

М. Ю. МУХИН, Е. Р. ФИЛАТОВА
*(Уральский федеральный университет им. первого Президента
России Б.Н. Ельцина, г. Екатеринбург, Россия)*

УДК 821.161.1-3(Чехов А. П.)
ББК Ш33(2Рос=Рус)5-8,44

ПРОЗА, ДРАМАТУРГИЯ, ПУБЛИЦИСТИКА И ПЕРЕПИСКА А.П. ЧЕХОВА В СОПОСТАВИТЕЛЬНО- СТАТИСТИЧЕСКОМ РАКУРСЕ

Аннотация: В статье рассматриваются лексические особенности прозы, драматургии, публицистики и переписки А.П. Чехова. На базе объемного корпуса текстов писателя (более 2,3 млн. слов) проведен сравнительный статистический анализ лексики. Проведена работа по лемматизации текстов, грамматической и частичной семантической разметке. Выявлены лексемы, наиболее характерные для каждого из четырех подкорпусов, а также слова, частотные для творчества Чехова в целом. Проведен тематический анализ, составлен индивидуально-авторский литературный тезаурус частотной лексики, даны идиостилевые комментарии к наиболее важным лексическим группам — таким, как «Эмоции», «Оценка», «Движение», «Родственные и семейные отношения», «Право», «Быт», «Медицина» и т.п. Сделаны выводы о различиях в лексическом составе и, как следствие, функциональных особенностях прозы, драматургии, публицистики и переписки А.П. Чехова. При этом письма и публицистические произведения как лингвистические источники выступают наравне с художественными произведениями. Проведенный анализ является экстенсивным с точки зрения большого объема материала и задает векторы направления филологического исследования. Полученные данные открывают перспективы для дальнейшего идиостилевого анализа творчества — и в целом наследия — писателя.

Ключевые слова: лексическая статистика, текстовый корпус, семантический анализ, А.П. Чехов, идиостиль, проза, драматургия, публицистика, переписка

1. Идиостилевые параметры все чаще получают сегодня статистическую аргументацию. Стилometрия развивается: выходят авторские частотные словари [Минц 1999; Частотный грамматико-семантический словарь... А.П. Чехова 2012; Частотный словарь рассказов А.П. Чехова 1999; Частотный словарь рассказов Л.Н. Андреева 1999; Шайкевич 2003], для изучения текста применяют математические модели и проводят сопоставитель-

ные исследования творчества разных авторов.

Художник слова часто является только поэтом, или прозаиком, или драматургом. Пробуя свои силы в пределах другого рода литературы, он редко создает объемно и много. К примеру, поэт Борис Пастернак как прозаик написал один большой роман «Доктор Живаго» и несколько других небольших произведений. Специализация литературного творчества является причиной того, что сопоставительное филологическое исследование обычно проводится на однородном материале творчества разных авторов или – в пределах наследия одного автора – на материале однотипных в родовом смысле произведений.

В то же время статистические модели более эффективны в применении к объемному материалу. В этом смысле любопытным межстилевым примером является для нас наследие А.П. Чехова. Объем его прозы, драматургии, публицистики и эпистолярных текстов достаточно большой, чтобы можно было говорить о статистической точности. Таким образом, мы предполагаем, что на основании лексико-статистического сопоставления этих четырех ипостасей одного писателя можно выявить как универсально-авторские черты, так и индивидуальные функционально-стилистические признаки.

Творчество писателя уже становилось объектом статистических измерений. В 1999 г. в СПбГУ опубликован Частотный словарь рассказов А.П. Чехова. В 2012 г. в Лаборатории общей и компьютерной лексикологии и лексикографии МГУ в результате огромной и кропотливой работы вышел Частотный грамматико-семантический словарь языка художественных произведений А.П. Чехова. Однако и этот замечательный словарь отражает лексический состав только прозы и драматургии. Нас же интересуют все направления чеховского письма, и, кроме того, в основе предлагаемой методологии – статистическое сопоставление нескольких лексических массивов.

Подкорпус прозы Чехова составляет 1 100 792 слова (47 % материала), драматургии – 176 629 слов (7,5 %), публицистики — 182 650 (7,8 %) и, наконец, переписки – 882 263 слова (37,7 %). Общий объем корпуса чеховских текстов, которым мы располагаем, – 2 342 334 слова.

Тексты лемматизированы при помощи программы Mystem

(URL: <https://tech.yandex.ru/mystem/>) с функцией контекстного снятия омонимии и переведены в базу данных. Следует заметить, что при наличии ошибок автоматизированного разрешения грамматической омонимии общее качество этой процедуры следует признать удовлетворительным, а некоторые системные ошибки были скорректированы вручную.

2. Аналогично модели сопоставительного статистического анализа, предложенной в [Мухин 2010], из каждого подкорпуса выбрано по 2000 самых частотных знаменательных лемм. Одинаковые по объему выборки позволяют скомпенсировать количественную разницу в исходных лексических массивах. Кроме того, отказ от рассмотрения служебных слов и местоимений объясняется «семантикоцентричностью» работы, задача которой – связать лексические частоты и текстовые стилевые черты.

Полученные списки сопоставлены, и выявлены слова, которые попадают в набор самых частотных только в одном из подкорпусов. Для прозы этот список составили 337 слов, для драматургии – 503, для публицистики – 613 и для переписки – 518 слов. Само такое распределение кажется вполне объяснимым: стиль прозы является лексически наиболее нейтральным с точки зрения общепотребительной лексики, а стиль публицистики и эпистолярного жанра – наоборот, что «маргинализует» перечень оригинальных слов и увеличивает его объем.

Кроме того, был выбран «универсальный» для писателя лексический массив (775 слов, часто встречающихся во всех чеховских подкорпусах). В верхней части этого списка ожидаемо фигурируют лексемы *быть, говорить, человек, мочь, еще, уже, сказать, знать, теперь, давать, год, рука, день, писать, очень, идти* и т.д. Эта лексика не является индивидуально-авторской, ее общезыковая частота высока, и она попадает в «пик» частотности практически любого протяженного текста. Однако в выбранном массиве есть слова, на которые нужно обратить внимание с точки зрения изучения именно чеховского идиостиля. Сравним относительные частоты чеховских лексем с данными, которые приводятся в «Новом частотном словаре русской лексики» [электрон. ресурс]. Некоторые из них по своей относительной частоте существенно (от 3 до 300 раз по индексу *ipt*, т.е. на миллион словоупотреблений) превосходят общезыковые

показатели.

Конечно, отчасти мы имеем дело с временными различиями в языке, так как в этой группе есть устаревшие варианты слов, историзмы и просто лексемы, употребительность которых за век значительно снизилась: *счастье, жалование, уезд, прислуга, верста, управа, господин, дама, непременно, почтенный, казенный*. Однако активное употребление Чеховым следующих групп слов нельзя объяснить приметами времени. Это широкий спектр оценочной лексики: *прелестный, милый, великолепный, умный, благородный, честный, порядочный, дурной, дурно, ужасный, ужасно, чепуха, пустяк, обыкновенно, необыкновенный*. Это слова, связанные с творчеством и перепиской (*рассказ, пьеса, публика, письмо, телеграмма*), лечением и здоровьем (*доктор, больной, здоровье, здоровый*), семейными отношениями (*женатый, семейство, невеста*), бытом (*чай, обед, сытый, табак, шляпа, шапка*), финансами (*дорого, даром, копейка*), погодой и восприятием тепла/холода (*погода, дождь, тепло, холодно*). Занятно, что в небольшой перечень «универсальных» чеховских эмотивов (*скука, скучно, скучный, виноватый*) попали три репрезентанта концепта «скука», значимость которого для чеховского творчества понимает любой филолог (см., например, об этом [Сопова 2014; Маханова, электрон. ресурс и др.]). Кроме того, значительно отличается от общеязыковой частота употребления писателем таких слов, как *сад, пьяный, громадный, лошадь, тотчас, мало-помалу*. Среди перечисленных слов много и других репрезентантов чеховских концептов – сошлемся здесь на работу [Виноградова 2008].

Рассматривая эти группы, мы намеренно отказались от анализа глагольной лексики. Специфика лемматизатора *Mystem* заключается в том, что он приводит глаголы совершенного вида к видовым парам несовершенного вида. Поэтому сопоставление наших данных с «Новым частотным словарем», в котором видовые пары разведены, не является корректным. В решении этой грамматической проблемы мы видим одну из частных перспектив исследования.

Подобная выборка имеет свои перспективы для дальнейшего филологического описания чеховского языка, ее можно использовать в качестве базы для идиостилевого словаря, в котором

рассматриваются индивидуально-авторские, а не общезначимые лексические признаки.

3. Перейдем к анализу частотных слов, характерных для четырех подкорпусов (из списков исключены имена собственные). Основу четырех оригинальных лексических списков составляют следующие слова:

- Проза: *крикнуть, проговаривать, послушаться, засмеяться, побежать, почувствовать, поручик, рукав, пробормотать, пятно, покраснеть, коляска, облако, схватывать, спальня, проходиться, бровь, погода, повертываться, неподвижно, побледнеть, удивление...*

- Драматургия (здесь топ списка естественным образом возглавляют фамилии и имена персонажей, но среди них встречаем и другие слова): *пауза, хор, занавес, целоваться, прочь, рыдать, замолчать, живо, уставать, напевать, душечка, отставать, замучиваться, испуганно, танцевать, мошенник, налево, вбегать, дружба, радостно...*

- Публицистика: *поселенец, ссыльный, селение, округ, арестант, колония, надзиратель, каторга, гилек, население, смотритель, тюремный, побег, житель, материк, южный, японец, возраст, тайга, сельский, приказ, долина, японский, администрация...*

- Переписка: *оборот, поклон, высылать, конверт, март, благополучный, привет, корректура, преданный, побывать, слать, сборник, издание, февраль, телеграфировать, повидаться, октябрь, гонорар, хранить, сердечный, печатать, экземпляр ...*

До проведения семантического анализа можно предположить, что наборы индивидуальных частотных слов в переписке должны отражать реальную жизнь писателя и круг его общения, в драматургии – специфику в первую очередь персонажной речи и авторских ремарок, в прозе – сочетание авторской и персонажной речи, с явным преобладанием речи авторской, в публицистике – социальные аспекты, которые волновали Чехова (в первую очередь связанные с его поездкой на остров Сахалин).

4. На следующем этапе проведена идеографическая классификация частотных слов, образующая литературный тезаурус, и выделены доминантные тематические группы лексики, наиболее характерные для сравниваемых подкорпусов. Семантическая

идентификация произведена с опорой на идеографические словари Уральской семантической школы (руководитель – проф. Л.Г. Бабенко).

Для прозы А.П. Чехова количественно наиболее актуальны сферы эмоций, речи и звучания, быта, движения, физиологических действий, оценки, социального статуса, частей тела человека и животного (*бровь, локоть, затылок, подбородок, морда, рожа, ладонь, борода, головка, крыло, глазик, ножка*).

Для драматургии – оценки, эмоций, религиозной и медицинской сфер, движения, звучания, родственных и семейных отношений. Кроме того, специфика рода литературы отражена в частотности таких слов, как *пауза, занавес, действующий, рампа, хор*.

Поскольку проза и драматургия, как направления художественной литературы, тематически близки, сопоставим лексические классы, которые для них актуальны.

Собственно эмоции, чувства в прозе и пьесах достаточно похожи: равнодушие и спокойствие, испуг, радость и грусть, раздражение и злость и т.д. Сравним: *удивление, ненависть, злоба, грустный, отвращение, рассердиться, испуганный, радостный, почувствовать, равнодушный, равнодушно* (проза); *испуганно, испугать, радостно, взволновать, раздражённо, нетерпеливо, покойно, спокойный, вспылить, злость, возмущать, смущённо, разлюбить, томиться, взволнованно, хладнокровно, тосковать, растрогать, меланхолия, гордость, самолюбие* (драматургия). Резко отличается проза разработанностью лексики внешнего выражения эмоций: *засмеяться, вспыхивать, сконфузиться, конфузиться, покраснеть, побледнеть, краснеть, бледнеть, нахмуриваться, зарыдать, плач, захохотать, поморщиться, морщиться, хмуриться, закричать, всхлипывать, выбежать, задрожать, замахать, взвизгивать* и др. В пьесах эту функцию выполняет разве что слово *рыдать* да еще ряд глаголов совмещенной семантики: *вбегать, вырывать, плясать*. В самом деле, визуальность драматургии, возможность актерского внешнего выражения делает необязательным использование лексических средств выражения эмоций. В прозе же это необходимо, иначе эмоциональное состояние персонажа читателю представить трудно.

По этой же причине, вероятно, в прозе значительно важнее, чем в драматургии, лексика речи и звучания: *крикнуть, проговаривать, послышаться, пробормотать, помолчать, забормотать, шёпот, прошептать, зашептать, закричать, вслух, запевать, проворчать, пение, доноситься, произносить, выговаривать, громкий, умолкнуть, кашлянуть, крякнуть* и др. В драматургии этот перечень значительно более бедный: *замолчать, напевать, высказывать, бредить, клеветать, слышный, щёлкать, звонить, свистеть, наигрывать, трещать*.

Зато оценочная лексика в драматургии реализована очень са­мо­бытно, так как в диалогах персонажей оценка должна быть эксплицитной: *мошенник, губить, погубить, подлость, подло, мерзавец, безнравственный, бесчестный, пошлость, пошлый, глупец, гадость, невежество, нудный, безобразие, мелочный, мелко, гнусность, неприлично, опротиветь, несправедливость, бесчестный, грешник, грешница, грешно, чучело, психопат* и, наоборот, *умница, умно, красавец, справедливый, великодушный, восхитительный, прелестный, милая, мило, чудно, изящный, молодчина, добродетель, искренность, справедливый, честность*. Этот перечень включает самые разные виды оценки: интеллектуальную, этическую, эстетическую, – отрицательную и положительную. В прозе же есть иные (и часто имплицитные) средства выражения оценки, поэтому этот лексический класс большим объемом и явными особенностями не отличается. Кстати, если опираться на частотную лексику, в своих письмах Чехов использовал не очень большое количество оригинальных оценочных слов. С положительной оценкой в переписке связаны лексемы *благополучный, недурно, симпатичный, хвалить, велико­лепно, превосходно, изумительный, недурной, сносно*, а с отрицательной – *скверный* и *скверно*. Рассмотрение единичных словоупотреблений [Величко 2010] резко увеличивает этот список, но требует подробного контекстологического анализа – в нашем случае на материале 2,3 млн. слов.

Именно в художественных текстах, в отличие от публицистики и писем, оказывается количественно выраженной лексика движения, перемещения. И если в прозе движение может быть значимо «само по себе»: *побежать, пройтись, повёртываться, оглянуться, заходить, нагибаться, направляться, шевелиться,* –

то в драматургии глаголы движения обычно служат выражению эмоции или физического состояния (*танцевать, вбегать, шататься, вырывать, пошатываться, плясать*).

Проза, в отличие от пьес, «наводнена» словами, обозначающими детали быта, обихода, – разнообразными мелочами, заполняющими художественное пространство. Это следующие лексические группы:

– предметы одежды и их части: *пуговица, рукав, цилиндр, воротник, сорочка, туфля*;

– различные предметы бытового обихода: *одеяло, подушка, спичка, трубка, очки, зеркало, сундук, ковёр, карандаш, ложка*;

– помещения, строения и их части: *амбар, ступень, калитка, спальня, потолок, крыльцо, забор, порог*;

– транспортные средства, приводимые в движение лошаадьми: *бричка, телега, сани, экипаж, коляска, воз*.

В публицистике Чехова определяющей и наиболее разработанной является правовая сфера, в которой четко выделяются группы «Человек» (*ссылный, арестант, подсудимый, преступник, ссылнокаторжный, каторжник, осуждённый, рецидивист, испытуемый, поселенец, беглый, виновный* и, с другой стороны, *надзиратель, смотритель, защитник, палач, жанدارм, заседатель, конвойный, полицейский*), «Место заключения» (*колония, камера, карцер, рудники, каторга, ссылка*), «Правовой процесс» (*надзор, заключение, следствие, казнь*) «Предмет, связанный с наказанием и заключением» (*розги, кандалы, плеть*), «Документ» (*приговор, протокол* и т.п.). Ориентация на географическую и фактографическую реалистичность определяет здесь также особую актуальность следующих лексических классов:

– населенный пункт (*селение, округ, поселение, слободка, местность, поселок*),

– природные объекты (*материк, тайга, долина, мыс, почва, устье, пролив, залив, речка, луг, побережье, океан, болото, приток, хребет, равнина, просека, бухта, падь*),

– сельское хозяйство (*сельскохозяйственный, урожай, пахотный, семя, огород, хлебопашество, растение, пшеница, сенокос, сеять, зерно, пашня, солома, сельский, картофель, почва*),

– медицина (*лазарет, цинга, эпидемия, сифилис, оспа*),

– нации и народы, человек по месту проживания (*гиляк, японец, китаец, татарин, москвич, сахалинец, аинка*).

– слова из разных сфер, обозначающие социальный статус человека, род занятий: *служащий, генерал-губернатор, защитник, инспектор, корреспондент, кассир, бухгалтер, агроном, артистка, антрепренёр, певец, плотник, промышленник, клоун* и др.

В целом чеховскую публицистику следует признать крайне оригинальным с лексико-семантической точки зрения подкорпусом.

В письмах А.П. Чехова, непосредственно отражающих жизнь и интересы писателя, наиболее реализованы сферы «Искусство, литературное творчество», «Медицина», «Передача информации», «Оценка», «Финансы», «Эмоции». Специфичной для эпистолярного жанра является группа личных обращений к жене: *дуся, дусик, лошадка, актрисуля*.

К сфере «Искусство, литературное творчество» в переписке относятся как слова, называющие литературные направления и жанры (*проза, сказка, стихотворение, сочинение, фельетон, заметка, рецензия*), так и во многом лексемы, указывающие на специфику издания литературного произведения (*корректурa, издание, печатать, экземпляр, рукопись, оттиск, обложка* и др.). Неслучайны здесь и слова сферы «Финансы»: *гонорар, уплачивать, расход, аванс, зарабатывать, платёж, безденежье, перевод*.

Слова, обслуживающие сферу «Передача информации», также характерны именно для эпистолярного жанра: *высылать, конверт, слать, телеграфировать, заказное (письмо), посылка, марка, бандероль, выписывать, извещать, курьер, адресоваться, переписка, телеграф, высылка, пересылка* и многие др.

Теперь сравним реализацию медицинской сферы и слов, указывающих на здоровье/нездоровье в письмах, драматургии и публицистике. В переписке явно выделяются подгруппы: болезненное состояние или выздоровление (*хворать, нездоровиться, поправляться, выздоравливать*), болезни (*геморрой, плеврит, тиф, инфлуэнца*), симптомы (*кашель, кровохаркание, воспаление*) и слово *клиника*. Проблемы Чехова со здоровьем начались ещё в десятилетнем возрасте, когда он начал чувствовать «тес-

нение в груди»; в гимназические и студенческие годы Чехов болел туберкулезным воспалением брюшины, а с 1884 г. страдал кровотечением из правого легкого. Чехов вел борьбу с болезнью на протяжении всей жизни, и вполне естественно, что эта тема нашла важное отражение в его письмах.

Для других подкорпусов такая лексика менее актуальна. В драматургии встречаем слова, обозначающие симптомы, необходимые, пожалуй, более для внешнего выражения внутреннего состояния (*сердцебиение, обморок, трещать, пошатываться, бредить*), а из названий болезней – только *подагру* и *ревматизм*. Для публицистики актуальны распространенные во времена Чехова и из-за этого социально значимые болезни *цинга, сифилис, оспа*, само слово *эпидемия*, а также место массового лечения – *лазарет*. В итоге сопоставление лексического материала наглядно показывает, как общие функциональные черты конкретного подкорпуса влияют на словоупотребление, и эта картина охватывает большое количество актуальных для каждого массива слов.

Чеховское наследие – это более чем удачный материал, позволяющий судить о лексической разнице литературных форм, к которым обращался один автор. Большой объем корпуса и, значит, повторяемость частотных слов объективируют применение статистических методов. А первичный анализ, основанный на сопоставлении массивов слов, позволяет говорить об универсально-текстовых («внечеховских»), универсально-авторских (идиостилевых) и индивидуальных чертах, которые свойственны каждому подкорпусу. Дальнейшее филологическое развитие такого исследования видится нам в подробном контекстологическом анализе и более глубокой интерпретации полученных данных.

ЛИТЕРАТУРА

Величко Е.В. Оценочность в языковой картине мира А.П. Чехова (на материале писем крымского периода) // Семинар «Мир языка А.П. Чехова», тр. IV междунар. конгресса «Русский язык: исторические судьбы и современность». – М., 2010. URL: http://www.philol.msu.ru/~rlc2010/abstracts/rlc2010_abstracts_sem22.pdf (дата обращения: 21.03.2016).

Виноградова М.С. Реконструкция языковой картины мира позднего Чехова: на материале художественной прозы 1898-1903 гг. Автореф. дис. ... канд. филол. наук. – Тверь, 2008.

Маханова Г.Е. Репрезентация концепта «скука» специализированными структурными схемами в художественной прозе и письмах А.П. Чехова. URL: <http://cyberleninka.ru/article/n/representatsiya-kontsepta-skuka-spetsializirovannymi-strukturnymi-shehami-v-hudozhestvennoy-proze-i-pismah-a-p-chehova> (дата обращения: 21.03.2016).

Минц З.Г. Частотный словарь цикла «Стихи о Прекрасной даме»; Частотный словарь «первого тома» лирики Ал. Блока // Минц З.Г. Блок и русский символизм: избр. тр.: в 3 кн. Кн. 1: Поэтика Александра Блока. – СПб., 1999. С. 581-627; 699-721.

Мухин М.Ю. Лексическая статистика и концептуальная система автора: М. Булгаков, В. Набоков, А. Платонов, М. Шолохов. – Екатеринбург, 2010. 232 с.

Новый частотный словарь русской лексики. Под ред. О.Н. Ляшевской, С.А. Шарова // Словари на основе Национального корпуса русского языка. URL: <http://dict.ruslang.ru/freq.php> (дата обращения: 21.03.2016).

Сопова С.П. Мотивные составляющие концепта «скука» в повести А. П. Чехова «Скучная история» // Сибирский филол. журнал. 2014. № 2. С. 43-47.

Частотный грамматико-семантический словарь языка художественных произведений А.П. Чехова с электронным приложением / Под общ. ред. А.А. Поликарпова. М., 2012.

Частотный словарь рассказов А.П. Чехова. – СПб., 1999.

Частотный словарь рассказов Л.Н. Андреева. – СПб., 2003.

Шайкевич А.Я., Андрущенко В.М., Ребецкая Н.А. Статистический словарь языка Достоевского. – М., 2003.

© Мухин М.Ю., 2016

© Филатова Е.Р., 2016