

ПРОЕКТЫ. ПРОГРАММЫ. ГИПОТЕЗЫ

Мухин М. Ю.
Екатеринбург, Россия
ORCID ID: 0000-0001-8716-9260
E-mail: mikhail.mukhin@urfu.ru

УДК 81'42
DOI 10.26170/FK19-04-01
ББК Ш105.51
ГСНТИ 17.07.31
Код ВАК 10.01.08

Мухин Н. Ю.
Екатеринбург, Россия
ORCID ID: 0000-0001-8679-9869

АВТОРСКИЙ ТЕЗАУРУС И СОЧЕТАЕМОСТЬ СЛОВ: ЛЕКСИКО-СТАТИСТИЧЕСКИЕ МОДЕЛИ ИНДИВИДУАЛЬНОГО СТИЛЯ

Аннотация. В статье рассматриваются возможности сравнительного лексико-статистического анализа, применяемого для изучения индивидуального стиля (идиостиля) автора художественного текста. В рамках концепции идиостиля как двухплановой сущности обсуждаются две исследовательские модели, связанные с выявлением индивидуально-авторского лексикона и особенностей авторской лексической сочетаемости.

Ключевые слова: лексическая статистика; стилометрия; художественный текст; идиостиль; тезаурус; лексическая сочетаемость; синтагматика; лексическая биграмма.

Приведены списки авторских частотных слов, сформированные на базе корпуса классической прозы XIX в. (произведений Л. Н. Толстого, Ф. М. Достоевского, А. П. Чехова, И. С. Тургенева и И. А. Гончарова). Данные списки получены на основании межтекстового и межстилевого статистического сопоставления. Согласно проведенному эксперименту, лексические списки практически без особых вариантов авторизуются филологической аудиторией. Определяется применимость лексической статистики для создания индивидуальных литературно-художественных тезаурусов. Последующая семантическая разметка дает возможность определить наиболее актуальные для разных авторов сферы действительности, отраженные в художественных текстах, а также сферы-лакуны. Такая модель, в частности, позволяет создать базу для значительно более объективной выборки материала идиостилевых словарей.

Рассмотрены принципы описания авторской лексической сочетаемости, связанные с формализованной систематизацией лексических биграмм – пар слов, употребленных в одном фразовом контексте. На материале произведений XX в. показана специфика выявления закономерностей авторской работы со словом. Представлено описание контекстных синтагматических связей на примере идиостиля В. В. Набокова, осуществленное в сопоставлении с творчеством современных ему писателей. Синтагматический анализ проведен на примере авторской сочетаемости наречия «вдруг», которое в романах Владимира Набокова обладает оригинальным контекстным окружением. Такой анализ по сути является фрагментом идиостилевого словаря сочетаемости, построенном на статистически выверенной выборке фразового материала.

Сделаны выводы о применимости сравнительного лексико-статистического анализа для построения идиостилевых моделей в современной стилометрии. Ключевой характеристикой этих моделей является неинтуитивный подход к оценке лексических параметров текстов и дополнительная объективация последующего филологического анализа авторского стиля.

Mukhin M. Yu.
Mukhin N. Yu.
Ekaterinburg, Russia

AUTHOR'S THESAURUS AND WORD COMPATIBILITY: LEXICOSTATISTICAL MODELS OF INDIVIDUAL STYLE

Abstract. The article considers the possibilities of comparative lexicostatistical analysis to study the individual style (idiostyle) of a literary text author. Within the framework of the concept of idiostyle as a two-dimensional essence two research models connected with the identification of individual-author lexicon and peculiarities of an author's lexicon compatibility are discussed.

Keywords: lexical statistics; stylometry; literary text; idiostyle; thesaurus; lexical compatibility; syntagmatics; lexical bigram.

Lists of author's frequency words formed based on the classical prose text corpus of the XIX century (works by L. N. Tolstoy, F. M. Dostoyevsky, A. P. Chekhov, I. S. Turgenev and I. A. Goncharov) are given. These lists were obtained on the basis of intertext and interstyled statistical comparison. According to the experiment, lexical lists are authorized by the philological audience without any specific variants. The applicability of lexical statistics for the creation of individual literary and artistic thesauruses is determined. Subsequent semantic mark-up makes it possible to determine the most actual spheres of reality for different authors, reflected in literary texts, as well as the spheres-lacunae. Such a model, in particular, allows selecting material for idiostylistic dictionaries in a more objective way.

The article considers the principles of the description of an author's lexical compatibility related to the formalized systematization of n-grams – pairs of words used in one phrase context. On the material of the XX century works the specificity of revealing the patterns of an author's work with a word is shown. The article presents the description of context syntagmatic relations on the example of V. V. Nabokov's idiostyle, carried out in comparison with the works of modern writers. Syntagmatic analysis is carried out on the example of the author's combination of adverb "vdrug" ("suddenly"), which in the novels of Vladimir Nabokov has an original contextual environment. Such an analysis is essentially a fragment of an idiostylistic dictionary of compatibility based on a statistically verified sample of phrasal material.

The authors draw conclusions about the applicability of comparative lexicostatistical analysis for the construction of idiostyle models in modern stylometry. A key characteristic of these models is a non-intuitive approach to the evaluation of lexical parameters of texts and additional objectification of the subsequent philological analysis of an author's style.

Благодарности: Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований в рамках научного проекта № 19-012-00104 «Формализация индивидуальной лексической сочетаемости как средство описания идиостилей: корпусное сопоставительное исследование классической прозы XIX в.».

Acknowledgments: The research was carried out with the financial support of the Russian Foundation for Basic Research in the framework of the scientific project No. 19-012-00104 "Formalization of individual lexical compatibility for idiostylistic profiling: a corpus-based comparative study of the 19th century classical prose".

Для цитирования: Мухин, М. Ю. Авторский тезаурус и сочетаемость слов: лексико-статистические модели индивидуального стиля / М. Ю. Мухин, Н. Ю. Мухин // Филологический класс. – 2019. – № 4 (58). – С. 8–15. DOI: 10.26170/FK19-04-01.

For citation: Mukhin, M. Yu., Mukhin, N. Yu. (2019). Author's Thesaurus and Word Compatibility: Lexicostatistical Models of Individual Style. In *Philological Class*. No. 4 (58), pp. 8–15. DOI: 10.26170/FK19-04-01.

1. Вводные соображения

Статистические методы в филологической сфере применяют давно, несмотря на известное гуманитарное сопротивление. Сегодня можно считать классическими проведенные в междисциплинарном ключе работы В. С. Баевского, М. Л. Гаспарова, А. П. Журавлева, Ю. Н. Караулова, А. Я. Шайкевича и многих других известных ученых. Среди зарубежных авторов в этом ряду R. H. Vaayen, J. F. Burrows, T. N. Corns, D. I. Holmes, D. L. Hoover, H. Love и др. Стилистические исследования, предполагающие проведение статистического анализа (в том числе исследования, связанные с авторизацией текста), составляют направление, получившее название «стилометрия» (или «стилеметрия»).

Важным стимулом развития стилометрии является современная корпусная лингвистика. Большие размеченные текстовые базы данных обеспечивают быстрый поиск и получение статистических сведений. Кроме того, программы автоматической обработки текстов (в первую очередь морфологического анализа) позволяют исследователю преодолеть психологический барьер при обращении к большим данным, которые содержат сотни миллионов слов. На этом фоне возникают вопросы о новой методологии филологических исследований, т. е. о грамотном построении исследовательской модели художественного текста и индивидуального стиля автора с использованием статистических данных. Математическая статистика, какими бы правильными ни были сами вычисления, выполняет в гуманитарной сфере обслуживающую функцию. Если модель в области изучения языка и текста построена неверно – никакое знание статистических инструментов не поможет выявить индивидуальные особенности языка и стиля автора.

Цель статьи – охарактеризовать два направления идиостилевого анализа на основании данных лексической статистики и в соответствии с ними представить результаты исследования, проводимого на материале классической прозы XIX и XX вв. К этим двум направлениям относятся:

- 1) выявление индивидуально-авторского лексикона и составление литературно-художественного тезауруса,
- 2) формализованное выявление индивидуальных особенностей авторской лексической сочетаемости.

По сути, эти направления являются наиболее очевидными для лингвистического изучения индивидуального стиля автора. Понимание идиостиля как двусторонней сущности, предполагающей сочетание «концептуальной системы автора» и «индивидуальной трансформации языковых выражений», находим, например, в работах В. А. Пищальниковой [Пищальникова 1992: 47]. Первое, как система смыслов художественного текста, отражающая концептуальную картину мира автора, дает представление об авторских тематических приоритетах (например, о том, какие сферы тезауруса в большей степени представлены в литературном творчестве). Второе, как система индивидуально-употребления языковых единиц в художественном тексте, – о специфике авторской работы со словом.

Сходным образом рассуждает Н. Е. Сулименко: «изучение коммуникативных аспектов лексического значения... связывается с двумя исходными постулатами:

- 1) представлением о реализации в тексте определенной концептуальной системы носителя языка;
- 2) признанием гибкого, подвижного характера лексического значения с заложенными в нем потенциальными преобразованиями» [Сулименко 2009: 52].

При этом в каком бы ракурсе филологической сферы ни применялась статистика, важнейшим исследовательским принципом, по нашему мнению, является сопоставление разных текстов и творчества разных авторов. В самом деле, при изучении индивидуального стиля автора любой исследователь, независимо от избираемой методологии, сталкивается с двумя подводными камнями. Во-первых, ряд стилевых признаков может относиться только к одному конкретному тексту, а не к разным произведениям автора. Например, если в романе Ф. М. Достоевского «Братья Карамазовы» – преимущественно в одном известном всем фраг-

менте – часто употребляется слово *инквизитор*, из этого не следует однозначный вывод о значимости этого слова для всего стиля автора. Во всяком случае, это касается других его романов, в которых это слово вообще не встречается. В «Войне и мире» в число частотных попадают слова *армия, раненый, плен, фланг, атака, пехота, отступление*, но совершенно понятно, что эта «военная» лексика обусловлена романским сюжетом и не является сквозной для всего творчества Л. Н. Толстого.

Во-вторых, у любого писателя встречаются не уникальные, а общезначимые стилевые признаки, характерные для многих, если даже не для всех, авторов. Поэтому статистические данные, сформированные на материале произведений одного писателя, всегда вызывают вопрос об их стилевой маркированности.

В филологических исследованиях нередко можно встретить рассуждения об индивидуальном стиле автора, основанные на наблюдениях каких-либо часто встречающихся у него слов, выражений, средств выразительности и т. д. К сожалению, помимо неучета двух отмеченных принципов, там, как правило, недостает и понимания того, что некое арифметическое количе-

ство наблюдаемых фактов отнюдь не обязательно обладает статистической значимостью, связанной с «большими числами». Таким образом, аналитические модели, представленные в данной статье, обязательно предполагают масштабное статистическое сопоставление текстов и их авторов.

2. Индивидуально-авторский лексикон и литературно-художественный тезаурус

Располагая корпусом прозаических текстов XIX в., мы можем привести список самых частотных слов знаменательных частей речи, которые употребляют разные писатели. Приведенный ниже набор слов в общих чертах повторяет сведения, которые можно почерпнуть из других источников [Гребенников 2006; НЧС http; и др.]: *быть, сказать, говорить, человек, мочь, знать, еще, рука, теперь, уже, становиться, лицо, глаз, хотеть, видеть, дело, вдруг, спрашивать, думать, очень, понимать, день, два, жизнь, голова, время, опять, раз, давать, много*. Здесь 30 слов, и даже если его значительным образом расширить, универсальный характер этого ряда не изменится. В частотных списках эти лексемы распола-

Таблица 1

Индивидуальная лексика в творчестве пяти авторов XIX в.

Автор 1	Автор 2	Автор 3	Автор 4	Автор 5
актер	анекдот	авторитет	ветерок	веселье
бас	благородство	быт	втайне	внешний
всенощная	болезненный	весло	глушь	естественно
высокоблагородие	бредить	влечение	гнездо	заседание
вяло	виновный	ворочаться	говаривать	интересовать
вялый	воплъ	ворчать	голосок	количество
дедушка	горячка	гостить	господский	командир
жидкий	дико	гребец	дворянский	кофей
жутко	исповедь	дремать	джентльмен	либеральный
завод	каторга	дружеский	дождик	министр
земский	личико	морской	дума	наблюдение
инженер	младенец	насилу	загадочный	неприятность
казенный	нелепость	нейти	картуз	несомненный
калоша	необычайный	нехотя	крестьянский	подвода
лечить	осведомляться	нянька	крылечко	предводитель
лечиться	покончить	остров	миловидный	предстоящий
мозг	полиция	пирог	мсье	приемная
нервно	порешить	плетень	несчастливый	принц
обидно	преступник	постройка	поссориться	притворство
обстановка	раздражаться	посуда	предаваться	радовать
почтовый	рассудок	почва	рай	свойство
пошлость	роковой	призрак	робость	сложный
разбойник	скандал	пустыня	родина	сочувствовать
сапожник	скверный	суетиться	сдержанный	счастливо
скотина	сострадание	сутки	соната	удовлетворение
сумерки	струсить	тепло	тучка	управлять
телеграмма	тварь	уныло	унылый	устройство
томить	тоненький	упираться	фортепьяно	физический
тощий	улика	хитрость	француженка	царь
чепуха	унижение	шаль	чепец	энергия

гаются в разном порядке и при этом все равно фигурируют в числе самых употребительных слов любого достаточно протяженного художественного текста. А. О. Гребенников – исследователь рассказов А. Чехова и Л. Андреева – считает, что «эти совпадения легко объясняются прежде всего жанровыми особенностями» [Гребенников 2006: 160]. Мы полагаем, что главная причина лексических совпадений здесь – не столько жанровые, сколько глобальные текстовые особенности, связанные с отражением значимых представлений человека о действительности. Легко понять, что не может быть никакой речи о соотносительности этих слов со стилем конкретного автора.

Теперь рассмотрим другие ряды слов (так же по 30 в каждом), которые выявлены путем статистического сопоставления на базе корпуса классической прозы (это избранные произведения Л. Н. Толстого, Ф. М. Достоевского, А. П. Чехова, И. С. Тургенева и А. И. Гончарова – всего около 4 миллионов слов). Эти ряды соотношены с творчеством пяти указанных авторов, и мы предложим читателю определить, с каким из писателей ассоциируются у него эти слова.

Пока читатель думает об авторизации этих рядов, объясним, как они получены. Списки слов обусловлены межтекстовым и межстилевым сопоставлением. Исходно каждый автор представлен в нашем корпусе четырьмя крупными произведениями:

И. С. Тургенев: «Рудин», «Дворянское гнездо», «Накануне», «Отцы и дети» (178 тыс. слов);

И. А. Гончаров: «Обыкновенная история», «Обломов», «Обрыв», «Фрегат „Паллада“» (708,7 тыс. слов);

Ф. М. Достоевский: «Преступление и наказание», «Идиот», «Бесы», «Братья Карамазовы» (877 тыс. слов);

Л. Н. Толстой: «Семейное счастье», «Война и мир», «Анна Каренина», «Воскресение» (880 тыс. слов).

Проза А. П. Чехова (в силу преобладания малых форм) разделена на четыре хронологически обусловленных фрагмента, представляющих разные периоды его творчества (всего 1100,6 тыс. слов).

По каждому автору можно привести аргументированное объяснение для включения в перечень именно этих произведений, но самое главное для рассуждения в этой статье, что авторских подкорпусов одинаковое количество – четыре. Каждый текст и каждый автор имеют равные возможности для попадания в выборку для осуществления последующего сопоставительного анализа.

В авторские списки включены слова, которые попадают в число частотных как минимум в двух произведениях данного автора и не более чем в одном чужом произведении. Эти критерии позволяют избежать влияния сюжета конкретного текста на лексическую статистику. Кроме того, они выводят из рассмотрения слова, часто встречающиеся у многих писателей и характерные для многих художественных текстов. Добавим, что из списков выше исключены имена собственные (имена, фамилии персонажей и топонимы), чтобы исключить возможность совсем очевидной авторизации. Сами слова отсортированы в таблице по алфавиту, а не по частоте встречаемости.

Небольшой эксперимент показывает, что среди этих слов специалисты-филологи (и литературоведы,

и лингвисты) явно видят единицы, «окрашенные» тем или иным стилем. Высказаны мнения о том, что даже не отдельные слова, а их соединения, общность создают некий «образ», «лицо» автора. В итоге лексические ряды распределяются по авторам следующим образом:

Автор 1 – А. П. Чехов,

Автор 2 – Ф. М. Достоевский,

Автор 3 – И. А. Гончаров,

Автор 4 – И. С. Тургенев,

Автор 5 – Л. Н. Толстой.

На наш субъективный взгляд, наиболее очевидным является определение А. П. Чехова и Ф. М. Достоевского. При этом мы, естественно, не претендуем здесь на абсолютную истину, и у каждого эксперта могут быть свои соображения о связи отдельных лексем и авторских стилей.

Использовать такие «маркированные» идиостилем языковые данные (естественно, в расширенном варианте) можно по-разному. Например, А. И. Новиков в книге «Семантика текста и ее формализация» предлагал возможный подход к системному выделению содержания текста, включающий построение денотативной структуры текста, выделение ключевых слов, определение связей денотатов [Новиков 1983: 173]. Естественным основанием, на котором возможно выявление признаков концептуальной системы автора, является тематическая (или идеографическая) группировка лексики.

«Концепция семантического пространства языка, – пишет Л. Г. Бабенко, – показывает, что уже невозможно изучение лексических множеств без выявления семантических связей и расстояний между ними» [Бабенко 2004: 19]. Поэтому семантическая разметка контекстов частотных слов и идеографический анализ приводит к построению индивидуальных литературных тезаурусов (см. [Михеев 2003, 2010; Мухин 2010]). Даже приведенные нами авторские небольшие ряды по 30 слов показывают выраженность в идиостиле Достоевского лексики категории «Право», а у Чехова – социальных статусов и отношений, сниженность лексики в творчестве того и другого писателя. Можно обратить внимание на характерные для Тургенева уменьшительно-ласкательные номинации (*ветерок, голосок, дождик, крылечко, тучка*) и другие детали. В предложенных списках явно различается или отчасти пересекается авторская эмотивная лексика: Чехов (*вялый, жутко, нервно, обидно, томить*), Достоевский (*вопл, раздражаться, сострадание, унижение*), Гончаров (*влечение, дружеский, насилие, уныло*), Тургенев (*несчастливый, предаваться, робость, унылый*), Толстой (*веселие, интересоваться, неприятность, радоваться, сочувствовать, счастливо*). Эти и другие наблюдения – естественно, с привлечением контекстологического анализа – могут стать отправной точкой для различных идиостилевых исследований.

Такой материал позволяет более объективно определять словники авторских словарей. Ю. Н. Караулов в предисловии к Словарю языка Ф. М. Достоевского (получившему подзаголовок «Идио-гlossарий») писал, что «первая задача, с которой столкнулись составители при построении базового словаря, – определение состава его словника, списка идио-гloss. На первом этапе основу выделения идио-гloss составил метод экс-

пертных оценок. Оценка давалась, исходя из знания текстов и убежденности эксперта, что выделенное им слово типично для Достоевского, что оно отражает некоторое существенное явление интеллектуально-духовной, эмоционально-душевной или материальной жизни, что оно может быть включено в качестве структурообразующего элемента в картину мира автора и что без этого слова трудно или даже невозможно адекватно отобразить содержание соответствующего текста» [Караулов [http](#)]. Несмотря на дальнейшее «тестирование» и «формально-аналитическую проверку» материала [там же] параметр «убежденности» эксперта в таком статусе слова формализовать затруднительно.

Исходный сравнительно-статистический анализ дает лексикографу неинтуитивную базу для формирования словника, а также набора значимых для автора концептов, образов, символов.

Итак, представление об индивидуально-авторской частотной лексике выводит нас на систематизированное представление концептуальных систем авторов. Такой аспект изучения идиостилей можно назвать лексико-парадигматическим.

Однако для идиостиля автора крайне важным также является индивидуальное употребление языковых единиц (т. е. синтагматический аспект). Как пишет В. А. Пищальникова, индивидуальные речевые «трансформации (операции речевой деятельности) обусловлены эстетической целью и отражают особенности эстетической деятельности автора» [Пищальникова 1992: 44]. Формализовать с помощью статистического анализа специфику авторской сочетаемости слов значительно сложнее, чем выявить особенности индивидуального лексикона.

3. Индивидуальные особенности авторской лексической сочетаемости

Авторская сочетаемость слов давно признана в филологии одним из важнейших признаков индивидуального стиля (см. работы Л. Г. Бабенко, Н. С. Болотновой, В. П. Григорьева, М. Ю. Михеева, В. А. Пищальниковой и многих др.). В научно-популярном дискурсе мы можем встретить такие конструкции, как «индивидуальная работа автора со словом (языком)», «особое (нетипичное) употребление слов» и т. п. По сути, давно обсуждается идея о том, что нестандартные сочетания слов порождают новые смыслы и создают эффекты воздействия на читателя. Или, говоря по-другому, оригинальная авторская сочетаемость слов способствует приращению текстовых смыслов.

В то же время остро стоит вопрос о систематизированном представлении синтагматического материала. Представим простую арифметику. Например, в романе «Война и мир» примерно 450 тыс. слов, и каждое из них взаимодействует с рядом других слов как минимум внутри предложения. В этом тексте около 29,5 тыс. предложений, и, значит, средняя длина предложения составляет около 15 слов. Если подсчитать только возможные пары слов, то в итоге количество межсловных связей в пределах одного предложения в таком произведении будет равно трем миллионам! Иначе говоря, системное представление лексической сочетаемости всегда ранее сталкивалось с труднопреодолимым ба-

рьером в виде множественности синтагматических связей.

Есть и еще одна, уже качественная, проблема. При таком количестве материала затруднительно проводить «вручную» семантический анализ, а его приемлемая автоматизация вряд ли возможна даже в отдаленном будущем. В итоге многочисленные филологические исследования по этой тематике построены на анализе частных идиостилей и частных контекстов, но не имеют системного характера.

Современные исследования лексической сочетаемости получили новый импульс в связи с обращением к так называемым «лексическим n -граммам», а в самом простом варианте – «лексическим биграммам» (контекстным парам слов), которые могут быть претендентами на устойчивые сочетания, коллокации, сложные термины – см. [АОТ [http](#); Браславский, Соколов 2008; Ягунова, Пивоварова 2010]. Поиск по n -граммам возможен также в Национальном корпусе русского языка¹.

Для нас важно, что из большого массива биграмм путем сопоставления авторских корпусов можно вывести индивидуальные синтагматические представления [Мухин 2015, 2019], т. е. формализовать специфику авторской сочетаемости слов. В то же время, если в объемных произведениях количество биграмм измеряется миллионами, необходимы прозрачные критерии ограничения материала.

Именно здесь нам пригодится перечень слов, которые являются общеупотребительными и частотными у многих авторов. В самом деле, чтобы выявлять особенности лексической сочетаемости, необходимо иметь единое базовое основание для сопоставления индивидуальных стилей. Исходя из этого можно систематизировать контекстное окружение слов, которые часто употребляются всеми авторами. Напомним, что в перечень таких слов входят лексемы *говорить, человек, знать, еще, рука, теперь, уже, лицо, глаз, хотеть, видеть, дело, вдруг, спрашивать, думать, очень, понимать, два, жизнь* и др. Реализация этой идеи также предполагает проведение сопоставительного анализа – в данном случае контекстных партнеров этих слов. Чем больше у автора оригинальных контекстов со словом, например, *жизнь*, по сравнению с другими писателями, тем, соответственно, более значима сочетаемость этого слова для авторского идиостиля. Таким образом, источником синтагматического анализа становится систематизация и сопоставительный анализ биграмм, в которые обязательно входит одно из слов, часто употребляемых всеми (или многими) авторами.

Рассмотрим пример такого анализа на материале корпуса прозаических произведений XX в. Так, в результате сопоставления с другими авторами было выявлено, что, судя по набору оригинальных контекстных партнеров, в произведениях Владимира Набокова особой и статистически выраженной лексической сочетаемостью характеризуются наречия *вдруг, опять, еще и ужé*. Сама по себе констатация этого факта не является чем-то особенным, хотя наречия в романах Владимира Набокова составляют 11,8% от всех знаменатель-

¹ Поиск по биграммам возможен здесь: URL: www.ruscorpora.ru/new/search-ngrams_2.html.

ных словоформ, что значительно больше, чем у многих других авторов. Тем более показательна вариативность сочетаемости слов, часто используемых всеми писателями, по которой произведения Набокова существенно опережают тексты его современников. В особенности это касается слова *вдруг* (94 оригинальных контекста). Для сравнения: в романах М. А. Шолохова, несмотря на их объем, – 19 таких случаев, а, например, у М. А. Булгакова и А. Платонова оригинальных контекстов этого слова просто нет. Иначе говоря, в творчестве каждого автора специфическую активность в плане сочетаемости с контекстными партнерами проявляет некоторая часть общего частотного фонда. Чтобы показать особенности этого материала, обратимся к обобщению реальных контекстов, извлеченных из произведений В. Набокова.

Будучи многозначным, наречие *вдруг* употребляет В. Набоковым преимущественно в основном значении – по Большому толковому словарю [БТС 2004], «внезапно, неожиданно». Оригинальная по сравнению с другими авторами сочетаемость слова *вдруг* не является в текстах Набокова бесконечно широкой и сводится к словам *замечать/заметить, вставать/встать, понимать/понять, услышать, звук* и др., которые участвуют в перечисленных ниже семантических отношениях. Основная семантика слова *вдруг* в целом проявляет интерес писателя к быстро протекающим неожиданным, случайным процессам и в описаниях, приведенных ниже, обозначается оператором «внезапность». Рассмотрим эти семантические отношения.

- Внезапность (*вдруг*) ↔ восприятие (слова *замечать/заметить, услышать, звук*).

Чаще всего неожиданными для набоковских персонажей оказываются зрительные впечатления, что отражено в различных произведениях: *Вдруг Кречмар заметил, что она жарко покраснела и встала* (КО¹); *Он заметил вдруг, что Турати уже не сидит, а стоит, заломив руки* (ЗЛ); *Тут Федор Константинович вдруг заметил скорбно-проникновенный, обремененный сочувствием взгляд Чернышевской...* (Д); *Драйвер вдруг с ужасом заметил, что на нем [манекене] галстук* (КДВ).

Неожиданными могут быть и звуки: *Однажды он [Горн] поперхнулся, Кречмар... вдруг услышал в конце овального стола странный звук – как будто шумное человеческое придыхание* (КО); *Франц, с беззвучным стоном откидываясь назад, услышал вдруг, как Марта хрипло засмеялась, прочистила горло и засмеялась опять* (КДВ); *Он услышал вдруг шум голосов, воющий звук парадной двери...* (ЗЛ); *И вдруг Лужин отчетливо услышал за своей спиной особый, деревянно-раскатытый звук, от которого стало жарко и невпопад стукнуло сердце* (ЗЛ). Во многих контекстах отражается, как неожиданное зрительное или слуховое впечатление вызывает эмоцию или догадку персонажа, т. е. внезапно пришедшее понимание чего-либо важного. Эту идею дополняет следующее соотношение:

- Внезапность (*вдруг*) ↔ понимание (глагол *понимать/понять*).

Неожиданность понимания, о которой пойдет речь, связана не просто с частным событием, рядовым интеллектуальным процессом: так человек может внезапно постичь смысл стихотворения, решение задачи и т. п. Герои же Набокова *вдруг* прозревают, постигают суть жизненно важных обстоятельств, фатальный смысл событий и т. п.: «*Опять галлюцинации*», – *тревожно подумал Кречмар и вдруг понял, что именно так его тревожило ночью, – да-да, вот эти странные звуки...* (КО); *Дом он, впрочем, узнал, – и опять были гости, гости, – но вдруг Лужин понял, что он просто вернулся в недавний сон...* (ЗЛ). Подобные конструкции можно найти в разных произведениях, и это говорит о том, что такие ситуации неслучайны, концептуально значимы для автора: *И вдруг она поняла, что нарочно медлит, стоя в пижаме перед зеркалом...* (ЗЛ); *И вдруг она поняла: пиджак...* [пиджак – символ мужа в бредовом состоянии героини] (КДВ); *...и вдруг она поняла, что тоже ищет формулу, официальное воплощение чувства, а дело совсем не в том* (ЗЛ).

- Внезапность (*вдруг*) ↔ движение (глагол *вставать/встать*).

Если неожиданное восприятие часто является причиной сильной эмоции или спонтанно принятого решения, то резкое движение, которое совершает персонаж, – следствием: *Она вдруг резко встала, почувствовала, что вот сейчас задохнется* [от ненависти к мужу]... (КДВ); *Вдруг он [Н. Г. Чернышевский] встал в ал, решив повидать ее немедленно; был, представим себе, октябрьский вечер, летели тучи...* (Д); *Мать молчала, – и вдруг, после второго блюда, встала и, стараясь скрыть дрожащее лицо, повторяя шепотом, что «это ничего, ничего, сейчас пройдет», – поспешно вышла* [о переживании измены мужа] (ЗЛ). Слово *вдруг* в этих случаях также является маркером важности, иногда судьбоносности происходящего.

На фоне рассмотренных контекстов показательно, что, по нашим данным, слово *вдруг* имеет несколько оригинальных системных контекстных партнеров только в текстах М. Шолохова: *вспыхнуть* (о свете и выражении эмоций), *дрогнуть* и *ощущать*, – и можно считать, что количество этих контекстов (в каждом случае 4–5) при большом суммарном объеме шолоховских произведений не является значимым для авторской синтагматики.

Представленное описание, основанное на предварительном статистическом сопоставлении, фактически является фрагментом словаря лексической сочетаемости в романах Владимира Набокова. Типичные для автора конструкции типа *вдруг понял, вдруг она поняла, вдруг она понимала* в массе своей вряд ли могут быть отмечены при обычном читательском восприятии текста. Между тем они часто возникают в особых фрагментах художественного текста – когда решается судьба героя, происходит сюжетный перелом и т. п.

Как пишет М. Ю. Михеев в статье «Компиляция или... языковые клише?», из текста «извлекаются характерные – отличающиеся от обычных – сочетания слов... кажущиеся достаточно а) редкими у других авторов и б) повторяющиеся в „контрольном массиве“ у того/тех авторов, принадлежность чьей руке данного текста мы хотим проверить» [Михеев 2010 http]. Исходное статистическое сопоставление позволяет объективировать то, что кажется читателю, – тем более, если

¹ Здесь и далее приняты сокращения названий романов В. В. Набокова: КО – «Камера обскура», ЗЛ – «Защита Лужина», Д – «Дар», КДВ – «Король, дама, валет».

отмеченное явление неоднократно повторяется в разных произведениях.

На основании предложенной модели сравнительного статистического и контекстологического анализа в данный момент проводится исследование русской прозы XIX в. – произведений уже перечисленных выше Л. Н. Толстого, Ф. М. Достоевского, И. С. Тургенева, А. И. Гончарова и А. П. Чехова. Планируется выявить синтагматические приоритеты пяти авторов. Для этого нужно определить набор слов, имеющих в творчестве каждого из них наибольшее количество употреблений с оригинальными контекстными партнерами, а формализация этой «синтагматической активности» позволяет построить синтагматический профиль – оригинальное количественное соотноше-

ние контекстных партнеров избранных частотных слов в различных произведениях конкретного писателя.

Выявленные семантические отношения внутри авторских лексических биграмм предполагается обобщить и представить в виде совокупности семантических описаний – своеобразного авторского синтагматического словаря, характерного для пяти писателей-классиков. Проект предполагает филологическую интерпретацию этого массива на фоне существующих филологических исследований классической литературы XIX в. В ходе этой интерпретации возможно подтверждение или, наоборот, коррекция выводов об идиостилиях, сделанных учеными ранее на основании значительно менее формализованного анализа.

ЛИТЕРАТУРА

- AOT (Автоматическая обработка текста). Поиск по биграммам. – URL: www.aot.ru/demo/bigrams.html (дата обращения: 02.10.2019).
- Бабенко Л. Г. Концепция идеографических словарей, разрабатываемая Уральской семантической школой // *Образ человека и человеческий фактор в языке: словарь, грамматика, текст*. – Екатеринбург, 2004. – С. 11–20.
- Браславский П. И., Соколов В. А. Сравнение пяти методов извлечения терминов произвольной длины // *Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конф. «Диалог»*. – М.: РГГУ, 2008. – Вып. 7 (14). – С. 67–74.
- БТС – Большой толковый словарь русского языка / под ред. С. А. Кузнецова. – СПб.: Норинт, 2004. – 1534 с.
- Гребенников А. О. Частотный словарь и образ мира писателя // *Словоупотребление и стиль писателя*. – СПб., 2006. – Вып. 3. – С. 156–163.
- Караулов Ю. Н. Предисловие // *Словарь языка Достоевского*. – URL: slovari.ru/default.aspx?s=0&p=307 (дата обращения: 02.10.2019).
- Михеев М. Ю. В мир Платонова – через его язык. Предположения, факты, истолкования, догадки. – М.: Изд-во МГУ, 2003. – 408 с.
- Михеев М. Ю. Компиляция или... языковые клише? Сравнивая характерные для авторского стиля наборы словосочетаний // *Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной международной конф. «Диалог»*. – М., 2010. – URL: www.dialog-21.ru/dialog2010/materials/html/52.htm (дата обращения: 02.10.2019).
- Мухин М. Ю. Индивидуальная лексическая сочетаемость и ее корпусная формализация // *Язык, культура, ментальность: проблемы и перспективы филологических исследований*. – Курск, 2019. – С. 310–317.
- Мухин М. Ю. Лексическая статистика и концептуальная система автора: М. Булгаков, В. Набоков, А. Платонов, М. Шолохов. – Екатеринбург: Изд-во Урал. ун-та, 2010. – 232 с.
- Мухин М. Ю. Нетипичная лексическая сочетаемость: формализация термина и анализ текста // *Уральский филологический вестник*. – 2015. – Вып. 4. – С. 105–115. – URL: journals.uspu.ru/attachments/article/1036/MAKET_KREATIV.pdf (дата обращения: 02.10.2019).
- НЧС – Новый частотный словарь русской лексики / под ред. О. Н. Ляшевской, С. А. Шарова. – URL: dict.ruslang.ru/freq.php (дата обращения: 02.10.2019).
- Новиков А. И. Семантика текста и ее формализация. – М.: Наука, 1983. – 216 с.
- Пищальникова В. А. Проблема идиостилия. Психолингвистический аспект. – Барнаул: Изд-во Алтайск. ун-та, 1992. – 73 с.
- Сулименко Н. Е. Текст и аспекты его лексического анализа. – М.: Флинта: Наука, 2009. – 396 с.
- Ягунова Е. В., Пивоварова Л. М. Природа коллокаций в русском языке. Опыт автоматического извлечения и классификации на материале новостных текстов // *Сб. НТИ. Сер. 2*. – М., 2010. – № 6. – URL: medialing.spbu.ru/upload/files/file_1394529742_4311.pdf (дата обращения: 02.10.2019).

REFERENCES

- AOT (*Avtomaticheskaya obrabotka teksta*). *Poisk po bigrammam* [NLP (Natural Language Processing). Bigram Search]. URL: www.aot.ru/demo/bigrams.html (mode of access: 02.10.2019).
- Babenco, L. G. (2004). *Kontseptsiya ideograficheskikh slovarei, razrabatyvaemaya Ural'skoi semanticheskoi shkoloj* [The Concept of Ideographic Dictionaries Developed by the Ural Semantic School]. In *Obraz cheloveka i chelovecheskii faktor v yazyke: slovar', grammatika, tekst*. Ekaterinburg, pp. 11–20.
- Braslavskii, P. I., Sokolov, V. A. (2008). *Sravnienie pyati metodov izvlecheniya terminov proizvol'noi dliny* [Comparison of Five Methods for Variable Length Term Extraction]. In *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Po materialam ezhegodnoi mezhdunarodnoi konferentsii «Dialog»*. Moscow, Rossiiskii gosudarstvennyi gumanitarnyi universitet. Issue 7 (14), pp. 67–74.
- Grebennikov, A. O. (2006). *Chastotnyi slovar' i obraz mira pisatelya* [Writer's Word Lists by Frequency and Perception of the World]. In *Slovoupotreblenie i stil' pisatelya*. St. Petersburg. Issue 3, pp. 156–163.
- Karaulov, Yu. N. *Predislovie* [Foreword]. In *Slovar' yazyka Dostoevskogo*. URL: slovari.ru/default.aspx?s=0&p=307 (mode of access: 02.10.2019).
- Kuznetsov, S. A. (Ed.). (2004). *Bol'shoi tolkovyi slovar' russkogo yazyka* [The Great Explanatory Dictionary of the Russian Language]. St. Petersburg, Norint. 1534 p.
- Lyashevskaya, O. N., Sharov, S. A. (Eds.). *Novyi chastotnyi slovar' russkoi leksiki* [The New Frequency Dictionary of the Russian Language]. URL: dict.ruslang.ru/freq.php (mode of access: 02.10.2019).
- Mikheev, M. Yu. (2003). *V mir Platonova – cherez ego yazyk. Predpolozheniya, fakty, istolkovaniya, dogadki* [Into the World of Platonov – Through His Language. Assumptions, Facts, Interpretations, Conjectures]. Moscow, Moskovskii gosudarstvennyi universitet. 408 p.
- Mikheev, M. Yu. (2010). *Kompilyatsiya ili... yazykovye klishe? Sravnivaya kharakternye dlya avtorskogo stilya nabory slovosochetanii* [Compilation or... Language Cliches? Comparing Author-Specific Phrase Sets]. In *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Po materialam ezhegodnoi mezhdunarodnoi konferentsii «Dialog»*. Moscow. URL: www.dialog-21.ru/dialog2010/materials/html/52.htm (mode of access: 02.10.2019).

Mukhin, M. Yu. (2010). *Leksicheskaya statistika i kontseptual'naya sistema avtora: M. Bulgakov, V. Nabokov, A. Platonov, M. Sholokhov* [Lexical Statistics and Author's Conceptual System: M. Bulgakov, V. Nabokov, A. Platonov, M. Sholokhov]. Ekaterinburg, Ural'skii federal'nyi universitet. 232 p.

Mukhin, M. Yu. (2015). Netipichnaya leksicheskaya sochetaemost': formalizatsiya termina i analiz teksta [Atypical Lexical Compatibility: Term Formalization and Text Analysis]. In *Ural'skii filologicheskii vestnik*. Issue 4, pp. 105–115. URL: journals.uspu.ru/attachments/article/1036/MAKET_KREATIV.pdf (mode of access: 02.10.2019).

Mukhin, M. Yu. (2019). Individual'naya leksicheskaya sochetaemost' i ee korpusnaya formalizatsiya [Individual Lexical Compatibility and Its Corpus Formalization]. In *Yazyk, kul'tura, mental'nost': problemy i perspektivy filologicheskikh issledovaniy*. Kursk, pp. 310–317.

Novikov, A. I. (1983). *Semantika teksta i ee formalizatsiya* [Semantics of the Text and Its Formalization]. Moscow, Nauka. 216 p.

Pishchal'nikova, V. A. (1992). *Problema idiostilya. Psikholingvisticheskii aspekt* [The Problem of Idiostyle. Psycholinguistic Aspect]. Barnaul, Altaiskii gosudarstvennyi universitet. 73 p.

Sulimenko, N. E. (2009). *Tekst i aspekty ego leksicheskogo analiza* [Text and Aspects of its Lexical Analysis]. Moscow, Flinta, Nauka. 396 p.

Yagunova, E. V., Pivovarova, L. M. (2010). Priroda kollokatsii v russkom yazyke. Opyt avtomaticheskogo izvlecheniya i klassifikatsii na materiale novostnykh tekstov [The Nature of Collocations in Russian Language. Experience of the Automatic Extraction and Classification on the Material of News Texts]. In *Sb. NTL, Seriya 2*. Moscow. No. 6. URL: medialing.spbu.ru/upload/files/file_1394529742_4311.pdf (mode of access: 02.10.2019).

Сведения об авторах

Мухин Михаил Юрьевич – доктор филологических наук, профессор кафедры фундаментальной и прикладной лингвистики и текстологии, Уральский федеральный университет им. первого Президента России Б. Н. Ельцина (Екатеринбург).

Адрес: 620083, Россия, Екатеринбург, пр. Ленина, 51.

E-mail: mikhail.mukhin@urfu.ru.

Мухин Николай Юрьевич – кандидат филологических наук, доцент кафедры иностранных языков и перевода, Уральский федеральный университет им. первого Президента России Б. Н. Ельцина; доцент кафедры теории и методики обучения лиц с ограниченными возможностями здоровья, Уральский государственный педагогический университет (Екатеринбург).

Адрес: 620002, Россия, Екатеринбург, ул. Мира, 19; 620017, Россия, Екатеринбург, пр. Космонавтов, 26.

E-mail: nfly2011@yandex.ru.

Author's information

Mukhin Mikhail Yurievich – Doctor of Philology, Professor of the Department of Fundamental and Applied Linguistics and Textology, Ural Federal University named after the first President of Russia B. N. Yeltsin (Ekaterinburg).

Mukhin Nikolay Yurievich – Candidate of Philology, Associate Professor of the Department of Foreign Languages and Translation, Ural Federal University named after the first President of Russia B. N. Yeltsin; Associate Professor of the Department of Theory and Methods of Teaching People with Disabilities, Ural State Pedagogical University (Ekaterinburg).