

УДК 811.1'33

МРНТИ 16.31.21

<https://doi.org/10.48371/PHILS.2025.1.76.023>

ПРИМЕНЕНИЕ МЕТОДОВ КОРПУСНОГО АНАЛИЗА ПРИ ИЗУЧЕНИИ МАНИПУЛЯТИВНОГО МЕДИАДИСКУРСА

*Шакенова М.Т.¹, Миятбекова З.У.², Джилкишиева М.С.³,
Баймаханбетов М.А.⁴

^{*1,2,3}НАО «Южно-Казахстанский университет им. М. Ауэзова»
Шымкент, Казахстан

⁴АО «Национальный центр исследования и оценки в образовании
«Талдау» имени А. Байтурсунова», Астана, Казахстан

Аннотация. В данной статье рассматривается применение методов корпусного анализа при изучении манипулятивного медиадискурса. Актуальность темы в контексте современного информационного общества, живущего в условиях постоянного информационного потока, безусловна, поскольку методы корпусной лингвистики позволяют работать с большими базами данных и выявлять закономерности в них. Всеми очевидным фактом является осознание высокой роли СМИ в формировании общественного мнения, а также информационно-психологического воздействия на него. В связи с этим существует острая необходимость в эффективных методах анализа большого массива медиатекстов, заполнивших все интернет-пространство. Авторы представляют подробный анализ теоретических и методологических аспектов корпусной лингвистики в контексте изучения медиадискурса, описывают опыт формирования малых текстовых корпусов и их применения при исследовании манипулятивного медиадискурса. Под корпусом текстов понимается структурированная по ряду параметров система собранных текстов на определенном языке (в нашем случае медиатекстов на русском и казахском языках) в электронной форме. Материалом исследования послужили медиатексты, собранные из различных русскоязычных и казахскоязычных информационных интернет-источников, включая онлайн-новости и социальные медиа, за последние 5 лет. Изучение подобных корпусов с применением методов корпусного анализа обеспечивают объективность при исследовании медиадискурса, так как они основаны на статистических данных и частотности употребления в медиатекстах языковых средств, обладающих манипулятивным потенциалом. Авторы подчеркивают значимость данного подхода для глубокого понимания языковых стратегий, используемых в медиатекстах, и выявления их воздействия на читательскую аудиторию.

Ключевые слова: корпусная лингвистика, корпус, корпусный анализ, медиадискурс, медиатекст, манипулятивный потенциал, тональность, коллокация

Введение

Манипулятивный медиадискурс – это форма медиакommunikации, направленная на информационно-психологическое воздействие на мнение и настроение большой аудитории через использование манипулятивных стратегий, языковых структур и контекста.

Корпусная лингвистика – это область лингвистики, которая использует большие объемы текстовых данных (корпусы) для изучения языковых явлений и выявления их закономерностей.

Методы, используемые при корпусном анализе – это методы исследования, основанные на анализе корпусов текстов с использованием статистических и лингвистических методов для выявления частотности употребления языковых явлений в текстовых корпусах.

В современном обществе существует высокая степень информационной перегрузки, медиадискурс играет ключевую роль в формировании общественного мнения и воздействии на массовое сознание. Средства массовой информации могут использовать различные манипулятивные стратегии для формирования определенных представлений и мнений в обществе. В условиях нескончаемого огромного информационного потока необходимы новые методы их изучения, позволяющие обрабатывать большие объемы текстов. Корпусный анализ помогает выявлять и анализировать языковые инструменты, используемые в этом процессе.

С целью разработки методики распознавания медиаманипулирования необходимым этапом является формирование корпусов текстов, на основе которых будет разработана и апробирована разрабатываемая методика. Наша цель – на основе созданных малых корпусов медиатекстов из русскоязычных и казахскоязычных интернет-источников с помощью методов корпусного анализа выявить манипулятивные элементы. Методологической базой изучения корпусов текстов стали работы таких ученых, как Д. Бибер [1], Г. Лич [2], Т. Варади [3], Дж. Синклэр [4], Х. Зинсмайстер, Л. Лемницер [5], [6], С. Грейнджер В.А. [6] Плунгян [6] и др. Рассмотрим некоторые аспекты изучения корпусов, описанных в данных работах. Так, Д. Бибер, один из ведущих специалистов в области корпусной лингвистики, предложил теоретический подход к изучению разнообразия различных вариантов английского языка с использованием корпусных данных. Им применен функциональный анализ текстов на больших корпусах [1]. В работах Г. Лича раскрываются основы корпусной лингвистики, им затронуты вопросы частотности, лексического разнообразия и использования корпусов в лингвистическом исследовании [2]. Большой интерес представляют работы Т. Варади, в которых описаны методы анализа больших текстовых корпусов для создания лексикографических ресурсов [3]. В работе Дж. Синклэр «Corpus, Concordance, Collocation», одного из основоположников корпусной лингвистики, предложено использование корпусов при изучении коллокаций. [4]. Хайке Зинсмайстер, Лотар Лемницер в своей работе «Korpuslinguistik» раскрыли методологические вопросы в создании и

описании корпусов на немецком языке, ими отмечено, что корпус состоит из первичных данных и метаданных [5], [6]. В работе Дж. Синклера «Corpus, Concordance, Collocation» подчеркивается о важности сохранения в корпусе текстов в естественном формате, то есть в том виде, как он непосредственно представлен в речи [7]. В статье «The Computer Learner Corpus: A Versatile New Source of Data for SLA Research» говорится о возможностях компьютерных корпусов английского языка для исследования процессов усвоения второго языка и улучшения методики преподавания языка [6].

В.А. Плунгян, российский лингвист, в своих работах в области корпусной лингвистики занимается проблемами семантики и структуры текстов. Его исследования часто фокусируются на аспектах значения слова и лексических групп в контексте корпусного анализа [8].

В Казахстане также активно разрабатываются исследования в области корпусной лингвистики и расширения Национального корпуса казахского языка. Так, в статье «Методы автоматизации текстов в национальном корпусе казахского языка» рассматриваются особенности учебного подкорпуса и способы семантизации и их автоматизации [9].

Контент-анализ существующей литературы позволил нам обобщить понятие корпуса текстов. Корпус текстов представляет собой совокупность электронных текстов по тем или иным критериям. Цель исследования – формирование и изучение малых корпусов медиатекстов с применением методов корпусного анализа. Задачи исследования: описать структуру сформированных двух малых корпусов медиатекстов; продемонстрировать применение тематического моделирования при описании корпусов; использовать частотный анализ слов с эксплицитно выраженной негативной семантикой для характеристики негативной тональности текстов, провести коллокационный анализ слов с имплицитно выраженной негативной семантикой в медиатекстах.

Описание материалов и методов

Исследование базируется на использовании современных методов корпусной лингвистики – методов анализа текстовых корпусов. Корпусный анализ предполагает исследование языка на основе больших текстовых данных. В качестве материала исследования использованы статьи аналитического характера, написанные на казахском и русском языках. Данные тексты отобраны нами из популярных казахстанских информационных порталов. В процессе работы была проведена фильтрация текстов с учетом освещения актуальных вопросов и тем. В настоящее время отобрано 1964 русскоязычных и 1962 казахскоязычных медиатекстов. Основными методами исследования являются методы статистического и контекстного анализа для выделения языковых средств, обладающих манипулятивным потенциалом, для определения их контекстуальной окраски и частотности использования в медиатекстах.

Результаты и обсуждение

Формирование малых корпусов медиатекстов

Манипулятивный медиадискурс представляет собой форму коммуникации, где через специально отобранные языковые средства оказывается скрытое информационно-психологическое воздействие на аудиторию, в результате которого у нее формируется выгодное манипулятору восприятие событий, идей или личностей. Для изучения манипулятивного медиадискурса с применением корпусного анализа нами сформированы два малых корпуса медиатекстов: на казахском и русском языках.

В выборку текстов для корпуса были включены казахскоязычные и русскоязычные публикации открытых казахстанских информационных интернет-источников (новостные интернет-порталы и официальные СМИ). При отборе источников мы основывались на популярные республиканские источники: <https://baq.kz>, <https://kaz.inform.kz>, <https://qazaqstan.tv>, <https://rezonans.asia>, <https://www.ktk.kz>, <https://forbes.kz>, <https://orda.kz>, <https://tengrinews.kz> (см. таблицу 1).

Одним из важных параметров составления корпуса является сбалансированность. Данный параметр требует пропорционально и достаточное количество текстов, позволяющих отобразить необходимые признаки/категории исследуемой области. В основном соблюдение сбалансированности осуществляется за счет пропорционального распределения всех признаков/категории, где каждой категории присваивается весовая доля, определяющая количество текстов [10]. Мы попытались учесть данный критерий по источникам текстовой информации (см. рисунок 1).

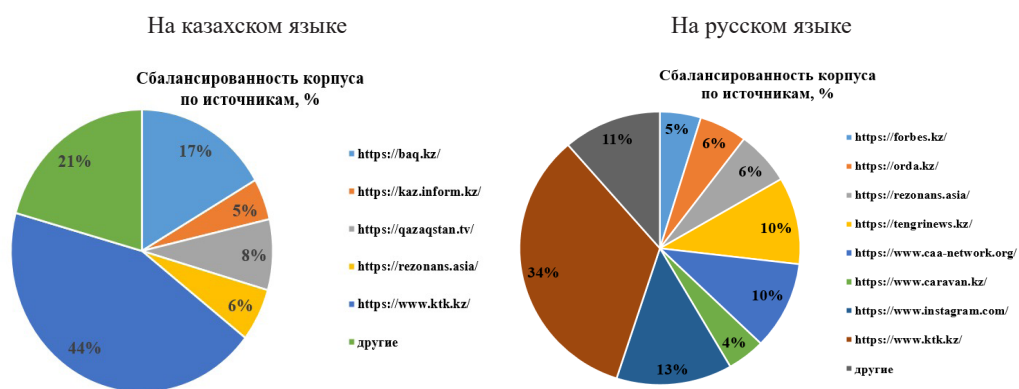


Рисунок 1 – Сбалансированность корпуса по источникам, %

Однако следует отметить, что оба малых корпуса медиатекстов требуют дополнительных доработок с учетом сбалансированности для пропорционального соотношения количества текстов по источникам. В обоих корпусах преобладают тексты из источника <https://www.ktk.kz>: 44%

– на казахском языке, 34% – на русском языке. Корпус медиатекстов на казахском языке следует дополнить текстами, в том числе из <https://qazaqstan.tv>, <https://rezonans.asia>, <https://kaz.inform.kz> и других источников. Корпус русскоязычных медиатекстов также необходимо дополнить статьями из <https://qazaqstan.tv>, <https://rezonans.asia>, <https://forbes.kz>, <https://orda.kz>. Над этой задачей по увеличению сбалансированности по данному параметру в настоящее время наша исследовательская группа продолжает работать.

Применение метода тематического моделирования корпусов

Методы корпусного анализа играют важную роль в изучении манипулятивного медиадискурса, так как они позволяют исследователям систематизировать, анализировать и интерпретировать большие объемы текстовых данных. Одним из способов корпусного анализа является идентификация ключевых слов, позволяющих выявить термины, используемые в текстах корпуса. В процессе тематического моделирования был использован апробированный ранее [11] инструментарий, разработанный на основе латентного размещения Дирихле LDA [12]. Алгоритм включает следующие шаги: вероятностная модель коллекции документов, преобразование текстов в нижний регистр для исключения нескольких копий одних и тех же слов; удаление стоп-слов (союзы, предлоги) и очень редких слов; токенизация слов; приведение слов к начальной форме путем использования лемматизации; создание словаря и корпуса; обучение модели. Затем запускается множество моделей LDA с разными значениями количества тем и определяется тема с наибольшим значением когерентности. Мерой интерпретируемости тем считается мера, показывающая частотность слов в текстах. Для оценивания согласованности использовалась поточечная взаимная информация (pointwise mutual information, PMI):

$$PMI_t = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i}^k PMI(w_i, w_j),$$

где w_i – ое слово.

После нахождения наилучшей модели выводятся ключевые слова для каждой темы. Данный инструментарий был применен к сформированным нами двум малым корпусам медиатекстов. Ниже в таблице 1 приведены результаты тематического моделирования.

Таблица 1 – Результаты тематического моделирования.

Корпус медиатекстов на казахском языке

Название топика	Список ключевых слов
Topic: 1 Мектеп	<i>мектеп, жағдайда, байланысты, балалар, мүмкін, қажет, бала</i>

Торіс: 2 Қазақстандағы жоғарғы оқу	<i>қазақ, халықаралық, атындағы, жоғары, оқу, студент, қатысқан, өнер, әлемдік</i>
Торіс: 3 Қазақстанның спортшылары Азия ойындары	<i>Шығатын, мекемесінің, қола, тарады, алқасының, Қазақстаннан, күміс, Азия</i>
Торіс: 4 Тұрғын үй және сауда мәселесі	<i>су, үй, бастап, сатып, астам, ақша, тұрғын, қаржы, ішінде, қайта</i>
Торіс: 5 Әйел және бала мәселесіне қатысты қылмыстық істер	<i>жастағы, сот, әйел, істер, қылмыстық, іс, бас, тұрғыны, бала, анасы</i>
Торіс: 6 Жол көлігі апатына ұшырағандарға көмек көрсету	<i>жол, ауруханаға, саны, жүргізушісі, жағдай, көлік, көлігінің, апаты»</i>
Торіс: 7 Өрт болған жағдайдағы әрекет	<i>екен, қалған, тіпті, өзі, келген», өрт, жатқан</i>
Торіс: 8 Нарықтық экономика жағдайындағы азық-түлік мәселесі	<i>азық, түлік, долларға, астық, деңгейлі, өсімдік, сәйкесінше, банк, банктің, нарығына</i>
Торіс: 9 Денсаулық сақтау мәселесі	<i>ауырлау, шағымдану, қауіпті, мазасыздық, депрессия зиян, тәзімділік, жоғалту, әлсіреу</i>
Торіс: 10 Президент және мемлекет	<i>президент, мемлекет, Касым-Жомарт, Тоқаев, басшысы, экономикалық, ел</i>
Торіс: 11 Полиция және тергеу жұмысы	<i>оқиға, полиция, тергеу, жастағы алматы, баспасөз, бас, сағат, алған, белгісіз</i>
Торіс: 12 Әлеуметтік жағдай және зейнетақы	<i>зейнетақы, БЖЗҚ, зейнеткерлікке, төменгі, базалық, төмендету, қамқорлық, толығырақ, күнкөріс</i>

Корпус медиатекстов на русском языке

Название топика	Список ключевых слов
Торіс: 1 Жилищно-коммунальные и дорожные вопросы	<i>аким, дорога, район, акимат, житель, Алматы, город, участок, водитель, село</i>
Торіс: 2 Торговля	<i>цена, продукт, взрыв, хлеб, авиакомпания, тонна, рейс, магазин, товар</i>
Торіс: 3 Российско-украинские отношения	<i>российский, отношение, экономический, граница, Украина, военный, глава, Республика</i>
Торіс: 4 Криминал и суд	<i>суд, дело, уголовный, полиция, расследование, орган, служба, организация, преступление</i>
Торіс: 5 Повышение цен и курс валюты	<i>бензин, нефть, доллар, рейтинг, цена, детский, скачок, курс</i>
Торіс: 6 Промышленность и производство	<i>завод, актюбинский, показатель, павлодарский, карагандинский, предприятие, место, мангистауская</i>
Торіс: 7 Международные отношения	<i>британский, допускаться, Афганистан, музей, огонь, фильм, талибан, Лондон, космический, перевод</i>
Торіс: 8 Образование	<i>школа, образование, студент, школьник, учебный, вуз, учитель, педагог, обучение, университет</i>

Торіс: 9 Языковая политика	<i>язык, организация, казахский, государственный, решение, общество, движение, центральный, народ, национальный</i>
Торіс: 10 Здравоохранение: коронавирус и вакцинация	<i>коронавирус, врач, вакцина, здравоохранение, вакцинация, маска, доза, больница, вирус, пневмония</i>
Торіс: 11 Выборы	<i>партия, кандидат, выборы, комиссия, избирательный, пост, предвыборный, выдвинуть</i>
Торіс: 12 Строительство и инфраструктура	<i>проект, строительство, комплекс, ожидать, экологический, инфраструктура, объект, поручение, инвестор</i>

Тематическое моделирование данных корпусов позволило выделить ключевые темы, отражающие актуальные социальные, экономические и политические вопросы в Казахстане. В собранных медиатекстах на казахском языке основное внимание уделяется вопросам образования, социальной защиты, транспорта, экономики, темам, связанным с женщинами и детьми, криминалом и чрезвычайными ситуациями. В корпусе русскоязычных медиатекстов выделены темы на жилищно-коммунальные и дорожные проблемы, торговлю, промышленность, международные отношения, политику и здравоохранение. Таким образом, тематическое моделирование в обоих корпусах отражает схожие проблемы, медиаповестка объединяет казахскоязычную и русскоязычную аудитории.

Применение метода частотного анализа на основе словарного подхода на примере слов с эксплицитно выраженной негативной семантикой

Методы частотного анализа используются для выявления закономерностей в использовании тех или иных языковых средств, направленных на оказание информационно-психологического воздействия в манипулятивном медиадискурсе. Анализ собранного нами корпуса показал, что в нем доминирует употребление слов с эксплицитно выраженной негативной семантикой. Выявление частотности данных слов стало возможным при использовании словарного подхода на основе раннее составленного нами словаря слов с негативным значением. Предварительно слова в словаре и текстах обрабатываются для улучшения точности анализа (удаление пунктуации, символов, стоп-слов, токенизация и лемматизация). Каждое слово из текста сравнивается со словарем. Если слово совпадает с элементом словаря, оно фиксируется как слово с негативной семантикой. Затем для каждого текста фиксируются все найденные совпадения и определяется общее количество слов с негативной семантикой в корпусе.

Анализ выявил, что наиболее часто в текстах используются следующие слова с негативной семантикой (нами указаны топ-20 слов): *ауыр – 149; қатты – 120; мәселе – 108; қауіпті – 105; төмен – 104; қиын – 94; күдікті – 92; заңсыз – 89; тыйым – 81; сыбайлас – 80; айыппұл – 77; күрделі – 74; қауіп – 70; өлім – 68; есірткі – 67; ауру – 62; қылмыс – 57; соғыс – 53;*

қорғаныс – 53; инфляция – 47; проблема – 585, война – 337, против – 259, угроза – 176, конфликт – 164, нарушение – 159, санкция – 138, талибан – 135, насилие – 130, преступление – 121, негативный – 114, кризис – 106, терроризм – 98, дефицит – 92, смерть – 89, зависимость – 86, коррупция – 85, погибнуть – 83, штраф – 77, жертва – 73.

В качестве примера приведем фрагмент текста, в котором активно используются слова с негативной семантикой: Әйелдер мен балаларға қатысты **зорлық-зомбылық**, Әлемде болып жатқан жағдаяттар, экономикалық **қиындықтар** мен **жұмыссыздық** адамдар арасындағы **агрессия** деңгейінің өсуіне алып келді. Бұндай **агрессивті** топтар алдымен әйелдер мен балаларға тіпті ер адамдарға да қысым көрсетеді. **Covid-19 пандемиясы** басталғалы бері балалар мен әйелдерге көрсетілетін **қылмыстық зорлық-зомбылық** деңгейінің өзектілігі артып келеді. Әрине, қыздар мен әйелдерге, балалармен әлсіз жандарға жасалатын **зомбылық пандемияға** дейін де болған. Бірақ, **карантин, жұмыссыздық пен тұрмыстық қажеттіліктер зорлық** көрсету деңгейінің өсуіне алып келді. Бұл жаһандық күрмеуі **қиын мәселе** (<https://qazdauiri.kz/news/aielder-men-balalarga-qatysty-zorlyq-zombylyq>).

В данном фрагменте используются слова с негативной семантикой для эмоционального воздействия и акцентирования внимания на остроте проблемы женского и детского насилия. Негативная лексика вызывает тревогу, усиливает драматизацию, подчеркивает серьезность данного вопроса в стране и мире..

Украинский **кризис** затронул Казахстан в сложный период развития, когда страна только начала восстанавливаться после воздействия **пандемии COVID-19**, а затем и **трагических январских событий**. Влияние **войны** в Украине на экономику Казахстана: что поменялось за два года? **Война** в Украине изменила экономические и стратегические отношения между многими странами мира, включая Казахстан и другие государства в Центральной Азии. Экономические **последствия конфликта** непосредственно затронули не только Россию и Украину, но и многие другие страны в регионе и за его пределами, чья экономика взаимосвязана с ними. В ответ на действия России западные страны во главе с США и ЕС ввели расширенный пакет **санкций**. Подобные **санкции** в последние несколько десятилетий стали основным инструментом в разрешении международной экономической и политической **напряженности** между странами. Хотя **санкции** были разработаны так, чтобы напрямую затронуть только Россию, их влияние заметно ощутилось и в соседних с Россией странах, включая Казахстан, из-за высокой степени экономической **зависимости** между ними. К ряду компаний, предположительно содействовавших обходу первичных **санкций** Россией, также были применены вторичные санкции. Эксперты практически единогласно прогнозировали высокую **инфляцию** и процентные ставки, которые будут сопровождаться

низким ростом экономики и *снижением* доходов населения, особенно его социально уязвимой части. (<https://inbusiness.kz/ru/news/vliyanie-vojnuy-v-ukraine-na-ekonomiku-kazahstana-cto-pomenyalos-za-dva-goda>). В данном тексте негативная лексика подчеркивает серьезность влияния украинского кризиса на Казахстан, показывает нестабильность и сложность ситуации, экономическую и политическую уязвимость Казахстана в сложившихся условиях и важность решения данного вопроса.

Следует отметить, что частотный анализ на основе словарного подхода имеет ряд ограничений при анализе текстов, поскольку при таком подходе не учитываются контекст, полисемия, ирония, идиомы и др. В связи с этим предлагаем его дополнить коллокационным анализом.

Применение метода коллокационного анализа на примере слов с имплицитно выраженной негативной семантикой

Выше мы рассмотрели слова с эксплицитно выраженной негативной семантикой. Но нередко нейтральные слова или слова с позитивной окраской в манипулятивных текстах содержат имплицитную негативную оценочность. Такую трансформацию и учет соседних слов и выражений в контексте можно определить посредством коллокационного анализа.

Данный анализ применен в ряде современных исследований, в которых отмечается, что контекстуальная оценочность обнаруживается при анализе в корпусе текстов коллокатов – окружения слова в контексте [4], [10]. Т.Б. Радбиль исследует явление наведенной оценочности, которую рассматривает путем анализа слова *народный* в его различных коллокациях на материале Национального корпуса русского языка. Как отмечает исследователь, слово «оценивается носителями национального языка как безусловно положительное свойство», включая такие смыслы, как: «имеющий ценностную (духовную или этическую) значимость», «соответствующий идеалам и ценностям народа», «истинный, правильный, такой, какой должен быть» и пр.: *хранитель народной мудрости; многовековые традиции народного благочестия; эффективную систему народного просвещения и развить народное самосознание* [13]. Однако в определенных репрезентативных контекстах это слово приобретает наведенную негативную оценочность. К примеру, в контексте употребления этого прилагательного с определяемыми словами, выражающими какие-либо негативные стороны жизни (грех, порок, недостаток, слабость): *на самом-то деле реальное народное зло – это сам чиновник и есть; Поговорите о народных грехах с ээками; Проблема неправильных капиталов, образующихся за счет разложения общества, народного разврата, пьянства...* [13]. Данный пример, описанный Т.Б. Радбилем, показывает, что коллокационный анализ позволяет рассмотреть вторичные, имплицитно выраженные негативные смыслы.

Подобных слову *народный* примеров достаточно много. Выявить

их можно путем применения коллокационного анализа. Приведем в качестве примера слово *ту*, которое в казахском языке имеет эксплицитно выраженную положительную семантику: *Керей-Жәнібек хан шаңырағын көтеріп, іргесін бекіткен күннен бастап, тәуелсіздіктің көк туы қайта желбірегенге дейінгі аралықта қазақ халқы қаншама қиындықты басынан өткізді* (<https://alashainasy.kz/aleumet/auyildyi-ansaymyiz-brak-kalada-turgyimyiz-keled-125738//aleumet/auyildyi-ansaymyiz-brak-kalada-turgyimyiz-keled-125738//aleumet/auyildyi-ansaymyiz-brak-kalada-turgyimyiz-keled-125738/>). В данном примере слово ТУ имеет положительное значение. Контекстной реализации позитивного значения способствует коллокация *көк* и *тәуелсіздіктің* (*Тәуелсіздіктің көк туы* – голубой флаг Независимости). Однако в ряде реперзентативных контекстов в слове *ту* реализуется его отрицательный оценочный потенциал. *Орыс тілді БАҚ РФ айдап салуымен «мемлекеттік тіл өз қызметін орындай алмауы» деген сұлтауды тағы да ту етіп көтере* бастады. *Армандары – Беларусьтегідей орыс тілін мемлекеттік қызметтегілер мен бюджеттегілер арасындағы тілді білмеушілер мен білгісі келмеушілерге байланысты қолданылатын жаза түрлері қазақ мемлекетін құруда қолданылмауы керек екен-мыс* (<https://abai.kz/post/103894>). В данном примере слово *ту* употреблено в значении «расшуметься», которое включает негативную коннотацию. Иными словами, приведенный пример анализа слова *ту* в казахском языке показывает, что это слово, обладая базовой позитивной семантикой (символ независимости и чести), может приобретать негативные оттенки в сочетаниях с определенными коллокациями, например, обозначая конфликтность или чрезмерное возмущение.

Коллокационный анализ слов *январь* и *январский* позволил выявить у данных слов имплицитно негативную семантику, которая реализуется в определенных коллокациях: *Кровавый январь: родные погибших просят депутатов Европарламента о международном расследовании; 24 августа близкие пострадавших во время январских событий сначала пришли в прокуратуру с требованием объективного расследования* (<https://orda.kz/blizkie-pogibshih-vo-vremya-krovavogo-yanvaryu-prosyat-evroparlament-nachat-mezhdunarodnoe-rassledovanie/>); *Январская трагедия, в которой каждому казахстанцу очевидны самые страшные преступления против человека и государства* (<https://orda.kz/posle-nazarbaeva-kak-skoro-v-kazahstane-rojavitsya-povaya-etika/>). Слова *январь* и *январский*, связанные с январскими событиями, демонстрируют отрицательную оценочность в вышеприведенных примерах.

Приведем еще один пример: слово *өсу* / *рост*, которое эксплицитно содержит положительную оценку, однако в определенных коллокациях (*агрессия деңгейінің өсу, зорлық көрсету деңгейінің өсу; рост проблем, рост цен* и др.) оно приобретает негативную коннотацию. Например: *Әлемде болып жатқан жағдаяттар, экономикалық қиындықтар мен*

жұмыссыздық адамдар арасындағы агрессия деңгейінің өсуіне алып келді. Бірақ, карантин, жұмыссыздық пен тұрмыстық қажеттіліктер зорлық көрсету деңгейінің өсуіне алып келді. Бұл жаһандық күрмеуі қиын мәселе ().Поскольку Россия и Украина являются крупными экспортёрами ряда сырьевых товаров на глобальные рынки, война вызвала как энергетический, так и продовольственный кризисы, вызвав сбои в цепочках поставок, рост цен и процентных ставок, а также повлияв на темпы роста мировой экономики (<https://inbusiness.kz/ru/news/vliyanie-vojny-v-ukraine-na-ekonomiku-kazahstana-chto-pomenyalos-za-dva-goda>).

Таким образом, коллокационный анализ слов корпуса дополняет частотный анализ на основе словарного подхода, может помочь выявить не только эксплицитные значения слов, но и возникающие имплицитные коннотации в тех или иных коллокациях, учитывать взаимодействие слов с их окружением, что позволяет обнаружить вторичные, скрытые, имплицитные смыслы и наведенные значения.

Заключение

Изучение манипулятивного медиадискурса с использованием методов корпусного анализа является неотъемлемой частью современных лингвистических исследований и приобретает особую актуальность в условиях современного информационного общества. Проведенное исследование показало значимость корпусного анализа для выявления манипулятивных элементов в медиадискурсе. Были созданы корпуса текстов на казахском и русском языках, отражающие медиаповестку Казахстана. Тематическое моделирование выявило основные проблемы и акценты в освещении: социальные вопросы, экономика, международные отношения и криминал. Анализ частотности негативной лексики и коллокационных связей продемонстрировал, как через использование определенных слов и контекстов создается эмоциональное воздействие, усиливающее манипулятивный эффект. Результаты подтвердили важность коллокационного анализа, так как даже нейтральные слова могут приобретать негативный оттенок. Разработанные корпуса и применение методов корпусного анализа создают основу для дальнейших исследований медиадискурса и могут применяться в системах автоматического анализа текстов для распознавания медиаманипуляции.

Информация о грантовом финансировании

Данное исследование финансировалось Комитетом по науке Министерства науки и высшего образования Республики Казахстан (грант № АР19679821 «Разработка методики распознавания медиаманипулирования и продвижения медиаграмотности в условиях информационного пространства»).

ЛИТЕРАТУРА

[1] Biber D. Representativeness in Corpus Design // *Literary and Linguistic Computing*. – 1993. – Vol. 8.4. – P. 241–257.

[2] Leech G. New resources, or just better old ones? The Holy Grail of representativeness // In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), *Corpus linguistics and the web*. – Amsterdam: Rodopi, 2007. – P. 133-149.

[3] Váradi T. The linguistic relevance of corpus linguistics // In *Proceedings of the Corpus Linguistics 2001 Conference*. UCREL Technical Papers. – 2001. – Vol. 13. – P. 587-593.

[4] Sinclair J.M. *Corpus, Concordance, Collocation*. – Oxford : Oxford univ. press, 1991. – 179 p.

[5] Lemnitzer L., Zinsmeister H. *Korpuslinguistik. Eine Einführung*. 3. Auflage. – Tübingen: Narr Verlag, 2015. – 163 S.

[6] Lemnitzer L. *Korpuslinguistik: Eine Einführung [Corpus Linguistics: Introduction]* / L. Lemnitzer, H. Zinsmeister. – Tübingen, 2006. – 213 p.

[7] Granger S. The computer learner Corpus: a versatile new source of data for SLA research // *Learner English on Computer*. – L., 1998. – 894 p.

[8] Плунгян В.А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики // *РЯНО*. – 2008. – №16 (2). – С.7-20.

[9] Амирбекова А.Б., Конырова А.Т., Кайырбекова У.С. Методы автоматизации текстов в национальном корпусе казахского языка // *Известия КазУМОИМЯ имени Абылай хана*. — 2023. Том 70. – № 3. <https://doi.org/10.48371/PHILS.2023.70.3.001>

[10] МакЭнери, Т. и Харди, А. *Корпусная лингвистика: метод, теория и практика*. – Кембридж: Издательство Кембриджского университета, 2012. *Corpus Linguistics. Method, Theory and Practice*. <http://corpora.lancs.ac.uk/clmtp/> (data obrashhenija: 09.09.2023).

[11] Baimakhanbetov M. Determination of the optimal number of topics in the LDA model when working with large arrays of text data // 2023 IEEE INTERNATIONAL CONFERENCE ON SMART INFORMATION SYSTEMS AND TECHNOLOGIES, eCF Paper Id: IEEESIST2023-176, 2023

[12] Blei D. M., Ng A. Y., Jordan M. I. Latent Dirichlet allocation // *Journal of Machine Learning Research*. – 2003. – Vol. 3. – P. 993–1022.

[13] Радбиль Т.Б. Наведенная оценочность как объект лингвокультурологического изучения: опыт корпусного анализа интернет-медиа // *Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 6: Языкознание*. – 2023. – № 3. – С. 95-105. DOI: 10.31249/ling/2023.03.07.

REFERENCES

- [1] Biber D. Representativeness in Corpus Design // *Literary and Linguistic Computing*. – 1993. – Vol. 8.4. – P. 241–257.
- [2] Leech G. New resources, or just better old ones? The Holy Grail of representativeness // In M. Hundt, N. Nesselhauf, & C. Biewer (Eds.), *Corpus linguistics and the web*. – Amsterdam: Rodopi, 2007. – P. 133-149.
- [3] Váradi T. The linguistic relevance of corpus linguistics // *Proceedings of the Corpus Linguistics 2001 Conference*. UCREL Technical Papers. – 2001. – Vol. 13. – P. 587-593.
- [4] Sinclair J.M. *Corpus, Concordance, Collocation*. – Oxford : Oxford univ. press, 1991. – 179 p.
- [5] Lemnitzer L., Zinsmeister H. *Korpuslinguistik. Eine Einführung*. 3. Auflage. – Tübingen: Narr Verlag, 2015. – 163 S.
- [6] Lemnitzer L. *Korpuslinguistik: Eine Einführung [Corpus Linguistics: Introduction]* / L. Lemnitzer, H. Zinsmeister. – Tübingen, 2006. – 213 p.
- [7] Granger S. The computer learner Corpus: a versatile new source of data for SLA research // *Learner English on Computer*. – L., 1998. – 894 p.
- [8] Plungyan V.A. Korpus kak instrument i kak ideologiya: o nekotorykh urokakh sovremennoj korpusnoj lingvistiki // *RYANO*. – 2008. – № 16 (2). – S. 7-20. [in Rus.]
- [9] Amirbekova A.B., Konyrova A.T., Kajyrbekova U.S. Metody avtomatizacii tekstov v nacional'nom korpusе kazakhskogo yazyka // *Izvestiya KaZUMOIMYA imeni Abylaj khana*. – 2023. Tom 70. – № 3. <https://doi.org/10.48371/PHILS.2023.70.3.001>
- [10] MaKEhneri, T. i Khardi, A. *Korpusnaya lingvistika: metod, teoriya i praktika*. – Kembridzh: Izdatel'stvo Kembridzhskogo universiteta, 2012. *Corpus Linguistics. Method, Theory and Practice* <http://corpora.lancs.ac.uk/clmtp/> (data obrashhenija: 09.09.2023).
- [11] Baimakhanbetov M. Determination of the optimal number of topics in the LDA model when working with large arrays of text data // *2023 IEEE INTERNATIONAL CONFERENCE ON SMART INFORMATION SYSTEMS AND TECHNOLOGIES*, eCF Paper Id: IEEESIST2023-176, 2023
- [12] Blei D. M., Ng A. Y., Jordan M. I. Latent Dirichlet allocation // *Journal of Machine Learning Research*. – 2003. – Vol. 3. – P. 993–1022.
- [13] Radbil' T.B. Navedennaya ocenochnost' kak ob'ekt lingvokul'turologicheskogo izucheniya: opyt korpusnogo analiza internet-media // *Social'nye i gumanitarnye nauki. Otechestvennaya i zarubezhnaya literatura. Seriya 6: Yazykoznanie*. – 2023. – № 3. - S. 95-105. DOI: 10.31249/ling/2023.03.07.

**МАНИПУЛЯТИВТІ БАҚ ДИСКУРСЫН ЗЕРТТЕУДЕ КОРПУСТЫҚ
ТАЛДАУ ӘДІСТЕРІН ҚОЛДАНУ**

*Шакенова М.Т.¹, Миятбекова З.У.², Джилкишиева М.С.³, Баймаханбетов М.А.⁴

^{1,2,3} «М.Әуезов атындағы Оңтүстік Қазақстан университеті» ҰАҚ
Шымкент, Қазақстан

⁴ «А. Байтұрсынов атындағы талдаудағы ұлттық зерттеу және бағалау орталығы» АҚ, Астана, Қазақстан

Андатпа. Бұл мақалада манипуляциялық медиа дискурсты зерттеуде корпусстық талдау әдістерін пайдалану қарастырылады. Үздіксіз ақпарат ағыны жағдайында өмір сүретін заманауи ақпараттық қоғам жағдайында тақырыптың өзектілігі сөзсіз, өйткені корпус лингвистикасының әдістері үлкен мәліметтер базасымен жұмыс істеуге және олардағы заңдылықтарды анықтауға мүмкіндік береді. Бұқаралық ақпарат құралдарының қоғамдық пікірді қалыптастырудағы жоғары рөлін, сондай-ақ оған ақпараттық-психологиялық ықпалын сезіну айқын факт болып табылады. Осыған байланысты бүкіл интернет кеңістігін толтырған медиамәтіндердің үлкен массивін талдаудың тиімді әдістерінің шұғыл қажеттілігі туындайды. Авторлар медиа дискурсты зерттеу контекстінде корпус лингвистикасының теориялық және әдістемелік аспектілеріне егжей-тегжейлі талдау жасайды, шағын мәтіндік корпуссты қалыптастыру тәжірибесін және оларды манипуляциялық медиа дискурсты зерттеуде пайдалануды сипаттайды. Мәтіндік корпус дегеніміз – электрондық түрдегі бірқатар параметрлер бойынша (біздің жағдайда орыс және қазақ тілдеріндегі БАҚ мәтіндері) құрылымдалған белгілі бір тілдегі жинақталған мәтіндер жүйесі. Зерттеу материалы соңғы бес жылда әртүрлі орыс тілді және қазақ тілді интернет-ақпарат көздерінен, соның ішінде желідегі жаңалықтар мен әлеуметтік желілерден жиналған медиа мәтіндер болды. Мұндай корпуссты талдау әдістерін қолдана отырып зерттеу медиа дискурсты зерттеудегі объективтілікті қамтамасыз етеді, өйткені олар статистикалық деректерге және медиамәтіндерде манипуляциялық әлеуеті бар лингвистикалық құралдарды қолдану жиілігіне негізделген. Авторлар медиамәтіндерде қолданылатын тілдік стратегияларды терең түсіну және олардың оқырман қауымға әсерін анықтау үшін бұл тәсілдің маңыздылығын атап көрсетті.

Тірек сөздер: корпусстық лингвистикасы, корпус, корпусстық талдау, медиамәтін, медиа дискурс, манипуляциялық әлеует, тоналдылық, коллокация

**APPLICATION OF CORPUS ANALYSIS METHODS IN STUDYING
MANIPULATIVE MEDIA DISCOURSE**

*Shakenova M.T.¹, Miyatbekova Z.U.², Dzhilkishieva M.S.³, Baymakhanbetov M.A.⁴

^{1,2,3} NAO “M. Auezov South Kazakhstan University”, Shymkent, Kazakhstan

⁴JSC “A. Baitursunov National Center for Research and Evaluation in Education “Taldau”, Astana, Kazakhstan

Abstract. This article applies corpus analysis methods to the study of manipulative media discourse. The relevance of the topic in the ninth modern information society, living in conditions of constant information flow, is unconditional, since the methods of corpus linguistics make it possible to work with certain databases and identify patterns in them. The obvious fact is the awareness of the high role of the media, which find themselves outside of public opinion, as well as the informational and psychological impact on it. In this regard, there is an urgent need for effective methods for analyzing a large array of medical texts that have filled the entire Internet space. The authors provide a detailed analysis of the concepts and methodological aspects of corpus linguistics in the second course of media discourse, a little experience in the formation of small text corpora and their use in conducting manipulative media discourse. A corpus of texts is understood as a structure, based on a number of parameters, of a system of collected texts in a specific language (in our case, media texts in English and Kazakh) in electronic form. The research material is media texts collected from various Russian-language and Kazakh-language Internet sources of information, including online news and social media, over the past 5 years. Nevertheless, the study of corpora using corpus analysis methods ensures objectivity when conducting media discourse, since they are based on statistical data and the frequency of use of linguistic means with manipulative potential in media texts. The authors emphasize the importance of this for a deep understanding of linguistic traditions, influence in media texts, and identify their influence on the readership.

Keywords: corpus linguistics, corpus, corpus analysis, media text, media discourse, manipulative potential, tonality, collocation

Статья поступила: 14 декабря 2023

Информация об авторах:

Шакенова Майгуль Тулегеновна – кандидат филологических наук, доцент, Южно-Казахстанский университет им. М. Ауэзова, Шымкент, Казахстан, e-mail: maigul1379@mail.ru, orcid.org/0000-0002-1329-4939

Миятбекова Зилиха Усмагамбетовна – кандидат филологических наук, доцент, Южно-Казахстанский университет им. М. Ауэзова, Шымкент,

Шакенова М.Т., Миятбекова З.У., Джилкишиева М.С., Баймаханбетов М.А.

Казахстан, e-mail: zlixamiatbekova@mail.ru, orcid.org/0000-0001-5787-3417

Джилкишиева Майра Сейтжапаровна – доктор педагогических наук, профессор, Южно-Казахстанский университет им. М. Ауэзова Шымкент, Казахстан, e-mail: maira-taraz@mail.ru, orcid.org/0000-0002-6372-8141

Баймаханбетов Мухит Абилкасимович – магистр, главный научный сотрудник, Национальный центр исследования и оценки в образовании «Талдау» имени А. Байтурсунова, Астана, Казахстан, e-mail: m.baimakhanbetov@mail.ru, orcid.org/0000-0003-4978-2704

Авторлар туралы мәлімет:

Шәкенова Майгүл Төлегенқызы – филология ғылымдарының кандидаты, доцент, М. Әуезов атындағы Оңтүстік Қазақстан университеті, Шымкент, Қазақстан, e-mail: maigul1379@mail.ru, orcid.org/0000-0002-1329-4939

Миятбекова Зилиха Усмагамбетқызы – филология ғылымдарының кандидаты, доцент, М. Әуезов атындағы Оңтүстік Қазақстан университеті, Шымкент, Қазақстан, e-mail: zlixamiatbekova@mail.ru, orcid.org/0000-0001-5787-3417

Джилкишиева Майра Сейтжапарқызы – педагогика ғылымдарының докторы, профессор, М. Әуезов атындағы Оңтүстік Қазақстан университеті, Шымкент, Қазақстан, e-mail: maira-taraz@mail.ru, orcid.org/0000-0002-6372-8141

Баймаханбетов Мұхит Әбілқасымұлы – магистр, бас талдаушы, Ахмет Байтұрсынұлы атындағы «Талдау» ұлттық зерттеулер және білімді бағалау орталығы, Астана, Қазақстан, e-mail: m.baimakhanbetov@mail.ru, orcid.org/0000-0003-4978-2704

Information about the authors:

Shakenova Maigul – Candidate of Philological Sciences, Associate Professor, M. Auezov South Kazakhstan University, Shymkent, Kazakhstan, e-mail: maigul1379@mail.ru, orcid.org/0000-0002-1329-4939

Miyatbekova Zilikha – Candidate of Philological Sciences, Associate Professor, M. Auezov South Kazakhstan University, Shymkent, Kazakhstan, e-mail: zlixamiatbekova@mail.ru, orcid.org/0000-0001-5787-3417

Jilkishiyeva Maira – Doctor of Pedagogical Sciences, Professor, M. Auezov South Kazakhstan University Shymkent, Kazakhstan, e-mail: maira-taraz@mail.ru, orcid.org/0000-0002-6372-8141

Baymakhanbetov – Master, Chief Researcher, Akhmet Baitursynuly National Centre for Research and Evaluation of Education “Taldau”, Astana, Kazakhstan, e-mail: m.baimakhanbetov@mail.ru, orcid.org/0000-0003-4978-2704