



Научно-исследовательский журнал «Modern Humanities Success / Успехи гуманитарных наук»
<https://mhs-journal.ru>
2025, № 12 / 2025, Iss. 12 <https://mhs-journal.ru/archives/category/publications>
Научная статья / Original article
Шифр научной специальности: 5.9.8. Теоретическая, прикладная и сравнительно-сопоставительная лингвистика (филологические науки)
УДК 811.521

Возможности применения корпусных методов в рамках исследований манипулятивного дискурса (на основе материалов на японском языке)

¹ Середенко В.М., ¹ Вьюнов А.С.,
¹ Военный университет имени князя Александра Невского МО РФ

Аннотация: в статье рассматриваются возможности применения корпусных методов в исследовании манипулятивного дискурса на материале японского языка. В работе подчёркивается, что использование корпусных технологий обеспечивает объективность и воспроизводимость результатов, позволяет определить частотные закономерности, эмоциональную тональность и динамику лексического выбора при освещении международных и внутривнутриполитических событий, связанных с Россией. Приводится обзор основных корпусов японского языка, созданных Национальным институтом японского языка и лингвистики (NINJAL), и анализируются их возможности для исследований в области политического и медийного дискурса. Авторы обращают внимание на ограничения существующих ресурсов: устаревшие данные, закрытый доступ к современным корпусам и отсутствие удобных пользовательских интерфейсов для анализа. Особое внимание уделяется практическим аспектам построения собственного корпуса японских медийных текстов, что обусловлено необходимостью работы с актуальными материалами. Рассматриваются доступные инструменты морфологического и синтаксического анализа японского языка.

Результаты исследования демонстрируют перспективность применения корпусных методов для выявления стратегий речевого воздействия и особенностей формирования общественного мнения в японских СМИ. Полученные данные могут быть использованы при разработке методик анализа медийного дискурса, а также при создании инструментов для исследований информационного пространства.

Ключевые слова: корпусная лингвистика, манипулятивный дискурс, японский язык, медиадискурс, лингвистический анализ, MeCab, SudachiPy, NINJAL, СМИ Японии

Для цитирования: Середенко В.М., Вьюнов А.С. Возможности применения корпусных методов в рамках исследований манипулятивного дискурса (на основе материалов на японском языке) // Modern Humanities Success. 2025. № 12. С. 69 – 74.

Поступила в редакцию: 8 июля 2025 г.; Одобрена после рецензирования: 11 октября 2025 г.; Принята к публикации: 18 ноября 2025 г.

The possibilities of applying corpus-based methods to manipulative discourse research (based on Japanese-language materials)

¹ Seredenko V.M., ¹ Vyunov A.S.,
¹ Prince Alexander Nevsky Military University Ministry of Defense of the Russian Federation

Abstract: this article examines the potential of applying corpus-based methods to the study of manipulative discourse using Japanese-language materials. The paper emphasizes that the use of corpus-based technologies ensures the objectivity and reproducibility of results, allowing us to identify frequency patterns, emotional tonality, and the dynamics of lexical choice when covering international and domestic political events related to Russia. This article provides an overview of the main Japanese-language corpora created by the National Institute of Japanese Language and Linguistics (NINJAL) and analyzes their potential for research in political and media discourse. The authors highlight the limitations of existing resources, including outdated data, restricted access to modern corpora, and the lack of user-friendly interfaces for analysis. Particular attention is

paid to the practical aspects of constructing a corpus of Japanese media texts, which is necessitated by the need to work with relevant materials. Available tools for morphological and syntactic analysis of the Japanese language are considered.

The results of the study demonstrate the potential of using corpus-based methods to identify speech influence strategies and the specifics of public opinion formation in Japanese media. The data obtained can be used in developing methods for analyzing media discourse, as well as in creating tools for researching the information space.

Keywords: corpus linguistics, manipulative discourse, the Japanese language, media discourse, linguistic analysis, MeCab, SudachiPy, NINJAL, Japanese media

For citation: Seredenko V.M., Vyunov A.S. The possibilities of applying corpus-based methods to manipulative discourse research (based on Japanese-language materials). Modern Humanities Success. 2025. 12. P. 69 – 74.

The article was submitted: July 8, 2025; Approved after reviewing: October 11, 2025; Accepted for publication: November 18, 2025.

Введение

Современная лингвистика характеризуется активным внедрением цифровых технологий и методологий обработки больших массивов данных, что открывает новые возможности для анализа речевого воздействия и манипулятивных стратегий в дискурсе. Одним из наиболее перспективных инструментов исследований в этой сфере являются корпусные методы, позволяющие выявлять закономерности языкового употребления на основе больших объемов данных. Несмотря на то, что корпусные методы широко применяются в исследованиях по английскому, русскому, японскому и другим языкам, потенциал корпусных методов в рамках анализа манипулятивного дискурса на японском языке остаётся недостаточно изученным.

Японский язык, обладающий высокой степенью контекстуальной обусловленности, сложной системой вежливости и разнообразием языковых средств выражения коммуникативно-прагматических значений, представляет особый интерес для исследования форм речевого воздействия. При этом манипулятивный дискурс в японской коммуникации нередко реализуется через эвфемизацию, недоговорённость и стратегическое использование грамматических и лексических средств, что требует точных и системных методов анализа. В настоящее время СМИ Японии активно освещают международную и внутривосточную деятельность России, а информационная политика Японии приобретает черты антироссийской направленности, о чем свидетельствуют факты, представленные в докладе МИД России «О действии (бездействии) властей Италии, ФРГ и Японии, в результате которых разрушается и фальсифицируется история, оправдывается фашизм и его пособники» [9]. Так, например, в сентябре 2025 года Генеральный секретарь правительства Японии Ёсимаса Хаяси назвал расширение санкционных мер против России «вкладом в международ-

ные усилия ради мира, в том числе в решение проблем вокруг Украины» [10].

Применение корпусных методов исследования может позволить систематизировать материалы, публикуемые в СМИ Японии, а затем провести качественный анализ: наиболее точно выявить «горячие» темы, определить эмоциональную окрашенность текстов, проследить изменение тональности лексики на основе больших объемов данных. Результаты анализа дадут возможность выявить лингвистические средства, которые используют СМИ в рамках манипулятивного дискурса, а также проследить этапы формирования и некоторые черты общественного мнения у аудитории национальных СМИ Японии по отношению к России. Таким образом, актуальность данного исследования обусловлена необходимостью:

- интеграции корпусных методов в изучение манипулятивного дискурса;
- расширения эмпирической базы за счёт анализа японского языкового материала;
- разработки методологических подходов к идентификации и описанию манипулятивных стратегий.

Материалы и методы исследований

Методы корпусной лингвистики открывают широкие перспективы для исследований манипулятивного дискурса на японском языке. Корпусная лингвистика занимается вопросами машинной обработки текстов с опорой на эмпирические данные. Она позволяет обрабатывать естественный язык и тем самым проводить исследования социальных медиа, медиадискурса, диалектов, языки меньшинств, создавать мультимодальные корпуса при помощи компьютерных средств. Достижения современной корпусной лингвистики способствуют развитию технологий искусственного интеллекта [5]. В настоящее время методы корпусной лингвистики используются в том числе для изучения коммуникативной личности политиков [6], для исследования медиаобраза государства [4], а также

для междисциплинарных дискурсивных исследований [1].

А. Цуладзе, рассматривая манипуляцию с точки зрения психологии, определяет ее как «воздействие на человека (группу людей) с целью побуждения к действиям, противоречащим его (их) собственным интересам» [7]. Т.М. Голубева, чье исследование направлено в большей степени на языковые средства манипуляции, говорит о том, что языковая манипуляция – это «вид скрытого речевого воздействия, совершаемого в интересах манипулятора и направленного на внедрение в модель мира реципиента новых знаний, мнений, отношений и/или модифицирование уже существующих посредством различного рода стратегий» [3]. В нашей работе под манипулятивным дискурсом предлагаем понимать «корпус текстовых и словарных единиц, подпадающих под определение «манипулятивный»» [2]. Под «манипулятивной» мы намерены рассматривать словарную единицу или текст, в которых просматривается «отрицательная интенциональность адресанта и скрытый (неявный для адресата) характер воздействия» [2]. По мнению И.В. Беляевой, манипулятивный дискурс занимает промежуточное положение между достоверной информацией и ложью, а к единицам манипулятивного дискурса прежде всего относятся тексты политической и коммерческой рекламы, агитационные тексты политического дискурса и телевизионные политические передачи [2]. На наш взгляд, исходя из нынешних реалий, складывающихся в информационном поле, к единицам манипулятивного дискурса стоит добавить новостные тексты СМИ, в которых также можно проследить наличие вербальных средств, отражающих как отрицательную интенциональность адресанта, так и скрытый характер воздействия на адресата.

В ходе поиска корпусных материалов для проведения исследования, мы обратились к данным с сайта Государственного института японского языка и лингвистики (National Institute of Japanese Language and Linguistics, 国立国語研究所), который предоставляет доступ к любому из представленных на нем корпусов:

1) Сбалансированный корпус современного письменного японского языка (BCCJW, Balanced Corpus of Contemporary Written Japanese, 現代日本語書き言葉均衡コーパス). В NINJAL объявили о работе над обновлением корпуса BCCJW до версии BCCJW2, которая будет включать 200 млн слов в период с 2006 по 2025 год. Выход новой версии запланирован на конец 2025 – начало 2026 года [11].

2) Корпус диалектов японского языка (COJADS, Corpus of Japanese Dialects, 日本語諸方言コーパス);

3) Корпус спонтанной речи на японском языке (CSJ, Corpus of Spontaneous Japanese, 日本語話し言葉コーパス);

4) Корпус повседневного японского языка (CEJC, Corpus of Everyday Japanese Conversation, 日本語日常会話コーパス);

5) Международный корпус японского языка как иностранного (I-JAS, International Corpus of Japanese as a Second Language, 多言語母語の日本語学習者横断コーパス);

6) Разговорный корпус Университета Нагоя (NUCC, Nagoya University Conversation Corpus, 名古屋大学会話コーパス);

7) Корпус разговоров на рабочем месте Гэн-Нити-Кэн (CWPC, Gen-Nichi-Ken Corpus of Workplace Conversation, 現日研・職場談話コーパス);

8) Веб-корпус японского языка (NWJC, NINJAL Web Japanese Corpus, 国語研日本語ウェブコーパス);

9) Корпус современного японского языка (CMJ, Corpus of Modern Japanese, 近代語のコーパス);

10) Корпус устной речи периода Сёва (SSC, Showa Speech Corpus, 昭和話し言葉コーパス);

11) Диахронический корпус японского языка (CHJ, Corpus of Historical Japanese, 日本語歴史コーパス);

12) Корпус письменной речи периодов Сёва-Хэйсэй (SHC, Showa-Heisei Corpus of written Japanese, 昭和・平成書き言葉コーパス);

13) Синхронический корпус японского языка как иностранного, изучаемого в Пекине (B-JAS, Beijing Corpus of Japanese as a Second Language, 北京日本語学習者縦断コーパス);

14) База данных слов, классифицированных по семантическому принципу (WLSP, Word List by Semantic Principles, 分類語彙表).

Вышеизложенный список корпусов японского языка представлен на английской версии сайта Государственного института японского языка и лингвистики. Если перейти на японскую версию сайта, то список будет дополнен еще 7 корпусами и поисковой системой для корпуса BCCJW. Таким образом, расширенный список будет дополнен следующими корпусами:

1) Корпус детского повседневного японского языка (CEJC-Child, 子ども版日本語日常会話コーパス);

2) Корпус изучающих японский язык как иностранный (родной корейский/китайский) (C-JAS, 中国語・韓国語母語の日本語学習者横断コーパス);

3) Открытый диахронический корпус японского языка (OpenCH, オープン CHJ);

4) Корпус естественных разговоров 1000 носителей японского языка (BTSJ, BTSJ1000 人日本語自然会話コーパス);

5) Корпус устного фольклора на языке айнов (Ainu Folklore, アイヌ語口承文芸コーパス);

6) Корпус древнеяпонского языка, подготовленный совместно Оксфордским университетом и Государственным институтом японского языка и лингвистики (ONCOJ, オックスフォード・NINJAL 上代日本語コーパス);

7) Корпус современного японского языка с семантической и синтаксической разметкой (NPCMJ, 統語・意味解析情報付き現代日本語コーパス).

8) Поисковая система NINJAL-LWP для корпуса BCCJW (NLB, NINJAL-LWP for BCCJW).

Результаты и обсуждения

Перечисленные корпуса японского языка представляют бесценную базу для всевозможных исследований японского языка методами компьютерной лингвистики. Поскольку наше исследование сосредоточено, в основном, на анализе новостных материалов, публикуемых национальными СМИ Японии, среди которых NHK, Asahi и Sankei, из перечисленных корпусов для нас подходят только два: «Сбалансированный корпус современного письменного японского языка (BCCJW)» и «Веб-корпус японского языка (NWJC)». Однако, когда нам потребовались данные этих корпусов, то мы столкнулись с рядом

проблем: в настоящее время последние данные в BCCJW датированы 2008 годом, а NWJC с 29 февраля 2024 года может быть использован только корпоративными клиентами после подписания соглашения о совместных исследованиях. Эти проблемы и обусловили необходимость формирования собственного корпуса для проведения исследования.

Формирование корпуса данных требует предварительной подготовки. В настоящее время существует немало инструментов, позволяющих работать с данными на японском языке, включая инструменты на основе искусственного интеллекта (далее – ИИ). Однако при работе с корпусными данными исследователь точно знает, к какой сфере деятельности принадлежит анализируемый текст, при работе же с ИИ выдаваемый результат не позволяет его сопоставить с первоисточником [8]. Поэтому для детального анализа текстов с сохранением метаданных, формирование традиционных корпусов остается незаменимой процедурой. Для обработки текстов на японском языке, а также преобразования текстов в полноценные корпуса существует немало инструментов. Такие инструменты делятся на те, которые требуют навыков работы с языками программирования, и те, которые обладают собственным графическим или веб-интерфейсом и подобных навыков не требуют (табл. 1).

Таблица 1

Инструменты для работы с корпусными данными на японском языке.

Table 1

Tools for working with Japanese corpus data.

	Инструменты, требующие навыков программирования	Инструменты, не требующие навыков программирования
1.	MeCab	Chunagon
2.	SudachiPy	Shonagon
3.	Janome	Tatoeba
4.	Juman++/ KNP	Kyoto Corpus Viewer
5.	CaboCha	Sketch Engine

Наиболее популярным и известным инструментом для работы с японским текстом в компьютерной лингвистике является MeCab (Matrix-based Evaluator of Candidate Boundaries). Данный инструмент был разработан Тадао Кудо в Национальном институте информатики Японии. В настоящее время используется для морфологического анализа и токенизации японского текста. Он способен автоматически определять границы слов и делить строку на токены, сопровождая их лингвистическими признаками. Для каждого слова MeCab указывает часть речи, базовую форму и

чтение. При работе с MeCab могут быть использованы различные словари: IPADic (подходит для общего анализа), UniDic (более современный и совместим с BCCJW), NEologd (расширенный словарь с новыми словами, включая имена, неологизмы и интернет-лексику). Существенным минусом данного инструмента является отсутствие удобного пользовательского интерфейса, его использование в настоящее время возможно только через командную строку.

SudachiPy представляет собой современный морфоанализатор на языке Python, разработанный

для более точной токенизации с учетом контекста. Он поддерживает три уровня разбиения (А, В, С) – от мелких морфем до устойчивых выражений. Включает собственный словарь SudachiDict, который постоянно обновляется. SudachiPy часто используется в сочетании со spaCy и GiNZA для построения полных NLP-пайплайнов.

Janome – полностью реализованная на Python библиотека для морфологического анализа, не требующая установки внешних зависимостей. Подходит для небольших проектов и образовательных задач. Выдает сегментацию, базовую форму и часть речи, что делает её хорошей отправной точкой для начинающих. Работает медленнее, чем MeCab или Sudachi, но проще в установке и интеграции.

Juman++/KNP – системы морфологического и синтаксического анализа, разработанные Киотским университетом. Juman++ выполняет сегментацию и определение частей речи, а KNP – анализ синтаксических зависимостей. Используется в исследовательских проектах и корпусах вроде Kyoto Corpus. Требуется базовых навыков работы с командной строкой и файловыми форматами вывода.

CaboCha представляет собой парсер зависимостей, построенный поверх MeCab, определяет синтаксическую структуру предложений. Использует модель машинного обучения для выделения грамматических связей между словами. Применяется в корпусных и синтаксических исследованиях. Выводит результаты в виде дерева зависимостей, удобного для последующего анализа.

Chuunagon – платный онлайн-интерфейс для работы с корпусами, созданный на базе BCCWJ и других коллекций. Shonagon – бесплатный аналог. Оба инструмента позволяют искать слова, словоформы, части речи и анализировать контексты. Имеют понятный интерфейс и подходят для пользователей без навыков программирования. Поддерживают экспорт результатов поиска в таблицы. Работают с корпусами, представленными на сайте NINJAL.

Tatoeba – веб-платформа с параллельными предложениями на различных языках, включая японский. Удобна для контрастивного анализа,

перевода и обучения языку. Пользователь может искать по ключевым словам, языковым парам или грамматическим структурам. Не требует установки и навыков программирования.

Kyoto Corpus Viewer – графический интерфейс для работы с Kyoto Text Corpus, синтаксически размеченным корпусом японского языка. Позволяет визуализировать структуры зависимостей, просматривать примеры и выполнять фильтрацию по частям речи. Часто используется в учебных целях. Не требует написания кода, но предоставляет глубокие лингвистические данные.

Sketch Engine – коммерческая платформа для корпусной лингвистики с поддержкой японского языка. Позволяет искать слова, строить коллокации и частотные списки. Имеет интуитивный интерфейс и используется в академических и прикладных исследованиях. Не требует знаний кода, поддерживает сложные поисковые шаблоны, но требует регистрации на сайте и подписки.

Выводы

Результаты анализа доступных корпусов и инструментов для проведения исследования японского медиадискурса позволили выявить ряд проблем:

- корпусы текстов японского языка, находящиеся в открытом доступе, не позволяют провести исследование наиболее актуальных материалов. Под «актуальными» мы рассматриваем материалы, не старше 5 лет;
- корпусы на японском языке, датируемые 2020 годом, найденные на платформе Sketch Engine, но для доступа к ним требуется оформление платной подписки;
- отсутствие пользовательского интерфейса у большинства инструментов для подготовительной работы с данными на японском языке значительно усложняет проведение исследований в этой области;
- существует необходимость создания собственных корпусов с текстами на японском языке и пользовательских интерфейсов для вспомогательных инструментов в целях проведения исследований на основе актуальных материалов.

Список источников

1. Аракелова А.Р., Ширяева Т.А. Дискурс отмены и аксиология: как и почему СМИ вербализуют ценностные концепты // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2025. Вып. 1 (895). С. 17 – 24.
2. Беляева И.В. Феномен речевой манипуляции: лингвоюридические аспекты: автореф. ... дис. д-р. филол. наук. Ростов-на-Дону, 2009. 39 с.
3. Голубева Т.М. Языковая манипуляция в предвыборном дискурсе: на материале американского варианта английского языка: автореф. дис. ... канд. филол. наук. Нижний Новгород, 2009. 22 с.

4. Махракова А.В. Анализ современных тенденций в исследованиях медиаобраза государства средствами корпусных и нейросетевых компонентов // Современное педагогическое образование. 2025. № 6. С. 400 – 405.
5. Микаллеф Л.О. Лингвистика нейросетей как парадигма современной науки о языке // Мир науки, культуры, образования. 2025. № 1 (110). С. 467 – 469.
6. Мухортов Д.С., Жовнер Е.А. Методология изучения коммуникативной личности политика // Вестник Российского университета дружбы народов. Серия: Теория языка. Семиотика. Семантика. 2025. Т. 16. № 1. С. 278 – 290. <https://doi.org/10.22363/2313-2299-2025-16-1-278-290>
7. Цуладзе А. Большая манипулятивная игра. М.: Алгоритм, 2000. 336 с.
8. Peter Crosthwaite, Vit Baisa Generative AI and the end of corpus-assisted data-driven learning? Not so fast! // Applied Corpus Linguistics. 2023. Vol. 3 Issue 3. URL: <https://doi.org/10.1016/j.acorp.2023.100066>
9. Интернет-портал // Российская газета URL: <https://rg.ru>
10. Интернет-портал // РИА Новости URL: <https://ria.ru>
11. Интернет-портал // National Institute for Japanese Language and Linguistics URL: <https://www.ninjal.ac.jp>

References

1. Arakelova A.R., Shiryaeva T.A. Discourse of cancellation and axiology: how and why the media verbalize value concepts. Bulletin of Moscow State Linguistic University. Humanities. 2025. Issue 1 (895). P. 17 – 24.
2. Belyaeva I.V. The phenomenon of speech manipulation: lingualegal aspects: Abstract of a PhD thesis in Philology. Rostov-on-Don, 2009. 39 p.
3. Golubeva T.M. Language manipulation in pre-election discourse: Based on the American version of English: Abstract of a PhD thesis in Philology. Nizhny Novgorod, 2009. 22 p.
4. Makhrakova A.V. Analysis of Current Trends in Research of the Media Image of the State Using Corpus and Neural Network Components. Modern Pedagogical Education. 2025. No. 6. P. 400 – 405.
5. Micallef L.O. Neural Network Linguistics as a Paradigm of Modern Language Science. The World of Science, Culture, Education. 2025. No. 1 (110). P. 467 – 469.
6. Mukhortov D.S., Zhovner E.A. Methodology for Studying the Communicative Personality of a Politician. Bulletin of the Peoples' Friendship University of Russia. Series: Language Theory. Semiotics. Semantics. 2025. Vol. 16. No. 1. P. 278 – 290. <https://doi.org/10.22363/2313-2299-2025-16-1-278-290>
7. Tsuladze A. The Great Manipulative Game. Moscow: Algorithm, 2000. 336 p.
8. Peter Crosthwaite, Vit Baisa. Generative AI and the End of Corpus-Assisted Data-Driven Learning? Not So Fast! Applied Corpus Linguistics. 2023. Vol. 3 Issue 3. URL: <https://doi.org/10.1016/j.acorp.2023.100066>
9. Internet portal. Rossiyskaya Gazeta URL: <https://rg.ru>
10. Internet portal. RIA Novosti URL: <https://ria.ru>
11. Internet portal. National Institute for Japanese Language and Linguistics URL: <https://www.ninjal.ac.jp>

Информация об авторах

Середенко В.М., кандидат филологических наук, старший преподаватель, Военный университет имени князя Александра Невского МО РФ, ichi210@mail.ru

Вьюнов А.С., преподаватель, Военный университет имени князя Александра Невского МО РФ, viunart@yandex.ru

© Середенко В.М., Вьюнов А.С., 2025