

УДК 81-112.2

Doi 10.26456/vtfilol/2024.2.117

ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ ЛЕКСИКИ НА ПРИМЕРЕ ПОВЕСТВОВАТЕЛЬНЫХ ТЕКСТОВ

С.А. Ермаков

Воронежский государственный лесотехнический университет
им. Г.Ф. Морозова, г. Воронеж

В статье описывается функциональный анализ лексики на примере словников из произведений современных российских писателей. Функциональный параметр является составной частью параметрического анализа лексики (ПАЛ), дающего представление о месте слова в иерархии лексико-семантической системы языка.

Ключевые слова: историческая лексикология, параметрический анализ лексики, периферия, лемма, словоформа.

Данная статья является первой в серии публикаций, посвященных описанию параметрического анализа лексики (ПАЛ) на примере русской нарративной прозы. Выбор повествовательных текстов как объектов исследования обусловлен их исторической универсальностью.

Исследование является составляющей частью эксперимента по системной стратификации лексики словника, представляющего собой часть более крупного изыскания, посвященного изучению лексико-семантической системы русского языка.

Цель данной работы – в рамках параметрического анализа лексики (ПАЛ) провести функциональный анализ лексики словника.

Для достижения поставленной цели необходимо выполнить следующие **задачи**:

- создать электронный конкорданс словника и провести его лемматизацию;
- выделить МинТеМ (минимально текстуально-маркированную) лексику словника;
- определить функциональный вес каждого слова.

В качестве **объекта** исследования выбраны фрагменты следующих текстов:

- повествовательный текст (1266 слов) из романа-житие «Лавр» [2], опубликованный в 2012 году, автор произведения – писатель и литературовед, доктор филологических наук Евгений Германович Водолазкин;

- повествовательный текст (1339 слов) из исторического романа «Обитель» [5], опубликованный в 2014 году, автор произведения – писатель, филолог, публицист, общественный и политический деятель Евгений Николаевич (Захар) Прилепин;

- повествовательный текст (1446 слов) из романа «Каменный мост» [6], опубликованный в 2009 году, автор произведения – писатель, журналист, редактор и киносценарист Александр Михайлович Терехов.

При составлении конкорданса главной задачей является «отображение всех зарегистрированных в некоторых текстах случаев употребления слова» [3: 85]. Конкорданс текста создается при помощи программ PrProcess и Sentences, затем загружается в электронные таблицы Microsoft Excel и после обработки макросом представляет собой организованный перечень всех словоформ словника. Для каждой словоформы указываются все контексты, демонстрирующие ее употребление в тексте.

Далее проводится лемматизация – преобразование всех слов словника в их словарную форму (существительные – именительный падеж, единственное число; прилагательные – именительный падеж, единственное число, мужской род; глаголы, причастия, деепричастия – глагол в инфинитиве).

В результате предложение «*В дедовой избе Арсений любил запах*» [2: 10] в конкордансе текста будет иметь следующий вид, например:

Таблица 1

Словофор	Лемма	Слово	Слово	Слово	Слово	Слово	Слово
		1	2	3	4	5	6
В	<i>В</i>	<i>В</i>	<i>дедово</i>	<i>избе</i>	<i>Арсени</i>	<i>любил</i>	<i>запах.</i>
			<i>й</i>		<i>й</i>		
<i>дедовой</i>	<i>ДЕДОВ</i>	<i>В</i>	<i>В</i>	<i>избе</i>	<i>Арсени</i>	<i>любил</i>	<i>запах.</i>
					<i>й</i>		
<i>избе</i>	<i>ИЗБА</i>	<i>В</i>	<i>дедово</i>	<i>В</i>	<i>Арсени</i>	<i>любил</i>	<i>запах.</i>
			<i>й</i>		<i>й</i>		
<i>Арсений</i>	<i>АРСЕНИ</i>	<i>В</i>	<i>дедово</i>	<i>избе</i>	<i>В</i>	<i>любил</i>	<i>запах.</i>
	<i>Й</i>		<i>й</i>				
<i>любил</i>	<i>ЛЮБИТ</i>	<i>В</i>	<i>дедово</i>	<i>избе</i>	<i>Арсени</i>	<i>В</i>	<i>запах.</i>
	<i>Ь</i>		<i>й</i>		<i>й</i>		
<i>запах.</i>	<i>ЗАПАХ</i>	<i>В</i>	<i>дедово</i>	<i>избе</i>	<i>Арсени</i>	<i>любил</i>	<i>В</i>
			<i>й</i>		<i>й</i>		

Конкорданс текста является чрезвычайно полезным и важным инструментом, который востребован на всех этапах практической работы.

Выделение МинТеМ (минимально текстуально-маркированной) лексики проводится с использованием инструментария маркемологии (см. Артемова [1], Кретов [4], Фаустов [8]).

В своей работе мы ориентируемся на Индекс Текстуальной Маркированности слова (ИнТеМ), вычисляемый как разность между частотным весом словоформы (Ч-весом) в тексте и её весом по длине (Д-весом), который колеблется в интервале от +1 до -1. Слова с ИнТеМом близким к +1 связаны со спецификой текста, тогда как слова с ИнТеМом близким к -1 относятся к периферии текста и словаря, а вот слова с ИнТеМом близким к 0 относятся к ядру языка и являются наименее маркированными.

Таким образом, чтобы извлечь из текста МинТеМ лексику, необходимо исключить лексику с ИнТеМами, максимальными по модулю.

Границей, отделяющей МинТеМ-лексику от периферии логично считать среднее значение ИнТеМа по модулю (в нашем случае среднее значение: 0,58917 для словника «Лавр», 0,62738 для словника «Обитель» и 0,69642 для словника «Каменный мост»): то, что меньше этого значения, относится к МинТеМ-лексике, то, что больше – к периферии.

В исследовании рассматриваются только значимые части речи, такие как существительные, прилагательные и глаголы, соответственно, из МинТеМ-лексики удаляются:

- служебные слова, наречия, местоимения, числительные, имена собственные, например: В, ЕСЛИ, ДВА, ТАМ, ХРИСТОФОР и т.д.;
- текстуально маркированная лексика, например: ПЕСЬЕГОЛОВЕЦ, КИТОВРАС, ПОСТРЕЛ и т.д.;
- все леммы с Частотой 1 (и Ч-весом 0).

В результате проведенной работы была отобрана МинТеМ-лексика в составе: словник «Лавр» – 91 слово, словник «Обитель» – 70 слов и словник «Каменный мост» – 88 слов.

Функциональный вес каждого слова рассчитывался по *средней длине его словоформ*.

Суммировав частоты всех словоформ слова, мы получили *частоту леммы*. Например, в исследуемом словнике словоформы слова ПЕРО (ПЕРЬЯ, ПЕРЬЕВ, ПЕРЬЯМИ и т.д.) встречаются 6 раз, соответственно частота леммы ПЕРО равна 6.

Для корректного вычисления длины леммы следует провести практическую транскрипцию словоформ, цель которой одна: число символов в транскрипции должно быть равно числу звуков в произношении. Правила замены указаны в Таблице 1. Например, если имеется слово СОЛОВЬИ, то в транскрипции оно будет выглядеть как СОЛОВЬИИ, ЯЗЫК превратится в ЙАЗЫК, УЧИТЬ будет УЧИТ, а ЕСЛИ станет ЙЭСЛИ.

Таблица 2. Транскрипция звуков

Буквы	ь/ье	ь/ё	ь/и	ь/ю	ь/я	#/Ve	#/Vё	#/Vю	#/Vя	ь/ь
Звуки	йэ	йо	йи	йу	йа	йэ	йо	йу	йа	Ø

V – любой гласный, # – начало слова, Ø – удаление (последняя операция).

Суммирование произведений частоты всех словоформ слова на их длину даёт *суммарную длину леммы*. Например, слово РУЧЕЙ представлено в словнике двумя словоформами РУЧЬЮВ (6 звуков) и РУЧЬАМИ (7 звуков), суммарная длина которых равна 13 звукам.

Округляя до целого числа частное при делении суммарной длины леммы на её частоту, мы получили *среднюю длину словоформы слова*. Например, слово ЛЕС представлено в словнике 3 раза словоформой ЛЕСУ (4 звука*3=12) и 1 раз словоформой ЛЕСАМ (5 звуков). Суммарная длина этих словоформ составляет 17 звуков. При делении суммарной длины леммы (17) на её частоту (4) и округлении результата (4,25) до целого числа, получаем среднюю длину словоформы слова, представленного леммой ЛЕС 4 звука.

Динамика распределения слов по средней длине словоформы в словниках «Лавр», «Обитель» и «Каменный мост» представлена на Рис. 1:

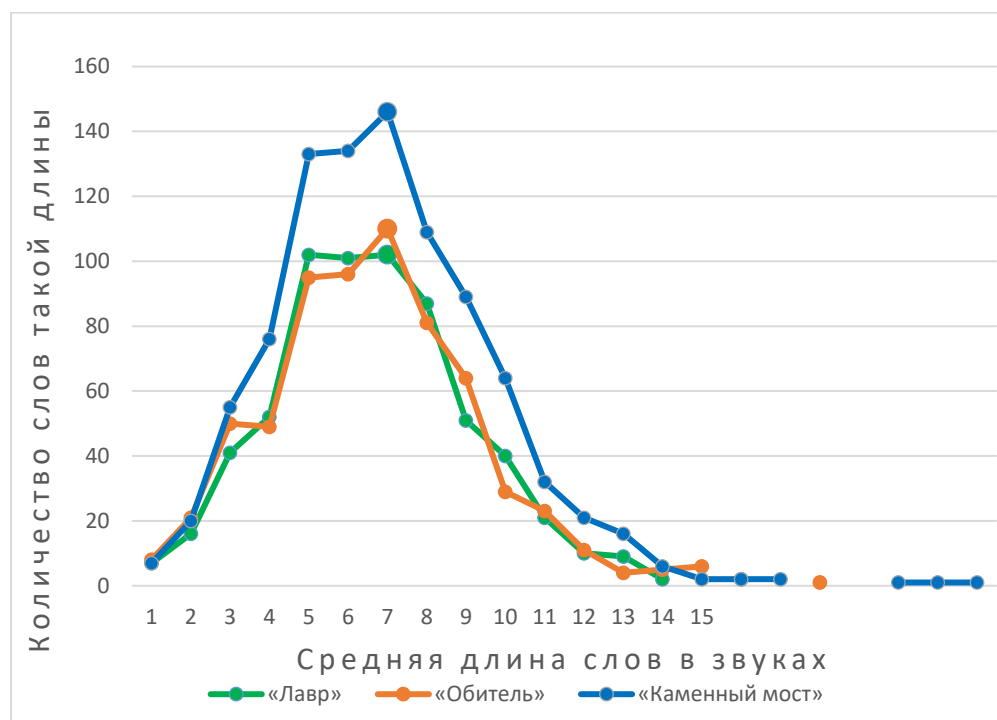


Рис. 1. Распределение слов по длине

График показывает, что наиболее часто встречаются слова со средней длиной словоформы от 5 до 7 звуков («Лавр»: 5 – 102, 6 – 101, 7 – 102; «Обитель»: 5 – 95, 6 – 96, 7 – 110; «Каменный мост»: 5 – 133, 6 – 134, 7 – 146).

Наличие плато в распределении свидетельствует о его качественной неоднородности: о соединении в нем двух качественно

различных распределений. Можно предположить, что это распределение диалогической речи персонажей с модой 5 и монологической речи автора – с модой 7. В этом отношении выразительно наличие в тексте «Лавр» двух мод: 5 звуков – 102 случая и 7 звуков – 102 случая.

По формуле, предложенной В. Т. Титовым [7]: $ВЕС = (\Sigma r - R1-i) / \Sigma r$, (где Σr – сумма единиц всех рангов, $R1-i$ – сумма единиц от первого до данного), был определен вес каждой словоформы МинТеМ лексики словника. Результаты вычислений представлены ниже:

Таблица 3. Функциональные веса словников текстов «Лавр», «Обитель», «Каменный мост»

Текст: Длина-ср. сл/формы	«Лавр»			«Обитель»			«Каменный мост»		
	Слов	Накопл.	Ф- вес	Слов	Накопл.	Ф- вес	Слов	Накопл.	Ф- вес
1	7	7	0,9891	8	8	0,9878	7	7	0,9924
2	16	23	0,9642	21	29	0,9556	20	27	0,9706
3	41	64	0,9005	50	79	0,8790	55	82	0,9106
4	52	116	0,8196	49	128	0,8040	76	158	0,8277
5	102	218	0,6610	95	223	0,6585	133	291	0,6827
6	101	319	0,5039	96	319	0,5115	134	425	0,5365
7	102	421	0,3453	110	429	0,3430	146	571	0,3773
8	87	508	0,2100	81	510	0,2190	109	680	0,2585
9	51	559	0,1306	64	574	0,1210	89	769	0,1614
10	40	599	0,0684	29	603	0,0766	64	833	0,0916
11	21	620	0,0358	23	626	0,0414	32	865	0,0567
12	10	630	0,0202	11	637	0,0245	21	886	0,0338
13	9	639	0,0062	4	641	0,0184	16	902	0,0164
14	2	641	0,0031	5	646	0,0107	6	908	0,0098
15				6	652	0,0015	2	910	0,0076
16	2	643	0				2	912	0,0055
17							2	914	0,0033
18				1	653	0			
19									
20							1	915	0,0022
21							1	916	0,0011
22							1	917	0

На основании проведенного исследования можно отметить некоторые закономерности и сделать определенные выводы:

- слова со средней длиной от 5 до 7 звуков встречаются наиболее часто, что свидетельствует об оптимальности их длины для коммуникации и при передаче смысла;

- чем короче слово, тем оно имеет больший функциональный вес в системе языка;

- методы и приемы, используемые при определении функционального веса лексики словника, являются действенным инструментом и могут быть применены в дальнейшей работе, направленной на исследование лексико-семантической системы русского языка.

Список литературы

1. Артемова О. Г. Языковые ключи к английской литературе от Шекспира до Фаулза / Научн. ред. А. А. Кретов. Воронеж: Наука-Юнипресс, 2020. (Серия: Библиотека маркемологии. Т. 4). 596 с.
2. Водолазкин Е. Г. Лавр / Е. Г. Водолазкин - М. : «Издательство АСТ», 2012. - 448 с.
3. Доминикан А. И. Конкорданс как инструмент для изучения значения слова / Вестник ТвГУ. Серия «Филология». 2016. № 4. С. 85–89
4. Кретов А. А. Метод формального выделения тематически нейтральной лексики (на примере старославянских текстов) / Вестник Воронежского государственного университета. Серия: лингвистика и межкультурная коммуникация. 2007, № 1. С. 81–90.
5. Прилепин З. Обитель / З. Прилепин. – М : Редакция Елены Шубиной : АСТ, 2014. 745 с.
6. Терехов А. М. Каменный мост. / А. М. Терехов. – М : Издательство «АСТ» / «Редакция Елены Шубиной», 2021. 832 с.
7. Титов В. Т. Частная количественная лексикология романских языков / В. Т. Титов. – Воронеж : Изд-во Воронежского гос. ун-та, 2004. 551 с.
8. Фаустов А. А., Кретов А. А. Понятие маркемы и предварительные итоги маркемного анализа русской литературы / Вестник Воронежского государственного университета. Серия: лингвистика и межкультурная коммуникация. 2017, № 4. С. 16–31.

Об авторе:

ЕРМАКОВ Сергей Алексеевич – кандидат филологических наук, доцент кафедры иностранных языков, Воронежский государственный лесотехнический университет им. Г.Ф. Морозова (394087, Воронежская область, г. Воронеж, улица Тимирязева, дом 8), e-mail: sergey-ermakov.1811@mail.ru

FUNCTIONAL ANALYSIS OF VOCABULARY ON THE EXAMPLE OF NARRATIVE TEXTS

S.A. Ermakov

Voronezh State University of Forestry and Technologies
Named after G.F. Morozov, Voronezh

This article describes the functional analysis of vocabulary on the example of texts from the works of modern Russian writers. The functional parameter is an integral part of the parametric analysis of vocabulary (PAL), which gives an idea of the place of a word in the hierarchy of the lexico-semantic system of the language.

Keywords: *historical lexicology, parametric analysis of vocabulary, periphery, lemma, word form.*

About the Author:

ERMAKOV Sergey Alekseevich – Candidate of Philology, Associate Professor of the Foreign Languages Department, Voronezh State University of Forestry and Technologies Named after G.F. Morozov (Ulitsa Timiryazeva, 8, Voronezh, Voronezh Oblast, 394036), e-mail: sergey-ermakov.1811@mail.ru

© Ермаков С.А., 2024

Статья поступила в редакцию 15.01.24
Подписана в печать 20.02.24