

**А.Б. Борунов, В.Т. Малыгин**

*Московский информационно-технологический университет - Московский архитектурно-строительный институт, г. Москва  
Финансовый университет при Правительстве Российской Федерации,  
Владимирский филиал, г. Владимир*

## **СТАТИСТИЧЕСКИЕ МЕТОДЫ И ПРИЁМЫ ЛИНГВИСТИЧЕСКОГО ИЗУЧЕНИЯ ЯЗЫКА ПИСАТЕЛЯ**

**(на материале художественной англоязычной прозы)**

### **STATISTICAL AND LINGUISTIC TECHNIQUES OF WRITER'S LANGUAGE EXPLORATION (on the material of the English prose)**

*Ключевые слова: художественная проза, статистический анализ, идиостиль, лингвостатистика, количественные методы, Джон Уиндем, Рэй Брэдбери*

*Keywords: literary prose, statistical research, ideostyle, linguostatistics, quantity methods, John Wyndhem, Ray Bradbury*

В современных лингвистических исследованиях всё большее внимание уделяется формализации описания языковых явлений, которая обуславливается большими возможностями автоматизированной обработки текстов. Как справедливо отмечает М.М. Мухин, «стремление к формализации содержания текста приводит исследователей к системному описанию не только смысла конкретного произведения, но и специфики целых индивидуально-авторских художественных систем» (Мухин, 2011). При этом важнейшим моментом методики анализа является сопоставление произведений разных писателей или нескольких текстов одного автора.

Лингвостатистические исследования художественных текстов позволяют выявить смысловую структуру текста, а также концептуальные и стилистические особенности авторского идиостиля и его отличия от других художественных систем (Мухин, 2011).

Л.Л. Нелюбин тоже отмечает, что «в направлении получения общих характеристик текста и его функционального стиля, лингвистические объекты исследуются с применением традиционных статистических методов для изучения произведений художественной литературы и системы языка в целом, способствуя определению общих статистических законов применительно к языку в его статике и динамике» (Нелюбин, 2003: 320).

Вместе с тем, нельзя не отметить, что некоторые учёные несколько скептически относятся к статистическим методам при исследовании художественных текстов. К примеру, М. Хэллидей пишет, что «адекватный анализ индивидуального стиля писателя представляет большую ценность, чем просто подсчёт лингвистически выделенных единиц» (Хэллидей, 1980: 136). Однако одно лишь упоминание о всё возрастающем интересе, например, к составлению частотных словарей произведений отдельных

авторов красноречиво говорит о том, что лингвостатистика уже прочно завоевала себе место под солнцем.

Первые исследования идиостиля связаны с именами Ю.Н. Тынянова, Ю.Н. Караулова и В.В. Виноградова, которые в середине XX века занимались изучением языковой личности. В.В. Виноградов дал следующую дефиницию данному понятию: «идиостиль – есть система эстетически-творческого подбора, осмысления и расположения символов» (Виноградов, 1980: 358). В свою очередь Ю.А. Беличиков отмечал, что «языковое мастерство писателя, его индивидуальную манеру письма называют индивидуально-авторским стилем» (Беличиков, 1987: 423).

В данной статье предпринимается попытка выделения черт индивидуального стиля писателей Д. Уиндема и Р. Брэдбери посредством обработки выбранных для анализа лексической синтагматики текстов специализированным компьютерным обеспечением и последующей интерпретации полученных данных на основе их сопоставительного анализа.

Фактическим языковым материалом исследования являются:

- 1) англоязычный рассказ британского писателя-фантаста Джона Уиндема «Поиски наугад» («Random quest», 1961);
- 2) англоязычный рассказ американского писателя-фантаста Рэя Брэдбери «И грянул гром» («A sound of Thunder», 1952).

Отбор фактического материала продиктован стремлением к достижению максимальной объективности исследования, для чего намеренно были выбраны два рассказа разных авторов на английском языке, написанные в одинаковом жанре и примерно в одно и то же время. Данные произведения позволяют провести статистическое исследование в сопоставительном плане, сравнив полученные результаты.

Опишем кратко методику проведенного статистического исследования текстов, которая условно может быть разделена на пять стадий:

- стадия 1: формулировка целей и задач лингвостатистического исследования: подсчёт слов в произведении, определение числа уникальных слов, частотный анализ, выделение ключевых и семантически значимых слов, морфологический анализ фрагментов произведений, анализ структурно-вероятностных моделей текста, стилистический анализ, установление средней длины слова, проведение различных лингвостатистических экспериментов, выбор специализированных компьютерных программ;
- стадия 2: отбор фактического англоязычного материала, перевод печатных текстов в электронный корпус;
- стадия 3: обработка данных с помощью специализированных компьютерных программ;
- стадия 4: сведение и систематизация полученной информации, представление полученных данных в табличном формате;
- стадия 5: анализ статистических данных и формулирование выводов по итогам работы.

В настоящее время при анализе художественных текстов используются, в основном, методы анализа формальных характеристик письменной речи. Данные характеристики представляют собой частоту встречаемости тех или иных единиц речи в исследуемом тексте (частотность). Довольно часто представляется трудным выделить в тексте идиостиль писателя, так как авторы при использовании тех или иных приёмов достижения художественности текстов невольно уподобляются манере изложения других писателей, нередко даже в некоторой степени копируя её, что ведёт к размытости их индивидуального стиля. Именно поэтому статистические методы изучения текста являются незаменимыми в данном случае, так как они позволяют автоматизировано, при помощи специализированного программного обеспечения, выделить в тексте сугубо авторские особенности. К примеру, если иметь в виду частеречные характеристики текста, то естественно каждый автор каждый автор употребляет в своих произведениях разное количество существительных, глаголов, прилагательных и т.д., оперируя нередко определённым набором «любимых» слов.

Проиллюстрируем это конкретными примерами. Рассмотрим, например, в указанных текстах вербализацию концепта «удивление». Так, у Д. Уиндема в рассказе «Поиски наугад» используется не столь часто встречающийся для обозначения удивления глагол «to jolt» (4 раза во всём тексте). Концепт удивления передаётся также глаголом «to daze» (5 раз) и прилагательным «curious» (9 раз). Причём это прилагательное употребляется в значениях «странный», «удивительный», «удивлённый», а не «любопытный». Прилагательное «odd» в значении «странный», «необычный» встречается в тексте 7 раз. Интересно, что наиболее часто употребляемое для обозначения удивления слово «surprise» используется в этом рассказе относительно редко – всего 3 раза.

Концепт «омерзительный» выражается у Р. Брэдли словом «slime» (использовано 4 раза, что составляет 0.1 % от всех употреблённых в тексте слов).

Эти наблюдения подтверждают, что, будучи ментальной репрезентацией, которая определяет, как вещи связаны между собой и как они категоризируются (Бабушкин, 1996: 103), концепты формируют индивидуальные концептосферы писателей, отражая их внутренний мир и образность мышления.

Статистические методы в изучении языка писателя позволяют выявить лексические средства объективации его индивидуальных концептосфер, то есть всю совокупность слов и понятий, наиболее часто используемых им в своих произведениях. Если предположить, что, как правило, писатель выбирает необходимое ему для достижения определенного художественного эффекта определённое слово из синонимического ряда подсознательно, то становится очевидной возможность исследования художественных текстов на уровне подсознания автора. Использование лингвостатистики позволяет

при этом говорить о достаточно высокой степени объективности такой методологии.

По лингвостатистической методике А.Н. Баранова, описанной в его книге «Введение в прикладную лингвистику», характеристики идиостиля автора устанавливаются путём количественного анализа квазисинонимичных лексем, то есть близких по значению слов, словосочетаний или фразеологизмов. Также автор отмечает, что «информацию о стиле писателя можно получить из частотных характеристик служебных и модальных слов в тексте произведения» (Баранов, 2001: 347).

Кроме того, применение методов статистики при изучении художественных текстов влечёт за собой появление и развитие новых научных направлений в языкознании, в частности, в лингвостатистике, стилометрии, структурной и прикладной лингвистике.

Обратим внимание на то, что данные, полученные лингвостатистическими методами, являются точными и устойчивыми на протяжении длительного периода. Так, было выявлено, что доля служебных слов в текстах остаётся неизменной в течение всей творческой жизни автора. Данный параметр может помочь не только в определении индивидуального стиля писателя, но и при установлении личности автора путём литературоведческой экспертизы.

Как уже было отмечено ранее, статистика текста затрагивает бессознательную часть творчества писателя, то есть статистика с наибольшим успехом анализирует именно ту часть произведения, которая была написана без сознательных раздумий. Например, частота употребления редких и иностранных слов может свидетельствовать об уровне интеллекта автора. Тем не менее, этот фактор легко может быть установлен и на сознательном уровне. Если автор сознательно «обогащает» своё произведение иностранными словами, постоянно заглядывая в словарь, то статистика не сможет использовать этот фактор при определении идиостиля писателя. То есть, для его объективной оценки при помощи статистических методов необходимо соблюдать следующие требования:

- 1) показатель авторской индивидуальности должен характеризоваться подсознательным уровнем своего происхождения; например, таким показателем может быть количество использованных автором в тексте служебных слов, поскольку вряд ли какому автору придёт в голову считать количество предлогов, частиц и союзов в своих текстах;
- 2) показатель должен иметь постоянное числовое значение в произведениях определённого автора;
- 3) показатель должен иметь очевидные отличительные значения и характеристики в сравнении с текстами других писателей.

Вместе с тем, работа лишь с сухими статистическими данными при исследовании индивидуального стиля писателя может не привести к ожидаемому результату, так как при частотном анализе текста практически у

всех писателей выявляется примерно одинаковый набор сверхчастотных лексических единиц.

Приведём списки из 20 ключевых слов (полученных нами в ходе анализа через программу Seozor) из проанализированных нами текстов «Поиски наугад» (Уиндем) и «И грянул гром» (Брэдбери).

У Уиндема этот список выглядит так: *the, and, was, you, that, but, had, not, she, for, with, there, have, his, said, one, her, all then, did.*

А у Брэдбэри список несколько иной: *the, and, his, you, that, was, this, not, they, all, said, time, like, there, its, machine, with, can, one, out.*

Как видим, 13 из 20 слов из обоих списков совпадают. По различающимся 7 словам вряд ли можно сделать какие-то серьёзные выводы как о стиле авторов, так и о хотя бы общем содержании произведения. Сопоставим эти списки из 7 слов:

в тексте «Поиски наугад»: *had, she, for, have, her, then, did.*

в тексте «И грянул гром»: *time, like, its, machine, can, out, this.*

Приведенные выше отдельные лексемы не позволяют нам говорить о каких-либо особенностях идиостиля авторов указанных текстов. Они могут служить лишь некими сигналами о возможных персонажах или, скажем, о теме произведения, но не более того. Так, мы могли бы предположить, что в проанализированном тексте из произведения «Поиски наугад» присутствует женский персонаж (*на основании местоимений «she», «her»*), а в тексте из произведения «И грянул гром» женские персонажи отсутствуют, так как в данном списке наиболее частотных слов нет местоимений женского рода.

Анализ текстов этих рассказов показал, что наше предположение оправдалось. Кроме того, в ряду 7 указанных ключевых слов из рассказа Брэдбэри стоят два существительных «*time*» («время») и «*machine*» («машина»). Можно ли предположить, что в данном рассказе речь идёт о некой «машине времени»? В данном конкретном случае так оно и есть, но значит ли это, что такой набор лексем всегда будет связан с такой темой и с таким содержанием? Разумеется, нет. А значит только на этом основании невозможно делать какие-либо скоропалительные выводы ни о содержании текстов, ни тем более о каком-то своеобразии идиостиля их авторов. И хотя словосочетание «машина времени» может направить нас на мысль, что автор, употребивший в своём рассказе это словосочетание никто иной, как фантаст, это само по себе не более чем фантазия, не представляющая никакой научной ценности.

Проанализируем список выявленных нами семантически значимых наиболее часто употребляемых слов в указанных произведениях, в круглых скобках указана частотность слова.

У Уиндема это: *little* – «маленький» (45), *man* – «человек» (44), *name* – «имя» (39), *doctor* – «доктор» (37), *happen* – «случаться» (33), *young* – «молодой» (29), *feel* – «чувствовать» (28), *way* – «путь» (24), *moment* – «момент» (22), *seem* – «казаться» (21), *call* – «звонить» (18), *different* – «другой» (18), *marry* – «жениться» (16), *own* – «собственный» (16), *end* –

«конец» (15), *enquiry* – «запрос» (15), *place* – «место» (14), *plane* – «плоскость» (14), *woman* – «женщина» (14), *book* – «книга» (14).

А у Брэдбери это: *time* – «время» (26), *machine* – «машина» (23), *path* – «тропинка» (20), *hand* – «рука» (15), *monster* – «чудовище» (12), *mouse* – «мышь» (11), *safari* – «сафари» (11), *body* – «тело» (9), *change* – «меняться» (9), *feel* – «чувствовать» (9), *kill* – «убивать» (9), *step* – «шаг» (9), *ten* – «десять» (9), *desk* – «стол» (8), *gun* – «ружьё» (8), *jungle* – «джунгли» (8), *million* – «миллион» (8), *mouth* – «рот» (8), *move* – «двигаться» (8), *rifle* – «винтовка» (8). Это говорит о том, что статистика, применяющаяся при исследовании художественных произведений, должна быть относительна, а не опираться лишь на сухие факты.

Список слов по Уиндему носит довольно абстрактный характер. Судя только по приведенным лексемам, невозможно сказать, о чём этот рассказ. Можно только предположить, что речь идёт как минимум о трёх действующих лицах: доктор, молодой человек и женщина. А вот по списку слов из рассказа Брэдбери можно предположительно даже обозначить его содержание: путешествие на машине времени (*time*, *machine*, *path*), появление чудовища (*monster*), сафари, то есть охота (*safari*), убийство (*kill*, *gun*) и заканчивается всё винтовкой (*rifle*), то есть, скорее всего, выстрелом. И это всё соответствует действительному содержанию рассказа.

Чем же можно объяснить столь существенную разницу двух наших выборок? Данный факт можно объяснить объёмом текста. Объём «Поиски наугад» составляет примерно 15 000 слов, в то время как «И грянул гром» состоит всего из 4500 слов, то есть объём произведения Брэдбери почти в 3,5 раза меньше объёма произведения Уиндема. А это говорит о том, что значимым здесь является сам факт наличия у Брэдбери более богатых лексических возможностей для объективации его индивидуальной концептосферы.

Исследуемые тексты характеризуются также некоторыми особенностями в использовании их авторами функционально-смысловых типов речи. Для Дж. Уиндема характерно повествование с элементами рассуждения, а Рэй Брэдбери больше тяготеет к повествованию с элементами описания. Кроме того, у Дж. Уиндема в рассказе «Поиски наугад» присутствуют фрагменты научного функционального стиля.

Приведём результаты статистического анализа употребления частей речи в исследуемых рассказах Уиндема и Брэдбери.

Для анализа были выбраны из данных рассказов по три фрагмента объёмом в 200 слов, относящихся к одному функциональному стилю и функционально-смысловому типу речи. В каждом из фрагментов было подсчитано количество имён существительных, прилагательных, числительных, имён собственных, глаголов и служебных частей речи. Затем мы сложили полученные показатели и вывели среднее арифметическое по всем фрагментам для каждой части речи отдельно по каждому рассказу. В представленной таблице 1 сведены среднеарифметические данные о

количестве частей речи на каждые 200 слов в рассказах «Поиски наугад» и «И грянул гром».

Таблица 1. Среднеарифметические данные о количестве частей речи на каждые 200 слов

Часть речи	Джон Уиндем «Поиски наугад»	Рэй Брэдбери «И грянул гром»	Разница (%)
Имя существительное	45 (22.5%)	52 (26%)	7%
Имя прилагательное	17 (8.5%)	23 (11.5%)	6%
Глагол	30 (15%)	29 (14.5%)	1%
Местоимение	37 (18.5%)	20 (10%)	17%
Наречие	15 (7.5%)	6 (3%)	9%
Служебные слова (союзы, предлоги, частицы, артикли)	55 (27.5%)	63 (31.5%)	8%
Имя числительное	1 (0.5%)	2 (1%)	1%
Имя собственное	2 (1%)	6 (3%)	4%

Мы видим, что для обоих рассказов характерно большое количество глаголов и имён существительных, что свидетельствует о принадлежности обоих текстов в основном к одному функционально-смысловому типу речи – повествованию.

В рассказе Брэдбери «И грянул гром» встречается большее количество прилагательных (11,5 %), так как прилагательные делают художественный текст красивее, образнее, ярче и, как правило, выступая в роли эпитетов, характеризует предметы, явления, героев, подчёркивая их индивидуальные особенности, качества и свойства. Это также свидетельствует о принадлежности рассказа Брэдбери к типу повествования с элементами описания.

Наличие большего числа имён собственных в произведении Р. Брэдбери объясняется тем, что он ведёт повествование от третьего лица, в то время как в произведении Дж. Уиндема основное повествование идёт от первого лица, и лишь небольшие фрагменты в конце и начале рассказа ведутся от третьего лица, а в основном текст строится в форме диалогов.

Обилие наречий в рассказе «Поиски наугад» может рассматриваться как еще одно важное и яркое средство оценочной характеристики действий героев, реалистичности описания и художественной выразительности авторской речи. Можно предположить, что главный герой этого рассказа является мыслящим человеком, пытающимся объективно оценить окружающую его обстановку. Всему рассказу в целом свойственна большая оценочность явлений, чем для рассказа «И грянул гром». Это объясняется тем, что произведение Дж. Уиндема хоть и является рассказом в жанре фантастики, но тяготеет скорее к реализму. Обилие наречий (7,5% против 3 % в другом исследуемом рассказе) компенсирует меньшее количество прилагательных (8,5 % против 11,5 %), придавая тексту экспрессивный,

эмоционально-окрашенный характер. Меньшее количество прилагательных у Уиндема приводит к тому, что сокращается длина предложений.

Таким образом, лингвостатистический анализ художественных произведений двух писателей-фантастов, живших и творивших в одну эпоху (середина XX века), позволил выявить некоторые специфические черты их идиостиля, нашедших своё выражение в свойственном каждому из них своеобразии лексической объективации их индивидуальных концептосфер.

### Литература

1. Бабушкин А.П. Типы концептов в лексико-фразеологической семантике языка. – Воронеж: изд-во Воронеж. гос. ун-та, 1996. – 103 с.
2. Баранов А.Н. Введение в прикладную лингвистику. – М.: Эдиториал УРСС, 2001. - 347 с.
3. Бельчиков Ю.А. Стиль в языкознании // Литературный энциклопедический словарь. - М.: Советская энциклопедия, 1987. – С. 442 - 443.
4. Виноградов В.В. Избранные труды. О языке художественной прозы. – М.: Наука, 1980. – 358 с.
5. Карасик В.И. Лингвосемиотическое моделирование ценностей // Политическая лингвистика, - Екатеринбург. - Вып. 1(39), 2012.– С. 43 - 50.
6. Лихачев Д.С. Концептосфера русского языка. – М.: Известия ОРЯ, Серия литературы и языка. - Т. 52, №1, 1993. – С. 3 - 9.
7. Мухин М.Ю. Лексическая статистика и идиостиль автора. // Вестник Южно-Уральского государственного университета. Серия: Лингвистика – Выпуск №2 (135), 2009. – С. 51 - 55.
8. Нелюбин Л.Л. Толковый переводоведческий словарь. - М.: Флинта, 2003. – 320 с.
9. Носорева Н.Н., Чернова Ю.С. Лингвостатистический подход в изучении литературных направлений // Символ науки. – М. – Вып. № 12-1, 2015. – С. 223 - 225.
10. Сурина В.Н. Понятие концепта и концептосферы – М.: Молодой ученый. - № 5. Т.2, 2010. – С. 43 - 46.
11. Хэллiday М.А.К. Лингвистическая функция и литературный стиль. // Новое в зарубежной лингвистике. Лингвостилистика. – М.: Прогресс. – Вып. 9, 1980. – С. 118 - 147.
12. Семантический анализ текста онлайн. - [Электронный ресурс]. – Электрон. данные. – Режим доступа <http://advego.ru/text/seo/> свободный. Заглавие с экрана, дата обращения: 16.10.2016.
13. A Sound of Thunder. Рассказ Рэя Брэдли. - [Электронный ресурс]. – Электрон. данные. – Режим доступа: <http://raybradbury.ru/library/story/52/8/0/> Заглавие с экрана, дата обращения: 20.10.2016.
14. Seo-анализ текста. - [Электронный ресурс]. – Электрон. данные. – Режим доступа <https://text.ru/seo>, свободный. Заглавие с экрана, дата обращения: 30.06.2016.



15. Wyndham J. The Infinite moment. – New York: Ballantine books, 1961.  
- 159 p.

**(0,6 п.л.)**