

**К ВОПРОСУ ОБ ИССЛЕДОВАНИИ
УСТОЙЧИВЫХ СЛОВЕСНЫХ КОМПЛЕКСОВ
С ИСПОЛЬЗОВАНИЕМ ЭЛЕКТРОННЫХ ЛИНГВИСТИЧЕСКИХ РЕСУРСОВ**

И.Г. Гомонова
(Гомель, Беларусь)

В статье рассматриваются компьютерные лингвистические ресурсы (электронные словари и корпуса текстов) как инструмент для исследования устойчивых словесных комплексов

Ключевые слова: устойчивый словесный комплекс, фразеологизм, электронный словарь, корпус текстов

Computer linguistic resources (electronic dictionaries and corpora of texts) as a tool for the study of stable verbal complexes are discussed in this article.

Key words: stable verbal complex, idiom, electronic dictionary, corpus of texts

Большинство современных лингвистических исследований осуществляется с применением компьютерных технологий, которые помогают усовершенствовать и ускорить процесс получения, обработки и передачи информации. Информационные технологии в лингвистике – «это совокупность законов, методов и средств получения, хранения, передачи, распространения, преобразования информации о языке и законах его функционирования с помощью компьютеров» [1, 8].

В исследовании устойчивых словесных комплексов (УСК) использование информационных технологий становится важным и неотъемлемым компонентом, оптимизирующим весь процесс. Основными типами ресурсов, которые привлекаются при изучении вышеназванных языковых единиц, являются электронные словари и корпуса текстов.

Электронный словарь – это словарь в компьютере или другом электронном устройстве; с технической точки зрения это компьютерная база данных, которая содержит особым образом закодированные словарные статьи, позволяющие осуществлять быстрый поиск. Электронные словари обладают рядом существенных преимуществ по сравнению с традиционными словарями: больший объем; автоматический поиск; высокая скорость поиска; возможность поиска единицы в любой форме; одновременный поиск не только по названию словарной статьи, но и по всему огромному объему словарей и др. «Важно отметить, что, несмотря на электронную форму словарей, пользователю доступна вся информация, имеющаяся в печатных словарях, в том числе и структурная. Кроме того, именно электронная форма словаря позволяет существенно ускорить поиск словарной статьи, а также открывает перед человеком принципиально новые возможности в работе над переводом текстов» [2, 78]. В то же время электронные словари сохранили некоторые недостатки обычных словарей. Разрешение многозначности и снятие омонимии – главные проблемы компьютерной лексикографии. Кроме того, существует проблема неполноты словаря, а также поддержания словаря в актуальном состоянии.

Богатые коллекции словарей на русском языке представлены на сайтах и порталах «Словари и энциклопедии на Академике» [3], «Грамота.ру» [4], «Культура письменной речи» [5], «Словопедия» [6] и многих других.

Сайт «Словари и энциклопедии на Академике» [3] содержит огромное количество словарей из разных сфер (техника, медицина, естествознание, философия и многие другие). Перечень

лингвистических словарей тоже очень внушительный: различные словари и справочники по правописанию; толковые словари (7 источников); этимологические и исторические словари (8 источников); стилистические и ортологические словари; словари синонимов, антонимов и др. На сайте представлен самый значительный, по нашим наблюдениям, блок фразеологических и паремиологических словарей: «Толково-фразеологический словарь» М. Ю. Михельсона (в том числе в оригинальной орфографии); «Фразеологический словарь русского языка» (электронный вариант подготовлен на основе «Большого фразеологического словаря русского языка» В. Н. Телия); «Фразеологический словарь русского литературного языка» (электронный вариант подготовлен на основе одноименного словаря А. И. Федорова); «Большой словарь русских поговорок» (электронный вариант подготовлен на основе одноименного словаря В. М. Мокиенко и Т. Г. Никитиной); В. И. Даль. «Пословицы русского народа»; «Словарь народной фразеологии» (электронный вариант подготовлен на основе словаря

В. Кузмича «Жгучий глагол: Словарь народной фразеологии») и др. Кроме того, дополнительный материал о фразеологических и паремиологических единицах можно получить из других лексикографических источников, представленных на сайте. Удобство его построения в том, что можно осуществлять поиск прямо с главной страницы, причем по выбранным источникам или по всем ресурсам одновременно.

Задаем, например, для поиска УСК *седьмая вода на киселе* и получаем возможность анализировать материалы целого ряда словарей.

В «Толково-фразеологическом словаре» М. Ю. Михельсона приводятся толкование выражения; контексты его употребления; синонимичные выражения:

Седьмая вода на киселе _дальняя родня (чуть не по Адаму)‘.

Сравн.: *В Сибири... родство, свойство и кумовство считается... чуть не до двадцатого колена. Седьмая вода на киселе, десятая вода на квасине и всякая с боку припека из роду из племени не выкидается* (Мельников. На горах).

См.: *На его бабушке сарафан горел, а мой дедушка пришел да руки погрел; Наши собаки из одного корыта лакали; Моя бабушка с его дедушкой вместе капусту садили; Наша Марина вашей Катерине двоюродная Прасковья; Нашему слесарю _ двоюродный кузнец.*

«Фразеологический словарь русского литературного языка» содержит толкование выражения; контексты его употребления; стилистические пометы и информацию о варианте выражения, связанном с меной его компонента:

Седьмая (десятая) вода на киселе разг. экспрес. _очень дальний родственник‘. Напр.: _ *Да наследники-то какие: троюродные, седьмая вода на киселе* (Гончаров. Обломов); _ *Вот какой у неё защитник нашёлся? Что он ей, сват или брат? Десятая вода на киселе?..*

(В. Амигулов. Этот странный Иван).

«Большой словарь русских поговорок» содержит информацию о разных вариантах выражения, связанных с меной его компонентов (диалектные варианты сопровождаются соответствующей пометой); толкование выражения; его стилистическую характеристику:

Девятая (десятая, седьмая) вода на киселе народн. ирон. _об очень дальнем родственнике, отдалённом родстве‘;

Седьмая вода на гуще (на гущенке) перм. ирон. _то же, что девятая вода на киселе‘;

Десятая (седьмая) водина на квасине (на киселине) пск. шутол. ирон. *—*то же, что девятая вода на киселе'.

В «Справочнике по фразеологии» приводятся толкование выражения и информация о его происхождении:

Седьмая вода на киселе *—*о человеке, находящемся в крайне отдаленном родстве с кемлибо'. Происхождение этого оборота связано с изготовлением киселя, представляющего собой жидкую студенистую массу. Если кисель долго стоит, то он теряет свои вкусовые качества и на нем появляется слой воды. Седьмая вода – вода, появившаяся на длительно стоявшем киселе седьмой раз, отчего ее вкус ничего общего с киселем уже не имеет. Числительное *семь* употреблено здесь обобщенно-символически, как и в других русских пословицах и поговорках, поэтому оно варьируется. Возможен также вариант *десятая вода на киселе*.

В словарях синонимов приводятся следующие близкие по значению слова и выражения: *дальняя родня; десятая вода на киселе; нашему забору двоюродный плетень; нашему слесарю двоюродный кузнец; родственник; четвероуродный*.

Кроме того, если переходить по ссылкам сайта, можно получить и другую информацию, связанную с анализируемым УСК.

Несомненный интерес у исследователя вызовет сайт «Фразеология.ру», содержащий электронный «Фразеологический словарь», построенный на материале «Словаря современного русского литературного языка» в 17-ти томах, который «отражает ранний исторический этап в научном осмыслении русской фразеологии» [7]. В нем, как пишут создатели «Фразеологического словаря» А. Н. Тихонов, А. Г. Ломов, А. В. Королькова, «совмещены оба подхода в понимании состава фразеологической системы: узкий (как совокупность идиоматических выражений) и широкий (как совокупность всех типов устойчивых сочетаний, включая составные наименования, терминологические сочетания, разнообразные типы устойчивых конструкций с семантически слитными и неслитными компонентами)» [7].

Практически все «заромбовые» выражения «Словаря современного русского литературного языка» в 17-ти томах вошли во «Фразеологический словарь», представленный на сайте «Фразеология.ру». Составители словаря «не ставили цели сократить словарные статьи семнадцатитомника с точки зрения своего (субъективного!) понимания фразеологии, оставив право на вынесение вердикта читателю...» [7].

Представлен в данном словаре и УСК *седьмая вода на киселе*, приводятся его толкование и контекст употребления. Особенностью подачи анализируемого фразеологизма в данном словаре является иная, чем в других словарях, стилистическая характеристика и наличие в качестве варианта нумеративного компонента числительного *девятая* (приводится также в «Большом словаре русских поговорок»):

Седьмая (девятая, десятая) вода на киселе (простореч., шутол.) *—*о дальнем родстве'. *Пошли торопливые расспросы: кто смоленский, кто полтавский, кто донской. Нашлись земляки, почти родичи, «седьмая вода на киселе»* (Казакевич).

Вторым типом электронных ресурсов, востребованных в настоящее время исследователями устойчивых словесных комплексов, являются корпуса текстов. Обращение к корпусам текстов открывает для фразеологических исследований ряд дополнительных возможностей. Работа с корпусом позволяет определить частотность исследуемых фразеологических единиц; уточнить стандартную с точки зрения современного узуса начальную форму фразеологизма; определить набор наиболее значимых вариантов каждого фразеологизма; выявить многозначность фразеологических единиц и уточнить их конкретные значения; определить типичные для каждого фразеологизма трансформации; охарактеризовать типичное для исследуемых фразеологизмов

окружение и виды контекстов, в которых их использование воспринимается как наиболее естественное [8]. Решить эти и другие задачи можно благодаря богатому репрезентативному эмпирическому материалу, представленным в корпусе реальным контекстам со «встроенными» в них устойчивыми словесными комплексами. Экстралингвистическая информация о текстах, содержащих УСК, позволяет сделать выводы о представленности определенной фразеологической или паремиологической единицы в текстах разной стилистической, жанровой, хронологической и т. д. маркированности.

Перечисленными возможностями не исчерпывается специфика применения корпусных методов в исследованиях УСК. Набор решаемых с помощью корпуса задач в каждом конкретном случае может быть различным и зависит от целого ряда факторов (в первую очередь от цели исследования и состава фразеологического и паремиологического материала).

Так, запросы разного типа (*вода + на + кисель; седьмой + кисель; десятый + кисель*), оформленные в Национальном корпусе русского языка [9], позволяют выявить более 90 контекстов с УСК *седьмая вода на киселе*, его вариантами и трансформациями.

Наиболее типичной для данного УСК является мена нумеративного компонента: варианты с компонентом *седьмая* составляют 79 % анализируемого материала, с компонентом *десятая* – 10 %, с другими компонентами – 11 %. При этом вариант с компонентом *девятая*, получивший лексикографическое закрепление, представлен в текстах НКРЯ только один раз: *Повертел в руках, переписал в какой-то кондуит все данные, необходимые для отыскания подозрительной личности. – Она вам – девятая вода на киселе* (Павел Сиркес. Труба исхода (1990–1999)). Кроме того, контексты демонстрируют трансформации УСК с нумеративами *вторая, третья, пятая, двенадцатая, двадцатая*, а также с адъективными и местоименными компонентами (*родственная, симбирская, какая-то, бог весть какая*): *Приехал Любин какой-то десятиуродный брат, одним словом, двадцатая вода на киселе* (В. С. Новицкая. Безмятежные годы (1912)); *И вот как раз той зимой подруга написала маме, что ее троюродный или четвероуродный... короче, какаято вода на киселе едет в Питер на курсы обучения чему-то или усовершенствования чего-то* (Нина Катерли. Дневник сломанной куклы // «Звезда», 2001). Мена компонента *на киселе*, отраженная в «Большом словаре русских поговорок» и имеющая диалектный характер, в текстах НКРЯ не представлена. Единичными, по данным корпуса, являются другие трансформации УСК (усечение состава и инверсия компонентов): *Я, сударь, вам не про то; я вот хотел слово вымолвить, – ведь она, сударь, мне-то почти из родни, то есть из дальней, примером, как говорится, седьмая вода, то есть уж не побрезгайте словом нашим, сударь, люди мы темные – да сызмалетства такая!* (Ф. М. Достоевский. Хозяйка (1847)); – *Этотто, что на монетке, он тому Наполеону кем же доводится? – На киселе седьмая вода. Иван Дмитриевич вынул часы, щелкнул крышкой* (Леонид Юзефович. Костюм Арлекина (2001)).

Большинство контекстов корпуса включает в себя анализируемый УСК в значении, отраженном во фразеологических словарях (*о дальнем родстве*), при этом, как правило, определение степени родства субъектов не ограничивается одним устойчивым сочетанием, в контексте содержатся и другие пояснения на этот счет: *Это был Штерн – десятая вода на киселе, её троюродный или четвероуродный дядя* (Ю. О. Домбровский. Факультет ненужных вещей, часть 5 (1978)); *Таким образом, он стал моим зятем, а я – каким-то там очень далеким, седьмая вода на киселе, троюродный плетень нашему забору, но все же родственником Марка Твена* (Владимир Матлин. Марк Твен и я (2003) // «Вестник США», 2003.11.26). Материалы корпуса позволяют также выявить употребление данного УСК в других значениях. Напр., *о никчемном человеке*: *И ничем вроде бы не примечательный, так себе человечико, плевый, десятая вода на киселе, пена с сыворотки* (Владимир Личутин. Любостай

(1987)).

Интерес у исследователя может вызвать и информация о том, у какого автора и в каком произведении впервые (по данным корпуса) употреблен тот или иной УСК. В нашем случае это роман Ф. М. Достоевского «Бедные люди» (1846): *Уж как бы там ни было, а я вам хоть дальний родной, хоть, по пословице, и седьмая вода на киселе, а все-таки родственник и теперь ближайший родственник и покровитель; ибо там, где вы ближе всего имели право ис-*

кать покровительства и защиты, нашли вы предательство и обиду. Пользуясь метаразметкой и статистикой корпуса, можно также получить сведения о распределении УСК по годам, о количестве его употреблений разными авторами, в текстах разных стилей, жанров и др. Таким образом, использование корпусных методов позволяет проводить исследование устойчивых словесных комплексов, которое основано на большом количестве реальных контекстов (с учетом их метатекстовых характеристик) и, следовательно, имеет объективный характер.