, то вершина V становится структурной и терминальной вершиной, в которой, как в контейнере, порождается менее чем *n* сетей, каждая из которых состоит из одного корня с меткой класса – T и попарно различными индивидуальными метками – элементами множества $\{d_1, \dots, d_n\}$.

Если активной вершине V сети N соответствует начальная вершина ***простой кванторной формулы без переменных*** F (рис. 5а), метка которой имеет вид ∃! M T, где M – описание (конечного или бесконечного) множества, а T – термин, то вершина V становится простой и терминальной вершиной, а ее меткой – один из элементов множества.

Если активной вершине V сети N соответствует начальная вершина ***простой кванторной формулы без переменных*** F (рис. 5б), метка которой имеет вид ∃[] I T, где I – конечный целый (вещественный) интервал, а T – термин, то вершина V становится простой и терминальной вершиной, а ее меткой – целый (вещественный) интервал, являющийся подынтервалом интервала I. После этого вершина V перестает быть активной.

Если активной вершине V сети N соответствует начальная вершина ***унарной формулы без переменных*** F (рис. 6), то в сети N порождается дуга с меткой T, выходящая из вершины V. Вершина V перестает быть активной, а вершина, в которую эта дуга входит, становится активной. Ей ставится в соответствие вершина, в которую входит дуга в унарной формуле без переменных F.

Если активной вершине V сети N соответствует начальная вершина W ***пропозициональной формулы без переменных*** F (рис. 7), вершина W имеет пропозициональную метку &, из вершины W выходит *n* дуг с метками $T_1, \dots, T_n$, то вершина V становится простой, порождается *n* дуг, выходящих из вершины V, с теми же метками. Вершина V перестает быть активной, а вершины, в которые эти дуги входят, становятся активными. Им ставятся в соответствие вершины, в которые входят дуги с теми же метками в пропозициональной формуле без переменных F. Если дуга в пропозициональной формуле без переменных F имеет метку факультативности, то дуга в сети N с той же меткой может и не порождаться. Если все дуги в пропозициональной формуле без переменных F имеют метки факультативности и в сети не порождается ни одной дуги, выходящей из вершины V, то такая ситуация считается ошибочной.

Если активной вершине V сети N соответствует начальная вершина W ***пропозициональной формулы без переменных*** F (рис. 8), вершина W имеет пропозициональную метку ∨, из вершины W выходит *n* дуг с метками $T_1, \dots, T_n$, то в сети N порождаются дуги числом не больше *n*, выходящие из вершины V, а метки этих дуг являются некоторыми попарно различными элементами множества $T_1, \dots, T_n$. Вершина V перестает быть активной, а вершины, в которые входят порожденные дуги, становятся активными. Им ставятся в соответствие вершины, в которые входят дуги с теми же метками в пропозициональной формуле без переменных F.

Если активной вершине V сети N соответствует начальная вершина W ***пропозициональной формулы без переменных*** F (рис. 9), вершина W имеет пропозициональную метку |, из вершины W выходит *n* дуг с метками $T_1, \dots, T_n$, то в сети N порождается дуга, выходящая из вершины V с меткой – одним из элементов множества $T_1, \dots, T_n$. Вершина V перестает быть активной, а вершина, в которую входит порожденная дуга, становится активной. Ей ставится в соответствие вершина, в которую входит дуга с той же меткой в пропозициональной формуле без переменных F.

Если активной вершине V сети N соответствует начальная вершина W ***структурной кванторной формулы без переменных*** F (рис. 10а), метка которой имеет вид ∀ FS T, где FS – описание конечного множества $\{d_1, \dots, d_n\}$, а T – термин, то вершина V становится структурной вершиной, в которой, как в контейнере, порождается *n* вершин, каждая из которых имеет метку класса – T и индивидуальную метку – элемент множества $\{d_1, \dots, d_n\}$ (индивидуальные метки попарно различны). Вершина V перестает быть активной, а вновь порожденные вершины становятся активными. Каждой из них ставится в соответствие начальная вершина формулы без переменных F_1.

Если активной вершине V сети N соответствует начальная вершина W ***структурной кванторной***

формулы без переменных F (рис. 10б), метка которой имеет вид ∃ M T, где M – описание (конечного или бесконечного) множества, а T – термин, то вершина V становится структурной вершиной, в которой, как в контейнере, порождается конечное множество вершин, каждая из которых имеет метку класса – T и индивидуальную метку – элемент множества (индивидуальные метки попарно различны). Вершина V перестает быть активной, а вновь порожденные вершины становятся активными. Каждой из них ставится в соответствие начальная вершина формулы без переменных F_1.

Если активной вершине V сети N соответствует начальная вершина W ***структурной кванторной формулы без переменных*** F (рис. 11а), метка которой имеет вид ∃2 M T, где M – описание (конечного или бесконечного) множества, а T – термин, то вершина V становится структурной вершиной, в которой, как в контейнере, порождается не менее двух вершин, каждая из которых имеет метку класса – T и индивидуальную метку – элемент множества (индивидуальные метки попарно различны). Вершина V перестает быть активной, а вновь порожденные вершины становятся активными. Каждой из них ставится в соответствие начальная вершина формулы без переменных F_1.

Если активной вершине V сети N соответствует начальная вершина W ***структурной кванторной формулы без переменных*** F (рис. 11б), метка которой имеет вид ∃? FS T, где FS – описание конечного множества $\{d_1, \dots, d_n\}$, а T – термин, то вершина V становится структурной вершиной, в которой, как в контейнере, порождается менее *n* вершин, каждая из которых имеет метку класса – T и индивидуальную метку – элемент множества $\{d_1, \dots, d_n\}$ (индивидуальные метки попарно различны). Вершина V перестает быть активной, а вновь порожденные вершины становятся активными. Каждой из них ставится в соответствие начальная вершина формулы без переменных F_1.

Грамматика, определенная в настоящем разделе, является контекстно-свободной в том смысле, что при порождении сети N из любой ее активной вершины V, соответствующей начальной вершине формулы без переменных F, никакой контекст в сети N или в какой-либо другой сети не учитывается.

Пример 1. На рис. 12 представлена контекстно-свободная грамматика языка описания простых баз наблюдений [4], а на рис. 13 – сеть, порожденная по этой грамматике. Информация (простая база наблюдений) состоит из множества наблюдений, каждое из которых имеет уникальное название и возможные значения – множество названий этих значений. Сеть, представляющая эту информацию, состоит из одной структурной вершины с меткой класса "База наблюдений" и индивидуальной меткой – названием конкретной базы наблюдений (в данном примере – «эндокринология»). Эта структурная вершина, как контейнер, содержит упорядоченное множество сетей. Корень каждой сети является простой вершиной с меткой класса "Наблюдение" и индивидуальной меткой – названием наблюдения (в примере – «аппетит», «алопеция», …). Из этого корня выходит единственная дуга с меткой "Возможные значения", входящая в структурную вершину, которая, как контейнер, содержит упорядоченное множество сетей. Каждая сеть состоит только из корня – простой вершины с меткой класса "Значение" и индивидуальной меткой – названием значения.

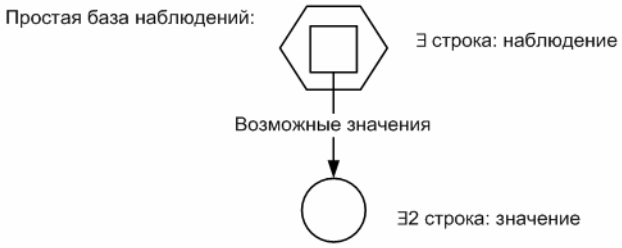

Рис. 12. Контекстно-свободная грамматика языка описания простых баз наблюдений

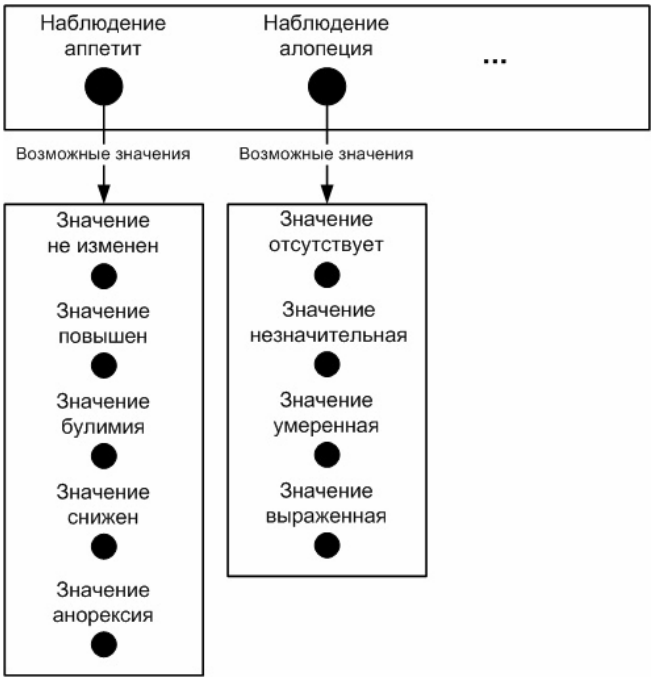

Рис. 13. Иерархическая семантическая сеть, представляющая простую базу наблюдений «эндокринология»

.

Пример 2. На рис. 14 представлена контекстно-свободная грамматика языка описания реальных баз наблюдений [4]. Реальная база наблюдений представляет собой рекурсивную структуру. Она состоит из групп и/или наблюдений, причем каждая группа, в свою очередь, состоит из групп и/или наблюдений. Наблюдение может иметь качественные значения, количественные (целые или вещественные) значения или составные значения. Составные значения состоят из характеристик, каждая из которых может иметь качественные значения, количественные (целые или вещественные) значения или составные значения.

Рис. 14. Контекстно-свободная грамматика языка описания реальных баз наблюдений

ЗАКЛЮЧЕНИЕ

В настоящей статье описан метаязык для задания контекстно-свободных грамматик искусственных языков в виде (бесконечного) подмножества множества семантических сетей. Грамматика искусственного языка представляет собой логическую формулу этого языка. Логическими формулами являются размеченные графы, в каждом из которых задана одна вершина, называемая начальной вершиной этой формулы. Выделены следующие группы логических формул: простая формула без переменных, простая кванторная формула без переменных, унарная формула без переменных, пропозициональная формула без переменных, структурная кванторная формула без переменных. Представлены логическая и порождающая семантики языка.

СПИСОК ЛИТЕРАТУРЫ

1. Семантика языков программирования: Сборник статей / пер. с англ. А.Н. Бирюкова, В.А. Серебрякова; под общ. ред. В.М. Курочкина. – М.: Мир, 1980. – 395 с.
2. Ахо А., Ульман Дж. Теория синтаксического анализа, перевода и компиляции. Т. 1. Синтаксический анализ. – М.: Мир, 1978. – 612 с.
3. Мельчук И.А. Опыт теории лингвистических моделей «Смысл <–> Текст»: Семантика. Синтаксис. – М.: Наука, 1974. – 314 с.
4. Никифорова Н.Ю., Черняховская М.Ю. Базы наблюдений в компьютерном банке медицинского знания // Материалы Седьмой Международной научно-технической конференции "Искусственный интеллект. Интеллектуальные и многопроцессорные системы", Кацивели, 24–29 сентября 2006. – Таганрог: Изд-во ТРТУ, 2006. Т. 1. – С. 76–80.

Материал поступил в редакцию 23.11.12.

Сведения об авторах

ГРИБОВА Валерия Викторовна – доктор технических наук, старший научный сотрудник, заведующий лабораторией интеллектуальных систем Института автоматики и процессов управления Дальневосточного отделения Российской академии наук (ИАПУ ДВО РАН), г. Владивосток
E-mail: gribova@iacp.dvo.ru

КЛЕЩЕВ Александр Сергеевич – доктор физико-математических наук, профессор, главный научный сотрудник ИАПУ ДВО РАН, г. Владивосток
E-mail: kleschev@iacp.dvo.ru

КРЫЛОВ Дмитрий Александрович – ведущий инженер-программист лаборатории интеллектуальных систем ИАПУ ДВО РАН, г. Владивосток
\e-mail: dmalkr@gmail.com

УДК 81’374.2’373.46

Е.В. Падучева

Онтологические категории имен речи: проблема категориальной неоднозначности*

Задача работы – выявление онтологических категорий в тематическом классе имен речи. Основные категории, к которым принадлежат отглагольные имена речи, – это действие, текст, содержание, деятельность. Выявлены типы регулярной многозначности: действие/текст (как у слова «заявление»), действие/содержание (как у слова «требование»), текст/содержание (как у слова «инструкция»). Имена деятельности, типа «злословие», часто ведут себя как имена массы.

Ключевые слова: *онтология, тематические классы, акциональные классы, регулярная многозначность, действие, текст*

1. ОНТОЛОГИЧЕСКАЯ КАТЕГОРИЯ И ТЕМАТИЧЕСКИЙ КЛАСС

Статья посвящена семантической классификации имен речи. Среди них большинство составляют отглагольные и отпредикатные имена (такие как *высказывание, намек*). Поэтому для начала следует сказать два слова о классификациях глаголов и предикатов, которые разработаны более полно и могут служить естественной основой дальнейшего. К настоящему времени сложилась традиция различать две семантические классификации глагольной лексики (см., например, [1]).

1. С одной стороны, есть традиционная тематическая классификация, которая различает **тематические** классы. Тематические классы – это наследники семантических полей Трира и других ранних опытов этого рода. Тематическими являются классы глаголов **движения, речи, воли, восприятия, знания/мнения, эмоций, физического воздействия, создания, существования, звука** и т.д. – их несколько десятков. См., например, [2].

2. С другой стороны, есть классификация глаголов по Вендлеру–Маслову, которая порождает **акциональные** классы, наследники аспектуальных классов Зено Вендлера (см. [3–5]. Это **действия** (accomplishments)**, деятельности** (activities)**, состояния, процессы** и др. (Термин акциональный класс обоснован тем, что центральным здесь является класс **действие** – action.) В применении к поствендлеровским классам используется, помимо термина акциональный класс, термин **таксономическая** (или **онтологическая**) **категория**. См. [1, 6].

Категории определяют грамматические аспекты поведения слова, в том числе – парадигму аспектуальных значений, залог и пассивные валентности. И только актантная структура слова ориентируется на тематический класс.

Переходя теперь к именам речи, можно сказать, что свой тематический класс они наследуют от глаголов, а что касается онтологических категорий, то они у имен во многом свои. Задача работы состоит в том, чтобы выявить для имен речи набор онтологических категорий – такой, чтобы общность категории у слов предопределяла в какой-то степени общность их смысла. Существенно то, что неоднозначность слова часто можно представить как мену категории. Неоднозначность слова, представленная в толковых словарях, – это, в большой своей части, **регулярная многозначность** по Ю.Д. Апресяну [7]. И частая разновидность неоднозначности – это неоднозначная категориальная принадлежность. Так обстоит дело у глаголов; можно думать, так оно обстоит и у имен.

Имена речи обнаруживают тесную связь с именами ментальной сферы, что естественно, поскольку слова выражают мысль (ср. известную максиму «Say what you think!»). Будем считать, что слово относится к ментальной сфере, если ментальным является его и с х о д н о е значение. Например, слово *аргумент* = ‘то, что подтверждает’ относится к ментальной сфере, хотя аргументы, как правило, формулируются словами какого-то языка. Т.е. в класс имен речи слова с ментальным значением попадают только при наличии существенного речевого компонента, например, *тезис, вывод*.

В центре внимания в этой работе находятся **предикатные** имена речи, такие как *просьба* (от *просить*). Слово *язык* и слова, в толковании которых участвует слово *язык* (например, *интонация*), остаются за пределами рассмотрения.

* Данная работа была выполнена при финансовой поддержке РГНФ, грант № 11-04-00488а «Акциональные классы и актантные типы предикатных имен: семантика, грамматика, словарь»

Основы тематической классификации глаголов (и, соответственно, отглагольных имен) можно считать заложенными (см., например, семантические классы глаголов, используемые в Национальном корпусе русского языка (http://www.ruscorpora.ru/). Категориальная же классификация пока полностью не установилась – ни для глаголов, ни тем более для имен.

По ассортименту тематических классов отглагольные имена вполне вкладываются в глаголы. А категории отглагольных имен лишь отчасти повторяют категории глаголов – нельзя забывать, что имя существительное, как бы оно ни сохраняло смысл глагола, представляет любую субстанцию как нечто предметное.

В работе [8] были выявлены основные категории в классе предикатных имен эмоций – это состояния, отношения, чувства, свойства и события. Задача нашей работы – выявить категории отглагольных имен речи.

Отглагольные имена речи, согласно толковым словарям, часто неоднозначны. Так, в контексте *публичное заявление* имя *заявление* – это действие, а в сочетании *подать заявление* это текст, вещь. Работа имеет целью показать, что неоднозначность предикатных имен речи во многих случаях является меной онтологической категории, и в этом смысле – регулярна. С семантической точки зрения имена категории – это, как правило, метонимический сдвиг, т.е. перераспределение фокуса внимания среди компонентов толкования. Это положение намечено в работе [7] и развито в [6, 9].

Регулярная многозначность с успехом переосмысляется как семантическая деривация [9–11]. Именно в этом «динамическом» ключе она рассматривается в данной работе: как правило, имеет смысл говорить об исходной и производной категории слова. У слова *комментарий*, например, отчетливо противопоставлены употребления в категориях Действие, Содержание и Текст, и исходной является категория Действие: *комментарий* – это:

1) действие <по глаголу *комментировать* = ‘разъяснять содержание интеллектуального объекта’> (*комментариям не было конца*);

2) содержание действия = результат: то, что разъясняет интеллектуальный объект (*его комментарий состоит в том, что …*);

3) текст <содержащий разъяснение> (*рукопись была издана с подробными комментариями*).

В пределах имен одного тематического класса главный вид семантической деривации – это мена категории. Однако, что касается имен речи (в отличие, например, от имен эмоции), то им как раз свойственна не только категориальная, но и тематическая многозначность. Дело в том, что между семантической сферой речи и ментально-эмоциональной сферой имеется тесная связь. В самом деле, мысль обычно выражается словами, а чувство – хоть и не всегда, но часто. Даже воля примыкает к тематической сфере речи (*Какие будут пожелания?*).

Слово *справка* имеет, по МАС (Малый академический словарь), три значения: действие (*телефон для справок*), сведение (*собирает о нем справки*); документ (*справка с места работы*).

Тематическая принадлежность слова может быть связана с категориальной. Так, слово *выражение* в категории Действие имеет не речевое значение: *избрать экстравагантный способ выражения чувств*. Оно же может иметь категорию Содержание: *Это было выражением внутренних противоречий*. И только в категории Текст (в значении ‘речевой оборот’) оно относится к тематической сфере речи (*Откуда выражение* “*точить лясы*”?). Слово *вопрос* может выражать речевое действие – сфера речи (*отвечаю на твой вопрос*), а может иметь значение ‘проблема’ (*решить вопрос*) – ментальная сфера.

Семантика имен (как и глаголов) речи обладает свойством, которое можно назвать политематичностью. Тематический класс слова определяется наличием в его семантической структуре того или иного семантического компонента, и у имен речи их оказывается как минимум два, ср. *опровержение* (речь и мысль), *жалоба* (речь и эмоция). Множественность категорий слова – это неоднозначность, поскольку в каждом данном контексте слово, как правило, выступает в одной из своих категорий[1]. Между тем, политематичность – это не многозначность. Так, *жалоба* – это одновременно эмоция и речь; *повеление* – это одновременно речь и воля.

Для глаголов и имен сферы речи политематичность заложена в основу самого тематического класса [9, с. 356]. Дело в том, что основные тематические классы глаголов объединены единством цели и, соответственно, р е з у л ь т а т а действия (ср. глаголы деформации, контакта, создания). Между тем класс глаголов речи основан на единстве с р е д с т в а. Средство фиксирует у глаголов речи семантический компонент и тему речь. Но кроме того, есть цель речевого действия, и ей соответствует другой тематический компонент. Так, слово *убеждение* в категории Действие политематично: оно имеет компонент «воздействие на ментальное состояние» (от ‘верить’) и компонент «речь» (воздействие словом). В категории Содержание (*убеждение* (X-а), ‘то, в чем X убежден’) на первый план выходит компонент «вера», но «речь» тоже присутствует. В семантике глагола *жаловаться* два компонента, из которых ассертивным является в одном контексте один, в другом – другой; но значение слова неизменно и включает оба компонента:

(a) – Он вам *не жаловался* на меня? = ‘не говорил, что плохо’;

(b) – Как вам живется? – Я *не жалуюсь* = ‘не чувствую, что плохо’.

Выходит, что между тематическими классами речь, ментальная сфера, эмоция и даже воля нельзя провести четких границ. Можно было бы основываться на том, к какому тематическому классу относится слово в своем исходном значении. Например, слово *аргумент* = ‘то, что подтверждает’ относится прежде всего к ментальной сфере, хотя аргумент формулируется словами какого-то языка. То же верно про слово *аргументация*. Однако надежнее будет отнести такого рода слова к двум (или более) классам. Так, слово *высказывание* (в значении ‘суждение’) вхо-

[1] Есть исключение, о котором будет сказано ниже.

дит в два тематических класса – речь и знание/мнение. Имена *пожелание, претензия, просьба, упрек* тоже относятся к двум тематическим классам – эмоция и речь. А слова *протест, осуждение, одобрение*, которые обозначают эмоциональное или ментальное отношение, не включены в тематический класс речи, хотя эти отношения обычно выражаются словами, а их имена возможны в контексте *выразить* и даже *высказать*. Разница в том, что первый ряд слов имеет категорию Содержание, и это содержание может быть выражено словами. Например: *мой упрек состоит в том, что …*; а слова второго ряда этой категории не имеют: **мой протест состоит в том, что … .*

Слово *отговорка*, несмотря на его словообразовательную структуру, естественно отнести к ментальной сфере. Но, в принципе, в классе речи есть и слова с существенным ментальным компонентом, например, *тезис, консультация*.

Как правило, значение слова данного тематического класса интересует нас с точностью до категории. Но иногда необходимы более дробные деления. Так, МАС различает у слова *перекличка* следующие два значения: 1) криком давать о себе знать; 2) действие с целью узнать, все ли присутствуют. Оба эти значения относятся к категории Действие. Однако только второе значение дает производную категорию Мероприятие [12]: *иду на перекличку*.

2. КАТЕГОРИИ И ИХ ДИАГНОСТИКИ

Исходная гипотеза состоит в том, что онтологическая категория имени речи в существенной степени предсказывается набором его сочетаемостных возможностей. Так что проверка принадлежности имени к данной категории осуществляется с помощью тестов на сочетаемость. Не хотелось бы, однако, чтобы принадлежность имени к данной категории проявлялась только в сочетаемости. Можно думать, от этого мы защищены исходной предпосылкой о том, что общая сочетаемость предопределена общностью семантики. Некоторые важные аспекты сочетаемости имен и глаголов были описаны в рамках модели Смысл ⇔ Текст [13].

И.А. Мельчук характеризует лексические функции типа Oper, Func и Labor («лексико-функциональные» глаголы) как «синтаксические оформители описания ситуации с помощью существительного, ее называющего» [13, с. 93]. Эти слова названы в [8] вербализирующими операторами. На их основе оказалось возможным построить диагностики для категорий имен эмоций. Для имен речи лексические функции тоже играют роль.

Для речевой деятельности человека базовым является понятие высказывания. И слово *высказывание* категориально неоднозначно. Это может быть речевое действие (произведенное в определенный момент, в определенном месте) и содержание речевого действия, ср. (1а) и (1б):

(1) а. Его высказывание было не ко времени [не ко времени было действие];

б. Его высказывание было противоречиво [противоречиво – содержание, т.е. то, что он высказал].

Но причислить употребление категориально неоднозначного слова к той или иной категории не всегда легко.

Тест на принадлежность слова X к категории Содержание – это возможность выступать в конструкции «X состоит в том, что…» (*состоять* считается в МСТ выражением лексической функции Copul, см. [7, с. 166]). Например, слово *обвинение* имеет категорию Содержание, поскольку возможно сочетание *мое обвинение состоит в том, что он* (*отступил*, *не поддержал* и т.п.); слово *инструкция* имеет категорию Содержание, поскольку можно сказать *инструкция состоит в том, что… .*

Сложнее с категорией Действие. Допустимость сочетания с глаголом, выражающим лексическую функцию Oper1, не всегда диагностирует имя как действие. Так, *высказал похвалу ≈ похвалил*:

Сергей Степашин высказал похвалу в наш адрес, заметив, что… .

Но отсюда не следует, что слово *похвала* имеет здесь категорию Действие; например, можно сказать *высказал соображение, что …*, но *соображение* не действие. Слово *похвала* в принципе имеет категорию Действие (*Вещуньина с похвал вскружилась голова*; *Похвала вскружила ей голову*), но не обязательно имеет ее в контексте *высказать*. Ср. также *отповедь*:

Валерий Отчерцов *дал отповедь* Рязанову, заявив, что «в отдельно взятой области невозможно создать газовый рынок. [«Известия», 2002.02.11]

Обычно пишут, что это стихотворение *есть отповедь* идеалиста-шеллингианца материализму. [«Наш современник», 2003.12.15]

Правда, попытка МЭРТ включить пункт об «излишней централизации» в программу правительства уже *вызвала отповедь* Белого дома: министерство обвинили в покушении на конституционные права председателя правительства. [«Газета», 2003]

Имена категории Действие имеют широкую сочетаемость. Например, они возможны в каузативных контекстах (*исповедь облегчила меня*), временных (*после допроса*); специально речевые действия диагностирует контекст глагола *слышать* (*вопрос, просьбу*). Принимаем, что категория у слов типа *выговор, нагоняй* в контекстах *сделал* и *получил* тоже Действие: *получил выговор* – как *получил удар* (так в МСТ).

Слово *утверждение* имеет категорию Действие в контексте (2а) и категорию Содержание в контекстах (2б,в)[2]:

(2) а. Кстати, я не спешу с *утверждением*, что сегодня это невозможно повторить. [И.Э. Кио. Иллюзии без иллюзий (1995–1999)]

б. в какой степени обосновано *утверждение*, что …;

в. ложью является *утверждение*, что … .

Между тем в предложениях из (3) категория слова *утверждение* не идентифицируется однозначно:

(3) а. Вы согласны с *утверждением*, что наш кинематограф потерпел крах? [Культура, 2002.04.01]

б. И у меня вызвало недоумение *утверждение*, что такой синдром «раньше никогда не встречался». [Медицина и здоровье (форум) (2005)]

[2] Здесь и далее примеры с годом – из Национального корпуса русского языка.

Две категории – Действие и Содержание – есть и у частных типов высказываний – у слов *просьба, совет, требование, приказ, жалоба*. Так, слово *просьба* в (4а) обозначает действие, а в (4б) – его содержание (то, о чем просят):

(4) а. Нас буквально *одолевали просьбами* друзья и знакомые, чтобы забронировать в кассе театра билеты или получить контрамарку. [И.А. Архипова. Музыка жизни (1996)]

б. <…> высказал *просьбу* к томским музейщикам – усилить роль Сталина. [Культура, 2002.04.08]

Два значения, соответствующие двум категориям слов типа *просьба*, иногда рассматриваются как третье, которое представляет собой дизъюнкцию двух. Я предпочту считать слова типа *просьба* неоднозначными. Дело осложняется, правда, тем, что слова типа *просьба*, в отличие от слов типа *обвинение*, могут выступать одновременно, т.е. в одном предложении, в обоих своих значениях. Так, в примере (4б) *просьба* – и действие, и содержание. Тем самым они противоречат сформулированному выше тезису о том, что в каждом данном контексте слово может выступать в нескольких тематических классах, но в одной из своих категорий. Установлено, однако, что для двух значений, которые метонимически связаны, такое совмещение значений не исключено, см. [9, с. 416; 14].

Содержание бывает не только у речевого действия, но и у текста (в том числе – предложения): содержание – это прежде всего смысл текста. Так, в *принес письмо* слово *письмо* имеет категорию Текст, а текст имеет содержание. Но слово *весть* в *принес весть* имеет категорию Содержание и не имеет категории Текст. Так что содержание не обязательно выражается текстом и тем более не обязательно выражается речевым действием. Содержанием высказывания может быть суждение, чувство, оценка, просьба и прочее.

Действие, Содержание, Текст – это три очевидных категории, к которым могут принадлежать имена речи. Принадлежность к категории Содержание диагностируется контекстом «…состоит в том, что/чтобы…». Иначе: тест на принадлежность слова X к категории Содержание – это возможность выступать в конструкции «X состоит в том, что…». Например, для слова *обвинение* это будет сочетание *обвинение состоит в том, что* (*убил*, *поддержал* и т.п.); для слова *просьба* – сочетание *просьба состоит в том, чтобы* (*усилить, укрепить* и т.п.). Другой критерий принадлежности слова X к категории Содержание – допустимость описания его значения в соответствии со словообразовательными моделями типа S2 или S3 (в смысле модели Смысл ⇔ Текст): *просьба* – это то, о чем просят.

Слова *пожелание, претензия, ответ, упрек, обвинение* имеют, как и имена типа *просьба*, помимо категории Действие, категорию Содержание, поскольку допустимы сочетания *мое пожелание* (*претензия, просьба, упрек*) *состоит в том, что…*:

обвинение – действие (*сыпал обвинениями*)/содержание = ‘то, в чем обвиняют’, т.е. S3 (*ему предъявили обвинение в убийстве*; *Сирия отвергает обвинения в поддержке международного терроризма*);

Эти слова – дериваты типа S2 или S3 от соответствующих глаголов второго участника в смысле модели Смысл ⇔ Текст). Их значение описывается соответствующими словообразовательными моделями. Категория Содержание – это шаг на пути к субстантивации глагольного значения.

Напротив, слова *одобрение, осуждение, протест, похвала, восхищение, порицание, согласие* имеют категорию Действие (*после одобрения решения кабинетом министров Совета Европы*), но не имеют категории Содержание – они невозможны в составе конструкции *состоит в том, что…* Еще вопрос, можно ли эти слова назвать речевыми в категории Действие.

Для слова *опровержение* Корпус не дает примеров сочетания *опровержение состоит*; пример из Яндекса (*опровержение состоит в сообщении о несоответствии действительности*) неадекватен. Так что категории Содержание у слова *опровержение* в классе речи нет. Оно имеет категорию Содержание в ментальном тематическом классе: *этот факт служит опровержением известного тезиса…* (*опровержение* = ‘то, что опровергает’, S1).

Возможность присоединять актант, выражающий содержание, у разных имен содержания разная. Слово *обвинение* способно присоединять актант Содержание: *Сирия отвергает обвинения в поддержке международного терроризма*; то же касается слова *требование* (*требование освободить*). А слова *оговор*, *оскорбление, отзыв* – нет (притом, что все они имеют категорию Содержание): **не верьте оговору, что...*

Содержание может быть не только у речевого действия, но и у текста (письменного и устного), в том числе – у предложения или слова. Содержание для текста – это не всегда то же, что смысл. Так, у слова *заповедь* различаются категории Текст и Осуществление:

(5) *заповедь* – текст ‘предписание’ (усвоил заповедь: гость – лицо священное. [И. Грекова. На испытаниях (1967)])/осуществление <того, что предписано> (заповедей Божьих не соблюдали –Real).

Слово может иметь категорию Содержание, не имея ни категории Текст, ни категории Действие. Так, слово *весть* в сочетании *принес весть* имеет категорию Содержание – в отличие от слова *письмо*, которое в сочетании *принес письмо* имеет категорию Текст.

Содержание может быть конкретизировано – как суждение, чувство, оценка, предписание, совет, просьба, вопрос и прочее. Но это находится за рамками нашего рассмотрения, поскольку задача состоит в том, чтобы выявить категории самого верхнего уровня.

Тест на категорию Текст – это возможность участия в конструкциях с глаголами *говориться, гласить*. Так, у слова *наказ*, помимо категории Действие (*И он дал наказ всем комендантам – смотреть востро…* [Ю.Н. Тынянов. Восковая персона (1930)]), есть категория Содержание, потому что можно сказать «в чем состоит наказ?», и категория Текст, поскольку возможно «в наказе говорится…».

Имена с категорией Текст сочетаются с глаголами материального плана, например, с глаголами перемещения и создания.

Текст – это носитель содержания и результат речевого действия. Обычно речевое действие создает текст (ср. *высказывание*). Но не всегда. Так, слово *зарок* имеет категорию Содержание (*зарок был – не видеться с ней никогда*), но не категорию Текст (**зарок гласит*). Содержание – это S2 от глагола, выражающего предписание (этот глагол для слова *зарок* не сохранился). Т.е. категория Содержание возникает тогда, когда исходная категория – Действие. Исходное значение – действие (поскольку имена отглагольные); результат – содержание, и – текст, если содержание мыслится как вещь.

Текст может быть не только письменный (вещь, например, письмо), но и устный (услышал приветствие). Есть отдельная категория: ‘официальный документ’ – гипоним категории Текст (пример – заявление).

Не все имена речи имеют категорию Содержание. Чистый пример – слово *приглашение*: у него есть категория Действие, но нет категории Содержание (**приглашение состояло в том, что…*) – только Действие и Текст.

Возьмем также слово *намек*. У него есть категория Действие (– Это был намек? <вопрос мне про мое речевое действие: «Оно было намеком?»; *быть* здесь – Опер1>), но нет категории Содержание. Нельзя спросить: «В чем состоит твой намек?» – надо: «На что ты намекаешь?». Значит, у него вторая категория – Суждение.

Категория Текст может быть исходным значением слова (как у слова *письмо* или *нота*; ср. *направил ноту*).

Исключительное свойство имен с категорией Текст (впрочем, не всех!) – это type-token distinction, противопоставление индивидного и экземплярного тождества.

Действия бывают чисто речевые (они выражаются перформативными глаголами и образованными от них именами, как обещание) и не чисто речевые, как угроза: глагол угрожать перлокутивный. Возможно, это различие отражается на структуре многозначности имени: имя, мотивированное чисто речевым глаголом, как просить, обычно совмещает категории Действие и Содержание, а у намек категории Содержание нет. Впрочем, *угроза* тоже совмещает категории!

Отдельным контекстом, выявляющим определенный аспект значения имени, можно было бы считать лексическую функцию Real (со значением ‘выполнить требование, содержащееся в имени ситуации’, см. [12, с. 97]. Можно сказать: *выполнил обещание*/*команду, наказ, условие, договор*; *сдержал клятву*, *удовлетворил претензию*/*просьбу*, *предсказание сбылось, соответствует условию,* но нельзя: **выполнил возражение*, *комплимент*, **удовлетворил намек*, **обвинение сбылось*. Категория Осуществление, соответствующая именно этому контексту, оказалась малочисленной; в примере (5) она приписана слову *заповедь* в контексте глагола *соблюдать*.

Категорию Способ действия имеют, например, имена *отзыв*, *изложение* (*Эта история в твоем изложении прозвучала слабо*).

Слова с категорией Сведение составляют подкласс содержаний: сведение – это передаваемое содержание. Категорию Сведение как единственную (т.е. не производную ни от Действие/Содержание, ни от Текст) имеют слова *весть, известие, новость, информация,* само *сведение* = передаваемое/получаемое содержание; *молва* = ‘слухи’ – сведение (дошла молва; молва уверяла – *устар.*).

Связь семантики имени с семантикой глагола принимается во внимание, даже если связь эта чисто семантическая, ср. *гипотеза* и *предполагать*.

Казалось бы, речевое действие всегда создает текст: текст – это результат действия. Однако текст не всегда стоит между действием и содержанием. Например, слово *просьба* имеет категории Действие и Содержание, но не имеет категории Текст.

Итак, содержание может быть у речевого действия и у текста. Категория Текст может соответствовать исходному значению слова (см. выше: *письмо, нота*) или производному. Так, *заявление* – это прежде всего речевое действие и только во-вторых – официальный текст, т.е. документ. Не исключено, что значение в категории Содержание и в категории Текст образованы у слова *заявление* от разных значений глагола *заявить*.

Категории Действие или Текст не в такой степени нуждаются в диагностиках (т.е. тестах на принадлежность), как Содержание.

Наша классификация имен отражает только роль объекта в речевой ситуации и не касается собственно сущности объекта. Между тем может быть важно, например, что слово *информация* принадлежит к именам массы, а слово *текст* – к именам индивидов.

3. КАТЕГОРИАЛЬНЫЕ КЛАССЫ ОТГЛАГОЛЬНЫХ ИМЕН РЕЧИ

Как видно из сказанного, в классе имен речи широко распространена категориальная неоднозначность – гораздо шире, чем, например, у имен эмоций. Возьмем то же слово *высказывание*. Оно имеет неоднозначность действие (произведенное в определенный момент, в определенном месте)/содержание. При этом провести границу между этими двумя употреблениями слова *высказывание*, равно как и слов, обозначающих разновидности высказывания, не всегда легко. Например, *реплика* – это одновременно речевое действие и его содержание:

Прошу не подавать *реплик* относительно моего внешнего вида.

В самом деле, *подать реплику* значит ‘сделать высказывание’, т.е. ‘осуществить действие’; в то же время группа с предлогом *относительно* характеризует содержание. Нельзя сказать: *его реплика состояла в том, что …*; однако можно сказать: *его реплика сводилась к тому, что …*, охарактеризовав таким образом ее содержание. Такое совмещение, впрочем, типично для метонимически связанных значений (см. выше).

Поскольку неоднозначность действие/содержание разрешается только контекстом, может возникнуть желание ввести обобщающую категорию Высказывание для имен, допускающих оба употребления – как действие и как содержание. Однако это не даст возможности описать различия в значении слов, которые по-разному совмещают в своем значении эти три категории. Так что все три категории рассматриваются как различные, а совмещение категорий в одном слове – как отдельный семантический факт.

Следуя традиции (см. [7]), мы трактуем возможный категориальный спектр слова как его неоднозначность. Хотя можно представить его как сочетаемостный потенциал, т.е. считать, что в разных контекстах срабатывают разные аспекты единого семантического комплекса. Но для такой постановки задачи не разработан концептуальный аппарат.

Ниже следуют перечни слов с категориальной неоднозначностью разных типов, а также перечни имен с однозначной категориальной принадлежностью.

Класс 1. Имена с неоднозначностью действие/содержание

В этот класс входят имена, мотивированные глаголами речевого действия, которым свойственна категориальная многозначность действие/содержание. Категория Содержание – это либо дериват S2 или S3 от глагола, мотивирующего данное имя [13, с. 97], как в (1), либо условие искренности речевого действия, как в (2):

(1) *обвинение* – действие (сыпал обвинениями)/содержание = ‘то, в чем обвиняют’, S3 (ему предъявили обвинение в убийстве; Сирия отвергает обвинения в поддержке международного терроризма);

(2) *похвала* – действие (рассыпался в похвалах)/содержание: одобрение (письмо выражало похвалу уму и сердцу дочери; сказал в похвалу).

Возможность присоединять актант, выражающий содержание, у разных имен содержания разная. Слово *обвинение* способно присоединять актант Содержание: *Сирия отвергает обвинения в поддержке международного терроризма*; а слово *оговор* – нет: **не верьте оговору, что...* Но *оговор* признается именем содержания, как и *обвинение*.

Имена класса 1 могут употребляться и в каких-то других категориях, в том числе – в категориях, обусловленных переходом имени в другой тематический класс. Например, у слова *ответ* есть, в числе прочих, категория Действие, вызванное чем-то:

(3) теракты, ставшие *ответом* на это решение.

Некоторые имена класса 1 могут употребляться также в категории Текст (например, *наказ, жалоба*). Здесь существенно направление семантической деривации: действие первично.

Ниже следует перечень имен с неоднозначностью типа действие/содержание. В фигурных скобках дано значение или категория, при которых имя уже не принадлежит к семантической сфере речи.

возражение – действие (*не терпит возражений, не слушал возражений*) / содержание <довод> (*мое возражение состоит в том, что…;* то, чем возражаю)

высказывание – действие (*сделал ряд высказываний; после этого высказывания*) / содержание (*Да и в последующем в этих сводках нередко приводились высказывания, сводившиеся, по существу, к одной мысли: «Фюрер прав, заключив договор с Москвой».* [А. Чаковский. Блокада (1968)])

вопрос – действие <требующее ответа> (*посыпались вопросы*) / содержание (*Он не ответил на вопрос, был ли там мальчик*) / {проблема}, требующая решения (*возник вопрос*) / {дело}, зависящее от чего-то (*вопрос времени*)

добавление – действие <по глаголу *добавлять*, лишь в частности речевое> (*Вынужден сделать добавление к фразе русского философа В.В. Розанова*) / содержание, т.е. то, что добавлено, S2, текст (*с разными добавлениями*)

жалоба – действие <выражение неудовольствия> / содержание (*моя жалоба состоит в том, что…*) / официальный текст (*направил жалобу*)

зарок – действие (*дал зарок*) <обещание не делать> / содержание речевого действия (*зарок был – не видеться с ней*)

заявление – действие <высказывание, решительное по тону> (*сделать заявление для прессы*) / содержание (*в заявлении говорится*) / официальный текст с просьбой (*подал заявление*); NB: содержание м.б. разное: распоряжение, просьба

клятва – действие <с ритуальным моментом: обещание> (*дал клятву отомстить*) / содержание, т.е. то, что обещано сделать, S2 (*сдержал клятву*)

команда – действие <приказание> / содержание, т.е. то, что приказано (*выполнил команду*) / {группа людей} (*команда разведчиков, пожарная команда*)

комплимент – действие <речевое; с лестным содержанием> (*сказать, сделать комплимент*; *Наворотил мерзостей и еще требует комплиментов, скотина!* [Ю. Трифонов. Дом на набережной (1976)]) / содержание: похвала (*это мне комплимент; Она улыбнулась; комплимент ей очень понравился.* [И.Ф. Анненский. Вторая книга отражений (1909)]; хотя сомнительно: *его комплимент состоял в том, что…*)

назидание – действие (*сказал в назидание; не нуждаюсь в назиданиях*) / содержание (состоит в том, что…)

намек – действие <речевое, со специфическим свойством: неявно выраженный смысл> (*– Это был намек?* <вопрос мне про мое действие>) / содержание (*тем самым он сделал намек, что может поддержать применение силы*)[3] / {слабое подобие} (*без всякого намека на талию*), метафорический перенос

наказ – действие (*дал наказ*) / содержание (*нарушил наказ*) / текст (*из больницы писал и передавал наказы*)

комментарий – действие <по глаголу *комментировать* = ‘разъяснять содержание интеллектуального объекта’> (*комментариям не было конца*) / содержание действия = результат: то, что разъясняет

[3] NB: метонимическое совмещение двух значений в одном употреблении слова: *сделал намек* – действие, *намек, что…* – содержание.

интеллектуальный объект (*его комментарий сводится к тому, что…*) / текст <содержащий разъяснение> (*рукопись была издана с подробными комментариями*); второй вариант: *комментарий* – текст <содержащий разъяснения к другому тексту> / действие <разъяснение> (*комментариям не было конца*) / содержание, т.е. то, чем разъясняют

напутствие – действие (*После напутствия родителей <…> невеста кланялась всем присутствующим*; однако *сделал напутствие) [Народное творчество, 2004, № 5] / содержание (*сказал в напутствие = в качестве напутствия; выполнил напутствие* = ‘то, чем напутствовали’)

наставление – действие (*после всех наставлений*) / содержание речевого действия *наставлять*: совет, рекомендация (*…читал когда-то наставленья* (Пушкин); *наставленья, куда ехать*)

обвинение – действие (*сыпал обвинениями*) / содержание = ‘то, в чем обвиняют’, S3 (*ему предъявили обвинение в убийстве; Сирия отвергает обвинения в поддержке международного терроризма*)

обет – действие с ритуальным моментом <обещание> (*дал обет*) / содержание, т.е. то, что обещано, S2 (*выполнил обет*)

обещание – действие (*несмотря на свое обещание*) / содержание, S2 (*выполнил обещание*)

объяснение – действие по глаголу *объяснять* (*начал объяснение, выслушивал объяснения*) / содержание = то, чем X объясняет Y, S3, / {ментальное значение} (*этому есть разумное объяснение; объяснение состоит в том, что…; понял объяснение* = ‘понял, в чем состоит объяснение’) / {деятельность} по глаголу *объясняться* ‘выяснять отношения’, не чисто речевая (*у матушки происходило объяснение с отцом*)

оговор – действие <ложное обвинение> (*воспрепятствовать оговору; не испугалась оговора*) / содержание (*не верьте оговору* = ‘тому, в чем обвиняют’)

оскорбление – действие (*он ответит за оскорбление*) / содержание (*в этих словах усматривали оскорбление*)

острота – действие (*ответил остротой, сыпал остротами*) / содержание (*это была меткая острота*) = ‘высказывание (остроумное)’; однако теста на категорию Содержание острота не проходит: **острота состояла в том, что...* Может быть, исходное значение здесь не действие, а текст.

ответ – действие (*дал ответ*) / содержание (*ответ состоял в том, что…*) / {действие-реакция}, отклик (*теракты, ставшие ответом на это решение*); NB: метафора.

отзыв – действие (*воздерживался от отзывов*) / содержание <суждение, содержащее оценку> (*его отзыв отрицательный*) / текст

опровержение – действие (*А он-то думал, что последует опровержение его словам.* [Ф. Искандер. Чик чтит обычаи (1967)]) / содержание ‘то, что опровергает’ / текст (*Газеты опубликовали опровержение собственной статьи!* [А. Тарасов. Миллионер (2004)]) / {ментальное значение} (*Он или знает его, или найдет опровержение за доской.* [64 – Шахматное обозрение, 2004.08.15])

повеление – действие (*было слышно, как он отдавал повеления; по его повелению был задержан*) / содержание = ‘то, что велено’ (*выполняла повеления мужа*)

подсказка – действие (*благодаря подсказке*) / содержание (*подсказка состоит в том, чтобы…*)

пожелание – действие (*высказал пожелание*, т.е. пожелал) / содержание (*мое пожелание состояло в том, чтобы…*); NB: в отличие от слова *желание*, которое не является действием, а следовательно и содержанием (речевого акта)

попрек – действие (*попреками в гроб вгонит*) / содержание: неодобрение (*слова звучали как попрек*)

похвала – действие (*рассыпался в похвалах*) / содержание: одобрение (*письмо содержало похвалу уму и сердцу дочери; сказал в похвалу*)

пояснение – действие / содержание

предложение – действие (*посыпались предложения*) / содержание = то, что предложено (*мое предложение – пойти на выставку*)

предсказание – действие (*сделал предсказание*) / содержание, т.е. то, что предсказано (*предсказание сбылось*)

претензия – действие: заявление недовольства положением дел с сознанием права требовать исправления (*предъявил претензию к качеству сборки*) / содержание: недостатки (*Исполнитель удовлетворяет претензию в течение 10-ти дней после ее получения.* [Договор о туристических услугах (2003)]) / {состояние: эмоциональное} (*он на тебя в претензии*)

приказ – действие (*отдает приказы*) / содержание (*не подчинились приказу* = ‘тому, что приказано’) / текст: документ (*в приказе говорится*)

приказание – действие (*кучеру отдано приказание*) / содержание (*выполняла приказание* = ‘то, что приказано’)

прорицание – действие (*он обладает даром прорицания*) / содержание, т.е. то, что предсказано (*прорицанье их [волхвов] осуществить*, МАС; *прорицания сбудутся*).

просьба – действие (*сразу последовала просьба*) / содержание, т.е. то, о чем просят (*просьба состояла в том, чтобы…*)

разъяснение – действие (*сделал разъяснение; ждут разъяснения*) / содержание (*разъяснение состоит в том, что…*)

распоряжение – действие (*сделал распоряжение; вплоть до особого распоряжения*) / содержание (*распоряжение состоит в том, что…*)

реплика – действие (*раздавались реплики*) / содержание (*его реплики были однообразными; реплики относительно их внешнего вида*)

рекомендация – действие (*после такой рекомендации*) / содержание ‘то, что рекомендуется’ (*В «Единой России» с интересом отнеслись к рекомендации Г. Павловского.* [Аргументы и факты, 2003.06.11]; *К рекомендации «не крепить груз веревками» стоит прислушаться, но выполнить ее часто не удается.* [Автопилот, 2002.04.15] / текст (*няня вынула рекомендации*)

совет – действие (*дал совет* = ‘посоветовал’) / содержание (*совет состоит в том, что…*) / мероприятие (*иду на совет*)

требование – действие (*посыпались требования*) / содержание, т.е. то, чего требуют (*требование прекратить*)

упоминание – действие (*упоминание о еде поколебало его решимость*) / содержание = то, что упоминается (*встречаются упоминания о темных его сторонах*)

упрек – действие (*сделал ей упрек, что…*) / содержание (*мой упрек состоит в том, что…*)

уточнение – действие (*последовало немедленное уточнение*) / содержание (*уточнение состоит в том, что…*)

К тому же классу можно отнести слова *признание, призыв, проклятье, пророчество, божба,* ср. *Она легко поверила божбе Теребилова, будто бы в Псков она вернется засветло* [А. Дмитриев. Дорога обратно // Знамя, №1, 2001].

В классе имен с неоднозначностью действие/содержание можно выделить два слова, способных выступать также в категории Сведение, т.е. передаваемое содержание:

сообщение – действие (*Нефть подорожала после сообщения об отказе Багдада сотрудничать с инспекторами ООН по вооружению.* [Финансовая Россия, 2002.09.19]) / сведение = передаваемое содержание (*весь мир облетело сообщение*)

указание – действие (*необходимо указание лица, заключающего договор*) / сведение = передаваемое содержание (*получил ценные указания*)

Категорию Сведение как единственную (т.е. непроизводную ни от Действие/Содержание, ни от Текст) имеют слова *весть, известие, информация*.

Класс 2. Имена с неоднозначностью действие/текст

Категория Текст задается сочетаемостью имени с глаголами более материального плана – перемещения, создания вещи. Чистый пример – слово *приглашение*: у него нет категории Содержание: **приглашение состояло в том, что…;* только Действие и Текст. Если категория Содержание дает себя знать как отдельный компонент смысла слова со значением действия (см., например, *жалоба*), то слово относится к классу 1. Затруднение вызывает слово *отзыв*, у которого категория Текст явно производна от Содержание. Оно причислено к действиям, хотя его употребления в этой категории маргинальны.

Имеется тонкое различие между содержанием (речевого действия) и смыслом текста. Казалось бы, содержание не может быть выражено иначе как текстом, т.е. быть смыслом текста. Однако текст – это прежде всего материальный предмет, и если действие порождает (или способно породить) материальный предмет, то у имени действия есть вторая категория – Текст, а категории Содержание может не быть.

апелляция – действие: обжалование / обращение за поддержкой (*Апелляция к религиозным ценностям может дать заметный эффект.* [Еженедельный журнал, 2003.04.08] / текст: документ (*подать апелляцию*)

апология – действие: защита, восхваление (*нам не нужна апология прозападной ориентации*) / текст (*…не прилагаю здесь ни своего послужного списка, <…> ни собственной апологии.* [А.С. Пушкин. Мои замечания об русском театре (1820)])

вызов – действие (*необходим вызов врача*) / текст: документ <заявление, требование> (*получил вызов*) / {протест}, готовность к борьбе (*смотрела с вызовом*) / {challenge}

выступление – действие (*его выступление было встречено овациями*) / текст (*в своем выступлении*)

добавление – действие <по глаголу *добавлять*, лишь в частности речевое> (*вынужден сделать добавление к фразе русского философа В.В. Розанова*) / то, что добавлено, S2, текст (*с разными добавлениями*)

доклад – действие (*готовиться к докладу; без доклада не входить*) / текст (*текст доклада занимает 25 страниц*)

донос – действие (*Филарет сделал донос на Павского, будто бы он лютеранин.* [А.С. Пушкин. Записные книжки (1815–1836)]) / текст (*…всех преподавателей заставлял писать на себя доносы.* [А. Азольский. Облдрамтеатр // Новый Мир, 1997, № 11])

завещание – действие <предсмертное распоряжение> / текст: документ, содержащий распоряжение об имуществе / содержание (метафора) (*прочесть архив Цветаевой как негласное завещание понять «душу»*)

извещение – действие <сообщение информации кому-то> (*необходимо извещение участников процесса*) / текст: документ, содержащий это сообщение[4] (*…накануне пришло извещение, что к экзаменам не допущен.* [П. Сиркес. Труба исхода (1990–1999)])

интервью – действие (*взял интервью у Иванова*) / текст (*опубликовано интервью*) / мероприятие (*иду на интервью*)

исповедь – действие (*исповедь облегчила меня; хотелось продолжать исповедь*) / текст (*что-то недоговоренное было в ее исповеди*) / мероприятие (*иду на исповедь*)

молитва – действие ритуальное (*чеченцы завершили молитву*) / текст (*прочел молитву*) / мероприятие (*Зовут к молитве муэдзины* (М.Ю. Лермонтов))

обращение – действие (*выступил с обращением к согражданам*) / текст (*в обращении говорится*)

объявление – действие (*объявлением о помолвке он нанес ей страшный удар; сделал объявление*) / текст <содержащий сведение> (*повесил объявление*)

описание – действие (*перейду к описанию места действия*) / текст <результат действия> (*1-я глава – это описание*)

[4] Категории Содержание слово *извещение* в современном языке не имеет; содержанием извещения может быть *сообщение* или *известие*.

оповещение – действие (*начал оповещение*) / текст <содержащий сведения> (*прислал оповещение*)

отчет – действие (*потребовали отчета*) / текст: документ (*подал отчет*)

пересказ – действие (*потребовали пересказа содержания*) / текст, письменный или устный (*написать пересказ*)

поздравление – действие (*…к 2 час. собрались наверху для поздравления дам.* [Николай II. Дневники 1894–1896 (1894–1896)]) / текст (*послал поздравление*)

поучение – действие (*литература не ставит себе целью поучение; Чапаев продолжал поучение*) / текст (*процитируем поучение старца Иосифа с горы Афон*); скорее, деятельность, так как нет глагола *поучить* в значении ‘прочесть поучение’ и часто во мн. числе (*начинаются поучения*)

приглашение – действие (*началось с приглашения в ресторан; на приглашение спеть ответила отказом*) / текст (*вот лежит приглашение*)

призыв – действие (*Судебные уставы предусматривали призыв присяжных как в палаты, так и в Сенат.* [Отечественные записки, 2003]) / текст (*зачитал призыв из анонимной листовки*); в контексте «это не есть призыв к слепой вере» можно было бы усмотреть категорию Содержание = ‘то, что призывает’; но скорее это действие.

разрешение – действие (*ждал разрешения на выезд, попросил, получил разрешение; от кого зависит разрешение*) / официальный текст: документ (*вынул из кармана разрешение*)

резюме – действие (*делаю резюме*) / текст: излагающий суть сказанного (*в своем резюме*)

уведомление – действие (*необходимо было уведомление начальства о…*) / текст (*получил уведомление о переводе*)

ходатайство – действие (*почему вы прекратили ходатайство о разводе?*) / текст: документ (*направил ходатайство*)

Класс 3. Имена с неоднозначностью / текст /содержание

Многозначность текст/содержание может показаться странной, поскольку у текста всегда есть содержание, ср. слово *тезис*. Категория Содержание возникает у имени только тогда, когда его исходная категория не Текст, а Действие. Класс немногочисленный.

заповедь – содержание: предписание (*заповеди Священного Корана не позволяют мусульманину творить зло и насилие*) = ‘то, что предписано’ (*заповедей не соблюдали*) / текст (*заповедь не сотвори себе кумира*) / осуществление ‘то, что предписано’ (*заповедей Божьих не соблюдали*)

запрос – текст: документ, содержащий требование дать сведения / осуществление (*удовлетворять запросы*)

заявка – текст: документ (*заявка на изобретение; подать, рассмотреть, отправить*) / содержание ‘требование’ (*Глебыч, на тебя есть заявка*)

инструкция – текст (*в инструкции сказано*) / содержание <речевого действия или текста; сведения о действиях> (*получить инструкцию*) <действие называется *инструктирование*>

Класс 4. Имена категории Действие

Действия бывают чисто речевые (они выражаются перформативными глаголами и образованными от них именами, как *обещание*) и не чисто речевые, как *угрожать*. Это отражается на структуре многозначности имени: имя, мотивированное чисто речевым глаголом, как *просить*, обычно совмещает категории Действие и Содержание. Имена данного класса обычно обозначают не чисто речевое действие – иначе была бы многозначность типа действие/содержание или действие/текст.

благодарность – действие: выражение словами положительной эмоции специального вида > (*выразил благодарность, получил благодарность*, как *получил удар* = стал адресатом действия «выражение благодарности»; *рассыпался в благодарностях; наговорил благодарностей*) / {чувство} (*с благодарностью*)

восклицание – действие <выражающее чувство, возможно, в словах> (*не мог удержаться от радостного восклицания*)

выговор – действие <выражающее словами отрицательную оценку> (*сделал выговор; получил выговор*: стал адресатом действия «выговор», как *получил удар*) / {способ произнесения звуков языка}

выступление – действие (*после выступления*)

извинение – действие <по глаголу *извиняться*: просьба о прощении> (*пробормотал извинения*) / S1 (*извинением мне служит*) NB: неречевое значение

излияние – процесс вытекания (*излияния базальта*) / действие <выражение чувств> (*я осторожно прервал его излияния*) NB: речевое значение производное

моление – действие <по глаголу *молиться*> / действие <по глаголу *молить*> (*А она только воссылает моления за своих врагов…* [М.А. Булгаков. Театральный роман (1936–1937)])

нагоняй – действие <выражающее отрицательную оценку> (*сделал нагоняй; получил нагоняй* – как *получил удар* = ‘стал адресатом действия’)

напоминание – действие (*только после напоминания*) / то, что напомнило, S1: ментальное значение (*это послужило напоминанием*); NB: возможность неагентивного употребления глагола *напоминать*[5]

насмешка – действие <не чисто речевое> (*осыпали насмешками, подвергался насмешкам*) / {слабое подобие} (*насмешка над здравым смыслом*), метафорический перенос

обжалование – действие <не чисто речевое: подать жалобу> (*обжалованию не подлежит*)

одобрение – действие (*вызвал, высказал одобрение; получила одобрение мастеров; после одобрения закона верхней палатой)* / {содержание: нере-

[5] Глагол *напоминать* является речевым лишь в производном значении, в отличие от *спрашивать* или *обещать*; и не удивительно, что его S1 не относится к классу имен речи.

чевое значение} (*его слова/жест выражал одобрение*);

оправдание – действие (*выслушивать оправдания*) / {ментальное значение} (*служит оправданием*)

осуждение – действие (*высказал/выразил осуждение; подверг осуждению*); нет категории Содержание (*осуждение состоит в том, что…)

отказ – действие <не чисто речевое> (*нас удивил его отказ; на просьбу ответил отказом*)

отклик – действие <ответ на зов> (*наконец донесся ленивый отклик*) / {ответная реакция} (*беды людей вызывали в нем сочувственный отклик*), метафора

отповедь – действие <ответ, содержащий выговор> (*ожидая отповеди за свою игривость*)

покаяние – действие (*удивил неуместным покаянием; ждать от них покаяния не приходится*) / деятельность (*остаток дней моих возмогу провести в покаянии*)

порицание – действие (*воздерживался от порицаний, не высказывал порицания*); NB: нет категории Содержание: **порицание состояло в том, что…*

предостережение – действие речевое / {действие неречевое}; это не действие, а каузатор: (*выговор будет предостережением для других*)

предупреждение – действие речевое / {действие неречевое} (*это будет тебе предупреждением*)

приветствие – действие, не обязательно речевое (*услышал сзади приветствие*)

признание – действие (*после такого признания; требовать признания*); NB: нет категории Содержание (*признание состояло в том, что…)

придирка – действие (*я ушла с репетиции после очередной придирки*)

проклятье – действие: крайнее осуждение (*предал проклятью; снял проклятье*)

угроза – действие: обещание причинить зло (*не поверил его угрозе*) / каузатор: ‘то, что может причинить зло’ (*лисица – это угроза охотничьему хозяйству*) / опасность причинения зла (*возникла угроза наводнения*), NB: деагентивация

утешение – действие <не чисто речевое> (*мне было противно утешение; не было опыта в утешении*) / {то, что (или тот, кто) утешает}, неречевое значение (*мне слезы утешенье*).

Класс 5. Имена категории Деятельность

Деятельность часто бывает не чисто речевая. Диагностические контексты: *во время, после*. Имена деятельности охотно употребляются также в категории Мероприятие. В отличие от действия (речевого), у Деятельности лишь в редких случаях есть дериват Текст. Возможно, неслучайно слова *клевета*, *дезинформация*, *злословие* (ср. *по этим поводам будет много злословия*) являются именами массы.

беседа – деятельность / {группа людей}, устар., см. МАС: *…распустил свою пьяную беседу* (С. Аксаков) <ср. тот же семантический переход в слове *команда*>

величание – деятельность (чествование песнями) / текст: песня

дебаты – деятельность / мероприятие (*участвовал в дебатах*)

дезинформация – деятельность: распространение ложной информации (*профессионалы дезинформации*) / сведение: ложное <имя массы> (*Эта дезинформация была запущена специально теми же спецслужбами.* [А. Яковлев. Омут памяти. Т. 2 (2001)])

диалог – деятельность (*минутный диалог его успокоил*) / текст (*диалоги Платона*)

диктовка – деятельность (*продолжалась диктовка; писать под диктовку*)

дискуссия – деятельность / мероприятие (*присутствовал на дискуссии*)

диспут – деятельность / мероприятие

допрос – деятельность / мероприятие (*иду на допрос*)

заикание – деятельность (*на мое заикание он картинно развел руки*) / {неспособность} (*заиканием страдают миллионы людей*)

злословие – деятельность (*Дабы не быть заподозренной в злословии, приведу некоторые примеры.* [Отечественные записки, 2003])

клевета – деятельность: распространение ложного обвинения (*клевета наказывается; осудили за клевету*) / ложная информация <имя массы> (*Клеветы хватает везде* [Е. Козырева. Дамская охота (2001)]); NB: *клевета* не содержание, поскольку невозможно «состоит в том» и нет актанта Содержание, как у *обвинение*);

коммуникация – деятельность: общение (*действия участников коммуникации*) / {линия связи} (*перерезать коммуникации противника*)

консультация – деятельность / мероприятие (*иду на консультацию*)

кривотолки – деятельность (*отмена <совещания> вызовет кривотолки*)

критика – деятельность: обсуждение (*ударились в критику*) / {сообщество людей}

моление – деятельность <по глаголу *молиться*> (*совершали моления; А она только воссылает моления за своих врагов…* [М.А. Булгаков. Театральный роман (1936–1937)])

отговоры – деятельность: многократные просьбы не делать (*несмотря на все отговоры*)

перекличка – деятельность / мероприятие (*участвовал в перекличке*)

похвальба – деятельность (*прервал эту похвальбу*)

прения – деятельность / мероприятие (*участвовал в прениях*)

прощание – деятельность (*во время прощания, перед прощанием*) / ?? (*сказать на прощанье*)

рассуждение – деятельность / текст

речь – деятельность (*я слышу речь не мальчика, но мужа*) / текст (*в своей речи подчеркнул*)

собеседование – деятельность / мероприятие (*иду на собеседование*)

увещевание – деятельность (*Ну, сейчас будет увещевание пастырем заблудшей в гордыне овцы…* (В. Гроссман); *после долгого увещевания; послан для увещевания;* формы СВ у глагола нет: *увещать* – НСВ)

хвастовство – деятельность (*мужчина склонен к хвастовству*)

Категорию Деятельность имеют также слова:

агитация, воздыхание, лепет, лесть, мольба, опрос, переговоры, пересуды, полемика, разговор, словоизвержение, спор, толки, читка, чтение.

Если в семантику имени входит оценка, то часто трудно отделить имя деятельности от имени содержания. Ниже имена в списке (а) скорее выражают оценку деятельности, в списке (б) – скорее оценку содержания или текста (при этом оценка может делать невозможной для имени содержания перефразировку с «состоять»):

(а) болтовня, водотолчение, воркотня, городьба, демагогия, злословие, краснобайство, кривотолки, нападки, описательство, празднословие, сквернословие (*казаки упражняются в сквернословии*), славословие, словоблудие, суесловие, тары-бары, трезвон, трепотня, турусы, шумиха (*шумиха не утихала*), треп, трескотня.

(б) ахинея, бодяга, болтология, бред, брехня, вода, враки, вранье, грубость, жвачка, колкость, многословие (*министр не терпел многословия*), небывальщина, небылица нелепица неправда, околесица, отсебятина, пустословие, резкость, россказни, сальность, скабрезность, сплетня, тарабарщина, телега, утка, чепуха (*сказал чепуху*), чушь, укол, экивок.

Класс 6. Имена категории Содержание

весть – содержание речевого действия или текста <а именно, сведение, т.е. сообщаемое содержание> (*принес весть*)

известие – содержание речевого действия или текста <= сведение> (*известие удивило; не имели известий; Синоптик Прохоров принес известие о том, что на берег идет ненастье.* [Ю. Рытхэу. Сегодня в моде пиликены (1970–1977)])

информация – содержание <речевого действия или текста> <= сведение> (*передать информацию*) (действие – информирование)

завет – содержание речевого действия <наставление> (*выполнять заветы*, но *давать заветы; ср. *оставил завет*)

пророчество – содержание ‘то, что предсказано’ (*сбылось пророчество*)

утверждение – содержание (*суть моего утверждения состоит в том, что…*) / {действие не речевое по глаголу *утвердить*} (*приговор послали государю на утверждение*); значение действие речевое отсутствует (*я начал <речь> с утверждения о том, что…*).

Класс 7. Имена категории Текст

донесение – текст (*написать, получить, прослушать донесение; в донесении говорилось*); *писать донесение* = ‘создавать текст’: *в донесении говорилось>*

каламбур – текст (не вещь); именно текст – его содержание неотделимо от формы и несущественно

любезность – текст <выражающий нечто любезное> (*говорить любезности*) / {свойство} (не речевое: любезного человека); NB: речевое значение производное

представление – {действие неречевое} / текст: документ (*представление к награде*) / {ментально-перцептивное значение}

продолжение – {действие–деятельность: не речевые} (*рекомендуется продолжение начатой работы*) / текст (*продолжение следует*)

Слово *нравоучение* (= наставление по поводу морали: *старался избегать нравоучений*) не способно выступать в категории Действие (нет глагола **нравоучать*); оно имеет только категорию Текст (*…не понимает басни, если в конце ее не находит нравоучения*. [М.Ю. Лермонтов. Герой нашего времени (1839–1841)]

Категорию Текст имеют также следующие слова (некоторые не являются отглагольными именами):

анонс, аншлаг, апелляция, апология, вердикт, воззвание, декларация, дифирамб, заклинание, запись, заповедь, история (*рассказал историю* – как *рассказал анекдот*), манифест, монолог, мораль, нотация, обжалование, обращение (*выступил с обращением к согражданам*), панегирик, пасквиль, повествование, показание, послание, постановление, предписание, рапорт, резолюция, рекомендация, реляция, репортаж, речь, тост, увещевание, удостоверение, уговор, ультиматум, филиппика, ходатайство, экскурс, экспертиза.

Примыкают к именам текста имена жанров и имена печатных изданий:

анекдот, анналы, байка, басня, повесть, прибаутка, присказка, присловье, притча, рассказ, роман, сказка, скороговорка, считалка; книга, брошюра, газета, журнал.

Остался ряд слов, которые трудно уложить в намеченные рамки, но их немного.

ЗАКЛЮЧЕНИЕ

Проведенное исследование показывает, что понятие онтологической категории является полезным инструментом при описании сочетаемости предикатных имен.

Другой вывод состоит в том, что онтологическая категория, как и тематический класс, являются важным орудием в борьбе с неоднозначностью. При этом онтологическая категория и тематический класс взаимодействуют друг с другом. Так, слово *опровержение* в категории Действие/Текст принадлежит к сфере речи (Статья была скандальной, даже Утесов написал *опровержение*), а в категории Содержание – и к сфере речи, и к ментальной сфере: опровержением может служить не только текст, но и факт (Этот факт служит *опровержением* известного тезиса, что движущей силой развития науки являются исключительно практические потребности. [В.А. Успенский. Что такое аксиоматический метод? (2002)])

Не полностью отражены противопоставления, связанные с возможностью речи восприниматься на слух. Так, слово *ропот* означает ‘недовольство, выражаемое негромко’; недовольство – это эмоциональное отношение; выражаемое – значит речь; не-

громко – значит звук. Поэтому *раздавался ропот* гораздо лучше, чем **раздался упрек*.

По словарям, *упрек* – это ‘выражение недовольства/неодобрения – обычно словами’ (как *упрекнул* = ‘выразил недовольство/неодобрение’); его онтологическая категория – Действие:

С повинной головой он пришел к ней, в ее квартирку на Басманной, попросил помощи, *выслушал мягкие упреки* [А. Азольский. Лопушок // Новый Мир, № 8, 1998]

А его тематический класс – имя речи: мягкими были выражения недовольства; выслушал он слова, которыми она выражала недовольство.

Но в сочетаниях типа *высказать* (*сделать, бросить*) слово *упрек* может интерпретироваться двояко. Можно считать, что *упрек* в контексте этих полувспомогательных глаголов сохраняет значение действия: *сделал упрек* = совершил <речевое> действие; т.е. значение ‘действие’ выражено дважды – в глаголе и в имени (как в *сделать неправильное действие*). И тогда *упрек* относится к тематическому классу имен речи. С другой стороны, можно считать, что в таких сочетаниях на долю слова *упрек* остается только ‘недовольство’: *сделал* (*бросил, высказал*) *упрек* = ‘высказал <свое> состояние: недоволен’, и тогда *упрек* имеет онтологическую категорию Содержание, а тематически является именем эмоции.

Тем самым взаимосвязь ментальной, речевой и эмоциональной сферы становится неоспоримым фактом. Механизм изменения тематического класса через сдвиг акцентного статуса компонентов в толковании нуждается в дальнейшем исследовании.

СПИСОК ЛИТЕРАТУРЫ

1. Падучева Е. В. О параметрах лексического значения глагола: онтологическая категория и тематический класс // Русский язык сегодня. Т. 3. Проблемы русской лексикографии / ред. Л. П. Крысин. – М.: Азбуковник, 2004. – С. 213–238.
2. Толковый словарь русских глаголов: идеографическое описание. Английские эквиваленты. Синонимы. Антонимы / под общ. ред. Л. Г. Бабенко. – М., 1999.
3. Vendler Z. Linguistics in Philosophy. – Ithaca, N. Y.: Cornell Univ. Press, 1967.
4. Маслов Ю. С. Вид и лексическое значение глагола в русском языке // Изв. АН СССР. Сер. лит. и яз. – 1948. – Т. 7, № 4. – С. 303-316.
5. Падучева Е. В. Лексическая аспектуальность и классификация предикатов по Маслову-Вендлеру // Вопросы языкознания. – 2009. – № 6. – С. 3-21. – URL: http://lexicograph.ruslang.ru/TextPdf2/Maslov_Vendler_final.pdf
6. Падучева Е. В. Семантические исследования: Семантика времени и вида в русском языке. Семантика нарратива. – М.: Языки русской культуры, 1996. – URL: http://lexicograph.ruslang.ru/TextPdf1/paducheva1985.pdf
7. Апресян Ю. Д. Лексическая семантика: Синонимические средства языка. – М.: Наука, 1974.
8. Ляшевская О. Н., Падучева Е.В. Онтологические категории имен эмоций // НТИ. Сер. 2. – 2011. – № 5. – С. 23-31. – URL: http://lexicograph.ruslang.ru/ TextPdf1/ emotions_teor_NTI_final3.pdf
9. Падучева Е. В. Динамические модели в семантике лексики. – М.: ЯСК, 2004.
10. Кустова Г. И. Типы производных значений и механизмы языкового расширения. – М.: ЯСК, 2004.
11. Розина Р. И. Семантическое развитие слова в русском литературном языке и в современном сленге. – М.: Азбуковник, 2005.
12. Муравенко Е. В. О случаях нетривиального соответствия семантических и синтаксических валентностей глагола // Семиотика и информатика. Вып. 36. – М.: Языки рус. культуры: Рус. словари, 1998. – С. 71–81.
13. Мельчук И. А. Опыт теории лингвистических моделей «Смысл ⇔Текст». – М.: Наука, 1974.
14. Перцова Н. Н. Формализация толкования слова. – М.: Изд-во МГУ, 1988.

Материал поступил в редакцию 28.11.12.

Сведения об авторе

ПАДУЧЕВА Елена Викторовна – доктор филологических наук, профессор, главный научный сотрудник ВИНИТИ РАН, Москва
E-mail: oli@viniti.ru

УДК 81’374.2’373.46

В.И. Хайруллин

Корпусный словарь фобиотерминологии

Рассматривается поле терминов страха и один из способов их представления в многоязычном словаре. Дается краткая классификация фобий и обосновывается необходимость лингвистического описания и систематизации терминологических единиц данной области. Представляется проект по выявлению и описанию корпуса фобиотерминологии и дается характеристика корпусного словаря, составленного авторским коллективом.

Ключевые слова: *корпусная лингвистика, словарь, термин, фобия, язык, единица*

В настоящей статье представлен результат большого исследовательского проекта авторского коллектива кафедры делового иностранного языка и перевода Башкирского государственного университета, а именно корпус терминов, описывающих различные проявления одной из базовых эмоций – страха – в английском, русском и башкирском языках.

Проблемы корпусной лингвистики уже были предметом рассмотрения в публикациях сборника «Научно-техническая информация». В частности, нам знакомы работы В.И. Подлесской, А.А. Кибрик, В.Л. Цукановой, Н.А. Коротаева, В.Г. Хуршудян, З.В. Ефимовой [1].

Тематика, выбранная нами, не случайна – фобии привлекали внимание величайших умов человечества – от Еврипида в V веке до н.э., Вильяма Шекспира, Френсиса Бэкона в XVI веке, до Зигмунда Фрейда в XX веке и многих других представителей самых различных профессиональных областей – Альфреда Хичкока, Стивена Кинга, Грэма Грина и даже Брижит Бардо. Одним из наиболее известных высказываний относительно интересующего нас концепта и термина, является следующее, принадлежащее Франклину Д. Рузвельту: «The only thing we have to fear is fear itself» («Единственное, чего мы должны страшиться, - это самого страха») [2, с. 127], в котором тавтология подчеркивает потенциальную самоцентричность данного понятия, т. е. мысль о том, что чувство страха может быть направлено на страх. Приводимая цитата в перефразированном виде выглядит следующим образом: «Бойтесь бояться», что очень оптимистично, поскольку термин отрицательной эмоции, дважды используемый в пределах одного высказывания, создает положительный эффект.

Вместе с тем, современный опыт показывает, что число как фобий, так и их наименований увеличивается в значительной прогрессии и это делает необходимым профессионально классифицировать и упорядочивать информацию, природа которой может быть весьма различна [3]. Специалисты различают фобии по трем основным категориям: социальные фобии, при которых человек испытывает страх перед возможными социальными или профессиональными контактами; панические расстройства, характеризующиеся периодическими приступами патологического страха без какой-либо видимой причины, это так называемые «беспричинные» страхи; специфические фобии, такие как непреодолимая боязнь змей, замкнутых пространств, высоты и проч. Последняя из перечисленных категорий представляется наименее сложной как для понимания ее природы, так и для способов ее преодоления.

Актуальность проблем фобий очевидна для многих областей знания – психологии, медицины, социологии и – что для нас особенно важно – лингвистики, поскольку вновь выявляемые фобии, равно как и уже известные, требуют лингвистической классификации и описания, т. е. разработки адекватных наименований, которые бы логично и точно представляли на языковом уровне характер страха.

Наша работа актуальна также в том плане, что она выполнена в русле корпусной лингвистики – одного из наиболее перспективных и активно разрабатываемых разделов современной науки о языке. По мнению авторов Программы фундаментальных исследований Президиума РАН «Корпусная лингвистика», корпус является не только необходимым инструментом исследования языка, но и необходимым компонентом его описания, поскольку к классической паре «словарь – грамматика» в настоящее время добавился третий элемент, а именно корпус. В связи с этим правомерно говорить о корпусных словарях, выполненных по отношению к конкретному фиксированному корпусу [4].

Таким конкретным фиксированным корпусом, по отношению к которому выполнена наша работа, может быть, во-первых, корпус русского языка и, во-вторых, корпус башкирского языка.

Разработанный нами корпусный словарь мы для удобства изложения называем «терминологическим корпусом», «корпусом фобиотерминологии», «корпусом терминов» или «корпусом».

Научная новизна работы состоит в том, что впервые в отечественном языкознании предпринимается попытка представления терминов страха в трех разносистемных языках, относящихся, кроме того, к различным языковым группам. Этими языками являются английский (аналитический язык, германская

группа), русский (синтетический язык, славянская группа) и башкирский (агглютинативный, тюркский язык). Это хорошо согласуется с положениями общей научно-практической задачи российской науки, состоящей в целенаправленном развитии корпусных методов описания языка с особым упором на описание русского языка и других языков народов Российской Федерации, что позволит осуществить создание интегрированной системы электронных корпусов русского языка и языков народов России [4], в данном случае – башкирского языка.

В настоящем проекте английский язык следует рассматривать, во-первых, как исходный, из которого заимствуется большая часть фобиотерминологии, и, во-вторых, как своего рода лингвистический фон, позволяющий наиболее ярко представить особенности разрабатываемого поля.

Следует подчеркнуть, что до настоящего времени исследований такого широкого плана не предпринималось и, соответственно, работы, представляющие результаты таких исследований, не публиковались. Более того, результаты поиска в Национальном корпусе русского языка показали, что подавляющее большинство терминов страха отсутствует в Национальном корпусе, включающем более 300 миллионов слов. Во всяком случае, по нашим запросам, как показала поисковая программа, «ничего не найдено» [5].

Теоретически значимо то, что настоящий корпус позволяет объединить и свести в свои рамки термины непреодолимого страха, относящиеся ко всем трем перечисленным выше категориям, а именно термины социальных, «беспричинных» и специфических фобий. Это важно, так как целый ряд фобий и, соответственно, терминов этих фобий связаны друг с другом, например, такая связь наблюдается между некоторыми социальными и специфическими страхами (иатрофобия и клаустрофобия, т. е. непреодолимая боязнь врачей и страх перед замкнутыми пространствами). Кроме того, некоторые из фобий и их терминов относятся одновременно к разным категориям, например, хрометофобия, страх перед деньгами, в равной мере имеет характеристику как социальной, так и специфической принадлежности.

Практическая ценность нашей работы видится в том, что она может быть широко использована на самых разных социально-профессиональных уровнях – от университетской студенческой аудитории при изучении курсов лексикологии различных языков (английского, русского, башкирского), курсов стилистики и теории перевода, при обучении будущих психологов, врачей и социологов, - до практикующих специалистов (переводчиков, врачей, психологов, социологов) при их работе с представителями разных языков и культур.

Материалом нашего исследования послужили профессиональные словари, лингвистические двуязычные словари, интернет-ресурсы.

При отборе языкового материала исходным языком был английский, причем практиковалась методика качественного подхода к лексическим единицам, состоящем в отборе системы терминов-понятий фобий, встречающихся в этом языке. Исходной точкой служила логическая систематизация явлений непреодолимого страха и выявление номинативных единиц в соответствии с этой систематизацией. Помимо терминологических корпусов, по этому принципу строятся также терминологические словари и словари-тезаурусы [6, с. 224].

Объем корпуса составляет 536 терминов, что далеко не исчерпывает всего поля фобиотерминологии, поскольку в данной области наблюдается ярко выраженная тенденция к образованию новых единиц.

Термины, составляющие настоящий корпус, подразделяются на три основные группы: узуальные, т. е. широко известные, в том числе имеющие прозрачную этимологию (например, аэрофобия - патологическая боязнь полетов на летательных аппаратах, гидрофобия - патологическая боязнь воды); узуально-профессиональные – это менее известные, в том числе узкопрофессиональные (например, иофобия - патологическая боязнь отравления, кенофобия - патологическая боязнь пустоты, свободных пространств); окказиональные, т. е. образованные для «единичного» использования, возможно, для использования в юмористическом контексте (например, магейрокофобия - патологическая боязнь приготовления пищи, параскаведекатриафобия - патологическая боязнь "пятницы тринадцатого").

С точки зрения языков, используемых в процессе терминообразования, корпус также неоднороден. Бóльшая часть терминов произведена от древнегреческих основ, например, *panophobia* – панофобия, патологическая боязнь всего окружающего (от греч. *pas*, *pan* – все); *anthrophobia*, *anthophobia* – антрофобия, патологическая боязнь цветов (от греч. *anthēra* – цветок). Вторую относительно немногочисленную группу составляют термины, имеющие латинскую основу, например, *altophobia* – альтофобия, непреодолимый страх перед высотой (от лат. *altus* – высокий). В третью группу входят единичные нерегулярные термины с англо-германской основой, например, *angrophobia* –ангрофобия, патологическая боязнь гнева (от англ. *anger* – гнев, от др.-сканд. *angr* – печаль); *snakephobia* – снэйкофобия, патологическая боязнь змей (от англ. *snake* – змея, от др.-англ. *snaca*, др.-ниж.-герм. *snake*, др.-сканд. *snákr*, *snókr* – змея) [7, с. 29, 36, 40, 795, 1083]. Последний способ терминообразования мы называем «эклектичным», при котором в морфологической структуре сочетаются основы современного (в данном случае – английского: *snake*) и классического (в данном случае – древнегреческого: *phobia*) языков. Сочетание латинской и древнегреческой основ не представляется «эклектичным» по той причине, что оба языка являются классическими.

Примечательно, что ряд терминов имеют синонимы в пределах корпуса. К их числу, в частности, относятся *snakephobia* (снэйкофобия) = *ophidiophobia* (офидиофобия) – боязнь змей; *nebulaphobia* (небулафобия) = *homichlophobia* (гомихлофобия) – боязнь тумана; *koniophobia* (кониофобия) = *amathophobia* (аматофобия) – боязнь пыли; *maieusiophobia* (майев-

сиофобия) = *lockiophobia* (локиофобия) = *parturiphobia* (партурифобия) - боязнь рождения ребенка.

Из приводимых примеров следует, что термины с англо-германской основой могут иметь в качестве синонимов регулярные термины, т. е. термины, образованные на древнегреческой основе. На наш взгляд, если в ряду синонимов есть единицы, образованные способом эклектики, и единицы, образованные стандартным способом (сочетание основ из классических языков), то предпочтительными являются термины, образованные вторым способом, так называемые классические термины, при образовании которых не было допущено смешения современных и классических языков. Однако «эклектичные» термины также должны быть представлены в корпусе, поскольку они имеют распространение в языке, были выявлены и зарегистрированы.

Анализ корпуса терминов фобий по алфавиту показывает, что не все буквы в одинаковой степени продуктивны в плане терминообразования. Наибольшее число терминов образовано с начальной латинской *A* (68 терминов), *P* (63 термина), *C* (51 термин). Наименьшее количество – с начальной *J* (2 термина), *U* (2 термина), *W* (2 термина). Буквы *Q* и *Y* характеризуются нулевой продуктивностью – на эти буквы не зарегистрировано ни одной терминологической единицы.

Представленная в настоящей работе характеристика корпусного словаря терминов непреодолимого страха позволит потенциальным пользователям оценить его практическую значимость, ценность и безусловную необходимость.

СПИСОК ЛИТЕРАТУРЫ

1. Корпус устной речи: дискурсивная транскрипция // НТИ. Сер. 2. – 2007. – № 2. – С. 2-52.
2. The Oxford Dictionary of Quotations. - Oxford: University Press, 2001. - 508 p.
3. Белоногов Г.Г., Гиляревский Р.С., Хорошилов А.А. О природе информации // НТИ. Сер. 2. - 2009. - № 1. - С. 1-7.
4. Корпусная лингвистика. Программа фундаментальных исследований Президиума РАН. – URL: www.corpling-ran.ru (дата обращения 10.12.2012)
5. Национальный корпус русского языка. – URL: www.ruscorpora.ru (дата обращения 10.12.2012)
6. Креленштейн Н.С. Словарь-минимум базового языка // Преподавание иностранных языков и его лингвистические основы. – М.: Наука, 1972. – С. 222–236.
7. The Concise Oxford Dictionary. – Oxford: Clarendon Press, 1978. – 1368 p.

Материал поступил в редакцию 21.01.13.

Сведения об авторе

ХАЙРУЛЛИН Владимир Ихсанович – доктор филологических наук, профессор, заведующий кафедрой делового иностранного языка и перевода Института права Башкирского государственного университета, г. Уфа
E-mail: vladimir-blt@mail.ru

УДК 81’322’373.46

К.А.Васильченко

Терминологические словосочетания и проблема их определения

Рассматриваются словосочетания, имеющие распространение в поле юридической терминологии. Дается анализ определений данных терминологических единиц. Материалом исследования являются французский, русский и английский языки.

Ключевые слова: *термин, синтагма, словосочетание, функции, референция*

Терминология имеет широкое распространение в современном мире. Социальные институты, совершенствуя правовые отношения, пополняют и развивают свои терминологические системы. Одной из областей, имеющих несомненный интерес для лингвистики, является юридическая терминология. На сегодняшний день весьма актуальной остается проблема разграничения терминов, их состава и структуры. Например, лингвистический энциклопедически словарь отмечает такие особенности термина, как: системность, наличие дефиниции (для большинства терминов), тенденция к моносемантичности в пределах своего терминологического поля, отсутствие экспрессии, стилистическая нейтральность [1, с. 508]. В языкознании определение терминов вписывается в обычные приемы, т. е. оно может исходить из следующих опций: определять сущность, определять лексические единицы (исходить из формы слова и ее системного включения) или определять процессуальное действие, что ведет к построению «теоретических» предметов.

В качестве терминов могут выступать как отдельные лексические единицы, так и сочетания слов.

Словосочетания – это единицы более высокого уровня, чем уровень слова. Лексические единицы в словосочетании функционируют как часть относительно сложного целого, в рамках которого более полно проявляются присущие отдельному слову свойства индивидуализации и конкретизации. Семантика словосочетания, которое наделено номинативной функцией, предопределяется как семантическими особенностями составляющих его имен, так и спецификой предлога, если таковой имеется. При этом нужно иметь в виду, что лексико-семантическая организация словосочетания отличается от лексико-семантической организации предложения [2, с. 11-13]. Характерно, что субстантивные словосочетания значительно расширяют номинативные потенции языка в целом.

Исследуемые нами синтаксические единицы во французской лингвистической традиции обозначаются как синтагмы, в понятие которой специалисты вкладывают различное содержание. Мы, в соответствие с задачами настоящей работы, следуем определению, которое под синтагмой понимает двусоставное языковое образование, обладающее линейным характером и выполняющее функции идентификации и дифференциации. В отечественной лингвистике терминологическое словосочетание рассматривается как сочетание «стержневого» и зависимого слова, что соответствует понятию детерминативной синтагмы. Если стержневое слово выполняет функцию отождествления, то зависимое служит для различения терминологических словосочетаний от других терминологических словосочетаний, например:

analyse criminologique – криминологический анализ
analyse juridique – юридический анализ
annexion partielle – частичная аннексия
annexion totale – полная аннексия

В приведенных именных терминологических словосочетаниях существительное используется для того, чтобы производить идентификацию, тогда как прилагательное уточняет, дифференцирует, характеризует действие. В речи словосочетания такого типа проявляют себя как целостные единицы, предназначенные для референтного употребления, они используются для указания на отдельный объект действительности.

Какие же свойства определенных дескрипций способствуют референтному употреблению? Изыскания таких известных лингвистов и классиков, как Г. Фреге, П. Стросон, Б. Рассел и др. привели к пониманию того факта, что в основе референциальных свойств определенных дескрипций находится понятие пресуппозиции. При этом предполагается, что экзистенциальные пресуппозиции связаны со всеми лексическими единицами. Прагматическая пресуппозиция как определенных, так и неопределенных дескрипций предполагает существование класса объектов. На этом этапе референциальные характеристики дескрипций идентичны и зависят от свойств существительного. Различие же в употреблении определенных и неопределенных дескрипций заключается в следующем: определенные дескрипции семантически предполагают существование и единственность объекта. Используя определенную дескрипцию, говорящий с помощью самого этого выражения заставляет слушающего разделить с ним уверенность в существовании и единственности объекта. По этому поводу специалисты отмечают, что в «естественных языках, где не существует абстрактных переменных, система пресуппозиций... избавляет говорящего от необходимости утверждать в каждом акте высказывания су-

ществование объектов референции» [3, с.125]. Референтное употребление является следствием указания существования и единичности, свойственного определенным дескрипциям.

Референция неоднородна. В связи с этим Ж. Клебер выделяет три типа референции: номинативную, присущую именам собственным; указательную, свойственную дейктикам; и описательную референцию, характерную для определенных дескрипций [4, с. 78]. Последние позволяют слушающему осуществить первичную, а затем и повторную идентификацию объекта. Именные словосочетания, являясь сжатой, экономичной грамматической структурой, находят широкое распространение в языке права. При употреблении в сопровождении определенного артикля, указательного или притяжательного детерминатива, словосочетания способны выступать в качестве определенных дескрипций.

Анализ материала французско-русского юридического словаря [5] позволяет отметить, что большую часть имен существительных составляют номинализации или существительные абстрактные, например: *base (f)* – основание, *cas (m)* - случай, *détournement (m)* – 1) похищение, присвоение (*вверенного имущества*); 2) злоупотребление (*властью, служебным положением*); 3) обход (*нормативных актов, установленных ограничений*).

Именно абстрактный характер семантики стержневого слова обусловливает то, что терминология обращается именно к словосочетаниям.

Рассмотрим примеры французских юридических терминов, в состав которых входит концепт *action (f)*: *action civile, action de concert, action confessoire, action estimatoire, etc*. Определяющее слово является указателем отнесенности к данной терминологической системе. Отметим, что в определении данных терминов присутствует единица *action*:

action confessoire. Action réelle qui tend à la reconnaissance d'un droit à une servitude.
action estimatoire. Action par laquelle l'acquéreur d'une chose demande une diminution du prix en raison de vices cachés.
action en justice. Pouvoir reconnu aux sujets de droit de s'adresser à la justice pour obtenir le respect de leurs droits ou de leurs intérets légitimes [6, с. 16].

Интерес представляет следующее толкование данного термина:

action (*f) en justice* 1. Право на предъявление иска; 2. Заявление (в суд), исковое заявление.
В строгом смысле это отнюдь не иск, а именно право на предъявление иска. Синонимом является *droit d'agir en justice*.
Второе (и, отметим, не совсем точное) значение – заявление в суд. Точный термин во французском языке – *demande* (*f) en justice* [7, с. 9].

В современном французском языке номинальная синтагма имеет тенденцию к распространению, стержневое имя существительное находится в позиции перед зависимым словом, т. е. преобладает прогрессивный порядок слов. В именном словосочетании русского языка мы наблюдаем постпозицию существительного, что хорошо прослеживается на примере следующих терминов:

contrat (*m) accessoire* – дополнительный договор
contrat (*m) bancaire* – банковский договор

При сопоставлении французского термина с его русским эквивалентом мы обнаруживаем, что глагольные словосочетания при переводе сохраняют свое значение, при этом в переводе используется прямой эквивалент, сравним:

глагол + имя существительное:

prendre des mesures	принять меры
munir les pouvoirs	предъявить полномочия
remplir les obligations	выполнять обязательства
conférer la responsabilité	возлагать ответственность
accepter les obligations	принимать на себя обязательства
établir le règlement	устанавливать правила,

глагол + имя существительное с предлогом:

entrer en vigueur	вступить в силу
agir en application	действовать на основании
prendre en considération	принять во внимание.

Помимо глагольных довольно частотны номинативные и предложные конструкции:

exécution des mesures	проведение мер (род. падеж)
sur recommandation	по рекомендации
en vertu de	на основании
sur l'initiative de	по инициативе.

Проведенный нами анализ позволяет сделать вывод о том, что во французском языке отношения между именами существительными в рамках терминологических словосочетаний выражаются при помощи предлогов, в русском же языке эти отношения выражены падежной формой. Это отчасти объясняется большей аналитичностью французского языка по сравнению с русским. Можно предположить, что первый занимает «промежуточное» положение между сильно выраженной аналитикой английского и синтетикой русского языка. Более того, некоторые из терминологических словосочетаний, имеющих распространение в английском и французском языках, являются как бы «зеркальным отражением» друг друга, например, *sur l'initiative de* (фр.) – *on the initiative of* (англ.), *prendre en consideration* (фр.) – *take into consideration* (англ.) и др.

К одному из достоинств терминологических дефиниций исследователи относят выраженную понятийную определенность [8, с. 27], что не всегда наблюдается в рассматриваемом нами поле. Причина видится в так называемой fuzziness [9, с. 135], т.е. характерной терминологической «размытости», нечеткости, что отмечается рядом специалистов. Мы не усматриваем в этом противоречия, поскольку диалектический характер термина позволяет в рамках одного понятия объединять, казалось бы, противоположные черты, с одной

стороны. С другой, как известно, недостатки – это продолжение достоинств: четкая понятийная определенность терминологической дефиниции сохраняется при определении термина в пределах одной области знания, например, при определении юридического термина «закон». Полагаем, что особый статус терминологических словосочетаний требует появления современных словарей. [10, с. 2]. Четкость данного термина «стирается», «погашается» при его использовании в других областях знания, скажем, в социологии (например, «законы развития общества») или физики (например, «законы Ньютона»). Однако данная проблема, представляющая несомненный научный интерес, может явиться предметом отдельного исследования.

СПИСОК ЛИТЕРАТУРЫ

1. Лингвистический энциклопедический словарь / гл. ред. В. Н. Ярцева, – М.: Советская энциклопедия, 1990. – 685 с.
2. Арутюнова Н.Д. Предложение и его смысл. – М.: Наука, 1976. – 383 с.
3. Николаева М.А. О логико-семантических работах современных французских лингвистов // Вопросы языкознания. – 1987. – № 6. – С. 121-133.
4. Kleiber G. Dénomination et relations dénominatives // Langages. – 1984. – №. 76. – P. 77–94.
5. Мачковский Г.И. Французско-русский юридический словарь. Ок.35 000 терминов. 4-е изд. – М.: РУССО, 2002. – 440 с.
6. Guillien R., Vincent J. Termes juridiques. Paris. 1995. – 583 p.
7. Шлепнев Д.Н. Французский язык для юристов: толковый словарь. – Ростов-на-Дону: Феникс, 2006. – 288 с.
8. Хайруллин В.И. Теории перевода: Восток и Запад // НТИ. Сер. 2. – 2008. – № 8. – С. 26-28.
9. Pokorn N.K. In Defence of Fuzziness // The Metalanguage of Translation. – Amsterdam/ Philadelphia: John Benjamins, 2009. – P. 135–144.
10. Белоногов Г. Г., Хорошилов А. А. Автоматизация составления англо-русских двуязычных фразеологических словарей по массивам двуязычных текстов (билингв) // НТИ. Сер. 2. – 2010. – № 5. – С. 1–8.

Материал поступил в редакцию 21.01.13.

Сведения об авторе

ВАСИЛЬЧЕНКО Кирилл Алексеевич – аспирант кафедры делового иностранного языка и перевода Института права Башкирского государственного университета, г. Уфа.
E-mail: kyril_korol@mail.ru

УВАЖАЕМЫЕ ЧИТАТЕЛИ!

ЦЕНТР НАУЧНО-ИНФОРМАЦИОННОГО ОБСЛУЖИВАНИЯ ВИНИТИ РАН

ПРЕДОСТАВЛЯЕТ КОПИИ ПЕРВОИСТОЧНИКОВ

ВИНИТИ РАН осуществляет обслуживание копиями первоисточников, хранящихся в фонде научно-технической литературы ВИНИТИ, в фондах других библиотек, а также в доступных ВИНИТИ электронных ресурсах.

Фонд научно-технической литературы ВИНИТИ включает более 2 млн изданий по точным, естественным и техническим наукам, в т.ч.:

– отечественные и иностранные периодические и продолжающиеся издания – с 1987 г.;

– отечественные книги – с 1987 г.;

– иностранные книги – с 1991 г.;

– рукописи, депонированные в ВИНИТИ, – с 1962 г.

Заказы на бумажные или электронные копии первоисточников принимает Центр научно-информационного обслуживания (ЦНИО) ВИНИТИ. ЦНИО ВИНИТИ обслуживает коллективных (организации и учреждения) и индивидуальных пользователей.

Формы обслуживания:

– абонементная (на основе договоров и предоплаты);

– разовые заказы (с предоплатой заказа по счету);

– индивидуальная форма обслуживания в читальном зале ЦНИО ВИНИТИ.

На сайте ВИНИТИ (http://www.viniti.ru) представлен полный Электронный каталог научно-технической литературы (http://catalog.viniti.ru), зарегистрированной в ВИНИТИ с 1994 г. Доступ для просмотра и поиска по Каталогу свободный. Постоянные абоненты ЦНИО ВИНИТИ, имеющие логин и пароль для работы с Каталогом, могут делать заказ копий непосредственно через Каталог.

Услуги по изготовлению копий первоисточников из фондов других библиотек предоставляются только постоянным абонентам. Место хранения первоисточников указывается в Электронном каталоге.

За подробной информацией обращаться по адресу:

125190, Россия, Москва, ул. Усиевича, 20, ВИНИТИ РАН. ЦНИО

Телефоны: 8 (499)155-42-43, 155-42-09, 152-54-59

Факс: 8 (499) 943-00-60

E-mail: cnio@viniti.ru; **URL:** http://www.viniti.ru