

Санкт-Петербургский государственный университет
Кафедра математической лингвистики

Направление: «Лингвистика»

Образовательная программа: «Прикладная, экспериментальная
и математическая лингвистика (английский язык)»

**АТТРИБУЦИЯ АНОНИМНЫХ И ПСЕВДОНИМНЫХ
ТЕКСТОВ НА ПРИМЕРЕ РАНИХ РАБОТ ДОВЛАТОВА**

Выпускная квалификационная работа
соискателя на степень бакалавра филологии

ПАВЛОВА ДАНИЛА АНДРЕЕВИЧА

Научный руководитель:

д.ф.н., профессор

Марусенко Михаил Александрович

Рецензент:

д.ф.н., профессор

Доброзракова Галина Александровна

Санкт-Петербург
2016

Оглавление

Введение	4
Глава 1. Методы атрибуции	7
1.1 История вопроса атрибуции различных текстов	7
1.2 Методы атрибуции	8
1.3 Метод распознавания образов	13
2.1 История вопроса	18
2.2 Формирование атрибуционной гипотезы	24
2.3 Формирование априорных классов	25
2.4 Описание априорных классов	30
2.5 Определение информативного набора параметров	33
2.6 Определение объема выборки	39
2.7 Детерминированный алгоритм распознавания	40
2.8 Вероятностный алгоритм распознавания	43
2.9 Интерпретация результатов атрибуции	46

Введение

Тема данной работы состоит в применении методов автоматической атрибуции анонимных и псевдонимных текстов на примере произведений, приписываемых С.Довлатову.

Актуальность работы определяется историей проблемы установления авторства различных текстов, развитием методов атрибуции и математическим моделированием текстов, а также исследованием раннего творчества С.Довлатова. Разрешение проблемы авторства необходимо, в том числе, и при составлении собраний сочинений писателя, биография и произведения которого приобретают все большую популярность в последние годы.

Объект исследования данной работы – характеристики индивидуальных авторских стилей С.Довлатова, а также других потенциальных авторов спорных произведений - В.Циона и А.А.Матюшкина-Герке.

Предмет исследования – псевдонимные фельетоны и юморески, опубликованные в периодических изданиях в конце 60-х начале 70-х гг. XX века, авторство которых приписывают С.Довлатову.

Цель работы - решение проблемы авторства ранних юмористических работ С.Довлатова, опубликованных в газетах Эстонии на рубеже 60-х - 70-х гг. XX века с помощью математического метода атрибуции анонимных и псевдонимных произведений.

Для достижения поставленной цели необходимо решить ряд **задач**:

1. провести обзор истории развития методов атрибуции текстов со спорным авторством;
2. оценить современное состояние проблемы атрибуции;
3. изучить применение методов распознавания образов;
4. рассмотреть историю вопроса авторства ранних произведений С.Довлатова и сформировать атрибуционную гипотезу;

5. осуществить проверку атрибуционной гипотезы.

Методы: анализ, синтез, эксперимент. Для проверки атрибуционной гипотезы и достижения цели работы будет использован метод распознавания образов, математическое моделирование текстов на естественном языке с помощью математической статистики и вероятностного подхода.

Теоретическая значимость работы состоит в применении теории распознавания образов и описании индивидуального авторского стиля методами многомерной классификации с помощью стилеразличающего набора параметров. Результаты исследования вносят вклад в популяризацию применения методов распознавания образов для атрибуции художественных произведения.

Практическая значимость данного исследования состоит в проверке литературоведческой гипотезы об авторстве некоторых псевдонимных произведений С.Довлатовым, что важно для более полного анализа раннего творчества писателя.

Основные разделы.

Структура работы обусловлена обозначенными предметом, целью и задачами исследования. Данная работа состоит из введения, двух глав – теоретической и практической – заключения и библиографического перечня источников в алфавитном порядке.

Введение раскрывает тему дипломной работы, ее актуальность, теоретическую и практическую значимость, а также определяет объект, предмет, цель, задачи и методы исследования.

Глава 1 содержит теоретическое обоснование применяемого метода и состоит из следующих разделов:

- 1) История вопроса атрибуции различных текстов – обзор проблемы установления авторства художественных произведений, самые известные случаи спорного авторства;

- 2) Методы атрибуции – обзор методов, которые применялись для атрибуции текстов, оценка современного состояния проблемы;
- 3) Метод распознавания образов – описание применяемого метода, этапы процедуры атрибуции

Глава 2 посвящена практическому применению метода распознавания образов для атрибуции псевдонимных текстов, приписываемых С.Довлатову. Данная глава включает в себя следующие разделы:

- 1) История вопроса – анализ проблемы авторства ранних произведений С.Довлатова, обоснование выбранных объектов атрибуции;
- 2) Формирование атрибуционной гипотезы – обоснование выбора альтернативных авторов;
- 3) Формирование априорных классов – отбор текстов, соответствующих псевдонимным по жанру, стилю и времени написания;
- 4) Описание классов методами теории распознавания образов - описание классов из априорного алфавита классов на языке параметров из априорного словаря параметров;
- 5) Определение информативного набора параметров – определение релевантных и нерелевантных для различения априорных классов параметров;
- 6) Определение объема выборки - уточнение объёма выборки априорных классов для установления координат эталонов классов;
- 7) Процедура выполнения детерминированного алгоритма распознавания;
- 8) Процедура выполнения вероятностного алгоритма распознавания;

9) Интерпретация результатов – проверка соответствия полученных результатов исходной литературно-критической гипотезе либо альтернативным гипотезам.

В заключении подводятся итоги исследования, формулируются окончательные выводы по рассматриваемой теме.

Глава 1. Методы атрибуции

1.1 История вопроса атрибуции различных текстов

Проблема авторства анонимных и псевдонимных произведений является одной из старейших филологических задач. До изобретения книгопечатания (середина XV века), все литературные произведения распространялись в виде рукописей. Они подписывались автором только в редких случаях. Так, по сей день спорным является авторство произведений древнегреческих философов Платона и Аристотеля, поскольку до нашего времени не дошло ни одного автографа от произведений античной литературы.

В средневековой литературе почти каждое произведение имело сложную историю и целый ряд авторов, причем часто древнейший из дошедших до нас списков отделялся несколькими столетиями от времени создания произведения. Так, существуют различные версии об авторе главного памятника средневековой русской литературы конца XII века «Слова о полку Игореве». Несколько гипотез были выдвинуты различными исследователями и об авторстве французского романа двенадцатого века "Вильгельм Английский", приписываемого Кретьену де Труа.

После появления книгопечатания проблема атрибуции текстов не исчезла. Вопросы возникают в случаях, когда большая часть произведений автора или не была опубликована при его жизни, или есть неточности и искажения, либо допущенные при публикации, либо сделанные вынуждено из-за цензуры. Неопубликованные в печати произведения могли существовать в качестве многочисленных списков, ни один из которых нельзя признать достоверным, как в случае с комедией в стихах "Горе от ума" А.С. Грибоедова.

Отсутствие оригинальных рукописей, точных данных о создании произведения и необычная биография писателя могут вызывать сомнения об истинном авторстве тех или иных работ, возникают гипотезы о литературных мистификациях известных произведений. Например, не утихают споры по поводу исследований об авторстве романа «Тихий Дон» М.А.Шолохова и работ У.Шекспира.

Вопросы определения истинного авторства приобрели особую актуальность на этапе изучения и издания наследия русских писателей-классиков. В разные годы возникали сомнения в подлинности произведений Н.А.Некрасова, В.В.Маяковского, А.П. Чехова, М.А. Булгакова. Также проводились исследования по атрибуции работ иностранных писателей, в частности Эмиля Ажара.

1.2 Методы атрибуции

В классификации методов атрибуции, составленной академиком В.Виноградовым в 1961 г., выделялись 5 субъективных и 6 объективных принципа атрибуции. Виноградов относил группу объективных принципов к традиционной лингвистике, а лингвостатистические методы атрибуции в рамках его исследования имели лишь вспомогательную роль, т.к.

действенных методов определения авторства, основанных на формальном математическом аппарате, в начале 60-х годов XX века еще не существовало. Однако именно лингвостатистический анализ с использованием математических методов является наиболее эффективным, т.к. количественная оценка однородности состава и структуры сравниваемых текстов намного объективнее, чем любое стилистическое исследование.

До 70-х годов двадцатого века среди атрибуционных методов «доминировали историко-документальные и филологические исследования. Для выявления авторских особенностей применялся субъективный подход атрибуции, в соответствии с которым отбирались внешние детали авторского стиля, такие как любимые слова, термины, выражения» [12].

Впервые математико-статистические методы были применены в конце XIX века для атрибуции античных произведений. В работах *Кэмпбелля* и *Люгославского* (1867, 1897 годы) использовались методы выбора "оригинальных" слов и определения позиционного расположения определяющих и определяемых слов.

Впервые в отечественной истории математический аппарат для решения задач атрибуции текстов был применен *Н. А. Морозовым*. В статье «Лингвистические спектры», опубликованной в 1915г., он, в отличие от предшествующих исследователей, опирающихся на частотность знаменательных слов в тексте, использует при установлении индивидуального стиля писателя число служебных слов, поскольку они не связаны с темой и содержанием произведения. Так, при проведении атрибуции Морозов вычислял комбинации употреблений предлогов «в», «на» и частицы «не». Недостатком метода Морозова является то, что он не выходил за рамки анализа лексики: при увеличении объема текста,

количественное значение рассматриваемых характеристик сводится к средним показателям для русского языка [14].

С 60–70-х годов XX века лингвоматематические методы для описания индивидуального стиля автора применяются все чаще, разрабатываются принципы применения методов статистики к лексике и грамматике.

Так, в исследованиях А.Якубайтис, А.Н.Скляревича рассматривалось число повторений какой-либо части речи в научно-технических, поэтических и драматических текстах. В результате ученые пришли к выводу о структурности систем частей речи и считали, что достоверность результатов атрибуции повышется при увеличении числа рассматриваемых признаков.

Исследование А.П. Василевича, опубликованное в 1981г. было посвящено употреблению цветонаименований в поэзии и прозе 19-20 веков. Основой анализа являлись индекс лексической оригинальности (соотношение частотности редких слов и частотности частых), индекс морфологической оригинальности (соотношение сложных и простых слов) и индекс насыщенности (частотность употребления цветонаименований). Результатом исследования стала характеристика «живописности» писателей разных эпох.

Монография *Ю.Тулдавы* (1987 г.) обобщила опыт квантитативно-лингвистических исследований. В данной работе Ю.Тулдава сформулировал принципы системности и вероятностно-статистического характера организации лексики, а также выдвинул идею о взаимосвязи признаков, что и стало основой при разработке математического аппарата для оценки взаимосвязи параметров.

В 70-е – 80-е годы XX века в отечественной лингвистике был проведен ряд исследований, посвященных квантитативно-структурному изучению текстов на синтаксическом уровне [11]. Это было связано с

пониманием стиля как структурно-синтаксической категории. При этом использование синтаксического анализа подразумевает анализ текста как на синтаксическом, так и на лексическом, фразеологическом и морфологическом уровнях.

В этот период были проведены исследования методов стилистической диагностики, в центре которых был анализ графов синтаксических связей. Они были представлены в работах *И.П. Севбо* и *Г.Я. Мартыненко*. Однако диагностические параметры И.П. Севбо, и меры сложности Г.Я. Мартыненко, были связаны с характеристиками предложения, а не текста. Поэтому построение усредненного графа для целого текста и характеристики авторского

стиля на его основе не может быть эффективным методом установления индивидуального авторского стиля.

С конца 70-х г.г. XX века по настоящее время активно развивается компьютерная техника и программное обеспечение. Все больше исследователей проявляют интерес к применению вычислительной техники для автоматической обработки данных при анализе текстов, как на синтаксическом, так и на грамматическом и лексическом уровнях.

Так, в исследовании норвежского филолога *Г.Хемсо* (1978) впервые была применена методика атрибуции, в основе которой лежал автоматизированный анализ текста. Исследователь пытался определить авторство текста, используя параметр "средняя длина предложения" с помощью частотных словарей и статистических данных. Позже данная методика была опровергнута как статистически бессмысленная. Специально разработанная компьютерная программа применялась и *Л.В. Миловым* при работе с древними текстами. Он производил обработку текста с помощью построения графов «сильных связей» по матрице частот парной встречаемости грамматических классов слов.

Одно из последних исследований по атрибуции текстов, также основанное на анализе лексики, было проведено *Домиником Лаббе*. В 2001 году он предложил формулу вычисления "межтекстового расстояния". В его исследовании лексический анализ текстового материала выполнен при помощи автоматического аппарата морфологического анализа. Каждое слово представлялось в виде записи, в которой было три компонента: словоформа, вокабула и соответствующая часть речи. Полученные результаты и были основой при вычислении «межтекстового расстояния» двух текстов, т.е. происходило определение мер близости или удаленности объектов атрибуции друг от друга. «Исследования Д. Лаббе обнаруживают серьезные недостатки методологического и статистико-вероятностного характера, связанные в первую очередь с недостоверностью результатов атрибуции, осуществленной на основе анализа одного лишь лексического уровня» [12].

Подавляющее большинство исследований индивидуальных характеристик авторского стиля сводится к анализу лексического уровня, лексики синтаксиса и поэтической семантики. Разрабатываются методы автоматизации грамматического и морфологического анализа. Однако для описания текста необходим анализ его структуры, синтаксиса. Эффективный анализ индивидуального стиля автора в синтаксическом аспекте был впервые разработан *М.А. Марусенко* в 1990г. При атрибуции анонимных и псевдонимных произведений он использует методы распознавания образов, где применяется многомерная классификация признаков для определения индивидуальных характеристик авторского стиля. Данная методика позволяет охарактеризовать текст в его совокупности, а не анализировать отдельные предложения на различных уровнях языковой системы. Рассмотрим ее подробнее.

1.3 Метод распознавания образов

Впервые применение методов распознавания образов для атрибуции анонимных и псевдонимных произведений было описано в 1990г. в монографии М.А. Марусенко [7].

В данной работе текст рассматривается как сложный лингвистический объект, который можно описать широким набором элементов с помощью

многоуровневого анализа. В основе данного метода определения авторства анонимных и псевдонимных произведений лежит принцип многомерного статистического анализа, а именно теория распознавания образов.

В терминах распознавания образов индивидуальный авторский стиль рассматривается как структурная синтаксическая категория. Стиль выражается, прежде всего, в синтаксисе языка: наборе конструкций, их расположении и взаимосвязи, представляющей общее целое. Таким образом, стиль определяется как «набор свойств (параметров), характеризующих состав, способы объединения и статистико-вероятностные закономерности употребления речевых средств, образующих данную разновидность языка» [7, с. 17-18]. Набором свойств, характеризующих структуру текста в синтаксическом аспекте, становится в данном случае совокупность информативных параметров, чей состав определяется путем выполнения специальной процедуры отбора информативных параметров для каждого конкретного случая.

Процедура атрибуции разделяется на три этапа [8, с 7]:

1. Формирование литературно-критической атрибуционной гипотезы, которое выполняется методами традиционного филологического анализа;

2. Проверка литературно-критической гипотезы с помощью средств теории распознавания образов.
3. Интерпретация результатов проверки атрибуционной гипотезы.

Гипотеза считается статистически подтвержденной, если результаты распознавания согласуются с исходной литературно-критической атрибуционной гипотезой (при установленном уровне значимости). В противном случае гипотеза считается опровергнутой, и проверяется либо альтернативная гипотеза, либо переформулированная исходная гипотеза [9].

При реализации такой схемы атрибуции статистико-вероятностные методы анализа языка и стиля используются лишь в качестве вспомогательных средств для проверки исходной атрибуционной гипотезы. Соответственно результаты анализа зависят от обоснованности данной гипотезы, методы распознавания образов лишь устанавливают наличие или отсутствие статистически значимых различий между атрибутируемым текстом и корпусом текстов предполагаемого автора.

Проверка литературно-критической гипотезы происходит в несколько этапов с использованием определенного комплекса процедур в строгой последовательности [8, с 8-17]:

1. **Определение априорного алфавита классов.**

Состав априорных классов определяется требованиями временной и жанровой однородности, а объем измеряется в основных единицах синтаксиса — предложениях. При необходимости возможно уточнение априорного алфавита классов, формируется рабочий алфавит классов.

2. **Определение априорного словаря параметров.**

Определяется полный перечень параметров, характеризующих тексты, для которых разрабатывается система. Включаются параметры,

необходимые и достаточные для полного описания классов на языке данных параметров. Единицей параметрического описания априорных классов является предложение.

3. Описание классов из априорного алфавита классов на языке параметров из априорного словаря параметров.

Каждому лингвистическому объекту, подвергаемому анализу в целях стилистической диагностики, ставится в соответствие математический объект p , характеризующийся n -мерным вектором, где n – число параметров. Описаниями классов на языке параметров являются их эталоны – точки, равноудаленные от точек, соответствующих объектам, принадлежащим каждому из классов. Описание объектов атрибуции на языке априорного словаря параметров происходит с помощью ручного анализа данных.

4. Определение информативного набора параметров.

Этот этап атрибуции заключается в выделении из имеющегося информационного параметрического пространства необходимого и достаточного числа параметров для отнесения объекта к классу, благодаря чему удаляются лишние параметры. Процедура свертывания параметрического пространства производится путем обработки корреляционной матрицы связей параметров. Содержательным критерием информативности набора параметров служат слабая корреляция информативных параметров между собой и их сильная корреляция с остальными параметрами, не входящими в данную группу. Таким образом, вычисляются средняя внутригрупповая корреляция параметра, средняя внегрупповая корреляция параметра и значение критерия его эффективности.

Для определения информативного набора параметров проводится описание априорных классов на языке параметров из априорного словаря параметров. Случайным образом формируются выборки объемом 100 предложений каждого априорного класса. Выбор информативных

параметров предполагает разбиение априорного набора параметров на два подмножества: релевантных и нерелевантных для различения априорных классов параметров. Релевантность определяется по t -критерию Стьюдента, пороговое значение которого равно 1,96 (при уровне значимости = 0,05).

5. Определение объема выборки

При определении координат распознаваемых объектов желательно провести сплошное статистическое обследование, т.к. распознаваемые объекты зачастую имеют небольшой объем авторского текста. Для определения координат эталонов классов, наоборот, целесообразно применить выборочный метод, если позволяют объемы тестов априорных классов. Объем выборки

определяется выборочным средним значением признака для совокупности, выборочным средним квадратичным отклонением, долей отбора и стандартной ошибкой оценки среднего совокупности.

6. Детерминированный алгоритм распознавания

Алгоритм распознавания предусматривает двухступенчатую процедуру распознавания: детерминированное и вероятностное. В качестве функции детерминированного алгоритма используется t -критерий Стьюдента, критическое значение которого при $\alpha=0,05$ равно 1,96. Соответственно, если для параметра значение t -критерия при сравнении двух объектов меньше 1,96, то можно с вероятностью 0,95 говорить о принадлежности данных объектов к одному классу и наоборот. Используя выбранные с помощью такого метода параметры, можно составить решающее правило детерминированного алгоритма. Если объект по всем информативным параметрам относится к классу А и не относится к классу В, то он относится к классу А и наоборот. Если же для данного объекта по всем параметрам получен "отказ" или есть хотя бы два параметра, по одному из которых объект относится к классу А, а по другому к классу В,

то результатом будет отказ от распознавания и остановка детерминированного алгоритма. Далее в действие вступает вероятностный алгоритм.

7. Вероятностный алгоритм распознавания.

Здесь в качестве функции расстояния между объектами в многомерном параметрическом пространстве и априорными классами используется взвешенное евклидово расстояние. Принадлежность объекта атрибуции одному из априорных классов определяется на основе вероятностного подхода.

8. Интерпретация полученных результатов.

Подтверждение или опровержение первоначальной литературно-критической гипотезы, проверка альтернативных гипотез.

Глава 2. Атрибуция псевдонимных текстов, приписываемых С.Довлатову

В вопросе установления авторства спорного произведения необходимо охарактеризовать его с биографической, идеологической и стилистической точек зрения, поскольку каждое произведение - это результат творчества писателя в определённый момент его жизни, литературной деятельности в конкретную историческую эпоху. При этом постановка задачи и применение конечных результатов атрибуции осуществляется в рамках литературоведения, а аппарат и методы анализа относятся к математической сфере.

Неотъемлемой частью атрибуции будет являться установление стиля псевдонимного произведения и сравнение его со стилем предполагаемых авторов. Поскольку индивидуальный авторский стиль может отличаться в зависимости от жанра произведения и изменяться с годами, произведения все рассматриваемые произведения, как псевдонимные, так и работы потенциальных авторов, имеют жанровую однородность и принадлежат одному историческому периоду - второй половине 60-х и первой половине 70-х гг. XX века.

2.1 История вопроса

К вопросу атрибуции ранних работ С.Д.Довлатова специалисты обратились после «Вторых Довлатовских чтений», проходивших в 2011г. в Санкт-Петербурге. На встрече с участниками конференции главный редактор журнала «Звезда» и друг Довлатова А.Арьев раскрыл один из псевдонимов писателя – «С.Адер». Как показало дальнейшее исследование, существует большое количество псевдонимов, которые, как считают некоторые литературоведы, также принадлежат Довлатову. Под этими псевдонимами в периодической печати публиковались репортажи,

юморески и фельетоны. Юморески Довлатова появлялись в специальных сатирико-юмористических рубриках в журналах и газетах. Писатель еще до эмиграции публиковал короткие юмористические и сатирические произведения в конце 1960-х и в 1970-е годы в журналах «Крокодил», «Аврора», а также в эстонских газетах «Советская Эстония», «Вечерний Таллин», «Моряк Эстонии», но они остались практически незамеченными специалистами.

Журналистская карьера Довлатова началась с работы в студенческой многотиражной газете «За кадры верфям» Ленинградского кораблестроительного института. В этот период, с 1965 по 1969, молодой журналист публиковал отчеты, очерки, интервью, репортажи о жизни университета в том числе используя псевдоним «Д.Сергеев» [2]. В 1968 Довлатов совместно с Б. Тараторкиным выпускал юмористическую страницу "На полубаке». В 1970г. юмореска Довлатова "А вы читали Кафку?" была опубликована в журнале «Аврора» в юмористическом разделе «СЛОН». Довлатов печатался в «СЛОНе» еще 3 раза. Миниатюра «Дозвонился» появилась в печати в 1975, «С Ивановым шутки плохи» – в 1976 и, наконец, юмореска «Клад», которую автор подписал псевдонимом «Далматов», – в 1977. Несколько раз произведения Довлатова печатались в журнале «Крокодил»: в 1968 вышел рассказ «Когда-то жили в горах», в 1969 – рассказ «Победители», в 1971 – две комические миниатюры «У реки» и «Спасите наши уши!», а в 1972 – «Счастливчик».

Но большинство своих фельетонов и юморесок Довлатов создал и опубликовал, работая в Таллине в 1972-1975г.г. Он сотрудничал с такими газетами, как: «Моряк Эстонии» (где работал ответственным секретарём), «Советская Эстония», (работал в отделе информации с 1973) и «Вечерний Таллин» (был внештатным корреспондентом).

Первые псевдонимные публикации появляются в газете «Советская Эстония» осенью 1972 (к примеру, юмореска «Энтузиаст», за подписью

В.Сергеев). А подпись «Довлатов» можно увидеть в газете только 8 декабря 1972 года под очерком «Благодарите моего коня». В дальнейшем писатель использует как свое имя, так и несколько псевдонимов для публикации репортажей, рецензий, очерков, фельетонов и юморесок. Такая политика поддерживалась главным редактором издания Г.Турунком, создавалась видимость большого числа сотрудников. Так, в газете можно встретить такие псевдонимы, как: С.Адер, С.Кивистик, Братья Тривзоровы, О.Михайлов, Герб, О.Корюк, Д.Сильд.

Одновременно Довлатов работает в газете «Вечерний Таллин», публикуясь в сатирической рубрике «Пчелка». Здесь можно встретить две юморески, опубликованные под именем Довлатова («На равных», 1972 и «Иванов не виноват!», 1973), а также множество юмористических произведений под псевдонимами Л.Бурых, О.Михалыч, И.Олеп, Н.Кох, И.Кузьмич, Х.Сирв, Х.Сире, Братья Тривзоровы.

С конца 1972 в газете «Молодежь Эстонии» и с января по апрель 1973 в газете «Моряк Эстонии» появляются рецензии, очерки, корреспонденции, интервью, репортажи и юморески под фамилией «Довлатов». Во время короткого сотрудничества с газетой «Моряк Эстонии» Довлатов успел опубликовать в разделе «Пьянству – бой!» юмореску «Заведу себе бар...» под псевдонимом Д.Сергеев.

Сергеев - один из первых псевдонимов писателя. Работая в 1965-1969г.г. в ленинградской газете «За кадры верфям» под различными публикациями он подписывался «Д.Сергеев». Тайну этого псевдонима раскрыл бывший сотрудник данной газеты В.Степанов.

В газете «Советская Эстония» чаще сего появляется псевдоним «С.Адер». Как уже отмечалось выше, о том, что псевдоним принадлежит Довлатову, стало известно в 2011г., когда на «Вторых Довлатовских чтениях» об этом рассказал редактор журнала «Звезда» и друг Довлатова А.Арьев.

Есть доказательства принадлежности еще одного псевдонима Довлатову. Юмореска «Как я бросил пить» под именем «С.Кивистик», напечатанная в «Советской Эстонии» в 1973, была опубликована годом ранее в газете ЛОМО «Знамя прогресса» под собственной фамилией писателя. Тексты двух произведений отличаются незначительно (например, в таллинском варианте нет подзаголовка «Быль»).

Также считается, что «Далматов» – это псевдоним Довлатова, точнее, его искаженная фамилия (в жизни его фамилию часто искажали, о чем он неоднократно упоминал). В его последней повести «Фиалиал» главный герой - Далматов.

Как отмечают литературоведы, юморески и фельетоны, написанные Довлатовым, и предположительно принадлежащие ему, имеют схожие особенности. Среди них можно выделить: повествование от первого лица, краткость абзацев, диалогичность, неожиданная концовка (пуант), аллюзии из произведений русской классической литературы, стихотворные элементы, повторяемость имен персонажей в разных произведениях, схожие тематики (изображение человеческих пороков, абсурдных жизненных ситуаций, критика работников сферы обслуживания) [1].

На основании литературоведческих исследований специалисты говорят о принадлежности следующих юморесок перу Довлатова: тексты из рубрики «Юмореска» (газета «Советская Эстония») под псевдонимами О.Корюк, Д.Сильд, С.Кивистик, Герб; юморески и фельетоны, опубликованные в рубрике «Пчелка» газеты «Вечерний Таллин», под псевдонимами Л.Бурых, О.Михайлов, О.Михалыч, И.Олен, Н.Кох, И.Кузьмич, Х.Сирв, Х.Сире, О.Летберг [1].

Таким образом, в качестве объектов атрибуции выступают произведения, приведенные в таблице 1:

Табл.1 Объекты атрибуции

№	Произведение	Мощность N,
1	В. Сергеев. Энтузиаст // Советская Эстония. 03.09.1972. С.4	6
2	Д. Сергеев. Заведу себе бар... // Моряк Эстонии. 21.04.1973. С. 4	33
3	О. Михайлов. Первый «заяц» // Советская Эстония. 01.01.1973. С.4	17
4	О. Михайлов. Здравсьте, это Вася говорит... // Вечерний Таллин. 16.01.1974. №13 (443). С. 3	11
5	О. Михалыч. Астрономия // Вечерний Таллин. 09.02.1974. №34 (464). С. 3	5
6	Д. Сильд. Случай с Барашкиным // Советская Эстония. 28.01.1973. С. 4.	19
7	С. Кивистик. Как я бросил пить // Советская Эстония. 11.02.1973. С. 4.	29
8	О. Корюк. Звонок из Антимира // Советская Эстония 25.03.1973. С. 4	15
9	Х. Сирв. Да здравствует петух! // Вечерний Таллин. №71. 21.10.1972. С. 3.	27
10	Х. Сире. Человек изголодался // Вечерний Таллин. 14.09.1974. №215 (645). С. 3.	31
11	Л. Бурых. Труба // Вечерний Таллин. 28.10.1972. С.3.	9
12	Л. Бурых. Не лезь в бутылку. (Полуфантастическая история) // Вечерний Таллин. 01.11.1972. С.3.	14
13	Л. Бурых. Не тот уровень // Вечерний Таллин. 18.11.1972. С.3.	5
14	Л. Бурых. Такая специфика // Вечерний Таллин. №106. 06.12.1972. С.3.	8

15	Л. Бурых. Последний сапожник // Вечерний Таллин. №11 (138). 13.01.1973. С.3.	4
16	Л. Бурых. Ха-ха! // Вечерний Таллин. 28.02.1973. С.3.	4
17	Л. Бурых. Мужские хлопоты // Вечерний Таллин. 07.03.1974. С.4	47
18	И. Олеп. Почти что драма про некую маму, ее малолетнего сына Игната и директора лесокombината // Вечерний Таллин. 05.01.1974. С.3	15
19	Н. Кох. Сколько лет бабушке? // Вечерний Таллин. 13.03.1974. С.3	24
20	И. Кузьмич. Муха // Вечерний Таллин. 27.10.1973. С.3	12
21	И. Кузьмич. В ожидании генерала. (Действительное происшествие) // Вечерний Таллин. № 119 (549). 25.05.1974. С.3	17
22	И.Кузьмич. Сон // Вечерний Таллин. 14.11.1973. №264 (391). С.3	21
23	И.Кузьмич. История // Вечерний Таллин. 15.12.1973. №290 (417). С.3	18
24	Братья Тривзоровы. Зонт // Советская Эстония. 02.09.1973. С. 4.	16
25	Братья Тривзоровы. Пончо // Вечерний Таллин. 12.09.1973. С.3.	5
26	Братья Тривзоровы. Хочу всё знать! // Вечерний Таллин. 10.10.1973. С. 3.	17
27	О. Летберг. Чуткость // Вечерний Таллин. 12.09.1973. С. 3	16

Тексты данных произведений представлены в приложении 1 «Объекты атрибуции».

2.2 Формирование атрибуционной гипотезы

При поиске возможных альтернативных авторов данных псевдонимных произведений обращает на себя внимание писатель, близкий Довлатову по работе в Ленинграде – А.А.Матюшкин-Герке, который вел юмористический «журнал в журнале» «СЛОН» в «Авроре», где несколько раз публиковался Довлатов. В тот же период Матюшкин-Герке работал и в газете «Молодежь Эстонии», где появлялись его юмористические рассказы (например, юмореска «Наглядный урок», 1974). Как отмечают литературоведы, можно заметить значительное сходство его произведений с довлатовскими миниатюрами. Например, использование одинаковых имён персонажей (Иван Петрович, Петр Иванович, Петров и т.п., характерные для юморесок Довлатова); малый объем; повествование от 1 лица; неожиданная концовка – (пуанта); диалогичность; тематика (о сфере обслуживания, о выигрыше в лотерею, о неисправном телефоне-автомате, о попадании в медвытрезвитель и т.д.).

Те же особенности можно обнаружить и у другого автора – В.Циона, произведения которого публиковались в тот же период в газете «Вечерний Таллин». При этом «Цион» не может быть псевдонимом Довлатова по той причине, что фельетоны под этой фамилией появились в самом первом выпуске рубрики «Пчелка», вышедшем в августе 1972 г., а Довлатов приехал в Таллин только сентябре.

Таким образом, в число возможных авторов могут входить В.Цион, А.А.Матюшкин-Герке и неопределенное число неизвестных авторов.

Нулевую гипотезу (H_0) можно сформулировать так: все псевдонимные произведения принадлежат Сергею Довлатову. (27 работ при 15 псевдонимах).

Соответственно, альтернативными гипотезами будут следующие:

(H₁) - все работы принадлежат не Довлатову, а одному или нескольким авторам (А.Матюшкину-Герке или В.Циону);

(H₂) - часть произведений принадлежит Довлатову, а часть – еще одному или нескольким писателям.

В соответствии с применяемой методикой для проверки атрибуционной гипотезы необходимо:

- определить априорный и информативный словари параметров,
- сформировать априорные классы,
- определить их координаты и координаты атрибутируемых объектов,
- произвести распознавание объектов с оценкой качества классификации.

2.3 Формирование априорных классов

Состав априорных классов должен соответствовать требованиям жанровой, стилевой и хронологической однородности. Для этого были отобраны фельетоны и юморески, публиковавшиеся соответствующими авторами в периодических изданиях с 1968 по 1977 гг. Сформированный таким образом априорный алфавит классов, их мощность (число предложений авторской речи) и объем (число текстов) представлены в таблице 2.

Табл.2 Объем априорных классов.

№	Класс	Объем (V)	Мощность (N)
---	-------	-----------	--------------

1	Довлатов	29	749
2	Цион	11	319
3	Матюшкин-Герке	15	391

Состав и структура априорных классов К1 (Довлатов), К2 (Цион), К3 (Матюшкин-Герке) описаны в таблицах 3, 4, 5 соответственно.

Табл.3 Состав и структура априорного класса К1 (Довлатов).

№	Произведение	Мощность, N
1	Сергей Далматов. Клад // Аврора. №2. 1977. С. 78	37
2	С. Адер. Победа над Геннадием Шатковым. (Спортивная быль) // Советская Эстония. 07.07.1973. С. 4	14
3	С. Адер. Все на свете неспроста // Советская Эстония. 05.08.1973. С. 4.	45
4	С. Адер. Новая жизнь // Советская Эстония. 12.08.1973. С. 4.	25
5	С. Адер. Чудесная находка // Советская Эстония. 18.08.1973. С. 4.	24
6	С. Адер. Свое слово в искусстве // Советская Эстония. 25.08.1973. С. 4.	19
7	С. Адер. Трое вышли из леса // Советская Эстония. 01.09.1973. С. 4.	12
8	С. Адер. Ничего страшного! // Советская Эстония. 25.11.1973. С. 4.	19
9	С. Адер. Экономический эффект, или рассказ о том, как Никеша Квакин пальтишко покупал // Советская Эстония 09.12.1973. С. 4.	30
10	С. Адер. Хитрый Булкин // Советская Эстония. 26.05.1974. С. 3.	34
11	С. Адер. Зигзаг удачи. // Советская Эстония. 05.02.1975. С. 3.	14

12	С. Адер. Герой // Советская Эстония 09.02.1975. С. 3.	1
13	С. Довлатов. Когда-то мы жили в горах // Крокодил. № 26. 1968. С. 11.	67
14	С. Довлатов. Победители // Крокодил. №2. 1969. С. 10.	26
15	С. Довлатов. У реки // Крокодил. №19. 1971. С. 8-9.	31
16	С. Довлатов. Спасите наши уши! // Крокодил. №25. 1971. С. 14.	27
17	С. Довлатов. Счастливчик // Крокодил. №2. 1972. С. 13.	20
18	С. Довлатов. А вы читал Кафку? // Аврора. №7. 1970. С. 78	9
19	С. Довлатов. Дозвонился // Аврора. №10.1975. С. 78.	41
20	С. Довлатов. С Ивановым шутки плохи // Аврора. №4. 1976. С. 76.	23
21	С. Довлатов. Сила убеждения // Советская Эстония. 07.01.1973. С. 4.	4
22	С. Довлатов. Испорченный телефон. // Советская Эстония. 13.01.1973. С. 4.	50
23	С. Довлатов. С Ивановым шутки плохи // Советская Эстония. 01.04.1973. С. 4.	22
24	С. Довлатов. Осторожно – Фантомас // Советская Эстония. 24.04.1973. С. 4.	17
25	С. Довлатов. Вперед, к динозаврам! // Советская Эстония. 08.07.1973. С. 4.	15
26	С. Довлатов. Обыкновенное чудо. // Советская Эстония. 30.09.1973. С. 4.	19
27	С. Довлатов. Бессмертный дядя Вася? // Советская Эстония. 11.02.1975. С. 2.	83

28	С. Довлатов. На равных. // Вечерний Таллин. 13.12.1972. №112. С. 3.	18
29	С. Довлатов. Иванов не виноват! // Вечерний Таллин. №111(238). 16.05.1973. С. 3.	3

Тексты произведений С.Д.Довлатова представлены в приложении 2 «Априорный класс К1 (Довлатов)».

Табл. 4. Состав и структура априорного класса К2 (Цион).

№	Произведение	Мощность, N
1	В. Цион. Взаимопонимание // Вечерний Таллин. 21.11.73. №270 (397). С.3	5
2	В. Цион. А у нас в квартире газ. А у вас? // Вечерний Таллин. 08.09.1973. №219(337). С.3	48
3	В. Цион. В конце месяца // Вечерний Таллин. 29.09.1973. № 228(355). С.3	5
4	В. Цион. Долгий ящик // Вечерний Таллин. 09.08.1972. С.3	16
5	В. Цион. Кто хочет – тот добьется! // Вечерний Таллин. 09.06.1973. С.3	54
6	В. Цион. "Не забывайте!" // Вечерний Таллин. 16.05.1973. С.3	68
7	В. Цион. Феноменальная память // Вечерний Таллин. 26.08.1972. С.3	30
8	В. Цион. Как учили // Вечерний Таллин. 03.10.1973. №231(358). С.3	0
9	В. Цион. Ситуация // Вечерний Таллин. 06.10.1973. №234(361). С.3	15
10	В. Цион. Все прекрасно // Вечерний Таллин. 13.10.1973. №240(367). С.4	26
11	В. Цион. А это чревато... // Вечерний Таллин. 23.05.1973. С.3	52

Тексты произведений В.Циона представлены в приложении 3 «Априорный класс К2 (Цион)».

Табл. 5. Состав и структура априорного класса К3 "Матюшкин-Герке".

№	Произведение	Мощность, N
1	А. А. Матюшкин-Герке. Наглядный урок // Молодежь Эстонии. 10.10.1974. С. 3	4
2	А. А. Матюшкин-Герке. Приносим свои извинения // Приятное с полезным. - Л., 1986. С. 102-105	35
3	А. А. Матюшкин-Герке. Вратарь // Приятное с полезным. - Л., 1986. С. 19-21	48
4	А. А. Матюшкин-Герке. Счастливчик // Приятное с полезным. - Л., 1986. С. 40-41	14
5	А. А. Матюшкин-Герке. Шай-бу, шай-бу! // Приятное с полезным. - Л., 1986. С. 17	8
6	А. А. Матюшкин-Герке. Первая роль // Приятное с полезным. - Л., 1986. С. 4-6	14
7	А. А. Матюшкин-Герке. Почему я не люблю ходить в гости // Приятное с полезным. - Л., 1986. С. 43-45	27
8	А. А. Матюшкин-Герке. День здоровья // Приятное с полезным. - Л., 1986. С. 49-50	18
9	А. А. Матюшкин-Герке. Средство общения // Приятное с полезным. - Л., 1986. С. 41-43	25
10	А. А. Матюшкин-Герке. Приятное с полезным // Приятное с полезным. - Л., 1986. С. 30-31	47
11	А. А. Матюшкин-Герке. Учитель // Градусы в углу. - Л., 1987. С. 86-93	76

12	А. А. Матюшкин-Герке. Дебют // Приятное с полезным. - Л., 1986. С. 83-84	12
13	А. А. Матюшкин-Герке. Рубль // Приятное с полезным. - Л., 1986. С. 114-116	25
14	А. А. Матюшкин-Герке. Визит // Приятное с полезным. - Л., 1986. С. 124-126	19
15	А. А. Матюшкин-Герке. Клад // Приятное с полезным. - Л., 1986. С. 128-130	19

Тексты произведений А.А. Матюшкина-Герке представлены в приложении 4 «Априорный класс К3 (Матюшкин-Герке)».

2.4 Описание априорных классов

Описание классов из априорного алфавита классов происходит на языке параметров из априорного словаря параметров. Для начала для каждого априорного класса сделаем выборку из 100 предложений авторской речи случайным образом (см приложение 5 «Выборка текстов С.Довлатова», приложение 6 «Выборка текстов В. Циона» и приложение 7 «Выборка текстов А.А.Матюшкина-Герке»). Охарактеризуем предложения каждой выборки по 56 параметрам из априорного словаря параметров¹ (см. таблицу 6).

Табл. 6. Априорный словарь параметров.

X01	число слов в цельном предложении
-----	----------------------------------

¹ Марусенко М.А. Атрибуция анонимных и псевдонимных литературных произведений методами теории распознавания образов. [Текст] - Л., 1990. С.71-72.

X02	число графем в цельном предложении
X03	число слов в простом самостоятельном предложении
X04	число элементарных предложений в цельном предложении
X05	число главных предложений
X06	число сочиненных предложений
X07	число сочиненных предложений без спрягаемой формы
X08	число подчиненных предложений
X09	число подчиненных предложений 1-й степени
X10	число подчиненных предложений 2-й степени
X11	число подчиненных предложений 3-й степени
X12	число подчиненных предложений 4-й и высших степеней
X13	число элементарных предложений без номинативного подлежащего
X14	число подчиненных предложений без спрягаемой формы глагола
X15	число вставных предложений
X16	число охватывающих предложений
X17	число слов 1-й группы (знаменательных)
X18	число слов 2-й группы (служебных)
X19	число существительных
X20	число прилагательных
X21	число местоимений
X22	число спрягаемых форм глагола
X23	число именных форм глагола
X24	число наречий
X25	число предлогов
X26	число союзов
X27	число подчинительных союзов
X28	число сочинительных союзов
X29	число предикативов (кратких прилагательных и предикативных наречий)
X30	число слов в аккузативе (прямых дополнений)
X31	число косвенных дополнений (число слов в дативе)

X32	число подлежащих
X33	число местоимений - подлежащих
X34	число групп однородных членов
X35	число членов однородных групп
X36	число однородных сказуемых
X37	число однородных групп дополнений
X38	число причастных оборотов
X39	число членов причастных оборотов
X40	число распространенных причастных определений
X41	число членов распространенных причастных оборотов
X42	число согласованных определений
X43	число причастий - согласованных определений
X44	число несогласованных определений
X45	число существительных - несогласованных определений
X46	число обособленных членов
X47	число членов в группах обособленных членов
X48	число абсолютных оборотов
X49	число членов абсолютных оборотов
X50	число инфинитивных оборотов
X51	число членов инфинитивных оборотов
X52	число существительных без группы
X53	число групп существительных
X54	число членов групп существительных
X55	число знаменательных слов в группах существительных
X56	число служебных слов в группах существительных

Результаты вычисления значений данных параметров для каждого предложения авторской речи трех априорных классов занесем в таблицы, составив матрицы априорных классов (см. приложение 8 «Матрица априорного класса К1 (Довлатов)», приложение 9 «Матрица априорного

класса К2 (Цион)» и приложение 10 «Матрица априорного класса К3 (Матюшкин-Герке)»).

2.5 Определение информативного набора параметров

Вычислим для каждого параметра среднее значение и стандартное отклонение в диапазоне данных 100 предложений выборки каждого априорного класса. Результаты вычислений представлены в таблице 7.

Табл. 7. Результаты вычислений среднего значения и стандартного отклонения для параметров априорного набора параметров.

Параметры	Довлатов		Цион		Матюшкин-Герке	
	Среднее значение	Стандартное отклонение	Среднее значение	Стандартное отклонение	Среднее значение	Стандартное отклонение
X1	6,66	3,934	7,99	5,454	11,4 3	8,02
X2	35,2 2	20,288	40,9 2	28,139	61,9 3	42,419
X3	6,17	3,385	5,57	3,841	8	4,843
X4	1,15	0,435	1,43	0,756	1,46	0,797

X5	1,11	0,399	0,9	0,362	1,13	0,418
X6	0,16	0,615	0,39	0,815	0,24	0,712
X7	0	0	0,08	0,339	0,04	0,243
X8	0,04	0,197	0,15	0,458	0,33	0,667
X9	0,04	0,197	0,12	0,356	0,26	0,525
X10	0	0	0,03	0,171	0,04	0,197
X11	0	0	0	0	0,01	0,1
X12	0	0	0,01	0,1	0	0
X13	0,4	0,586	0,49	0,732	0,34	0,555
X14	0	0	0	0	0,03	0,223
X15	0	0	0,03	0,171	0	0
X16	0	0	0,01	0,1	0	0
X17	5,22	3,129	6,21	4,142	8,76	6,035
X18	1,44	1,25	1,78	1,845	2,7	2,303
X19	2,14	1,891	2,25	1,93	3,17	2,257
X20	0,51	0,759	0,68	0,898	0,85	0,989
X21	0,63	0,734	0,73	0,93	1,07	1,289
X22	1,25	0,77	1,27	1,043	1,68	1,254
X23	0,15	0,359	0,22	0,596	0,53	0,858
X24	0,47	0,674	0,54	0,797	0,82	0,925
X25	0,75	0,716	0,83	1,016	1,41	1,272
X26	0,43	0,671	0,7	0,948	0,96	1,034
X27	0,06	0,239	0,2	0,471	0,4	0,696
X28	0,4	0,636	0,5	0,847	0,57	0,655
X29	0	0	0,11	0,345	0,04	0,243
X30	0,61	1,004	0,55	0,796	0,87	1,125
X31	0,06	0,239	0,13	0,338	0,15	0,435
X32	0,73	0,633	0,73	0,815	1,11	0,84
X33	0,24	0,429	0,29	0,537	0,32	0,51
X34	0,25	0,52	0,24	0,452	0,33	0,514

X35	0,58	1,319	0,58	1,112	0,75	1,306
X36	0,33	0,865	0,37	0,917	0,58	1,103
X37	0,18	0,642	0,02	0,141	0,09	0,452
X38	0	0	0,03	0,171	0,04	0,197
X39	0	0	0,11	0,65	0,2	1,146
X40	0	0	0	0	0,03	0,171
X41	0	0	0	0	0,05	0,359
X42	0,65	0,869	0,63	1,012	0,79	1,057
X43	0	0	0,04	0,197	0,07	0,256
X44	0,04	0,197	0,04	0,243	0,14	0,472
X45	0	0	0,09	0,9	0,11	1,1
X46	0,08	0,273	0,09	0,288	0,27	0,489
X47	0,17	0,667	0,2	0,804	0,82	1,731
X48	0	0	0	0	0	0
X49	0	0	0	0	0	0
X50	0	0	0	0	0	0
X51	0	0	0	0	0	0
X52	0,92	1,041	0,81	0,961	1,39	1,483
X53	0,73	0,863	1,04	1,044	1,2	1,11
X54	2,14	2,667	2,95	3,301	3,51	3,889
X55	1,79	2,212	2,22	2,611	2,81	3,126
X56	0,35	0,592	0,73	0,908	0,7	1,04

Для определения информативного набора параметров вычислим Т-критерий Стьюдента для каждого параметра каждой из пар априорных классов К1/К2, К2/К3, К1/К3 по формуле:

где – среднее значение параметра первого априорного класса, - среднее значение параметра второго априорного класса, - стандартное отклонение параметра первого априорного класса, - стандартное отклонение параметра второго априорного класса, , – количество предложений (равное 100). Результаты для пар К2 (Цион) / К3(Герке), К2 (Цион) / К1 (Довлатов) и К3 (Герке) / К1 (Довлатов) представлены в таблице 8.

Табл. 8 Попарное вычисление значения t-критерия Стьюдента

Параметры	К2/К3	К2/К1	К3/К1
X1	2,894	1,900	4,986
X2	1,010	0,823	3,196
X3	3,928	1,203	3,056
X4	0,366	3,613	3,799
X5	6,061	5,557	0,461
X6	1,733	2,697	1,067
X7	1,160	2,360	1,646
X8	3,582	2,382	4,323
X9	3,650	2,223	4,160
X10	0,572	1,754	2,030
X11	100,000	-	1,000
X12	1,000	1,000	-
X13	1,966	1,174	1,018
X14	60,327	-	1,345
X15	1,754	1,754	-
X16	1,000	1,000	-
X17	3,276	1,933	5,047

X18	3,873	1,699	5,124
X19	3,771	0,481	3,939
X20	1,707	1,779	3,249
X21	3,102	1,016	3,277
X22	3,416	0,181	3,274
X23	4,630	1,150	4,363
X24	3,173	0,831	3,607
X25	4,924	0,750	4,988
X26	2,465	2,719	4,912
X27	3,850	2,937	4,845
X28	0,787	1,127	2,444
X29	1,995	3,188	1,646
X30	3,469	0,669	2,121
X31	0,560	2,037	2,042
X32	4,291	0,000	4,318
X33	0,533	0,900	1,514
X34	1,881	0,209	1,479
X35	1,325	0,000	1,149
X36	2,022	0,403	2,123
X37	4,336	8,781	1,825
X38	0,572	1,754	2,030
X39	1,152	1,692	1,745
X40	102,596	-	1,754
X41	38,795	-	1,393
X42	1,424	0,184	1,236
X43	1,474	2,030	2,734
X44	3,770	0,000	2,101
X45	0,196	1,000	1,000
X46	5,771	0,338	3,827
X47	5,618		3,661

X48	-	-	-
X49	-	-	-
X50	-	-	-
X51	-	-	-
X52	4,911	1,029	2,953
X53	1,371	2,772	3,968
X54	1,163	2,019	2,978
X55	1,644	1,387	2,821
X56	0,295	4,029	3,256

Полужирным шрифтом выделены значения, превышающие 1,96. Они являются статистически значимыми при критическом значении = 0,05. Как можно видеть из представленной выше таблицы существует четыре параметра, значения которых превышает 1,96 у всех трех пар априорных классов. Они и составят информативный набор параметров (см табл. 9). Все остальные параметры являются нерелевантными для различения данных классов.

Табл. 9. Информативный набор параметров.

X08	число подчиненных предложений
X09	число подчиненных предложений 1-й степени
X26	число союзов
X27	число подчинительных союзов

Поскольку объемы атрибутируемых текстов невелики, воспользуемся методом сплошного обследования. Это повысит точность определения координат объектов. Проведем сплошной анализ объектов атрибуции (см. табл.1) по параметрам из информативного набора параметров. Вычислим среднее значение и стандартное отклонение четырех параметров на объеме всех предложений каждого из

рассматриваемых псевдонимных текстов. Полученные матрицы значений представлены в приложении 11 «Матрицы объектов атрибуции» (листы 1-27).

2.6 Определение объема выборки

В связи с большими объемами классов воспользуемся выборочным методом при определении координат эталонов классов, уточнив объём выборки априорных классов по формуле:

где $\alpha = 0.05$, N – мощность априорного класса, \bar{x} – среднее значение параметра априорного класса, σ – стандартное отклонение параметра первого априорного класса.

Результаты вычислений объема выборки для каждого параметра из информативного набора параметров представлены в таблицах 10, 11, 12. В качестве окончательного объема выборки для каждого априорного класса возьмем наибольшее значение.

Табл. 10. Определение объема выборки класса K1 (Довлатов)

Параметры	Среднее значение	Стандартное отклонение	Объем выборки
X8	0,040	0,197	696
X9	0,040	0,197	696
X26	0,430	0,671	423
X27	0,060	0,239	670

Табл. 11. Определение объема выборки класса К2 (Цион)

Параметры	Среднее значение	Стандартное отклонение	Объем выборки
X8	0,150	0,458	294
X9	0,120	0,356	293
X26	0,700	0,948	222
X27	0,200	0,471	279

Табл. 12. Определение объема выборки класса К3 (Матюшкин-Герке)

Параметры	Среднее значение	Стандартное отклонение	Объем выборки
X8	0,330	0,667	316
X9	0,260	0,525	316
X26	0,960	1,034	212
X27	0,400	0,696	296

Таким образом, объем выборки класса К1 равен 696 предложениям, К2 - 294, К3 - 316.

2.7 Детерминированный алгоритм распознавания

Для выполнения детерминированного алгоритма распознавания сначала необходимо установить координаты эталонов априорных классов. Увеличим матрицы соответствующих априорных классов до полученных значений выборки и вычислим значения параметров из информативного набора параметров. Полученные результаты занесем в таблицы (см. приложение 8 «Матрица априорного класса К1 (Довлатов)», приложение 9

«Матрица априорного класса К2 (Цион)» и приложение 10 «Матрица априорного класса К3 (Матюшкин-Герке)»).

Вычислим для каждого параметра из информативного набора параметров среднее значение и стандартное отклонение в диапазоне полученных объемом выборок. Результаты вычислений представлены в таблице 13.

Табл. 13. Результаты вычислений среднего значения и стандартного отклонения для параметров информативного набора параметров.

К1 (Довлатов)		
Параметры	Среднее значение	Стандартное отклонение
X8	0,079	0,300
X9	0,074	0,273
X26	0,435	0,690
X27	0,112	0,354
К2 (Цион)		
Параметры	Среднее значение	Стандартное отклонение
X8	0,1423	0,495
X9	0,122	0,411
X26	0,632	0,964
X27	0,180	0,540
К3 (Матюшкин-Герке)		
Параметры	Среднее значение	Стандартное отклонение
X8	0,322	0,692
X9	0,262	0,549
X26	0,949	1,172
X27	0,395	0,733

Используя данные значения и результаты, полученные при вычислении среднего значения и стандартного отклонения параметров каждого из объектов атрибуции, вычислим t -критерий Стьюдента. Значение t -критерия вычисляется для каждой пары объект атрибуции/потенциальный автор. Для того, чтобы атрибутируемый текст можно было однозначно отнести к одному из априорных классов необходимо, чтобы значение T -критерия было меньше 1,96 при уровне значимости = 0,05 для всех параметров в одной из пар объект/автор и больше 1,96 в двух других парах. Полученные результаты для каждого объекта атрибуции представлены в приложении 11 «Матрицы объектов атрибуции» (листы 1-27). Пример вычисления T -критерия представлен в таблице 8:

Табл. 14. Детерминированный алгоритм распознавания объекта 7.

Параметры	К2 (Цион)	К3 (Матюшкин - Герке)	К1 (Довлатов)
X8	1,338	4,155	0,208
X9	1,183	3,440	0,119
X26	1,166	3,505	0,400
X27	0,593	3,382	0,395

Полужирным шрифтом выделены значения, не превышающие 1.96. Как можно заметить, объект 7 можно отнести как к классу 2, так и к классу 1, но нельзя отнести к классу 3. Рассмотрев результаты детерминированного алгоритма распознавания всех объектов атрибуции, можно заметить, что похожая ситуация наблюдается для объектов 9, 13, 18, 19, 21, 23.

Однако ни один из объектов нельзя однозначно отнести к одному из априорных классов. Таким образом, в данном случае наблюдается отказ от распознавания и остановка детерминированного алгоритма. Для проведения атрибуции псевдонимных текстов и решения поставленной

задачи по установлению их авторства необходимо воспользоваться вероятностным алгоритмом.

2.8 Вероятностный алгоритм распознавания.

Вероятностный алгоритм распознавания приписывает вероятность принадлежность объекта к определенному классу и вычисляется по формуле:

где d_{ij} - расстояние между объектом и i -м классом, d_i - расстояние между объектом и остальными классами.

В качестве функции расстояния между объектом атрибуции и априорным классом в многомерном параметрическом пространстве используется взвешенное евклидово расстояние, которое вычисляется по формуле:

где n – размерность евклидова пространства, a и b – точки в пространстве с координатами a_j , b_j , весовой коэффициент w_j j -й переменной.

Результаты вычислений представлены в приложении 11 «Матрицы объектов атрибуции» (листы 1-27) и в приложении 12 «Матрица расстояний».

Решающее правило вероятностного алгоритма заключается в том, что объект можно однозначно отнести к определенному классу в том случае, если вероятность такого исхода больше 0,5. Если вероятность

такого исхода меньше 0,5, то объект нельзя однозначно отнести к данному классу, можно лишь говорить о большей/меньшей доли вероятности данного исхода.

Результаты работы вероятностного алгоритма представлены в таблице 9:

Табл. 15. Вероятностный алгоритм распознавания.

Объекты	Классы		
	К2 (Цион)	К3 (Матюшкин - Герке)	К1 (Довлатов)
1	0,311	0,139	0,550
2	0,513	0,153	0,334
3	0,492	0,201	0,306
4	0,335	0,145	0,521
5	0,193	0,666	0,141
6	0,314	0,181	0,505
7	0,219	0,064	0,717
8	0,294	0,095	0,611
9	0,178	0,064	0,758
10	0,578	0,193	0,228
11	0,267	0,518	0,215
12	0,358	0,130	0,512
13	0,426	0,280	0,295
14	0,306	0,516	0,177
15	0,292	0,150	0,558
16	0,445	0,209	0,346
17	0,529	0,189	0,282
18	0,432	0,141	0,428
19	0,120	0,043	0,836

20	0,576	0,209	0,215
21	0,252	0,105	0,643
22	0,264	0,531	0,205
23	0,191	0,069	0,740
24	0,421	0,259	0,321
25	0,301	0,161	0,538
26	0,172	0,708	0,120
27	0,295	0,156	0,549

Полужирным шрифтом в таблице выделены наибольшие значения, позволяющие отнести объект к определенному классу. В большинстве случаев эти значения превышают 0,5, а значит данные тексты можно однозначно отнести к определенному автору. Однако объекты 3, 13, 16, и 24 имеют вероятность меньшую 0,5, но больше 0,4. В данных случаях можно говорить, что эти объекты с большой долей вероятности принадлежат соответствующим классам. В случае с объектом 18 однозначного решения вынести нельзя, слишком мала разница вероятности принадлежности классу 1 (Довлатов) - 0,428 и классу 2 (Цион) - 0,432.

Таким образом, по результатам вероятностного алгоритма можно составить матрицу решений (см табл. 10), где принадлежность объекта определенному классу обозначается как 1, а не принадлежность - 0.

Табл. 16. Матрица решений

Объект ы	Классы		
	К1 (Цион)	К2 (Матюшкин-Герке)	К3 (Довлатов)
1	0	0	1
2	1	0	0
3	1	0	0

4	0	0	1
5	0	1	0
6	0	0	1
7	0	0	1
8	0	0	1
9	0	0	1
10	1	0	0
11	0	1	0
12	0	0	1
13	1	0	0
14	0	1	0
15	0	0	1
16	1	0	0
17	1	0	0
18	0	0	0
19	0	0	1
20	1	0	0
21	0	0	1
22	0	1	0
23	0	0	1
24	1	0	0
25	0	0	1
26	0	1	0
27	0	0	1

2.9 Интерпретация результатов атрибуции

Анализ данных, содержащихся в таблице 10 позволяет отвергнуть нулевую гипотезу, в которой утверждалось то, что все псевдонимные произведения принадлежат Сергею Довлатову. Нашла подтверждение альтернативная гипотеза (H₂) - часть произведений принадлежит С.Довлатову, а часть – еще одному или нескольким писателям.

Таким образом, перу Сергея Довлатова принадлежат следующие произведения «Энтузиаст» (Советская Эстония. 03.09.1972. С.4), «Здрасьте, это Вася говорит...» (Вечерний Таллин. 16.01.1974. №13 (443). С. 3), «Случай с Барашкиным» (Советская Эстония. 28.01.1973. С. 4), «Как я бросил пить» (Советская Эстония. 11.02.1973. С. 4), «Звонок из Антимира» (Советская Эстония 25.03.1973. С. 4), «Да здравствует петух!» (Вечерний Таллин. №71. 21.10.1972. С.3), «Не лезь в бутылку. (Полуфантастическая история)» (Вечерний Таллин. 01.11.1972. С.3), «Последний сапожник» (Вечерний Таллин. №11 (138). 13.01.1973. С.3), С. 3), «Сколько лет бабушке?» (Вечерний Таллин. 13.03.1974. С.3), «В ожидании генерала. (Действительное происшествие)» (Вечерний Таллин. № 119 (549). 25.05.1974. С.3), «История» (Вечерний Таллин. 15.12.1973. №290 (417). С.3), «Пончо» (Вечерний Таллин. 12.09.1973. С.3), «Чуткость» (Вечерний Таллин. 12.09.1973. С. 3). Всего 13 произведений.

Владимир Цион является автором публикаций «Заведу себе бар...» (Морьяк Эстонии. 21.04.1973. С. 4), «Мужские хлопоты» (Вечерний Таллин. 07.03.1974. С.4), «Муха» (Вечерний Таллин. 27.10.1973. С.3), «Первый «заяц» (Советская Эстония. 01.01.1973. С.4), «Человек изголодался» (Вечерний Таллин. 14.09.1974. №215 (645) С. 3), «Не тот уровень» (Вечерний Таллин. 18.11.1972. С.3), «Ха-ха!» (Вечерний Таллин. 28.02.1973. С.3), «Зонт» (Советская Эстония. 02.09.1973. С. 4). Всего 8 произведений.

Александр Матюшкин-Герке написал «Сон» (Вечерний Таллин. 14.11.1973. №264 (391), «Астрономия» (Вечерний Таллин. 09.02.1974. №34

(464). С. 3), «Труба» (Вечерний Таллин. 28.10.1972. С.3), «Такая специфика» (Вечерний Таллин. №106. 06.12.1972. С.3), «Хочу всё знать!» (Вечерний Таллин. 10.10.1973. С. 3). Всего пять произведений.

Авторство фельетона «Почти что драма про некую маму, ее малолетнего сына Игната и директора лесокombината.», опубликованного 05.01.1974 в газете «Вечерний Таллин» под псевдонимом «И. Олеп» достоверно установить не удалось.

Заключение

В данной работе был представлен опыт применения математических методов атрибуции текстов на примере проблемы авторства псевдонимных фельетонов и юморесок, приписываемых С.Довлатову.

Обзор истории данного вопроса позволил сформулировать атрибуционную гипотезу, отражающую возможность написания данных спорных произведений В.Ционом, А.А.Матюшкиным-Герке и С.Довлатовым. Класс объектов атрибуции составили 27 спорных произведений. Три априорных класса (К1 «Довлатов» - 29 текстов, К2 «Цион» - 11 текстов, К3 «Матюшкин-Герке» - 15 текстов) были описаны на языке 56 априорных параметров. Информативный набор параметров составили 4 параметра: X08 (число подчиненных предложений), X09

(число подчиненных предложений 1-й степени), X26 (число союзов) и X27 (число подчинительных союзов).

В результате проведенной атрибуции, работы детерминированного и вероятностного алгоритмов, подтвердилась альтернативная гипотеза о том, что часть атрибутируемых произведений принадлежит С.Довлатову, а часть – еще одному или нескольким писателям. Интерпретация полученных результатов показала, что с высокой долей вероятности 5 текстов могут быть отнесены Александру Матюшкину-Герке, 8 - Владимиру Циону и 13 - Сергею Довлатову. Авторство одного из объектов атрибуции (фельетона «Почти что драма про некую маму, ее малолетнего сына Игната и директора лесокombината.») точно установить не удалось.

Таким образом, заявленная цель была достигнута. Практическая и научная значимость данной работы состоит в том, что она является первой попыткой описания индивидуального стиля Сергея Довлатова с помощью математических методов, а также первым исследованием раннего творчества писателя с использованием теории распознавания образов. Полученные результаты могут стать примером для дальнейших исследований при изучении творчества писателя.

Список литературы

- 1.Доброзракова Г.А. Псевдонимные юморески и фельетоны Сергея Довлатова. К вопросу атрибуции // Вопросы литературы. Вып. 4. 2014. С.21-54
- 2.Доброзракова Г.А. Журналистская деятельность С. Довлатова (доэмигрантский период) // Мир науки, культуры, образования. Вып. 1 (56). 2016. С. 311-313
- 3.Виноградов В. В. Проблема авторства и теория стилей. - М.: Гослитиздат, 1961. - 614 с.
- 4.Мартыненко Г. Я. Основы стилеметрии. - Л.: Изд-во Ленингр. ун-та, 1988. - 173 с.

5. Мартыненко Г. Я. Многомерный синтаксический анализ художественной прозы // Структурная и прикладная лингвистика: межвуз. сборник / под ред. А. С. Герда. — Л.: Изд-во Ленингр. ун-та, 1983. Вып.2. - С. 47-61.
6. Марусенко М.А. Атрибуция анонимных и псевдонимных литературных произведений методами распознавания образов: дис. ... д-ра филолог. наук: 10.02.19 / М.А. Марусенко. – Л.: Изд-во ЛГУ, 1990. - 370 С. + Автореф. 26 С.
7. Марусенко М.А. Атрибуция анонимных и псевдонимных литературных произведений методами теории распознавания образов. - Л.: Изд-во Ленингр. ун-та, 1990. 164 с.
8. Марусенко М.А., Бессонов Б.Л., Богданова Л.М., Аникин М.А., Мясоедова Н.Е. В поисках потерянного автора. Этюды атрибуции. - СПб.: Филол. фак. С.-Петербур. гос. ун-та, 2001. - 209 с.
9. Методы атрибуции / Применяемая методика. - [Электронный ресурс]. URL: <http://gary.corneille-moliere.com/?p=page15&m=main&l=rus> (дата обращения: 10.05.2016)
10. Милов Л. В., Бородкин Л. И., Иванова Т. И. и др. От Нестора до Фонвизина. Новые методы определения авторства. - М., 1994.
11. Понятие атрибуции - [Электронный ресурс]. URL: http://corneille-moliere.com/?p=attribution_methods&m=main&l=rus (дата обращения: 10.05.2016)
12. Родионова Е.С. Методы атрибуции художественных текстов // Структурная и прикладная лингвистика. Вып. №7: Межвуз. сб. / Под. Ред. А.С.Герда. – СПб.: Изд-во С-Петер. Ун-та. 2008. С.118-127.
13. Родионова Е.С. Отбор информативных параметров при атрибуции стихотворных пьес Мольера // Материалы XXXVI Международной филологической конференции (12 – 17 марта 2007 г.). – СПб: Филол.

- фак. С.-Петербур. гос. ун-та, 2007. – Вып. 10: Прикладная и математическая лингвистика / под ред. Т. Г. Скребцовой. С. 67–74
14. Родионова Е.С. Параметризация стилей: отбор информативных параметров при атрибуции пьес Мольера // Вестник Санкт-Петербургского университета. Сер 9. Филология. Востоковедение. Журналистика. Вып. 2 – Ч. 2. - 2007. С 61-67.
15. Севбо И. П. Графическое представление синтаксических структур и стилистическая диагностика. - Киев: Наукова думка, 1981. - 192 с.
16. Тулдава Ю. А. Проблемы и методы квантитативно-системного исследования лексики. - Тарту: Тартуский гос. ун-т, 1987. -203 с.
17. Якубайтис Т. А. Вероятностная атрибуция типа текста по нескольким морфологическим признакам. - Рига: ИЭВТ, 1982. - 53 с.