

## РАЗДЕЛ V. ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ

2. Правая граница интервала для допустимого отклонения, равного  $\pm 2$  минуты определится из выражения:

$$l_2 = \frac{1}{\frac{DT_i=Lk}{Lk}} = \frac{Lk}{DT_i=Lk} = \frac{Lk}{4} \quad (17)$$

3. Правая граница интервала для допустимого отклонения, равного  $\pm 3$  минуты определится из выражения:

$$l_3 = \frac{9}{4 \frac{DT_i=Lk}{Lk}} = \frac{9Lk}{4DT_i=Lk} = \frac{9Lk}{16} \quad (18)$$

### Выводы

Рассмотренные подходы позволяют учесть изменившуюся транспортную ситуацию, связанную с условиями дорожного движения на участке маршрута. В случаях, когда допущено отставание от графика по причине сложных условий движения, возможности уменьшения величины отставания резко ограничены. Следовательно, величина отставания от расписания, которую можно устранить водителю транспортного средства зависит как от протяженности оставшейся части маршрута, так и от конкретных условий движения. Исходя из этого, в новых условиях дорожного движения допустимое отклонение должно формироваться исходя из следующего принципа: "водитель не должен считаться нарушителем расписания, если его отклоне-

ние определилось действием не контролируемых системой управления факторов, действие которых превышало возможности водителя по регулированию движения".

### СПИСОК ЛИТЕРАТУРЫ

1. Адрющенко, В.А. Теория систем автоматического управления. [Текст] / В.А. Адрющенко // Л.: Изд-во Ленинградского ун-та, 1990. — 251 с.
2. Афанасьев, Л.Л. Единая транспортная система и автомобильные перевозки. [Текст] / Л.Л.Афанасьев, Н.Б.Островский, С.М. Цукерберг - М.Транспорт, 1984. — 200 с.
3. Афанасьев, Л.Л. Автомобильные перевозки. [Текст] / Л.Л.Афанасьев, С.М. Цукерберг - М., Транспорт, 1973.
4. Официальный сайт Федеральной службы государственной статистики: [Электронный ресурс]: статистика. - Режим доступа: <http://www.gks.ru>. — Загл. с экрана.

*Директор по научному сопровождению проектов ЗАО "НПП Транснавигация" (г. Москва) к.т.н., **Богумил В.Н.** тел. (495) 783-54-85, [tm@transnavi.ru](mailto:tm@transnavi.ru); Докторант кафедры «Транспортная телематика» Московского автомобильно-дорожного государственного технического университета (МАДИ) к.т.н., доцент **Ефименко Д.Б.** [telematica@madi.ru](mailto:telematica@madi.ru)*

УДК 004.934:004.912:008:001.8

## ОПЫТ ИССЛЕДОВАНИЯ ЧАСТОТНЫХ ХАРАКТЕРИСТИК ХУДОЖЕСТВЕННЫХ ТЕКСТОВ: ЧАСТЬ I - ОПРЕДЕЛЕНИЕ АВТОРСТВА, ИСТОРИЧЕСКИХ И КУЛЬТУРНЫХ ОСОБЕННОСТЕЙ

О.В. Головань, А.В. Ишков

В статье описано использование оригинального программного комплекса (программы для ЭВМ LangFracDim и базы данных DB Language Fractal Dimension, взаимодействующих через SQL-сервер Fire Bird) для исследования частотных характеристик отдельных художественных текстов и их корпусов, установления количественных параметров функциональной связи между частотой и рангом слов в тексте и определение на ее основе авторства, исторических и культурных особенностей текстов.

**Ключевые слова:** математическая лингвистика, частота и ранг слова, закон Ципфа.

### Введение

Количественные методы исследования текстового материала часто используются в прикладных областях лингвистики и основываются на проведении различных подсчетов и косвенных измерениях единиц любого уровня. Поскольку количественные методы в лингвистике часто непосредственно опираются на математическую статистику, то часто их называют также статистическими методами

[1]. Основным объектом применения количественных методов является сам художественный текст, как объединенная смысловой связью последовательность знаковых единиц и его количественные характеристики, которые могут дать определенную информацию, как о самом тексте, так и его характеристиках, связанных с индивидуальными особенностями автора, стиля текста, вкладываемого в него смысла и информации, а также особен-

ностями самого языка, используемого для порождения текста, который, в свою очередь, претерпевает различные трансформации под воздействием объективных культурных и исторических процессов [2].

Сложность и многоаспектность рассмотренных отношений между текстом и влияющими на него факторами сильно усложняет задачу количественного исследования особенностей различного текстового материала, особенно, установления степени влияния отдельных факторов на статистические закономерности текста [3].

В настоящей статье описан один из вариантов количественного метода исследования художественного текстового материала, основанный на выявлении в нем определенных частотных закономерностей и измерении их характеристик.

#### Экспериментальная часть

Для реализации метода разработан программный комплекс, позволяющий производить анализ и разбор как отдельных текстов, так и их корпусов на отдельные слова, занесение их в базу данных (БД), извлечение из БД по определенным признакам, автоматическое пополнение БД при увеличении объема исследованного текстологического материала, а также позволяющий проводить определение частотных характеристик получающихся массивов [4].

В комплекс входят программы для ЭВМ «Фрактальная размерность языка (LangFracDim)» и специализированная база данных «БД Фрактальная размерность языка (DB Language Fractal Dimension)» [5, 6].

Исходными объектами для работы комплекса являются электронные текстовые файлы, слова из которых, после обработки программой, заносятся в базу данных БД (DB Language Fractal Dimension) через используемый нами SQL-сервер Fire Bird v. 1.0.

#### Результаты и их обсуждение

ПрЭВМ LangFracDim предназначена для определения частотных характеристик текста или языка и корреляций между ними. Определение производится путем составления частотного словаря корпуса либо отдельно взятого текста и определения ранга и частоты слов. Программа позволяет проводить группировку слов в БД по определенным признакам.

Определение корреляций между частотными характеристиками основывается на установлении зависимости между частотой и рангом слова по закону Ципфа [7], параметры

которой, после линейаризации, определяются методом наименьших квадратов (МНК):

$$C = k \cdot P^\alpha, \quad (1)$$

где  $C$  - частота встречаемости слов в тексте;  $k$  - коэффициент пропорциональности;  $P$  - ранг слов;  $\alpha$  - степень развитости и наполняемости текста различными лексическими единицами (для корпусов большинства современных текстов и «естественного» языка близка к  $-1$ ).

Специальная организация работы основного алгоритма программы и ее взаимодействия с БД позволяет ставить в соответствие каждому слову его уникальный номер (ID), связанный с темой и словарем, что позволяет избежать повтора одинаковых слов и ускорять извлечение слова из БД SQL-сервером, а связь ID с определенной темой или словарем не только ускоряет работу комплекса ПО - БД, но позволяет осуществлять программе «LangFracDim» оперативное построение таблиц и графиков зависимостей частоты от ранга для слов определенной темы, словаря или общего списка слов, находящихся в БД (рисунок 1).

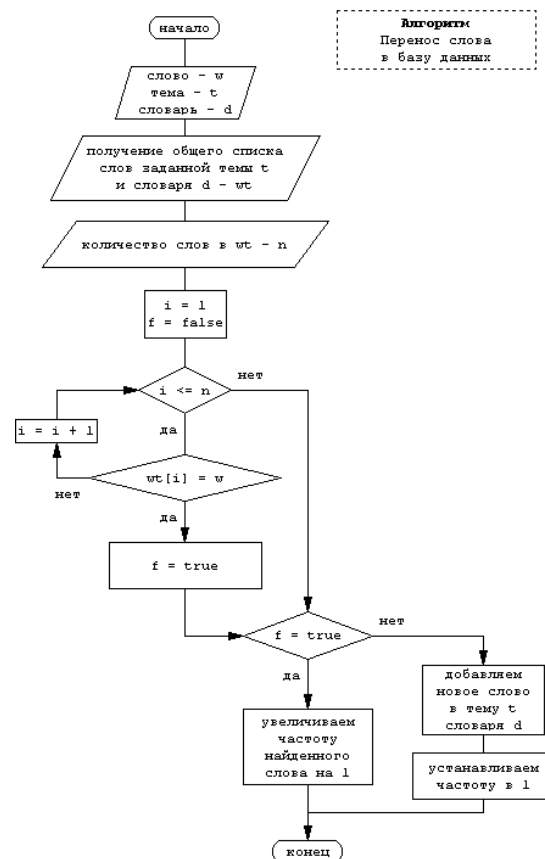


Рисунок 1 - Блок-схема основного алгоритма переноса слова в БД

Получаемая в результате работы программы LangFracDim функциональная зависимость частоты от ранга слова в тексте или корпусе - частотограмма, как оказалось, содержит в себе информацию, связанную с авторскими, историческими и культурными особенностями текста [8].

Например, на соответствующей логарифмической зависимости (1), полученной нами для рассказа В.М. Гаршина «Четыре дня», легко выявляются репрезентанты «трагического», которые могут быть найдены и в независимом когнитивном эксперименте [9] - они проявляются в разных «одночастотных» областях (площадках) частотограммы (рисунок 2).

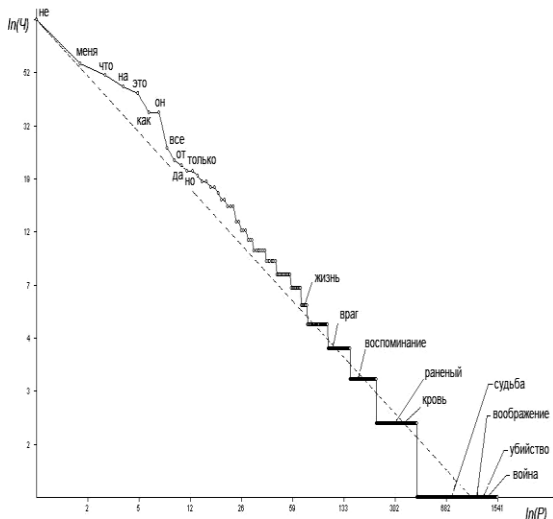


Рисунок 2 - Частотограмма рассказа В.М. Гаршина «Четыре дня»

Рассмотрим поведение функции (1) подробно. Высокие частоты и низкие ранги в тексте рассмотренного рассказа имеют различные служебные части речи, от частоты 86 (частица «не») вплоть до частоты 12 смыслозначимых лексем в тексте не обнаруживается. Первые смыслозначимые лексемы «быть» и «здесь» имеют частоту употребления в тексте рассказа 12.

В группе слов с одинаковой частотой 12 в рассказе «Четыре дня» находятся лексемы: «что-то», «него» и «быть».

В группе слов с частотой 5 находятся уже 29 лексем, в группе с частотой 4 – 43 лексемы, среди которых первый репрезентант «трагического» - «враг», в группе с частотой 3 – 76 лексем, среди которых еще один репрезентант – «воспоминание», с частотой 2 - 203, тут находятся репрезентанты «раненый» и «кровь», и, наконец, в группе с

частотой 1 находится уже 1111 лексем – 72,2 % всех слов текста, среди которых остальные репрезентанты «трагического», обнаруженные нами и в психолингвистическом эксперименте: «судьба», «воображение», «убийство» и «война».

Функциональная зависимость  $\chi=f(P)$  для рассказа «Четыре дня» в логарифмических координатах аппроксимируется прямой (пунктирная линия тренда на рис. 2), описываемой следующим уравнением:

$$\ln(\chi) = -0,6238 \times \ln(P) + 4,2906, \quad (2)$$

с  $\sigma(x)$  экспериментальных точек от теоретической кривой 0,0832 и коэффициентом линейной корреляции  $r = -0,9024$ , свидетельствующими о существовании сильной обратной связи между частотой и рангом слов в произведении.

Для использования параметров зависимости (1) при определении авторства текста нами были исследованы и другие произведения В.М. Гаршина: очерк «Очень коротенький роман» (1878); рассказ «Трус» (1879); рассказ «Из воспоминаний рядового Иванова» (1883) и автобиографический очерк «Аяслярское дело» [10].

Если теперь, расположить найденные нами коэффициенты пропорциональности и показатели степени в уравнении закона Ципфа (1) для четырех произведений в порядке возрастания общего количества лексем (N) в соответствующих текстах (813, 1538, 2018, 2525, 5225), получим следующие ряды: для  $k$  19,4531, 73,0102, 54,4132, 105,4045, 191,8485; для  $\alpha$  -0,4812, -0,6238, -0,5604, -0,6335, -0,6495 (рисунок 3).

Из рисунка 3 виден в целом согласованный характер изменения первой величины, и асимптотическое изменение показателя степени  $\alpha$  в законе Ципфа, который для рассмотренных нами произведений одного автора находится в интервале от -0,4812 до -0,6495, приближаясь к последнему значению.

### Выводы

1. Авторские особенности текстов количественно проявляются в показателе степени в уравнении закона Ципфа для тематической выборки произведений, что подтверждается предельным характером зависимости  $\alpha = f(N)$ .
2. Обнаруженные и зафиксированные в художественных текстах В.М. Гаршина составляющие культурно-языкового концепта «трагическое» определяют его функционирование в русской языковой культуре конца XIX в.

## ОПЫТ ИССЛЕДОВАНИЯ ЧАСТОТНЫХ ХАРАКТЕРИСТИК ХУДОЖЕСТВЕННЫХ ТЕКСТОВ: ЧАСТЬ II - РЕПРЕЗЕНТАЦИЯ ВРЕМЕНИ В РУССКОЙ И АНГЛИЙСКОЙ КУЛЬТУРЕ

3. Составление и исследование частотных словарей текстов и корпусов, построение соответствующих частотограмм и поиск особых точек функции  $Ч=f(P)$  позволяют устанавливать авторские, исторические и культурные особенности художественных текстов.

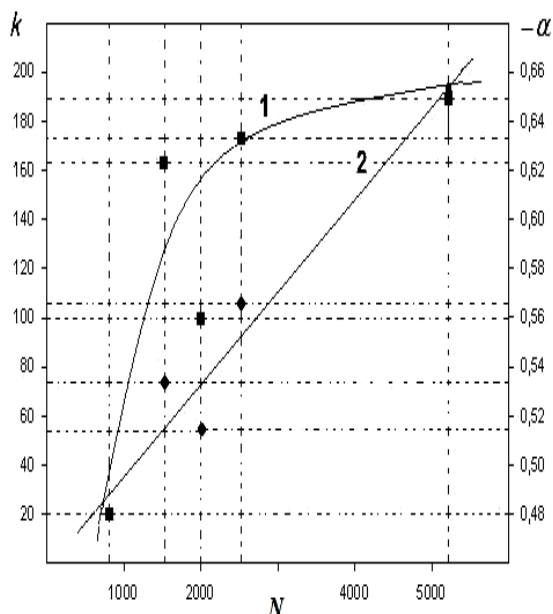


Рисунок 3 - Зависимость коэффициентов  $\alpha$  (1) и  $k$  (2) в уравнении закона Ципфа от количества лек-

сем для произведений В.М. Гаршина, репрезентирующих «трагическое»

### СПИСОК ЛИТЕРАТУРЫ

1. Андреев Н.Д. Статистико-комбинаторные методы в теоретическом и прикладном языковедении. [Текст] / Н.Д. Андреев / - СПб.: 1997.
2. Лотман Ю.М. Структура художественного текста. [Текст] / Ю.М. Лотман / - М., 1970.
3. Гришунин А. Л. Исследовательские аспекты текстологии. [Текст] / А.Л. Гришунин / - М., 1998.
4. Головань О.В. // Ползуновский альманах. 2008. № 2. С. 153-156.
5. Св-во № 2005610982 (RU) от 22.04.2005 / Головань О.В. // Оpubл. Бюлл. № 2. 2005.
6. Св-во № 2005620308 (RU) от 28.11.2005 / Головань О.В. // Оpubл. Бюлл. № 4. 2005.
7. Zipf G.K. The psycho-biology of language. Boston, 1935.
8. Вежбицкая А. Язык. Культура. Познание. [Текст] / А. Вежбицкая / - М.: 1997.
9. Головань О.В. Семантико-ассоциативная структура концепта «война». [Текст] / О.В. Головань / - Барнаул, 2001.
10. Гаршин В.М. Сочинения. [Текст] / В.М. Гаршин / - Москва, 1955.

К.ф.н., доцент, Головань О.В., e-mail: oleg6888@rambler.ru - Алтайский государственный технический университет им. И.И. Ползунова; д.т.н., профессор Ишков А.В., тел. (3852)-49-00-85 - Алтайский государственный аграрный университет (г. Барнаул).

УДК 004.934:004.912:008:001.8

## ОПЫТ ИССЛЕДОВАНИЯ ЧАСТОТНЫХ ХАРАКТЕРИСТИК ХУДОЖЕСТВЕННЫХ ТЕКСТОВ: ЧАСТЬ II - РЕПРЕЗЕНТАЦИЯ ВРЕМЕНИ В РУССКОЙ И АНГЛИЙСКОЙ КУЛЬТУРЕ

О.В. Головань, А.В. Ишков

Исследование и измерение времени в художественном произведении является одной из самых сложных задач математической лингвистики. В сообщении описана оригинальная методика поиска и структурно-функционального анализа лексем-репрезентантов времени в художественном тексте - «культурологический хронометр». Показано, что представление времени и его характеристики объективно различаются в русской и англоязычной культурах.

**Ключевые слова:** математическая лингвистика, частота и ранг слова, время, частотный спектр.

### Введение

Ранее нами сообщалось об успешной разработке оригинального способа количественного исследования текстового художественного материала, основанного на выявлении культурно- и смыслозначимых компонентов текстов и их корпусов при выявлении частотных закономерностей и установлении их количественных характеристик [1]. В первой части исследования, приведенной ранее

в этом номере, было показано использование этого способа для определения авторства, исторических и культурных особенностей художественного текста.

Целью настоящей работы являлось сравнительное исследование репрезентации времени в русской и английской культурах.

### Экспериментальная часть

Для исследования был использован программный комплекс, состоящий из про-