

УДК 81'322.2

## КВАНТИТАТИВНЫЙ АНАЛИЗ СЛОВАРЯ ПРОИЗВЕДЕНИЙ В. ОВЕЧКИНА

*Ю.В. Веревкина*

*Соискатель ученой степени кандидата филологических наук,  
кафедра русского языка  
e-mail: [zolotosony@yandex.ru](mailto:zolotosony@yandex.ru)*

*Курский государственный университет*

*В статье рассматривается опыт лексикографирования произведений В. Овечкина, составление индивидуального авторского словаря, приводится количественный и качественный анализ лексических единиц авторского идиолекта.*

***Ключевые слова:** лексикография, квантитативный анализ, коэффициент лексического разнообразия, частотность лексических единиц.*

«Авторский словарь, аккумулирующий всю лексику произведений определенного автора или заданные составителем ее слои, выступает как репрезентант языка данного индивидуума, являясь одновременно и продуктом исследовательского подхода, и инструментом анализа» [Караулов 2003: 4].

Существуют разные теоретические принципы создания словарей писателей, идеи и словарные проекты, способы лексикографического отражения особенностей индивидуальных авторских стилей, позволяющие воссоздать лексикографические портреты мастеров художественного слова.

Мы представляем первый опыт лексикографирования произведений В.В. Овечкина, жизнь и творчество которого связаны с Курским краем. Исследование выполнено в русле научного проекта кафедры русского языка КГУ «Курское слово». В статье рассмотрены количественные параметры составленного нами словника разножанровых произведений писателя: повести «С фронтовым приветом», 18 рассказов, 5 очерков. Наша цель – выявление своеобразия авторского лексикона В.В. Овечкина на фоне лексикографических продуктов, полученных другими исследователями в ходе изучения языка курских писателей [Засорина 1977, Ашчихо 2009, Кризская 2009, Разумова 2007].

Принятый порядок подачи материала дает наглядную картину лексического разнообразия, позволяет составить представление о полноте лексикографического описания творчества автора.

Богатый опыт русской авторской лексикографии служит основой для нашего исследования, для изучения вопросов теории авторской лексикографии, составления словаря идиолекта писателя.

Авторские словари могут решать «синхронические задачи изучения особенностей авторского языка на фоне литературного языка эпохи или задачи сопоставительного анализа двух и более идиолектов авторов-современников; это может быть диахронически ориентированное исследование, позволяющее проследить эволюционные сдвиги и преобразования в жизни национального литературного языка на материале произведений авторов, принадлежащих разным историческим эпохам» [Караулов 2003: 4].

В решении всех задач, которые стоят перед исследователем языка творческой личности, основная нагрузка принадлежит словарю, составленному по текстам этого автора.

Главной образующей авторских словарей является «языковая единица или иной показатель языковой структуры, который, образуя вход в словарь или зону словарной статьи, выступает в функции лексикографического параметра» [там же: 6].

Не менее важным является понятие «языковой личности», под которой имеется в виду, «во-первых, любой носитель языка, охарактеризованный на основе анализа произведенных им текстов с точки зрения использования в этих текстах системных средств данного языка для отражения видения им окружающей действительности (картины мира) и для достижения определенных целей в этом мире; во-вторых, языковая личность – это инструмент лингвистического анализа, комплекс приемов, позволяющий описать языковую способность индивида и соединить системное представление его языка с функциональным анализом произведенных им текстов» (цит. по: [Никандрова 2010: 15–16]). Дело в том, что к автору изучаемых текстов применимы характеристики, которые используются при изучении среднего носителя языка.

Описанная в произведениях В. Овечкина эпоха ушла в прошлое, унеся с собой целый культурно-исторический пласт жизни нашего народа. Все, что есть в его произведениях, было частью жизни страны, нескольких поколений. Созданный нами словник представляет собой попытку зафиксировать лексический образ времени, раскрыть специфику культурно-исторического феномена – языка советской эпохи.

«Термин «квантитативная лингвистика» характеризует междисциплинарное направление в прикладных исследованиях, в котором в качестве основного инструмента изучения языка и речи используются количественные или статистические методы анализа» [Баранов <http>]. Квантитативный анализ подразумевает использование подсчетов и измерений при изучении языка писателя. Основой данного метода является математическая статистика, поэтому его можно назвать статистическим методом. В лингвистике он применяется для характеристики единиц любого уровня. Метод квантитативного анализа используется для описания и классификации текстов, позволяет

соотносить сегменты текстов как объекты подсчета с единицами языка, что немаловажно при изучении писательского лексикона.

Математические статистические методы помогут нам в «распределении конкретных языковых элементов по разрядам относительной частоты их употребления» [Звегинцев 1962: 137]. Данный подход к изучению лексикона позволяет «количественно описывать поведение лексем в тексте, частоту употребления единиц, их распределение в текстах разного жанра, сочетаемость с другими единицами и т.п.» [Бембеев 2012: 163].

Концепция составления словника нацелена на создание «бумажной» версии частотного словника и его электронного варианта. Программа «NewSlov» позволяет проводить изучение обширных массивов текстов, дает возможность доступа к текстовым данным, допуская многократное к ним обращение, служит залогом описания особенностей функционирования лексем в условиях контекста.

Квантитативный анализ предполагает определение общего количества лексем и словоупотреблений, определение средней частоты слова, коэффициента лексического разнообразия, что позволяет дать наиболее полную характеристику писательского лексикона.

Словник идиолекта В. Овечкина, составленный на материале трехтомного собрания сочинений, включает 22285 лексем, зафиксированных в 328545 словоупотреблениях. Полученные данные дают возможность определить коэффициент лексического разнообразия словаря писателя (С) и среднюю частоту слова (F ср.). Величина С зависит от общего объема словника и определяется отношением общего количества лексем словника (V) к количеству словоупотреблений (N):

$$C = \frac{V}{N} = \frac{22285}{328545} = 0,07$$

Обратной коэффициенту лексического значения слова является величина, обозначающая лексическое разнообразие:

$$F_{ср.} = \frac{N}{V} = \frac{328545}{22285} = 14,74$$

Итак, коэффициент лексического разнообразия словаря В. Овечкина равен 0,07, средняя частота слова равна 14,74.

Сопоставим полученные показатели с результатами лексикографирования творчества других курских писателей и общезыковыми данными (Таблица 1).

Таблица 1.

	Частотный словарь под ред. Л.Н. Засориной	Словник прозы А.П. Гайдара	Словник художественной прозы К.Д. Воробьева	Словник публицистики Е.И. Носова	Словник лексикона В.В. Овечкина
Фср.	26,9	18,1	13,6	5,7	14,74

Показатели средней частоты слова в словниках разнятся, что обусловлено разными объемами выборки. Близкими являются только показатели «Словника художественной прозы К.Д. Воробьева» и «Словника лексикона В.В. Овечкина» (13,6 и 14,74); это объясняется сравнительно равными объемами выборки, смежной тематикой поля их творчества.

Нами составлены более частные словники в зависимости от жанровой природы произведений В.В. Овечкина. Все словники различны по объему слов и количеству словоупотреблений. Приведем полученные количественные данные в виде таблицы (Таблица 2).

Таблица 2.

	Рассказы	Очерки	Повесть	Пьесы	Статьи
Кол-во слов	12358	11560	6717	6291	5794
Кол-во с/у	94386	111782	38972	50095	33417

Сопоставим коэффициенты лексического разнообразия словаря писателя (С) и среднюю частоту слова (F ср.) всех словников (Таблица 3).

Таблица 3.

	Рассказы	Очерки	Повесть	Пьесы	Статьи
Коэффициент лексического разнообразия словаря писателя (С)	0,13	0,10	0,17	0,13	0,17
Средняя частота слова (F ср.)	7,64	9,65	5,80	7,96	5,77

Мы получили количественные показатели, позволяющие выявить качественные особенности авторского словаря в разножанровых произведениях, динамические особенности лексикона В.В. Овечкина.

Анализируя количественные показатели лексического разнообразия и частоты слова, делаем следующие выводы:

1. Наиболее высокий коэффициент лексического разнообразия и наименьшая повторяемость слов отмечены в повести и статьях писателя. Причем количество слов и словоупотреблений словника повести и общего

словника статей приблизительно равны. Причиной является публицистическая направленность произведений.

2. Средний показатель в рассказах и пьесах (количество слов и словоупотреблений в рассказах в 2 раза больше). Объясняется это единой темой колхозного строительства.

3. Наименьшие показатели лексического разнообразия получены в ходе анализа очерков. Возможно, это связано с тем, что все 5 очерков имеют единую тему, объединены местом действия и общими героями.

Таким образом, тенденция увеличения коэффициента лексического разнообразия словаря писателя и уменьшение повторяемости лексических единиц прослеживается в зависимости от тематики произведений и аудитории, на которую они ориентированы.

Задача частотного словаря – ранжировать слова по их частоте и определить лексическое ядро писательского идиолекта. Определение средней частоты слова и коэффициента лексического разнообразия позволяет распределить лексемы по нескольким группам.

1. Нарах legomena (слова с единичным словоупотреблением).
2. Низкочастотная лексика (слова с частотой употребления от 2 до 3).
3. Частотная лексика (слова с частотой употребления от 4 до 50).
4. Высокочастотная лексика (слова с частотой употребления от 51 до 9871 – наибольшего показателя частотности).

Эти параметры одинаковы для всех составленных нами словарей, потому что подобные закономерности прослеживаются в каждом из них.

Представим полученные показатели и их процентное соотношение в виде таблицы (Таблица 4).

Таблица 4.1. Показатели и их процентное соотношение в общем словнике:

Степень частотности слов	Количество лексем	% от объема словника	Количество словоупотреблений	% от объема выборки
Слова с частотой 1	9345	41,93%	9345	2,84%
Низкочастотная лексика	5559	24,95%	13063	3,98%
Частотная лексика	6569	29,48%	78583	23,92%
Высокочастотная лексика	812	3,64%	227554	69,26%

Таблица 4.2. Показатели и их процентное соотношение в словнике рассказов:

Степень частотности слов	Количество лексем	% от объема словника	Количество словоупотреблений	% от объема выборки
Слова с частотой 1	6132	49,62%	6132	6,50%
Низкочастотная лексика	3075	24,88%	7169	7,60%
Частотная лексика	2918	23,61%	31059	32,90%
Высокочастотная лексика	233	1,89%	50026	53,00%

Таблица 4.3. Показатели и их процентное соотношение  
в словнике очерков:

Степень частотности слов	Количество лексем	% от объема словника	Количество словоупотреблений	% от объема выборки
Слова с частотой 1	5494	47,53%	5494	4,91%
Низкочастотная лексика	2888	24,98%	6752	6,04%
Частотная лексика	2889	24,99%	32089	28,71%
Высокочастотная лексика	289	2,5%	67447	60,34%

Таблица 4.4. Показатели и их процентное соотношение  
в словнике повести:

Степень частотности слов	Количество лексем	% от объема словника	Количество словоупотреблений	% от объема выборки
Слова с частотой 1	3601	53,61%	3601	9,24%
Низкочастотная лексика	1598	23,79%	3706	9,51%
Частотная лексика	1409	20,98%	14225	36,50%
Высокочастотная лексика	109	1,62%	17440	44,75%

Таблица 4.5. Показатели и их процентное соотношение  
в словнике пьес:

Степень частотности слов	Количество лексем	% от объема словника	Количество словоупотреблений	% от объема выборки
Слова с частотой 1	3208	50,99%	3208	6,41%
Низкочастотная лексика	1493	23,73%	3503	6,99%
Частотная лексика	1433	22,78%	16336	32,61%
Высокочастотная лексика	157	2,50%	27048	53,99%

Таблица 4.6. Показатели и их процентное соотношение  
в словнике статей:

Степень частотности слов	Количество лексем	% от объема словника	Количество словоупотреблений	% от объема выборки
Слова с частотой 1	3140	54,19%	3140	9,40%
Низкочастотная лексика	1395	24,08%	3219	9,63%
Частотная лексика	1168	20,16%	12267	36,71%
Высокочастотная лексика	91	1,57%	14791	44,26%

Наименьшая частота слова в любом словнике – 1 словоупотребление. Редкие слова – это один из основных показателей богатства и разнообразия писательского лексикона, поэтому им следует уделить особое внимание.

Представим в виде таблицы общую сравнительную характеристику слов с частотой 1 в рассматриваемых словарях (Таблица 5):

Таблица 5.

Степень частотности слов	Частотный словарь под ред. Л.Н. Засориной	Словник прозы А.П. Гайдара	Словник художественной прозы К.Д. Воробьева	Словник публицистики Е.И. Носова	Словник лексикона В.В. Овечкина
% от объема словника	34,1%	31,6%	41,4%	50,5%	41,93%

В словнике В. Овечкина слова *hарах legomena* составляют чуть меньше половины всей выборки. Схожие показатели видим в словнике К.Д. Воробьева, немногим больше этот процент в словнике Е.И. Носова. Подобная разница мотивирована, на наш взгляд, разницей в тематике, которую затрагивали писатели в своих произведениях: чем более привержен автор к публицистичности, военной тематике, тем меньше процент слов с единичным словоупотреблением (Таблица 6).

Таблица 6.

Степень частотности слов	Общий словарь	Словник рассказов	Словник очерков	Словник повести	Словник пьес	Словник статей
% от объема словника	41,93%	49,62%	47,53%	53,61%	50,99%	54,19%

Из представленных данных видно, что процент слов *hарах legomena* в отдельных словниках колеблется в пределах от 47,53% до 54,19%, что составляет практически половину или немногим больше половины выборки каждого из них, что зависит от жанровой и тематической принадлежности (рассказы, очерки и пьесы посвящены описанию устройства колхозной жизни, повесть полностью пронизана военной тематикой, а статьи обращены к писательской аудитории). В общем словнике слова с единичным словоупотреблением составляют чуть меньше половины всех слов – 41,93%, в то время как в общезыковом употреблении – 34,1%. В произведениях В. Овечкина данный показатель значительно выше установленной средней величины, что еще раз подчеркивает лексическое разнообразие авторского словаря.

Каждый текст, равно как и каждый словник, предполагает собственную тематику и лексическую базу, определенные лексемы которой употребляются намного чаще среднего. Частотные показатели, полученные на основе выборки из корпуса всех анализируемых текстов, отражают их специфику.

Приведём данные о процентном соотношении частотной лексики словников В. Овечкина разных жанров в виде сводной таблицы (Таблица 7).

Таблица 7.

Степень частотности слов	Общий словарь	Словник рассказов	Словник очерков	Словник повести	Словник пьес	Словник статей
Низкочастотная лексика	24,95%	24,88%	24,98%	23,79%	23,73%	24,08%
Частотная лексика	29,48%	23,61%	24,99%	20,98%	22,78%	20,16%
Высокочастотная лексика	3,64%	1,89%	2,5%	1,62%	2,50%	1,57%

Разница показателей низкочастотной, частотной и высокочастотной лексики во всех словарях незначительна; это позволяет говорить о некотором постоянстве автора при выборе лексических средств, не прослеживается резкой динамики, дающей право судить об определенных этапах развития идиолекта. Объясняется это жанровым делением при составлении отдельных словарей, а не хронологическим.

Сравним процентное соотношение частотной лексики лексикона В. Овечкина с аналогичными данными по творчеству других курских писателей (Таблица 8).

Таблица 8.

Степень частотности слов	Словник прозы А.П. Гайдара	Словник художественной прозы К.Д. Воробьева	Словник публицистики Е.И. Носова	Словник лексикона В.В. Овечкина
Низкочастотная лексика	48,2%	46,7%	34,9%	24,95%
Частотная лексика	18,6%	8,7%	10,7%	29,48%
Высокочастотная лексика	1,5%	3,2%	1,1%	3,64%

Подобная разница возникает по причине разных параметров определения границ между видами лексики. Так, например, в словнике лексикона Гайдара в эту категорию входят лексемы, имеющие частоту от 2 до 10 (следует отметить, что выделенную в нем среднечастотную (от 11 до 25) и высокочастотную (от 26 до 149) лексику в нашей таблице мы отнесли к частотной лексике; сверхвысокочастотную мы отнесли к высокочастотной: от 150 и выше); в словнике К.Д. Воробьева низкочастотная лексика от 2 до 13, частотная от 14 до 50, высокочастотная от 51 до и выше; в словнике Е.И. Носова низкочастотная от 2 до 5, частотная от 6 до 49, высокочастотная от 50 и более). В словнике В. Овечкина границы низкочастотной лексики определены на основе общей закономерности, прослеживающейся во всех блоках, от 2 до 3 словоупотреблений, так как существует некая разница в количестве слов



со словоупотреблением 3 и 4, поэтому порог низкочастотной лексики ниже, чем у других писателей. Блок частотной лексики, соответственно, больше, чем у других авторов за счет более широкого спектра слов (начиная со слов со словоупотреблением 4), входящих в эту категорию. Показатель высокочастотной лексики незначительно отличается от показателей остальных авторов, взятых для сравнения.

Таким образом, на материале трехтомного издания произведений В. Овечкина мы получили количественный анализ его лексикона, что позволяет выявить качественную индивидуальность авторского идиостиля, а также дать сравнительную характеристику его лексикона на фоне лексиконов других курских писателей. Значительный объем словника показывает богатство и разнообразие словаря В. Овечкина, что выражается в общем количестве слов и словоупотреблений, в которых они зафиксированы, а также в высоком показателе слов *hapaх legomen*, что говорит о высоком художественном уровне писателя как мастера слова.

### ***Библиографический список***

*Ашчихо И.В.* Количественный и качественный анализ словника А.П. Гайдара // Курское слово. Вып. 6. Курск: Курский гос. ун-т, 2009. 80 с.

*Баранов А.Н.* Введение в прикладную лингвистику: учебное пособие [Электронный ресурс]. М.: Эдиториал УРСС, 2001. 360 с. URL: <http://rykovodstvo.ru/exspl/35408/index.html> (дата обращения: 07.11.18).

*Бембеев Е.В.* Опыт квантитативной обработки текста на старокалмыцком языке: количественные характеристики // *Oriental studies* (Вестник Калмыцкого института гуманитарных исследований РАН). 2012. № 2. С. 163–167.

*Засорина Л.Н.* Частотный словарь русского языка. М.: Русский язык, 1977. 936 с.

*Звегинцев В.А.* Очерки по общему языкознанию. М.: Изд-во МГУ им. М.В. Ломоносова, 1962. 384 с.

*Караулов Ю.Н., Гинзбург Е.Л.* Опыт типологизации авторских словарей // *Русская авторская лексикография XIX–XX веков*. М.: Азбуковник, 2003. С. 4–16.

*Кризская Т.В.* Язык художественной прозы К.Д. Воробьева: дис. ... канд. филол. наук: 10.02.01. Курск, 2009. 192 с.

*Никандрова И.А.* Языковая личность персонажа: лингвистический аспект исследования // *Вестник Вятского государственного гуманитарного университета*. 2010. № 2 (2). С. 15–19.

*Овечкин В.В.* Собрание сочинений в 3-х томах. М.: Худож. лит., 1989. Т. 1–3.

*Разумова М.А.* Лексика публицистического дискурса Е.И. Носова:  
дис. ... канд. филол. наук: 10.02.01. Курск, 2007. 307 с.