

Борунов Артем Борисович

ПАРАМЕТРИЧЕСКАЯ СТРАТИФИКАЦИЯ ЛЕКСИКИ АВТОРСКОГО АНГЛОЯЗЫЧНОГО КОРПУСА (НА МАТЕРИАЛЕ АВТОРСКОГО КОРПУСА АМЕРИКАНСКОГО ПИСАТЕЛЯ РЭГУ Н. МИТРЫ)

Предметом исследования является параметрическая стратификация авторского англоязычного корпуса Рэгу Н. Митры. Объект исследования - авторский корпус объемом 414 311 единиц. Автор рассматривает функциональный параметр лексики корпуса - длина, частотность, а также выборка единиц, относящихся к ядру корпуса. Полученные данные позволяют выделить специфику Рэгу Н. Митры как автора, а также дают возможность построить частотные авторские словари.

Адрес статьи: www.gramota.net/materials/2/2016/12-1/19.html

Источник

Филологические науки. Вопросы теории и практики

Тамбов: Грамота, 2016. № 12(66): в 4-х ч. Ч. 1. С. 65-68. ISSN 1997-2911.

Адрес журнала: www.gramota.net/editions/2.html

Содержание данного номера журнала: www.gramota.net/materials/2/2016/12-1/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: phil@gramota.net

COGNITIVE CONSTRUCTION OF CONCESSIVENESS

Bezzemel'naya Olesya Alekseevna

*Ufa Law Institute of the Ministry of Internal Affairs of the Russian Federation
olesik_2009@mail.ru*

The article examines the cognitive construction of concessiveness. Concessiveness is directly linked to the conceptualization of worldview and thinking. The concept of “concessiveness” encloses semantics of the previous fragment in a certain frame by the recognition of the simultaneous existence of a contradictory thought, fact, and action, along with what has been reported previously, contrary to it. Concessiveness reflects such communicative categories of English communicative consciousness as ambiguity, tolerance, non-categoricity and non-conflictiness.

Key words and phrases: concessiveness; causality; concessive nature; adversative nature; contrariety; language worldview.

УДК 81-139

Предметом исследования является параметрическая стратификация авторского англоязычного корпуса Рэгу Н. Митры. Объект исследования – авторский корпус объемом 414 311 единиц. Автор рассматривает функциональный параметр лексики корпуса – длина, частотность, а также выборка единиц, относящихся к ядру корпуса. Полученные данные позволяют выделить специфику Рэгу Н. Митры как автора, а также дают возможность построить частотные авторские словари.

Ключевые слова и фразы: квантитативный анализ; авторский корпус; параметрические исследования; Рэгу Н. Митра; основное частотное ядро; функциональный параметр.

Борунов Артем Борисович

*Московский информационно-технологический университет –
Московский архитектурно-строительный институт
borunov.artem@yandex.ru*

**ПАРАМЕТРИЧЕСКАЯ СТРАТИФИКАЦИЯ ЛЕКСИКИ
АВТОРСКОГО АНГЛОЯЗЫЧНОГО КОРПУСА
(НА МАТЕРИАЛЕ АВТОРСКОГО КОРПУСА АМЕРИКАНСКОГО ПИСАТЕЛЯ РЭГУ Н. МИТРЫ)**

Параметрические исследования лексики довольно успешно проводят различные учёные как в России, так и за рубежом. Целью параметрического изучения лексики является статистическая обработка текста, анализ полученных результатов и их интерпретация для составления алфавитно-частотных глоссариев, выделения особенностей словоупотребления и авторских индивидуальных черт, для определения авторства. Чтобы показать важность данного направления исследований, назовем наиболее крупные работы последних лет, которые были проведены учёными научной школы Воронежского государственного университета под руководством профессора А. А. Кротова, такие как, например, исследование параметрических ядер германских языков О. М. Воевудской [2] и Д. С. Воевудского [3], параметрический анализ лексики немецкого языка Т. А. Казаковой [4], параметрическое исследование лексики русского языка Ю. А. Стародубцевой [7] и др. Среди ряда вышеперечисленных работ особо выделим масштабные параметрические исследования лексики германских языков, проведенные О. М. Воевудской и Д. С. Воевудским. К частным вопросам и проблемам квантитативной лексикологии, корпусной лингвистики, лингвостатистики отечественные лингвисты проявляют всё больший интерес, к примеру, назовем теоретические статьи А. Б. Борунова и В. Т. Малыгина [1], А. В. Кашеевой [5], В. А. Яцко [10] и многих др.

Корпусные исследования дают богатый материал, а использование компьютерных программ при обработке текста позволяет мгновенно получить точные данные для интерпретации. Обработка авторских корпусов имеет давнюю традицию как в мировом, так и в отечественном языкознании, например, в 1956-1961 гг. был опубликован Словарь языка Пушкина [6], а также словари отдельных произведений, например Частотный словарь романа Л. Н. Толстого «Война и мир» [9], опубликованный в 1978 г., и ряд других работ. Данное направление исследования, несомненно, актуально, так как в наши дни внимание исследователей всё чаще привлекают корпусы современных писателей, иногда неизвестных широкому читателю, что отнюдь не снижает важности изучения полученного материала, который в дальнейшем может использоваться в лексикографии, в том числе и авторской.

Для выявления параметрических характеристик лексики прибегнем к параметрическому анализу лексико-семантической системы языка, который развивали в своих работах В. Т. Титов и О. М. Воевудская [2; 8]. В их работах, при исследовании параметрических ядер романских и германских языков, лексические единицы отбираются по четырем параметрам, а именно:

- 1) функциональный параметр (употребительность);
- 2) синтагматический параметр (сочетаемость);
- 3) эпидигматический параметр (многозначность);
- 4) парадигматический параметр (вхождение в синонимические ряды).

Принимая во внимание ограниченные объемы статьи, мы решили, беря за основу параметры анализа параметрических ядер, выделенные О. М. Воеводской, ограничиться только функциональным параметром и выделить количественные характеристики параметрического ядра исследуемого текста.

Материалом данного исследования является авторский англоязычный корпус, созданный нами путем перевода произведений автора в электронный вид, вычитки материала и объединения в единый корпус. В авторский корпус были включены следующие произведения Рэгу Н. Митры:

- 1) «Очень банальная страсть» («A Very Insipid Passion» [12]);
- 2) «Грехопадение» («Impute Fall to Sin» [16]);
- 3) «Дождь из теней» («A Rain Full of Ghosts» [11]);
- 4) «Если бы не смерть» («If there wasn't death» [15]);

а также 2 фрагмента других произведений писателя, представленных в свободном доступе в сети Интернет:

- 5) «As in the falling of an eyelid» [13];
- 6) «At The Davies» [14].

Полученный файл с авторским корпусом был загружен нами в программу «WordSmith Tools» [17]. Файл электронного корпуса текстов Р. Н. Митры включает в себя 414 311 слов, которые имеют различный индекс частотности.

В данной статье рассмотрим функциональный параметр лексики авторского корпуса. Как известно, показателем функциональной активности слова является его длина в буквах. Известный американский лингвист Джордж Ципф установил закономерность: частотность слов обратно пропорциональна их длине: чем короче слово, тем (при прочих равных условиях) чаще оно употребляется, и наоборот [18]. Построим автоматически в программе «WordSmith Tools» Таблицу 1, в которую занесены полученные данные.

Таблица 1.

Распределение лексики по длине слов в буквах

1- буквен-ные слова	2- буквен-ные слова	3- буквен-ные слова	4- буквен-ные слова	5- буквен-ные слова	6- буквен-ные слова	7- буквен-ные слова
23 498	71 264	99 945	78 327	44 295	33 420	25 601
8- буквен-ные слова	9- буквен-ные слова	10- буквен-ные слова	11- буквен-ные слова	12- буквен-ные слова	13- буквен-ные слова	14- буквен-ные слова
16 185	10 305	6 270	2 888	1 660	775	239
15- буквен-ные слова	16- буквен-ные слова	17- буквен-ные слова	18- буквен-ные слова	19- буквен-ные слова	20- буквен-ные слова	21- буквен-ные слова
101	35	12	0	0	3	0

Отметим, что при анализе был использован весь авторский корпус: артикли, местоимения, числительные, наречия, служебные слова, а также все онимы. В данном случае мы не ставим перед собой задачу целенаправленно исследовать лексико-семантический уровень языка, поэтому преднамеренно не выделяли из статистических данных аппеллятивы. Статистика программы «WordSmith Tools» показывает частотность словоупотребления, так если, к примеру, артикль «a» (однобуквенное слово) встречается в тексте 1000 раз, значит, в статистическом отчете-таблице отображается «1-буквенное слово – 1000».

Данные указывают на то, что наиболее частотными в произведениях Р. Н. Митры являются 3-буквенные слова, что может быть объяснено включением в подсчет определенного артикля «the» (24% от всего корпуса). Количество слов, состоящих из 1, 2 и 3 букв, составляет около 47% корпуса, что объясняется обилием служебных слов: артиклей, предлогов, союзов, состоящих в английском языке преимущественно из 1-3 букв. Отметим присутствие в корпусе 16- (например, *AUTOBIOGRAPHICAL* «автобиографический», *CHARACTEROLOGICAL* «характерологический»), 17- (например, *CIRCUMFERENTIALLY* «по кольцу», *INCONSEQUENTIALLY* «нелогично») и 20- (например, *UNCHARACTERISTICALLY* «нетипично») буквенных слов, составляющих около 0,01% от всего корпуса. Как видно, данные слова представлены, как правило, дериватами, образованными с помощью префиксов и аффиксов. При проведении дальнейшего исследования было бы интересно сравнить количество *n*-буквенных слов в авторском корпусе Р. Н. Митры с процентным соотношением *n*-буквенных слов английского языка, а также с авторским корпусом другого писателя.

Выделим основное ядро авторского корпуса Р. Н. Митры, в которое входят 932 единицы, процент употребления которых 0,01% и выше от общего текста и частотность употребления 42 и выше. Единицы с частотностью 41 и ниже составляют процент от текста менее 0,01, что не поддается исчислению программой «WordSmith Tools». Среди основных 932 единиц мы выделим значимые, а именно:

- а) антропонимы;
- б) имена существительные;
- в) смысловые глаголы.

Основные антропонимы литературных произведений в творчестве Р. Н. Митры представлены в Таблице 2, в которую были внесены только имена основных героев произведений.

Таблица 2.

Антропони́мы в исследуемом корпусе

Антропоним	Частотность употребления	% от общей доли корпуса
<i>Halley</i> (Хейли)	1830	0,44
<i>Sandy</i> (Сэнди)	1603	0,39
<i>Martin</i> (Мартин)	881	0,21
<i>Dob</i> (Доб)	858	0,21
<i>Dobbelia</i> (Доббеля)	617	0,15
<i>Bloorwoise</i> (Блурвайз)	409	0,1
<i>Willard</i> (Уиллард)	246	0,06
<i>Smith</i> (Смит)	227	0,05
<i>Alexander</i> (Александр)	57	0,01

В произведениях Рэгу Н. Митры нарратором является Мартин, именно от его лица ведется повествование. Его друзьями являются Доб (Доббеля), Снди Блурвайз (Alexander) и Хейли. Имена данных героев являются наиболее часто употребляемыми (1,62% всего текста). Имя «Halley Willard» употребляется чаще других в силу того, что он является руководителем полицейского управления. Что касается персонажа «Sandy Bloorwoise», тут мы видим, что чаще фигурирует в личном общении, его называют по имени «Сэнди» (частотность 1603), а в официальном общении по фамилии – всего 409, так же как и с персонажем «Dob» (Dobbelia Smith) – в личном общении – 858, а официально полным именем – 617 (не считая употребления фамилии).

В Таблице 3 приведём список имен существительных (преимущественно только единственное число, исключение сделано только для существительных, образующих множественное число супплетивно) и смысловых глаголов с частотностью более 41 и процентом от общего текста 0,01 и выше. Они, наряду с антропонимами, формируют основное ядро авторского корпуса писателя. Все остальные единицы: прилагательные, местоимения, формы, образованные от глаголов, предлоги и т.п. в таблицу не были включены. В первом столбике цифра указывает порядковый номер слова в списке частотного употребления, сформированного программой «Wordsmith Tools».

Таблица 3.

Имена существительные и смысловые глаголы в корпусе (часть таблицы)

<i>N</i>	<i>Word</i>	<i>Freq.</i>	<i>%</i>
50	<i>LIKE</i> (любить)	1225	0,30
54	<i>KNOW</i> (знать)	1142	0,28
57	<i>TIME</i> (время)	1089	0,26
68	<i>MAN</i> (человек)	838	0,20
70	<i>GET</i> (получать)	834	0,20
	...		
910	<i>COURT</i> (суд)	42	0,01
911	<i>DIE</i> (умирать)	42	0,01
915	<i>HIDE</i> (прятать)	42	0,01
920	<i>PICK</i> (фразовый глагол)	42	0,01
923	<i>RING</i> (кольцо, звонить)	42	0,01
925	<i>SIGN</i> (подписывать, знак)	42	0,01
927	<i>STEP</i> (шаг, шагать)	42	0,01
931	<i>WONDER</i> (спрашивать)	42	0,01

Из таблицы видно, после анализа 932 исходных единиц, методом ручной выборки, данный список был уменьшен до 314 единиц, что составляет 33,6% от первоначального объёма. Таким образом, параметрическое ядро авторского корпуса Рэгу Н. Митры состоит из 314 единиц, составляющих базовый лексикон писателя.

Назовём, опираясь на данные, приведенные в таблице, состав основного ядра авторского корпуса, которое состоит из:

а) общих часто употребляемых единиц, таких как *go* (идти), *come* (приходить), *like* (любить), *stop* (останавливаться) и т.п.;

б) специфических единиц, употребление которых обуславливается спецификой текста – детективного романа, героями которого являются не только полицейские, но и врачи. Среди таких специфических частотных единиц авторского корпуса назовем, например, *inspector* (инспектор), *police* (полиция), *autopsy* (вскрытие), *die* (умирать), *sergeant* (сержант), *therapy* (терапия), *accident* (происшествие), *hospital* (больница), *psychiatrist* (психиатр), *murder* (убийца), *crime* (преступление), *drug* (наркотик), *money* (деньги) и др. Частотность названных единиц обусловлена тематикой произведений писателя;

в) лексических единиц, употребление которых обусловлено авторским замыслом (частная медицинская практика одного из героев – доктора Мартина, жизнь Сэнди, встречи друзей на квартире Мартина, любовь к еде и алкоголю). Приведем данные единицы: *table* (стол), *coffee* (кофе), *piece* (кусочек), *smell* (запах), *physician* (врач), *England* (Англия), *wine* (вино), *whiskey* (виски), *scotch* (скотч, виски), *lunch* (ланч), *dinner* (обед), *depression* (депрессия), *food* (еда), *bottle* (бутылка), *apartment* (квартира), *glass* (стакан), *psychiatrist* (психиатр) и др.

Таким образом, можно сказать, что основное частотное ядро англоязычного авторского корпуса Рэгу Н. Митры состоит из 314 единиц, которые можно разделить на 3 группы – а), б), в). Однако блок а) является общеупотребительной лексикой, лексика блока б) продиктована жанровыми особенностями, а вот блок в) можно интерпретировать как особенности жанровой манеры писателя. Спецификой Р. Н. Митры как автора является отражение его профессиональной деятельности (он врач-психиатр) как в образе доктора Мартина (врач-психиатр, любящий старые детективы), так и в особом выборе лексических единиц для повествования.

Список литературы

1. **Борунов А. Б., Малыгин В. Т.** Исследование англоязычного авторского корпуса Рэгу Н. Митры: опыт обработки текста компьютерной программой «WordSmith Tools» [Электронный ресурс]. URL: http://tverlingua.ru/archive/043/04_043.pdf (дата обращения: 10.08.2016).
2. **Воеводская О. М.** Концепция идеографического словаря основного лексического фонда германских языков: дисс. ... д. филол. н. Воронеж, 2015. 450 с.
3. **Воеводский Д. С.** Парадигматическая стратификация лексики нидерландского языка // Вестник Воронежского государственного университета. Серия «Лингвистика и межкультурная коммуникация». 2013. № 1. С. 111-114.
4. **Казакова Т. А.** Параметрический анализ немецкой лексики: дисс. ... к. филол. н. Воронеж, 2008. 254 с.
5. **Кашеева А. В.** Квантитативные и качественные методы исследования в прикладной лингвистике // Социально-экономические явления и процессы. 2013. № 3 (049). С. 155-162.
6. **Словарь языка Пушкина:** в 4-х т. М.: ГИС, 1956-1961. Т. 1. 806 с.; Т. 2. 896 с.; Т. 3. 1070 с.; Т. 4. 1045 с.
7. **Стародубцева Ю. А.** Исследование параметрического ядра русской лексики по данным MAC-2 // Вестник ВГУ. Серия «Лингвистика и межкультурная коммуникация». 2015. № 4. С. 82-91.
8. **Титов В. Т.** Частная квантитативная лексикология романских языков: монография. Воронеж: Издательство Воронежского государственного университета, 2004. 552 с.
9. **Частотный словарь романа Л. Н. Толстого «Война и мир».** Тула: Изд-во ТГПИ им. Л. Н. Толстого, 1978. 380 с.
10. **Яцко В. А.** Предметная область компьютерной лингвистики // Вестник ИГЛУ. 2014. № 2 (27). С. 24-35.
11. **Mitra R. N.** A Rain Full of Ghosts. Baltimore: Publish America, 2004. 366 p.
12. **Mitra R. N.** A Very Insipid Passion. M.: Manager, 2002. 336 p.
13. **Mitra R. N.** As in the falling of an eyelid (отрывок части книги) [Электронный ресурс]. URL: <http://shibahill.tripod.com/eyelid.html> (дата обращения: 20.02.2016).
14. **Mitra R. N.** At The Davies: A Novel of Medical Life (отрывок части книги) [Электронный ресурс]. URL: <http://shibahill.tripod.com/atthedavies.html> (дата обращения: 20.02.2016).
15. **Mitra R. N.** If there wasn't death. Denver, Colorado: Outskirts Press Inc., 2007. 230 p.
16. **Mitra R. N.** Impute Fall to Sin. M.: Manager, 2005. 336 p.
17. **WordSmith Tools** [Электронный ресурс]. URL: <http://www.lexically.net/wordsmith/index.html> (дата обращения: 20.02.2016).
18. **Zipf G. K.** Relative frequency as a determinant of phonetic change // Harvard Studies in Classical Philology. Cambridge, Mass.: Harvard University Press, 1929. № 40. P. 2-42.

PARAMETRIC STRATIFICATION OF THE AUTHOR'S ENGLISH-LANGUAGE CORPUS VOCABULARY (BY THE MATERIAL OF THE AMERICAN WRITER RAGHU N. MITRAS'S AUTHOR'S CORPUS)

Borunov Artem Borisovich

*Moscow University of Information and Technology -
Moscow Institute of Architecture and Construction
borunov.artem@yandex.ru*

The subject of this study is the parametric stratification of Raghu N. Mitra's author's English-language corpus. The object of the study is the author's corpus containing 414 311 units. The author examines the functional parameter of the corpus vocabulary – the length, frequency and sampling of units, related to the core of the corpus. These data highlight Raghu N. Mitra's specificity as the author, as well as provide an opportunity to compile frequency author's dictionaries.

Key words and phrases: quantitative analysis; author's corpus; parametric study; Raghu N. Mitra; main frequency core; functional parameter.