

УДК 821.512.145(03)

## ЛИНГВОГРАФИЯ И ТЕКСТЫ: ИНВЕНТАРИЗУЮЩИЕ ЯЗЫКОВЫЕ СПРАВОЧНИКИ

*К.Р. Галиуллин, Р.Н. Каримуллина*

### Аннотация

Рассматриваются принципы и особенности составления, информационный потенциал языковых справочников, создаваемых в Казанском государственном университете и описывающих текстовые источники татарского и русского языков (словари языка Г. Тукая, М. Джалиля, татарских пословиц и поговорок и др.; словари языка русских пословиц и поговорок, русской поэзии XIX века, писем М.В. Ломоносова, произведений Г.Р. Державина, памятников Казанского края XVI – XVII веков и др.), словарных комплексов, которые объединяют традиционные (бумажные) и электронные компоненты.

**Ключевые слова:** лингвография, инвентаризирующий языковой справочник, электронный словарь, текстовые источники, русский язык, татарский язык, интернет-лингвография.

---

Для многих областей русской и татарской лингвистики чрезвычайно актуальной является задача введения в научный оборот языковых материалов, содержащихся в огромном массиве разнообразных текстов. Решению этой задачи во многом способствуют национальные корпуса (машинные фонды) языков. Существенную помощь в регистрации, системном описании могут оказать также различные инвентаризирующие языковые справочники.

В Казанском университете разрабатывается несколько лингвографических проектов, ставящих своей задачей словарную регистрацию материалов текстовых источников татарского языка (словари языка Габдуллы Тукая [1], Мусы Джалиля [2 (словарь слов и словоформ); 3 (словарь словоформ)], татарских пословиц и поговорок и др.) и русского языка (словари языка русских пословиц и поговорок [4], русской поэзии XIX века, писем М.В. Ломоносова [5], произведений Г.Р. Державина, памятников Казанского края XVI – XVII веков [6, 7], документов, связанных с русско-восточными отношениями XVI – XVII веков, и др.

В состав названных словарей входят конкорданс, частотный и обратный слово-/словоформоуказатели, а для справочников, описывающих поэтические тексты, также словарь рифм.

Конкорданс является одним из перспективных средств инвентаризации материалов текстов. Компьютерная реализация конкорданса (в виде CD- или интернет-варианта) позволяет эффективно справиться с задачей абсолютно пол-

ного цитирования, которая до недавнего времени представлялась невыполнимой<sup>1</sup>.

В ряде случаев полнота цитирования достигается и в рамках традиционного (бумажного) словаря. Это реализуется путем включения в корпус словаря описываемых текстов (см., например, индексированную текстотеку стихотворений цикла «Моабитские тетради» в [2, с. 214–251]).

Точная адресация (номер текста, страницы издания или строфы, строки) позволяет ускорить доступ к расширенному контексту (в том числе в используемом академическом и в других изданиях).

Базовый вариант статьи конкорданса (слово/словоформа + цитата + адрес) дополняется сведениями о частоте, в таком виде конкорданс выполняет функцию и алфавитно-частотного словаря; см., например:

**АШКЫНУ** (6) ▲ А.<sup>(3)</sup> [АС]<sup>2</sup>

- ← **ашкына** (1) ■ Эллэ кая ашкына. .... Δ 36:4/4 ○ Сэг
- ← **ашкынды** (1) ■ Карбикэ ярсып язга ашкынды, .... Δ 85:10/1 ○ КК
- ← **ашкынып** (4) ■ Һәм карый ашкынып ..... Δ 9,1:2/3 ○ СЧ
  - Ул ашкынып көтэ сөйгәнән. .... Δ 43:15/2 ○ Күл
  - Бар да, ашкынып, .... Δ 48:12/4 ○ Сер
  - Шулай ашкынып ..... Δ 48:11/4 ○ Сер

<...> [2]

**ЯТУ** (2)

- Тыныч яту гүрдә. .... ● 3-6:4
- Шул хәлендә яту кайда тыныч кына! ..... ● 91-13:4

<...> [1] (подробнее о структуре словарных статей по материалам словаря

словоформ см. [9]).

Важной частью статей конкорданса является справочная зона, которая способствует представлению в нем элементов других типов словарей – толкового, энциклопедического.

Семантизирующая часть формируется сравнительными материалами из источников двух типов:

а) отражающих сведения, которые относятся ко времени создания описываемого источника; они демонстрируют особенности употребления слов в языке того периода, знакомят со словарными описаниями того времени, с другими употреблениями этой единицы (см., например, данные Малого энциклопедического словаря Брокгауза-Ефрона (1907–1909 гг.) в справочной части статей [1] или материалы Словаря русского языка XVIII века и Словаря Академии Российской в статьях [5]):

**ЧЫТАЛҖА** (2) ▲ Ч. – шәһәр исеме [БЕ (Чаталджа)] <...> [1]

**ЛИЦЕНЦИАТ**<sup>(-)</sup> / **ЛИЦЕНЦИЯТ** (1) ▼ XVIII: Л. [ЛомПСС, 11: 276]; Л. / Л.  
/ ... 'Лицо, имеющее ученую степень ниже докторской; кандидат в доктора' [С18]  
<...> [5]

б) отражающих данные современного периода; они показывают некоторые особенности употребления слова в настоящее время: графический облик лек-

<sup>1</sup> Ср.: «Цитирование в словаре может быть более или менее исчерпывающим, но никогда абсолютно исчерпывающим, иначе объем словаря превысит объем лексикографируемого произведения во много раз» [8, с. 19].

<sup>2</sup> Сокращения, представленные в примерах из словарей, не расшифровываются.

сической единицы, зафиксированные в источнике варианты, количество значений, наличие омонимии:

**КАРБИКЭ** (3) ▲ **К.**<sup>(1)</sup> [АС] <...> [2]

Толкования приводятся для единиц ограниченного употребления (устаревших, диалектных и т. п.). Одним из оснований для включения в словарную статью пояснения, толкования является отсутствие слова в современных словарях литературного языка:

**КҮЛЭНКЭ** (1) ▲ 'күлэгә' [ТТДС-1: 282] <...> [1]

**МӘХКҮМ** (1) ▲ **М.** 'хөкем ителгән' [ГТРАС] <...> [2]

Следует особо сказать о языке поэзии, которая «всегда иносказательна» [10, с. 341]. На данном этапе не ставится (а без специального изучения и не может ставиться) задача эксплицирования в словаре индивидуально-авторского содержания единиц поэтического текста. Это задача особых справочников – словарей образов, тропов или толковых словарей, описывающих особенности писательского идиолекта. В рассматриваемых же словарях помощь в раскрытии текстового значения слова призваны оказать цитаты.

Повышению информационного потенциала словарей в немалой степени способствуют отсылочные статьи и отсылки внутри основных статей. В конкордансе [1] это:

- отсылки (в том числе и перекрестные) к вариантам:

**ТЕФҮ** (1) ▲ ↔ **тфү** <...>

**ТФҮ** (2) ▲ ↔ **тефү** <...>

**ГАЛӘЛГОМЬЯ** (1) ▲ 'сукурларча' [ГФ; МС (галәлгомья)]; ↔ **галәлгомья** <...>

**ГАЛӘЛГОМЬЯ** (1) ▲ → **галәлгомья** <...>

- отсылки к неначальным компонентам сложных слов, префиксальных производных или сверхсловных единиц, которые приведены в тексте в слитном написании:

**алмадым** → **житалмадым, жөгралмадым, тоталмадым, түзәлмәдем, утралмадым**

**аяк** → **әл-аяк, кыргаяк, тимераяк, яланаяк**

**талып** → **арып-талып**

- отсылки к статье, где приведены материалы, поясняющие данную единицу:

**ТӘРӘХХЕМ** (2) ▲ 'ачыну, кызгану; мәрхәмәт итү' [ГФ; МС] <...>

**ТӘРӘХХЕМЛӘР** (1) ▲ → **тәрәххем** <...>

- отсылки к другим частям словаря; например, конкорданс к поэтическим текстам дополняется ссылками на соответствующие единицы, которые содержатся в конкордансе, описывающем материалы заголовочных частей стихотворений (названия произведений, эпиграфы, посвящения и др.):

**бичара** → 3.1

**Пушкин** → 3.1

Частотные словари предоставляют сведения об употребительности описываемых единиц, чаще всего слов и словоформ, однако объектами описания могут выступать также другие элементы анализируемого текста: буквы, буквосочетания, слоги, словосочетания, предложения и др.; см., например, частотные характеристики (в виде абсолютной и относительной частот и ранга) букв

(гласных/согласных, начальных/конечных) и слогов в тексте, списке словоформ и списке слов [4, с. 175–180].

Минимальный набор компонентов частотного языкового справочника – описываемая единица + частота (как правило, абсолютная).

Наиболее популярными являются алфавитно-частотные и частотно-алфавитные языковые справочники.

В рамках разрабатываемых словарных проектов функцию алфавитно-частотного словаря, как было отмечено выше, выполняет конкорданс, все статьи которого содержат сведения о частоте заголовочных (в ряде случаев – и внутрисклассовых) единиц – слов или словоформ; см. выше примеры статей конкордансов в [2 (словарь слов и словоформ); 1 (словарь словоформ)]:

В частотно-алфавитном словаре единицы располагаются в порядке убывания частоты:

мин (310)	бер 636/6
ул <sup>II</sup> (227)	бу 613/8
булу (166)	дә 519/1
син (154)	ул 509
белән (137)	мин 401
һәм (115) <...> [2]	да 400/5 <...> [1]

В ряде случаев общая частота может дополняться сведениями о количестве случаев особого, своеобразного употребления описываемой единицы в тексте или частоте этой единицы в определенном подмножестве (определенных подмножествах) текстов; см., например, сведения о количестве текстовых словоформ, зафиксированных в произведении «Мөбәрәк тәсбих өзелде», прозаическом по форме, но включенном Г. Тукаем, а вслед за ним и составителями академического издания в поэтический сборник (см. в примере из [1] частотную характеристику, приведенную после косой линии).

Абсолютная частота может дополняться сведениями об относительной и накопленной относительной частотах, а также о ранге описываемых единиц; эти данные могут быть представлены в специальных разделах (см., например, соответствующие данные в таблице, посвященной частотному распределению словоформ, в [5, с. 198–199]):

Ранг	Абсолютная частота	Относительная частота	Количество словоформ	Количество текстовых словоформ	Накопленная относительная частота
1	878	4.42%	1	878	4.42%
2	576	2.90%	1	576	7.32%
3	366	1.84%	1	366	9.17%
4	299	1.51%	1	299	10.67%
5	244	1.23%	1	244	11.90%

<...>

Указанные частотные характеристики особенно ценны для сравнительно-сопоставительных исследований идиолектов разных авторов или языка разных произведений / разных периодов творчества одного автора.

В большинстве случаев пользователи частотного справочника обращаются к единицам, находящимся в его верхней части (то есть к наиболее распростра-

ненным). Однако в писательских словарях особый интерес вызывают и редкие слова, которые в немалой степени отражают индивидуально-авторские особенности стиля [8, с. 70; 11].

Обратный словарь представляет собой перечень слов/словоформ, расположенных в обратном (инверсионном) алфавитном порядке с учетом конечных букв описываемых единиц.

Этот вид языкового справочника предоставляет сведения об образовании и функционировании единиц, имеющих сходные финали: тех или иных форм слов (со сходными словоизменительными аффиксами), тех или иных производных (со сходными постфиксальными морфемами), служит источником для исследований по словоизменению и словообразованию, для изучения особенностей употребления форм слов и дериватов в писательском идиолекте.

Инверсионный список может дополняться:

- сведениями о частоте слова или словоформы (см., например, обратный словоформоуказатель в [7, с. 226–255]):

<...> (2) помещик  
 (2) стройщик  
 (5) розсылщик  
 (2) посылщик  
 (1) заимщик  
 (1) оханщик  
 <...>

- данными о количестве слов/словоформ с определенной финалью; эти сведения могут быть вынесены в особый раздел (см., например, табл. 4 в разделе «Квантитативные характеристики» в [4, с. 177–178]);

- указаниями на варианты (см., например, обратный словоуказатель в [4, с. 141–148]):

<...> свадьба	<...> глаголь
=* женьдба	<sup>1</sup> слать
/ похвальба	=* <sup>2</sup> слать
/ стрельба	*послать
/ цельба	// стлать
/ женитьба	посылать
<...>	<...>

В приведенных примерах косые линии указывают на наличие вариантов (количество линий соответствует количеству вариантов). Знаком равенства помечены варианты. Выносными цифрами снабжены омонимы. Звездочкой отмечены слова, которые встречаются в тексте в начальной форме или в форме, сходной с начальной (например, для части существительных это форма винительного падежа);

- отсылками к другим разделам словаря (например, к словарю рифм) и др.

Как показывает анализ, значительная часть единиц в перечне слов и перечне словоформ совпадает. Так, из 2187 слов и вариантов, зафиксированных в названном выше обратном словоуказателе, 1030 выступают как текстовые формы в источнике данного словаря – в сборнике пословиц Петровской галереи; в [2] совпадают приблизительно 20% единиц. Поэтому для ряда языковых справоч-

ников, в которых единицами описания являются слова и словоформы, целесообразной представляется совмещенная форма инверсария – обратный указатель слов и словоформ (см., например, в словаре цикла «Моабитские тетради» М. Джалиля [2, с. 191–212]):

<...> чаба	<...> сөюен
= <b>нава</b>	= <sup>1, II</sup> <b>ин</b>
= <b>дэва</b>	= <b>жин</b>
= <b>ага</b>	= <b>кин</b>
<b>дага</b>	= <b>тин</b>
кага	= <sup>1, II</sup> <b>мон</b>
юбкага	= <sup>1, II, III</sup> <b>сон</b>
аптекага	= <b>ун</b>
капкага	сыйпавын
<...>	<...>

Полужирным шрифтом набраны слова, обычным – словоформы. Слова, совпадающие с текстовой словоформой, сопровождаются знаком равенства. Выносные римские цифры указывают на наличие омонимов.

Специфическим для поэтической лингвографии видом языкового справочника является словарь рифм.

Словари рифм, которые созданы или создаются в рамках рассматриваемых словарных проектов, в настоящее время имеют две основные формы – прямой алфавитный и инверсионный алфавитный; см., например, указатели рифм в стихотворных текстах, содержащихся в письмах М.В. Ломоносова [5, с. 196–197]:

1) прямой алфавитный список

<...> пе'ть ↔ взлете'ть – #24/492:1, 2  
 побежа'л ↔ ва'л – #24/491:21, 22  
 подвигнул ↔ дости'гнул – #24/491:13, 12  
 посвяща'ет ↔ наполня'ет – #11/471:8, 9  
 руко'ю ↔ собо'ю – #24/491:19, 20 <...>

2) инверсионный алфавитный список

<...> #11/471:8, 9 – посвяща'ет ↔ наполня'ет  
 #24/491:10, 11 – горы' ↔ бугры'  
 #24/491:11, 10 – бугры' ↔ горы'  
 #24/491:19, 20 – руко'ю ↔ собо'ю  
 #24/492:3, 4 – горо'ю ↔ ного'ю <...>

Ряды, которые образованы рифмующимися словоформами, в данном издании снабжены адресами соответствующих стихотворных строк, включающими номера текста, страницы издания, стихотворной строки. Указание места ударения позволяет судить о типе рифмы, а ссылка на номер строки – о конфигурации рифм. Ряды могут также сопровождаться другими характеристиками рифмы.

Для большинства рассматриваемых языковых справочников разрабатываются электронные версии, среди которых важное место занимают интернет-словари, являющиеся, по мнению составителей, наиболее перспективными и удобными (см., например, интернет-версии языковых справочников на портале Казанского государственного университета в разделе «Казанский лингвогра-

фический фонд» – <http://www.klf.ksu.ru>). Особо следует отметить глобальную обобщественность материалов сетевых словарей.

Электронный словарь, с одной стороны, хранилище всевозможных данных, с другой – эффективный инструмент научного поиска (см. [12]).

Электронные версии словарей позволяют по желанию пользователя устанавливать объем текстового массива для обработки и работать как со всей совокупностью включенных в базу текстов, так и с различными ее подмножествами, а также с отдельно взятым текстом. Кроме того, они существенно расширяют возможности поиска, причем не только по одному параметру, но и по их совокупности.

Результаты проведенных исследований показывают значительный информационный потенциал инвентаризирующих языковых справочников, перспективность и эффективность использования их для многостороннего описания материалов текстов.

Работы по формированию компьютерных фондов, подготовке словарей, изданию и размещению языковых справочников в Интернете поддержаны, в частности, Федеральной целевой научно-технической программой «Исследования и разработки по приоритетным направлениям развития науки и техники» на 2002–2006 гг.; Российским гуманитарным научным фондом (проекты «Машинный фонд татарского языка: словарный подфонд», «Компьютерная поддержка русской лексикографии XVIII века»; «Большой корпус русского языка XVIII века», проект № 07-04-12147в; «Комплексный фонд русскоязычных памятников Казанского края XVI – XVII веков: текстовый и словарный подфонды», проект № 08-04-12146в); Российским фондом фундаментальных исследований (в рамках проектов № 05-07-90376, 02-07-90230 и др.); Культурным центром имени Дж. Неру при Посольстве Индии в Российской Федерации; Федеральной целевой программой «Русский язык» (проект «Компьютерный лингвографический фонд русского языка»), Аналитической ведомственной целевой программой «Развитие научного потенциала высшей школы (2009–2010 гг.)» (рег. номер 2.2.1.1/6944), Республиканской целевой программой «Русский язык в Татарстане», а также Комитетом по реализации Закона «О языках народов Республики Татарстан» при Кабинете Министров Республики Татарстан и др.

### Summary

*K.R. Galiullin, R.N. Karimullina. Linguography and Texts: Inventorying Language Reference Books.*

The article views principles and features of making reference books which are being created in Kazan State University, along with their informational potential. The reference books describe textual sources of the Russian and Tatar languages, such as the vocabulary of G. Tukay's, M. Jalil's language, Tatar proverbs and sayings; the vocabulary of Russian proverbs and sayings, Russian poetry language (19th century), M.V. Lomonosov's letters language, G.R. Derzhavin's works of literature language, the language of written monuments of Kazan Region (16–17th centuries), dictionary complexes, uniting traditional (paper) and electronic components.

**Key words:** linguography, inventorying language reference, electronic dictionary, textual sources, Russian language, Tatar language, internet-linguography.

## Литература

1. *Галиуллин К.Р., Каримуллина Р.Н.* Габдулла Тукай сүзлеге. Шигърият теле = Габдулла Тукай. Словарь языка поэзии: 2 кит.— Казан: Мәгариф, 2009.
2. *Галиуллин К.Р., Каримуллина Р.Н., Минһажәева Л.С.* Муса Жәлил. «Моабит дэфтәрләр» теле: сүзлек = Муса Джалиль. Язык «Моабитских тетрадей»: словарь. – Казан: Мәгариф, 2006. – 255 б.
3. *Галиуллин К.Р., Каримуллина Р.Н., Минһажәева Л.С.* Муса Жәлилнең «Моабит дэфтәрләр» сүзлек = «Моабитские тетради» Мусы Джалиля: словарь. – Казан: Казан дәүләт ун-ты, 2006. – 224 б.
4. *Галиуллин К.Р., Мартъянов Д.А.* Словарь языка русских пословиц и поговорок конца XVII – первой половины XVIII века: сб. пословиц Петровской галереи. – Казань: Казан. гос. ун-т, 2006. – 188 с.
5. *Галиуллин К.Р., Каримуллина Г.Н., Каримуллина Р.Н., Гизатуллина А.Р., Каримуллина А.Н., Мартъянов Д.А.* Язык писем М.В. Ломоносова: материалы для словаря. – Казань: Изд-во Казан. ун-та, 2007. – 204 с. (интернет-компонент: <http://www.klf.ksu.ru/lomonosov>).
6. *Исламова Э.А., Галиуллин К.Р.* Казанский край: словарь памятников XVI века. – Казань: Изд-во Казан. ун-та, 2000. – 286 с.
7. *Галиуллин К.Р., Гизатуллина А.Р.* Казанский край: словарь языка памятников первой четверти XVII века. – Казань: Изд-во Казан. ун-та, 2008. – 276 с. (интернет-компонент: <http://www.klf.ksu.ru/kazan>).
8. *Карпова О.М.* Словари языка писателей. – М.: Изд-во Моск. полиграф. ин-та, 1989. – 107 с.
9. *Галиуллин К.Р., Каримуллина Р.Н.* О словарном описании языка произведений Мусы Джалиля: «Моабитские тетради» // Учен. зап. Казан. гос. ун-та. Сер. Гуманит. науки. – 2006. – Т. 148, кн. 3. – С. 55–64.
10. *Потебня А.А.* Эстетика и поэтика. – М.: Искусство, 1976. – 614 с.
11. Русская авторская лексикография XIX – XX веков: антология / Сост. Е.Л. Гинзбург и др. – М.: Азбуковник, 2003. – 511 с.
12. Компьютерная лингвография / Науч. ред. Н.К. Замов, К.Р. Галиуллин. – Казань: Изд-во Казан. ун-та, 1995. – 119 с. (интернет-версия: [http://www.ksu.ru/f10/publications/1995/comp\\_ling.php](http://www.ksu.ru/f10/publications/1995/comp_ling.php)).

Поступила в редакцию  
10.10.08

---

**Галиуллин Камилъ Рахимович** – доктор филологических наук, профессор, заведующий кафедрой теоретической и прикладной лингвистики Казанского государственного университета.

E-mail: [kamil.galiullin@ksu.ru](mailto:kamil.galiullin@ksu.ru); [kamil.galiullin@mail.ru](mailto:kamil.galiullin@mail.ru)

**Каримуллина Резеда Нурутдиновна** – ассистент кафедры теоретической и прикладной лингвистики Казанского государственного университета.

E-mail: [re\\_ka@mail.ru](mailto:re_ka@mail.ru)