

«Центральный Комитет Всесоюзной Коммунистической партии (большевиков) постановил»: сравнительный анализ отчетов ЦК VI-XXVIII съездов

Безрукова А.И.

Аннотация

Данная статья посвящена изучению советского авторитетного дискурса на примере Отчетных докладов ЦК в соответствии с концепцией, выдвинутой А. Юрчаком в книге «Это было навсегда, пока не кончилось. Последнее советское поколение» [4]. Главная цель исследования – подтвердить или опровергнуть тезис Юрчака о самовоспроизводстве и ретрансляции дискурса в период позднего социализма с помощью цифровых методов.

В ходе исследования в текстах Отчетных докладов ЦК периода позднего социализма были выявлены процессы лексической и стилистической унификации советского политического языка, приведшие к не просто его нормализации и стандартизации, но и анонимизации. Также для отчетов ЦК было посчитано косинусное расстояние между ними и получено, что все тексты, начиная со Сталина, лексико-семантически близки друг другу.

Подобные результаты позволяют говорить о том, что на рубеже 1920 – 1930-х годов в СССР произошла «сталинизация» дискурса – появилась негласная норма того, как должны были быть написаны официальные тексты. Поддержание этой нормы в период позднего социализма и привело к тому, что «различные тексты ... все больше походили на цитаты из неких предыдущих текстов, а значит, все больше походили друг на друга» [4].

Keywords: советский политический язык, авторитетный дискурс, ритуализованная речь, ретрансляция дискурса, векторные модели, Юрчак, digital humanities

1 Введение

В книге «Это было навсегда, пока не кончилось. Последнее советское поколение» антрополог Алексей Юрчак выдвинул тезис о гипернормализации советского политического языка («авторитетного дискурса») во времена позднего социализма (1950 – 1980-е). Явление «гипернормализации» по Юрчаку включает в себя не только нарастающую нормализацию, стандартизацию и цитируемость форм авторитетного языка (когда фразы, выражения и даже

целые отрывки копировались из старых текстов в новые и из одного контекста в другой практически без изменений), но и их постепенное усложнение за счет эффекта «снежного кома», когда желание оставаться в рамках тщательного воспроизводства языковых оборотов этого языка приводило к не просто детальному копированию их формы, но копированию этой формы «с запасом». Как следствие, в советском дискурсе коренным образом изменилась и позиция автора – из производителя дискурса он превратился в ретранслятора уже существующего знания. Подобный вектор развития советского дискурса, обусловленный вышеперечисленными характеристиками, привел к тому, что он перестал описывать реальность, так как произошел перформативный сдвиг от констатирующей составляющей смысла к его перформативной составляющей.

Концепция Юрчака до сих пор вызывает много споров в профессиональном сообществе, так как предоставляет новую точку зрения на исследования Советского Союза, уходя от имеющих место в существовавших до этого культурологических и социальных исследованиях бинарных оппозиций. Она также дает свое объяснение такому событию как крах СССР в 1991 году, акцентируя внимание на незаметных для многих внутренних преобразованиях, протекавших в советской системе во время позднего социализма, которые в итоге сделали его возможным и даже необратимым – по мнению Юрчака, именно из-за повсеместного повторения одних и тех же форм авторитетного языка советская система выглядела для многих монотонной, единой и неизменной, в то время как на самом деле, из-за утраты основного свойства дискурса – описания реальности – была хрупкой настолько, что к её краху все были подготовлены.

Данное исследование не ставит перед собой задачу подтвердить все, что пишет о советском дискурсе Юрчак – вместо этого, оно концентрируется лишь на советском политическом языке (как его составляющей) и гипотезе о ретрансляции и самовоспроизводстве советского авторитетного дискурса, то есть появлении нормы языка, когда «различные тексты, написанные на этом языке, все больше походили на цитаты из неких предыдущих текстов, а значит, все больше походили друг на друга».

Советский язык был объектом научного интереса с самого момента своего появления: как в 1920-х годах (С.О. Карцевский [7], В.Б. Шкловский [6], Эйхенбаум Б.М. [5] и Ю.Н. Тынянов [9]), так и в наше время (С. Коткин [2], И. Г. Земцов [1], Н. Б. Вахтин, А.А. Кожаева [3], Т.В. Савина [8]), многие исследователи анализировали его и его влияние на жизнь и мышление советского гражданского общества, являющееся ключом к пониманию этого периода мировой истории.

2 Постановка задачи

Главной задачей этого исследования является подтверждение или опровержение гипотезы А. Юрчака о самовоспроизводстве и ретрансляции советского

авторитетного дискурса периода позднего социализма путем анализа Отчетных докладов ЦК как его примеров, а также оценки их близости при помощи цифровых методов. Параллельно будут выявлены достоинства и недостатки данного подхода и в целом оценена возможность его применения к изучению советского дискурса.

Приведенные цифры (в силу их воспроизводимости) и выводы могут быть использованы в дальнейших исследованиях лингвистических советологов с целью изучения специфики русского языка в советский период.

3 Анализируемые данные

Так как для адекватной оценки того, что происходило с советским авторитетным дискурсом в 1950–1980-е годы, необходимо понимать как то, что было до этого периода, так и то, что было после, в качестве изучаемых данных были взяты тексты Отчетных докладов (до 1934 года - Политических отчетов) ЦК в период с VI по XXVIII съезд. Подобные границы обусловлены общим числом съездов и тем, что до VI съезда процедура чтения доклада ЦК значительно отличалась от принятой позднее - в частности, на V съезде докладчиков было трое (от меньшевиков, от большевиков и от Бунда).

Все данные были взяты из открытых источников и предварительно обработаны для того, чтобы они:

1. адекватно читались программой (то есть были удалены все таблицы)
2. отражали только то, что говорилось докладчиком (то есть были убраны примечания редактора)
3. не содержали лишних данных (то есть были убраны комментарии из зала и ссылки)

4 Применяемые методы

Главным инструментом исследования является язык Python - именно его возможности позволили провести как анализ текстов, так и оценку их близости.

Для анализа были использованы следующие библиотеки: nltk и codecs (для обработки естественного языка); numpy и pandas (для подсчета соответствующих метрик); matplotlib и wordcloud (для визуализации результатов).

Для оценки близости отчетов ЦК были использованы следующие библиотеки: nltk и rymorphu3 (для обработки естественного языка); numpy и pandas (для хранения текстов в виде базы данных); gensim и sklearn (для использования нейросети Word2vec и методов машинного обучения: алгоритмов k-means и k-NN); supervenn, matplotlib и plotly (для визуализации результатов)

5 Ход исследования

Настоящее исследование было разделено на два различных по задачам блока: в ходе экспериментов первого блока изучались стилистические и лексические особенности текстов, а в ходе экспериментов второго блока ставилась цель проверить утверждение о ретрансляции дискурса путем проведения тестов на лексико-семантическую близость текстов.

Первый блок исследования представляет собой следующую цепочку действий:

1. Чтение данных из файлов и их предобработка (все слова были приведены к начальной форме)
2. Подсчет различных текстовых характеристик: количества запятых на 1000 символов текста, лексического разнообразия, частоты употребления заданных штамповых лексем и знаков препинания
3. Построение облаков слов для каждого текста

Второй блок исследования представляет собой следующую цепочку действий:

1. Чтение данных из файлов (по одному на каждый съезд) и их предобработка (все слова были приведены к начальной форме, знаки препинания - убраны)
2. Создание корпуса, в котором хранились тексты и некоторые данные о них (год написания, автор)
3. Эксперимент 1, в ходе которого была проведена кластеризация текстов отчетов. В ходе эксперимента были использованы TfidfVectorizer (для получения вещественного представления текстов), PCA (для понижения размерности) и KMeans (для выполнения задачи кластеризации)
4. Эксперимент 2, в ходе которого было проведено разделение текстов отчетов на два класса «Ленин/Сталин». В ходе эксперимента были использованы TfidfVectorizer (для получения вещественного представления текстов) и KNeighborsClassifier (для выполнения задачи классификации)
5. Эксперимент 3, в ходе которого была посчитана косинусная близость между всеми текстами Отчетных докладов ЦК и найдены «описывающие» слова. Векторы текстов считались как сумма векторов употребляющихся в них слов. В ходе эксперимента была использована векторная модель word2vec, предобученная на НКРЯ (для векторизации текстов)

6. Эксперимент 4, в ходе которого была посчитана косинусная близость между авторами и найдены «описывающие» слова. В качестве текстов брались все отчеты, принадлежавшие одному автору, векторы считались так же, как и в предыдущем эксперименте. В ходе эксперимента была использована векторная модель word2vec, предобученная на НКРЯ (для векторизации текста)

6 Результаты

6.1 Анализ советского политического языка

Во второй главе книги «Это было навсегда пока не кончилось» [4], «Гегемония формы», утверждается, что «в результате всеобщей имитации и копирования из советского авторитетного языка постепенно вытеснялся авторский голос. Авторитетный язык стал анонимным языком». Иначе говоря, по мнению Юрчака, в период позднего социализма «авторский голос» не проявлялся ни на уровне смысла (того, *что* говорил человек), ни на уровне формы (того, *как* он говорил).

Главный вопрос исследования: из-за чего возникало это ощущение монолитности? Могло ли быть так, что подобный эффект достигался только за счет использования авторами одного и того же лексикона, в то время как их индивидуальность проявлялась, но в других аспектах: использовании знаков препинания или построении предложений? Или же анонимизация произошла на всех уровнях текста, и авторы действительно перестали быть различимыми между собой?

Для того, чтобы понять это, было необходимо провести комплексный анализ советского политического языка и оценить степень его нормализации, постаравшись найти как общие тренды, так и индивидуальные особенности у его носителей.

И те и другие проявляются на многих уровнях текста - в синтаксисе, лексике, семантике. Иногда это заметные читателю приемы, но большая их часть все же остается вне поля его зрения. Такие исследования как стилометрический анализ (и родственные ему) выявляют то, что достаточно сложно заметить или посчитать вручную. Это, например, средняя длина предложения или лексическое разнообразие автора.

6.1.1 «Облака слов»

Чтобы оценить степень нормализации советского политического языка на уровне лексики, мы воспользовались методом wordcloud (облако слов). Облако слов - это простой, но мощный объект визуального представления текста, который отображает наиболее часто встречающиеся слова более крупными и жирными буквами и разными цветами.

В ходе исследования были построены и проанализированы облака слов для всех изучаемых съездов, представленные на Рис. 1. Так, например, лексикон Ленина полон слов из военной риторики («армия», «война», «передышка»), глаголов («мочь», «стать», «сделать»), вовлекающих слушателей местоимений («мы», «наш»), в то время как лексикон Брежнева менее эмоционально окрашен и по своему характеру номинативен: чаще употребляются слова «партия» и «народ», акцент сместился на экономическую сферу жизни страны («производство», «развитие», «работа», «экономический»), глаголов практически нет. Влияние времени в текстах также проявляется: это и «НЭП» у Зиновьева, и «США» у Хрущева, и «перестройка» у Горбачева.

Если пронаблюдать за ситуацией в хронологическом порядке, можно заметить, что в то время как лексикон Ленина а) разнообразен (наборы слов для различных речей пересекаются, но не совпадают) и б) уникален (самые популярные слова советского политического языка впервые употребляются не им, а Зиновьевым или Сталиным), у его последователей, начиная со Сталина, уже появляется стандартный набор фраз и слов, кочующих от автора к автору, и общие тренды (снижение числа глаголов в речи, меньшее эмоциональное вовлечение). Пересечений между текстами также становится все больше – и если в отчетах с XXIII по XXVI съезд это обусловлено тем, что на всех из них докладчиком был один и тот же человек, то вот огромное число пересечений у выступлений на XVI (Сталин), XIX (Маленков), XX (Хрущев) и XXIII (Брежнев) съездах явно свидетельствует о формировании и закреплении в политическом языке особых слов и оборотов, без которых невозможно представить ни один официальный текст.

Таким образом, несмотря на то, что в докладах проявляются следы уникальности и актуальности (отражаются исторические процессы), главное место в них все же занимают слова из некоего общего словаря, без которого ни одно из обращений не мыслимо, за счет чего и создается эффект «унификации» дискурса.

6.1.2 Лексическое разнообразие

Еще одним статистическим параметром, значения которого помогают отделить одного автора от другого, является лексическое разнообразие – то есть соотношение числа уникальных слов к их общему количеству.

На Рис. 2 видно, что среди полученных графиков есть как те, что лежат в одном диапазоне (Хрущев и Брежнев, Ленин и Зиновьев), так и те, что сильно отличаются по уровню лексического разнообразия между собой (Горбачев, Маленков и Сталин). Это говорит о том, что в общей совокупности отчеты различимы, но если сконцентрировать внимание лишь на периоде позднего социализма (с момента смерти Сталина и до прихода Горбачева), то с точки зрения лексического разнообразия отличий между докладами ЦК не найдено.

Рис. 1: Облака слов для VI-XXVIII съездов

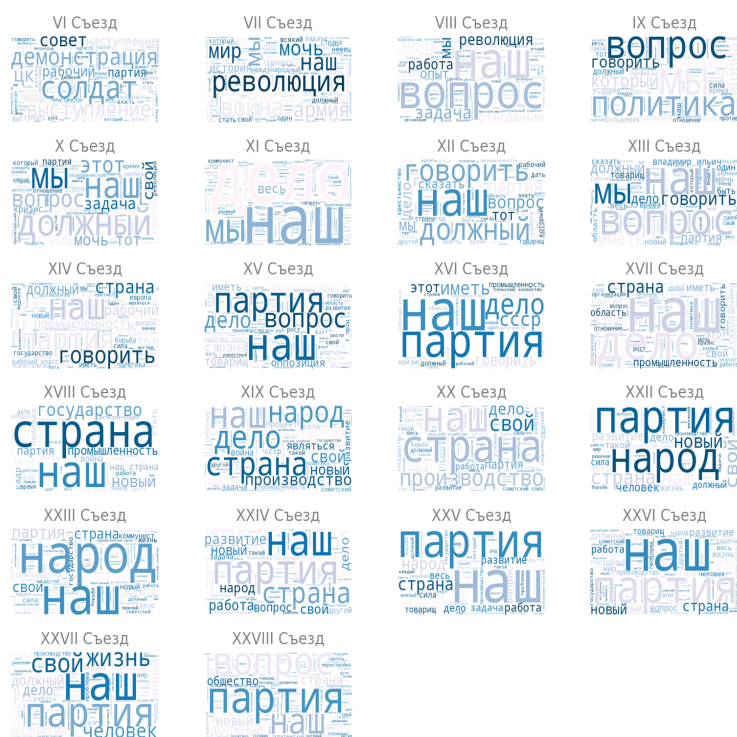
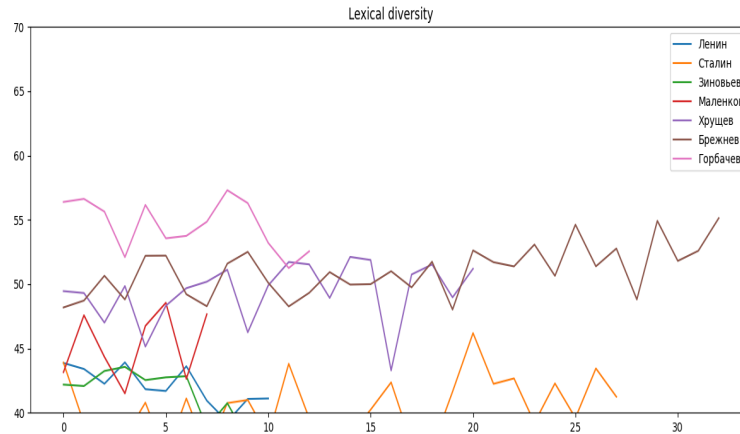


Рис. 2: Лексическое разнообразие авторов



6.1.3 Исследование пунктуации

Еще один уровень текста, на котором может проявиться «авторский голос» - пунктуационный, то есть через знаки препинания. То, как человек использует запятые и точки, многое может сказать о стиле его письма.

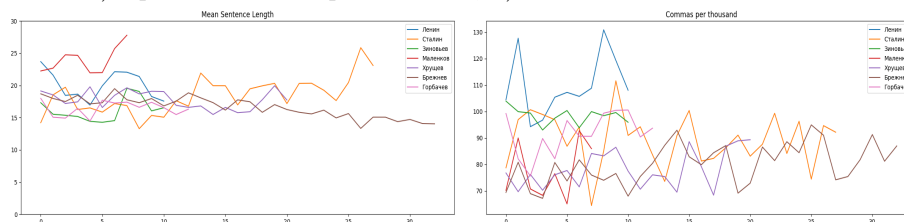
Для анализа пунктуационных особенностей отчетов были рассмотрены две метрики: изменение средней длины предложения и изменение частотности употребления запятых на 1000 символов текста для каждого автора. Если графики будут находиться на одном уровне, то можно будет говорить о том, что нормализация затронула и то, как авторы строили предложения. Иначе же мы будем вынуждены признать, что на пунктуационном уровне в текстах проявилась, пусть и скрытно, авторская индивидуальность.

Как мы видим из Рис. 3а, у большей части авторов средняя длина предложения колеблется между 15 и 20, сохраняя некоторую стабильность. Однако есть и выбивающиеся из общего тренда графики - например, у Маленкова значения не опускаются ниже 20, а порой превышают и 25. Также стоит отметить, что в отчетах Ленина и Сталина есть место некоторой индивидуальности - у первого средняя длина предложения колеблется между значениями, равными 16 и 24 (то есть превышает «среднее по больнице»), а у второго необычен характер графика, так как он растет от начала к концу (то есть средняя длина предложения от отчета к отчету у Сталина увеличивалась).

Рис. 3б демонстрирует схожую картину - несмотря на отличие от большинства графиков Ленина, Зиновьева и Горбачева, все остальные расположены в общем диапазоне от 70 до 90 запятых на тысячу знаков. Стоит отдельно отметить и нестабильность графиков - во всех наблюдаются колебания с большим разбросом (у Сталина, например, он достигает пятидесяти знаков).

Если же смотреть лишь на Хрущева и Брежнева, то ни на Рис. 3а, ни на Рис. 3б между их графиками не выявлено никаких заметных различий, позволяющих отделить одного от другого. Таким образом, можно сказать, что на пунктуационном уровне текста авторский стиль в период позднего социализма практически не проявлялся.

Рис. 3: а) Средняя длина предложения; б) Число запятых на 1000 символов



6.1.4 Частотный анализ

Наконец, посмотрим на частоту употребления знаков препинания, стоп-слов и слов-штампов советской политической речи.

На Рис. 4а никаких особых различий в употреблении знаков препинания и часто употребляемых в речи слов (союзов, предлогов, местоимений) не обнаруживается - характеры графиков одинаковы. Есть лишь одно отклонение - слово «что» Лениным употребляется гораздо чаще, чем его последователями.

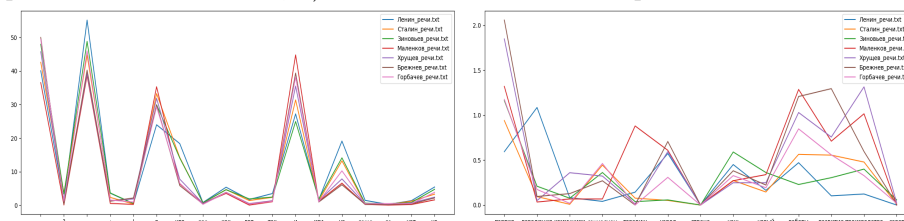
А вот на Рис. 4б различия в частоте употребления штамповых слов уже гораздо заметнее - так, Маленков чаще других употреблял слово «товарищ» (то есть, обращался непосредственно к аудитории или упоминал кого-то), Ленин и Зиновьев практически не говорили о «производстве», Хрущев больше остальных упоминал «коммунизм», а Брежнев акцентировал внимание слушателей на «развитии» страны.

Таким образом, в совокупности у авторов (Ленин, Сталин, Зиновьев, Маленков, Хрущев, Брежнев, Горбачев) присутствуют четкие различия в частоте использования слов-штампов и, порой, пунктуации. И даже если рассматривать только графики Хрущева и Брежнева - авторов советского дискурса в период позднего социализма - то можно найти штамповые лексемы, с точки зрения частоты употребления которых они различаются.

6.2 Близость текстов

Подсчет близости текстов - достаточно трудоемкая и в некотором смысле «спорная» задача, так как для неё не существует идеального метода решения. Дело в том, что термин «близость» для текстов можно определить по-разному - под ней можно понимать и семантическую (смысловую) близость,

Рис. 4: Графики частотности употребления распространенных а) знаков препинания и стоп-слов; б) слов-штампов советской речи



и жанровую близость (принадлежность одному жанру), и стилометрическую близость (принадлежность одному автору). С каждым из видов определения близости связано множество алгоритмов машинного обучения, способных решить соответствующую задачу подсчета сходства между двумя текстами.

В ходе нашего исследования для получения более полной и подробной картины было принято решение использовать несколько подходов: «явный» (где считается заданная метрика близости) и «неявный» (где тексты объединяются в группы/классы). В качестве «неявных» методов были взяты решения задач классификации и кластеризации текстов с помощью алгоритмов k-NN и k-Means соответственно. В качестве «явного» метода был взят подсчет косинусного расстояния между текстами.

6.2.1 Оценка близости текстов через задачу кластеризации

В ходе первого эксперимента тексты Отчетных докладов ЦК и их объединения по авторам были кластеризованы с помощью метода k-средних (k-means).

Результатом эксперимента является следующее распределение отчетов на кластеры: в кластер № 1 вошли доклады ЦК с VI по XVII съезд, в кластер № 2 - все остальные. Таким образом, алгоритм разделил отчеты на условно «раннесоциалистические» и «позднесоциалистические».

Разделение авторов полностью совпало с разделением Отчетных докладов ЦК - Ленин, Зиновьев и Сталин попали в кластер № 2, а Маленков, Хрущев, Брежнев и Горбачев - в кластер № 1.

6.2.2 Оценка близости текстов через задачу классификации

В предыдущем эксперименте было выявлено, что доклады периода позднего социализма отделимы от предшествовавших им - так, алгоритм объединил в одном кластере отчеты Маленкова, Хрущева, Брежнева и Горбачева, а в другом - речи Ленина, Зиновьева и Сталина. Однако сами по себе эти результаты могут сказать о малом - о влиянии времени, разве что. Гораздо интереснее понять, к кому из авторов-предшественников, к Ленину или к Сталину, были ближе их последователи. Если к первому, то можно будет говорить о том,

что это - результат возвращения к «заветам Ильича». Иначе же мы получим подтверждение закрепления «сталинской» версии советского политического языка его последователями.

В ходе второго эксперимента алгоритму k-NN, обучившемуся на текстах Ленина и Сталина, предлагалось определить остальные тексты в один из этих двух классов.

Результатом эксперимента стало то, что все тексты, кроме речей Зиновьева на XII и XIII съездах и отчета Горбачева на XXVIII съезде, были определены как «сталинские» - то есть, несмотря на большую разницу во времени, развенчание культа личности и возвращение к «заветам Ильича», настоящего поворота дискурса к прошлому не случилось. Наоборот, его «сталинская» версия просуществовала вплоть до 1986 года. Это явление мы назовем «сталинизацией» дискурса и подробнее поговорим о нем в разделе, посвященном выводам исследования.

6.2.3 Оценка близости текстов через подсчет косинусного расстояния

Определим лексико-семантическую близость текстов так: как известно, два слова называются семантически близкими, если они обладают схожим смыслом (например, «государство» и «страна», «шляпа» и «шапка», «ребенок» и «младенец»); тексты же будем считать **лексико-семантически близкими**, если они составлены из семантически близких слов (пусть и в другом порядке и с другим смыслом).

Такое определение позволяет нам говорить как о близости предложений «Наша страна заботится о детях» и «Наше государство опекает младенцев» (близость достигается за счет смысла), так и о близости предложений «Наша страна заботится о детях» и «Дети заботятся о нашей стране» (близость достигается за счет использования тех же слов).

За меру лексико-семантической близости текстов мы можем взять косинусное расстояние между векторами, являющимися суммой векторных представлений соответствующих слов («векторами текстов»), полученными с помощью векторной модели word2vec. Достоинствами этого подхода можно считать а) независимость косинусной меры от мер векторов, на которые оказывают влияние длины соответствующих текстов; и б) интерпретируемость результатов - если векторы предложений близки с точки зрения косинусной меры, то это могло быть вызвано только синонимичностью или совпадением слов в них.

Подобное определение близости и метода помогает нам сформулировать четкий критерий, согласно которому будет проверяться тезис Юрчака о ретрансляции советского авторитетного дискурса. Раз уж утверждается, что в языке периода позднего социализма проходили процессы самовоспроизводства и цитируемости, то тексты этого периода должны быть лексико-семантически близкими за счет использования одинаковой или синонимичной лексики.

Преимуществом этого подхода является то, что а) учитывается как лексическое, так и семантическое соответствие текстов; и б) из-за основополагающей идеи машинного обучения можно провести аналогию между восприятием речи машиной и слушающим её человеком. Есть, однако, и изъян у этого подхода - сами по себе цифры мало приближат нас к пониманию протекавших в советском дискурсе процессов, необходимо сравнение с другими странами. Поэтому было принято решение провести подсчет косинусной близости не только для текстов Отчетного доклада ЦК, но и для его «американского аналога» - зачитываемого Конгрессу обращения президента США «О положении страны».

Рис. 5: Косинусная близость, посчитанная для Отчетных докладов ЦК

	XXVIII Съезд	XXVII Съезд	XXVI Съезд	XXV Съезд	XXIV Съезд	XXIII Съезд	XXII Съезд	XX Съезд	XIX Съезд	XVIII Съезд	XVII Съезд	XVI Съезд	XV Съезд	XIV Съезд	XIII Съезд	XII Съезд	XI Съезд	X Съезд	IX Съезд	VIII Съезд	VII Съезд	VI Съезд
XXVIII Съезд	1.0	0.99	0.98	0.98	0.98	0.97	0.96	0.96	0.95	0.95	0.95	0.94	0.95	0.94	0.94	0.92	0.89	0.95	0.9	0.96	0.84	0.77
XXVII Съезд	0.99	1.0	0.99	0.99	0.99	0.98	0.97	0.97	0.97	0.96	0.96	0.95	0.96	0.94	0.92	0.9	0.87	0.94	0.88	0.94	0.82	0.72
XXVI Съезд	0.98	0.99	1.0	1.0	1.0	0.99	0.99	0.99	0.99	0.98	0.97	0.96	0.97	0.96	0.93	0.92	0.88	0.95	0.89	0.94	0.82	0.76
XXV Съезд	0.98	0.99	1.0	1.0	1.0	0.99	0.99	0.99	0.98	0.97	0.97	0.96	0.97	0.96	0.93	0.92	0.88	0.95	0.89	0.95	0.83	0.75
XXIV Съезд	0.98	0.99	1.0	1.0	1.0	0.99	0.99	0.99	0.99	0.98	0.98	0.97	0.97	0.96	0.93	0.92	0.87	0.94	0.88	0.94	0.81	0.76
XXIII Съезд	0.97	0.98	0.99	0.99	0.99	1.0	1.0	0.99	0.99	0.98	0.98	0.97	0.97	0.97	0.93	0.92	0.89	0.94	0.89	0.94	0.85	0.77
XXII Съезд	0.96	0.97	0.99	0.99	0.99	1.0	1.0	0.99	0.99	0.98	0.98	0.98	0.98	0.98	0.94	0.93	0.89	0.95	0.9	0.94	0.86	0.76
XX Съезд	0.96	0.97	0.99	0.99	0.99	0.99	0.99	1.0	1.0	0.98	0.99	0.98	0.98	0.98	0.92	0.91	0.87	0.94	0.87	0.93	0.81	0.74
XIX Съезд	0.95	0.97	0.99	0.98	0.99	0.99	0.99	1.0	1.0	0.99	0.99	0.98	0.98	0.98	0.92	0.91	0.87	0.94	0.87	0.92	0.82	0.74
XVIII Съезд	0.95	0.96	0.98	0.97	0.98	0.98	0.99	0.98	0.99	1.0	0.99	0.98	0.99	0.99	0.95	0.94	0.9	0.96	0.92	0.95	0.87	0.77
XVII Съезд	0.95	0.96	0.97	0.97	0.98	0.98	0.98	0.99	0.99	0.99	1.0	0.99	0.99	0.99	0.95	0.94	0.9	0.96	0.92	0.95	0.85	0.75
XVI Съезд	0.94	0.95	0.96	0.96	0.97	0.97	0.98	0.98	0.98	0.98	0.98	0.99	1.0	0.99	0.94	0.93	0.88	0.95	0.89	0.93	0.84	0.74
XV Съезд	0.95	0.96	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98	0.99	0.99	1.0	0.99	0.96	0.95	0.91	0.97	0.92	0.95	0.86	0.77
XIV Съезд	0.94	0.94	0.96	0.96	0.96	0.97	0.98	0.97	0.97	0.99	0.99	0.99	0.99	1.0	0.97	0.96	0.91	0.96	0.92	0.94	0.87	0.76
XIII Съезд	0.94	0.92	0.93	0.93	0.93	0.93	0.94	0.92	0.92	0.95	0.95	0.94	0.96	0.97	1.0	0.99	0.97	0.98	0.95	0.97	0.88	0.79
XII Съезд	0.92	0.9	0.92	0.92	0.92	0.92	0.93	0.91	0.91	0.94	0.94	0.93	0.95	0.96	0.99	1.0	0.97	0.97	0.96	0.97	0.89	0.8
XI Съезд	0.89	0.87	0.88	0.88	0.87	0.89	0.89	0.87	0.87	0.9	0.9	0.88	0.91	0.91	0.97	0.97	1.0	0.95	0.95	0.96	0.91	0.77
X Съезд	0.95	0.94	0.95	0.95	0.94	0.94	0.95	0.94	0.94	0.96	0.96	0.95	0.97	0.96	0.98	0.97	0.95	1.0	0.96	0.97	0.92	0.79
IX Съезд	0.9	0.88	0.89	0.89	0.88	0.89	0.9	0.87	0.87	0.92	0.9	0.89	0.92	0.92	0.95	0.96	0.95	0.96	1.0	0.95	0.95	0.83
VIII Съезд	0.96	0.94	0.94	0.95	0.94	0.94	0.94	0.93	0.92	0.95	0.94	0.93	0.95	0.94	0.97	0.97	0.96	0.97	0.95	1.0	0.9	0.83
VII Съезд	0.84	0.82	0.82	0.83	0.81	0.85	0.86	0.81	0.82	0.87	0.85	0.84	0.86	0.87	0.88	0.89	0.91	0.92	0.95	0.9	1.0	0.77
VI Съезд	0.77	0.72	0.76	0.75	0.76	0.77	0.76	0.74	0.74	0.77	0.75	0.74	0.77	0.76	0.79	0.8	0.77	0.79	0.83	0.83	0.77	1.0

В ходе анализа лексико-семантической близости текстов были получены результаты, представленные на Рис. 5 и 6. И пусть цифры мало о чем говорят сами по себе, но в совокупности и в сравнении они дают более глубокое понимание того, что происходило с советским политическим языком в этот период.

На соответствующей тепловой карте (Рис. 5) можно выделить три «прямоугольника близости»: один образуют речи Ленина и Зиновьева (VIII - XIII съезды), другой - речи Сталина, Маленкова, Хрущева и Брежнева (XIV-XXVI съезды), третий - речи Брежнева и Горбачева (XXIV-XXVIII съезды). Полученные прямоугольники слегка различаются по «насыщенности» (разбросу значений внутри них) - в то время как в первом преобладают значения 0.97 – 0.96, во втором и третьем же большая их часть лежит в диапазоне 0.99 – 0.98.

Также имеет место некоторая унификация распределения близостей Отчетных докладов вне прямоугольников - значения косинусных расстояний между отдельно взятым текстом Ленина и отчетами его последователей (не считая Зиновьева) примерно одинаковы. Так, например, доклады Ленина на

четных (VIII и X) съездах близки всем отчетам Сталина, Хрущева и Брежнева - значения косинусного расстояния лежат в диапазонах 0.92–0.96 и 0.94–0.97 соответственно, - а вот его же доклады на нечетных съездах (VII, IX и XI) - далеки, косинусное расстояние не превосходит 0.91.

Поэтому можно сказать, что влияние времени практически не проявляется - разве что в сильном удалении от доклада на VI съезде всех остальных текстов, что логично, так как он был зачитан еще до Октябрьской революции. Общее отхождение от речей Ленина нельзя причислить к влиянию времени, так как а) отдельные отчеты, написанные им, последователям достаточно близки; и б) цифры расстояний устойчивы к разнице во времени - от отчета Ленина на XI съезде доклад Горбачева на XXVIII съезде (разница - 68 лет) удален даже меньше, чем речь Сталина на XVI съезде (разница - 8 лет).

Рис. 6: Косинусная близость, посчитанная для обращений «О положении страны»

	Bush 1990	Reagan 1986	Reagan 1981	Ford 1976	Nixon 1971	Johnson 1966	Eisenhower 1966	Truman 1956	Roosevelt 1952	Roosevelt 1939	Hoover 1930	Coolidge 1927	Coolidge 1925	Coolidge 1924	Coolidge 1923	Harding 1922	Harding 1921	Wilson 1920	Wilson 1919	Wilson 1918	Wilson 1917	
Bush 1990	1.0	0.98	0.89	0.96	0.97	0.97	0.94	0.91	0.97	0.95	0.93	0.85	0.9	0.9	0.9	0.9	0.91	0.91	0.91	0.92	0.93	0.93
Reagan 1986	0.98	1.0	0.9	0.97	0.97	0.98	0.95	0.93	0.97	0.96	0.94	0.87	0.91	0.92	0.92	0.92	0.93	0.93	0.93	0.93	0.94	0.94
Reagan 1981	0.89	0.9	1.0	0.97	0.93	0.93	0.95	0.95	0.92	0.93	0.94	0.95	0.95	0.94	0.94	0.93	0.93	0.93	0.93	0.92	0.9	0.88
Ford 1976	0.96	0.97	0.97	1.0	0.98	0.98	0.98	0.97	0.97	0.98	0.97	0.94	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.95	0.92
Nixon 1971	0.97	0.97	0.93	0.98	1.0	0.98	0.96	0.95	0.97	0.96	0.96	0.9	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.92
Johnson 1966	0.97	0.98	0.93	0.98	0.98	1.0	0.98	0.96	0.98	0.98	0.97	0.91	0.95	0.95	0.95	0.95	0.96	0.95	0.95	0.96	0.96	0.95
Kennedy 1961	0.94	0.95	0.95	0.98	0.96	0.98	1.0	0.98	0.97	0.98	0.98	0.95	0.96	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.93
Eisenhower 1956	0.91	0.93	0.95	0.97	0.95	0.96	0.98	1.0	0.95	0.97	0.98	0.97	0.98	0.98	0.98	0.98	0.97	0.97	0.96	0.97	0.95	0.9
Truman 1952	0.97	0.97	0.92	0.97	0.97	0.98	0.97	0.95	1.0	0.98	0.96	0.9	0.94	0.94	0.95	0.94	0.95	0.95	0.95	0.96	0.97	0.95
Roosevelt 1939	0.95	0.96	0.93	0.98	0.96	0.98	0.98	0.97	0.98	1.0	0.98	0.93	0.95	0.97	0.97	0.97	0.97	0.97	0.97	0.98	0.97	0.96
Roosevelt 1934	0.93	0.94	0.94	0.97	0.96	0.97	0.98	0.98	0.96	0.98	1.0	0.96	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.97	0.98	0.93
Hoover 1930	0.85	0.87	0.95	0.94	0.9	0.91	0.95	0.97	0.9	0.93	0.96	1.0	0.98	0.97	0.97	0.96	0.96	0.96	0.95	0.95	0.92	0.85
Coolidge 1927	0.9	0.91	0.95	0.96	0.94	0.95	0.96	0.98	0.94	0.95	0.98	0.98	1.0	0.99	0.99	0.99	0.98	0.98	0.97	0.97	0.96	0.89
Coolidge 1925	0.9	0.92	0.94	0.96	0.94	0.95	0.97	0.98	0.94	0.97	0.98	0.97	0.99	1.0	0.99	0.99	0.99	0.99	0.98	0.98	0.96	0.93
Coolidge 1924	0.9	0.92	0.94	0.96	0.94	0.95	0.97	0.98	0.95	0.97	0.99	0.97	0.99	0.99	1.0	0.99	0.99	0.99	0.98	0.99	0.96	0.92
Coolidge 1923	0.9	0.92	0.93	0.96	0.94	0.95	0.97	0.98	0.94	0.97	0.98	0.96	0.99	0.99	0.99	1.0	0.99	0.99	0.98	0.98	0.96	0.92
Harding 1922	0.91	0.93	0.93	0.96	0.94	0.96	0.97	0.97	0.95	0.97	0.98	0.96	0.98	0.99	0.99	0.99	1.0	0.99	0.97	0.99	0.97	0.94
Harding 1921	0.91	0.93	0.93	0.96	0.94	0.95	0.97	0.97	0.95	0.97	0.98	0.96	0.98	0.99	0.99	0.99	0.99	1.0	0.98	0.99	0.97	0.94
Wilson 1920	0.91	0.93	0.93	0.96	0.94	0.95	0.97	0.96	0.95	0.97	0.97	0.95	0.97	0.98	0.98	0.98	0.97	0.98	1.0	0.98	0.97	0.94
Wilson 1919	0.92	0.93	0.92	0.96	0.94	0.96	0.97	0.97	0.96	0.98	0.98	0.95	0.97	0.98	0.99	0.98	0.99	0.99	0.98	1.0	0.97	0.95
Wilson 1918	0.93	0.94	0.9	0.95	0.94	0.96	0.97	0.95	0.97	0.97	0.97	0.92	0.96	0.96	0.96	0.96	0.97	0.97	0.97	0.97	1.0	0.96
Wilson 1917	0.93	0.94	0.88	0.92	0.92	0.95	0.93	0.9	0.95	0.96	0.93	0.85	0.89	0.93	0.92	0.92	0.94	0.94	0.94	0.95	0.96	1.0

В Америке же картина разительно другая - несмотря на наличие двух прямоугольников близости («Вильсон - Гардинг - Кулидж» и «Рузвельт - Труман - Эйзенхауэр - Кеннеди - Джонсон - Никсон - Форд»), их «насыщенность» гораздо ниже - среднее значение внутри первого равно 0.98 (и вызвано это, скорее всего, тем, что четыре речи Кулиджа близки друг другу), а внутри второго - 0.95 – 0.96. Влияние времени проявляется не только в удаленности от обращений предшественников речей Рейгана и Буша-старшего (и от обращений последователей речи Вильсона в 1917 году), но и в том, что удаление происходит постепенно - так, например, значения косинусной близости речи Кулиджа в 1925 году поначалу достигают 0.98 (Рузвельт в 1934 году, Эйзенхауэр в 1956 году), но потом постепенно снижаются до 0.9 (0.95 - у Джонсона, 0.94 - у Никсона, 0.92 - у Рейгана в 1986 году). Также менее заметно себя проявляет, хотя и присутствует, унификация распределения расстояний вне прямоугольников - отдельно взятая

речь Гувера в 1930 году относительно близка речам Эйзенхауэра, Кеннеди и Форда (значения 0.94–0.97), но далека от речей Никсона, Джонсона и Трумана (значения 0.9 – 0.91).

Таким образом, насыщенность тепловой карты Отчетных докладов ЦК и результаты её сравнения с тепловой картой для обращения «О положении страны» подтверждают выводы Юрчака о ретрансляции дискурса - советский авторитетный дискурс действительно проявляет себя в рамках эксперимента как менее разнообразный и более «унифицированный» и ритуализованный. Если рассматривать речи, зачитанные в 1950-1980-х годах, то с точки зрения лексико-семантической текстовой близости они практически не отличимы друг от друга - косинусные расстояния между ними колеблются в диапазоне 0.98 – 1.0, что соответствует полной смысловой идентичности текстов. Отдельно стоит отметить, что обнаруженный в ходе исследования процесс унификации начался не сразу, а лишь в середине 1920-х годов, с момента прихода Сталина к власти, а потому этот эффект невозможно приписать общей специфике коммунистического учения.

Подобная картина может быть вызвана как минимум двумя сторонними факторами: а) тем, что несколько речей подряд читали одни и те же люди - так, например, косинусные значения близости речей Брежнева на XXIV-XXVI съездах равны 1 (то есть, соответствующие векторы направлены абсолютно одинаково); и б) тем, что все авторы принадлежали одной партии, диктующей свои взгляды на основные экономические и политические вопросы. Поэтому, чтобы снизить влияние эффектов несменяемости власти в течение долгого времени и однопартийного строя в СССР, были проведены дополнительные эксперименты.

В первом определялась близость непосредственно авторов - его результаты представлены на Рис. 7. Как можно увидеть, тепловая карта из Рис. 5 во многом повторяется на Рис. 7а - вновь выделяются те же самые «прямоугольники близости», распределение вне их повторяется. Из незамеченного ранее - больше подчеркнута уникальность Горбачева, проявляющаяся в удалении его докладов от докладов Хрущева и Сталина.

Во втором эксперименте, результаты которого представлены на Рис. 7б и 7с, рассматривалось распределение близостей обращений «О положении страны» лишь для членов одной партии - республиканцев и демократов.

На Рис. 7б, посвященном республиканцам, можно выявить прямоугольник близости «Гардинг - Кулидж - Гувер - Эйзенхауэр», значения в котором колеблются от 0.99 до 0.96, и вложенный в него более насыщенный прямоугольник «Гардинг - Кулидж» (в нем значения близости не падают ниже 0.98). Интересно, что эти прямоугольники по-разному себя проявляют в отношении близости с последователями - в то время как близости Кулиджа и Гардинга с президентами после них повторяют друг друга, распределение косинусных расстояний Гувера и Эйзенхауэра уже никак нельзя назвать одинаковыми. В остальном же наблюдаются уже отмеченные паттерны - влияние времени проявляется в плавном удалении последователей от предшественников, замечен

большой разброс у авторов второй половины XX века.

На Рис. 7с, посвященном демократам, нет ни одного значения косинусной близости выше 0.98 (для разных авторов). Однако, несмотря на это, можно выделить два прямоугольника близости - «Джонсон - Кеннеди - Труман - Рузвельт» и «Рузвельт - Вильсон», причем в обоих прямоугольниках значения лежат на промежутке 0.97 – 0.98. Влияние времени проявляется лишь у Джонсона, значения близости с которым уменьшаются слева направо, у остальных авторов подобного удаления не наблюдается - так, например, речи Вильсона в 1918 году и обращение Рузвельта в 1939 году близки большинству авторов-демократов.

Наконец, чтобы оценить семантическую составляющую полученных значений близости, для всех текстов и для всех авторов были построены пересечения «описывающих» их доклады слов (слов, чьи векторы находятся наиболее близко к вектору всего текста). Для XIX съезда «описывающие» слова выглядят так: *государство, организация, развитие, система, условие, экономический, страна, деятельность, производство, предприятие*. Для XXVI съезда это следующий набор: *организация, развитие, государство, деятельность, условие, система, экономический, страна, частность, политический*. А для XXVII съезда - такой: *развитие, система, государство, деятельность, организация, условие, экономический, принцип, отношение, достижение*. Если сравнивать их со словами, полученными через облако слов, то можно увидеть, что наборы частично пересекаются, но не совпадают - то есть, это не наборы самых частых в употреблении слов, а все же некоторые смысловые приближения текстов.

Результаты подсчета их пересечений для всех Отчетных докладов представлены на Рис. 8а. Можно увидеть, что большинство наборов пересекаются в словах под номерами 1-8. В первый раз этот набор появился у Сталина на XVII съезде, у Ленина и Зиновьева же слова из него практически не встречались. Также стоит отметить постепенное снижение разнообразия наборов от доклада к докладу у одного автора - в то время как у Ленина описывающие слова VII и VIII съездов пересекаются лишь дважды, у Брежнева пересечений между докладами на XXIV и XXV съездах уже 9.

На Рис. 8б точно такое же пересечение было построено для авторов и оно подтверждает результаты, полученные при подсчете для них косинусной близости. Вновь продемонстрировано, что Ленин и Зиновьев имеют с последователями достаточно небольшое количество пересечений, в то время как Сталин проявляет себя как некий «трендсеттер», направивший дискурс по курсу, который впоследствии был поддержан остальными.

Таким образом, с точки зрения лексико-семантической близости текстов, в советском авторитетном дискурсе были замечены следующие паттерны:

- Так как и с точки зрения алгоритма k-NN, и с точки зрения косинусного расстояния речи Сталина, Маленкова, Хрущева, Брежнева и Горбачева близки друг другу, а их близость речам Ленина и Зиновьева распреде-

лена практически одинаково, можно говорить о унификации советского дискурса.

- Ощущение монолитности и незыблемости дискурса, описанное многими свидетелями того времени, находит подтверждение в цифрах - алгоритм считает векторы речей практически идентичными (а сами речи определяются в один класс/кластер)
- Влияние времени проявляется неравномерно - например, отдаление между отчетами на XXVII и XIV съездах (разница в 61 год) меньше, чем между отчетами на XIV и на VI съездах (разница в 7 лет).
- Имеет место такое явление как «сталинизация» дискурса - остановка его развития где-то в 1930-1950х годах, выражающаяся в близости к текстам Сталина текстов его последователей и общей схожести.
- «Точкой слома» советского авторитетного дискурса на примере Отчетных докладов ЦК можно назвать XIV съезд, первый послереволюционный доклад Сталина.
- Возвращения к «заветам Ильича» после смерти Сталина так и не случилось - с точки зрения косинусной близости все остальные авторы отделились от него даже больше.

7 Выводы

Подведем итоги исследования. О чем говорят полученные результаты? Во-первых, ими было подтверждено явление ретрансляции и самовоспроизводства дискурса - появление и поддержание абстрактной нормы советского официального языка, выражающейся не только и не столько в лексиконе, сколько в стиле письма. Во-вторых, они показывают, какие уровни текста были подчинены нормализации, а на каких еще оставался простор для проявления «авторского голоса». В третьих, они объясняют, за счет чего у советских граждан могло возникать ощущение монолитности и неизменности дискурса в период позднего социализма. И наконец, полученные результаты говорят о том, что выявленная лексическая, синтаксическая и стилометрическая схожесть Хрущева и Брежнева - скорее аномалия, а не регулярно наблюдаемый паттерн, так как в совокупности авторы Отчетных докладов ЦК все же различимы между собой: с точки зрения лексического разнообразия, средней длины предложения или употребления слов «партия» и «производство».

Таким образом, можно считать, что проведенное исследование не только подтвердило гипотезу Юрчака о ретрансляции и самовоспроизводстве советского авторитетного дискурса в период позднего социализма, но и предоставило численные доказательства существования этих явлений.

Список литературы

- [1] Zemtsov I., *Soviet political language*, London : Overseas publ. interchange, 1985.
- [2] Kotkin S., *Magnetic Mountain: Stalinism as a Civilization*, University of California Press, 1995.
- [3] Кожаева А.А., Речевые практики подражания официальному советскому дискурсу в письмах «лишенцев» во власть (1918–1936 годы) (на материалах Западной Сибири), *Исторический курьер*, (6):29–38, 2022, <https://doi.org/10.31518/2618-9100-2022-6-2>, available from: <http://www.jstor.org/stable/2030894>.
- [4] Юрчак А.В., *Это было навсегда, пока не кончилось. Последнее советское поколение*, М.: Новое Литературное Обозрение, 2014.
- [5] Эйхенбаум Б.М., Основные стилевые тенденции в речи Ленина, *ЛЕФ*, (1), 1924.
- [6] Шкловский В.Б., Ленин как деканонизатор, *ЛЕФ*, (1), 1924.
- [7] Карцевский С. О., *Язык, война и революция*, Берлин : Рус. универс. изд-во, 1923.
- [8] Савина Т.В., Говорить по-большевистски? Язык партийца в литературе «оттепели», *Исторический курьер*, (6):164–179, 2022, <https://doi.org/10.31518/2618-9100-2022-6-13>, available from: <http://istkurier.ru/data/2022/ISTKURIER-2022-6-13.pdf>.
- [9] Тынянов Ю.Н., Словарь Ленина-полемиста, *ЛЕФ*, (1), 1924.

Рис. 7: Косинусная близость, посчитанная для авторов: а) в СССР б) в США (Республиканская партия) с) в США (Демократическая партия)

	Горбачев	Брежнев	Хрущев	Маленков	Сталин	Зиновьев	Ленин
Горбачев	1.0	0.99	0.97	0.97	0.96	0.92	0.92
Брежнев	0.99	1.0	0.99	0.99	0.98	0.93	0.92
Хрущев	0.97	0.99	1.0	0.99	0.99	0.93	0.92
Маленков	0.97	0.99	0.99	1.0	0.99	0.92	0.9
Сталин	0.96	0.98	0.99	0.99	1.0	0.96	0.94
Зиновьев	0.92	0.93	0.93	0.92	0.96	1.0	0.98
Ленин	0.92	0.92	0.92	0.9	0.94	0.98	1.0

	Bush 1990	Reigan 1986	Reigan 1981	Ford 1976	Nixon 1971	Eisenhower 1956	Hoover 1930	Coolidge 1927	Coolidge 1925	Coolidge 1924	Coolidge 1923	Harding 1922	Harding 1921
Bush 1990	1.0	0.98	0.89	0.96	0.97	0.92	0.85	0.9	0.9	0.91	0.9	0.91	0.91
Reigan 1986	0.98	1.0	0.9	0.97	0.97	0.93	0.87	0.91	0.92	0.92	0.92	0.93	0.93
Reigan 1981	0.89	0.9	1.0	0.96	0.93	0.95	0.95	0.95	0.94	0.94	0.94	0.93	0.93
Ford 1976	0.96	0.97	0.96	1.0	0.98	0.97	0.94	0.96	0.96	0.96	0.96	0.96	0.96
Nixon 1971	0.97	0.97	0.93	0.98	1.0	0.95	0.9	0.94	0.94	0.94	0.94	0.94	0.94
Eisenhower 1956	0.92	0.93	0.95	0.97	0.95	1.0	0.97	0.98	0.98	0.98	0.98	0.97	0.97
Hoover 1930	0.85	0.87	0.95	0.94	0.9	0.97	1.0	0.98	0.97	0.97	0.96	0.96	0.96
Coolidge 1927	0.9	0.91	0.95	0.96	0.94	0.98	0.98	1.0	0.99	0.99	0.99	0.98	0.98
Coolidge 1925	0.9	0.92	0.94	0.96	0.94	0.98	0.97	0.99	1.0	0.99	0.99	0.99	0.99
Coolidge 1924	0.91	0.92	0.94	0.96	0.94	0.98	0.97	0.99	0.99	1.0	0.99	0.99	0.99
Coolidge 1923	0.9	0.92	0.94	0.96	0.94	0.98	0.96	0.99	0.99	0.99	1.0	0.99	0.99
Harding 1922	0.91	0.93	0.93	0.96	0.94	0.97	0.96	0.98	0.99	0.99	0.99	1.0	0.99
Harding 1921	0.91	0.93	0.93	0.96	0.94	0.97	0.96	0.98	0.99	0.99	0.99	0.99	1.0

	Johnson 1966	Kennedy 1961	Truman 1952	Roosevelt 1939	Roosevelt 1934	Wilson 1920	Wilson 1919	Wilson 1918	Wilson 1917
Johnson 1966	1.0	0.98	0.98	0.98	0.97	0.95	0.96	0.96	0.95
Kennedy 1961	0.98	1.0	0.97	0.98	0.98	0.97	0.97	0.97	0.93
Truman 1952	0.98	0.97	1.0	0.98	0.96	0.95	0.96	0.97	0.95
Roosevelt 1939	0.98	0.98	0.98	1.0	0.98	0.97	0.98	0.97	0.96
Roosevelt 1934	0.97	0.98	0.96	0.98	1.0	0.97	0.98	0.97	0.93
Wilson 1920	0.95	0.97	0.95	0.97	0.97	1.0	0.98	0.97	0.94
Wilson 1919	0.96	0.97	0.96	0.98	0.98	0.98	1.0	0.97	0.95
Wilson 1918	0.96	0.97	0.97	0.97	0.97	0.97	0.97	1.0	0.96
Wilson 1917	0.95	0.93	0.95	0.96	0.93	0.94	0.95	0.96	1.0

Рис. 8: Пересечение «описывающих» слов для а) съездов; б) авторов

