

# Датасет и разработка инструмента учебной аналитики для извлечения проявлений студенческой агентности из текстов отзывов на MOOK

Юлия Дюличева

Статья поступила  
в редакцию  
в январе 2023 г.

**Дюличева Юлия Юрьевна** — кандидат физико-математических наук, доцент кафедры прикладной математики, Крымский федеральный университет им. В.И. Вернадского. Адрес: 295007 Симферополь, просп. Академика Вернадского, 4. E-mail: [dyulicheva\\_yu@mail.ru](mailto:dyulicheva_yu@mail.ru). ORCID: <https://orcid.org/0000-0003-1314-5367>

Аннотация

Исследование посвящено разработке способов автоматического обнаружения в текстах отзывов на MOOK проявлений составляющих и источников студенческой агентности, а также описаний внутренних и внешних трансформаций, происходящих в процессе изучения MOOK. Для извлечения описаний, соответствующих индивидуальным, реляционным и контекстуальным источникам студенческой агентности, сформирован датасет из 3445 англоязычных комментариев к наиболее популярным курсам по математике, представленным на платформе *Udemu*, а также дополнительно выделены 1787 комментариев на практикоориентированные MOOK и MOOK по предпринимательству для извлечения описаний внутренней и внешней трансформации у слушателей MOOK. Для извлечения из текстов отзывов на MOOK ключевых слов и их сочетаний предложен методологический подход, основанный на применении методов обработки естественного языка, таких как тематическое моделирование, анализ тональности и анализ частотности N-грамм. На основании этих ключевых слов и их сочетаний описаны проявления составляющих индивидуального источника студенческой агентности в виде самооэффективности, усиления чувства уверенности при решении задач и мотивации; составляющих реляционного источника в виде поддержки и сопровождения онлайн-курса со стороны тьютора с помощью быстрых ответов и хорошо структурированного учебного материала; составляющих контекстуального источника в виде предоставления возможности принимать решение при выборе альтернативных онлайн-курсов. На том же эмпирическом материале описаны проявления внутренней трансформации обучающихся как переход от преодоления страхов, неуверенности, разрешения проблем в усвоении контента MOOK к пониманию цели обучения, а также проявления внешней трансформации в виде создания нового или изменения структуры существующего продукта, стартапа или бизнеса через изменение мышления.

Ключевые слова

MOOK, учебная аналитика, студенческая агентность, тональность, униграммы, биграммы, тематическое моделирование

Для цитирования

Дюличева Ю.Ю. (2024) Датасет и разработка инструмента учебной аналитики для извлечения проявлений студенческой агентности из текстов отзывов на MOOK. *Вопросы образования / Educational Studies Moscow*, № 1, сс. 288–313. <https://doi.org/10.17323/vo-2024-16677>

## Dataset and Development of Learning Analytic Tool to Extract Manifestations of Students' Agency from Texts of Comments from MOOCs

Yulia Dyulichева

**Yulia Yu. Dyulichева** — Candidate of Sciences in Physics and Mathematics, Associate Professor at the Department of Applied Mathematics, V.I. Vernadsky Crimean Federal University. Address: 4, Academic Vernadsky Ave., 295007 Simferopol, Russian Federation. E-mail: dyulicheva\_yu@mail.ru. ORCID: <https://orcid.org/0000-0003-1314-5367>

**Abstract** The study is devoted to the automatic identification of manifestations of various components and sources of student agency from the texts of reviews of MOOCs, as well as descriptions of internal and external transformation among students in the process of studying MOOCs. To extract descriptions corresponding to individual, relational and contextual sources of students' agency, a dataset of 3445 English-language comments on the most popular mathematics courses presented on the Ude-my platform was generated, and additionally 1787 comments on practice-oriented MOOCs and entrepreneurship MOOCs were extracted to understand the descriptions corresponding manifestation of internal and external transformation in MOOC listeners. The paper proposes a methodological approach based on the use of natural language processing methods such as topic modeling, sentiment analysis and N-gram frequency analysis for extracting keywords and their combinations from MOOCs' comments texts to describe the manifestation of the components of an individual source of student agency in the form of self-efficacy, increased sense of confidence in solving problems and motivation; components of the relational source in the form of support and accompaniment of the online course by the tutor with the help of quick answers and well-structured educational material; components of the contextual source in the form of the ability to make decisions when choosing alternative online courses, as well as descriptions of the manifestation of internal transformation of students, expressed in the transition from internal struggle — overcoming fears, uncertainty, difficulties in perceiving MOOC content to understanding the purpose of learning and external transformation, expressed in the texts of comments on MOOCs in the form of creating a new or changing the structure of an existing product, startup or business through a change in thinking.

**Keywords** MOOC, learning analytics, student agency, sentiment analysis, unigrams, bigrams, topic modeling

**For citing** Dyulichева Yu.Yu. (2024) Dataset i razrabotka instrumenta uchebnoy analitiki dlya izvlecheniya proyavleniy studencheskoy agentnosti iz tekstov otzyvov na MOOK [Dataset and Development of Learning Analytic Tool to Extract Manifestations of Students' Agency from Texts of Comments from MOOCs]. *Voprosy obrazovaniya / Educational Studies Moscow*, no 1, pp. 288–313. <https://doi.org/10.17323/vo-2024-16677>

Агентность стала в последние годы одним из наиболее активно развивающихся направлений исследований в разных областях науки: в *Scopus* заметный рост числа публикаций по этой тематике наблюдается с 2016 г. [Сорокин, Зыкова, 2021]. Исследования показывают, что агентность играет ключевую роль в процессе обучения в течение всей жизни и способствует развитию навы-

ков, помогающих адаптироваться к быстро изменяющимся условиям жизни и высокому уровню неопределенности, — таких как креативность и умение взаимодействовать с другими людьми [Littleton, Taylor, Eteläpelto, 2011]. Исследования агентности, особенно активизировавшиеся сегодня, имеют богатую традицию: Д. Дьюи связывал понятие агентности со свободой выбора и обдумыванием (принятием решений) [Dewey, 1922], Л.С. Выготский рассматривал проявление агентности при социальном взаимодействии [Vygotsky, 1980], А. Бандура ассоциировал студенческую агентность с самоэффективностью и способностью корректировать действия для достижения цели [Bandura, 1986; 1999], Л. Ахерн определяла агентность как социокультурно опосредованную способность к действию [Ahearn, 2001]. В проекте Организации экономического сотрудничества и развития *OECD Future of Education and Skills 2030*<sup>1</sup> сформулировано представление о студенческой агентности как о сложном явлении, сочетающем идентичность, чувство принадлежности, мотивацию, надежду, самоэффективность, мышление роста и чувство цели, а также способность трансформировать учебный процесс, учебную среду и стратегии обучения [Mameli, Passini, 2018; Matos et al., 2018]. Многие исследователи придерживаются мнения, что агентность присуща группам людей, а следовательно, требует навыков общения и умения выражать свои идеи посредством языка [Nogueira, 2017].

Поскольку агентность проявляется в социальном взаимодействии, а средством такого взаимодействия является язык, исследователи изучают связь агентности в детском возрасте с социокультурными характеристиками речевых паттернов, принятых в семье [Fogle, King, 2013]. Применительно к системе образования анализируется воздействие тех или иных речевых приемов на обучаемого и использование учителями разных речевых паттернов при общении с обучающимися, различающимися уровнем академической успешности [Allington, 1980; Johnston, 2004]. В частности, выделены типы бесед с обучающимися, которые побуждают их к проявлению агентности, т.е. к активному действию, к самостоятельности и ответственности в поведении, к применению своих знаний и к принятию решений [Johnston, 2004. P. 10]. В студенческой агентности исследователи выделяют нравственную составляющую и связывают ее проявления с умением критически мыслить и задавать вопросы о том, что обучающемуся следует делать и имеет ли он на это право [Leadbeater, 2017]. При исследовании лингвистической репрезентации и реализации агентности выделены два уровня агентности: эго-утверждающая агентность как способность самовыражаться и влиять на других с помощью речи, обнаруживая свое присутствие и добиваясь

<sup>1</sup> <https://www.oecd.org/education/2030-project/teaching-and-learning/learning-student-agency/>

признания от других людей, и деятельно-устанавливающая агентность как способность к действию [Duranti, 2005].

В данном исследовании мы оцениваем возможность извлечь из текстов отзывов на массовые открытые онлайн-курсы (МООК) описания различных составляющих студенческой агентности и проявлений агентности как перехода от знаний к действиям, направленным на создание нового или улучшение существующего курса, электронной книги, плагина, веб-сайта, бизнеса и т.п. В организации этого исследования мы руководствуемся подходом П. Джонстона, который изучал проявления агентности в диалоге и в речи [Johnston, 2004], а также представлениями П. Джааскела с соавторами о студенческой агентности как о студенческом опыте, который проявляется в разных контекстах, в том числе включающих возможности и ресурсы для организации целенаправленного и осмысленного обучения [Jääskelä et al., 2020], и вслед за Л. Макколи и К. Кинг рассматриваем агентность как инициативу и способность совершать действия, приводящие к значимым изменениям в себе или в окружающей среде [McCauley, King, 2021].

Цель данного исследования состоит в изучении отзывов на МООК как возможной эмпирической базы для извлечения лингвистических конструкций, описывающих проявление агентности, и разработке инструмента учебной аналитики для извлечения из текстов отзывов на МООК описаний различных типов активностей, которые связаны с источниками студенческой агентности, и описаний внутренней (преодоление ограничений и изменение мышления) и внешней (создание нового и изменение существующего) трансформации.

В настоящее время автору неизвестны исследования, которые были бы направлены на разработку способов автоматического извлечения источников и составляющих студенческой агентности из текстов отзывов на МООК. Между тем разработка новых инструментов учебной аналитики на основе интеллектуального анализа образовательных данных способствует принятию обоснованных педагогических решений и облегчает организацию мониторинга мнений и настроений обучающихся, позволяет оценить связи между удовлетворенностью обучающихся, их академической успешностью и студенческой агентностью, а также выявлять факторы, ограничивающие проявление студенческой агентности и разрабатывать стратегии взаимодействия между тьюторами и обучающимися для реализации агентности [Jääskelä et al., 2020]. Сегодня возможности получения тьюторами обратной связи от обучающихся на платформах МООК весьма ограничены, тьюторы испытывают потребность в аналитических инструментах, которые позволили бы им проанализировать те опции, которые предоставляет тот или иной курс для стимулирования активности обучающихся: их личностного роста и перехода от накопления знаний к

конструктивному действию и самореализации. Таким образом, разработка новых инструментов учебной аналитики на основе эмпирических данных позволит получить всестороннее представление о проявлениях студенческой агентности и их взаимосвязи с академической успешностью, а также даст возможность тьюторам онлайн-курсов скорректировать контент MOOK с таким расчетом, чтобы он стимулировал проявления агентности у обучающихся в современных быстро меняющихся условиях жизни.

## **1. Обзор литературы**

Исследование языка студенческой агентности, в частности речевых паттернов, используемых учителями в целях поддержки студенческой агентности и применяемых для организации социальных взаимодействий типа «учитель — обучающийся» и «обучающийся — обучающийся», а также средств выражения студентами в текстах своих идей, проявления креативности и способности применять знания в реальной жизни — одна из наиболее актуальных задач компьютерной лингвистики в сфере анализа текстовых образовательных данных.

### **1.1. Исследования составляющих студенческой агентности и оценочные шкалы**

Рассмотрим некоторые модели и шкалы агентности, которые дают представление об агентности как о многомерном понятии. Исследователи предлагают разные варианты списка размерностей (источников) агентности и анализируют ее проявления в устной и письменной речи. М. Вон рассматривает студенческую агентность в трех измерениях: он выделяет диспозиционную, мотивационную и позиционную размерность [Vaughn, 2020]. Диспозиционная размерность студенческой агентности проявляется в способности индивидуумов действовать и преобразовывать среду. Мотивационная размерность связана с контролем действий, их корректировкой при осуществлении выбора и принятии решений. Позиционная размерность проявляется в социальном взаимодействии умением обсуждать и договариваться, отстаивать свою точку зрения. М. Вон приводит примеры побуждения учителем учащихся на уроке литературы к реализации всех трех размерностей агентности через постановку продуманных вопросов и организацию социального взаимодействия [Vaughn, 2020]. Чтобы проявить агентность в любой из трех ее размерностей, учащийся должен уметь сформулировать, представить и донести до окружающих свои идеи — а значит, проявления агентности неразрывно связаны с речевыми конструкциями, в которых осуществляются рассуждения обучающихся и структуризация диалога между учителем и обучающимся. Таким образом, выявление и использование учителями определенных речевых паттернов может способствовать проявлению, поддержке и развитию студенческой агентности.

М. Вон выделяет также контексты, в которых может находить свое выражение студенческая агентность: моральный (имеет ли право обучающийся на определенное действие и как оно отразится на окружающих), социальный (как обучающемуся строить отношения в обществе, как находить общий язык с незнакомыми людьми), экономический (какие есть возможности у обучающегося для влияния на экономику) и креативный (как воплотить свои инновационные идеи, способствующие развитию общества) [Vaughn, 2020].

Факторный анализ созданной финскими исследователями Шкалы агентности студентов университетов (*Agency of University Students, AUS*) позволил выделить индивидуальные, реляционные и контекстуальные источники агентности при изучении курса [Jääskelä et al., 2016]. Интерес и мотивация, самооффективность, убежденность в компетентности и активность участия рассматриваются как составляющие индивидуального источника агентности. Равные условия обучения для всех студентов, поддержка преподавателей, коллег и атмосфера доверия — это составляющие реляционного источника агентности. Наконец, возможность влиять на условия обучения и делать выбор авторы относят к составляющим контекстуального источника агентности.

В книге, посвященной способам усиления студенческой агентности, К. Зайсер с коллегами, в частности, приводят примеры фраз, связанных с составляющими студенческой агентности (табл. 1) [Zeiser, Scholz, Cirks, 2018].

Таблица 1. Тестовые фразы, соответствующие составляющим студенческой агентности [Zeiser, Scholz, Cirks, 2018]

Составляющие студенческой агентности	Тестовые фразы
Самооффективность	Описание возможности достижения важных целей
Устойчивость/неустойчивость интересов	Описание ситуации, в которой новые идеи и проекты отвлекают от актуальных
Устойчивость усилий	Описание стремления к завершению всего начатого
Контроль происходящих событий	Описание ситуации, позволяющей судить о восприятии обучающимся предсказуемости событий, о его способности понимать, что произойдет дальше
Ориентированность на мастерство	Описание причин действий, например, «Выполняя классную работу, так как нравится изучать новое»
Метакогнитивная саморегуляция	Описание поиска подтверждений правильного понимания изученного: обучающийся задает вопросы, ответы на которые подтверждают правильное понимание учебного материала
Саморегуляция обучения	Описание умений правильно распределять время, в частности обучающийся умеет выделять время на выполнение домашних заданий

Составляющие студенческой агентности	Тестовые фразы
Ориентированность на будущее	Описание ситуации, позволяющей судить об уверенности студента в полезности всего, что он изучает в классе, для его дальнейшего успеха

Итак, для формирования универсальных компетенций обучающихся и поддержки студенческой агентности необходимо ясное представление о контекстах проявления студенческой агентности и способах их обнаружения в речи обучающихся. Поскольку обратная связь от слушателей MOOK ограничена, отзывы на тот или иной курс могут рассматриваться как важный источник данных об удовлетворенности слушателей содержанием MOOK, о настроениях обучающихся и проявлениях студенческой агентности.

## 1.2. Разработка инструментов учебной аналитики на основе анализа текстов отзывов на MOOK

В последнее время растет число исследований, посвященных разработке инструментов учебной аналитики на основе анализа текстовых образовательных данных: от изучения мнений и настроений обучающихся по отзывам и комментариям на форумах и платформах MOOK [Kastrati et al., 2021] до оценки ментального здоровья [Nazari et al., 2023] и мониторинга агентности обучающихся с помощью методов машинного обучения. Анализ студенческой активности на основе опросника *Approaches and Study Skills Inventory for Students (ASSIST)* позволил исследовать проявления агентности обучающихся в процессе преодоления ограничений, возникающих в ходе достижения целей [Andre, 2021]. Перспективными направлениями применения инструментов учебной аналитики, в том числе с использованием искусственного интеллекта, являются исследования взаимосвязи студенческой агентности с академической успешностью и удовлетворенностью обучением и выявление факторов, ограничивающих проявление студенческой агентности [Heilala, 2022]. Учебная аналитика служит также для поиска ресурсов, способствующих поддержке студенческой агентности в вузах, и исследования возможностей применения результатов такого анализа на этапе педагогического планирования [Saarela et al., 2021].

Для извлечения закономерностей из текстовых образовательных данных используются различные способы обработки естественного языка: от простых статистических методов до методов тематического моделирования. Автоматическому анализу текстов комментариев предшествует процесс разметки, т.е. отнесения комментариев и отдельных предложений из них к тому или иному классу. Разметка датасета часто оказывается сложной и трудоемкой задачей, для решения которой используются либо подходы на основе краудсорсинга, либо средства с примени-

ем методов машинного обучения. И в первом и во втором случае неизбежны случаи ошибочной разметки, которые возникают из-за неоднозначности текста и отсутствия в нем признаков принадлежности предложений к какой-либо группе, из-за субъективности такой разметки или из-за несовершенства алгоритмов обработки естественного языка.

Методы тематического моделирования хорошо зарекомендовали себя при разработке инструментов учебной аналитики, поскольку они позволяют выделять группы текстов со схожей тематикой, описываемой с помощью извлеченных ключевых слов. По сути, эти ключевые слова представляют собой закономерности, характерные для семантически схожих текстов. Применение тематического моделирования на основе латентного размещения Дирихле (*Latent Dirichlet Allocation, LDA*) для анализа публикаций по тематике смешанного обучения позволило выделить такие основные темы, как результаты обучения, формирование оценки, факторы учителя, математическое и медицинское образование и т.п. [Yin, Yuan, 2022]. Такого рода исследования позволяют получить квинт-эссенцию содержания научных публикаций по различным направлениям в сфере образования, понять, чему исследователи уделяют наибольшее внимание, и выделить нерешенные проблемы. На основе тематической модели, реализованной в *python*-библиотеке *Mallet*, разработано персонализированное предложение учебного контента пациентам, страдающим диабетом [Kandula et al., 2011].

Анализ частотности N-грамм и построение тем на основе латентного размещения Дирихле применены к текстам рецензируемых статей по теме «Влияние пандемии COVID-19 на электронное обучение». Результаты тематического моделирования позволили выявить в статьях широкий круг тем, связанных с потребностями в обучении, социальным влиянием, организацией онлайн-обучения, в том числе в виртуальных классах, проблемами ментального здоровья, выявлением факторов удовлетворенности онлайн-обучением, особенностями онлайн-обучения по разным дисциплинам и т. д. [Gurcan, Dalveren, Derawi, 2022]. Сходные темы выявлены на основе тематического моделирования с помощью *python*-библиотеки *gensim* и в другом исследовании публикаций в журналах из сферы образования [Ekin et al., 2021]. Передовые образовательные технологии, связанные с персонализацией обучения, мобильным обучением, обучением с применением знаний из робототехники и т.п., извлечены и описаны с помощью структурного тематического моделирования из публикаций в журнале *Computers & Education* [Chen et al., 2020].

Исследование высказываний в социальных сетях по тематике образования помогает определить отношение общества к тем или иным методологиям и технологиям, используемым в образовании, исследовать причины недовольства и тревожностей, воз-



никающих в связи с внедрением онлайн-образования. Для изучения мнений об онлайн-образовании индийские исследователи извлекли твитты по ключевым словам данной тематики и данные гугл-форм, в которых обучающиеся описывали позитивные и негативные стороны онлайн-обучения, и проанализировали их с помощью *python*-библиотеки *TextBlob* для определения тональности текстов — позитивной, негативной или нейтральной, а также применили методы тематического моделирования: латентно-семантический анализ (*latent semantic analysis*, LSA) и LDA [Devi et al., 2022]. На основе тематического моделирования из твиттов удалось выделить темы, связанные с проблемами образования, социальными и этическими вопросами, применением технологий и обсуждением проблем с ментальным здоровьем [Alhazmi, 2022]. С помощью исследования тональности и извлечения тем из текстов в сфере образования удастся исключить влияние фейковых новостей и дезинформации в твитах на результаты анализа текстов твитов [Waheeb, Khan, Shang, 2022].

Средствами тематического моделирования исследовано содержание электронных писем обучающихся до и после пандемии COVID-19 [Tadeo, Yoo, 2022]. Выделение кластеров тем и их представление в виде сетевого графа с помощью инструмента *InfraNodus* позволило проследить увеличение числа сообщений, содержание которых связано с организацией и проведением тестов, составлением вопросов и оценкой знаний, а также выявить острую потребность учителей в поддержке.

Одно из активно разрабатываемых направлений исследований в образовании — оценивание знаний учителей и влияния этих знаний на методику преподавания и академическую успешность обучающихся на основе формальных подходов. Разработан способ оценивания знаний учителей путем изучения текстов развернутых ответов с помощью тематического моделирования [Corrig-Gencturk, Choi, Cohen, 2022].

Для выявления фокуса обсуждения в тех или иных онлайн-сообществах, принадлежащих к сфере образования, — в блогах, на форумах — и для облегчения поиска нужного сообщества предложен подход, основанный на извлечении значимых термов тем и построении сети ассоциаций между терминами [Li et al., 2015]. Извлечение тем из названий и описаний открытых образовательных ресурсов позволяет разрабатывать эффективные системы рекомендаций в отношении MOOK и образовательного контента. С помощью вероятностной тематической модели LDA извлечены темы из названий образовательных ресурсов на *Youtube* и сопоставлены со списком навыков в области анализа данных. Результаты тематического моделирования предлагается использовать для построения индивидуальной образовательной траектории обучающихся [Molavi, Tavakoli, Kismihók, 2020].

Специальные инструменты учебной аналитики разрабатываются для сопровождения онлайн-курсов и автоматической организации поддержки обучающихся на форумах [Wong, Wong, Hindle, 2019]. Благодаря применению иерархического процесса Дирихле и базовой тематической модели LDA стало возможным сопоставление контента MOOK с сообщениями на форумах. Тематическое моделирование применяется для организации обратной связи с обучающимися, оценивания их знаний и составления индивидуального учебного плана [He et al., 2017]. Исследователи анализируют также возможности анализа текстов студенческих рефлексивных журналов для разработки критериев оценивания и инструментов учебной аналитики на основе тематического моделирования [Chen et al., 2016].

С учетом современных достижений в разработке инструментов учебной аналитики на основе тематического моделирования и частотного анализа текстовых данных, в данном исследовании мы предлагаем методологию для извлечения составляющих студенческой агентности на основе латентного размещения Дирихле и метода тематического моделирования *Bertopic*. Она позволяет выделить кластеры предложений по тематикам и получить их описания с помощью наиболее релевантных слов. Там, где из-за малочисленности данных не удалось применить методы тематического моделирования, для получения описаний составляющих студенческой агентности и описаний внутренней и внешней трансформации использовался частотный анализ.

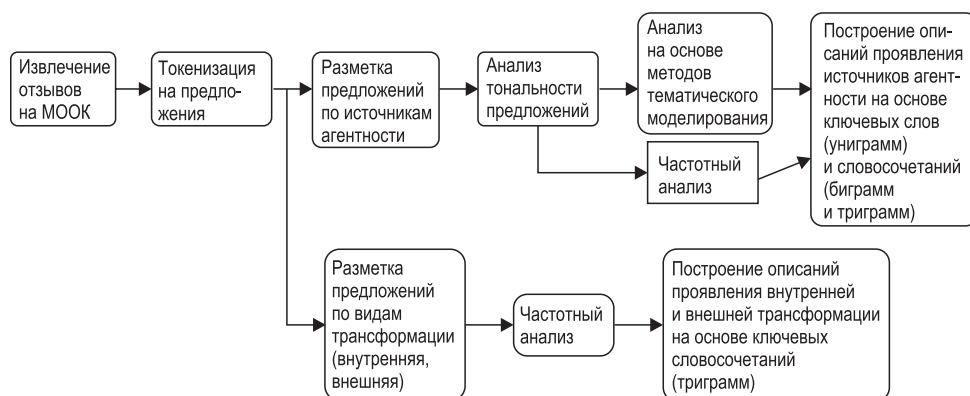
## 2. Методология исследования

Для исследования проявления источников студенческой агентности и проявления внутренней и внешней трансформации в текстах отзывов на MOOK в данной работе применяются такие методы обработки естественного языка, как анализ тональности предложений, частотный анализ словосочетаний, тематическое моделирование для выделения тем и ключевых слов, а также их сочетаний — биграмм и триграмм.

Для исследования проявления источников студенческой агентности в текстах отзывов на MOOK с платформы *Udemy* извлечены 3445 англоязычных отзывов о наиболее популярных курсах по математике, а для исследования описаний внутренней и внешней трансформации — 1787 отзывов на практикоориентированные MOOK и MOOK по предпринимательству. Основные этапы исследования представлены на рис. 1.

Извлечение отзывов с платформы *Udemy* реализовано на основе возможностей *python*-библиотеки *scrapy*. С помощью *python*-библиотеки *nltk* проведена токенизация текстов отзывов на предложения, после чего вручную отобраны предложения, относящиеся к проявлению различных источников агентности по AUS-шкале и выполнена разметка датасета с отнесением каждо-

Рис. 1. Основные этапы извлечения описаний проявления источников студенческой агентности и описаний внутренней и внешней трансформации



го предложения к одному из источников студенческой агентности. После токенизации и разметки для предложений по классам источников студенческой агентности выполнен анализ тональности с помощью метода VADER, основанного на словарях и правилах и реализованного в *python*-библиотеке *vaderSentiment*. В ходе анализа тональности предложения по каждому источнику агентности распределены на три класса: позитивные, негативные и нейтральные. Далее для каждого источника агентности и каждой тональности требовалось выделить ключевые слова и сочетания слов, описывающие причину позитивного, негативного или нейтрального отношения. Целям такой группировки предложений отвечают методы тематического моделирования и частотный анализ предложений. В данном исследовании для построения более полных описаний на основе групп предложений с общей тематикой использованы два метода тематического моделирования — метод латентного размещения Дирихле, реализованный в *python*-библиотеке *gensim*, и метод тематического моделирования *BERTopic*, реализованный в одноименной *python*-библиотеке. Латентное размещение Дирихле представляет собой вероятностную тематическую модель, позволяющую извлекать темы с учетом распределения тем в корпусе и основанную на вычислении вероятности появления пары «документ — слово» [Zhang, Miao, Gao, 2013]. *BERTopic* дает возможность строить плотные кластеры ключевых слов выделенных тем на основе архитектуры «трансформер» и меры *c-TF-IDF* [Grootendorst, 2022]. В данном исследовании применялись оба метода для обеспечения разнообразия извлеченных закономерностей, используемых для описания составляющих студенческой агентности.

Качество результатов тематического моделирования оценивалось с помощью показателя когерентности. Когерентность осно-

вана на вычислении вероятности совместного появления слов в теме. Чем выше значение когерентности, тем легче пользователю интерпретировать результаты модели [Коршунов, Гомзин, 2012].

В силу малочисленности предложений, содержащих описание проявления контекстуального источника агентности и описания внутренней и внешней трансформации, для их исследования применен частотный анализ. Из предложений, содержащих описания трансформации, выделены ключевые триграммы, и результаты частотного анализа визуализированы на основе облаков триграмм с помощью *python*-библиотеки *wordcloud*.

**3. Результаты исследования**  
**3.1. Исследование проявления источников студенческой агентности в текстах отзывов на MOOK**

На рис. 2 представлен фрагмент собранного датасета с отзывами на MOOK.

Рис. 2. Фрагмент англоязычного датасета с отзывами на MOOK (после перевода отзывов на русский язык)

Отзывы
Курс наполнен большим количеством информации, превосходящей мои ожидания; особенно раздел научных обозначений, который оказывается очень полезным. Теперь я чувствую себя более уверенно в математике. Инструктор говорит с явным акцентом, очень легко слушать. Большое спасибо, Криста Кинг
Этот математический курс подготовил меня к алгебре 1. Содержание хорошо составлено, а концепции преподаются таким образом, чтобы мой мозг легко их усваивал. Учебники и заметки потрясающие, и я никогда не чувствовал себя настолько уверенно в основах, прежде чем пройти этот курс. Благодаря последовательности я смог пройти курс чуть менее чем за два месяца, хорошо разбираясь в каждом разделе. Спасибо, мисс Кинг, за организацию этих курсов!
Отличный учитель, четкие объяснения, именно то, что я искал для освежения знаний, но также отлично подходит для людей, которые хотят изучать предмет впервые. Спасибо, Криста!
Криста — хороший учитель — удивительно, как многому мы научились, пройдя один этот курс. Она также обладает очень позитивной атмосферой, которая вовлекает вас в образ мыслей, способный сделать все возможное. Она действительно растягивает ваш ум и заставляет работать от простых вопросов к сложным. Мы на других курсах Кристи Кинг!
Абсолютно гениально для курса повышения квалификации. Теперь, когда у меня есть четкое понимание основ, я чувствую себя готовым заняться более сложной математикой

Вслед за [Jääskelä et al., 2016] мы выделяем три класса источников агентности: индивидуальный (класс 1), реляционный (класс 2) и контекстуальный (класс 3). К индивидуальному источнику относятся интерес, самоэффективность, убежденность в компетентности и активное участие, к реляционному — одинаковые условия для всех в обучении, поддержка со стороны учителей и сверстников, атмосфера доверия, к контекстуальному источнику (источнику участия) — возможность влияния и возможность выбора.

В результате токенизации отзывов выделены 308 предложений с указанием их принадлежности к одному из трех классов источников студенческой агентности (рис. 3).

**Рис. 3. Фрагмент датасета с англоязычными предложениями после разметки по классам — источникам студенческой агентности (после перевода предложений на русский язык)**

Предложения	Класс
Мне очень понравилось проходить этот курс, и после его прохождения я значительно улучшил свою математическую зрелость и способности	1
Кроме того, она предоставляет все необходимые детали для обучения	2
Благодаря этому асинхронному видеоформату я узнал больше, чем в более традиционной среде	1
Спустя два месяца я решаю эти задачи и чувствую себя уверенно, приступая к линейной алгебре и исчислению	1
Я даже начал легко применять то, что узнал из курса базовой математики, в своей повседневной жизни	1

Примеры распределения предложений из исследуемого датасета по классам в соответствии с источниками студенческой агентности представлены в табл. 2, причем к тому или иному классу отнесены как предложения с описанием определенных проявлений источников студенческой агентности, так и с описанием их отсутствия.

**Таблица 2. Примеры предложений из датасета с разметкой по источникам студенческой агентности**

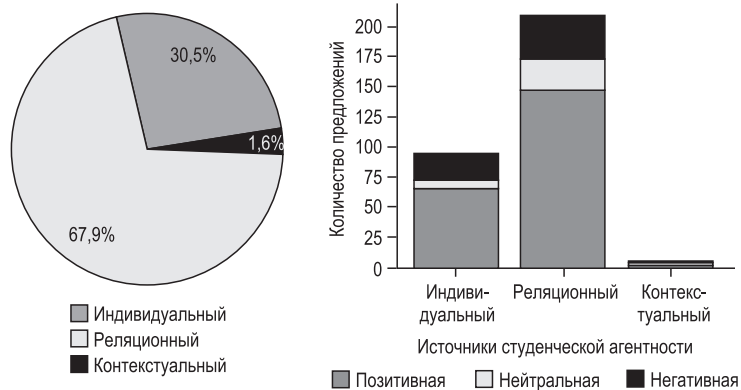
Пример предложения из датасета	Источник студенческой агентности на основе разметки вручную
Я значительно улучшил свою математическую зрелость и способности после прохождения курса	Класс 1, индивидуальные источники — проявление самоэффективности
Это разочаровывает и часто деморализует процесс обучения	Класс 1, индивидуальные источники — отсутствие самоэффективности
Кроме того, она предоставляет все необходимые детали для обучения	Класс 2, реляционные источники — проявление поддержки учителя
Я застрял, потому что даже объяснение ответов сбивает с толку	Класс 2, реляционные источники — отсутствие поддержки со стороны тьютора
Лучший выбор, который я мог сделать в своей жизни, — покупка ее курсов	Класс 3, контекстуальные источники — проявление возможности выбора
В комментариях к этому есть много жалоб, которые были признаны и согласованы с разработчиком, но, к сожалению, курс остается неизменным	Класс 3, контекстуальные источники — отсутствие возможности влияния, изменения ситуации

Как видно из примеров, представленных в табл. 2, некоторые фразы подтверждают составляющую агентности, а некоторые свидетельствуют о ее отсутствии, поэтому интерес представля-

ет исследование тональности (эмоциональной окраски) фраз, на основании которых выявляется студенческая агентность, и выделение тем и ключевых слов, характерных для тех или иных размерностей студенческой агентности, с учетом тональности. Для классификации выделенных предложений по видам тональности — позитивные, негативные или нейтральные — применялась *python*-библиотеки *VaderSentiment*.

На рис. 4 слева приведены результаты распределения предложений из отзывов на MOOK по источникам студенческой агентности, а справа — распределение предложений по видам тональности с учетом принадлежности к тому или иному классу источников агентности. Из рисунка видно, что в проанализированном массиве отзывов преобладают предложения, отражающие реляционные (67,9%) и индивидуальные (30,5%) источники агентности, а предложения, в которых пользователи описывают свои возможности влиять на контент MOOK или выбор MOOK, чрезвычайно малочисленны (менее 2% от общего числа предложений в отзывах). Эти предложения представляют контекстуальный источник агентности. При этом в исследуемом массиве преобладают позитивные по тональности предложения, содержащие описание проявлений индивидуального и реляционного источников агентности.

Рис. 4. Распределение предложений из отзывов на MOOK по источникам агентности (слева) и по тональности с учетом источников агентности (справа)



Для выявления составляющих студенческой агентности в отзывах на MOOK с учетом их тональности проведено тематическое моделирование на основе LDA с помощью *python*-библиотеки *gensim*. Качество выделенных тем и ключевых слов оценивалось по показателю когерентности. Оптимальное число тем определялось с учетом значений когерентности, а затем из тем выделялись словосочетания — биграммы.

Таблица 3. **Результаты выделения словосочетаний из ключевых слов на основе латентного размещения Дирихле**

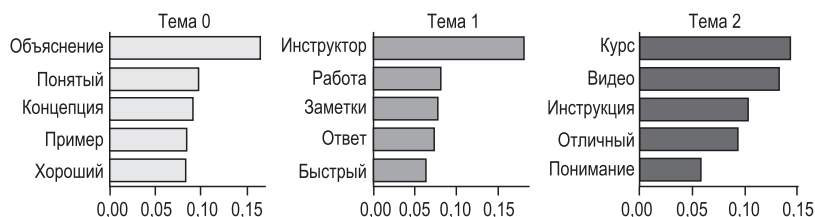
Источник студенческой агентности и тональность	Словосочетания из ключевых слов выделенных на основе LDA тем для описания составляющих студенческой агентности	Когерентность
Индивидуальный источник с позитивной тональностью	Вовлекающий курс, безболезненное самообучение, мотивирующая лекция, наращивание потенциала, совершенное решение, улучшенный навык, легкое понимание, помогающее обучение	0,5355
Индивидуальный источник с негативной тональностью	Пробелы (в) знаниях, необходимый навык, мысли (о) карьере	0,5595
Реляционный источник с позитивной тональностью	Хороший учитель, отличная практика, много примеров, понятное объяснение, шаг (за) шагом, доступность информации, быстрый ответ, легко следовать, хорошо структурирован, поступающие вопросы, понятный материал	0,4931
Реляционный источник с негативной тональностью	Сложная викторина, неправильный тест, асинхронное видео, необходимость объяснения	0,5991

Результаты выделения биграмм для разных размерностей студенческой агентности представлены в табл. 3. Словосочетания из ключевых слов, выделенные на основе латентного размещения Дирихле, дают представление о тех характеристиках MOOK, которые вызвали у обучающихся позитивное или негативное отношение при описании индивидуального и реляционного источника студенческой агентности. Например, позитивное отношение обучающихся при описании индивидуального источника агентности вызывает наличие мотивирующих лекций, возможностей личностного роста через наращивание собственного потенциала, улучшение навыков.

Результаты построения с помощью *python*-библиотеки *BERTopic* плотных кластеров ключевых слов для выделенных тем из предложений с позитивной тональностью, связанных с реляционным источником студенческой агентности, показаны на рис. 5. Реляционный источник студенческой агентности составляет поддержка инструктора MOOK, и она проявляется в основном в предоставлении обучающимся хорошо разработанных учебных материалов (видео, заметки, инструкции, примеры, концепции и т.п.), а также в обеспечении их быстрыми ответами от инструктора.

Ввиду малочисленности в датасете предложений, связанных с отдельными источниками студенческой агентности определенной тональности, извлечь темы и ключевые слова средствами тематического моделирования не удалось, поэтому дополнительно реализован частотный анализ N-грамм (униграмм, биграмм и триграмм).

Рис. 5. Пример извлечения тем и их ключевых слов из позитивных предложений, связанных с реляционным источником студенческой агентности



Результаты частотного анализа униграмм, биграмм и триграмм, представленные в табл. 4, позволили выявить такие составляющие студенческой агентности, как проявление самооэффективности в виде повышения математической зрелости, совершенствования навыков, понимания и появления чувства уверенности при решении задач; поддержка от инструкторов в виде хорошо структурированного и спланированного контента и быстрых ответов; ориентированность на будущее в виде появления знаний, помогающих в карьере; способность свободно принимать решения, в том числе выбирать другие курсы/лекции и т.п.

Таблица 4. Топ униграмм, биграмм, триграмм и их частоты для описания проявления составляющих студенческой агентности

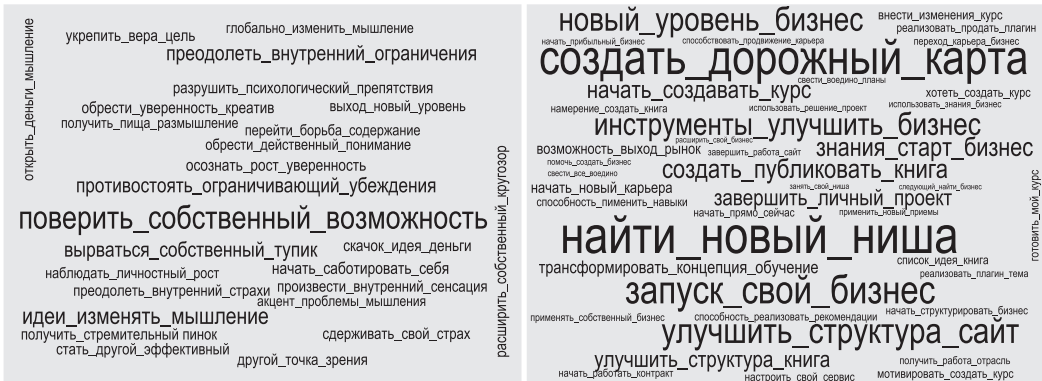
Униграммы	Биграммы	Триграммы
Анализ предложений с позитивной тональностью, относящихся к индивидуальному источнику студенческой агентности		
Помогать/помогающий — 24,54%, понять/понимание — 23,08%, обучение/обучаться — 23,08%, улучшать/увеличивать — 13,85%, наслаждаться — 12,31%	Улучшить понимание — 11,69%, чувствовать уверенность — 9,74%, решать задачи — 7,79%, помогать карьере — 6,49%	Улучшить математическую(ие, ое) зрелость/навыки/понимание — 15,38%, действительно наслаждаться обучением — 8,14%, быть/стать способным делать/решать — 7,24%, определено/экстремально помогло — 11,69%
Анализ предложений с позитивной тональностью, относящихся к реляционному источнику студенческой агентности		
Материал/видео/концепция — 21,18%, объяснение — 13,79%, понятное (понимание)/легкое — 18,22%, вопрос — 8,87%	Понятное объяснение — 8,49%, легкое понимание — 7,08%, хорошо организованный — 6,6%, хороший баланс — 6,13%, быстрый ответ — 5,66%, легко следовать — 5,19%	Курс/лекция хорошо организован(а)/спланирован(а) — 18,22%, заметки к видео/секциям — 16,03%
Анализ предложений с позитивной тональностью, относящихся к концептуальному источнику студенческой агентности		
Выбор — 5,26%, жизнь — 5,26%, покупка — 5,26%	Обучают похоже — 5%, полностью бесплатно — 5%, академия Хана — 5%	Сделан лучший выбор — 5,56%, свободно принять решение — 5,56%



### 3.2. Исследование проявления внутренней и внешней трансформации в текстах отзывов на MOOK

Следуя определению агентности как инициативы и способности действовать таким образом, чтобы произвести значимые изменения в себе или окружающей среде [McCauley, King, 2021], будем называть изменения в себе внутренней трансформацией, а изменения, которые создает актор в окружающей среде, — внешней. При анализе текстов 1787 англоязычных отзывов на практикоориентированные MOOK и MOOK по предпринимательству выделены проявления внутренней трансформации как перехода от внутренней борьбы (страхов, неуверенности, сложностей восприятия контента MOOK) к пониманию цели и внешней трансформации в виде создания нового или изменения структуры существующего продукта, стартапа или бизнеса через перестройку мышления. Пример извлеченных из отзывов предложений с описанием внутренней трансформации: «Это действительно помогло разрушить психологические препятствия, с которыми сталкиваются многие предприниматели при попытке начать свой собственный бизнес и создать дорожную карту»; «Это хороший базовый курс, который помог мне противостоять некоторым ограничивающим убеждениям о деньгах, о себе и так далее и узнать все, что нужно, чтобы стать предпринимателем и перейти к созданию собственного бизнеса»; «Я перешел от борьбы с содержанием и путем к точному пониманию того, что я хочу сделать и как это сделать». Примеры предложений с описанием внешней трансформации: «Я сам создал несколько бизнесов и думал, что опыта достаточно, чтобы стать коучем, но пройденный курс фактически структурировал мои мысли, позволил создать четкую интеллект-карту для изменения моего бизнеса»; «Я использовал рассмотренные техники для улучшения своего бизнеса и нашел свою нишу на рынке»; «Этот курс позволил построить четкую структуру и вывести мой бизнес на новый уровень». После токенизации из текстов отзывов на MOOK выделены 108 предложений, содержащих описание внутренней и внешней трансформации. Далее из предложений выделены триграммы, содержащие описание внутренней (34 триграммы) и внешней (74 триграммы) трансформации. Для исследования этих триграмм реализован частотный анализ с визуализацией результатов анализа в виде облаков триграмм, представленных на рис. 6. Как видно из рис. 6, в описаниях внутренних трансформаций упоминаются изменения мышления, преодоление страхов и внутренних ограничений, появление веры в собственные возможности и т.п., а описания внешней трансформации содержат информацию о переходе от знаний к действиям, а также о выходе на новый уровень понимания проблем и запуске своего бизнеса, о поиске новой ниши, об улучшении существующего продукта, например сайта, книги.

Рис. 6. Визуализация частотности триграмм для описания внутренней и внешней трансформации



#### 4. Обсуждение результатов и направления дальнейших исследований

Составляющими студенческой агентности являются самоэффективность, мотивация, рост уверенности в возможности применения полученных знаний и навыков для достижений целей, способствующих благополучию в обществе [Kim, 2022]. В данном исследовании на основании анализа текстов отзывов на MOOK удалось выделить описания лишь некоторых составляющих понятия «студенческая агентность». В частности, на основе частотного анализа и инструментов тематического моделирования из текстов выделены описания, относящиеся к проявлению различных источников студенческой агентности, а также описания, свидетельствующие о внутренней (изменение образа мышления) трансформации обучающихся и внешней трансформации (создание нового или изменение структуры существующего продукта, стартапа или бизнеса через изменение мышления).

Выделение тем и ключевых слов с помощью методов тематического моделирования с учетом тональности и результатов частотного анализа позволили понять, какие характеристики или ресурсы онлайн-курса вызвали у пользователей позитивное, а какие — негативное отношение к тем или иным источникам студенческой агентности, и осуществить мониторинг внутренних и внешних трансформаций обучающихся как основных составляющих студенческой агентности.

Перечислим некоторые трудности и ограничения, связанные с формированием датасета.

1. Существенную трудность составляет подбор эмпирического материала, соответствующего исследуемой тематике. Перед формированием датасетов мы ознакомились с отзывами обучающихся на разных платформах, таких как *Udemy*, *Coursera*, *EdX* и *Stepik*, но только на платформе *Udemy* обнаружили отзывы, наиболее полно соответствующие исследуемой тематике. С плат-

формы *Udemy* извлечены 5232 англоязычных отзыва, но менее 10% предложений из этих отзывов содержали описания, связанные с источниками агентности и трансформациями. В комментариях преобладали общие фразы, например «прекрасный курс», «да», «курс понравился».

2. Разметка текста — как вручную, так и с помощью математических методов — сопряжена с трудностями из-за неоднозначности критериев отнесения комментария или отдельного предложения к тому или иному классу проявления студенческой агентности.

3. Процесс отбора предложений, связанных с исследуемой предметной областью, достаточно трудоемкий.

4. Число отзывов на MOOK ограничено, и для формирования датасета из них нужно выбрать только отзывы с определенной спецификой, так что эти датасеты оказываются весьма скромного размера. Тем не менее тексты, находящиеся в открытом доступе в интернете, представляют несомненный интерес для исследователей, поскольку в них пользователи описывают свои мысли, чувства и отношение к какой-либо проблеме, и эти описания дают больше информации, чем результаты обследования обучающихся с помощью опросников с ограниченным количеством вопросов и вариантов ответа [Copur-Gencturk, Choi, Cohen, 2022].

5. Исследование проявлений отдельных составляющих источников студенческой агентности в комментариях на MOOK затруднено, поскольку такие проявления не существуют в явном виде. Например, о наличии поддержки тьютора как проявления одной из составляющих реляционного источника студенческой агентности судят по предоставлению тьютором учебных материалов, дополнительных пояснений, заметок, а также по структурированности онлайн-курса. Проявлений такой составляющей реляционного источника студенческой агентности, как взаимодействие со сверстниками или другими слушателями онлайн-курсов, в комментариях на MOOK в данном исследовании не выявлено. Отзывы, свидетельствующие о наличии контекстуальных источников агентности, обнаружены в незначительных количествах, в основном их малочисленность обусловлена отсутствием у пользователей возможности повлиять на контент курса и устранить неточности и опечатки или выбрать другой MOOK, например бесплатный онлайн-курс с похожим контентом. Отзывы на MOOK приходится считать весьма слабой эмпирической базой для исследования понятия «агентность» в полном объеме, включающем трансформирующую агентность как действия, приводящие к изменению структур [Сорокин, 2021]. Анализ отзывов на MOOK показывает, что обучающиеся редко описывают результаты обучения на MOOK в виде появления новых идей или намерения к действию, еще реже встречается описание создания какого-либо продукта или услуги даже в отзывах на практикоориентированные MOOK, цель

которых состоит в разработке собственного курса MOOK, электронной книги, программного продукта, бизнеса и т.п.

В дальнейшем для расширения датасета автор планирует обратиться к комментариям на форумах MOOK, где можно ожидать проявления в текстах такой составляющей реляционного источника студенческой агентности, как описание взаимодействия и взаимной поддержки между слушателями онлайн-курсов. Предложенная методология может быть применена для исследования проявления студенческой агентности в русскоязычных отзывах на MOOK с подбором необходимых *python*-библиотек для анализа текстовых данных.

Вниманию исследователей можно предложить целый ряд нерешенных и фрагментарно рассмотренных вопросов, связанных с изучением проявления студенческой агентности на основе анализа текстов отзывов или комментариев обучающихся в социальных сетях и на платформах MOOK посредством применения методов обработки естественного языка и компьютерной лингвистики: выявление речевых конструкций учителей, направленных на стимулирование и поддержку проявления студенческой агентности; автоматическое выделение из текстов различных составляющих студенческой агентности; разработка формальных подходов для оценивания студенческой агентности и ее взаимосвязи с удовлетворенностью MOOK и академической успешностью обучающихся.

**5. Заключение** Агентность является сложным многомерным понятием. Для исследования, в частности, студенческой агентности посредством описания ее размерностей и составляющих на основе извлеченных из текстов речевых паттернов требуется разрабатывать способы формирования специализированных датасетов и новые методы анализа текстов, речевых паттернов и особенностей языка. Данную работу можно рассматривать как стартовое исследование возможностей автоматического извлечения из текстов отзывов на MOOK закономерностей, позволяющих описывать некоторые составляющие студенческой агентности. В дальнейшем на основании таких закономерностей могут быть выработаны критерии для оценивания уровня проявления студенческой агентности без необходимости проведения дополнительных опросов, что особенно важно при разработке инструментов учебной аналитики MOOK.

С помощью методов тематического моделирования и частотного анализа предложений из текстов отзывов на MOOK удалось извлечь ключевые слова и их сочетания для описания проявлений индивидуального источника студенческой агентности в виде самооценки, усиления чувства уверенности при решении задач и мотивации; проявлений реляционного источника в виде поддержки со стороны тьютора с помощью быстрых ответов, хо-

рошо структурированного учебного материала и его сопровождения; проявлений контекстуального источника в виде возможности принимать решение при выборе онлайн-курсов. Из этих текстов удалось также извлечь предложения, позволяющие описать внутреннюю трансформацию обучающихся, которая состоит в преодолении страхов и внутренних ограничений для достижения целей, и внешнюю трансформацию, проявляющуюся в переходе от знаний к созиданию.

Для дальнейшего изучения студенческой агентности, ее размерностей и составляющих требуются более значительные объемы эмпирического материала. При этом полученные в данной работе результаты могут быть легко перенесены и использованы при анализе датасетов больших данных, в том числе и для исследования комментариев на русском языке с применением соответствующих *python*-библиотек для обработки естественного языка.

## Литература

1. Коршунов А., Гомзин А. (2012) Тематическое моделирование текстов на естественном языке. *Труды Института системного программирования РАН*, т. 23. <https://doi.org/10.15514/ISPRAS-2012-23-13>
2. Сорокин П.С. (2021) «Трансформирующая агентность» как предмет социологического анализа: современные дискуссии и роль образования. *RUDN Journal of Sociology*, vol. 21, no 1, pp. 124–138. <https://doi.org/10.22363/2313-2272-2021-21-1-124-138>
3. Сорокин П.С., Зыкова А.В. (2021) «Трансформирующая агентность» как предмет исследований и разработок в XXI веке: обзор и интерпретация международного опыта. *Мониторинг общественного мнения: экономические и социальные перемены*, № 5, сс. 216–241. <https://doi.org/10.14515/monitoring.2021.5.1858>
4. Ahearn L.M. (2001) Language and Agency. *Annual Review of Anthropology*, vol. 30, no 1, pp. 109–137. <https://doi.org/10.1146/annurev.anthro.30.1.109>
5. Alhazmi H. (2022) Detection of Students' Problems in Distance Education Using Topic Modeling and Machine Learning. *Future Internet*, vol. 14, no 6, Article no 170. <https://doi.org/10.3390/fi14060170>
6. Allington R.L. (1980) Teacher Interruption Behaviors during Primary-Grade Oral Reading. *Journal of Educational Psychology*, vol. 72, no 3, pp. 371–377. <https://doi.org/10.1037/0022-0663.72.3.371>
7. Andre J. (2021) *Can Learning Analytics Increase Agency and Transform Digital Learning?* Paper presented at the conference “Digital Transformation and Higher Education: When Challenges are Opportunities” (September, 2021, Hanoi, Vietnam).
8. Bandura A. (1999) Social Cognitive Theory: An Agentic Perspective. *Asian Journal of Social Psychology*, vol. 2, no 1, pp. 21–41. <https://doi.org/10.1111/1467-839x.00024>
9. Bandura A. (1986) *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice Hall.
10. Chen Y., Yu B., Zhang X., Yu Y. (2016) Topic Modeling for Evaluating Students' Reflective Writing. Proceedings of the *Sixth International Conference on Learning Analytics & Knowledge — LAK '16 (Edinburgh, UK, 2016, April 25–29)*. <http://dx.doi.org/10.1145/2883851.2883951>
11. Chen X., Zou D., Cheng G., Xie H. (2020) Detecting Latent Topics and Trends in Educational Technologies over Four Decades Using Structural Topic Mode-

- ling: A Retrospective of All Volumes of Computers & Education. *Computers & Education*, vol. 151, July, Article no 103855. <https://doi.org/10.1016/j.compedu.2020.103855>
12. Copur-Gencturk Y., Choi H., Cohen A. (2022) Investigating Teachers' Understanding through Topic Modeling: A Promising Approach to Studying Teachers' Knowledge. *Journal of Mathematics Teacher Education*, vol. 26, pp. 281–302. <https://doi.org/10.1007/s10857-021-09529-w>
  13. Devi S., Dhavale C., Moharkar L., Khanvilkar S. (2022) Impact of Online Education and Sentiment Analysis from Twitter Data Using Topic Modeling Algorithms. *International Journal of Applied Sciences and Smart Technologies*, vol. 4, no 1, pp. 21–34. <https://doi.org/10.24071/ijasst.v4i1.4637>
  14. Dewey J. (1922) *Human Nature and Conduct: An Introduction to Social Psychology*. New York, NY: Henry Holt and Co. <https://doi.org/10.1037/14663-000>
  15. Duranti A. (2005) Agency in Language. *A Companion to Linguistic Anthropology* (ed. A. Duranti), Malden, MA: Blackwell, pp. 449–473. <https://doi.org/10.1002/9780470996522.ch20>
  16. Ekin C.Ç., Çakici M., Şener E., Türker S., Altanlar S. (2021) Research Trends Analysis in Educational Journal Publications on COVID-19 Using Descriptive and Text Mining Methods: Preliminary Analysis. *European Journal of Science and Technology*, iss. 29, pp. 432–437. <https://doi.org/10.31590/ejosat.1036109>
  17. Fogle L.W., King K.A. (2013) Child Agency and Language Policy in Transnational Families. *Issues in Applied Linguistics*, vol. 19. <https://doi.org/10.5070/I4190005288>
  18. Grootendorst M. (2022) *BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure*. <https://doi.org/10.48550/arXiv.2203.05794>
  19. Gurcan F., Dalveren G.G., Derawi M. (2022) COVID-19 and e-Learning: An Exploratory Analysis of Research Topics and Interests in e-Learning during the Pandemic. *IEEE Access*, vol. 10, pp. 123349–123357. <https://doi.org/10.1109/access.2022.3224034>
  20. He J., Rubinstein B.I.P., Bailey J., Zhang R., Milligan S. (2017) Topic Response: A Marriage of Topic Modelling and Rasch Modelling for Automatic Measurement in MOOCs. *ArXiv: 607.08720v2*.
  21. Heilala V. (2022) *Learning Analytics with Learning and Analytics: Advancing Student Agency Analytics* (PhD Thesis), Jyväskylä: University of Jyväskylä.
  22. Jääskelä P., Heilala V., Kärkkäinen T., Häkkinen P. (2020) Student Agency Analytics: Learning Analytics as a Tool for Analysing Student Agency in Higher Education. *Behaviour & Information Technology*, vol. 40, no 8, pp. 790–808. <https://doi.org/10.1080/0144929x.2020.1725130>
  23. Jääskelä P., Poikkeus A., Vasalampi K., Valleala U.M., Rasku-Puttonen H. (2016) Assessing Agency of University Students: Validation of the AUS Scale. *Studies in Higher Education*, vol. 42, no 11, pp. 2061–2079. <https://doi.org/10.1080/03075079.2015.1130693>
  24. Johnston P.H. (2004) *Choice Words: How Our Language Affects Children's Learning*. Portland, ME: Stenhouse.
  25. Kandula S., Curtis D., Hill B., Zeng-Treitler Q. (2011) Use of Topic Modeling for Recommending Relevant Education Material to Diabetic Patients. *AMIA Annual Symposium Proceedings Archive*, pp. 674–682.
  26. Kastrati Z., Dalipi F., Imran A.S., Nuci K.P., Wani M.A. (2021) Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study. *Applied Sciences*, vol. 11, no 9, Article no 3986. <https://doi.org/10.3390/app11093986>
  27. Kim J. (2022) Analyzing 'Student Agency' Embedded in the Discourse of Future Education: (Re)interpretation of the OECD Education 2030 Project. *The Journal of Curriculum Studies*, vol. 40, no 2, pp. 181–202. <https://doi.org/10.15708/kscs.40.2.8>

28. Leadbeater C. (2017) *Student Agency: Learning to Make a Difference*. East Melbourne, VIC: The Centre for Strategic Education.
29. Li Y., Zheng Y., Bao H., Liu Y. (2015) Towards Better Understanding of Hot Topics in Online Learning Communities. *Smart Learning Environments*, vol. 2, no 1, Article no 12. <https://doi.org/10.1186/s40561-015-0019-6>
30. Littleton K., Taylor S., Eteläpelto A. (2011) Special Issue Introduction: Creativity and Creative Work in Contemporary Working Contexts. *Vocations and Learning*, vol. 5, no 1, pp. 1–4. <https://doi.org/10.1007/s12186-011-9067-4>
31. Marni C., Passini S. (2018) Development and Validation of an Enlarged Version of the Student Agentic Engagement Scale. *Journal of Psychoeducational Assessment*, vol. 37, no 4, pp. 450–463. <https://doi.org/10.1177/0734282918757849>
32. Matos L., Reeve J., Herrera D., Claux M. (2018) Students' Agentic Engagement Predicts Longitudinal Increases in Perceived Autonomy-Supportive Teaching: The Squeaky Wheel Gets the Grease. *The Journal of Experimental Education*, vol. 86, no 4, pp. 579–596. <https://doi.org/10.1080/00220973.2018.1448746>
33. McCauley L., King K. (2021) *Human-Centered Learning and Student Agency: "Think Big, Start Small and Act Fast"*. Available at: <https://knowledgeworks.org/resources/human-centered-learning-student-agency/> (accessed 20 February 2024).
34. Molavi M., Tavakoli M., Kismihók G. (2020) Extracting Topics from Open Educational Resources. *Addressing Global Challenges and Quality Education. EC-TEL 2020. Lecture Notes in Computer Science* (eds C. Alario-Hoyos, M.J. Rodríguez-Triana, M. Scheffel, I. Arnedillo-Sánchez, S.M. Dennerlein), vol. 12315. Cham: Springer, pp. 455–460. [https://doi.org/10.1007/978-3-030-57717-9\\_44](https://doi.org/10.1007/978-3-030-57717-9_44)
35. Nazari A., Hossennia M., Garmaroudi G., Torkian S. (2023) *Social Media and Mental Health in Students: A Cross-Sectional Study during the COVID-19 Pandemic*. <https://doi.org/10.21203/rs.3.rs-2535464/v1>
36. Nogueira F.D. (2017) Reassembling the Social: An Introduction to Actor-Network Theory, Oxford university press, 2005. *Formação (Online)*, vol. 1, no 25, pp. 229–233. <https://doi.org/10.33081/formacao.v1i25.5251>
37. Saarela M., Heilala V., Jaaskela P., Rantakaulio A., Karkkainen T. (2021) Explainable Student Agency Analytics. *IEEE Access*, vol. 9, pp. 137444–137459. <https://doi.org/10.1109/access.2021.3116664>
38. Tadeo D.J., Yoo J. (2022) Topic Modeling of the Student Emails Sent before and during the Birth of COVID-19 in Physics and Math Classes. *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 18, no 10, Article no em2167. <https://doi.org/10.29333/ejmste/12455>
39. Vaughn M. (2020) What Is Student Agency and Why Is It Needed Now More Than Ever? *Theory Into Practice*, vol. 59, no 2, pp. 109–118. <https://doi.org/10.1080/00405841.2019.1702393>
40. Vygotsky L. (1980) *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA; London: Harvard University.
41. Waheeb S.A., Khan N.A., Shang X. (2022) Topic Modeling and Sentiment Analysis of Online Education in the COVID-19 Era Using Social Networks Based Datasets. *Electronics*, vol. 11, no 5, Article no 715. <https://doi.org/10.3390/electronics11050715>
42. Wong A.V., Wong K., Hindle A. (2019) Tracing Forum Posts to MOOC Content using Topic Analysis. *ArXiv: 1904.07307v1*.
43. Yin B., Yuan C. (2022) Detecting Latent Topics and Trends in Blended Learning Using LDA Topic Modeling. *Education and Information Technologies*, vol. 27, no 9, pp. 12689–12712. <https://doi.org/10.1007/s10639-022-11118-0>
44. Zeiser K., Scholz C., Cirks V. (2018) *Maximizing Student Agency: Implementing and Measuring Student-Centered Learning Practices*. Boston; Washington, DC; Oakland: American Institutes for Research.

45. Zhang Z., Miao D., Gao C. (2013) Short Text Classification Using Latent Dirichlet Allocation. *Journal of Computer Applications*, vol. 33, no 6, pp. 1587–1590. <https://doi.org/10.3724/sp.j.1087.2013.01587>

## References

- Ahearn L.M. (2001) Language and Agency. *Annual Review of Anthropology*, vol. 30, no 1, pp. 109–137. <https://doi.org/10.1146/annurev.anthro.30.1.109>
- Alhazmi H. (2022) Detection of Students' Problems in Distance Education Using Topic Modeling and Machine Learning. *Future Internet*, vol. 14, no 6, Article no 170. <https://doi.org/10.3390/fi14060170>
- Allington R.L. (1980) Teacher Interruption Behaviors during Primary-Grade Oral Reading. *Journal of Educational Psychology*, vol. 72, no 3, pp. 371–377. <https://doi.org/10.1037/0022-0663.72.3.371>
- Andre J. (2021) *Can Learning Analytics Increase Agency and Transform Digital Learning?* Paper presented at the conference “Digital Transformation and Higher Education: When Challenges are Opportunities” (September, 2021, Hanoi, Vietnam).
- Bandura A. (1999) Social Cognitive Theory: An Agentic Perspective. *Asian Journal of Social Psychology*, vol. 2, no 1, pp. 21–41. <https://doi.org/10.1111/1467-839x.00024>
- Bandura A. (1986) *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice Hall.
- Chen Y., Yu B., Zhang X., Yu Y. (2016) Topic Modeling for Evaluating Students' Reflective Writing. Proceedings of the *Sixth International Conference on Learning Analytics & Knowledge — LAK '16 (Edinburgh, UK, 2016, April 25–29)*. <http://dx.doi.org/10.1145/2883851.2883951>
- Chen X., Zou D., Cheng G., Xie H. (2020) Detecting Latent Topics and Trends in Educational Technologies over Four Decades Using Structural Topic Modeling: A Retrospective of All Volumes of *Computers & Education*. *Computers & Education*, vol. 151, July, Article no 103855. <https://doi.org/10.1016/j.compedu.2020.103855>
- Copur-Gencturk Y., Choi H., Cohen A. (2022) Investigating Teachers' Understanding through Topic Modeling: A Promising Approach to Studying Teachers' Knowledge. *Journal of Mathematics Teacher Education*, vol. 26, pp. 281–302. <https://doi.org/10.1007/s10857-021-09529-w>
- Devi S., Dhavale C., Moharkar L., Khanvilkar S. (2022) Impact of Online Education and Sentiment Analysis from Twitter Data Using Topic Modeling Algorithms. *International Journal of Applied Sciences and Smart Technologies*, vol. 4, no 1, pp. 21–34. <https://doi.org/10.24071/ijasst.v4i1.4637>
- Dewey J. (1922) *Human Nature and Conduct: An Introduction to Social Psychology*. New York, NY: Henry Holt and Co. <https://doi.org/10.1037/14663-000>
- Duranti A. (2005) Agency in Language. *A Companion to Linguistic Anthropology* (ed. A. Duranti), Malden, MA: Blackwell, pp. 449–473. <https://doi.org/10.1002/9780470996522.ch20>
- Ekin C.Ç., Çakici M., Şener E., Türker S., Altanlar S. (2021) Research Trends Analysis in Educational Journal Publications on COVID-19 Using Descriptive and Text Mining Methods: Preliminary Analysis. *European Journal of Science and Technology*, iss. 29, pp. 432–437. <https://doi.org/10.31590/ejosat.1036109>
- Fogle L.W., King K.A. (2013) Child Agency and Language Policy in Transnational Families. *Issues in Applied Linguistics*, vol. 19. <https://doi.org/10.5070/14190005288>
- Grootendorst M. (2022) *BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure*. <https://doi.org/10.48550/arXiv.2203.05794>
- Gurcan F., Dalveren G.G., Derawi M. (2022) COVID-19 and e-Learning: An Exploratory Analysis of Research Topics and Interests in e-Learning during the Pandemic. *IEEE Access*, vol. 10, pp. 123349–123357. <https://doi.org/10.1109/access.2022.3224034>



- He J., Rubinstein B.I.P., Bailey J., Zhang R., Milligan S. (2017) TopicResponse: A Marriage of Topic Modelling and Rasch Modelling for Automatic Measurement in MOOCs. *ArXiv: 607.08720v2*.
- Heilala V. (2022) *Learning Analytics with Learning and Analytics: Advancing Student Agency Analytics* (PhD Thesis), Jyväskylä: University of Jyväskylä.
- Jääskelä P., Heilala V., Kärkkäinen T., Häkkinen P. (2020) Student Agency Analytics: Learning Analytics as a Tool for Analysing Student Agency in Higher Education. *Behaviour & Information Technology*, vol. 40, no 8, pp. 790–808. <https://doi.org/10.1080/0144929x.2020.1725130>
- Jääskelä P., Poikkeus A., Vasalampi K., Valleala U.M., Rasku-Puttonen H. (2016) Assessing Agency of University Students: Validation of the AUS scale. *Studies in Higher Education*, vol. 42, no 11, pp. 2061–2079. <https://doi.org/10.1080/03075079.2015.1130693>
- Johnston P.H. (2004) *Choice Words: How Our Language Affects Children's Learning*. Portland, ME: Stenhouse.
- Kandula S., Curtis D., Hill B., Zeng-Treitler Q. (2011) Use of Topic Modeling for Recommending Relevant Education Material to Diabetic Patients. *AMIA Annual Symposium Proceedings Archive*, pp. 674–682.
- Kastrati Z., Dalipi F., Imran A.S., Nuci K.P., Wani M.A. (2021) Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study. *Applied Sciences*, vol. 11, no 9, Article no 3986. <https://doi.org/10.3390/app11093986>
- Kim J. (2022) Analyzing 'Student Agency' Embedded in the Discourse of Future Education: (Re)interpretation of the OECD Education 2030 Project. *The Journal of Curriculum Studies*, vol. 40, no 2, pp. 181–202. <https://doi.org/10.15708/jscs.40.2.8>
- Korshunov A., Gomzin A. (2012) Tematicheskoe modelirovanie tekstov na estestvennom yazyke [Topic Modeling in Natural Language Texts]. *Proceedings of the Institute for System Programming of the RAS*, vol. 23. <https://doi.org/10.15514/ISPRAS-2012-23-13>
- Leadbeater C. (2017) *Student Agency: Learning to Make a Difference*. East Melbourne, VIC: The Centre for Strategic Education.
- Li Y., Zheng Y., Bao H., Liu Y. (2015) Towards Better Understanding of Hot Topics in Online Learning Communities. *Smart Learning Environments*, vol. 2, no 1, Article no 12. <https://doi.org/10.1186/s40561-015-0019-6>
- Littleton K., Taylor S., Eteläpelto A. (2011) Special Issue Introduction: Creativity and Creative Work in Contemporary Working Contexts. *Vocations and Learning*, vol. 5, no 1, pp. 1–4. <https://doi.org/10.1007/s12186-011-9067-4>
- Mameli C., Passini S. (2018) Development and Validation of an Enlarged Version of the Student Agentic Engagement Scale. *Journal of Psychoeducational Assessment*, vol. 37, no 4, pp. 450–463. <https://doi.org/10.1177/0734282918757849>
- Matos L., Reeve J., Herrera D., Claux M. (2018) Students' Agentic Engagement Predicts Longitudinal Increases in Perceived Autonomy-Supportive Teaching: The Squeaky Wheel Gets the Grease. *The Journal of Experimental Education*, vol. 86, no 4, pp. 579–596. <https://doi.org/10.1080/00220973.2018.1448746>
- McCauley L., King K. (2021) *Human-Centered Learning and Student Agency: "Think Big, Start Small and Act Fast"*. Available at: <https://knowledgeworks.org/resources/human-centered-learning-student-agency/> (accessed 20 February 2024).
- Molavi M., Tavakoli M., Kismihók G. (2020) Extracting Topics from Open Educational Resources. *Addressing Global Challenges and Quality Education. EC-TEL 2020. Lecture Notes in Computer Science* (eds C. Alario-Hoyos, M.J. Rodríguez-Triana, M. Scheffel, I. Arnedillo-Sánchez, S.M. Dennerlein), vol. 12315. Cham: Springer, pp. 455–460. [https://doi.org/10.1007/978-3-030-57717-9\\_44](https://doi.org/10.1007/978-3-030-57717-9_44)

- Nazari A., Hossennia M., Garmaroudi G., Torkian S. (2023) *Social Media and Mental Health in Students: A Cross-Sectional Study during the COVID-19 Pandemic*. <https://doi.org/10.21203/rs.3.rs-2535464/v1>
- Nogueira F.D. (2017) Reassembling the Social: An Introduction to Actor-Network Theory, Oxford university press, 2005. *Formação (Online)*, vol. 1, no 25, pp. 229–233. <https://doi.org/10.33081/formacao.v1i25.5251>
- Saarela M., Heilala V., Jaaskela P., Rantakaulio A., Karkkainen T. (2021) Explainable Student Agency Analytics. *IEEE Access*, vol. 9, pp. 137444–137459. <https://doi.org/10.1109/access.2021.3116664>
- Sorokin P.S. (2021) “Transformiruyushchaya agentnost” kak predmet sotsiologicheskogo analiza: sovremennye diskussii i rol' obrazovaniya [“Transformative Agency” as an Object of Sociological Analysis: Contemporary Discussions and the Role of Education]. *RUDN Journal of Sociology*, vol. 21, no 1, pp. 124–138. <https://doi.org/10.22363/2313-2272-2021-21-1-124-138>
- Sorokin P.S., Zykova A.V. (2021) “Transformiruyushchaya agentnost” kak predmet issledovaniy i razrabotok v XXI veke: obzor i interpretatsiya mezhdunarodnogo opyta [“Transformative Agency” as a Subject of Research and Development in the 21st Century: A Review and Interpretation of International Experience]. *Monitoring of Public Opinion: Economic and Social Changes*, no 5, pp. 216–241. <https://doi.org/10.14515/monitoring.2021.5.1858>
- Tadeo D.J., Yoo J. (2022) Topic Modeling of the Student Emails Sent before and during the Birth of COVID-19 in Physics and Math Classes. *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 18, no 10, Article no em2167. <https://doi.org/10.29333/ejmste/12455>
- Vaughn M. (2020) What Is Student Agency and Why Is It Needed Now More Than Ever? *Theory Into Practice*, vol. 59, no 2, pp. 109–118. <https://doi.org/10.1080/00405841.2019.1702393>
- Vygotsky L. (1980) *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA; London: Harvard University.
- Waheeb S.A., Khan N.A., Shang X. (2022) Topic Modeling and Sentiment Analysis of Online Education in the COVID-19 Era Using Social Networks Based Datasets. *Electronics*, vol. 11, no 5, Article no 715. <https://doi.org/10.3390/electronics11050715>
- Wong A.V., Wong K., Hindle A. (2019) Tracing Forum Posts to MOOC Content using Topic Analysis. *ArXiv: 1904.07307v1*.
- Yin B., Yuan C. (2022) Detecting Latent Topics and Trends in Blended Learning Using LDA Topic Modeling. *Education and Information Technologies*, vol. 27, no 9, pp. 12689–12712. <https://doi.org/10.1007/s10639-022-11118-0>
- Zeiser K., Scholz C., Cirks V. (2018) *Maximizing Student Agency: Implementing and Measuring Student-Centered Learning Practices*. Boston; Washington, DC; Oakland: American Institutes for Research.
- Zhang Z., Miao D., Gao C. (2013) Short Text Classification Using Latent Dirichlet Allocation. *Journal of Computer Applications*, vol. 33, no 6, pp. 1587–1590. <https://doi.org/10.3724/sp.j.1087.2013.01587>