

УДК 81 '33

Катермина Вероника Викторовна

доктор филологических наук,
профессор
Кубанский государственный
университет
г. Краснодар, Россия

Veronika Katermina

Doctor of Philological Sciences
Professor
Kuban State University
Krasnodar, Russia
katermina_v@mail.ru

Гнедаш Анна Александровна

кандидат политических наук,
доцент
Кубанский государственный
университет
г. Краснодар, Россия

Anna Gnedash

Candidate of Political Sciences,
Associate Professor
Kuban State University
Krasnodar, Russia
anna_gnedash@inbox.ru

ОПЫТ АНАЛИЗА
БОЛЬШИХ ЛИНГВИСТИЧЕСКИХ ДАННЫХ ПОСРЕДСТВОМ ИИ
(на примере исследования российских ВК-сообществ
в сфере семьи и родительства)

Статья посвящена кейсу применения технологий искусственного интеллекта к анализу больших лингвистических данных российских сообществ «ВКонтакте», тематически ориентированных на семью и родительство. Корпус из ≈ 480 тыс. постов трех крупнейших групп («Родители и педагоги! Наши дети», «Родители в теме», «Я МАМА!™ Дети. Семья») был собран в трех временных точках (февраль 2021, 2022 и 2024 гг.) и подвергнут сквозной обработке на базе ИИ «Концептоскоп»: токенизация, лемматизация, LDA-моделирование тем, кластеризация SVO-связок, оценка полярности (TextBlob-RU) и построение когнитивных графов. Полученные метрики показали, что во всех сообществах семейный дискурс следует общей траектории «ценностный пафос → функциональная прагматика»: героико-патриотические и воспитательные нарративы 2021 г. постепенно вытесняются утилитарными лайфхаками и рекламными рекомендациями 2024 г.; эмоциональная амплитуда сплющивается к нейтральному уровню, а ролевая вертикаль «взрослый наставник/ребенок/герой» растворяется в детоцентричных инструкциях. Эти выводы подтверждают продуктивность ИИ-подхода для фиксации скрытых семантических и аффективных сдвигов и позволяют рассматривать цифровое пространство семьи как сервисно-ориентированный конструкт, управляемый логикой быстрых решений.

Ключевые слова: искусственный интеллект; большие лингвистические данные; социальные сети; семейный дискурс; ВКонтакте; тематическое моделирование; эмоциональный анализ; цифровая социология.

EXPERIENCE OF BIG LINGUISTIC DATA ANALYSIS VIA AI
(A Case Study of Russian VK Communities on Family and Parenthood)

The paper presents a methodological case of applying artificial-intelligence techniques to the analysis of big linguistic data drawn from Russian VKontakte communities focused on family and parenthood. A corpus of approximately 480 thousand posts from three largest groups

(“Parents and Teachers! Our Children,” “Parents in the Know,” and “I МАМА!™ Children. Family”) was compiled for three time-points-February 2021, 2022 and 2024 – and processed end-to-end: tokenisation, lemmatisation, topic modelling with LDA, SVO-clustering, sentiment scoring with TextBlob-RU and construction of cognitive graphs using graph neural networks. Quantitative metrics reveal a common trajectory in all communities from value-laden pathos to functional pragmatics: the heroic-patriotic and educational narratives of 2021 are gradually displaced by utilitarian life-hacks and commercial recommendations by 2024; emotional amplitude flattens toward neutrality, and the hierarchical “adult-mentor/child-hero” role pattern dissolves into child-centred instructions. These findings demonstrate the efficacy of AI-based NLP for detecting latent semantic and affective shifts and suggest that the digital representation of family is increasingly shaped as a service-oriented construct driven by the logic of instant solutions.

Key words: artificial intelligence; big linguistic data; social networks; family discourse; VKontakte; topic modelling; sentiment analysis; digital sociology.

Искусственный интеллект (ИИ), оформленный в качестве своеобразного междисциплинарного фокуса, вбирает в себя достижения компьютерных наук, логики, биологии, политологии, психологии и лингвистики, чем обуславливает способность одной и той же методики одновременно решать инженерные задачи и извлекать глубокие культурно-семиотические смыслы из языковой ткани цифровых коммуникаций [1, с. 270]; именно эта теоретическая всеядность позволяет трактовать ИИ не как вспомогательный инструмент, а как методологический фундамент гуманитарной аналитики больших текстовых массивов.

Благодаря подобной методологической рамке поток неструктурированных сообщений, будь то повседневные статусы, дискуссионные тренды или меметические реплики, трансформируется в иерархизированную совокупность сущностей, отношений и тематических кластеров: процедура автоматического извлечения знаний, укорененная в концепции Natural Language Understanding [2], сегментирует текст, фиксирует лексико-грамматические маркеры, устанавливает логические связи и складывает их в машино-читаемую базу, что, в свою очередь, делает возможной последующую статистико-семиотическую интерпретацию больших корпусов.

Поворотным технологическим моментом становится вовлечение больших языковых моделей (БЯМ), которые, сформированные предварительным обучением на терабайтах разнообразного текста, способны одновременно удерживать синтаксическую целостность, учитывать прагматический контекст и реконструировать скрытые мотивационные паттерны дискурса, тем самым задавая качественно иной стандарт лингвистической аналитики и позволяя без ручной разметки индексировать, категоризовать и эмоционально профилировать миллионы сообщений в режимах, близких к реальному времени [3, с. 25].

Примененные к российским сообществам «ВКонтакте», посвященным семье и родительству, методики Natural Language Processing (NLP, обработка естественного языка) демонстрируют, что масштабная, динамическая и жанрово размноженная речевая среда способна служить чувствительным баро-

метром культурных трансформаций. БЯМ вскрывают скрытые кластеры тем от «цифрового родительства» до пересмотра гендерных ролей, моделируют миграцию ценностных акцентов и позволяют предсказывать какие дискурсивные «узлы напряжения» обострятся в обозримом будущем, превращая лингвистический анализ Больших Данных в инструмент социокультурного прогнозирования.

Объект нашего исследования – цифровой дискурс российских ВК-сообществ, посвященных семье и родительству, то есть совокупность всех постов, опубликованных в трех крупнейших группах ВКонтакте («Родители и педагоги! Наши дети», «Родители в теме», «Я МАМА!TM Дети. Семья»).

Цель исследования – выявить динамику смысловых доминант, ценностных акцентов и эмоциональных регистров семейно-родительского дискурса ВК-сообществ в трех временных точках – февраль 2021 г. (до СВО), февраль 2022 г. (момент начала) и февраль 2024 г. (третья годовщина).

Этапы исследования

1. Сформировать корпус объемом 480 тыс. публикаций путем сплошной выгрузки данных через VK API и экспертной фильтрации ключевых слов «мама», «папа», «дети», «семья» и др.

2. Предобработать тексты: удалить пустые строки, стоп-слова, унифицировать регистр, выполнить лемматизацию.

3. Провести тематическое моделирование методом LDA и SVO-классификации, выделив устойчивые смысловые узлы.

4. Оценить эмоциональную полярность публикаций при помощи TextBlob-RU.

5. Построить когнитивные графы смыслов с использованием комплекса «Концептоскоп» на базе графовых нейросетей.

6. Сопоставить результаты трех временных срезов и интерпретировать тенденции (традиционализм → эмоциональная поддержка → прагматизм).

Исследование базируется на парадигме «Большие лингвистические данные» и когнитивно-лингвистическом подходе к дискурсивному анализу; используется подход Digital Sociology о социальных медиа как барометре общественных изменений; принцип Data-Driven Linguistics, где знания извлекаются напрямую из неструктурированных текстов.

Методы:

- сбор данных через VK API;
- токенизация, лемматизация и стоп-фильтрация на Python;
- тематическое моделирование LDA на матрице CountVectorizer;
- анализ SVO-связок и n-грамм для семантического профилирования;
- расчет эмоциональной полярности TextBlob-RU;
- построение когнитивных графов в «Концептоскоп», работающем на графовых нейронных сетях;
- диахронное сравнение трех корпусов с последующей качественной интерпретацией.

Результаты исследования: как трансформируются семейные репрезентации в трех ВК-сообществах.

«Родители и педагоги! Наши дети», «Родители в теме», «Я МАМА!™ Дети. Семья»

Каждая ВК-площадка прошла собственную, но типологически соотносимую траекторию «ценностный пафос → функциональная прагматика».

В Сообществе «Родители и педагоги! Наши дети» героико-педагогический дискурс 2021 года, насыщенный патриотическими аллюзиями и фигурой взрослого наставника, уже к 2024-у редуцируется до утилитарных чек-листов и детоцентричных лайфхаков. Сообщество «Родители в теме», изначально ориентированное на развернутые методички раннего развития, эволюционирует к форматам микро-подказок, поддерживая «эмоциональное плато» нулевой полярности и лишь эпизодически добавляя экологические акценты. Сообщество «Я МАМА!™ Дети. Семья», стартовавшее с «папоцентричной» лирики, демонстрирует наиболее резкий кульбит: к 2024 году половина контента превращается в каталог партнерских ссылок и функционирует уже как рекламный хештег в витрине маркетплейса. При этом для всех групп характерно статистически фиксируемое выравнивание эмоционального тонуса к нулю, что можно интерпретировать как аффективную усталость аудитории и переключение коммуникации на режим минимального когнитивного вложения.

Кросс-сопоставление показывает феномен аффективной нивелировки: распределения полярности, какими бы разными они ни были на входе, стремятся к нулю, а спектр эмоций сужается.

Гендерная перспектива подтверждает общий сдвиг от иерархий к растворению ролей. Если в 2021 г. вертикаль «взрослый наставник → ребёнок-герой» задает композицию постов «Родители и педагоги! Наши дети», то к 2024 г. взрослый субъект становится безымянным фасилитатором чужих лайфхаков, а ребенок формально превращается в единственного адресата инструкции. В Сообществе «Я МАМА!™ Дети. Семья» аналогичный процесс идет через стадию «гендерного тандема счастья» (2022 г.) к роли «советчика-покупателя», когда автор говорит читателю, что именно следует класть в корзину ради «гармоничного развития» ребенка. Сообщество «Родители в теме» изначально избегало жестких ролевых оппозиций, однако и здесь в 2024 г. фиксируется переход от индивидуального «родитель-эксперт» к коллективному «мы – команда, улучшающая мир», что эмпирически проявляется в росте запросов об имени или совместных экологических инициативах.

Качественная интерпретация этих сдвигов выводит нас к концепту семьи как сервиса. Когда эмоциональное поле выравнивается, а идеологические знаки изымаются, на первый план выходит прагматика времени и ресурса: быстрее найти совет, схему, ссылку; быстрее получить выгоду; быстрее адаптировать ребенка к бурной внешней среде. Именно этот сервисный вектор задает рост жанрового микротекста: длинная эссеистика стирается в пользу чек-листов и бейдж-постов, а доминирующий медиум смещается к изображению или лаконичному сторис-описанию.

Такое понимание открывает сразу две перспективы. С одной стороны, у гуманитарной науки появляется верифицируемый инструмент описания «сердобольного выхолащивания» ценностей: ИИ-модели позволяют количественно доказать, что престиж героических, патриотических или воспитательных нарративов снижается и замещается сервисной лексикой. С другой – политикам в сфере семейного благополучия и цифрового образования становится яснее, что попытки возвращать в публичный обиход «большой пафос» без учета потребности в быстрой практической выгоде могут наткнуться на барьер аффективного равнодушия.

У всех трех VK-сообществ эмотивный график сплюснулся почти до нуля. Нейросеть видит: вместо всплесков радости или гордости – ровная линейка информационных сообщений. Это похоже на «аффективную усталость»: людям нужно быстрое решение, а не патетика. Одновременно исчезли героические и патриотические слова; там, где раньше говорили о «подвиге» или «будущем защитнике», теперь сухо объясняют, как успеть сделать поделку до понедельника. Даже мем «папа – герой» выжил лишь в формате рекламного хэштега.

Получается, что в цифровом пространстве семья перестает быть «малой родиной» или «школой гражданина» и превращается в сервисный конструктор: набор быстрых инструкций, скидок и микрорешений. Именно искусственный интеллект раскладывает эту трансформацию по полочкам: от частот слов до диаграмм эмоций и делает видимой тенденцию, которая на глазах выглядела бы как обычная суэта ленты. То есть большие лингвистические данные + ИИ = инструмент, который позволяет заметить, как героический нарратив тихо сдает позиции прагматике, а семейная риторика становится руководством «сделай-сам» в реальном времени.

ЛИТЕРАТУРА

1. Сорокина С. Г. Искусственный интеллект в контексте междисциплинарных исследований языка // Вестник Кемеровского государственного университета. Серия: Гуманитарные и общественные науки. 2023. Т. 7. № 3. С. 267–280. DOI: <https://doi.org/10.21603/2542-1840-2023-7-3-267-280>.
2. Томарева И. Г. Взаимодействие ИИ и когнитивной лингвистики: актуальное положение и будущие направления // Международный научно-исследовательский журнал. 2024. № 6 (144). DOI: <https://doi.org/10.60797/IRJ.2024.144.86>.
3. Громов М. С., Чертовских М. Г. Большие языковые модели: текущее состояние, оценки и прогнозы // Международный бизнес. 2023. № 3 (5). С. 24–31.