



УДК 811.161.1'282(470.44)

## АНАЛИЗ ВЫСОКОЧАСТОТНОЙ ЛЕКСИКИ ДИАЛЕКТА (на материале корпуса диалектных текстов Саратовской области)

А. И. Буранова

Саратовский государственный университет  
E-mail: aburanova@list.ru

Статья посвящена изучению состава высокочастотной диалектной лексики и выделению особенностей языковой картины мира диалектоносителей.

**Ключевые слова:** диалект, количественный анализ, частота.

### Analysis of High-frequency Dialect Lexis (Based on Dialect Corpus of Saratov Region)

A. I. Buranova

The article is dedicated to the study of the composition of high-frequency dialect lexis and to the definition of linguistic world-image characteristics of dialect speakers.

**Key words:** dialect, quantitative analysis, frequency.

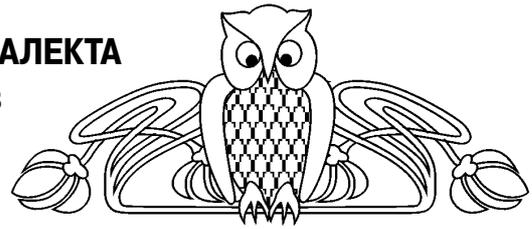
В современной лингвистике активно выполняются исследования с привлечением материалов частотных словарей. Применение данных об употребительности слова позволяет исследователям выяснить, насколько актуальным оно является для носителя языка. Анализ высокочастотных лексем помогает выявить наиболее существенные характеристики лексики отдельной языковой идиомы. Частотность употребления слова в речи является отражением его функционального веса в системе языка, а кроме того, тесно связана с «его парадигматической значимостью, важностью, существенностью для языка»<sup>1</sup>.

Высокочастотная лексика формирует ядро лексической системы. В ее состав входят лексемы, репрезентирующие базовые для представителей отдельной речевой культуры понятия и концепты.

Анализ высокочастотной лексики диалекта позволяет выявить некоторые характерные черты языковой картины мира носителей традиционной народной культуры.

В ходе исследования нами был проанализирован корпус диалектных текстов, записанных в Саратовской области, общим объемом 510 000 словоупотреблений. Данный массив текстов был обработан с помощью морфологического анализатора MyStem и программы лексикографического анализа текста KarGateka, созданной в ИФИ СПбГУ. В результате обработки был получен словарь и вычислена частота употребления каждой леммы, а также составлен список лемм в порядке убывания их частот.

В ходе анализа мы рассмотрим состав 500 наиболее частотных лемм словаря, состав-



ленного на основе диалектного корпуса, предварительно разбив их на зоны по 100 слов в каждой.

В состав **первой зоны** входят преимущественно служебные слова (союзы, предлоги, частицы), наиболее частотное из которых – союз *и* (3,5%). В данную зону также входит несколько глаголов: *быть, говорить, знать, давать, ходить, приезжать, работать, делать, рассказывать*. Группа существительных представлена леммами *год, человек, девчонка, время, бог*. В состав зоны входят прилагательные *другой* и *хороший*; наречные формы *сейчас, потом, здесь, туда, много*; практически все местоимения (наиболее частотными являются личные местоимения, особенно *я* (1,82%)); также зафиксированы числительные *один* и *два*. Следует отметить значительный перепад частот в данной зоне: если первая лексема имеет частоту 3,5%, то последняя – всего 0,14%.

Таким образом, первая зона образована служебными словами и небольшим количеством знаменательных слов, которые используются для обозначения наиболее общих и актуальных понятий и действий. Так, сюда входят глаголы движения, речевой деятельности; наречия времени и количества; существительные, служащие для обозначения человека и временных интервалов. Кроме того, в первой зоне присутствует существительное *бог*. Данный факт можно объяснить тем, что носители диалекта являются преимущественно верующими людьми. Высокая частотность личных местоимений в составе первой зоны свидетельствует о важности межличностных отношений для представителей народной речевой культуры и подтверждает положение об антропоцентричности диалектной картины мира.

Во **второй зоне** значительно увеличивается количество знаменательных слов и сокращается число служебных. При этом перепад частот является значительным: первая лексема данной зоны имеет частоту 0,14%, а последняя – 0,07%. Основную часть зоны составляют глаголы и существительные. Прилагательные не зафиксированы. В составе второй зоны можно выделить следующие группы существительных: термины родства (*мать, мама, бабушка, отец, сын, бабушка, муж*), обозначения животных (*лошадь, корова*), наименования построек (*дом, школа*), лексемы с религиозной тематикой (*церковь, господь*), лексемы, связанные с сельским хозяйством (*колхоз, земля*). Также в данную зону входят лексемы *война* и *жизнь*.



Среди глаголов здесь можно выделить следующие группы: глаголы движения (*пойти, прийти, уходить, поехать, приходить, ездить, выходить, уезжать*), глаголы чувственного восприятия и ментальной деятельности (*видеть, хотеть, глядеть, понимать*), глаголы состояния (*умирать, оставаться, сидеть, становиться, стоять*). Также во второй зоне зафиксированы глаголы *учиться* и *молиться*.

Таким образом, вторая зона образована более конкретными, однако нейтральными и общеупотребительными лексемами. Наиболее многочисленны термины родства и глаголы движения. Кроме того, как и в первой зоне, присутствуют лексемы, тематически связанные с религией. Высокая частотность лексемы *война* может объясняться тем, что все носители диалекта пережили данное событие, у многих на войне погибли близкие.

В **третью зону** входят преимущественно существительные и глаголы. Среди существительных можно выделить термины родства (*брат, дочь, сестра, дочка, родители, жена*), наименования частей тела (*нога, рука*), природных объектов (*речка, лес, вода*), поселений и их частей (*город, деревня, двор, улица*), пищевых продуктов (*хлеб, картошка*), религиозные термины (*вера, крест, святой*), наименования техники (*машина*). Достаточно частотными являются существительные *работа* и *правда*.

Среди глаголов, входящих в третью зону можно выделить следующие группы: глаголы ментальной деятельности и чувственного восприятия (*думать, помнить, забывать, бояться, любить, смотреть*), речевой деятельности (*звать, спрашивать*), интеллектуальной деятельности (*читать, писать*), глаголы, связанные с рыночными отношениями (*купить, продавать*); связанные с болезнью (*болеть, выхаживать*).

В составе данной зоны зафиксированы прилагательные *большой, маленький, больной, божий* и два числительных: *четыре, пять*.

Таким образом, третью зону образуют лексемы, связанные преимущественно с повседневной жизнью диалектоносителей. Следует отметить высокую частотность лексем, связанных тематически с болезнью. Данный факт объясняется тем, что носители диалекта – пожилые люди, которые зачастую имеют проблемы со здоровьем. Высокая частотность лексемы *правда* связана с актуальностью данного понятия для русской речевой культуры в целом и в особенности для народной речевой культуры.

**Четвертая зона** образована преимущественно существительными и глаголами, однако в ней резко возрастает количество прилагательных и наречий.

Так, выделяются следующие группы существительных: существительные с религиозной тематикой (*богородица, икона, батюшка, ангел, душа, молитва*), наименования частей тела (*голова*), существительные, связанные с сельским

хозяйством (*огород, скотина*); обозначающие временные интервалы (*день, зима*); обобщенные наименования людей (*народ, молодежь*).

Глаголы, входящие в четвертую зону, образуют следующие тематические группы: глаголы, связанные с эмоциональным состоянием (*верить, плакать*); глаголы движения (*подходить, сходить, дойти*); глаголы, связанные с рождением или смертью человека (*родить, родиться, помирать*); связанные с процессом принятия пищи (*есть, пить*).

Зафиксировано значительное количество наречий: состояния (*плохо*), времени (*скоро, всегда, давно, всегда*), места (*далеко, рядом*).

Также в эту зону входят прилагательные *живой, последний, скорый* и числительные *десять* и *семь*.

В целом можно отметить, что четвертая зона характеризуется наличием большого количества лексем с религиозной тематикой и временных идентификаторов и отсутствием союзов и предлогов.

В **пятой зоне** значительно увеличивается количество прилагательных. Так, зафиксированы прилагательные *старший, молодой, старый, интересный, прошлый, церковный, общий, земляной*.

Среди существительных выделяются те, которые обозначают учреждения (*больница, магазин*); наименования техники (*трактор, телевизор*); существительные с религиозной тематикой (*грех, бес*); термины родства (*дед, дядя, свекровь*); существительные, обозначающие период времени (*неделя, ночь*).

Среди глаголов, образующих пятую зону, могут быть выделены следующие подгруппы: глаголы движения (*увозить, пройти*); глаголы, связанные с ведением хозяйства (*посадить, сажать, доить*); обозначающие развлечение (*играть, гулять*); глаголы состояния (*спать, проживать*).

Сюда же входит ряд наречий: *вечером, утром, нынче, интересно*, и числительные *восемь, шесть*.

В целом пятая зона характеризуется увеличением количества прилагательных при полном отсутствии служебных частей речи.

Таким образом, служебные слова, преобладавшие в первой зоне, постепенно вытесняются знаменательными словами в следующих зонах и полностью отсутствуют в пятой зоне. Как отмечает Л. Н. Засорина, «наибольшей частотностью обладают “грамматические” слова, появление которых обусловлено строем языка, затем идут слова, отражающие тематику текстов»<sup>2</sup>.

В целом, анализ частотной лексики показал, что наиболее употребительная лексика является нейтральной, без стилистической окраски.

Результаты проведенного анализа частотных слов корпуса диалектных текстов, полагаем, можно сопоставить с данными, представленными в «Частотном словаре современного русского языка» (раздел «Частотный словарь живой устной речи»)»<sup>3</sup>. Подобное сопоставление дает возмож-



ность выявить, насколько специфично проанализированное частотное распределение лексем для диалектной речи.

Сопоставительный анализ словариков показал, что большое количество высокочастотных лексем с религиозной тематикой характерно для диалектной речи. Кроме того, лексем, связанные с сельским хозяйством, имеют большую частоту в речи диалектоносителей. Еще одной особенностью диалектной речи является высокая употребительность лексем *болеть*, *болезнь* и др. Кроме того, отличительной чертой словарика, составленного на основе корпуса диалектных текстов, является большое количество числительных. Данное распределение объясняется особенностью мышления диалектоносителей, проявляющейся в стремлении как можно более детально и точно описать какое-либо событие.

Анализ высокочастотной лексики позволяет выявить представления о мире, которые формируются в речи носителей диалекта. Человек представляется как физический организм (наименования частей тела, обозначения болезней и других состояний, пищи и процесса принятия пищи). Человек «социальный» представлен, прежде всего, в отношениях родства. Значимым является также его отношение к трудовой деятельности (высокая

частотность существительного *работа* и глаголов, связанных с ведением сельского хозяйства). Среди нравственных ориентиров выделяются религиозные и высшая ценность – *правда*. Преобладающими действиями для диалектоносителей являются движение и речь. Картина мира формируется через чувственное восприятие. Наиболее важными параметрами оценки становятся размерные (*большой* – *маленький*) и субъективные (*хороший* – *плохой*). Главная субстанция бытия для носителей диалекта – время.

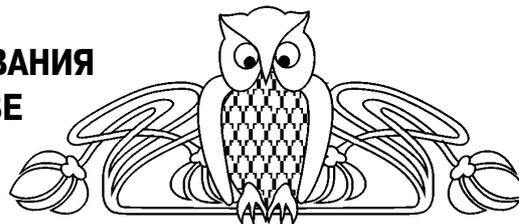
Таким образом, анализ частотной лексики позволяет пронаблюдать наиболее важные и актуальные для диалектоносителей понятия и концепты, выявить характерные особенности диалектной картины мира.

### Примечания

- <sup>1</sup> Гиндин С. Частота слова и его значимость в системе языка // Учен. зап. Тарт. ун-та. 1982. Вып. 658. С. 22.
- <sup>2</sup> Засорина Л. Частотный словарь русского языка. М., 1977. С. 930.
- <sup>3</sup> Ляшевская О., Шаров С. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М., 2009.

УДК [81'272:39] (571.12)

## О ПРОБЛЕМЕ ЭВОЛЮЦИИ ФУНКЦИОНИРОВАНИЯ ЯЗЫКОВ В ПОЛИЭТНИЧНОМ ПРОСТРАНСТВЕ ТЮМЕНСКОЙ ОБЛАСТИ: К ПОСТАНОВКЕ ПРОБЛЕМЫ



И. С. Карабулатова, К. К. Койше\*

Тюменский государственный университет

\*Заводоуковский филиал Тюменского государственного университета

E-mail: radogost2000@mail.ru, koishe.k.k@mail.ru

Авторы рассматривают проблему функционирования языков в контексте эволюционного развития языковой личности. Такой подход позволяет понять эволюцию формирования современной языковой личности в условиях полиэтнического характера глобализации.

**Ключевые слова:** языковая личность, функционирование языков, эволюция, полиэтническое пространство.

### The Issue of the Evolution of Languages in a Multiethnic Area of the Tyumen Region: to the Statement of the Problem

I. S. Karabulatova, K. K. Koyshe

The authors consider the problem of languages' functioning in the context of the evolution of linguistic personality. This approach allows us to understand the evolution of the development of modern linguistic personality in a multi-ethnic nature of globalization.

**Key words:** linguistic personality, languages' functioning, evolution, multi-ethnic area.

Тюменская область уже длительное время представляет интерес для исследователей как поликультурное, полиэтническое и полиязыковое образование со сложным устройством, в котором важную роль играют диаспоры, поэтому вопрос глубокого анализа эволюции языков и государственного регулирования их функционирования является чрезвычайно актуальным. Разнообразие этноязыкового ландшафта Тюменского региона обусловлено, с одной стороны, высоким уровнем представленности языков как России, так и постсоветского пространства, а с другой – тем, что в региональном тюркском пространстве доминируют татарский, азербайджанский и казахский языки, что обусловлено как исторически, так и современными экстралингвистическими факторами. Вопрос о смешении языков в процессе влияния одного языка на другой ставился еще в XI в. Так, М. Кашгари в своем словаре «Дивану