

ПРОБЛЕМА СЕМАНТИЧЕСКОГО ОТОЖДЕСТВЛЕНИЯ СИНХРОННЫХ И ДИАХРОННЫХ ВАРИАНТОВ СЛОВА

А. В. Кашкина

Воронежский государственный университет

Поступила в редакцию 23 марта 2012 г.

Аннотация: в статье рассматривается одна из проблем, возникающих при применении методики маркемного анализа – одновременное наличие в изучаемых текстах русских поэтов орфографических вариантов одного слова. Также исследуется диахронический аспект данной проблемы – отождествление устаревших лексем с их современными аналогами.

Ключевые слова: маркемный анализ, маркема, орфографический вариант, отождествление.

Abstract: the paper deals with one of the problems occurring in the process of markemic analysis, the simultaneous existence of several orthographical variants of a word in Russian poetic texts. In addition to this, the paper analyses the diachronic aspect of the problem, the identification of archaic lexemes with their modern equivalents.

Key words: markemic analysis, markeme, orthographical variant, identification.

Данная статья посвящена одному из вопросов, возникающих в процессе исследования русской поэзии XVIII – начала XXI вв. с помощью маркемного анализа. Маркемный анализ – разработанный профессором А. А. Кретовым метод исследования крупных текстовых массивов, основанный на понятиях маркемы и индекса тематической маркированности (ИнТеМ) – величины, зависящей от частоты встречаемости слова в тексте и его длины в звуках. Математически этот параметр представляет собой разность между частотным весом словоформы (Ч-вес) и ее функциональным весом (Ф-вес). Маркемами являются существительные с положительным индексом тематической маркированности, удовлетворяющие ряду грамматических и семантических критериев [1].

Пятьдесят первых по величине ИнТеМ маркем какого-либо автора составляют его маркемный список. Затем списки маркем поэтов, принадлежащих одному временному промежутку, сопоставляются для выявления общепериодической маркемной лексики. Нами рассматриваются следующие хронологические срезы: 1) XVIII в.; 2) первая треть XIX в.; 3) вторая треть XIX в.; 4) последняя треть XIX в.; 5) начало XX в. («Серебряный век»); 6) XX в. до 60-х гг.; 7) 60–80-е гг. XX в.; 8) конец XX – начало XXI вв. В каждом из указанных периодов анализируются произведения 15 авторов.

Поскольку обработка текстового материала ведется с помощью компьютерных программ и прежде всего «ТемАЛ» – программы тематического анализа лексики [2], выделение отдельных словоформ осу-

ществляется на основании их написания, а не значения. Таким образом, в маркемных списках поэтов, относящихся к какому-либо из изучаемых нами хронологических срезов, могут сосуществовать элементы, фактически являющиеся орфографическими вариантами [3, с. 208–210] одного слова (например, *молчание* и *молчанье*). Кроме того, возможна ситуация, когда в более раннем периоде слово представлено устаревшей формой, которая впоследствии вытесняется более современной. Так, в XVIII в. в русских поэтических текстах встречается маркема *спокойство*, полностью замещаемая эквивалентной формой *спокойствие* на более поздних этапах развития языка поэзии. Очевидно, что целесообразным решением с точки зрения удобства сопоставления как отдельных авторов, так и различных хронологических срезов будет сведение синхронных и диахронных вариантов слова к одной маркеме. Поиск единого подхода ко всем подобным случаям – цель данной работы.

Прежде всего необходимо выработать общую технику «сложения» эквивалентных форм слова. Для упрощения процесса отождествления словоформ в таблице, отображающей маркемный список каждого из анализируемых поэтов, вводится дополнительный (эталонный) столбец, содержащий форму, являющуюся нормативной для современного состояния языка. Когда возможны оба варианта, из них выбирается более употребительный в настоящее время (проверка может быть осуществлена, в частности, при помощи портала ГРАМОТА.РУ). Естественно, что в случае присутствия в маркемном списке какого-либо поэта двух слов с совпадающей

эталонной формой, необходимо их объединение. Табл. 1 – фрагмент стандартной таблицы представления маркемного состава поэтических текстов Державина до обработки.

Показатели частот маркем с одинаковой нормой складываются, и для полученного значения (в данном примере – 24) определяется частотный вес (в этом случае можно просто использовать сведения о частотном весе любого слова данного автора, имеющего аналогичную частоту, поскольку лексемы с одинаковой частотой имеют и равный Ч-вес). Функциональный вес полученной «суммы вариантов» может считаться равным таковому для формы, совпадающей с нормой. В результате получим (табл. 2).

Возможен и другой подход, при котором за длину «словосуммы» в звуках принимается средний показатель, получаемый по формуле

$$D_{\text{сум.}} = (D_1 \cdot Ч_1 + D_2 \cdot Ч_2) / (Ч_1 + Ч_2),$$

где $Ч_1, Ч_2$ – частоты объединяемых словоформ, D_1, D_2 – их длины.

Таким образом, для нашего примера получим $D_{\text{сум.}} = 7,291(6) \approx 7$, что соответствует длине эталонной формы. В то же время в некоторых других случаях эти числа (вычисленная по приведенной формуле длина суммы и длина эталонной формы) могут различаться, что в конечном итоге может привести к существованию в маркемных списках различных авторов одинаковых маркем с несовпадающими показателями длины. Поэтому, на наш взгляд, приравнение длины и функционального веса суммы маркем к соответствующим параметрам для эталонной формы является более рациональным при условии незначительного расхождения в количестве звуков (на один звук). В свою очередь, второй из предложенных нами методов оказывается уместным для отождествления вариантов слова, различающихся по длине более чем на один звук (например, *отважность – отвага*).

Мы рассмотрели процесс отождествления словоформ на основе маркемного списка одного автора, т.е. в синхронии. Здесь имеют место главным образом вариации -ье/-ие (*мгновенье – мгновение, молчанье – молчание* и др.), где за эталон принимается написание с -ие (кроме *счастье*). Такие расхождения формы не представляют особой сложности при сведении их к одному элементу маркемного списка, за исключением случая *воскресенье – воскресенье*, где в настоящее время наблюдается довольно четкое разграничение по семантике ('день недели' – 'возрождение'), однако еще в начале XX в. каждая из форм могла принимать оба значения, поэтому в данной ситуации необходимо обращение к контексту.

Другая довольно часто встречающаяся вариация – употребление авторами, принадлежащими к одному хронологическому срезу, как полногласных, так и неполногласных форм без семантических различий (*молодость – младость, голос – глас* и др.) – отмечается преимущественно на протяжении XVIII – XIX вв.

Обратимся теперь к диахроническому аспекту проблемы, актуальному для анализа маркемных списков поэтов XVIII в. В процессе сопоставления современных и устаревших словоформ могут возникать неоднозначные ситуации. Так, *пламень* легко отождествляется с *пламя, бедство – с бедствие*, тогда как *приятность* и *приятство* по семантике не полностью совпадают (у слова *приятность* нет значения 'расположение, приязнь').

Причина сложности в отождествлении интуитивно кажущихся нам сходными словоформ заключается в изменении их значения с течением времени. Например, слово *дружество*, во времена Пушкина эквивалентное *дружбе*, в XVIII в. имело три основных значения: «1. дружба, дружеские отношения. ... 2. дружеское общество, друзья. 3. союз, объединение людей (обычно по социальной, религиозной принадлежности)» [4]. Следовательно, в этой и подобных

Т а б л и ц а 1

Фрагмент списка маркем Г. Р. Державина до процедуры отождествления

Слово в текстах	Норма	Частота	Длина	Ч-вес	Ф-вес	ИнТеМ
Счастье	Счастье	7	8	0,946 422	0,779 919	0,166 503
Счастье	Счастье	17	7	0,983 344	0,838 322	0,145 022

Т а б л и ц а 2

Фрагмент списка маркем Г. Р. Державина после процедуры отождествления

Слово в текстах	Норма	Частота	Длина	Ч-вес	Ф-вес	ИнТеМ
Счастье	Счастье	24	7	0,988 503	0,838 322	0,150 181

ситуациях до проведения отождествления необходимо.

• Уточнить семантику словоформы по современному для нее словарю. В развитии лексики русского языка можно выделить три основных периода:

1) XVIII в. (для ознакомления с семантикой лексем авторов этого этапа развития русской поэзии можно обратиться к Словарю русского языка XVIII в. [4]);

2) XIX в. (в нашей работе разделяется на три хронологических среза, но в свете изучаемой проблемы может рассматриваться как единое целое – по Словарю Даля [5]);

3) Современное состояние (весь XX в. – по ССРЛЯ [6]).

• Проверить значение слова в контексте (т.е. в произведениях исследуемого автора).

Указанный подход дает возможность избежать ошибок в соотношении форм слов, принадлежащих разным хронологическим срезам.

В то же время в текстах поэтов XVIII в. встречается ряд маркем, для которых, по нашему мнению, описанная выше процедура отождествления и отыскания современного эквивалента не является целесообразной. Прежде всего это многозначные лексемы, которым в современном языке соответствует два или более слов. Например, для маркемы В. К. Тредиаковского *благочестие* эквивалентами являются *благоговение* и *набожность*. Отождествление маркемы такого рода лишь с одной из соответствующих ей современных лексем может привести к чрезмерно обобщенной интерпретации языка поэзии.

К словам, не требующим отождествления с более поздними эквивалентами, на наш взгляд, также могут быть отнесены редкие маркемы, в первую очередь индивидуальные (т.е. встречающиеся в некотором периоде только в произведениях одного автора). В качестве примера можно привести маркему *благосердность*, в XVIII в. обнаруживаемую лишь в поэтических текстах В. К. Тредиаковского. Необходимо отметить, что ее эквивалент *добросердечность* на более поздних этапах развития русской поэзии (например, в начале XX в.) в список маркем не попадает, т.е.

отсутствует сам элемент, с которым могло бы быть возможно отождествление.

Таким образом, для применения к какой-либо маркеме XVIII в. процедуры отождествления в диахронии необходимо, чтобы анализируемая лексема не являлась узкоспецифичной, свойственной одному конкретному представителю рассматриваемого периода и имела единственный эквивалент, являющийся маркемой в произведениях авторов, относящихся к более поздним этапам развития русской поэзии.

Предлагаемая нами в работе методика приведения орфографических и хронологических вариантов слова к эталонной форме, во-первых, в значительной степени унифицирует маркемные списки авторов разных периодов, что существенно упрощает их сопоставление и, во-вторых, сокращает число эквивалентных элементов маркемного списка, что позволяет получить более полную картину языка русской поэзии с точки зрения маркемологии.

ЛИТЕРАТУРА

1. *Кашкина А. В.* О классификации маркем / А. В. Кашкина // Вестн. Воронеж. гос. ун-та. Серия : Лингвистика и межкультурная коммуникация. – Воронеж, 2011. – № 1. – С. 97–100.

2. *Кретов А. А.* Программа выделения тематически маркированной лексики / А. А. Кретов, И. Е. Воронина, И. В. Попова. Зарегистрировано в Государственном информационном фонде неопубликованных документов ФГНУ «Центр информационных технологий и систем органов исполнительной власти» (№ 50201000004 от 11.01.2010).

3. *Ахманова О. С.* Очерки по общей и русской лексикологии / О. С. Ахманова. – М., 2004. – С. 208–210.

4. Словарь русского языка XVIII века / АН СССР. Ин-т рус. яз. ; гл. ред. Ю. С. Сорокин. – Л. : Наука, Лен. отд., 1984–1991. – Вып. 1–6 ; СПб. : Наука, 1992. – Вып. 7.

5. *Даль В. И.* Толковый словарь живого великорусского языка : Т. 1–4 / В. И. Даль. – М. : Рус. яз., 1989–1991.

6. Словарь современного русского литературного языка : в 17 т. / под ред. В. И. Чернышёва. – М. ; Л. : Изд-во АН СССР, 1948–1965.

Воронежский государственный университет

Кашкина А. В., преподаватель кафедры теории перевода и межкультурной коммуникации
E-mail: shtil72@yandex.ru

Voronezh State University

Kashkina A. V., Lecturer of the Translatology and Cross-Cultural Communication Department
E-mail: shtil72@yandex.ru