

ФОРМАЛИЗАЦИЯ В ЭЛЕКТРОННЫХ СЛОВАРЯХ

Ю. А. Ким

Одной из основных целей компьютерной лингвистики, выражаясь образно, является обучение компьютера естественному языку.

Поскольку физическая природа и принципы действия искусственного и человеческого интеллекта различны, языковые знания следует вводить в компьютер посредством специального объяснения, или формального описания.

В общем виде такое описание должно содержать факты, или единицы языка, равно как и правила их актуализации. Поскольку язык является сложной системой, описание проводится на нескольких уровнях, не упуская из виду такие его качества, как полнота и целостность. Компьютерная лексикография создает описания, применяемые в системах поиска информации и машинного перевода [1].

Такая практическая направленность требует иных принципов работы, чем при создании традиционных словарей. Безусловно, существует множество словарей, которые, по большому счету, просто представляют содержание бумажных изданий в электронной форме, но здесь о них не будет идти речь, потому как создаются они без глубокой формализации и, как правило, неприменимы в системах автоматической обработки языков.

Электронные словари, которые имеются в виду здесь, зачастую одноязычны и содержат не только лексические, но и грамматические значения, а также правила их актуализации. Подобный словарь снабжается указателями, позволяющими найти адекватное соответствие при комбинировании с любым другим аналогичным словарем.

По сравнению с традиционным словарем, меняется представление о лемме. Не только лексемы и омонимичные единицы и формы, но даже отдельные значения полисемичных слов составляют отдельные леммы. Таким образом, в этом описании многозначность замещается тотальной омонимией.

Описание леммы, уже не традиционная словарная статья, представляет собой набор меток, характеризующих как грамматическую, так и лексическую семантику леммы. Так, например, белорусское существительное *дуб* (*chêne*) (*m.*) можно описать следующим образом (используется французское программное обеспечение):

дуб <substantif> (имя) <nom commun> (нарицательное) <concret> (конкретное) <inanimé> (неодушевленное) <arbre> (дерево) <m.> (мужской) <nombrable> (исчислимое) <neutre> (нейтральное) <paradigme de flexion 1> (парадигма 1).

Итак, лемма характеризуется уникальной комбинацией меток, которые представляют ее сущность. Каждая метка (включая нулевую), следовательно, нужна для дифференциации леммы. Если в нашем примере опустить метку <arbre>, заменить <concret> на <substance> (вещество), <nombrable> на <non nombrable> (несчетное) и, наконец, <paradigme de flexion 1> на <paradigme de flexion 2>, то получится название материала, вещественное «дуб – древесина» (du bois de chêne). Если ввести <animé> (одушевленное) <réjoratif> (оскорбительное) <paradigme de flexion 3>, то соответствие во французском будет совсем иное, «imbécile».

Опираясь на совокупность меток, компьютер может выбрать нужную лексему. Парадигма – это описание грамматических трансформаций, присущих группе лексем, выполненное в текстовом или графическом виде. Текстовый вариант парадигмы слова дуб (дерево) выглядит примерно так:

ДУБ: Nom s/<E> + Gen s/a + Dat s/y + Acc s/<E> + Instr s/ам + Loc s/e + Nom p/ы + Gen p/oў + Dat p/ам + Acc p/ы + Instr p/амі + Loc p/ах.

<E> обозначает отсутствие действия. В зависимости от преобразований можно использовать другие операторы (удалить символ слева, <R> шаг вправо, <L> шаг влево и др.). Слева от «/» стоит конкретное грамматическое значение, справа – оператор и флексия. Этот маленький алгоритм позволяет по заявленному грамматическому значению и корню получить форму, и наоборот. Если такая схема охватывает определенное число единиц, ей присваивается имя (одной из лексем, например), в результате чего получается новая метка. Так, например, слово дом (maison) (f.), которое склоняется аналогично, получит метку: <flexion ДУБ>.

При переводе с белорусского на французский, компьютер сперва анализирует форму, а затем метки белорусской леммы, чтобы выйти на аналогичное сочетание меток, а затем – на эквивалент, во французском. А чтобы определить значение каждой метки, необходима опора на контекст [2]. *Сцены аздоблены дубам. Les murs sont ornés de chêne.* В частной грамматике леммы (имени древесины) должны быть указаны все возможные сочетания, в частности, с глаголом «orne» (украшать), которого не будет при описании дуба как дерева, растения.

Зато частная грамматика последнего включает двор (la cour): *На двары дуб. Dans la cour il y a un chêne.* Конечно, можно было предположить, что на дворе лежит вещество, древесина, но в таком случае, скорее всего, была бы лексема *скласці* (déposer, stocker). *На двары складзены дуб. Dans la cour on a stocké du bois de chêne.*

Исходя из парадигмы склонения, компьютер может, например, отличить имя вещества (древесина) от имени дерева (дуб). К примеру, предложение: *Тут няма дубу.* Компьютер распознает корень дуб и флек-

сию у, что может указывать на дательный падеж (le cas datif) – исчисляемое и вещественное или же родительный падеж – тогда это только имя вещественное. Далее, обнаруживается отрицательная частица *няма* и делается вывод, что это родительный падеж, а значит, имя вещественное, при этом перевод: такой *Ici il n'y a pas de bois de chêne*.

Некоторые другие подробности должны вводиться в формальное описание, например, интонемы. *Каравул* обычно обозначает *la garde*, но, с восклицательным знаком превращается в «*au secours !*» Узнавание формы требует анализ ее непосредственного окружения в высказывании, а зачастую и текста. Выполняется это, как и грамматические преобразования, конечным автоматом, который визуально можно представить в форме графика [1]. Этот автомат «примеряет» к форме метки в заданном порядке, а затем делает выводы и переходит в следующее состояние, вплоть до результата, пользуясь такими операциями, как, например, «если да, то А, если нет, то Б».

Например, в предложении *Ишоў дождж*, автомат распознает, что *дождж* в единственном числе, но, чтобы узнать падеж, нужно узнать роль слова в предложении. Имя идет после глагола, что не типично, но это единственный глагол и единственное имя, причем форма прошедшего времени согласуется в роде с существительным *дождж*, следовательно, это обратный порядок слов в предложении, когда субъект на втором месте, а предикат на первом.

Значит, *дождж* стоит в единственном числе, в именительном падеже. Также можно буквально перевести *ишоў*, как *allait*, но в частной грамматике глагола *aller*, как и существительного *pluie*, указано, что вместе они приобретают значение, которое не сводится к простой сумме их значений. Таким образом, нужно актуализировать эквивалентный предикат *pleuvoir+imparfait*. Когда начинаем спрягать *pleuvoir*, местоимение *il* вводится автоматически, а для обозначения imparfait – флексия (-ait): *Il pleuvait*.

Мы привели примеры некоторых задач, проблем, с которыми сталкиваются создатели электронных словарей для систем поиска информации или машинного перевода. Примеры взяты из практики использования многофункциональной среды для обработки естественного языка NooJ [3], общие принципы – из работ лаборатории fLexSem Барселонского автономного университета.

Литература

1. BLANCO X. Repères pour l'enseignement des langues sur objectifs spécifiques à partir de la lexicographie informatique // Études de linguistique appliquée 135, Paris 2004
2. BLANCO X. Linguistic Formalisation of Romance Languages in the Frame of the Inter-Gram Project // International Fifth Conference FASSBL-5, Bulgarian Academy of Sciences, Sofia 18-20 October 2006.

3. *SILBERZTEIN M.* NooJ: A cooperative, object-oriented architecture for NLP. INTEX pour la linguistique et le traitement automatique des langues // Cahiers de la MSH Ledoux. Besançon, 2004

КОНЦЕПТУАЛЬНАЯ ПОЛЕМИКА В.В.НАБОКОВА С Н.Г.ЧЕРНЫШЕВСКИМ В РОМАНЕ «ДАР»

А. Ю. Корзун

Роман «Дар» был закончен В.В.Набоковым в 1937г. и стал во многих отношениях центральным в его творчестве. В предисловии автор указывал: «[Эта книга] <...> была и останется последним романом, написанным мной по-русски. Ее героиня <...> русская литература» [1, с.3]. Главный персонаж романа – молодой писатель из среды русской эмиграции Федор Годунов-Чердынцев, книга которого – «Жизнь Чернышевского» – дается в четвертой главе «Дара».

Н.Г.Чернышевский в 1853г. написал магистерскую диссертацию «Эстетические отношения искусства к действительности» [4], в которой выдвинул ряд тезисов о роли и предназначении искусства в жизни человека и общества. Краткий сравнительный анализ эстетических установок В.В.Набокова и Н.Г.Чернышевского позволяет заметить коренное расхождение по ряду моментов. Чернышевский отказывал в праве на жизнь «искусству ради искусства» и фактически отрицал примат искусства над материальной жизнью. Набоков позиционировал себя как чистого художника, далекого от всего, кроме искусства. Чернышевский же утверждал: «Наука и искусство (поэзия) – «Handbuch» для начинающего изучать жизнь; их значение – приготовить к чтению источников и потом, от времени до времени, служить для справок» [4, с.159]. Набоков в эссе назвал подобные установки возвращением к «забытым образцам, мистериям и басням, вызывающим <...> зевоту у простого народа, но зато с должной силой восхваляющим добродетель и бичующим порок» [2, с.406]. Труды Чернышевского во многом определили понимание и отношение к искусству в Советской России. Набоков отзывался о культуре этой страны следующим образом: «В современной России – стране моральных уродов, улыбающихся рабов и тупоголовых громил – перестали замечать пошлость, поскольку в Советской России развилась своя, особая разновидность пошляка, сочетающего деспотизм с поддельной культурой» [2, с.388]. Тот факт, что именно Чернышевскому Набоков «отдал» самую большую главу своего последнего русского романа, закономерно ставит вопрос о наличии в «Даре» концептуальной полемики художника с философом.

Взгляд на «Дар» «сверху» позволяет рассмотреть роман как совокупность циркульно обозначенных смыслов, которые, если их растащить из-под одной обложки, могут показаться вполне