

**ЧАСТОТНЫЙ АНАЛИЗ ХУДОЖЕСТВЕННОГО
ПРОИЗВЕДЕНИЯ НА ПРИМЕРЕ ПОВЕСТИ В. БЫКОВА
«ВОЛЧЬЯ ЯМА»**

**FREQUENCY ANALYSIS OF A FICTION ON THE EXAMPLE
OF V. BYKOV'S STORY «WOLF PIT»**

Ю.Ю. Красовская

J.J. Krasowskaja

Белорусский государственный университет,
Минск, Беларусь, *hisradaar@gmail.com*
Belarusian State University,
Minsk, Belarus, *hisradaar@gmail.com*

Статья посвящена частотному анализу повести Василя Быкова «Волчья яма». Анализ индивидуальной языковой системы автора – идиолекта – является ключом к языковой картине мира писателя. Компьютерные методы в лингвистике могут быть применимы для широкой области текстов, в том числе – для художественных.

Ключевые слова: частотный словарь; частотность; Василь Быков; белорусская литература.

This article is devoted to the frequency analysis of Vasil Bykov's story "Wolf Pit". Analysis of the individual language system of the author – an idiolect – is the key to the linguistic picture of the world. Thus, computer methods in linguistics can be applied to a wide range of texts, including fiction.

Keywords: frequency dictionary; frequencies; Vasil Bykov; belarusian literature.

Частотность – термин квантитативной лингвистики, который употребляется для обозначения наиболее часто употребляемых единиц. Частотному анализу могут быть подвергнуты элементы любого уровня языка, чаще всего – лексемы (слова). Однако существуют частотные словари морфем, словоформ, словосочетаний. Так, например, частотный словарь морфем поможет определить наиболее продуктивные и распространенные словообразовательные морфемы языка.

Частотность определенного слова вычисляется по следующей формуле:

$$Freq(x) = \frac{Q(x)}{Q(all)}, \quad (1)$$

где $Freq(x)$ – частотность слова « x », $Q(x)$ – количество его употреблений « x », а $Q(all)$ – общее количество слов. Частотность может быть представлена в процентах или числом $0 \leq Freq(x) \leq 1$.

Широкое применение в частотном анализе текстов нашел закон Ципфа-Мандельброта – частотная закономерность всякого коррелированного сообщения. Он имеет следующий вид:

$$p(i) = \frac{K}{(B+i)^\gamma}, \quad (2)$$

$$M = KZ - B, \quad (3)$$

$K, B, \gamma - \text{const}$: $K = \frac{1}{\ln F_1}$, $B = \frac{K}{p_1}$, $\gamma = 1$, p_i, p_1 – частоты, соответственно, i -ого и 1-ого по рангу слов, F_1 – число вхождений самого частого слова, M – теоретический объем словаря текста длиной Z .

Для согласования зоны низких частот с формулой Мандельброта (3) допускается, что число разных слов n_F , каждое из которых встретилось на выборке объемом Z единиц ровно F раз, равно:

$$n_F = \frac{M}{F(F+1)} \quad (4)$$

Таким образом, слов, встретившихся ровно один раз каждое ($F = 1$), должно быть $n_1 = \frac{M}{2}$, т.е. половина всего словарного запаса M выборки.

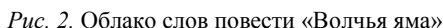
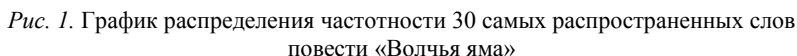
При частотном анализе произведения составляется частотный список, или частотный словарь. Частотные словари содержат статистические данные, широко используемые для решения различных языковых задач, таких как: определение динамичных средств словообразования и анализа языка, усовершенствование вопросов орфографии и графики, связанных с учётом статистических сведений о словарном составе и др.

Частотные словари могут издаваться не только в полном, но и в частичном виде – в таком случае публикуется часть словаря с наиболее употребительными лингвистическими единицами. В этих словарях наиболее частотные единицы представляет главную ценность, так как именно они покрывают значительную часть текста.

В данной работе материалом исследования была выбрана повесть Василя Быкова «Волчья яма». Оригинальный текст написан на белорусском языке, однако повесть была переведена на русский язык самим автором. Текст повести был взят из интернета из открытых источников – с сайта «Беларуская Палічка» [1]. Общий объём текста составил 18 536 словоупотреблений и 100 373 символов без учёта пробелов.

Процесс построения частотного словаря был автоматизирован при помощи программного обеспечения, написанного на языке Python.

При помощи библиотеки `matplotlib` можно визуализировать частоту распределения слов. Такой график для первых 30 самых распространенных слов будет иллюстрировать закон Ципфа (рисунок 1).



227

В повести «Волчья яма» 200 самых частотных слов встречаются 10 324 раз. То есть они покрывают 56% всего текста. Частотный словарь содержит 2 438 слов, которые упоминаются всего по одному разу – это составляет 58% всех записей в словаре. Такая статистика является яркой иллюстрацией закона Ципфа.

Глядя на полученные результаты анализа лексики, можно сделать следующие выводы:

1) Значительная доля самых частотных слов приходится на служебные части речи: местоимения, предлоги, частицы и союзы. Высокую частотность местоимений – *он, она, они, свой, это, сам* – можно объяснить большим количеством предложений со ссылками, где, например, персонаж не называется по имени, а вместо этого используется замещающее местоимение. В таких предложениях значение ссылки восстанавливается из контекста.

2) Ознакомившись со списком частотных самых частотных слов, можно сделать некоторые предположения о содержании повести для тех, кто с ней не знаком.

2.1) В повести «Волчья яма» практически не упоминаются имена собственные. Поэтому, чтобы сделать предположения о главных персонажах, стоит опираться на самые частотные одушевленные существительные: *солдат, дезертир, парень; бомж, рыбак; женщина*. Те одушевленные существительные, которые повторяются редко – *бабушка, старик* – относятся к второстепенным персонажам повести.

2.2) Глядя на самые частотные существительные, можно сделать несколько предположений о локациях, где разворачивается место действия повести:

– окрестности какого-либо водоёма; к этой местности можно отнести слова *река, речка, болото, берег*. Кроме того, это предположение подтверждается словами *лягушка, рыбак, крючок, вода*.

– сельская местность; в тексте повести достаточно часто упоминаются *хутор, деревня, поле*.

– лес; к этой категории можно отнести слова *бор, лес*, а также *дерево, сосна*.

Кроме того, можно выделить слова, которые не относятся ни к одной из перечисленных категорий, но также помогают построить догадки о месте действия повести: *зона, обрыв, край*.

2.3) Высокая частотность слова *зона* в сочетании с высокой частотностью слова *радиация* позволяет сделать предположение, что действие разворачивается неподалёку от т.н. зоны отчуждения. Такая догадка также немного ограничивает временные рамки событий произведения, перенося их во вторую половину 1980-х годов и позже.

2.4) Среди самых частотных глаголов оказался глагол-связка *быть*, модальные глаголы *мочь, хотеться, казаться*. Крайне высока частотность глаголов, обозначающих эмоциональное состояние или умственную деятельность: *чувствовать, знать, подумать, бояться*. Также распространены глаголы состояния: *ждать, лежать, сидеть, жить*. Частотность глаголов, обозначающих перемещение в пространстве – *прийти, поехать, идти* – умеренная. Глаголов, обозначающих активные действия, мало, они даже не попали в выборку самых частотных в облаке слов. Так, например, *застрелить, ударить, обрушить, перехватывать, защищать* упоминаются всего по одному разу. Можно сделать предположение, что темп повествования размеренный, каких-либо динамичных действий там мало или нет вовсе.

Таким образом, можно говорить о том, что частотный анализ произведения способен помочь сделать не только выводы о распределении частот в текстах, но и о приблизительном содержании анализируемого текста.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Быков, В. Волчья яма / В. Быков – Беларуская Палічка [Электронный ресурс]. – Режим доступа: https://knihi.com/Vasil_Bykau/Volcja_jama-rus.html – Дата доступа: 10.01.2023
2. Келер Р. Синергетическая лингвистика: структура и динамика лексики / Р. Келер [Электронный ресурс] – Режим доступа: <http://ubt.opus.hbz-nrw.de/volltexte/2007/413/pdf/synling.pdf> – Дата доступа: 10.01.2023
3. Маслов, В.П., Маслова Т.В. О законе Ципфа и ранговых распределениях в лингвистике и семиотике / В. П. Маслов, Т. В. Маслова // Матем. заметки, т. 80, выпуск 5. – 2006. – С. 718–732

ПРЕДПОСЫЛКИ ИССЛЕДОВАНИЯ ФАУНИСТИЧЕСКОЙ ЛЕКСИКИ

BACKGROUND TO THE STUDY OF FAUNISTIC VOCABULARY

Е.В. Левшикова

K. Leushykava

Белорусский государственный университет,
Минск, Беларусь, khamtsova@bsu.by
Belarusian State University,
Minsk, Belarus, khamtsova@bsu.by

В данной статье рассматривается лексическая группа «зоонимы». В ходе исследования данной лексико-семантической группы было установлено, что некоторые исследователи на следующей ступени абстракции группируют названия всех животных на основании семантического дифференциального признака «тип животного» в четыре родовидовые группы, противопоставленные по дифференциальному признаку: «Названия млекопитающих животных», «Названия птиц», «Названия рыб»,