

УДК 811.512.145'37:811.511.131'37

DOI: 10.18384/2310-7278-2016-5-137-144

ПОТЕНЦИАЛ КОРПУСНЫХ И БУМАЖНЫХ СЛОВАРЕЙ ДЛЯ ЛЕКСИКО-СЕМАНТИЧЕСКИХ ИССЛЕДОВАНИЙ (НА ПРИМЕРЕ СРЕДСТВ ВЫРАЖЕНИЯ ПОНЯТИЯ «ЖЁЛТЫЙ» В ТАТАРСКОМ И УДМУРТСКОМ ЯЗЫКАХ)

Тагирова Ф.И., Чумарина Г.Р., Акулина Л.В.

*Казанский инновационный университет им. В.Г. Тимирязова (ИЭУП)
420111, г. Казань, ул. Московская, д. 42, Российская Федерация*

Аннотация. В статье раскрывается потенциал корпусов и словарей как базы для исследований в области перевода и как инструмента для обучения переводу. Уделяется внимание и возможностям выявления семантических и функционально-стилистических особенностей лексики на примере слов, обозначающих жёлтый цвет, в татарском и удмуртском языках и определению их общих и специфических значений на основе анализа их отражения в корпусных и бумажных словарях.

Ключевые слова: электронные корпуса, татарский язык, удмуртский язык, цветообозначающая лексика

POTENTIAL OF CORPORA AND PAPER DICTIONARIES FOR LEXICAL-SEMANTIC RESEARCHES (ON THE EXAMPLE OF EXPRESSION MEANS OF THE CONCEPT "YELLOW" IN THE TATAR AND UDMURT LANGUAGES)

F. Tagirova, G. Chumarina, L. Akulina

*Timiryasov Kazan Innovative University
42, Moskovskaya St., Kazan, Russian Federation, 420111*

Abstract. The article reveals the potential of corpora and dictionaries as a basis for researches in the field of the translation and as a tool for training in the translation. The attention is paid to opportunities to detect the semantic and functional-stylistic features of lexicon on the example of the words denoting yellow color in the Tatar and Udmurt languages and to definition of their general and specific meanings on the basis of the analysis of their reflection in corpora and paper dictionaries.

Keywords: electronic corpora, Tatar language, Udmurt language, colour naming lexicon.

Составление и введение в обиход электронных корпусных словарей является неизбежным и очень значительным вкладом как в развитие самого языка, так и самых передовых методов и направлений лингвистической науки. Исследования на основе корпусов стали основной методологией, используемой во многих

отраслях лингвистики, так как обеспечивают самый широкий спектр исследований фактическими данными, отличающимися надёжностью, разнообразием и конкретностью. С 1990-х годов корпусные словари стали применяться в переводе. Корпусная лингвистика не является однородной методологией, она используется с разной степенью детализации и разной опорой на количественные и качественные методы и со следующими особенностями: распознаваемый машиной естественный язык, сбалансированный и репрезентативный проект корпуса, систематический и исчерпывающий анализ.

Относительно недавно началось серьёзное изучение потенциала корпусных словарей. Исследование национальных языков на достоверном материале с использованием современных технологий автоматической обработки текстов позволяет выработать новые подходы к решению актуальных проблем изучения и исследования национальных языков. Корпусный словарь и связанные с ним корпусные аналитические инструменты, которые становятся очень популярными в широком спектре исследовательских методик, представляют собой способ для проведения эмпирического исследования текстовых характеристик языка, наряду с традиционными исследованиями на основе бумажных носителей. Так, на их базе могут быть осуществлены самые разные изыскания: выявлены, прослежены и убедительно доказаны процессы семантической деривации, функциональные особенности лексики, квантитативные исследования различных единиц языка и т. п. За последние десятилетия кор-

пусные ресурсы значительно повлияли также на исследования и практику перевода. В практике перевода корпус имеет сильное влияние на то, что касается работы профессионалов-переводчиков, обучающихся и пользователей. Особенно интересной представляется возможность изучения особенностей переводческого поведения, ориентированного на нормы, что нам видится малоизученным на сегодняшний день. Более поздние, связанные с данной темой направления исследования рассматривают перевод в связи с изменениями языка, например, то, как развитие переводческих стилей и норм связаны с развитием языковых норм, или как влияют на перевод, в свою очередь, изменения языка. Корпусы служат и решению многих проблем, связанных с решением принципов точности, эквивалентности, тождества перевода и пр. Например, известно, что «языковое своеобразие любого текста, ориентированность его содержания на определённый языковой коллектив, обладающий лишь ему присущими «фоновыми» знаниями и культурно-историческими особенностями, не может быть с абсолютной полнотой «воссоздано» на другом языке [9]. Корпуса же в этом плане обеспечивают переводчика практически неограниченным материалом, отражающим семантические, контекстуальные и иные характеристики языковых единиц, что позволяет получить объективную картину их функционирования и максимально подчинить перевод требованиям вышеназванных принципов.

В данном случае на базе корпусных словарей проводится сравнительный анализ цветообозначающей лексики в татарском и удмуртском языках для

выявления существенных сходств и национальной специфики.

В обоих языках обозначения жёлтого цвета очень употребительны как для называния цвета окружающих предметов, так и в метафоризированных значениях, а также имеют неограниченную сочетаемость. В татарском языке для обозначения основного оттенка жёлтого цвета используется слово *сары*. В древнетюркском словаре лексема *сары* зафиксирована в следующих значениях: *sarı* 1. Жёлтый (МК I 374): *sarı* *altun ыгън кьтъль жёлтое золото, белое серебро* (Тон 48)... 2. бледный: *sarı* *tañ ata keldi занялся бледный рассвет...* 3. Соловый (*о масти лошади*)... [4, с. 488]. В «Кратком историко-этимологическом словаре татарского языка» происхождение лексемы представлено следующим образом: *сары* – ‘жёлтый’ – общетюрк. *сары, сарыг* ‘жёлтый, белёсый, светлый’ [1, с. 172].

А.Н. Кононов пишет, что в древности в некоторых тюркских языках, например, в хазарском и, возможно, в болгарском, слово *сары* (*г*) в форме *сар* имело значение (как в современном чувашском, ‘белый’, о чём свидетельствует название хазарской крепости Саркел – *сар* + *кел*: чув.: *кил* ‘дом’ [6, с. 175]. То есть имеются основания говорить о семантической деривации слова *сары*.

В удмуртском языке жёлтый цвет обозначается словом ‘жёлтый’: жёлтый || жёлто- || желток; *курег пуз* ~ яичный желток; ~ *буёл* жёлтая краска; ~ *какыт* диал. жёлтая бусинка; ~ *музьемъёс геогр.* желтозёмы [14, с. 49]; ~ *бубыли* бабочка-лимонница; ~ *вумульы бот.* кубышка жёлтая; ~ *гадь* мухоловка малая (птица); ~ *губи* лисичка (гриб); ~ *зольгыри жулан* (птица); ~ *кушман*

морковь; ~ *йыро кый* медянка (змея); ~ *моль* злаковая ржавчина; ~ *мильым диал.* блины; ~ *пислэг* мухоловка желтоспинная; ~ *сяська бот.* 1) купавка красильная; 2) ромашка жёлтая; ромашка душистая; ~ *чечег* 1) песчанка (птица); 2) чиж; ~ *шушы* желтозобик (птица); ~ *юбер* иволга [12]; *Kurek-piz kaik тьиһ тьиһ мед lo!* [15, с. 30] ‘Как желток яйца пусть будет жёлтым-жёлтым!’ и т. д.

Происхождение цветообозначения *çуж* ‘жёлтый’ в удмуртском языке восходит к общепермской форме *çuž*, а также к более раннему финно-угорскому корню *šace* ‘родиться’ [8, с. 312].

Для получения наиболее полного представления о функционировании данных единиц цветообозначения в татарском и удмуртском языках обратимся к электронным корпусам соответствующих языков. В настоящее время реализуются различные проекты корпусов, имеющие определённые сходства и различия в принципах работы и объёмах представленного материала. Для примера обратимся к наиболее раннему проекту «Туган тел».

Объём Татарского национального корпуса «Туган тел» на сентябрь 2013 года составляет более 26 миллионов словоупотреблений. Корпус содержит тексты различных жанров (художественная литература, тексты СМИ, тексты официальных документов, учебная литература, научные публикации и др.). Каждый документ имеет метаописание (авторы, их пол, выходные данные, даты создания, жанры, части, главы и др.). Тексты, включённые в корпус, снабжены морфологической разметкой (информация о части речи и грамматических характеристиках словоформы) [13].

Другой проект подобного рода – Письменный корпус татарского языка. Он на август 2016 г. насчитывает более 116 миллионов слов. Включает в себя различные тексты: художественные произведения, сборники научных трудов, монографии, газеты и журналы, религиозную литературу и др. Корпус снабжён системой сложного морфологического поиска, системой синтеза татарской речи, морфологической разметкой текста и т. п. [10].

Что касается лексемы *сары* ‘жёлтый’, то она представлена в Письменном корпусе татарского языка 6509 вхождениями, например: *Сары көзгә кадәр озын жэй бар.* [10] ‘До жёлтой осени есть длинное лето’; ... *аскы иренен чайничэйни һәм сары сумкасын элakterен урамга чыгып йөггерде* [10] ‘... покусывая нижнюю губу и схватив жёлтую сумочку, убежала на улицу’, и т. п. В Татарском национальном корпусе «Туган тел» со словом *сары* – 3220 вхождений, например: *Сары яфрак төшен бара иде* [13] ‘Жёлтый лист падал’; *Күкэйгә дәү әнисе дә шатланды, Сары тавыкны мактады, Таһирны мактады* [13] ‘Бабушка тоже обрадовалась снесённому яйцу, жёлтую курицу похвалила, Тагира похвалила также’ и т. п.

Корпус удмуртского языка в настоящий момент включает в себя около 7,3 миллионов словоупотреблений. Тексты корпуса были размечены с помощью автоматического морфологического анализатора, около 88% словоформ корпуса имеют грамматический разбор. Омонимия в корпусе не снималась, т. е. каждой словоформе приписаны все возможные варианты разбора без учёта контекста. Последнее обновление корпуса было произведено 6 марта 2016 г.

В текущей версии корпуса подавляющее большинство словоупотреблений приходится на прессу 2007–2015 гг. (91%). Остальные тексты представляют собой блоги (6%) и нехудожественные тексты (Новый завет, статьи из Википедии и эссеистика (3%)). В будущем планируется расширить текстовую базу и включить в корпус художественную литературу [7].

С цветообозначением *чуж* ‘жёлтый’ в удмуртском национальном корпусе представлено 367 вхождений, 206 документов, например: *Чебер адске ук: поезд кошкыку, усьтэм өс дорын чуж флажокен сылэ чебер дһсяськем нылмурт...* [7] ‘Красиво смотрится, когда отъезжает поезд, возле открытой двери стоит красиво одетая женщина с жёлтым флажком’. – *Айда тогда тьсытыд ымдэ муж краскаен буялом!* [7] ‘Давай тогда твой рот покрасим в жёлтый цвет’ и т. п.

С лингвокультурологической точки зрения материал публицистических текстов является активной сферой функционирования языковых единиц, в частности, слов, обозначающих цвет. Данные единицы отражают культурные стереотипы общества, их изменения, ценностные приоритеты и трансформации.

Сходство и специфичность семантических и сочетаемостных свойств изучаемых слов в татарском и удмуртском языках во многом обусловлены их многозначностью. В обоих языках лексемы *сары* / *чуж* совпадают в следующих значениях:

- жёлтый, цвет жёлтой краски;
- желток (яйца);
- желтуха;
- светлый (цвет волос) и др.

Одним из специфических свойств цветообозначения *жёлтый* в удмуртском языке в сопоставлении с татарским языком является то, что лексема *çуж* 'жёлтый' в лингвокультурологическом освещении соотносима с термином *çуж*, связанным с материнской линией родства: *çужанай* 'бабушка по матери'; *çужатай* 'дедушка по матери'; *çужмурт* 'брат матери'. Например: *Пичиатай Виталий Семёнович но пичианай Тамара Григорьевна куинь полэс çужатаен çужанай луизы ини* [7] 'Прадедушка Виталий Семёнович и прабабушка Тамара Григорьевна уже в третий раз стали бабушкой и дедушкой'; *Мон – çужатайлэн яратоно внукез* [11, с. 28] 'Я любимый внук бабушки' и др.

И в удмуртском и в татарском языках лексема *çуж* / *сары* 'жёлтый', а также его производные используются как информативный носитель коннотации 'старость', 'телесная ветхость', 'увядание', 'тоска', 'болезнь' и т. п., что можно увидеть на примерах: удм.: *Кукэ но вал, дыр, мэрччанлэс çуж йырсийэ. Кукэ но вал, дыр, мэрччанлэс çуж йырсийэ* [5, с. 69] 'Когда-то были, наверное, желтее расчёсанного волокна волосы. Когда-то были, наверное, желтее расчёсанного волокна волосы'; *Çуж али ной, вож али, çуж али ной, вож али. Çужзэ дисай – бэздылиз ной вожзэ дисай – бэздылиз* [5, с. 186] 'Желты ещё да зелены ещё, Желты ещё да зелены ещё. Жёлтое надела – выцвело (да), Зелёное надела – выцвело'; *Одигъёсызлэн сюлэмзы кутэ, мукетъёсызлэн муссы ужамысь дугдэ, çужектон аля, куинетһосызлэн йырвизызы сураське* [7] 'У одних сердце схватывает, у других печень перестаёт функционировать, желтуха появляется, а третьи с ума сходят' и т. д.

Тат.: Авыруның баишангыч билгеләре: хәлсезлек, аппетит бетү, эчкә газ җыелу, ашказаны тирәсендә авырлык тою, эч кибу, тән тиресе һәм күз агы саргаю, тән кычыту [13] 'Первоначальные симптомы болезни: слабость, отсутствие аппетита, газообразование, пожелтение кожных покровов и глазного белка'; *Сагынырга, саргаерга туры килер* [13] 'Скучать, тосковать придётся'; *Күз алмасы саргаю – бавыр чире* [13] 'Пожелтение глазного белка – заболевание печени' и др.

Наиболее распространённой и часто употребляемой формой функционирования цветообозначения 'жёлтый' как у удмуртского, так и у татарского народа является глагол-derivat от слова удм. *çуж-* и тат. *сары-çужектыны* и *саргаю* 'желтеть', причём как в прямом, так и в переносном значении, выражающем эмоциональное состояние тоски, грусти: *Тат.: Сары сагыштан агарды толларның толымнары* [3, с. 220] 'От жёлтой тоски поседела косы вдов'; *Аклыгы да, сафлыгы да болынарга түгелде. Саргаюлар сарысы да яктырттылар күңелне* [3, с. 40] 'И белизна, и чистота легли на луга. Желтизна тоски осветила душу'. Удм.: *Улмо чибор тәкдөы үкәж лыктэ тәкдөы Кам кузя. Мон висъыса кй çужекты, мон çужектһ тон понна* [11, с. 81] 'Пёстро-румяная, как яблоко, белая утка плывёт по белой Каме. Не от болезни я пожелтела, я пожелтела от тоски по тебе' и т. п.

Проведённые наблюдения подтверждают, что слова цветообозначения не только называют определённую область цветового спектра, но и имеют дополнительные значения, коннотации, представляющие собой концентрацию культурных знаний. Наблюда-

ния позволили также выявить общее и специфическое в семантике и сочетаемости прилагательных *сары / цуж*, принадлежащих неродственным языкам. Сходство на семантическом уровне обусловлено общечеловеческими закономерностями в психологии восприятия объективной реальности. Кроме того, сказывается постоянное контактирование и взаимовлияние языков, приведшие к типологическим и структурным сходствам и отразившиеся в том числе и на лексико-семантическом уровне. Специфичность же вызвана отличиями в ассоциативном восприятии и интерпретации окружающего мира, обусловленных наличием у татарского и удмуртского народов своих национальных реалий,

традиций и других историко-культурных факторов, повлиявших на формирование менталитета и языкового сознания каждого народа. То есть каждый народ в дополнение к символам и эталонам, общим с другими народами, имеет систему образов, характерных только для данной культурной общности и обусловленных индивидуальностью образного мышления конкретного народа, представляющего собой сложный ассоциативно-психологический процесс. Результаты анализа массивного материала корпусов ещё раз подтверждают данный факт. Технический потенциал корпусов даёт исследователям уникальную возможность сравнения, систематизации, анализа любых особенностей лексики.

ЛИТЕРАТУРА:

1. Әхмәтъянов Р.Г. Татар теленең кыскача тарихи-этимологик сүзлеге. Казан: Татар китап нәшрияты, 2001. 272 б.
2. Борисов Т.К. Песни южных вотяков. Ижевск: УдКнига, 1929. 100 с.
3. Галиев Ш.Г. Сиңа әйтер сүзем бар: шигырьләр, фәлсәфи фикерләр. Казан: Татарстан китап нәшрияты, 2008. 367 с.
4. Древнетюркский словарь. Л.: Наука, 1969. 676 с.
5. Кельмаков В.К. Образцы удмуртской речи: Северное наречие и срединные говоры / Науч. ред. М.Г. Агаманов. Ижевск: Удмуртия, 1981. 300 с.
6. Кононов А.Н. Семантика цветообозначений в тюркских языках // Тюркологический сборник. М.: Наука, 1978. С. 159-179.
7. Корпус удмуртского языка [Электронный ресурс]. URL: <http://web-corpora.net/UdmurtCorpus/search/>
8. Лыткин В.И., Гуляев В.С. Краткий этимологический словарь коми языка. Сыктывкар: Коми, 1999. 430 с.
9. Паршин А. Теория и практика перевода. М.: Высшая школа, 2008. 422 с.
10. Письменный корпус татарского языка [Электронный ресурс]. URL: <http://corpus.tatar/index.php>
11. Романов В.В. Ваёбыж кар: Кылбуръёс. Ижевск: Удмуртия, 2004. 192 с.
12. Русско-удмуртский, удмуртско-русский переводчик онлайн. [Электронный ресурс]. URL: <http://udmurinfo.ru/russko-udmurtskij-slovar>
13. Татарский национальный корпус «Туган тел». [Электронный ресурс]. URL: <http://web-corpora.net/TatarCorpus/>
14. Туганаев В.В., Воронцов П.И., Перевошиков А.П. Русско-удмуртский словарь географических терминов. Ижевск: Удмуртский институт истории, языка и литературы УрО РАН, 2003. 176 с.

15. Wichmann Y. *Wotjakische Chrestomathie mit Glossar*. Helsingfors: Druckerei Der Finnischen Literaturgesellschaft, 1901. 136 s.

REFERENCES

1. Әхмәтъянов Р.Г. Татар теленең кыскача тарихи-этимологик сүзлегә. Казан: Татар китап нәшрияты, 2001. 272 б.
2. Borisov T.K. *Pesni yuzhnykh votyakov* [Songs of southern votyaks]. Izhevsk, UdKniga, 1929. 100 p.
3. Галиев Ш.Г. Сиңа әйтер сүзем бар: шигырьләр, фәлсәфи фикерләр. Казан: Татарстан китап нәшрияты, 2008. 367 с.
4. *Drevnetyurkskii slovar'* [Ancient Turkic dictionary]. L., «Nauka», 1969. 676 p.
5. Kel'makov V.K. *Obraztsy udmurtskoi rechi: Severnoe narechie i sredinnye govory* [Samples of the Udmurt speech: Northern dialect and the median dialects]. Izhevsk, Udmurtiya, 1981. 300 p.
6. Kononov A.N. *Semantika tsvetooboznachenii v tyurkskikh yazykakh* [The semantics of the colour names in Turkic languages] // *Tyurkologicheskii sbornik* [Turkological collection]. M., «Nauka», glavnaya redaktsiya vostochnoi literatury, 1978. pp. 159-179
7. *Korpus udmurtskogo yazyka* [Elektronnyi resurs]. [The corpus of the Udmurt language [E-source]]. URL: <http://web-corpora.net/UdmurtCorpus/search/>
8. Lytkin V.I., Gulyaev V.S. *Kratkii etimologicheskii slovar' komi yazyka* [Brief etymological dictionary of Komi language]. Syktyvkar, Komi, 1999. 430 p.
9. Parshin A. *Teoriya i praktika perevoda* [Theory and practice of translation]. M., Vysshaya shkola, 2008. 422 p.
10. *Pis'mennyi korpus tatarskogo yazyka* [Elektronnyi resurs]. [The corpus of written Tatar language [E-source]]. URL: <http://corpus.tatar/index.php>
11. Романов В.В. *Ваёбыж кар: Кылбуръёс*. Ижевск: Удмуртия, 2004. 192 с.
12. *Russko-udmurtskii, udmurtsko-russkii perevodchik onlain*. [Elektronnyi resurs]. [Russian-Udmurt, Udmurt-Russian translator online. [E-source]]. URL: <http://udmurtinfo.ru/russko-udmurtskij-slovar>
13. *Tatarskii natsional'nyi korpus «Tugan tel»*. [Elektronnyi resurs]. [Tatar national corpus "Tugan tel". [E-source]]. URL: <http://web-corpora.net/TatarCorpus/>
14. Tuganaev V.V., Vorontsov P.I., Perevoshchikov A.P. *Russko-udmurtskii slovar' geograficheskikh terminov* [Russian-Udmurt dictionary of geographical terms]. Izhevsk, Udmurtskii institut istorii, yazyka i literatury UrO RAN, 2003. 176 p.
15. Wichmann Y. *Wotjakische Chrestomathie mit Glossar*. Helsingfors: Druckerei Der Finnischen Literaturgesellschaft, 1901. 136 s.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Тагирова Фяридя Инсановна – кандидат филологических наук, ведущий научный сотрудник, обособленное подразделение ГНБУ «Академия наук Республики Татарстан» «Институт языка, литературы и искусства им. Г. Ибрагимова АН РТ»; e-mail: feride2412@mail.ru

Чумарина Гузель Раисовна – кандидат филологических наук, доцент, доцент кафедры иностранных языков и перевода Казанского инновационного университета им. В.Г. Тимирязова (ИЭУП); e-mail: chumarina1@yandex.ru

Акулина Людмила Владимировна – старший менеджер деканата Институт развития дистанционных образовательных технологий Казанского инновационного университета им. В.Г. Тимирязова (ИЭУП); e-mail: akulinalv@mail.ru

INFORMATION ABOUT THE AUTHORS

Fyaridya Tagirova – candidate of Philological Sciences, leading research scientist of Ibragimov Institute of Language of Literature and Art of the Academy of Sciences of the Republic of Tatarstan;

e-mail: feride2412@mail.ru

Guzel Chumarina – candidate of Philological Sciences, associate professor, associate professor at the department of foreign languages and translation of the Timiryasov Kazan Innovative University;

e-mail: chumarina1@yandex.ru

Ljudmila Akulina – senior manager at the Dean's Office of the Institute of distance learning technologies of Timiryasov Kazan Innovative University;

e-mail: akulinalv@mail.ru

БИБЛИОГРАФИЧЕСКАЯ ССЫЛКА

Тагирова Ф.И., Чумарина Г.Р., Акулина Л.В. Потенциал корпусных и бумажных словарей для лексико-семантических исследований (на примере средств выражения понятия «жёлтый» в татарском и удмуртском языках) // Вестник Московского государственного областного университета. Серия: Русская филология. 2016. № 5. С. 137-144.

DOI: 10.18384/2310-7278-2016-5-137-144

BIBLIOGRAPHIC REFERENCE

F. Tagirova, G. Chumarina, L. Akulina. Potential of corpora and paper dictionaries for lexical-semantic researches (on the example of expression means of the concept "yellow" in the Tatar and Udmurt languages) // Bulletin of Moscow Region State University. Series: Russian philology. 2016. no. 5. pp. 137-144.

DOI: 10.18384/2310-7278-2016-5-137-144