**Санкт-Петербургский филиал федерального государственного автономного образовательного учреждения высшего образования   
"Национальный исследовательский университет   
"Высшая школа экономики"**

**Факультет: Санкт-Петербургская школа гуманитарных наук и искусств**

**Департамент филологии**

**КУРСОВАЯ РАБОТА**

**на тему «Через модернизм Бергмана к Тарковскому и Триеру: тематическое моделирование субтитров»**

**«** **Through the Modernism of Bergman to Tarkovsky and Trier: Topic Modeling of Subtitles»**

**Регистрационный номер \_\_\_\_\_\_\_\_\_**

**Процент авторского текста \_\_\_\_\_\_\_\_\_\_**

**Оценка научного руководителя \_\_\_\_\_\_\_\_\_\_**

**Выполнила:**

Студентка 3 курса

группы БФЛ 192

Марголина Анастасия Валерьевна

**Научный руководитель:**

Старший преподаватель,   
Климов А.А.

Санкт-Петербург

2022

ОГЛАВЛЕНИЕ

[ВВЕДЕНИЕ 3](#_Toc105272280)

[Глава 1. Датасет «Риторика фильмов Бергмана, Тарковского и Триера» 5](#_Toc105272281)

[1.1 Выбор материала 5](#_Toc105272282)

[1.2 Подготовка корпуса 7](#_Toc105272283)

[1.3 Анализ корпуса с помощью методов Python 9](#_Toc105272284)

[1.4 Средняя длина предложения 9](#_Toc105272285)

[1.5 Облако слов 11](#_Toc105272286)

[Глава 2. Тематическое моделирование «Риторики фильмов Бергмана, Тарковского и Триера» 16](#_Toc105272287)

[2.1. Выбор модели 16](#_Toc105272288)

[2.2. Тематическое моделирование для фильмов Ингмара Бергмана 18](#_Toc105272289)

[2.3. Тематическое моделирование для фильмов Андрея Тарковского 21](#_Toc105272290)

[2.4. Тематическое моделирование для фильмов Ларса фон Триера 24](#_Toc105272291)

[2.5. Тематическое моделирование всего корпуса «Риторика фильмов Бергмана, Тарковского и Триера» 27](#_Toc105272292)

ЗАКЛЮЧЕНИЕ [31](#_Toc105272293)

[СПИСОК ИСТОЧНИКОВ 33](#_Toc105272294)

## ВВЕДЕНИЕ

Тематическое моделирование на сегодняшний день – популярный метод анализа данных в сфере естественной обработки языка. В основном его применяют к большим, «социальным» корпусам: комментарии, рецензии, отзывы. Тем не менее, существуют и литературные исследования на основе алгоритмов тематического моделирования [Navarro-Colorado 2018; Uglanova 2020; Rha 2021]. В них авторы используют модель LDA, которая была представлена еще в 2003 году [Blei, Ng, Jordan 2003]. Этот алгоритм имеет определенные недостатки, особенно при применении к небольшому объему данных: он не учитывает контекст и нуждается в обширной предобработке текста. Из-за этого теряется доля результата, что не особо важно при изучении больших данных, но что значительно влияет на качество выводов для маленького датасета.

Это исследование базируется на созданном мною корпусе «Риторика фильмов Бергмана, Тарковского и Триера», который состоит из субтитров ключевых картин указанных режиссеров. Поскольку эти данные представляют собой короткие реплики, зачастую это одна фраза, восклицание или маленькое предложение из 5 слов, для моих целей я использую современную тематическую модель BERTopic. Преимущества BERTopic для работы с моим корпусом субтитров – отсутствие обязательной предобработки текста и учет контекста. Кроме того, из-за новизны этой модели существует мало академических работ с ее применением, в особенности к художественным данным.

В этой курсовой работе я совмещаю цифровые методы анализа с художественной интерпретацией фильмов. Три изучаемых режиссера представляют собой три разных стиля, три разные эпохи, три разные страны. С помощью тематического моделирования я нахожу латентные, скрытые темы из фильмографии каждого автора по отдельности, а также в их совокупности. Несмотря на то, что Бергмана, Тарковского и Триера объединяет личные взаимодействия, стилистические влияния, эти режиссеры снимали очень разные фильмы. Бергман, как модернист, зачастую соблюдал единство времени и места, снимая камерное кино; Тарковский, как промежуточный этап между направлениями, снимал медленное, при этом напряженное в действиях кино; Триер, как постмодернист, экспериментировал в каждой своей работе, не останавливаясь на каком-либо одном стиле. Такое разнообразие дает благоприятную базу для тематического моделирования: в разности лейтмотивов оно способно найти общность.

Первая глава этой курсовой работы – разведочный анализ с помощью языка Python: подсчет статистик и исследование их в контексте стиля автора/авторов. Эта часть значима, как по результатам и выводам, так и в качестве сравнения более простого метода изучения текстовых данных с более сложным (т.е. с тематическим моделированием). Во второй главе я представляю результаты модели BERTopic по корпусу «Риторика фильмов Бергмана, Тарковского и Триера». Важная часть этой главы – примеры визуализаций, полученных тем, а также применение разных интерактивных инструментов для более глубокого изучения собранных данных.

# Глава 1. Датасет «Риторика фильмов Бергмана, Тарковского и Триера»

# Выбор материала

Тематическое моделирование обычно относят к методам работы с большими данными, в основном с социологическими: отзывы, комментарии, твиты. Мое же исследование отличается от типичного представления topic modeling, так как материал этой курсовой работы – реплики персонажей фильмов, субтитры.

Я остановилась на фильмах трех режиссеров - Ингмара Бергмана, Андрея Тарковского и Ларса фон Триера. Метод тематического моделирования предсказывает «скрытые» связи между любыми текстами, а результат представляется в виде определенного количества топиков, расставленных по частоте их совместного употребления. Выбор этих трех авторов объясняется с одной стороны их взаимосвязью и отчасти схожестью, с другой - разницей в стиле, темах и даже направлениях их кино.

Примечательно, что авторов объединяют почти одинаковые истории. Тарковский вдохновлялся Бергманом. Фильм «Жертвоприношение» можно считать прямой отсылкой к Бергману: он снят в Готланде, где долгое время проживал шведский режиссер. При жизни Бергман смотрел некоторые работы Тарковского и даже называл его тем, «кто изобрел новый киноязык <…>, который показывает жизнь, как отражение, жизнь как сон» [Schlegel 2012, 24]. Триер хоть и оглядывался на своего скандинавского коллегу, самым важным режиссером для себя всегда отмечал Тарковского [Долин 2019, 21]. Самый скандальный фильм Триера «Антихрист» посвящен русскому режиссеру – это указано в финальный титрах. Первая картина Триера «Элемент преступления» был показан Тарковскому, но «тому она очень не понравилась» [Долин 2019, 21].

Стоит более эксплицитно обозначить мотивацию к исследованию датасета «Риторика фильмов Бергмана, Тарковского и Триера». Как было сказано ранее, ценность тематического моделирования состоит буквально «в сочетании несочетаемого». Кинематограф, как ниша искусства, имеет свою специфику. Авторское кино зачастую – очень *разное.* Так, у Триера можно найти как эротику на большом экране, так и фильм-катастрофу, у Тарковского как фильм о средневековом религиозном художнике, так и научно-фантастическую драму, у Бергмана как фильм-воспитание (по аналогии с романом-воспитание), так и камерный фильм о двух влюбленных. Это отличает киноиндустрию от прозы – огромное разнообразие тем и сеттингов у одного автора, из-за чего использование тематического моделирования по отношению к корпусу может дать продуктивный результат. Если лейтмотивы, например, всей библиографии Достоевского легко понять в силу единства времени и места большинства его произведений, то фильмы изучаемых авторов не так унитарны.

Кроме того, важно учитывать, что эти три автора относятся к разным временам, течениям и культурам. Бергман — это строгий, шведский модернист, еще придерживающийся определенных правил киноиндустрии. Ларс фон Триер – строгий, датский постмодернист, ломающий любые правила. Тарковского же стоит рассматривать, как связующее звено между этими двумя направлениями и авторами. Такой цифровой метод, как тематическое моделирование представляется удачным и удобным для изучения взаимосвязи этих режиссеров.

Для корпуса были отобраны не все фильмы режиссеров, чтобы исключить перевес в сторону одно из авторов. Меньше всего картин было снято Тарковским – их всего семь, поэтому они все вошли в датасет. Из фильмографий Бергмана и Триера были взяты лишь части. Ключевые параметры для входа в корпус были: *авторство*, то есть только те картины, где в роли режиссера, сценариста и продюсера был один человек; репрезентативность, то есть только ключевые работы и неповторяющийся сеттинг.

Также надо ответить на ключевой вопрос: можно ли приравнивать реплики персонажей, субтитры к целому фильму как концепту? Безусловно, кинематограф можно представить не только в текстовых данных, но и в визуальных. Кино – это не только монолог и диалог, но это и закадровая музыка, операторская работа, выражение лиц актеров. Тем не менее, реплики остаются главным наполнением современных фильмов и их удобно использовать для изучения с помощью цифровых методов. Вопрос, поставленный в начале этого абзаца, становится одним из главных для этого исследования; а в более широком контексте меня интересует - можно ли вообще изучать искусство дедуктивно?

## Подготовка корпуса

По итогу отбора в корпус входят 24 фильма. Распределение количества по режиссерам примерно равное: все 7 существующих картин Тарковского, 8 -Бергмана («Осенняя соната», «Персона», «Шепоты и крики», «Фанни и Александр», «Сцены из супружеской жизни», «Седьмая печать», «Причастие», «Земляничная поляна») и 9 – Триера («Идиоты», «Рассекая волны», «Танцующая в темноте», «Догвилль», «Самый главный босс», «Меланхолия», «Антихрист», «Нимфоманка», «Дом, который построил Джек»).

Стоит отметить, что все фильмы представлены на английском языке, поскольку он предстает как самый нейтральный для исследования. Только один из режиссеров, Бергман, снимал только на своем родном языке – шведском. У Тарковского помимо русского встречается работа также и на шведском, а картины Триера были сняты, как на датском, так и на английском.

Так как данные корпуса – субтитры, и они имеют особое разрешение srt, файлы подвергаются обработке и чистке перед созданием датафрейма. Помимо самих реплик персонажей в тексте присутствуют отметки времени, на которых какое-либо высказывание появляется на экране. С помощью специальной библиотеки pysubparser для Python, созданной для работы с субтитрами, убираются лишние символы и, в некоторых случаях, html-разметка, кодирующая шрифт. Так, после обработки файлов получается датафрейм, состоящий из трех атрибутов – имя автора, название фильма и сам текст.

Важный шаг, кроме того, обработать сам текст для дальнейшей работы с корпусом и для построения визуализаций – убрать пунктуацию и привести весь текст к нижнему регистру. Без этих действий результаты будут неточными, а данные будут менее пригодными для любого анализа с помощью Python. Таким образом, текст представляет собой длинную последовательность из слов, без каких-либо синтаксических признаков.

В следующей части главы представлено техническое описание полученного корпуса, а также базовый анализ с помощью Python: объем данных, наибольший и наименьший тексты; самые частые слова и коллокации, визуализации корпуса. Для этого было бы недостаточно первичной обработки, описанной ранее. Более расширенная подготовка текста включает в себя исключение стоп-слов и токенизация, что осуществляется с помощью библиотеки NLTK. Токенизация текста разделяет строки на словоподобные единицы, чтобы Python мог объективно оценивать каждое слово как отдельный компонент, а также обеспечивать структуру ранее беспорядочного текста. Стоп-слова же включают в себе предлоги, частицы, местоимения, которые в целом не важны для общей семантики, но они заметно «нагружают» текст и статистику по тексту.

Третий этап подготовки корпуса – лемматизация. Этот процесс заметно улучшает любое корпусное исследования, поскольку приводит слова к изначальным формам, то есть леммам. Существует также метод стемминга, который представляет собой сокращение слова до его стеммы, основы, тем не менее тематическое моделирование дает лучший результат именно в совокупности с лемматизацией [May, Van Durme, Cotterel 2019].

Лемматизацию можно проводить с помощью многих библиотек: TextBlob, spacy, genism, NLTK и так далее. Для исследования был выбран самый оптимальный вариант - лемматизация в совокупности с POS-tagging на NLTK. Поскольку такой метод сначала определяет часть речи исходя из контекста в предложении и только потом приводит слово к лемме, это помогает избежать омонимии и, следовательно, неправильного результата.

# Анализ корпуса с помощью методов Python

Полученный после очистки корпус уже можно статистически исследовать методами Python. Хоть исследование посвящено тематическому моделированию, такой базовый анализ может быть полезен для некоторых выводов и сравнений с результатами, которые предоставляет BERT.

Статистическое исследование корпуса состоит из двух пунктов: вычисление средней длины предложения и построение облака слов.

## Средняя длина предложения

Для начала я решила исследовать среднюю длину предложения по корпусу. В английском языке она составляет от 15 до 20 слов [The Acropolitan 2017]. Основа корпуса – реплики персонажей, в основном диалоги. Важно смотреть на параметр средней длины не только по всему корпусу, но и по каждому автору отдельно, так как с помощью этого можно обнаружить стилистические особенности: чьи персонажи более монологичны, а чьи более молчаливы и кратки. При исследовании длины предложения я смотрела две переменные для более объективного результата: медиану и среднее значение. В корпусе, из-за особенности данных, наблюдаются выбросы, которые значительно могут исказить среднее значение.

Для нахождения медианы я использовала диаграмму размаха, или box plot. На рисунке 1 показан результат для каждого автора после удаления значительной части выбросов.

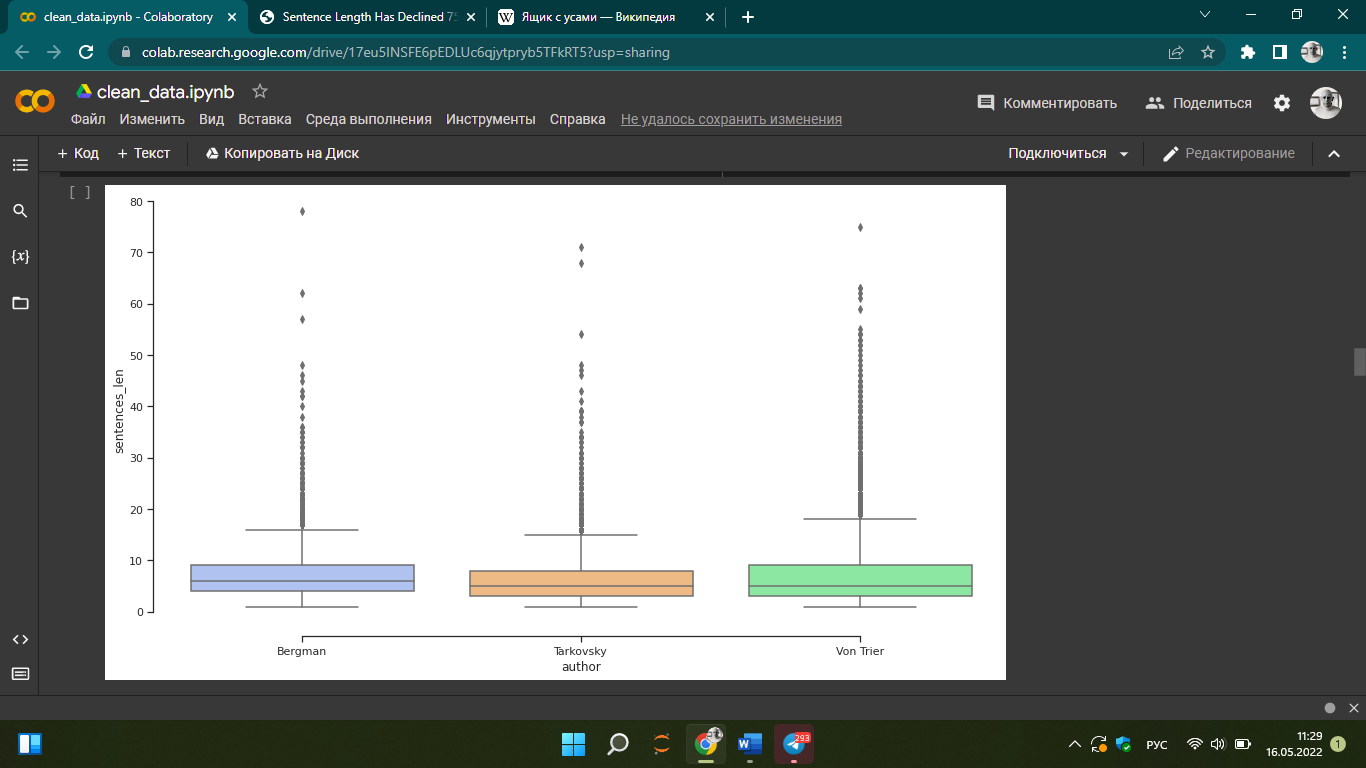


Рисунок 1

Как видно из результата, медиана по работам Бергмана находится выше всех. Это значение, в отличии от среднего, можно интерпретировать так: у шведского режиссера больше всего процент длинных предложений. Кроме того, можно заметить, что реплики из фильмов Тарковского немного уступают фильмам Триера.

Медиана по всему корпусу находится примерно на уровне 4-5 слов (рис.2). Этот результат неудивителен, даже в сравнении со средним значением по всему английскому языку (см. выше). Данные корпуса – устные высказывания, в основном в контексте разговора, что предполагает короткие реплики, ответы.

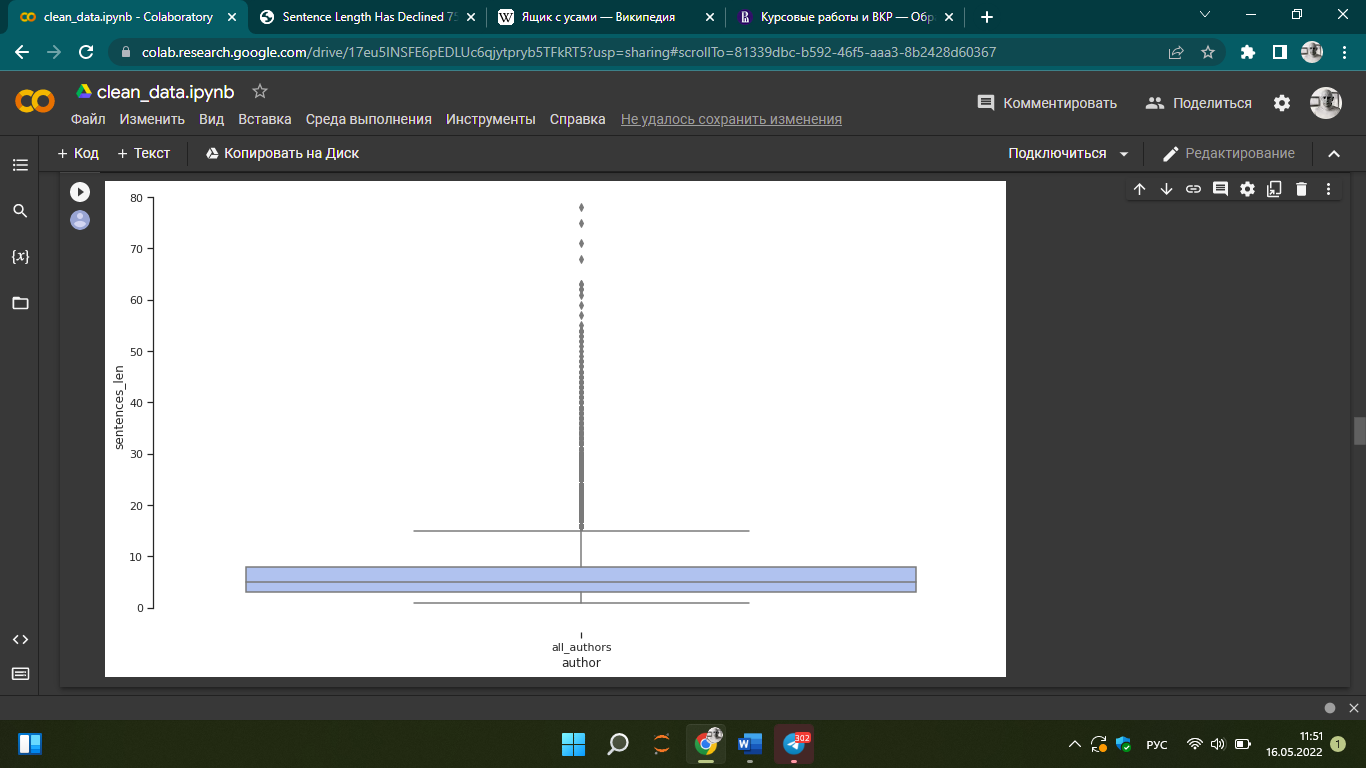


Рисунок 2

Как было сказано, стоит также посмотреть на среднее значение. Это дает более обширное представление по изучаемым данным. Средняя длина предложения была высчитана как для всего корпуса, так и для каждого автора по отдельности. Как и в случае с медианой, значение по фильмам Тарковского оказалось минимальным – 5.9. Тем не менее, при подсчёте этой переменной Бергман оказался на втором месте – 6.8, а Триер – на первом, со значением в 7.1 слов.

Итак, сравнение двух переменных – медианы и среднего значения – дают интересный результат. У Бергмана больший процент длинных предложений, тогда, как когда у Триера бывают очень длинные предложения. Это может быть связано с монологичным характером фильмов датского режиссера: зачастую, как, например, в фильме «Нимфоманка», герои долго высказывают свои мысли. Построение реплик персонажей у Триера больше похоже на прозу, внутренний монолог, исповедь. Бергман же, как модернист, строит свои фильмы более диалогично, но, судя по медиане, герои часто произносят относительно большие предложения. Интересно, что фильмы Тарковского при обоих подсчетах имеют минимальный результат. Герои этого автора действительно мало говорят или говорят короткими фразами. Работы Тарковского – созидательные и визуальные, они – наблюдение за жизнью и временем [Тарковский 1987].

## Облако слов

Помимо средней длинны предложений, безусловно, важно посмотреть на лексическую составляющую корпуса. Библиотека wordcloud позволяет визуализировать статистику самых частотных слов в корпусе. Мною было сделано 3 сценария, по каждому автору отдельно.

На рисунке 3 приставлен результат по разделу корпуса – фильмы Бергмана. В самых частотных словах находятся типичные для английских текстов слов – know, think, say и так далее. Это связано со спецификой формата данных – устная, диалогичная речь.



Рисунок 3

Далее я сделала облако слов для фильмов Тарковского (рис.4). Одно из самых заметных отличий от Бергмана – большой процент употреблений таких слов, как come и go, то есть буквально приходить и уходить. Действительно, такой результат отражает специфику фильмов Тарковского: герои всегда куда-то идут, убегают, возвращаются; у его героев будто нет определенного места, они всегда в движении. Сравнивая с фильмами Бергмана, где зачастую сцены – камерные, а персонажи могут почти весь хронометраж пробыть в одном помещении. У Тарковского нет похожих работ. Даже фильм «Жертвоприношение», которое был снят «в честь» Бергмана и в котором выражается некоторая замкнутость пространства – герои будто «заперты» на острове, - все равно динамичен. Александр, главный персонаж, постоянно бегает между своим домом и домом ведьмы, с которой он в итоге заключает сделку – обмен, жертва себя во благо спасения мира от ядерной войны.

Кроме того, в частотности употреблений слов come и go я вижу выражение цикличности фильмов Тарковского. Они имеют примерно одинаковый процент использований, то есть они рекурсивно заменяют друг друга, как пример «-1+1» будет всегда давать ноль. Если рассматривать все семь фильмов Андрея Тарковского можно заметить, что его фильмография начинается и заканчивается одним и тем же кадром ребенка, стоящего около одинокого, тонкого, почти больного дерева. Так, нарратив сначала «уходит», а затем «приходит» в изначальную точку, точку ноль.



Рисунок 4

Рисунок 5 показывает распределение по разделу датасета постмодерниста Ларса фон Триера. В нем можно заметить сходства с Ингмаром Бергманом: love, time, think находятся в верхних позициях по частоте употреблений, в отличии от Андрея Тарковского. Однако, по облаку слов Триера можно заметить и смену эпохи, в частности языковой. Такие слова, как right, one, look, really, thing используются в основном в качестве сленга. Стоит отметить, что Триер – единственный режиссер из моего исследования, который снимал многие свои знаковые фильмы на английском, а не только на своем родном языке.



Рисунок 5

Интересно рассмотреть представленные облака слов (рис. 3, рис.4, рис. 5) в их совокупности. Меня заинтересовало слово «god». Как можно видеть, у Андрея Тарковского из всех режиссеров «бог» употребляется чаще всего, к тому же, оно – одно из самых частотных из раздела с фильмами этого режиссера. В распределении Бергмана оно также присутствует, но заметно меньше, чем у Тарковского. Примечательно, что фильмы Триера предстают самыми «безбожными». Безусловно, это связано с переходом от модерна к постмодерну. Если раньше мировая философия была крайне сосредоточена на вопросе бога, то нынешнее время, 21 век – время критической теории. У Триера появляются чуждые искусству модерна понятия - sex, work, car, money.

Еще одно важное стилистическое различие – тема матери. В фильмах Бергмана и Тарковского часто встречается слово «mother», когда у Триера оно даже не помещается в облако слов. Вместо этого, у датского режиссера можно заметить слово «father». Однако, такое распределение не отражает реальное положение тем в картина. Скорее, настоящие лейтмотивы почти противоположны полученным результатам.

Так, для фильмов Тарковского фигура отца имеет такое же важное значение, как и фигура матери. Единственное отличие между этими двумя понятиями – мнимое присутствие отца, он присутствует в картинах своим отсутствием. В «Зеркале», где одна из главных ролей – это роль «матери», «отец» появляется всего лишь два раза (отсюда и малое употребление слова). Однако, отца нет – лишь буквально: он скорее выражается, например, через метавставки стихов Арсения Тарковского.

Материнский образ также важен и для Триера, что не «уловило» статистическое распределение. Как и в случае с Тарковским, «мать» не присутствует в фильмах буквально, а скорее эта фигура отражается на уровне метафоры. Почти в каждой картине Триера главный герой – это женщина. По сути, творчество этого режиссера можно описать, как попытку понять женское начало в себе. Фильм «Меланхолия», например, основан на личных переживаниях режиссера, а именно, на его депрессии. Главная героиня, как отмечает сам автор, - выражение его самого, его состояния и болезни [Carlsen 2011]. Рассматривая такую сублимацию с психоаналитической точки зрения, безусловно, женский образ в фильмах Триера можно приравнять к материнскому. По теории Жака Лакана человек, проходя стадию самоидентификации, смотрит на себя через «глаза матери» [Лакан 1949].

Как можно заметить, статистический анализ текстовых, художественных данных может дать как интересный и полезный результат, так и быть ошибочным, необъективным. Более осмысленный и неожиданный результат дает нейросетевая тематическая модель BERTopic, чему посвящена вся следующая глава моей исследовательской работы.

# Глава 2. Тематическое моделирование «Риторики фильмов Бергмана, Тарковского и Триера»

# 2.1. Выбор модели

Существуют разные методы тематического моделирования: от более простых алгоритмов, как LDA, до более сложных и новых моделей, как BigARTM. Все они генерируют осмысленные темы, которые представляют собой набор слов, схожих по семантическим свойствам.

Один из новейших подходов тематического моделирования – BERTopic [Grootendorst 2021]. Главное отличие этого метода, от других упомянутых в том, что он извлекает связное представление темы посредством разработки варианта TF-IDF на основе классов. Мера TF-IDF оценивает, насколько важно слово в документе из корпуса [Rajaraman 2011, 7]. BERTopic генерирует вложения в документах с помощью пред-обученной языковой модели, кластеризует эти сопоставления, а затем генерирует темы на основе меры TF-IDF [Grootendorst 2022, 1].

Преимущество BERTopic заключается в трехфазовой работе, что помогает достигнуть наилучшего результата. Как пишет автор, первый шаг заключается в создании вложений документов с использование обученной языковой модели. Такая мера позволяет получить информацию на уровне одного документа. Далее, модель создает семантически похожие кластеры документов, каждый из которых – отдельная тема. Заключительная фаза позволяет избежать результатов, основанных на центроидной перспективе [Abdalgader 2017]. Модель разрабатывает классовую версию TF-IDF, которая извлекает представление темы из каждого набора топиков [Grootendorst 2022, 2]. Кроме того, одними из главных преимуществ BERTopic – это удобство в использовании и новые интерактивные методы визуализации тем [Dua 2021].

Помимо вышеперечисленных преимуществ BERTopic, я использую эту модель из-за особенности собранного датасета. В отличии от других моделей, которые дают лучший результат только на больших данных, BERTopic, как отмечает автор, подходит для различных целей [Grootendorst 8, 2022]. Данные, которые я исследую, действительно, нетипичны для тематического моделирования. Это не только относительно небольшой датасет, но и формат этих данных – зачастую короткие фразы в контексте диалога. Поскольку BERTopic работает с уже обученной языковой моделью, результаты моего исследования, темы, которые алгоритм извлек, – последовательны и репрезентативны. Преимущество этой модели в том, что она может быть настроена под определенные задачи. Так, автор BERTopic в документации к алгоритму пишет, что выбор метода вложения [embedding] помогает создать тему даже с минимальным количеством доступных данных [Grootendorst, 2021].

Для своих целей я выбрала модель вложений – Sentence Transformer. Как пишет создатель BERTopic, этот трансформер дает наилучший результат [Grootendorst, 2021]. В отличие от более классических методов тематического моделирования, как, например, LDA, BERTopic учитывает контексты слов. Поэтому я не использую предобработку корпуса, описанную в первой главе: лемматизация, очищение от пунктуации и приведение к нижнему регистру. Преимущество выбранной мною модели тематического моделирования и модели вложений в том, что они максимально точно предсказывают топики из-за отсутствия «вторжения» в текстовые данные. Нетипично также и то, что при работе с этой моделью не следует убирать стоп-слова. Если удалить такие слова из корпуса, как это было сделано в первой главе моего исследования, модель не получит полный контекст. Вместо классической очистки текстовых данных от стоп-слов, я использую метод векторизации строки CountVectorizer. Он представляет весь текст как матрицу уникальных токенов и вместо удаления ненужных слов – «игнорирует» их.

В следующих разделах этой главы я представляю полученные темы: для каждого автора по отдельности, а затем для всего корпуса. Помимо результатов, я предлагаю художественный анализ этих тем в контексте со стилистическими особенностями режиссеров.

## 2.2. Тематическое моделирование для фильмов Ингмара Бергмана

BERTopic, как и другие популярные модели, позволяет визуализировать полученный результат с помощью разных интерактивных графиков. Они не только наглядны, но и практичны. Зачастую модели строят довольно много тем (от 50 до 150 и выше, в зависимости от размера данных), что может быть неудобно для подробного анализа. В BERTopic же есть инструменты, с помощью которых можно отследить похожие, близкие по семантике темы. В этом и последующих параграфах я представляю полученные результаты в виде иллюстраций, а также анализирую их, сопоставляя с *реальным* содержанием фильма, фильмографии.

На рисунке 6 показаны восемь тем. Они упорядочены по релевантности к документу. Поскольку BERTopic – это нейросетевая модель, при каждом новом запуске этого алгоритма получается немного отличающийся результат. Этого можно избежать, сохранив определенную модель. Прежде чем описывать все темы в целом, я хочу рассмотреть наиболее интересные и неожиданные варианты.

Первая тема, или «Topic 0» по-видимому относится к женским персонажам. С помощью отдельной команды я посмотрела полный набор слов к первой теме: loved, lena, took, actually, sat, nose, beautiful, girl, said, breasts. Как видно из развернутого списка, «Topic 0» посвящен феминности. Наличие в этой теме глаголов и наречия скорее всего объясняется непосредственной частотностью использования их либо с женским местоимением, либо с каким-либо женским именем (т.е. she said, she took, she actually, etc.). Интересно, что, таким образом, одна из главных тем для фильмографии Бергмана – это женщины. Тем не менее, это нельзя рассматривать как опережение Бергманом, режиссером 20 века, своего времени в феминистическом контексте. Наоборот, весь нулевой топик на самом деле лишь *проекция*, *мужской взгляд* на женщину: слова этой темы находятся вне самой девушки, это слова описания, нарратива и наблюдения за ней.

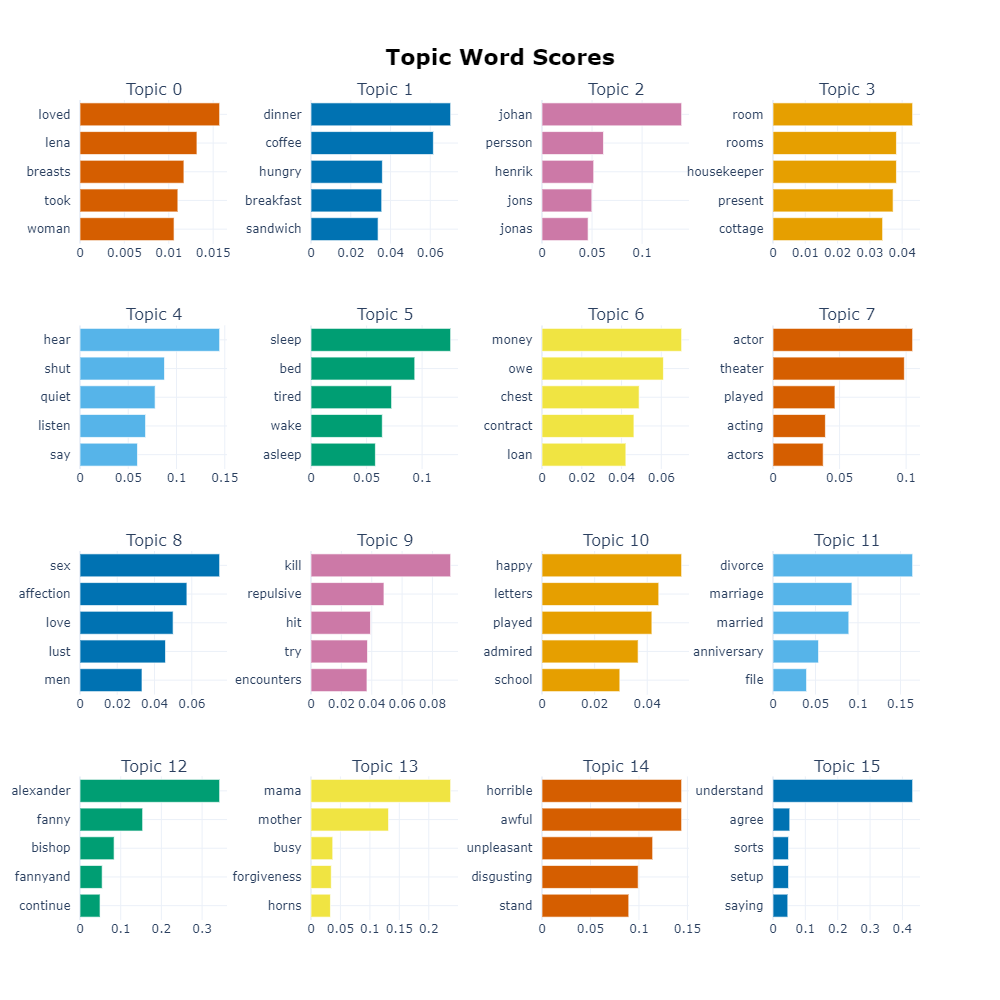


Рисунок 6

«Topic 5» и «Topic 7» значимы как отражения метанарративности, то есть рефлексивного свойства фильмов повествовать о себе: это темы сна и театра, соответственно. Зачастую кино сравнивается со сновидениями, и сам Бергман проводил такие аналогии [Schlegel 2012, 24]. Примечательно, что это отражается и на уровне лексики героев.

На рисунке 7 показаны еще восемь значимых темы. Суммарно по Бергману я представляю 16 топиков, хотя BERTopic сгенерировал 197. Оставшиеся результаты удобно анализировать, например, с помощью другого инструмента, который находит похожие темы по ключевому слову.

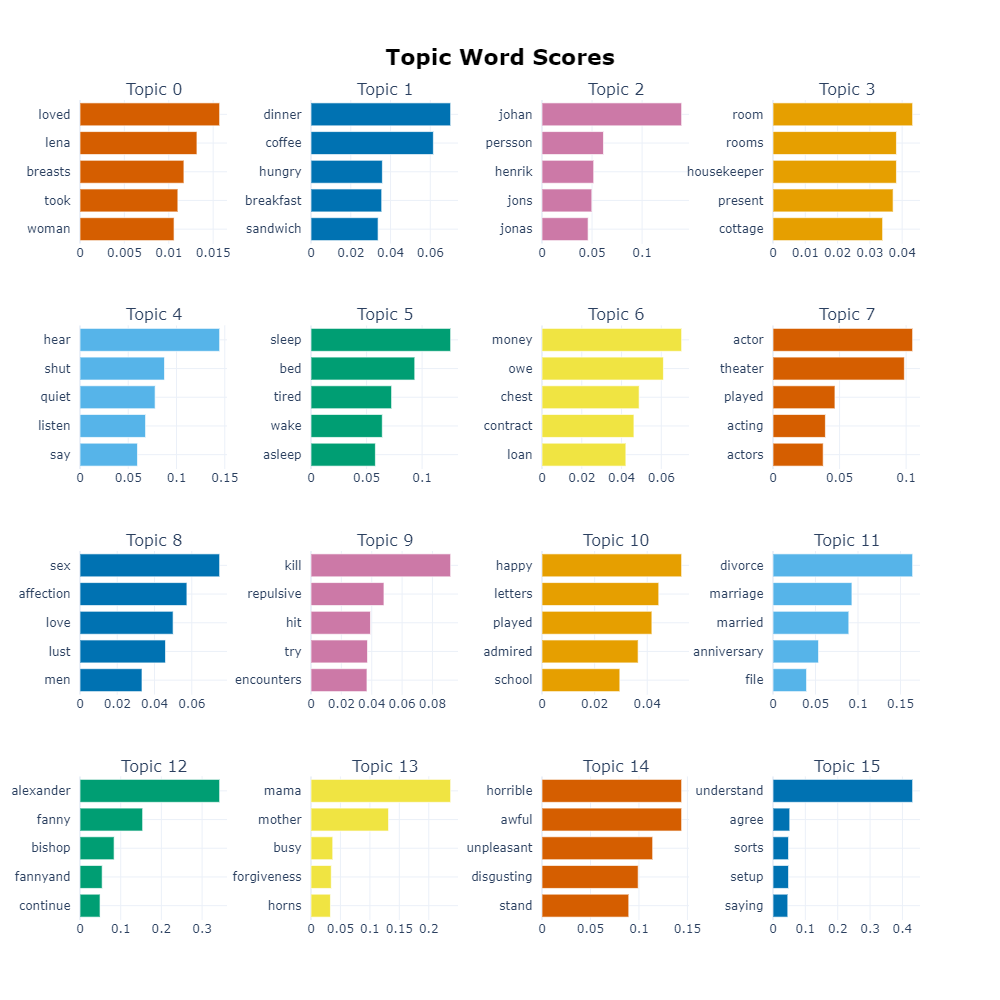


Рисунок 7

Из рисунка 7 стоит выделить две темы, которые я нахожу почти противоположными: тема под номером 8 и тема 13. В них наиболее ярко выражается тот *мужской взгляд*, о котором я говорила выше. В 8 теме такие понятия как секс, похоть, любовь стоят вместе со словом «мужчины», в теме же 13 мать, что тоже выражение феминности, связывается с занятостью и прощением.

С помощью встроенной в BERTopic функции по нахождению похожих тем я решила более детально изучить два термина «мужчины» и «женщины». Интересно, что некоторые топики для обоих слов одинаковы. Например, тема «children, grandchildren, support, grown, sake, pay workers, duties, conscientiously, custody» присутствует у обоих полов. Однако, BERTopic выделил также и две «женские» темы, одна из которой посвящена еде и готовке, а другая, как я это определяю, относится к обязанностям матери («wife, husband, seen, knowing, looking, playacting, responsible»).

Что же этот результат значит для настоящего исследования? Во-первых, это подтверждает гипотезу о *male’s gaze.* Несмотря на то, что Бергман часто снимал свои фильмы с главными персонажами-женщинами, они все равно *лишены* объективной репрезентативности – героини были созданы мужчиной. Во-вторых, такие результаты тематического моделирования, в том числе и вызванные ими выводы, на данном этапе говорят об эффективности BERTopic применительно к «художественным» данным. Рассматривая 16 тем с рисунка 6 и рисунка 7, можно заметить, что в них будто нет *автора,* то есть авторского стиля: ключевые слова топиков довольно абстрактные и нейтральные, по ним практически невозможно установить *настоящие* лейтмотивы фильмов, их сценарии. Так, этот метод цифрового анализа искусства позволяет занять отдаленную позицию, позицию «над» произведением.

# 2.3. Тематическое моделирование для фильмов Андрея Тарковского

Как и в случае с фильмами Бергмана, результаты по Андрею Тарковскому я представляю в виде двух иллюстраций по восемь самых релевантных тем по разделу корпуса. На рисунке 8 видно, что темы по фильмографии русского режиссера также абстрактны, тем не менее по ним более четко прослеживаются сценарии и *возможные* сеттинги. Так, например, «Topic 5» можно соотнести с наличием в фильмографии режиссера картины про космос, что в действительности подтверждается «Солярисом». Однако, темы цивилизации, свободы и мира присутствуют не только в упомянутой работе, но также во многих других – от «Андрея Рублева» до «Жертвоприношения». Это важно для понимания продуктивности и правильности работы выбранной мною тематической модели – BERTopic не искажает результат даже на маленьком корпусе из 7 фильмов. То есть на представленных темах (рис.8) не наблюдается перекос к какой-либо определенной тематике несмотря на то, что все фильмы, безусловно, очень разные, в том числе и по размеру.

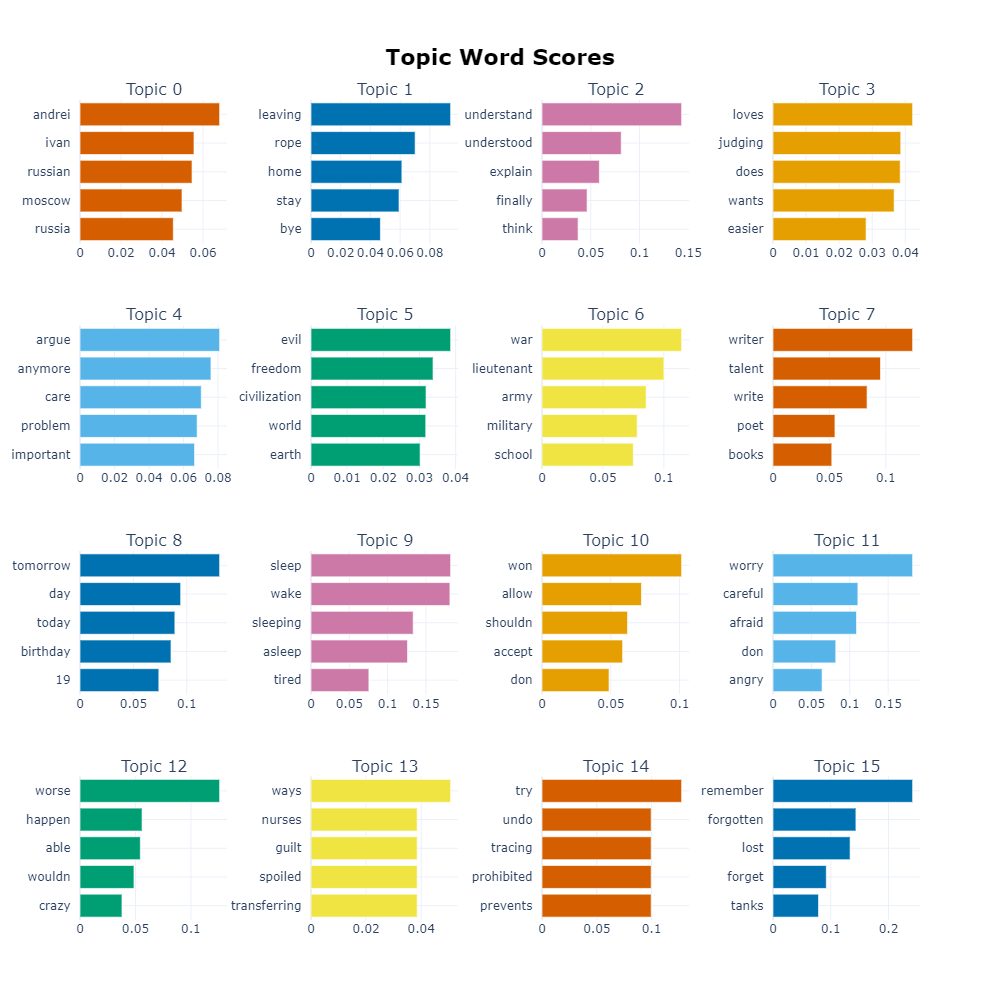


Рисунок 8

Перед тем как смотреть на все 16 тем в целом, я, как и в предыдущем параграфе, хочу сфокусироваться на некоторых, важных для анализа творчества Тарковского темах. Один из таких топиков – «Topic 7» (рис.8). Похожий результат был и у Бергмана, а именно тема про театр и актеров. У Тарковского же метанарратив, «фильм о фильме» - сравнение киноискусства с письмом. Примечательно, в своем теоретическом эссе «Запечатленное время» Андрей Тарковский критиковал *слияние* кинематографа с театром. Он разграничивал эти две парадигмы, считая, что кино, в отличии от простой «фиксации [на пленку] театрального зрелища» способно «запечатлеть реальность времени» [Тарковский, 25]. В этом проявляется отличие Тарковского от своего шведского коллеги, Бергмана. Кроме того, в этом контексте важно понимать и то, что Тарковский отрицал и литературность кинематографа. Тем не менее, самым близким жанром из всех искусств к кино он считал хокку [Тарковский, 27]

Еще один аргумент к метанарративности, полученный с помощью BERTopic, обнаруживается на интерактивной матрице подобия тем (рис.9). Это «тепловая» карта, сгенерированная по косинусному сходству, показывающая насколько какие-либо темы похожи друг на друга. Меня интересует тема 32, которая посвящена природным явлениям, и, в то же время, в которой можно найти слово «film». По матрице подобия наибольший балл схожести эта тема имеет с темой 9, которую можно охарактеризовать, как «сновидения».

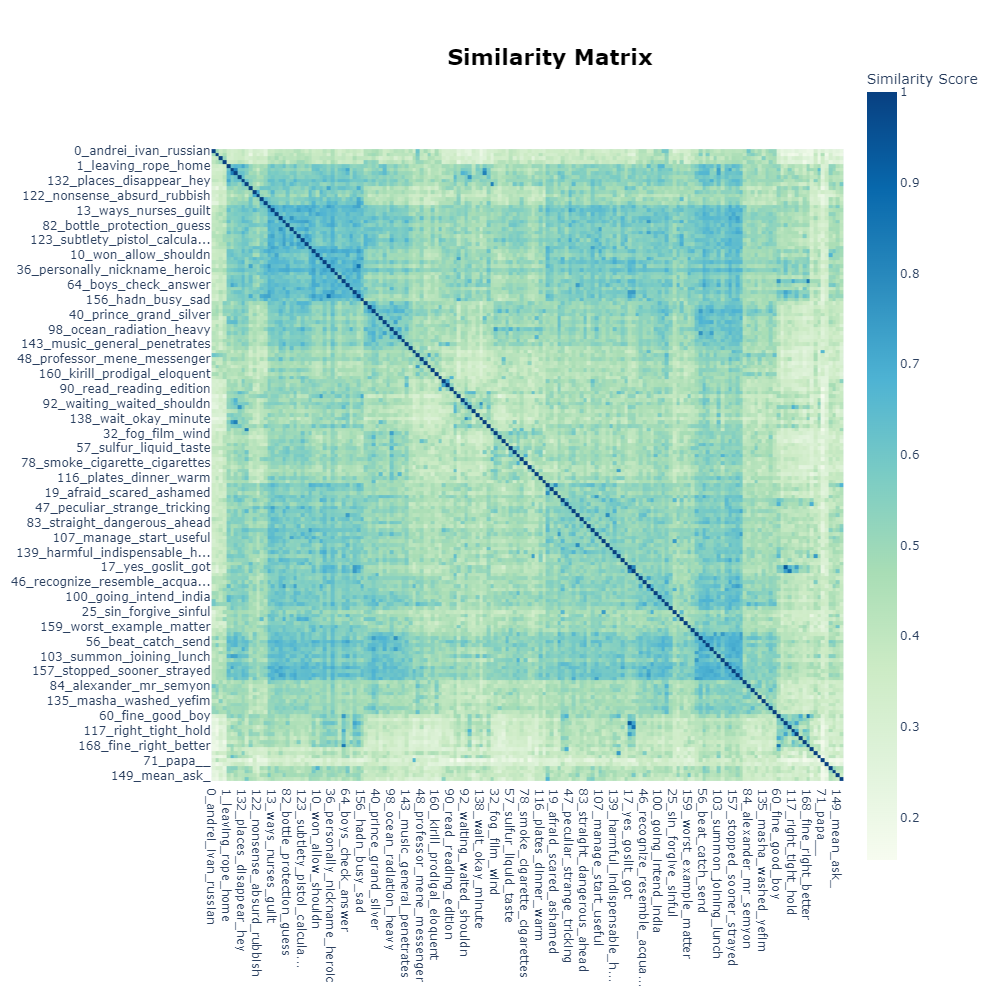


Рисунок 9

Так, теория Тарковского отражается и на лексическом уровне реплик его героев. Природа, сновидения – это важные и главные лейтмотивы в стилистике киноязыка Андрея Тарковского. Как писал сам автор, «дождь, огонь, вода, снег, роса, поземка — часть той материальной среды, в которой мы обитаем, правда жизни» [Тарковский, 98]. В этом контексте стоит снова вспомнить цитату Бергмана о киноязыке Тарковского, «в котором жизнь предстаёт как зеркало, как сон» [Schlegel 2012, 24].

В заключение, стоит рассмотреть все темы вкупе (рис.8, рис.10). На рисунке 10 представлена оставшаяся часть результатов по тематическому моделированию. Как и в первой главе (рис.4), в этих темах можно снова заметить выражение цикличности в лексическом контексте: «Тема 1» с восьмого рисунка – «stay», «leave»; Тема 8 с десятого рисунка – «tomorrow», «today»; «Тема 9»– «sleep», «wake»; Тема 15 – «remember», «forget». Кроме того, можно заметить глаголы, наречия, прилагательные, которые в основном выражают какие-либо негативные эмоции – wore, angry, worse, crazy, lost, evil, judging. Это говорит о тональности фильмов Тарковского: в основном они повествуют о травматических событиях, непонимании со стороны близких и со стороны Родины.

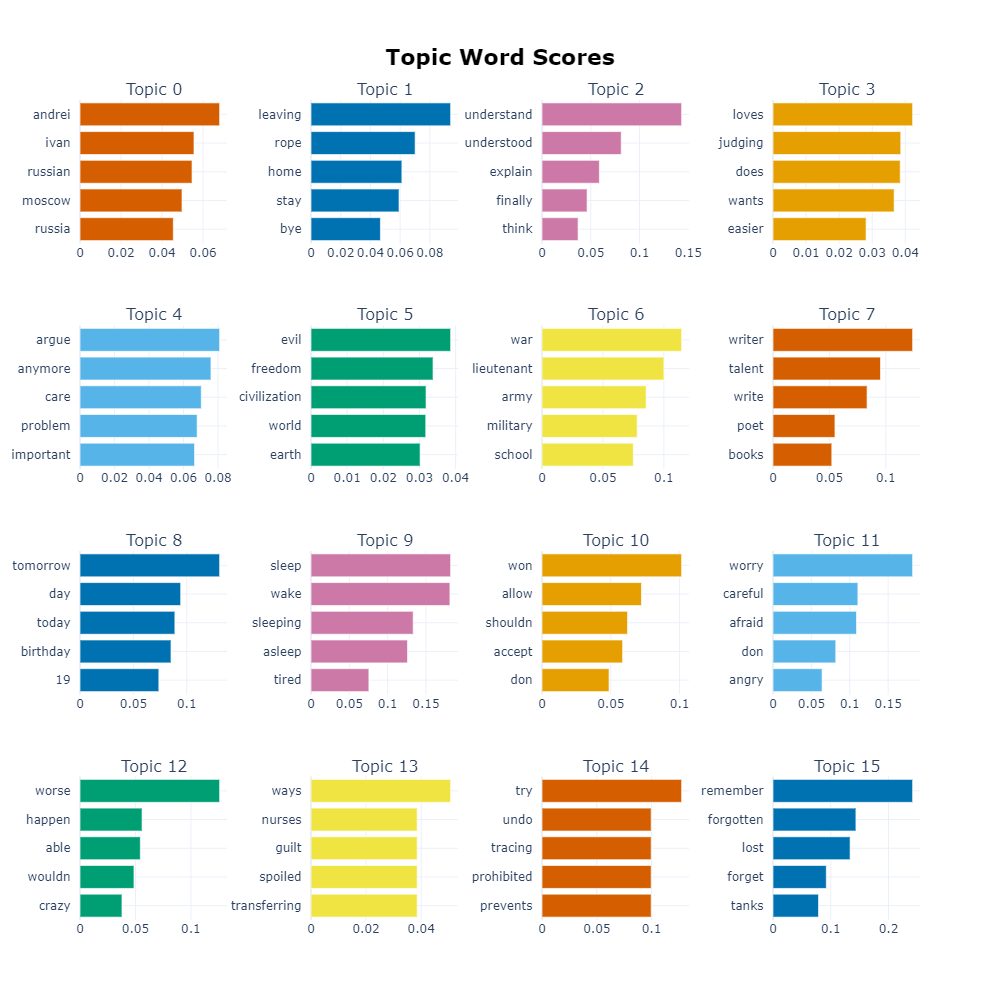


Рисунок 10

## 2.4. Тематическое моделирование для фильмов Ларса фон Триера

Самый неожиданный результат был получен по корпусу картин Триера. Фильмы этого режиссера, по изначальной гипотезе, представлялись самыми продуктивными для тематического моделирования, поскольку Триер – постмодернист без четкого стиля. Он снимал как по своему строгому манифесту «Догма 95» [Триер 1995], так и отказывался от определенных предписаний к киноформе. Каждая работа Триера – это уникальный сценарий, персонажи и сеттинг. Однако, именно эти обстоятельства повлияли на полученные результаты. В отличии от результатов по Бергману и Тарковскому, топики, сгенерированные по корпусу Триера, неабстрактны – зачастую по ним легко читается определенная картина, в которой та или иная тема главенствует. В этом разделе работы я разберу, с чем именно связано такое искажение, а также рассмотрю наиболее удачные темы в контексте киноязыка Ларса фон Триера.

Разница в результатах трех режиссеров показана в таблице 1. В ней находятся три графика-распределений удаленности тем по корпусам Бергмана, Тарковского и Триера, соответственно. Заметно, что график по датскому режиссеру отличается: если на первых двух темы распределены равномерно, зачастую с равной удаленностью друг от друга, то на последнем видно, что, в основном, темы образуют «независимые» кластеры.

Таблица 1 – сравнение трех графиков по авторам

|  |  |  |
| --- | --- | --- |
| Ингмар Бергман | Андрей Тарковский | Ларс фон Триер |
|  |  |  |

Кластеризация определенных тем по определенному фильму подтверждается и на лексическом уровне результатов. На рисунке 11 представлены восемь релевантных к корпусу Триера тем, из которых видно три очевидно относящихся к конкретным фильмам – Topic 1 и Topic 5 к «Нимфоманке», Topic 3 к «Догвиллю». Остальные результаты, однако, – связные, разнообразные темы. Во многих из них присутствуют слова, характерные для эпохи постмодерна: сленг, обозначения терминов относительно современных для кино, как например «phone» [телефон]. Можно заметить также пересечение в мотивах с предыдущими авторами. Шестой топик посвящен семье, в основном – отцу.

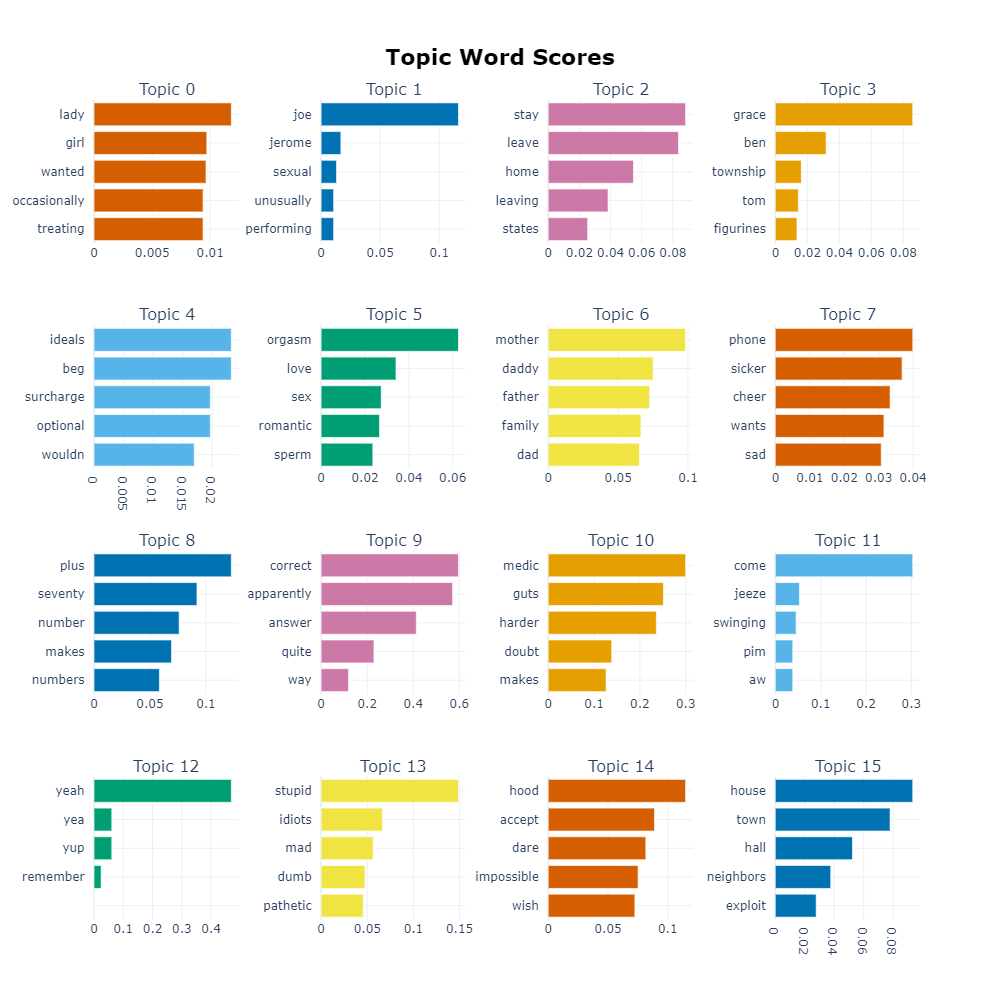


Рисунок 11

Сценарии Триера кажутся из всех исследуемых в этой работе режиссеров самыми разнообразными по действию и месту действия. Так почему же результаты именно по его кинокартинам оказались самыми «читаемыми»? Можно заметить, что его фильмы, по-видимому, меньше сходятся в лексике из-за несхожести сеттинга и лейтмотивов, а также преобладания определенных терминов в нескольких фильмах и практически полное отсутствие в других. В качестве примера, «Дом, который построил Джек» рассказывает о серийном убийце, соответственно в нем очень много лексики, связанной с убийствами, подготовками к ним. С помощью интерактивного графика распределения дистанции между темами (таблица 1) я установила кластер упомянутого фильма: в нем находятся такие слова, как «зло», «совершить», «убить», «серийный» и т.д. При этом, в этом кластере можно обнаружить тему, относящуюся к «Танцующей в темноте», так как одно из слов – имя главной героини, Сельмы. Это практически единственный фильм, помимо «Дом, который построил Джек», где присутствует акт убийства в качестве преступного действия, преступления *закона,* а не столько *морали.* Кроме того, я обнаружила еще несколько кластеров по 2-4 фильма со схожими или одинаковыми событиями, которые дают такую тесную группировку по темам.

Таким образом, моя гипотеза подтвердилась только частично. Фильмы Ларса фон Триера все же не такие обособленные друг от друга по своей лексике. Зачастую можно встретить пересечения на языковом уровне, которые отражают пересечение событий и мотивов из фильмов. Тем не менее, тематическое моделирование по картинам Триера дает сгруппированные, кластеризированые темы. Безусловно, это стоит связывать с направлением, к которому можно отнести датского режиссера – с постмодернизмом. Фильмы Бергмана и Тарковского, по-видимому, менее разнообразны лексически, из-за чего все их темы довольно равномерно удалены друг от друга. Безусловно, в них присутствуют узкие термины, которые либо отличают, либо связывают какие-либо картины вместе. В случае фильмографии Триера лексика становится более «расширенная», разговорная, индивидуальная. Примечательно, что даже с такими данными BERTopic смог построить связанные и последовательные темы, которые действительно отражают релевантные мотивы фильмов.

## 2.5. Тематическое моделирование всего корпуса «Риторика фильмов Бергмана, Тарковского и Триера»

Помимо тематического моделирования в контексте одного автора, безусловно, практическую ценность несет применение BERTopic ко всему составленному мною корпусу. Предыдущие результаты, сгенерированные по конкретной фильмографии, дают представление о творчестве одного режиссера в целом: как определенные темы, мотивы взаимодействуют друг с другом, какая лексика более характерна для конкретного автора. Темы, полученные для Бергмана, Тарковского и Триера по отдельности, предоставляют возможность рассмотреть стилистику, характерную не только для перечисленных авторов, но и для эпохи, которой они принадлежат. По этим трем результатам, отчасти, можно проследить развитие киноязыка диахронически. Однако, Бергмана, Тарковского, Триера стоит выделять в собственную экосистему, которая отличается как от мейнстрим-кинематографа, так и от фестивальных картин.

Как было показано в предыдущих параграфах, модель создает много разных тем, в чем-то пересекающихся между собой. На это также влияет и разнообразие сеттингов: каждый фильм каждого автора – уникален в персонажах, сюжете, лейтмотивах. Чтобы избежать большого разброса по темам и, следовательно, чтобы мои выводы по тематическому моделированию были более последовательны, я изменила настройки BERTopic, оставляя только 50 тем из 138. Почему же редуцирование до определенного количества топиков не ухудшает результат, а, наоборот, делает его более читаемым? Дело в том, что редукция не удаляет остальные темы, а *совмещает* наиболее похожие. Выбранное количество тем (50) представляется самым удачным, поскольку, если задать большее – изменения не ощутимы, если задать меньшее – темы становятся нечитаемыми.

Как и в предыдущих параграфах работы, основа выводов – результаты тематического моделирования по корпусу. Я так же смотрю только на 16 самых частотных и релевантных тем, что дает наилучшее представление об изучаемым текстовых данных.

Итак, на рисунке 12 представлены полученные темы по корпусу «Риторика фильмов Бергмана, Тарковского и Триера». В отличие от других частей работы, эти результаты я хочу рассмотреть вкупе, поскольку они составляют заключительную часть моего исследования фильмографии трех авторов.

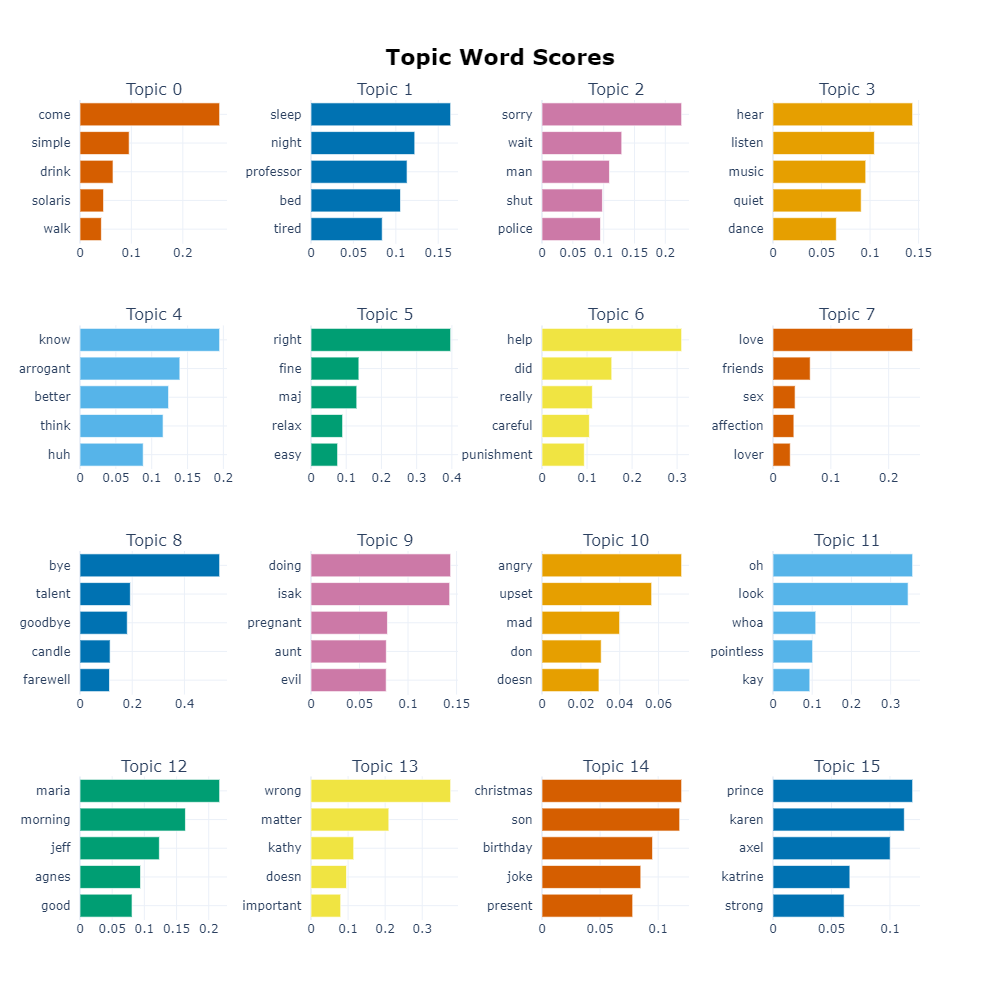


Рисунок 12

Несмотря на то, что темы (рис. 12) построены по всему корпусу, зачастую в них присутствуют имена и названия из разных фильмов (например, «maria», «solaris», «karen»). Это – небольшая погрешность модели, которая не влияет в значительной степени на выводы. В основном, темы последовательны и абстрактны. Во многих из них можно встретить глаголы императивного характера («listen», «come», «help»), что связано в первую очередь с форматом данных – устная, диалогичная речь. Однако, есть также и примечательные примеры тем.

Для всех авторов характерен лейтмотив сна – «Topic 1» (рис. 12), при этом он стоит одним из первых по значимости. Это можно трактовать, как метанарратив, который объединяет всех авторов. Кино зачастую сравнивают со сновидениями и грезами, а процесс просмотра фильма с погружением в глубокий сон. Такой результат примечателен, поскольку ни у одного из режиссеров нет фильма, в котором этот мотив был главенствующим. Из-за этого можно сделать вывод, что тема сна хоть и часто встречается, но она проходит едва замеченной, как бы вскользь.

Кроме того, меня интересует тема под номером 7, которую можно охарактеризовать, как «отношения». Несмотря на то, что в ней присутствуют разные виды взаимоотношений от дружбы до секса, самый главный и релевантный термин для этого топика – любовь, что показывает индекс значимости.

Темы 3 и 11 интересны, как отражение мультимедийности кинематографа. Первая названная тема посвящена звуку, во второй можно найти слово «look» (смотреть). Действительно, эти два чувства – ключевые для любого фильма, именно через них зритель воспринимает картину. Поэтому такой результат снова можно интерпретировать, как метанарратив, который прослеживается у трех авторов.

Таким образом, тематическое моделирование применительно к всему корпусу дает продуктивный результат. В темах с рисунка 12 можно проследить скрытые мотивы, которые присутствуют у трех авторов в совокупности. По сути, тематическая модель BERTopic смогла совместить несовместимое: модерн и постмодерн. В какой-то степени, полученные результаты – отражение кинематографа в целом, поскольку при всей «общности» Бергмана, Тарковского и Триера – это разные режиссеры, которых отличают выбираемые мотивы и стилистика.

ЗАКЛЮЧЕНИЕ

В этой исследовательской работе был рассмотрен корпус «Риторика фильмов Бергмана, Тарковского и Триера» с помощью цифровых методов. В перовой главе я исследовала собранные мною текстовые данные статистически, а также привела визуализации полученных результатов. На основе статистики были сделаны выводы о художественных особенностях авторов.

Вторая глава – центральная часть моего исследования. В ней я представляла результаты по тематическому моделированию: по каждому автору отдельно и по всему корпусу. Было установлено, что тематическое моделирование применительно к художественным данным – продуктивно. С помощью BERTopic были выявлены закономерности, отражающие определенную лексическую стилистику одного автора, и сделаны выводы про особенности кинематографа в целом.

Представленные мной результаты и выводы доказали, что дедуктивное изучение искусства сильно отличается от имманентного анализа. Конечно, тематические модели или статистика не могут уловить те черты искусства, которые базируются на чувствах и ассоциациях человека. Тем не менее, цифровые методы, а в частности тематическое моделирование, помогают обнаружить скрытые человеческому взгляду статистики. Как например, тема сна, которая наблюдается у каждого автора, но не прослеживается так явно при художественном разборе какого-либо фильма из корпуса. Для полного цифрового изучения кинематографа потребовались бы разные инструменты: от NLP до компьютерного зрения. В этом исследовании за основу были взяты лишь текстовые данные из фильмов. Но несмотря на такой маленький объём и спорный формат данных, были обнаружены интересные факты по фильмографиям режиссеров.

Изучение искусства на основе цифровых методов не претендует на объективность, как это не делает и человеческий взгляд. Но в отличии от классического анализа произведений, тематическое моделирование и статистическое исследование основываются на цифрах и предоставляют точные результаты, даже в такой хаотичной парадигме, как искусство. В современности представляется важным и рациональным использовать как традиционные методы искусствоведческого анализа, так и методы дальнего чтения.

## СПИСОК ИСТОЧНИКОВ

1. Abdalgader, Khaled. Clustering short text using a centroid-based lexical clustering algorithm // *International Journal of Computer Science*, 2017.
2. Abeer Abuzayed, Hend Al-Khalifa. BERT for Arabic Topic Modeling: An Experimental Study on BERTopic Technique. [Электронный ресурс] // Procedia Computer Science URL: [www.sciencedirect.com/science/article/pii/S1877050921012199](http://www.sciencedirect.com/science/article/pii/S1877050921012199) (дата обращения: 31.05.2022).
3. Bennet, Andrew et al. “Have you tried Neural Topic Models? Comparative Analysis of Neural and Non-Neural Topic Models with Application to COVID-19 Twitter Data” [Электронный ресурс] // arXiv URL: [arxiv.org/pdf/2105.10165.pdf](https://arxiv.org/pdf/2105.10165.pdf) (дата обращения: 31.05.2022).
4. Bird, Klein, Loper. Natural Language Processing with Python. California: O'Reilly Media, 2009.
5. Blei, David M.; Ng, Andrew Y.; Jordan, Michael. Latent Dirichlet allocation // *Journal of Machine Learning Research*, 2003.
6. Carlsen, Juul. The Only Redeeming Factor is the World Ending. [Электронный ресурс] // DET DANSKE FILMINSTITUT, URL: [www.dfi.dk/english/only-redeeming-factor-world-ending](http://www.dfi.dk/english/only-redeeming-factor-world-ending) (дата обращения: 31.05.2022).
7. Dua, Sejal. Dynamic Topic Modeling with BERTopic. [Электронный ресурс] // Medium, URL: towardsdatascience.com/dynamic-topic-modeling-with-bertopic-e5857e29f872#:~:text=The%20two%20greatest%20advantages%20to,its%20novel%20interactive%20visualization%20methods (дата обращения: 31.05.2022).
8. Grootendorst, Marten. BERTopic [Электронный ресурс]// GitHub, URL: maartengr.github.io/BERTopic/getting\_started/quickstart/quickstart.html
9. Grootendorst, Marten. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv*, 2022.
10. HyunSeung, Koh and Mark Fienup. “Topic Modeling as a Tool for Analyzing Library Chat Transcripts.” Information Technology & Libraries, no. 3, 2021, pp. 1-24
11. Jun Hyun, Ryoo, et al. “Do Spoilers Really Spoil? Using Topic Modeling to Measure the Effect of Spoiler Reviews on Box Office Revenue.” Journal of Marketing, no. 2, 2021, pp. 70-88.
12. May C., Cotterell R., Van Durme B. An Analysis of Lemmatization on Topic Models of Morphologically Rich Language [Электронный ресурс] // arXiv, URL: arxiv.org/pdf/1608.03995.pdf (дата обращения: 31.05.2022)
13. Navarro-Colorado, Borja. On Poetic Topic Modeling: Extracting Themes and Motifs from a Corpus of Spanish Poetry // *Frontiers in Digital Humanities*, 2018.
14. Rajaraman, A. Data Mining. [Электронный ресурс] // URL: i.stanford.edu/~ullman/mmds/ch1.pdf (дата обращения: 31.05.2022).
15. Rha, Lauren and Silver, Sean. Topic Modeling and Analysis: Comparing the Most Common Topics in 19th-Century Novels Written by Female Authors // *Aresty,* 2021.
16. Schlegel H.-J. Tarkovsky: Films, Stills, Polaroids and Writings. London: Thames and Hudson, 2012.
17. Schöch, Christof. “Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama.” [Электронный ресурс] // arXiv URL: [arxiv.org/ftp/arxiv/papers/2103/2103.13019.pdf](https://arxiv.org/ftp/arxiv/papers/2103/2103.13019.pdf) (дата обращения: 31.05.2022).
18. Shahbazi, Z. and Yung-Cheol Byun. “Topic prediction and knowledge discovery based on integrated topic modeling and deep neural networks approaches.” *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, no. 1, 2021.
19. Sharaff, A, et al. “Prospecting the Effect of Topic Modeling in Information Retrieval.” *International Journal on Semantic Web and Information Systems 2021*, no. 3, 2021.
20. The Acropolitan. Sentence Length Has Declined 75% in the Past 500 Years [Электронный ресурс] // Medium, URL: [medium.com/@theacropolitan/sentence-length-has-declined-75-in-the-past-500-years-2e40f80f589f#:~:text=On%20average%2C%20sentences%20today%20range,per%20sentence%20in%20some%20years](mailto:medium.com/@theacropolitan/sentence-length-has-declined-75-in-the-past-500-years-2e40f80f589f#:~:text=On%20average%2C%20sentences%20today%20range,per%20sentence%20in%20some%20years) (дата обращения: 31.05.2022).
21. Uglanova, Inna and Gius, Evelyn. The Order of Things. A Study on Topic Modelling of Literary Texts. [Электронный ресурс] // URL: ceur-ws.org/Vol-2723/long7.pdf. (дата обращения: 15.05.2022)
22. Воронцов К.В. Вероятностное тематическое моделирование. [Электронный ресурс] // Machine Learning, URL: [www.machinelearning.ru/wiki/images/2/22/voron-2013-ptm.pdf](http://www.machinelearning.ru/wiki/images/2/22/voron-2013-ptm.pdf) (дата обращения: 31.05.2022).
23. Долин, А. Ларс фон Триер. Контрольные работы. М: Новое литературное обозрение, 2019 г.
24. И. А. Ирхин, В. Г. Булатов, К. В. Воронцов, Аддитивная регуляризация тематических моделей с быстрой векторизацией текста, *Компьютерные исследования и моделирование*, 2020, том 12, выпуск 6, 1515–1528
25. Лакан Ж. Стадия Зеркала и ее роль в формировании функции Я [Электронный ресурс] // Психоанализ, URL: https://psychoanalysis.by/2019/03/26/%D1%81%D1%82%D0%B0%D1%82%D1%8C%D1%8F-%D0%B6-%D0%BB%D0%B0%D0%BA%D0%B0%D0%BD-%D1%81%D1%82%D0%B0%D0%B4%D0%B8%D1%8F-%D0%B7%D0%B5%D1%80%D0%BA%D0%B0%D0%BB%D0%B0-%D0%B8-%D0%B5%D0%B5-%D1%80%D0%BE%D0%BB/ (дата обращения: 31.05.2022).
26. Милкова М.А. Тематические модели как инструмент дальнего чтения. [Электронный ресурс] // Digital Economy, URL: digital-economy.ru/images/easyblog\_articles/520/DE-2019-01-06.pdf (дата обращения: 31.05.2022).
27. Моретти, Ф. Дальнее чтение / пер. с англ. А. Вдовина, О. Собчука, А. Шели. Науч. ред. перевода И. Кушнарева. — М.: Изд-во Института Гайдара, 2016.
28. Тарковский А.А. Запечатленное время [Электронный ресурс] // Royallib, URL: [docs.google.com/document/d/1XWVKSvNgE3A7QZ2Urg16PSBQC0e3-KCy/edit](https://docs.google.com/document/d/1XWVKSvNgE3A7QZ2Urg16PSBQC0e3-KCy/edit) (дата обращения: 31.05.2022).
29. Триер Л., Винтерберг Т. «Догма 95». Манифест [Электронный ресурс] //История Кино, URL: [kinoart.ru/texts/dogma-95-manifest](https://kinoart.ru/texts/dogma-95-manifest) (дата обращения: 31.05.2022).