Тема

Прикладные аспекты квантитативной лингвистики

План работы

1. Проблематика квантитативной лингвистики с теоретической и прикладной точек зрения

2. Основные области приложения структурно-вероятностной модели языка

3. Авторизация текста: пример экспертизы

Литература

# 1. Проблематика квантитативной лингвистики с теоретической и прикладной точек зрения

Название «квантитативная лингвистика» достаточно условно, хотя и довольно широко используется в современной научной литературе. Оно характеризует междисциплинарное направление в прикладных исследованиях, в котором в качестве основного инструмента изучения языка и речи используются количественные или статистические методы анализа. Иногда квантитативная (или количественная) лингвистика противопоставляется комбинаторной лингвистике. В последней доминирующую роль занимает «неколичественный» математический аппарат теория множеств, математическая логика, теория алгоритмов и т.д.

Применение количественных методов при описании функционирования языка мало чем отличается от использования аналогичного инструментария в естественных и гуманитарных науках. Привлечение методов измерения и подсчета языковых реализаций позволяет, однако, существенно модифицировать представление о языковой системе и возможностях ее функционирования. В этом отношении квантитативная лингвистика оказывается важнейшим фактором, влияющим на лингвистическую теорию. Например, в сфере грамматики теоретическая лингвистика, как правило, ограничивается констатацией существования в русском языке системы падежей. Со структурной точки зрения этого, быть может, и достаточно. Между тем за рамками обсуждения остается весьма существенная информация о том, как часто используются различные падежи, какова динамка использования различных падежей с течением времени. Исследование такого рода позволило бы выявить тенденции развития падежной системы и на основе этого даже сформулировать гипотезы о будущем состоянии грамматической системы русского языка.

Другой пример. С системной точки зрения в русском, английском и латинском языках имеется форма именительного падежа единственного числа личных местоимений. Однако в английском языке при глаголе эта форма местоимения практически всегда необходима, в русском — местоимение в этих случаях обычно представлено, а в латыни — как правило, отсутствует. Отсутствие достоверных количественных данных об этих языковых явлениях делает структурное описание явно недостаточным. Разумеется, имеются и смешанные случаи.

Близкие проблемы возникают и в сфере лексики. Обычные толковые словари не помещают в составе словарной статьи информации о частоте использования той или иной лексемы. Это связано с очень большим объемом работы, который надо проделать, чтобы для каждого слова указать хоть какие-то рамки частотности. Для пользователя словаря такая информация может оказаться очень важной, часто решающей для принятия решения об использовании слова. Ср., например, высокочастотные в публицистике идиомы с головы до ног/с ног до головы (48 вхождений на 21 млн словоупотреблений), целиком и полностью (49 вхождений на 21 млн), на все сто (42 вхождения на 21 млн), ровным счетом, ни больше ни меньше (71 вхождение на 21 млн), ни много, ни мало (133 вхождения на 21 млн) и весьма редкие для газетно-журнального стиля выражения море разливанное (9 вхождений на 21 млн), (и) стар и млад (8 вхождений на 21 млн), разные разности (1 вхождение на 21 млн)).

С теоретической точки зрения использование статистических методов в языкознании позволяет дополнить структурную модель языка вероятностным компонентом, то есть создать структурно-вероятностную модель, обладающую значительным объяснительным потенциалом. Эту сторону использования количественных методов следует считать приложением статистики в языкознании. К моделям такого рода относится, например, «модель жизненного цикла слова», предложенная А. А. Поликарповым . Проведенный им квантитативный анализ показал, что в достаточно значительной временной перспективе имеется явная тенденция к увеличению степени абстрактности значений у многозначного слова — чем позже возникает значение, тем оно более абстрактно. Разработанная количественная модель позволяет делать интересные предположение об относительном «возрасте» различных частей речи, тенденций развития лексической системы языка и т.д.

Из приведенного примера видно, что задача построения структурно вероятностной модели функционирования языка относится к теоретическим проблемам лингвистики и входит в компетенцию теории языка. В прикладной же области квантитативная лингвистика представлена прежде всего использованием фрагментов этой модели.

# 2. Основные области приложения структурно-вероятностной модели языка

Лингвистический мониторинг функционирования языка. Задача лингвистического мониторинга заключается в выявлении общих особенностей функционирования языковой системы в конкретном типе дискурса (научном, политическом дискурсе, текстах средств массовой информации и т.д.). В качестве предмета лингвистического мониторинга могут выступать такие феномены естественного языка, как типы языковых ошибок, сфера иностранных заимствований, новые слова и значения, новые (креативные, творческие — не конвенциональные) метафоры, тематическое распределение лексики (например, лексика временных и пространственных отношений, лексика выражения чувств и эмоций, спортивная лексика и т.д.), особенности использования в текстах тех или иных грамматических форм, синтаксических конструкций. Технология лингвистического мониторинга основывается на двух важнейших предпосылках: во-первых, на регулярности и периодичности анализируемых данных, и, во-вторых — на достаточно большом объеме привлекаемого материала, на репрезентативности выборки данных. В силу этого лингвистический мониторинг невозможен без соответствующего компьютерного обеспечения. Использование компьютерной технологии позволяет давать оценку исследуемому феномену, выявляя его распределение по времени, по источникам, авторам и т.д.

Информация о статистических закономерностях функционирования языковой системы лежит в основе некоторых методик анализа данных, разрабатываемых в политической лингвистике. К ним относится, в частности, методика контентанализа, используемая для выявления структуры и состояния общественного сознания. При помощи контентанализа появляется возможность по частоте употребления лексем реконструировать, например, ценностные ориентации общества, выявлять актуальные темы публичной политики, оценивать динамику изменения тематики политических дискуссий и т.д..

Компьютерное моделирования языка и речи. Другая важная область прикладного использования знаний о частоте использования тех или иных языковых структур — компьютерная лингвистика. Многие компьютерные программы, связанные с функционированием языка, используют алгоритмы, основывающиеся на данных о частоте употребления фонем, морфем, лексических единиц и синтаксических конструкций. Например, программы автоматической коррекции орфографии содержат словари, как правило, только наиболее частотных лексем. Редкие слова пользователь может вводить в свой индивидуальный словарь. Аналогичные словари используются в программах автоматического распознавания письменного текста и речи (типа Fine Reader). Абсолютная частота появления лексем (особенно терминологической лексики) используется в системах автоматического аннотирования и реферирования. Так, согласно статистико-дистрибутивному методу автоматического индексирования информативными для данного текста считаются скопления слов, расположенных достаточно близко друг от друга, частота которых превосходит некоторую пороговую величину, например, среднюю частоту слов в документе (метод ACSI-Matic).

Дешифровка кодированного текста. В процессе дешифровки также могут использоваться данные о частоте употребления графем, морфем и слов, а также их взаимном расположении. К настоящему времени разработаны продуктивные алгоритмы дешифровки, основанные на частоте и дистрибуции элементов кодированного текста; ср. деши-Фровочные алгоритмы Б. В. Сухотина, статистико-комбинаторный метод Н.Д.Андреева. Близки к задачам дешифровки формальные процедуры «открытия» морфемного состава неописанного языка, предложенные 3. Харрисом.

Авторизация атрибуция текста. Проблема авторизации текста относится к числу классических проблем филологического исследования. Часто она рассматривается в рамках «количественной стилистики» — стилеметрии. Авторизация включает как литературную, так и лингвистическую составляющую. В. В. Виноградов в книге «Проблема авторства и теория стилей» сформулировал типологию факторов атрибуции текста. К субъективным факторам он относит:

а) субъективно-коммерческие;

б) субъективно-конъюнктурные;

в) субъективно-эстетические;

г) субъективно-психологические;

д) субъективно-идеологические факторы.

Есть и объективные факторы:

а) документально-рукописные (археологические);

б) исторические (биографии, свидетельства современников);

в) историко-идеологические и сопоставительно-идеологические;

г) историко-стилистические;

д) художественно-стилистические;

е) лингвостилистические. Однако чисто филологическое направление авторизации не позволяет построить объективные операциональные критерии анализа и атрибуции текста. К сожалению, большинство факторов, на которые обращает внимание В. В. Виноградов, плохо формализуемы. Иными словами, разные эксперты, используя одни и те же факторы, могут сделать совершенно различные выводы.

Перспектива объективизации экспертного знания была обнаружена в использовании количественных, статистических методов анализа текста. Пионером в этой области стал Н. А. Морозов, перу которого принадлежит опубликованная в 1915 г. работа «Лингвистические спектры. Средство для отличия плагиатов от истинных произведений того или другого известного автора. Стилеметрический этюд». Существенно, что в квантитативном анализе Морозов предлагал опираться не на тематически связанную лексику слова, определяемые спецификой описываемого материала, его предметной и проблемной ориентацией, — а на служебные слова и слова тематически нейтральные. Дело в том, что именно особенности употребления служебных слов, лексем с общей семантикой, не привязанной к тематике художественного произведения, формируют авторский стиль и практически не поддаются имитации.

В настоящее время развитие методик авторизации текста наиболее продуктивно проходит в рамках стилеметрии. Лингвистические основания авторизации могут быть различны, но использование количественных методов анализа оказывается неизбежным. Одно из перспективных направлений в этой области — привлечение к авторизации текста теории распознавания образов. При таком подходе стиль описывается как пространство количественно выразимых параметров — средняя длина предложения, количество вложенных синтаксических структур, количество слов в предложении, количество предложений в абзаце и т.д. Далее каждый анализируемый текст выражается через вектор, координаты которого задаются значениями выбранных параметров. Сходство векторов определяет и сходство стилей.

Разрабатываются подходы, основанные на изучении количественных особенностей реализации синтаксических структур, а также на выявлении некоторых особенностей формальной структуры текста, связанных с выражением типов чужой и авторской речи. Соотношение чужой речи (прямой, смешанной, вложенной) с авторской также оказывается стилеобразующим фактором. Эта характеристика стиля отражена в «формально-пунктуационном» методе структуризации текста, который реализован в компьютерной системе DISSKOTE [Гринбаум 1996]. Ниже разбирается пример авторской экспертизы текста, основанный на методике анализа квазисинонимичных лексем.

# 3. Авторизация текста: пример экспертизы

Одна из наиболее распространенных областей использования знаний о статистических закономерностях языковых явлений — экспертиза авторства текста. Типологически можно представить следующие базовые ситуации экспертного анализа.

A. Множественная неопределенность. Имеется множество текстов или их фрагментов. Необходимо установить, скольким авторам принадлежат тексты, и атрибутировать каждый текст конкретному автору. Это, разумеется, наиболее сложный случай анализа.

Б. Сравнение по образцу. Имеется пример текста (текстов) некоторого автора X. Необходимо установить, является ли он и автором некоторого другого текста (текстов).

B. Конкуренция образцов. Имеются образцы текстов авторов X,Y,Z... . Необходимо установить, кто из них является автором текстов Т1,Т2,... ,Тn

Приводимая ниже экспертиза12) вписывается в рамки случая В. Речь шла о спорном авторстве. В качестве материала для исследования были получены тексты следующих произведений: «Следователь президента»; «Смоленская площадь»; «Безумные глаза»; «В погоне за невидимым убийцей»; «Незнакомец»; «Шакалы»; «Трудное решение».

Тексты были представлены в печатной и машиночитаемой форме (файлы в формате DOS TEXT). Выборочное сравнение файлов и страниц печатных текстов произведений показало, что они полностью совпадают с точностью до разметки гарнитуры и фафических вьщелений в файлах. При компьютерной обработке символы разметки не учитывались.

Исходная проблема экспертизы была сформулирована следующим образом. Автором произведений «Безумные глаза», «В погоне за невидимым убийцей», «Незнакомец» является Э. Плющихин, а произведений «Шакалы», «Трудное решение» — В. Непомнящий. Авторство произведений «Следователь президента», «Смоленская площадь» является предметом спора.

В процессе проведения экспертизы необходимо было подготовить ответ на следующие вопросы:

1. Есть ли специфические языковые особенности, отличающие произведения В. Непомнящего «Шакалы», «Трудное решение» от произведений Э. Плющихина «Безумные глаза», «В погоне за невидимым убийцей», «Незнакомец», а также от спорной группы произведений — «Следователь президента», «Смоленская площадь»?
2. Есть ли специфические языковые особенности, характерные одновременно для произведений Э.Плющихина «Безумные глаза», «В погоне за невидимым убийцей», «Незнакомец» и спорной группы произведений «Следователь президента», «Смоленская площадь», но не присущие произведениям В. Непомнящего «Шакалы» и «Трудное решение»?
3. Можно ли считать, что выявленные языковые особенности являются существенной чертой авторских стилей названных писателей и могут использоваться при установлении авторства художественного текста?

В исследовании по экспертизе использовалась методика количественного анализа квазисинонимичных лексем. Сущность методики заключается в выявлении авторских предпочтений в выборе из группы квазисинонимов — близких по значению слов или устойчивых словосочетаний (фразеологизмов). В литературоведении и структурной поэтике близкие методы привлекаются для характеристики стиля писателя и особенностей его видения мира. Интересную информацию об идиолекте писателя дает изучение частотных характеристик служебных и модальных слов. Так, частицы разве и неужели по-разному распределены в романах М. Булгакова «Мастер и Маргарита» и «Белая гвардия»: разве значительно чаще встречается в «Мастере и Маргарите», а неужели — наоборот. Значение частицы разве предполагает более активную, действенную позицию говорящего, подвергающего сомнению некоторое положение дел. В противоположность разве частица неужели скорее указывает на то, что некоторое положение дел практически принимается говорящим и он лишь недоумевает, удивляется, почему оно имеет место. Учитывая значение этих частиц, указанные факты распределения разве и неужели можно интерпретировать как лингвистический коррелят авторской позиции в изображении событий и действий героев: нечто вроде пассивного «изумления», «удивления» автора в «Белой гвардии» и при активном восприятии реальности в «Мастере и Маргарите».

Такие «всплески» распределения частот служебных и модальных слов характеризуют не только отдельные тексты художественной прозы, но и оказываются характерологическими для одного писателя, оказываясь поверхностным проявлением его идеологической и творческой позиции. Явная нестандартность распределения частоты некоторых служебных слов наблюдается и в произведениях Ф. М. Достоевского.

В художественных текстах Достоевского мы встречаемся с совершенно иной ситуацией: на одно употребление по меньшей мере приходится 342,5 употреблений по крайней мере (по крайней мере — 685 вхождений по 35 художественным текстам, по меньшей мере — 2 вхождения по 35 художественным текстам). Всего употреблений по меньшей мере три, но одно из них приходится на контекст, более подходящий для по крайней мере

В противоположность по меньшей мере, по крайней мере с несомненностью относится к словам Достоевского. Конкордансы на это словосочетание занимают более двухсот страниц текста.

Можно было бы подумать, что в русском языке времен Достоевского распределение между по крайней мере и по меньшей мере было именно таким: частота по крайней мере существенно превосходила частоту по меньшей мере. Однако это не так. Предварительный анализ корпуса текстов Гоголя (около трех мегабайт) показывает, что различие в частотности этих единиц не так велико, как у Достоевского: на 1 употребление по меньшей мере — 83 употребления по крайней мере. Весьма вероятно, что в сфере публицистики того времени различие в частоте употребления по меньшей мере и по крайней мере было еще меньшим и приближалось к современной норме. Здесь тоже можно видеть коррелят авторского видения мира, авторского стиля.

По полученным файлам произведений Э. Плющихина, В. Непомнящего и спорных произведений (для каждого текста в отдельности и потрем группам — «спорная группа», «группа произведений Плющихина», «группа произведений Непомнящего») были составлены словники с указанием абсолютной и относительной частоты употребления (общее количество словоупотреблений по произведениям и по группам см. в приложении к экспертизе). Для компьютерной обработки использовалась программа DIALEX и база данных ACCESS 7 в среде Windows.

Из сферы анализа были исключены слова с предметным значением, частота употребления которых определяется конкретной проблемной областью. Были проанализированы группы квазисинонимов для следующих слоев лексической системы языка:

* наречия,
* частицы,
* вводные слова и выражения,
* фразеологические выражения (идиомы),
* глаголы речи,
* союзы и союзные слова.

Рассмотрим последовательно примеры квазисинонимов из каждой группы лексем. Наречия степени. Первая проанализированная группа — наречия со значением неполноты проявления какого-либо свойства, характеристики — едва и немного. Эти наречия во многих контекстах близки по значению. Однако стилистические предпочтения авторов проявляются в тенденциях выбора этих слов. Относительная частота едва и немного для спорных произведений и произведений Плющихина практически идентична, однако она существенно отличается от относительной частоты употребления этих единиц у Непомнящего: едва = 0,007 % (спорные), 0,006% (Плющихин), 0,018% (Непомнящий); немного = 0,002% (спорные), 0,002% (Плющихин), 0,031 % (Непомнящий) (см. табл. 1).

Таблица 1

|  |  |  |  |
| --- | --- | --- | --- |
|  | Спорные произведения | Произведения Плющихина | Произведения Непомнящего |
| Абсолют. частота | Относит. частота | Абсолют. частота | Относит. частота | Абсолют. частота | Относит. частота |
| Едва Немного | 6+716)=13 2+1=3 | 0,007 % 0,002 % | 12+3+12=27 1+2+2=5 | 0,006% 0,002 % | 11+18=2925+25=50 | 0,018% 0,031 % |

Вторая группа исследованных наречий степени характеризует степень) ожидания некоторого события. К ним относятся лексемы вдруг, внезапно, неожиданно. Все эти наречия очень близки по значению. Тем самым их частотное распределение может рассматриваться как характеристика j авторского стиля. Анализ показывает, что относительные частоты употре-. бления наречий вдруг, внезапно и неожиданно в произведениях «спорной группы» и произведений Плющихина практически совпадают: вдруг — 0,052% (спорные), 0,05% (Плющихин); внезапно — 0,001 % (спорные), 0,000% (Плющихин); неожиданно — 0,007% (спорные), 0,006% (Плющихин) (см. табл. 2). Относительные частоты соответствующих наречий в группе произведений Непомнящего существенно отличаются от первых двух групп: вдруг — 0,065 %; внезапно — 0,005 %; неожиданно — 0,016 %.

Таблица 2

|  |  |  |  |
| --- | --- | --- | --- |
|  | Спорные произведения | Произведения Плющихина | Произведения Непомнящего |
| Абсолют. частота | Относ. частота | Абсолют. частота | Относ. частота | Абсолют. частота | Относ. частота |
| Вдруг Внезапно Неожиданно | 61+30=91 1+1=2 10+2=12 | 0,052% 0,001 % 0,007 % | 52+16+39=107 0+0+0=0 5+0+7=12 | 0,05%0,000 % 0,006% | 79+25=104 7+1=8 12+13=25 | 0,065 % 0,005 %0,016% |

Третья группа проанализированных наречий степени — лексемы очень и слишком. Частотное распределение этих единиц с определенностью указывает на сходство спорной группы произведений с произведениями Плющихина. С другой стороны, относительная частота употребления очень и слишком в текстах Непомнящего существенно выше — 0,17 % для очень (0,06 % и 0,05 % в спорной группе и в группе текстов Плющихина соответственно), 0,06% для слишком (0,01 % и 0,01 % в спорной группе и в группе текстов Плющихина) — см. табл. 3. Тенденция к использованию лексем, имеющих в значении идею «превышения нормы какого-то свойства», характерная для Непомнящего, проявляется и в частоте употребления количественного наречия много, ср. 0,06 % в текстах Непомнящего (абс. част. 100), 0,01 % (абс. част. 26) в спорной группе, 0,02% (абс. част. 42) в произведениях Плющихина.

Таблица 3

|  |  |  |  |
| --- | --- | --- | --- |
|  | Спорные произведения | Произведения Плющихина | Произведения Непомнящего |
| Абсолют. частота | Относ. частота | Абсолют. частота | Относ. частота | Абсолют. частота | Относ. частота |
| ОченьСлишком | 47+49=96 9+4=13 | 0,06% 0,01% | 32+36+42=110 3+3+8=14 | 0,05% 0,01% | 103+164=267 43+46=89 | 0,17% 0,06% |

Частицы. Из лексико-грамматического разряда частиц были проанализированы единицы тоже и также, лексема так (в разных значениях). Если по относительной частоте частицы также все группы произведений приблизительно равны, то по относительной частоте частицы тоже первая и вторая группы (спорные произведения и произведения Плющихина) объединяются и противопоставляются третьей — произведениям Непомнящего. Аналогично по частоте лексемы так противопоставлены спорная группа произведений и произведения Плющихина, с одной стороны, и произведения Непомнящего — с другой (см. табл.4).

Таблица 4

|  |  |  |  |
| --- | --- | --- | --- |
|  | Спорныепроизведения | ПроизведенияПлющихина | ПроизведенияНепомнящего |
| Абсолют.частота | Относ.частота | Абсолют.частота | Относ.частота | Абсолют.частота | Относ.частота |
| Тоже | 74+63=137 | 0,08% | 55+36+66=157 | 0,07% | 89+129=218 | 0,14% |
| Также | 13+29=42 | 0,02% | 11+5+15=31 | 0,01% | 5+21=26 | 0,02% |
| Так | 208+223=431 | 0,25% | 166+239+118=523 | 0,25% | 370+295=665 | 0,42% |

Вводные слова и выражения. Как известно, к числу вводных слов относятся лексемы, в которых говорящий выражает свое отношение к сказанному, в частности оценку достоверности-недостоверности пропозиции. Была исследована частота употребления лексических единиц конечно, разумеется, естественно, связанных с выражением уверенности в истинности утверждаемого (см. табл. 5). Легко видеть, что относительные частоты употребления этих выражений в группе спорных произведений и в произведениях Плющихина практически совпадают, причем из группы квазисинонимов конечно, разумеется, естественно предпочтение отдается конечно. В группе произведений Непомнящего относительная частота употребления конечно ниже, чем в первых двух группах, что компенсируется использованием вводных слов разумеется и естественно.

Таблица 5

|  |  |  |  |
| --- | --- | --- | --- |
|  | Спорные произведения | Произведения Плющихина | Произведения Непомнящего |
| Абсолют. | Относ. | Абсолют. | Относ. | Абсолют. | Относ. |
| частота | частота | частота | частота | частота | частота |
| Конечно | 114+86= 200 | 0,12% | 113+55+100= 268 | 0,13% | 88+77=165 | 0,10% |
| Разумеется | 0+0=0 | 0,00% | 0+0+0=0 | 0,00% | 33+13=46 | 0,03% |
| Естественно | 2+1=3 | 0,002 % | 0+0+0=0 | 0,00% | 16+3=19 | 0,01% |

Фразеологизмы. Специфические особенности использования фразеологизмов относятся к числу наиболее характерных стилевых особенностей. Для анализа была привлечена группа бранных идиом с компонентом черт, хер и хрен типа черт/хер/хрен его знает, черт/хер/хрен их знает и т. д. Было установлено, что эти выражения встречаются только в группе спорных произведений (абс. частота 8) и в текстах Плющихина (абс. частота 17).

Показательно использование довольно редкой идиомы тем паче: она не встречается в произведениях Непомнящего, но представлена в текстах спорной группы (5 употреблений) и в произведениях Плющихина (2 употребления).

Глаголы речи. Среди глаголов речи наибольший интерес представляют те лексемы, которые используются как авторские ремарки для введения речи персонажей. В имеющемся материале были проанализированы глаголы сказать и возразить. Несмотря на то, что по законам жанра (исследовались детективные романы) конфликтных ситуаций в спорных текстах и в произведениях Плющихина не меньше, чем в произведениях Непомнящего, возразить существенно чаще используется именно в последней группе, что является несомненным признаком авторского стиля — см. табл. 6.

Таблица 6

|  |  |  |  |
| --- | --- | --- | --- |
|  | Спорные произведения | Произведения Плющихина | Произведения Непомнящего |
|  | Абсолют. частота | Относ. частота | Абсолют. частота | Относ. частота | Абсолют. частота | Относ. частота |
| Сказать Возразить | 1059 2 | 0,61% 0,00% | 1087 0 | 0,51% 0,00% | 919 46 | 0,57% 0,03% |

Союзы и союзные слова. Анализировались также союзы и союзные слова поскольку, почему и если. Слово почему, кроме собственно союзного употребления, функционирует как вопросительное слово; для данного типа исследования это несущественно. По распределению частоты употребления и в этом случае группа произведений Плющихина существенно сближается с группой спорных произведений, а группа произведений Непомнящего противопоставлена первым двум группам, ср. для поскольку 0,03 % и 0,03 % vs. 0,00 %; для почему 0,05 % и 0,04 % vs. 0,13 %; для если 0,22 % и 0,23 % vs. 0,31 % (см. табл. 7).

Таблица 7

|  |  |  |  |
| --- | --- | --- | --- |
|  | Спорные произведения | Произведения Плющихина | Произведения Непомнящего |
|  | Абсолют. частота | Относ. частота | Абсолют. частота | Относ. частота | Абсолют. частота | Относ. частота |
| Поскольку Почему Если | 48 95 381 | 0,03% 0,05% 0,22% | 57 95 493 | 0,03% 0,04% 0,23% | 6 212 488 | 0,00% 0,13% 0,31% |

Проведенное исследование позволяет говорить о том, что существуют специфические языковые особенности, отличающие стиль произведений Э. Плющихина от стиля произведений В. Непомнящего. Эти особенности проявляются на различных уровнях лексической системы языка — в выборе наречий, вводных слов, частиц, фразеологизмов, союзов и союзных слов, а также глаголов речи. Характерные черты стиля Э. Плющихина прослеживаются в произведениях спорной группы «Следователь президента» и «Смоленская площадь». С другой стороны, характерные языковые черты стиля В. Непомнящего не обнаруживаются в произведениях спорной группы. В рамках данного ограниченного по объему материала (для предложенных текстов двух данных писателей) выявленные особенности авторского языка могут рассматриваться как существенный фактор в пользу признания авторства Э. Плющихина на произведения спорной группы.

# Литература

1. Виноградов В. В. Лингвистические основы научной критики текста Вопросы языкознания. 2004. №2, 3.
2. Виноградов В. В. Проблема авторства и теория стилей. М., 2005.
3. Марусенко М.А. Атрибуция анонимных и псевдонимных текстов методами прикладной лингвистики// Прикладное языкознание. СПб., 2008.
4. Пиотровский Р. Г., Бектаев К. Б., Пиотровская А. А. Математическая лингвистика. М. 2007.
5. Шайкевич А. Я. Дистрибутивно-статистический анализ в семантике Принципы и методы семантических исследований. М., 2006.
6. Шайкевич А. Я. Количественные методы Лингвистический энциклопедический словарь. М., 2008.