

УДК 811.161.1'367.6:811.581'367.6(045)

DOI: 10.18413/2313-8912-2026-12-2-0-2

Баркович А. А.<sup>1</sup>  
Ван Ц.<sup>2</sup>

Квалитативная и квантитативная специфика  
корпусной репрезентации частеречной  
категориальности русского и китайского языков

<sup>1</sup> Белорусский государственный университет  
пр-т Независимости, 4, Минск, 220004, Беларусь  
E-mail: [barkovichaa@gmail.com](mailto:barkovichaa@gmail.com)  
ORCID: 0000-0001-8469-8431

<sup>2</sup> Белорусский государственный университет  
пр-т Независимости, 4, Минск, 220004, Беларусь  
E-mail: [qing.wang@mail.ru](mailto:qing.wang@mail.ru)  
ORCID: 0009-0005-2248-256X

Статья поступила 30 сентября 2025 г.; принята 15 июня 2026 г.;  
опубликована 30 июня 2026 г.

**Аннотация:**

**Актуальность.** Данная статья посвящена анализу корпусной репрезентации частей речи в русском и китайском языках как объективному и актуальному ресурсу лингвистических исследований. Опыт обобщения в работе широкого спектра данных и метаданных является релевантным как для осмысления теоретических основ описания языка, так и для практико-ориентированного использования в прикладных междисциплинарных разработках; в дальнейшем он может быть продуктивно имплементирован в разноаспектную исследовательскую практику и развит. Научная значимость работы определяется тем, что корпусно-верифицированное сопоставление частеречных систем типологически удалённых языков позволяет не только уточнить статистические параметры их функционирования, но и аргументированно поставить вопрос о лингвистической состоятельности дискуссионных категорий – предикатива в русском языке и дифференцирующего слова в китайском.

**Проблематика.** И для русского, и для китайского языков проблематика совершенствования лингвистических категорий оказывается одинаково актуальной, что подтверждается модернизацией созданных для них частеречных систем в реалиях корпусной репрезентации. Такого рода данные позволяют интерпретировать частеречную категориальность как идентифицирующий и квалифицирующий признак метаязыковой практики.

**Методы.** Полученные посредством интегрированных в лингвистические корпуса компьютерных программ данные позволяют на материале принципиально стабильной институции частей речи преодолеть шаблонность и саморефлексию традиционных лингвистических описаний языка. В исследовании был задействован комплексный методологический подход, в том числе, квалитативный и квантитативный анализ, метод корпусных исследований.

**Результаты.** Подтверждена целесообразность метаязыкового позиционирования частеречной категориальности как лингвистической универсалии. В статье – по данным «Национального корпуса русского языка» и «Онлайн-корпуса» китайского

языка – рассмотрены квалитативные и квантитативные особенности функциональности языковых единиц в контексте их частеречной принадлежности. **Выводы.** Полученные корпусные данные о шаблонном реальноречевом функционировании единиц того или иного частеречного кластера в типологически далеких языках убедительно свидетельствуют о целесообразности классификации частей речи как метаязыкового инструментария. В свою очередь, проблематика развития частеречных систем русского и китайского языков является не менее лингвистически актуальной, что подтверждается насущностью выделения новых разрядов частей речи и их фактическим инкорпорированием в корпусную аннотацию. В данной связи показательно, что на основании корпусной репрезентации подтверждена лингвистическая состоятельность таких категорий, как предикатив в русском языке и дифференцирующее слово в китайском. Полученные результаты относятся прежде всего к высокочастотному сегменту лексики обоих языков. Они могут быть применены в практике корпусной аннотации, сопоставительной грамматике и преподавании русского и китайского языков как иностранных.

**Ключевые слова:** Корпусная репрезентация; Часть речи; Квантитативный анализ; Квалитативный анализ; Русский язык; Китайский язык; Частеречная категориальность

**Информация для цитирования:** Баркович А. А., Ван Ц. Квалитативная и квантитативная специфика корпусной репрезентации частеречной категориальности русского и китайского языков // Научный результат. Вопросы теоретической и прикладной лингвистики. 2026. Т. 12. № 2. С. 25–60. DOI: 10.18413/2313-8912-2026-12-2-0-2

УДК 811.161.1'367.6:811.581'367.6(045)

DOI: 10.18413/2313-8912-2026-12-2-0-2

Aliaksandr A. Barkovich<sup>1</sup>  
Qing Wang<sup>2</sup>

Qualitative and quantitative specificity of corpus  
representation of part-of-speech categoriality in Russian  
and Chinese

<sup>1</sup> Belarusian State University,  
4 Nezavisimosty Ave., Minsk, 220004, Belarus  
E-mail: [barkovichaa@gmail.com](mailto:barkovichaa@gmail.com)  
ORCID: 0000-0001-8469-8431

<sup>2</sup> Belarusian State University,  
4 Nezavisimosty Ave., Minsk, 220004, Belarus  
E-mail: [qing.wang@mail.ru](mailto:qing.wang@mail.ru)  
ORCID: 0009-0005-2248-256X

*Received 30 September 2025; accepted 15 June 2026; June 30 March 2026*

**Abstract:**

**Relevance.** This article analyzes corpus-based representations of parts of speech in Russian and Chinese as an objective and relevant resource for linguistic research. The experience of generalization of a wide range of data and metadata is relevant both for understanding the theoretical foundations of language description and for practice-

oriented use in applied interdisciplinary research. It can be productively implemented and further developed in multifaceted research. The scientific significance of the work is determined by the fact that a corpus-verified comparison of part-of-speech systems of typologically distant languages allows not only to clarify the statistical parameters of their functioning, but also to argue for the question of the linguistic validity of the discussion categories – the predicative in Russian and the differentiating word in Chinese.

**Problems.** Improving linguistic categories is equally relevant for both Russian and Chinese, as evidenced by the modernization of the parts-of-speech systems created for them in the context of corpus representation. Such data enables us to interpret part-of-speech as a feature that identifies and qualifies metalinguistic practice.

**Methods.** Data obtained through computer programs integrated into linguistic corpora enable us to overcome the commonplaceness and self-reflective nature of traditional linguistic descriptions of language using material from a fundamentally stable system of parts of speech. The study employed a comprehensive methodological approach, including qualitative and quantitative analysis and corpus-based technique.

**Results.** The feasibility of the metalinguistic positioning of part-of-speech categoricity as a linguistic universal is confirmed. Using data from the Russian National Corpus and the Online Corpus of Chinese, this article examines the qualitative and quantitative features of the functionality of linguistic units in the context of their part-of-speech affiliation.

**Conclusions.** The obtained corpus data on the units' real-speech functioning stereotypeness of a particular part-of-speech cluster in typologically distinct languages convincingly demonstrates the feasibility of classifying parts of speech as a metalinguistic tool. In turn, the development of part-of-speech systems in Russian and Chinese is no less linguistically pressing, as evidenced by the urgency of identifying new categories of parts of speech and their actual incorporation into the corpus annotation. In this regard, it is significant that the linguistic validity of such categories as the predicative in Russian and the differentiating word in Chinese was confirmed based on corpus representation. The knowledge gained primarily relates to the high-frequency segment of the vocabulary of both languages. The obtained results can be applied in the practice of corpus annotation, comparative grammar and teaching Russian and Chinese as foreign languages.

**Keywords:** Corpus representation; Part of speech; Quantitative analysis; Qualitative analysis; Russian; Chinese; Part-of-speech categorization

**How to cite:** Barkovich, A. A., Wang, Q. (2026). Qualitative and quantitative specificity of corpus representation of part-of-speech categoricity in Russian and Chinese, *Research Result. Theoretical and Applied Linguistics*, 12 (2), 25–60. DOI: 10.18413/2313-8912-2026-12-2-0-2

## 1. Введение

Современная лингвистика, благодаря компьютерным технологиям, приобрела новые возможности в верифицированной репрезентации структуры и функциональности языка как феномена (Barkovich, 2019). Это предполагает, в том числе, корпусную интерпретацию фундаментальных лингвистических

институций. Можно по-разному относиться к их обоснованности, но они созданы и востребованы: «Очевидная безграничность межъязыкового разнообразия породила два крайних подхода к языковым универсалиям. С одной стороны, следуя Аристотелю, Соссюру и Блумфилду в русле вариативности, можно идеализировать

многообразии, вплоть до полного отказа от универсалий. С другой стороны, разделяя увлеченность Платона, Соссюра и Хомского единообразием, можно игнорировать очевидные несоответствия как поверхностные, эфемерные, эпифеноменальные или воображаемые – принадлежащие эйдосу, языку, компетенции – и постулировать чрезвычайно сложные и абстрактные (или формальные) универсалии» (Givón, 1984 : 24).

К общепризнанным и основополагающим универсалиям языка, безусловно, относится и частеречная система. Трудно переоценить актуальность референтных данных о ней в изучении природы языка, особенно в формализации получаемых данных. В развитии системы знаний о языке соответствующие исследования играют важную роль, значимость которых в научной интерпретации современной коммуникации по-прежнему высока. Традиционно фундаментальной в обеспечении стабильной метаязыковой институциональности является частеречная система (Алпатов, 2018; Розенталь и др., 2010; Сунь; Ли, 1993 и др.). При этом, она призвана отражать логику функциональности весьма динамичного объекта моделирования, языка. Конечно, «... каждый язык имеет свою собственную схему» частей речи, характеризующуюся определенными «количеством, природой и необходимыми границами» (Sapir, 1921 : 119). Вместе с тем, изучение частеречной специфики каждого языка способствует общей систематизации релевантных знаний, как каждый фрагмент мозаики улучшает ее визуализацию. При этом сопоставление метаданных о типологически удаленных языках имеет особую ценность. Изучение лингвистической специфики разноструктурных языков является продуктивным, поскольку, пусть при этом анализируются и типологически достаточно далекие языковые системы, –

но они функционируют в общих для всех языков коммуникативных обстоятельствах и являются элементами единого и все более взаимосвязанного интеръязыкового континуума. Более того, поиск общих закономерностей развития типологически далеких языков очевидно целесообразен и по причине высокой вероятности того, что они могут быть экстраполированы и на описание множества более близких языковых систем.

Первые шаги по систематизации постепенно накапливающихся знаний о частях речи были сделаны античными мыслителями. Изначально в фокусе их внимания, естественно, оказался древнегреческий язык; позже внимание ученых концентрировалось на латыни, как языке науки, актуальном весьма протяженный период времени. Позже влияние античных грамматиков воплотилось в созданных по их образцу грамматических системах европейских языков – в них было инкорпорировано и адаптированное к их специфике учение о частях речи. По этому же пути пошли и грамматисты многих неевропейских языков, в том числе, – пусть и с некоторым опозданием – китайские ученые. Подобным же образом зародилась частеречная концепция для русского языка, ориентировавшаяся на уже апробированные шаблоны адаптации классической модели в западноевропейских языках. В русском языке части речи были системно идентифицированы М.В. Ломоносовым в его труде «Российская грамматика», впервые изданном в 1755 г. (Ломоносов, 1855). Для китайского языка система частей речи, как значимая новация, была представлена в работе «Объяснение правил письменного языка господина Ма» («马氏文通»), изданной в 1898 г. (Ma, 1983). Частеречная категориальность в описаниях русского и китайского языков прошла долгий путь становления и является до настоящего времени

дискуссионной. Доступные сегодня посредством корпусных методик данные позволяют приблизить функциональность частеречной системы к современным реалиям коммуникации.

Как металингвистическое представление о классах языковых единиц, *частеречная категориальность* характеризуется рядом семантически, формально и функционально значимых кодифицирующих признаков. Это не означает их полного подобия: в разных языках, особенно разноструктурных, совокупность таких признаков может быть разной, ее параметры и организация могут обладать выразительной уникальностью. В частности, для русского языка характерна флективность и развитая морфология, отражающая специфику многочисленных форм рода, числа, падежа, времени и т. д. В свою очередь, в контексте китайского языка важнейшую нагрузку несет синтаксическое позиционирование языковых единиц, по сути, организующее семантическую архитектуру текста. Лингвистически обоснованная дифференциация языковых единиц типичным образом оказывается дискуссионной на статистически репрезентативном материале: «В этом континууме нет четкой дифференциации ни по одной категории» (Ху, 2018 : 208). Тем не менее, частеречная система совершенно определенно является универсалией, свойственной научным представлениям о возможностях лингвистического моделирования языка как феномена, в целом.

Научная новизна настоящего исследования обеспечена как теоретическим обоснованием синтетического, количественно и качественно сбалансированного подхода к анализу репрезентативных языковых данных, так и практической реализацией методологически инновационной логики комплексного лингвистического исследования. В работе предложена методика формирования пропорционально

сопоставимых эталонных корпусов для изучения типологически удалённых языков на базе материала, изначально существенно различающегося по объёму. В соответствии с данным подходом, в процессе исследования было выполнено системное количественно-качественное сопоставление частеречных систем русского и китайского языков, обеспеченное референтными эталонными корпусами. На основе корректно агрегированных и интерпретированных корпусных данных была сформирована аргументационная база для единообразной параллельной оценки частеречного потенциала предикатива в русском и дифференцирующего слова в китайском языках.

## 2. Методы и материал

Практически любой естественный язык требует сложноорганизованной метаязыковой надстройки, что нередко обуславливает ее существенную вариативность. В данной связи, очевидна уместность и реалистичность корпусно-обеспеченного подхода к оценке фундаментальных грамматических проблем – в востребованном на сегодня дескриптивно-прескриптивном формате. Присущая ему достоверная фактография оказывается не только количественно, но и качественно продуктивной. Конечно, она не всегда подтверждает складывавшиеся долгие годы теоретические стереотипы, но зато позволяет уверенно решать прикладные задачи метаописания языка – особенно актуальные в контексте тотальной информатизации коммуникации. Вместе с тем, корпусный подход никак не отменяет сложившиеся шаблоны лингвистической практики: «Интуиция может быть чрезвычайно хорошим проводником при работе с корпусом. Проще говоря, интуиция лингвиста может предоставить ему основу для обоснованных догадок, позволяющую быстро и эффективно взаимодействовать с корпусом. Аналогичным образом, уже существующие теоретические подходы

могут помочь корпусному лингвисту» (McEnery and Hardie, 2012 : 161).

Количественная характеристика языковой специфики является лингвистически валидной и верифицируемой (Баркович, Ван, 2015). Статистика частотности языковых единиц, полученная из репрезентативных корпусов, таких как «Национальный корпус русского языка» («НКРЯ») и «Онлайн-корпус» китайского языка обладает высокой убедительностью для решения широкого круга лингвистических задач (НКРЯ; Онлайн-корпус). В частности, фактологически обеспеченный анализ частей речи по корпусным данным позволяет сформировать надежный фундамент для оценки их релевантности существующим грамматическим стереотипам (см., напр., Цао, 2011; Национальный корпус; Савчук и др., 2024 и др.). Качественная интерпретация полученной из репрезентативных корпусных источников статистики органично дополняется метаязыковым анализом эмпирически убедительных первичных данных. Такого рода методологический синкретизм полностью оправдан при рассмотрении фундаментальных основ языкознания, позволяя синтезировать новое знание. Частеречная категориальность – традиционно значимая лингвистическая база всех грамматических надстроек языка, что подтверждается ее интеграцией в метаописания таких типологически удаленных языков как русский и китайский. Корпусно-обеспеченный анализ современной речевой практики – целесообразная реализация современных возможностей лингвистической практики, в том числе, углубленного сопоставительного изучения разноструктурных языков.

Цель данного исследования – выявление и описание лингвистически релевантной качественной и количественной специфики корпусной репрезентации частеречной

категориальности русского и китайского языков в высокочастотном регистре.

Объектом исследования являются части речи русского и китайского языков, рассматриваемые как на уровне их качественных характеристик, так и на уровне их количественных показателей по данным корпусов.

Релевантными в данной связи задачами являются следующие:

1) обобщить сложившиеся в языкознании теоретические представления о сущности и особенностях систем частей речи для русского и китайского языков;

2) посредством корпусного инструментария получить объективную количественную информацию о фактическом распределении единиц русского и китайского языков по частеречным разрядам;

3) охарактеризовать качественную корреляцию распределения языковых единиц по частям речи с их функциональной спецификой в речевой практике;

4) соотнести данные о статистической дифференциации единиц русского и китайского языков разной частеречной принадлежности с имеющимися качественными обобщениями;

5) на основании корпусных данных оценить модификационный потенциал частеречных систем русского и китайского языков, в том числе, их метаязыковую сбалансированность.

Цель и задачи исследования обусловили задействование сбалансированного метаязыкового подхода с применением комплекса методов: обобщения, моделирования, структурного, контекстуального, качественного, количественного, компонентного, корпусного анализа, сопоставления.

Эмпирической базой исследования послужили следующие источники материала: русскоязычная специфика выявлялась по «Основному корпусу»

«Национального корпуса русского языка», китайскоязычная – по «Корпусу современного китайского языка» («现代汉语语料库») «Онлайн-корпуса». Оба ресурса являются достаточно репрезентативными: объем привлеченного в качестве базового русскоязычного материала превысил 389 млн. словоупотреблений, объем китайскоязычного – превысил 20 млн. словоупотреблений. Частеречная разметка в вышеупомянутых корпусах различается: «НКРЯ» оснащен полноценной морфологической аннотацией статистического характера; тогда как «Онлайн-корпус» воплощает преимущественно функциональный подход по модели, основанной на правилах, – что обусловлено типологической спецификой китайского языка, в частности, отсутствием в нем морфологического словоизменения. Жанровый состав корпусов также не является идентичным. Тем не менее, качество аутентичного материала задействованных корпусов позволило решить поставленные в исследовании задачи. Количественно объем «НКРЯ» существенно превышает объем «Онлайн-корпуса». Это обстоятельство было учтено в ходе исследования: целесообразное сопоставление проводилось на основе полностью сопоставимых данных эталонных корпусов, сформированных по единому шаблону репрезентации материала. Изначальная диспропорция объемов анализируемого материала в задействованных корпусах была нивелирована путем агрегации эталонных корпусов, составленных из 500 наиболее востребованных единиц для каждого языка. Данный порог был выбран как оптимальный с точки зрения баланса между репрезентативностью и управляемостью материала: он охватывает высокочастотный сегмент лексики, обеспечивающий анализ значимых статистических закономерностей, и, вместе с тем, допускает детальный

квалитативный анализ каждой единицы. Оба сформированных эталонных корпуса использовались как сопоставимые по объёму и принципам отбора базы данных, которые были проанализированы в соответствии с задачами исследования. В рамках исследования единицами материала являлись лексемы русского языка и сегментированные слова ( $\vec{w}$ ) китайского языка, идентифицируемые с учётом их частеречной разметки. Омонимичные и полифункциональные единицы (особенно применительно к китайскоязычному материалу) анализировались с учётом контекста и их преобладающей синтаксической функции в текстах выборки. Прошедшие автоматическую сегментацию и частеречную разметку омонимичные и полифункциональные единицы учитывались отдельно для каждого частеречного значения в соответствии с разметкой корпуса. Так, если одна и та же графическая форма аннотировалась в качестве репрезентанта разных частей речи, такие вхождения фиксировались как отдельные частеречные реализации. Для русскоязычного материала более детализированная и репрезентативная дополнительная выборка не понадобилась. Для углубленного анализа функциональности единиц китайского языка дополнительно на базе «Онлайн-корпуса» была сформирована расширенная выборка из 16 254 наиболее частотных единиц. Такой порог выборки обеспечил возможность анализа размеченных как частеречные классы совокупностей языковых единиц в составе не менее 50 контекстов. Показатель в 16 254 единиц был установлен экспериментальным путём как минимальное значение, при котором корпус-менеджер «Онлайн-корпуса» сформировал репрезентативную выборку, включившую все частеречные классы, в том числе не представленные либо недостаточно представленные в базовой выборке из 500 единиц. При этом, полученный в результате статистически

актуальный набор единиц с полной частеречной разметкой позволил избежать избыточной корпусной выборки, которая могла бы затруднить целесообразный анализ материала. Данная ситуация обусловлена большим знаковым разнообразием китайскоязычной графики, но на достоверность выводов для обоих языков не повлияла. Опционально идентичность исследованных языковых единиц сверялась с «Толковым словарем русского языка» и «Словарем современного китайского языка» (Ожегов и Шведова, 2006; Словарь современного китайского языка, 2016).

В целом, реализованная исследовательская стратегия позволила обеспечить необходимый уровень репрезентативности материала и объективно охарактеризовать специфику лингвистической интерпретации частеречной принадлежности языковых единиц русского и китайского языков. Полученные результаты позволили не только охарактеризовать совокупность референтных языковых единиц, но и задействовать их метаописание для формирования аргументированной позиции по возможной модернизации частеречных систем русского и китайского языков.

### 3. Результаты

Для того чтобы всесторонне изучить системы частей речи в русском и китайском языках, в исследовании потребовалось не только тщательно проанализировать квантитативные данные, извлечённые из языковых корпусов, но и обобщить полученные данные в квалитативном ключе, опираясь на имеющуюся теоретическую базу. Задействование встроенного в корпусные ресурсы аналитического инструментария полностью отвечает современному тренду лингвистики: «Поскольку тексты становятся онлайн доступными во всё возрастающих объёмах и на всё большем количестве языков, растёт потребность в надёжных методах обработки, которые

могут анализировать текст без дорогой и трудоёмкой адаптации к новым областям и жанрам» (Schütze, 1995 : 141). И «НКРЯ», и «Онлайн-корпус» в достаточной мере обеспечены актуальными на сегодняшний день аутентичными образцами речевой продукции, что позволяет оценить текущие реалии языкового функционирования. Вместе с тем, как показало проведенное исследование, статистика требует интерпретации и лишь в этом случае может обладать верифицированной научной ценностью, будучи совместимой с уже имеющимися теоретическими наработками. Реализованный в данном исследовании подход позволил получить новые сведения о том, как функционируют языковые единицы, как они взаимодействуют между собой и какими метаязыковыми свойствами обладают. Исследование дало возможность на статистически убедительном материале охарактеризовать распределение языковых единиц по частям речи в каждом из двух языков и сделать выводы относительно перспектив модификации частеречных систем для данных языков. В фокусе исследования оказались категории предикатива в русском языке и дифференцирующего слова в китайском языке, состоятельность частеречного потенциала которых была, в целом, подтверждена в синтетическом квантитативно-квалитативном аспекте.

#### 3.1. Квалитативная специфика частей речи русского и китайского языков

Обеспеченность языковой практики тем или иным набором частей речи убедительно свидетельствует о присутствии ее разновидностям определенном спектре модальности. Соответствующая квалитативная специфика, будучи референтной качественным аспектам языкового функционирования, в равной мере присуща и русскому, и китайскому языку, что позволяет ее интерпретировать в контексте данного исследования параллельно. Корпусные данные

ориентированы на языковую практику, которая и должна служить основанием широких метаязыковых обобщений: «В концепции, ориентированной на языковую практику, где грамматика напрямую основана на языковом опыте, нет типов данных, которые исключаются из рассмотрения, поскольку они считаются отражающими реализацию...» (Vybee, 2010 : 11). В данном аспекте, например, в случае насыщенности речи существительными и прилагательными, целесообразно предполагать статичность и описательность ее содержания. В свою очередь, насыщенность текстов глагольными формами позволяет уверенно констатировать динамику и хронологическое развитие содержания речевых фрагментов.

Рассмотрим, как это может быть проиллюстрировано и интерпретировано применительно к проблематике данного исследования. В частности, характерным для **модальности действительности / недействительности** оказывается наличие в контексте русскоязычных текстов определенных **прилагательных** или **наречий**. Так, фактически в русском языке соответствующая модальность зачастую маркируется единицами, прямо указывающими на оценку субъектом 'реальности' / 'нереальности' описываемой им ситуации. Например, прилагательные, специфичные для таких контекстов, можно условно разделить на две группы. Прилагательные, характерные модальности действительности подчёркивают 'реальность', 'достоверность', 'фактичность' описываемых событий: *реальный, действительный, подлинный, истинный, фактический, достоверный* и др. Прилагательные **модальности недействительности** указывают на 'нереальность', 'иллюзорность', 'вымышленность' описываемых событий: *мнимый, кажущийся, иллюзорный, фиктивный, ложный, воображаемый*, и др. Многие из таких прилагательных

имеют антонимичную пару. Наречия модальности действительности / недействительности, подобно соответствующим прилагательным, также, обладают ярко выраженной семантикой: *действительно, истинно, фактически* и др., с одной стороны, и *якобы, словно, вряд ли* и др. – с другой стороны. Факультативным маркером такой модальности выступают специфичные для оцениваемых ситуаций **глаголы**. Например, в русском языке:

*Преодолеть инстинкт, заставляющий животных спасаться бегством от истинной или мнимой опасности, невозможно.<sup>1</sup>*

*Она кажется слишком правдоподобной, и новички вряд ли бы стал интересоваться, существует ли на самом деле Любезность.*

Для китайского языка характерен аналогичный набор языковых средств. Здесь данную функцию выполняют преимущественно наречия (**副词**) и прилагательные (**形容词**), причём граница между ними зачастую размыта, а их частеречная идентичность зависит от контекста. Так, для формирования модальности действительности ключевыми являются следующие языковые единицы: *确实* ('действительно', 'в самом деле'), *的确* ('воистину', 'точно'), *真实* ('истинно'), *果然* ('и в самом деле', 'как и ожидалось'). О модальности недействительности свидетельствуют такие языковые единицы, как *似乎* ('кажется', 'будто'), *好像* ('как будто', 'словно'), *仿佛* ('словно', 'как будто'), *虚假* ('ложно'). При переводе данные единицы китайского языка в большинстве случаев передаются наречиями русского языка. Например:

<sup>1</sup> Здесь и далее русскоязычные примеры приводятся по данным «НКРЯ».

现在越来越多的材料证明，美洲大陆确实是在公元五世纪的时候被中国人首先发现的。<sup>2</sup> (Всё больше материалов свидетельствует о том, что американский континент действительно был впервые открыт китайцами в V веке нашей эры).

他似乎不知道这件事，但眼神出卖了他。(Будто он ничего не знал об этом, но взгляд его выдал).

那位收费的人看见这情况，似乎也有点怜悯之情，过来看了看孩子，说：“病得确实不轻！鼻子都有点扇了！”(Кассир, увидев это, словно проникся состраданием, подошёл посмотреть на ребёнка и сказал: “Болен действительно серьёзно – даже нос начал увеличиваться!”).

В последнем примере примечательно сочетание в одном контексте языковых единиц **确实** и **似乎** – выступающих как наречия. Здесь маркеры модальности действительности и модальности недействительности соседствуют, создавая сложную модальную семантику.

Важное типологическое различие между русским и китайским языками состоит в том, что в китайском модальность действительности / недействительности выражается исключительно лексически – через наречия и прилагательные, тогда как в русском языке существенным является частое присутствие в соответствующих контекстах глаголов в сослагательном наклонении. Это делает участвующие в формировании модальности действительности / недействительности прилагательные и наречия особенно значимыми для формирования модальности в китайской речи.

<sup>2</sup> Здесь и далее китайскоязычные примеры приводятся по данным «Онлайн-корпуса».

Подобным образом, сопровождаются специфичной глагольной лексикой контексты, которым присуща **потенциальная модальность** (с семантикой ‘возможности’ / ‘невозможности’); **дебитивная модальность** (с семантикой ‘необходимости’ и ‘долженствования’); **гипотетическая модальность** (с семантикой ‘предположительности’); **интенциональная модальность** (с семантикой ‘намерения’); **императивная модальность** (с семантикой ‘побудительности’); **оптативная модальность** (с семантикой ‘желательности’). Идентифицируемые – в плане присутствия в них единиц определенной частеречной категориальности – конструкции, содержащие глаголы, характерны для множества контекстов, фактически модально маркированных тем самым. Это, в равной мере, присуще и русскоязычным, и китайскоязычным контекстам. Например, в русском языке:

**Необходимо отметить, что несмотря на некоторые опасения со стороны правоохранительных органов, массовое мероприятие прошло очень спокойно, никаких серьезных инцидентов зафиксировано не было.**

**В целом можно сказать, что новые банковские системы денежных переводов работают по тем же принципам, что и гранды рынка переводов MG и WU.**

В китайском языке:

**基本民生的意义应该也包含扩大社会保障, 比如提高医疗费用报销比例, 扩大失业保险的覆盖面或者工伤保险覆盖面等。** (А обеспечение основного народного благосостояния должно включать в себя расширение социальной защиты, например, повышение оплаты гражданских медицинских расходов, расширение страхования от безработицы или производственного травматизма).

**据了解, 要将壁画完整地搬到电脑里, 前期需要采集海量信息, 后期数据拼接**

、整合、存储等一系列工作都需要人工手段完成。(Как стало известно, для того, чтобы перенести полную фреску в компьютер, прежде всего, **необходимо собрать** огромный объем реальных данных, затем специалистам **нужно проделать** работу по соединению элементов фрески и сохранению соответствующих данных в ручном режиме).

Для **персонализированной** речи свойственна **«местоименная»** насыщенность речевых фрагментов – при описании субъектности коммуникантов, характеристике их отношения к тому или иному действию или его признаку. Содержание речевых фрагментов может дифференцировано проецироваться посредством всего спектра местоимений. Для русского языка это, прежде всего личные местоимения, с помощью которых презентуется, в частности, позиция адресанта – посредством единиц *я* и *мы* (местоимения первого лица); позиция адресанта – посредством единиц *ты* и *вы* (местоимения второго лица); позиция наблюдателя – посредством единиц *он*, *она*, *оно* и *они* (местоимения третьего лица). Для китайского языка для соответствующих коммуникативных ситуаций актуален аналогичный перечень языковых средств: это местоимения *我* ('я') и *我们* ('мы'), как элементы категории местоимений первого лица, или **第一人称代词**; это *你* ('ты') и *您* ('вы') / *你们* ('вы'), как элементы категории местоимений второго лица, или **第二人称代词**, и местоимения *他* ('он'), *她* ('она'), *它* ('оно') и *他们*, *她们*, *它们* ('они'), как элементы категории местоимений третьего лица, или **第三人称代词**.

Например, в русском языке:

Из *тех, кто может свою судьбу перевернуть* и из *тех, кто плывет по течению*. Вот *ты* смог и *я* смог. А *Коля тот*, у него даже мечты нет!

А *ты* думаешь, *они* умеют? – хмыкнул Двинятин.

В китайском языке:

他说:

“**我**特别喜欢中华传统文化, 和中国朋友一起过春节非常开心, **我**还亲手写了福字给他们拜年, 祝**他们**幸福安康诸事顺遂。”  
(«**Мне** очень нравится китайская традиционная культура. **Я** неописуемо рад возможности отпраздновать Праздник Весны с китайскими друзьями. Чтобы поздравить китайских друзей с Новым годом, **я** написал китайский иероглиф «Фу», который должен принести **им** счастье, благополучие и удачу», – сказал юный житель Благовещенска).

你明天有事吗? 你能跟我们两一起去

◦ (У **тебя** завтра есть дела? **Ты** можешь пойти вместе с **нами**?).

В последнем примере использовано два местоимения: дважды *你* ('ты') и единожды *我们* ('мы'). При этом, здесь единица *我们* не включает в группу субъектов сообщения адресата: адресант под сочетанием иероглифов *我们* имеет в виду себя и постороннего человека (не собеседника). По этой причине используется именно комбинация иероглифов *我们*, а не, например, сочетание *咱们*, использование которого подразумевало бы включенность собеседника в состав субъектов, обозначаемых местоимением как *мы*.

**Временная** актуализация текста, в частности, при его ассоциации с определенными временными ориентирами, сопровождается своими специфическими **наречиями** и **предлогами**. Типичным образом описываемые события привязываются к моменту речи, а востребованными маркерами временной отнесенности коммуникационной ситуации выступают, в частности, такие единицы русского языка как *после*, *сейчас*, *до* и т. д. В китайском языке типичными в

подобных ситуациях языковыми единицами являются *已经* ('уже'), *刚刚* ('как только'), *当* ('в', 'на'), *从* ('с', 'от') и др. Например, в русском языке:

*Записать, не откладывая. Нет, потом, не сейчас. А иногда, ты знаешь, я просто боюсь начать писать о тебе.*

*Когда соус закипит, в него влить картофельный крахмал, разведенный мадерой (при вливании нужно размешивать соус лопаткой во избежание образования комков), и довести до кипения, после чего процедить через частое сито.*

В китайском языке:

*等我赶到医院的时候,他已经做完手术了,正在恢复室里休息。(Когда я добрался до больницы, ему уже сделали операцию и он отдыхал в палате восстановления).*

*她刚刚走出门口,电话就响了。*

*(Как только она вышла за дверь, тут же зазвонил телефон).*

В первом из вышеприведенных контекстов языковая единица *已经* решает задачу временной актуализации, формируя таксисное отношение предшествования: действие (завершение операции) совершилось раньше, чем наступил момент отсчёта времени (приезд субъекта). Далее уже частица *了* как показатель завершённости в постпозиции к глаголу конкретизирует это значение. Во втором контексте языковая единица *刚刚* также актуализирует событие – как только что завершившееся; далее частица *就* вводит следующее событие. В обоих контекстах наречия выступают как типичные средства временной актуализации. При этом, их использование компенсирует отсутствие в китайском языке глагольного времени как грамматической категории: там, где в русском языке используются видо-временные формы глагола, в китайском

задействуются лексические средства наподобие *已经* и *刚刚* – в сочетании с частицами.

**Локальная** актуализация той или иной коммуникационной ситуации может быть идентифицирована по фактам задействования в речи определенного ассортимента **наречий** и **предлогов**. Сложно представить, что носитель языка осознанно имплементирует в речевой поток какие-либо языковые единицы на основе их частеречной принадлежности, выбор языковых средств коммуникантами происходит интуитивно – у них, обычно, просто нет времени на подобные изыски. Тем не менее, сам факт использования в речи единиц определенного свойства, в том числе, соотносимо их ассоциированности с какой-либо частью речи достаточно красноречиво свидетельствует о выраженной таким образом модальности сообщения. В русском языке типичным образом о локальной актуализации содержания речевого фрагмента позволяет судить достаточно компактный набор референтных единиц, таких как *вверху*, *внизу*, *далеко*, *близко*, *здесь*, *там* и т. д. Для китайского языка не менее характерны свои языковые маркеры, принадлежащие, по определению, к узкому кругу наречий или предлогов – такие как *下面* ('под, внизу'), *上面* ('на, над, вверху'), *那里* ('там'), *这里* ('здесь') и др. Например, в русском языке:

*Там они вершат – здесь они отдыхают, совершенно инкогнито, в том смысле, что их никто и не пытается узнать.*

*Не знаю, близко ль, далеко ль, не знаю, В какой стране и при луне какой, Веселая, забытая, родная, Звучала ты, как песня за рекой.*

В китайском языке:

*伍月这时还算懂事,马上配合他说:*

*"他明天一早的飞机,要到西安参加书市。"*

但在桌子下面踢了严守——脚。严守吓了一跳，急忙把脚收了回来。(У Юэ благоразумно сразу же подыграла ему: «Он летит завтра ранним утром в Сиань, на книжную ярмарку». Но под столом она пнула Янь Шоуи. Тот вздрогнул и торопливо отдернул ногу назад).

大厦前脸有个院落，院落门口有座牌坊；沿着牌坊，拉着警戒线；警戒线处，有五六个门卫守着，不让人进。(Перед зданием находился дворик с аркой на входе, вдоль которой тянулась заградительная лента; у ограждения стояло человек пять-шесть охранников, никого не пропуская).

В первом примере присутствует 到 – предложный показатель направления ('в'); 在 – предложный показатель места ('под') дополнен локативным показателем со значением 'внизу' – 下面. Во втором – языковая единица 沿着, совмещающая функции наречия и предлога, вводит обстоятельство места с оттенком направленности ('вдоль чего-либо'); показатель 着 из ее состава указывает на 'длительное, протяжённое действие или положение'.

Таким образом, качественное обобщение языковых данных о присутствии тех или иных частей речи в тексте объективно свидетельствует о характере и специфике модальности и может быть описано посредством лингвистически корректных моделей.

### 3.2. Квантитативная специфика частей речи русского и китайского языков

При всей результативности качественного анализа речи посредством корпусной методики, не менее лингвистически значимым оказывается её квантитативный анализ на материале лингвистических корпусов. Корпусные данные о функциональности тех или иных частей речи обладают вполне определенной ценностью для уточнения самых разнообразных параметров и структуры языкового материала:

«Квантитативный анализ частей речи позволяет выявить не только стилистические, но и диахронические и жанровые закономерности» (Francis and Kučera, 1982 : 198). Они актуальны как при метаязыковом исследовании отдельных текстов, так и при оценке всего спектра языковых отношений на базе речевой практики. Как объективная проекция языковой системы, речь, в конечном итоге, отражает все действительно релевантные языку как объекту научного представления возможности его рефлексии. Извечная проблема заключается в том, чтобы их выявить и интерпретировать. Корпусный формат в данной связи является апробированным и надежным инструментарием: посредством его уже решались достаточно сложные метаязыковые задачи (Захаров, Богданова, 2011; Partee, 1987; Gries, 2009; Wang et al., 2025 и др.). Наряду с широким кругом насущных лингвистических задач, корпусы, благодаря так называемой POS-разметке, как компоненту морфологической аннотации, референтны и в статистической репрезентации закономерностей строения и функционирования системы частей речи. Ввиду их высокого типологического своеобразия, в контексте данного исследования системы частей речи для русского и китайского языков на базе корпусного материала целесообразно рассматривать дифференцированно.

#### 3.2.1. Интерпретация квантитативной специфики данных русского языка

Для получения лингвистически специфицированных данных о сравнительной статистике встречаемости языковых единиц русского языка в «Основном корпусе» «НКРЯ» был сформирован соответствующий запрос относительно 500 наиболее востребованных языковых единиц. Наличие в данном ресурсе разноплановой разметки изначально позволяло рассчитывать на интерпретацию

полученных данных в аспекте их грамматической идентичности – морфологическая разметка является базовой и основной практически для всех лингвистически компетентных корпусных ресурсов. Опираясь на верифицированную репрезентативность корпусных данных «НКРЯ», охватывающих широкий спектр речевой практики, из полученных данных был сформирован эталонный корпус языковых единиц. Анализ полученных

данных проводился по двухаспектной методике: оценивалась как востребованность лексем той или иной части речи, так и востребованность словоупотреблений единиц с привязкой к их частеречной идентичности. Данные были представлены как в абсолютном (в целых числах), так и в относительном (процентном), или удельном, выражении; перечень частей речи был представлен в виде рейтинга (Таблица 1).

**Таблица 1.** Сравнительная частотность единиц эталонного корпуса, представляющих разные части речи русского языка

**Table 1.** Comparative frequency of units in the reference corpus representing different parts of speech in the Russian language

Перечень наиболее распространенных в эталонном корпусе частей речи	Статистика выявленных лексем в абсолютном выражении	Статистика выявленных лексем в процентном выражении	Статистика выявленных словоупотреблений в абсолютном выражении	Статистика выявленных словоупотреблений в процентном выражении
существительное	161	32,53 %	27 481 950	13,47 %
глагол	99	20,00 %	23 250 959	11,40 %
наречие	57	11,51 %	12 661 920	6,20 %
прилагательное	48	9,70 %	7 167 531	3,51 %
местоимение	39	7,88 %	43 790 885	21,46 %
предлог	29	5,86 %	39 875 883	19,54 %
частица	20	4,04 %	14 341 685	7,03 %
союз	17	3,43 %	30 433 604	14,92 %
числительное	14	2,83 %	2 685 379	1,32 %
предикатив	5	1,01 %	1 297 353	0,64 %
вводное слово	5	1,01 %	715 656	0,35 %
междометие	1	0,20 %	330 245	0,16 %
ИТОГО	495*	100 %	204 033 050	100 %

\* Из общего объема наиболее частотных единиц русского языка (500) были исключены цифры (5).

Согласно данным эталонного корпуса наивысшие показатели частотности – как совокупность единиц определенной части речи – продемонстрировали **существительные** русского языка. Их совокупный удельный вес среди всего количества идентифицированных корпус-менеджером «НКРЯ» языковых единиц составил 32,53 %, что в абсолютном

выражении составило 161 лексему. Наиболее употребительной единицей из их числа оказалась лексема **год**, ее частотность составила 1 111 801 словоупотреблений – около 4 % от общего количества существительных. Например:

*Даже если на знакомство с каждым из этих ресурсов Сети затрачивать не более минуты, то на просмотр пятой*

части доступных русских серверов уйдет целый год рабочего времени.

В 1996 году повесть была опубликована в журнале «Октябрь».

Если принимать в расчет общее количество соответствующих определенной части речи словоупотреблений, самой востребованной из них следует признать *местоимение* – с показателем в 21,46 %. Из ассортимента местоимений наиболее востребованной единицей оказалось *я* – с показателем в 4 998 377 случаев использования, что составило более 11 % от общего объема соответствующей совокупности. Например:

*Поэтому считаю, что бороться надо на всех фронтах, в том числе и личными примерами – я, например, тоже не курю, никогда не курил, выступаю за здоровый образ жизни и всех к этому призываю.*

*Я не раз приводил аргументы, что он своими поправками пытается спихнуть груз ответственности на работников...*

Согласно полученным данным, была выявлена речевая реализация и пока не являющихся «официальными» частями речи предикатива и вводного слова. Наличие отдельных типов разметки в корпусе подтверждает реальность их потенциального позиционирования как частей речи. Языковые единицы вышеупомянутых категорий обнаруживаются в эталонном корпусе, являясь значительно более распространенными, чем, например, входящие в «академический» перечень частей речи русского языка междометия.

Однако очевидно, что необходимо принимать в расчет не только количественные показатели той или иной категории языковых единиц. При качественном рассмотрении функциональности в речи, в частности, *вводного слова*, выявляется его существенная парадигматически-синтагматическая ограниченность. Так, вводные слова совпадают по форме с

единицами других частей речи и не случайно обособляются на письме запятыми. Согласно сложившимся представлениям, вводным словом называется «слово, не связанное формально-грамматически с другими словами в предложении, не являющееся членом предложения и выражающее, например, отношение говорящего к высказываемому, отношение данной мысли к предыдущей или последующей речи или указывающее на то, что говорящий считает высказанную мысль чужой» (Ахманова, 1966: 69). Не обладая собственной содержательной значимостью, вводное слово является своеобразным комментарием к содержанию фрагмента речи, отражает эмоциональное состояние и прагматические установки говорящего. Фактически, оно не несет какой-либо существенной грамматической нагрузки в предложении – той среде, в контексте которой типичным образом проявляется грамматическая идентичность языковой единицы как части речи. Вся совокупность количественно-качественных характеристик вводного слова не дает достаточных оснований констатировать его соответствие частеречному статусу.

В свою очередь, проведенный анализ речевой функциональности *предикатива* достаточно убедительно подтверждает потенциал его частеречной категориальности. Представленность предикатива оказывается вполне репрезентативной: в количественном измерении предикатив занимает девятое место в общем рейтинге размеченных «НКРЯ» словоупотреблений – как репрезентантов частеречных категорий языковых единиц. В числе 500 наиболее частотных лексем в эталонном корпусе обнаруживается небольшой ассортимент предикативов – только пять: *можно, нельзя, нет, мало, нужно*. Например:

*Неужели нужно выделять кого-то из длинного ряда исполнителей, подчеркивать, как далась ему роль, или*

вообще сравнивать образы с историческими портретами?

Времени то *нет*, да и вообще понимаешь ли, эмансипация.

Вместе с тем, статистика говорит о высокой востребованности предикатива в речи: всего пять предикативов в квантитативном измерении дали суммарно 1 297 353 единицы в рейтинге частотности. Это, само по себе, веское основание для констатации существенного частеречного потенциала предикатива. При этом, квалитативный анализ речевой функциональности контекстов с присутствием предикатива не менее убедительно свидетельствует о его синтагматической состоятельности и парадигматической идентичности. По совокупности квалитативно-квантитативных характеристик подтверждается, в целом, обоснованность позиционирования предикатива как части речи. В грамматическом плане предикатив отличается от наречий прежде всего своей синтаксической функцией: он выступает исключительно в роли главного члена безличного предложения и не может выступать в позиции обстоятельства, типичной для наречий. В отличие от кратких прилагательных, предикативы не имеют полной формы и не изменяются по родам. От «категории состояния» в

традиционной трактовке предикатив отличается тем, что включает единицы, выражающие не только состояние (*нельзя, можно*), но и оценочную квалификацию (*нужно, мало*) и бытийную констатацию (*нет*). Устойчивость данных синтаксических и семантических признаков подтверждает правомерность выделения предикатива в самостоятельный частеречный класс.

### 3.2.2. Интерпретация квантитативной специфики данных китайского языка

В запросе корпусному менеджеру «Онлайн-корпуса» по поиску референтных данных в подкорпусе «Корпус современного китайского языка» целесообразным оказалось задание порога репрезентативности квантитативной выборки в 2 000 словоупотреблений. Такое ограничение позволило агрегировать действительно статистически значимые языковые единицы в релевантной «Статистической таблице по частотности» («词频统计表»). В свою очередь, полученные таким путем данные послужили базой эталонного корпуса для решения задач исследования: он был сформирован на базе 500 наиболее частотных языковых единиц корпусного ресурса (Таблица 2).

**Таблица 2.** Сравнительная частотность единиц эталонного корпуса, представляющих разные части речи китайского языка

**Table 2.** Comparative frequency of units in the reference corpus representing different parts of speech in the Chinese language

Перечень наиболее распространенных в эталонном корпусе частей речи	Статистика выявленных лексем в абсолютном выражении	Статистика выявленных лексем в процентном выражении	Статистика выявленных словоупотреблений в абсолютном выражении	Статистика выявленных словоупотреблений в процентном выражении
существительное	159	32,32 %	838 322	17,62 %
глагол	128	26,02 %	926 571	19,48 %
наречие	47	9,55 %	421 379	8,86 %
местоимение	37	7,52 %	465 729	9,79 %
предлог	30	6,10 %	365 330	7,68 %

союз	29	5,89 %	279 053	5,87 %
прилагательное	25	5,08 %	151 956	3,19 %
числительное	16	3,25 %	173 406	3,65 %
частица	15	3,05 %	1 057 440	22,23 %
счетное слово	6	1,22 %	77 637	1,63 %
ИТОГО	492**	100 %	4 756 823	100 %

\*\* Из общего объема наиболее частотных единиц китайского языка (500) были исключены цифры (8).

Корпусный анализ статистически репрезентативного материала позволил получить лингвистически информативные данные о распространенности тех или иных частей речи в китайском языке. Данные о специфике устной и письменной речи могут отличаться, но в корпусном масштабе и устная, и письменная речь достаточно объективно отражают метаязыковую идентичность языка, являясь надежным материалом для лингвистических обобщений. Полученные данные о фактическом речевом функционировании китайского языка не противоречат статистике русского языка – здесь также самыми распространенными являются языковые единицы, выполняющие номинативную функцию. Для китайского языка анализ речевой обусловленности языковых единиц особенно актуален: одна и та же единица может использоваться, например, и как существительное, и как глагол. Основанная на синтаксической функции языковых единиц корпусная разметка успешно решает возникающие при лингвистическом анализе дилеммы подобного рода. Так, по данным эталонного корпуса наиболее востребованы в китайском языке единицы в роли **существительных**, общий объем которых составил 32,32 % от общего количества исследованных языковых единиц. Этот показатель весьма близок данным русской речи – 32,53 %. Однако в китайском языке обнаруживаются свои приоритеты словоупотребления: здесь самым распространенным существительным оказалась единица **中** ('середина'), использованная 40 105 раз. Например:

他站在屋子的**中间**。(Он стоял посередине комнаты).

Помимо того, что иероглиф **中** может выступать в роли разных частей речи, его функциональность как частицы тоже достаточно разнообразна. Например, такая частица может быть локативным показателем (например, 'внутри') или аспектуальным показателем (например, 'в процессе'):

他把这件事深深地埋在**心中**·从不对人提起。(Он глубоко похоронил это в сердце и никогда никому не рассказывал).

谈判目前仍在**进行中**·双方尚未达成协议。(Переговоры в настоящее время всё ещё **продолжаются**, стороны пока не достигли соглашения).

В первом примере **中** – постпозитивный локативный показатель после существительного **心** ('сердце') – переводится предлогом *в*. Во втором примере **中** – постпозитивный аспектуальный показатель после единицы **进行** ('проводить', 'вести') – переводится глаголом *продолжаются*.

При этом многие содержащие единицу **中** контексты при переводе требуют достаточно сложных переводческих трансформаций. Например:  
 值得一提的是,在满洲里口岸此次**进口煤猛增中**,中国国有企业成为主力军。  
 (Стоит отметить, что на фоне стремительного увеличения объёмов ввозимого через Маньчжурию угля

главными потребителями стали китайские госпредприятия).

Отличается от реалий русского языка и китайскоязычная статистика о распространенности частей речи по количеству словоупотреблений. Здесь наиболее распространены *частицы* – в русском языке местоимения. В китайской речи частицы заняли 22,23 % от общего объема словоупотреблений. Наиболее распространенной единицей оказалась *的* – она была использована 744 863 раз. Она может определять, в том числе, специфику ‘принадлежности’. Например:

*这是我的书。* (Это моя книга).

В данном контексте частица *的* указывает на принадлежность предмета (книги) и точно соответствует русскому притяжательному местоимению *мой / моя / моё*, например в модели «чей-либо + существительное». В отличие от русского языка, где принадлежность выражается флексией или отдельным словом, в китайском языке *的* – единственный универсальный показатель соответствующих отношений, – не изменяющийся ни по роду, ни по числу, ни по падежу. Вообще, частица *的* – одна из самых многофункциональных единиц китайского языка. Она может выступать в нескольких совершенно разных ролях: как структурный элемент (атрибутивная связка): *我的书* – ‘моя книга’; как номинализатор: *红的* – ‘то, что красное’; как финализатор (в конце предложения для акцентирования): *他会来的* – ‘он придёт точно’; как локализатор, выделяя обстоятельство в составе конструкции *是...的*: *他是昨天来的* – ‘он приехал вчера’.

Подобные ситуации достаточно типичны в самых разных языках: «В языках, в которых независимые и зависимые формы различаются по парадигматической структуре, парадигмы независимых маркеров, как правило, богаче парадигм зависимых маркеров» (Siewierska, 2004 : 112). Характерно, что

такие содержащие *的* контексты, как например следующий, не могут переводиться на русский язык дословно:

*用于测试无人交通的路段位于北京中关村北部地区。* (Основной участок для проверки систем автономного вождения расположен в пекинском районе Чжунгуаньцунь).

Согласно информации, содержащейся в «Словаре современного китайского языка», являющегося, в свою очередь, официальным нормативным изданием Института языкознания Академии общественных наук КНР, в китайском языке 12 частей речи. Однако в сделанной по данным «Онлайн-корпуса» выборке – для наиболее востребованных 500 лексем – корпусным менеджером были идентифицированы единицы 10 частей речи. Вместе с тем, этот показатель характеризует лишь базовую выборку из 500 наиболее частотных единиц. Из числа «академически признанных» частей речи в выборку эталонного корпуса не попали звукоподражательные слова и междометия. Это не значит, что их нет в речи: для них выполнена разметка и они находятся в корпусных текстах при расширенном поиске. Расширенная выборка из 16 254 единиц (см. ниже) демонстрирует иное распределение частей речи, включая категории, отсутствующие в базовой выборке.

Например, корпус содержит контексты со следующими *междометиями* *哎呀* (‘Ах!’) и *哎哟* (‘Ой!’):

*哎呀，原来是你！* (Ах, оказывается, это ты!).

*哎哟！我肚子好疼！* (Ой! Как у меня болит живот!).

*到了咖啡厅，喝了一杯咖啡，严守一似乎突然想起什么，看了看表：*

*“哎哟，都十一点半了，我下午一点还得录像，该化妆去了。”* (В кафе они выпили по чашке кофе и тут Янь Шоуи, словно опомнившись, взглянул на часы: *Ой*, уже

половина двенадцатого. У меня в час съемка, еще нужно забежать в гримерку).

Междометие *哎呀* (āiyā) выражает 'неожиданное узнавание', 'приятное удивление'. Его ближайший русский эквивалент – *Ах!* Междометие *哎哟* (āiyō) выражает 'боль', 'физическое страдание' или 'досаду'. Его ближайший русский эквивалент – *Ой!* Строго говоря, в китайском языке нет полных аналогов *Ах!* и *Ой!* и функцию этих языковых единиц распределяют между собой междометия *啊* (a), *哦* (o), *哎呀* (āiyā) и *哎哟* (āiyō) – в зависимости от передаваемой эмоции: 'удивления', 'боли', 'восхищения' или 'узнавания'.

**Звукоподражательные слова** при увеличении порога выборки также в корпусе находятся. Например, это единицы *呼啦呼啦* (hūlāhūlā), *啪* (pā), *哐当* (kuāngdāng):

*冷风呼啦啦扫动着满地的枯叶纸屑，*

*更显出景象格外的凄冷。* (Холодный ветер с шелестом-хлопаньем «шурх-шурх» мёл по земле сухие листья и обрывки бумаги, отчего картина выглядела ещё более мрачной и холодной).

*啪！鞭子一挥，马儿立刻飞奔起来。*

(«Хлоп!» – кнут взвился, и лошадь тут же понеслась вперёд).

*“哐当”一声巨响，这么静的夜，我身体陡然一惊，心头提到嗓子边了。*

(Раздался оглушительный «бум!», и в такой тихой ночи я вздрогнул всем телом, аж сердце чуть не выпрыгнуло из груди).

*他把门 “哐当”*

*一声关上，才听到屋里传来沈雪像狼一样的嚎叫，接着是她痛哭的声音。* (Он захлопнул дверь – «бабах!» – и лишь тогда услышал, как из комнаты донёся вой Шэнь Сюэ, похожий на волчий, а следом – звуки её горького плача.)

Звукоподражательное слово *呼啦呼啦* выражает 'шум ветра', 'хлопанье флага', 'шелест крыльев', 'повторяющийся

хлопок' – передаёт протяжный повторяющийся звук. По-русски естественнее всего оно передается как *шурх-шурх*, *хлоп-хлоп* или *ишиш*. Звукоподражание *啪* означает 'резкий хлопок', 'выстрел', 'удар рукой' или 'падение предмета' – звук резкого однократного удара. По-русски: *хлоп*, *щёлк* или *бах*. Звукоподражательное слово *哐当* имитирует 'глухой удар по металлу или дереву', 'однократный резкий грохот', например, звук захлопывания двери или падения металлического предмета. По-русски ближе всего: *бабах* или *бум*. В последнем примере звукоподражательное слово *哐当* сопровождается единицей *一声*, означающей 'один звук' 'один раз', то есть вся конструкция "*哐当*" *一声* означает буквально 'с одним звуком «бабах»'. И это типичная для китайского языка модель формирования структуры предложения: *глагол + звукоподражательное слово + 一声 + результат действия*.

Вместе с тем, спектр аннотированных, но не попавших в массив наиболее востребованных категорий единиц китайского языка (для 500 единиц), не исчерпывается междометиями и звукоподражательными словами. Подобным им образом, для текстов корпуса предусмотрена аннотация такого знакового для китайского языка разряда лексики, как **дифференцирующее слово**. В качественном аспекте функциональность относящихся к данной категории языковых единиц уникальна и их использование в речи носит системный характер. В том числе, они идентифицируются в контекстах «Онлайн-корпуса». Например, здесь встречается дифференцирующее слово *女* ('женский'):

*教育女职工正确处理家庭关系上海闸北发电厂把教育女职工正确处理家庭关系作为建设精神文明的一个重要内容，对这方面的先进人物和先进事迹，进行广泛宣*

传·树立榜样。  
 (Шанхайская электростанция рассматривает обучение сотрудников правильному выстраиванию семейных отношений как важную часть духовно-нравственного воспитания. Передовые работницы и их достижения в этой области широко освещаются, а их примеры служат образцом для подражания).

Данный фрагмент текста с разметкой, будет выглядеть так:

教育/v 女/f 职工/n 正确/a 处理/v 家庭/n 关系/n 上海闸北发电厂/ni 把/p 教育/v 女/f 职工/n 正确/a 处理/v 家庭/n 关系/n 作为/v 建设/v 精神文明/n 的/u 一个/mq 重要/a 内容/n , /w 对/p 这/r 方面/n 的/u 先进/a 人物/n 和/c 先进/a 事迹/n , /w 进行/v 广泛/a 宣传/v , /w 树立/v 榜样/n 。/w.

Отличаясь от близкого к нему по форме прилагательного собственной функциональной идентичностью, дифференцирующее слово значимо как класс языковых единиц при рассмотрении многих теоретических и практических вопросов языкознания. Выделение такого рода единиц средствами корпусной разметки достаточно показательно. В текстах «Онлайн-корпуса» они находятся как специфическая категория, благодаря тегу *f*, позволяющему уверенно отличать дифференцирующие слова от близких к ним по атрибутивной значимости прилагательных. Так, в разметке приведенного выше предложения дифференцирующее слово *女* ('женский') как раз и снабжено специальным тегом *f*, который обеспечивает определяемому им иероглифу *职工* ('сотрудник' / 'работник') – помеченному тегом *n* как существительное – соответствующую коннотацию. Присвоенная данной языковой единице уже как *女职工*

('сотрудница' / 'работница') итоговая кодировка имеет вид *女/f 职工/n*. Аналогичным образом ведут себя и другие единицы данного класса. Так, например, дифференцирующее слово *野生* ('дикий') в сочетании *野生动物* ('дикое животное') идентифицировано как *野生/f* (в совокупности – *野生/f 动物/n*); единица *初级* ('начальный') в сочетании иероглифов *初级阶段* ('начальная стадия') получает разметку *初级/f* (в совокупности – *初级/f 阶段/n*) и т. д. Во всех подобных случаях тег *f* последовательно присваивается определяемому компоненту, что подтверждает системность данной категории в разметке корпуса.

В общем, в «Онлайн-корпусе китайского языка» используются следующие теги: *n* – 'нарицательное существительное'; *nt* – 'темпоратив'; *nd* – 'существительное со значением направления (или директив)'; *nl* – 'локатив'; *nh* – 'собственное имя'; *nhf* – 'фамилия'; *nhs* – 'имя'; *ns* – 'географическое название'; *nn* – 'название нации'; *ni* – 'название организации или учреждения'; *nz* – 'другое собственное имя'; *v* – 'глагол'; *vd* – 'глагол направления'; *vl* – 'глагол-связка'; *vi* – 'модальный глагол'; *a* – 'прилагательное'; *f* – 'дифференцирующее слово'; *m* – 'числительное'; *q* – 'счетное слово'; *d* – 'наречие'; *r* – 'местоимение'; *p* – 'предлог'; *c* – 'союз'; *u* – 'частица'; *e* – 'междометие'; *o* – 'звукоподражательное слово'; *j* – 'аббревиатура'; *w* – 'знак препинания'; *ws* – 'некитайский символ'; *wi* – 'другой неизвестный символ'. В частности, дифференцирующее слово здесь присутствует как категория метаописания и обозначается символом *f*, позволяя отграничить данные единицы от, например, прилагательных, обозначаемых символом *a*. Фактически, данный тегсет позволяет выполнять аннотацию

китайскоязычных текстов достаточно полно, что может быть проиллюстрировано следующими примерами.

服装<sup>n</sup> 裁剪<sup>v</sup> 是<sup>vl</sup> 一<sup>m</sup> 项<sup>q</sup> 较<sup>d</sup> 复杂<sup>a</sup> 的<sup>u</sup> 技术<sup>n</sup> , /w 初学者<sup>n</sup> 一般<sup>a</sup> 应从<sup>v</sup> 简单<sup>a</sup> 裁剪<sup>v</sup> 开始<sup>v</sup> , /w 可以<sup>vu</sup> 先<sup>d</sup> 学<sup>v</sup> 剪裁<sup>v</sup> 女<sup>f</sup> 裤<sup>n</sup> 。 /w (Кройка одежды является достаточно сложной технологией, как правило, начинают с простого, например с кройки **женских штанов**).

会上<sup>nl</sup> , /w 全国人大<sup>j</sup> 内务<sup>n</sup> 司法<sup>v</sup> 委员会<sup>n</sup> 副<sup>f</sup> 主任委员<sup>n</sup> 邹瑜作<sup>nh</sup> 了<sup>u</sup> 内务<sup>n</sup> 司法<sup>v</sup> 委员会<sup>n</sup> 关于<sup>p</sup> 最高人民法院<sup>ni</sup> 工作<sup>n</sup> 报告<sup>n</sup> 审议<sup>v</sup> 结果<sup>n</sup> 的<sup>u</sup> 报告<sup>n</sup> , /w 关于<sup>p</sup> 最高人民法院<sup>ni</sup> 工作<sup>n</sup> 报告<sup>n</sup> 审议<sup>v</sup> 结果<sup>n</sup> 的<sup>u</sup> 报告<sup>n</sup> 。 /w (На совещании **заместитель председателя Комиссии по внутренним делам и юстиции Всекитайского собрания народных представителей** Зоу Юй делал доклады о результатах рассмотрения Комиссией по внутренним делам и юстиции деятельности Верховного народного суда и о результатах деятельности Верховной народной прокуратуры).

В вышеприведенных примерах единицы 女 ('женский') и 副 ('заместитель') непосредственно сочетаются с существительными 裤 ('штаны') и 主任委员 ('председатель').

Характерно, что при позиционировании языковых единиц, обозначающих признаки предмета, – их местоположение в сочетании с дифференцирующими словами оказывается инвариантным – перед ними. Однако, при достаточно типичном добавлении частицы 的 место дифференцирующего слова в предложении

может становиться более свободным. В приведенных ниже примерах 复合 'сложный' и 高级 'высокоорганизованный' являются дифференцирующими словами. И в сопровождении данной частицы в предложении их необходимо размещать перед прилагательными – 总 ('общий') и 特殊 ('специфичный'), соответственно.

这<sup>r</sup> 段<sup>q</sup> 话<sup>n</sup> 的<sup>u</sup> 意思<sup>n</sup> , /w 即是<sup>v</sup> 说<sup>v</sup> , /w 历史<sup>n</sup> 是<sup>vl</sup> 社会<sup>n</sup> 的<sup>u</sup> 诸<sup>r</sup> 个人<sup>n</sup> 之<sup>u</sup> 意识<sup>n</sup> 的<sup>u</sup> 活动<sup>v</sup> 创造<sup>v</sup> 出来<sup>vd</sup> 的<sup>u</sup> , /w 历史<sup>n</sup> 的<sup>u</sup> 必然<sup>n</sup> , /w 即是<sup>v</sup> 诸<sup>r</sup> 个人<sup>n</sup> 的<sup>u</sup> 活动<sup>v</sup> 错综<sup>n</sup> 复合<sup>f</sup> 的<sup>u</sup> 总<sup>a</sup> 结果<sup>n</sup> 。 /w (Смысл данного высказывания заключается в том, что история обусловлена сознательной общественной деятельностью человека, а историческая необходимость представляет собой **сложный общий результат** деятельности человека).

其次<sup>c</sup> , /w 人类<sup>n</sup> 社会<sup>n</sup> 同<sup>p</sup> 自然界<sup>n</sup> 一样<sup>a</sup> , /w 也是<sup>v</sup> 物质<sup>n</sup> 运动<sup>n</sup> 形式<sup>n</sup> , /w 不过<sup>c</sup> 它<sup>r</sup> 又<sup>d</sup> 是<sup>vl</sup> 一<sup>m</sup> 种<sup>q</sup> 更<sup>d</sup> 高级<sup>f</sup> 的<sup>u</sup> 特殊<sup>a</sup> 的<sup>u</sup> 物质<sup>n</sup> 运动<sup>n</sup> 形式<sup>n</sup> 。 /w (Во-вторых, человеческое общество, как и природа, является формой движения материи, представляя более **высокоорганизованную специфичную** форму движения материи).

В грамматическом плане дифференцирующие слова обнаруживают ряд признаков, отличающих их как от прилагательных, так и от существительных в атрибутивной функции. В отличие от прилагательных, дифференцирующие слова не имеют степеней сравнения и не могут сочетаться с соответствующими наречиями (很, 非常 – 'очень'). В отличие от существительных в атрибутивной функции, они не требуют наличия структурной частицы 的 в препозиции к

определяемому слову. Дифференцирующие слова образуют закрытую совокупность языковых единиц, что является характерным признаком самодостаточной грамматикализованной категории. От существительных, выступающих в роли компонентов сложных слов – например, 钢笔 ('стальная ручка' как 'авторучка'), где иероглиф 钢 размечен тегом *n* как существительное – дифференцирующие слова отличаются тем, что сохраняют синтаксическую автономность. Они могут выступать определениями при разных существительных – в регулярной, продуктивной модели, например, 女装 ('женская одежда'), 女厕所 ('женский туалет'), 女演员 ('женщина-актриса'). При этом, компоненты сложных слов образуют с определяемым компонентом лексикализованное единство с непродуктивной, идиоматичной сочетаемостью. Эта регулярная продуктивная сочетаемость при отсутствии признаков прилагательного (степени сравнения, сочетаемость со степенными наречиями) и отсутствии признаков существительного в атрибутивной функции убедительно обуславливает выделение дифференцирующего слова в самостоятельный частеречный класс.

Статистическая значимость языковых единиц той или иной части речи зависит от множества факторов, которые не всегда могут быть учтены в какой-то универсальной схеме. Однако масштабируемость корпусных данных позволяет исследовать речевую практику на практически неограниченную глубину – в пределах собранной в корпусе совокупности текстов.

Экспериментальным путем целесообразный порог может быть найден даже без использования специальных программных фильтров. Логично предположить, что имплементация специальных тегов уровня частеречной разметки для дифференцирующего слова обусловлена его фактической значимостью для сбалансированной и непротиворечивой аннотации речи: для разработчиков спецификации корпусного менеджера «Онлайн-корпуса» функциональность дифференцирующего слова в речи очевидна. Для подтверждения данного предположения был проведен соответствующий эксперимент. Благо, что корпусная выборка может корректироваться и в программном обеспечении «Онлайн-корпуса» такая опция предусмотрена. Так, при обращении к описанной выше расширенной выборке из 16 254 единиц анализируемого подкорпуса, в выдаче оказывается более широкий спектр языковых единиц – по сравнению с данными эталонного корпуса. В перечне уже оказываются и междометия, и звукоподражательные слова, и, наконец, дифференцирующие слова: соответственно, 7 226 междометий, что составило 0,08 % относительной частотности; 2 176 звукоподражательных слов, что составило 0,03 % относительной частотности, и 2 794 дифференцирующих слов, что, также, составило 0,03 % относительной частотности (Таблица 3). Подобная избирательность речевой практики, на самом деле, вполне ординарна: в первом корпусе (достаточно репрезентативном «Брауновском корпусе») его создатели с удивлением не обнаружили множество ожидаемой лексики, например, лексемы *smile* (англ. 'улыбка').

**Таблица 3.** Сравнительная частотность единиц «Онлайн-корпуса», представляющих разные части речи китайского языка на материале расширенной выборки из 16 254 единиц

**Table 3.** Comparative frequency of units in the Online Corpus representing different parts of speech in Chinese, based on an expanded sample of 16,254 units

Перечень наиболее распространенных частей речи	Статистика выявленных словоупотреблений в абсолютном выражении	Статистика выявленных словоупотреблений в процентном выражении
существительное	2 573 011	29,35 %
глагол	2 299 497	26,23 %
наречие	630 701	7,20 %
местоимение	536 815	6,13 %
предлог	388 659	4,43 %
союз	327 045	3,73 %
прилагательное	548 764	6,26 %
числительное	220 321	2,51 %
частица	1 073 898	12,25 %
счетное слово	155 194	1,77 %
дифференцирующее слово	2 794	0,03 %
междометие	7 226	0,08 %
звукоподражательное слово	2 176	0,03 %
ИТОГО	8 766 101	100 %

Таким образом, демонстрируя значимую статистику сопоставимого с единицами из «официального» перечня частей речи словоупотребления, в частности для разрядов междометия и звукоподражательного слова, разряд дифференцирующего слова на их фоне выглядит вполне репрезентативно. Подобно междометию и звукоподражательному слову, дифференцирующее слово не лидирует в частеречном рейтинге (Таблица 3), однако и не является в нем избыточным элементом, подтверждая обоснованность своего корпусного позиционирования в ряду других частей речи. Этот тезис кроме вышеподтвержденной количественной обеспеченности – дифференцирующее слово не менее распространено, чем «официальные» части речи звукоподражательное слово и междометие – может быть подкреплен убедительной качественной аргументацией: важной в данном контексте представляется самостоятельная грамматическая значимость дифференцирующего слова в речи.

Полученные результаты показывают, что частеречный состав эталонных корпусов русского и китайского языка имеет как общие черты, так и различия. Так, по распространенности категорий показатели оказались, в общем, близкие: и в русскоязычном, и в китайскоязычном материале на первом месте *существительные*; второе и третье места по распространенности заняли, соответственно, *глаголы* и *наречия*. Тем не менее, сравнительная статистика общей массы словоупотреблений имеет принципиальное отличие: в русской речи наиболее востребовано являющееся знаменательной частью речи *местоимение*, в китайской – являющаяся служебной частью речи *частица*.

#### 4. Дискуссия

Выявляемые современными методами исследования речи противоречия в существующих метаязыковых шаблонах структурирования языка являются следствием его некогда интуитивного моделирования. В контексте тотального охвата языковой практики сегодня уже,

как минимум, нерационально рассматривать язык как застывший во времени артефакт. Накопленный багаж лингвистических знаний о законах и обычаях современной коммуникации обуславливает целесообразность не только заполнения многочисленных лакун, но и критический пересмотр многих уже устоявшихся стереотипов. Практика зачастую далека от традиционно удобных для описания идеальных схем. Технологичность и эмпиризм современной лингвистики безальтернативно свидетельствуют о том, что язык мало зависит от его описаний, а вот «... формальные структуры в грамматиках зависят от языка» (Croft, 2001: 34). Сегодня в науке очевидно смещение парадигмы от интуитивно-субъективных оценок к объективным, эмпирически обоснованным выводам. Этот переход стал возможен во многом благодаря развитию компьютерных технологий, задействование которых предоставляет исследователям верифицированные инструменты для анализа больших массивов языковых данных: «Как и проявления иных языковых системообразующих отношений реализация модификационного потенциала языка в компьютерно-опосредованной речевой сфере нередко отличается от сложившихся стереотипов, позволяя принять во внимание ресурсы и возможные направления развития той или иной семантики» (Баркович, 2018: 11).

Обоснованная коррекция лингвистических категорий не ставит под сомнение системность организации языка. Однако глубоко свойственная языку системность, в свою очередь, обуславливает тщательность проверки новых или уточненных языковых данных, принятие которых может существенно повлиять на стабильность самой системы. Сама по себе статистика, в том числе корпусная, не является универсальной материей, способной заполнить все пробелы в научном описании. Тем не

менее, игнорирование доступных посредством современных инструментов данных чревато схоластичностью: «В наши дни мы рискуем столкнуться с навязыванием нам проблемных аналитических систем, основанных на традициях языкового анализа, которые в прошлом целенаправленно отвергали корпусные данные. Сложность возникает из-за того, что те направления компьютерной лингвистики, которые раньше игнорировали корпусные данные, в последние годы радикально изменили своё отношение к корпусам, но сохранили модели языка, которые не подтверждаются имеющимися у них данными» (Sinclair, 1991: 22). Частеречная категориальность, несмотря на долгий срок ее эволюции и апробации, не является навсегда закрытой для совершенствования лингвистической моделью. А варианты ее модификации, как показало проведенное исследование, есть. Особенно убедительно такая аргументация подкрепляется сопоставительным анализом резервов грамматической организации для разноструктурных языков. Китайский язык, с его ярко выраженным аналитизмом, как оказалось, нуждается в частеречном метаописании не менее преимущественно синтетического русского. Части речи – фундаментальная категориальность языка: «В системе частей речи отражаются стадии развития данного языка, его грамматический строй» (Виноградов, 2001: 17). Конечно, частеречная категориальность разных языков обладает в разной степени выраженной спецификой и подвержена влиянию как собственно языковых, так и экстраязыковых факторов. Однако выверенность и адаптивность самой частеречной системы чрезвычайно важна для классификации и разноуровневой кодификации языковых единиц. Структура частей речи так или иначе учитывается при формулировке самых разных грамматических правил. Не менее актуален частеречный бэкграунд при анализе речевой функциональности

языковых единиц и описании их взаимодействия.

Корректность лингвистической интерпретации языка всегда была востребована, еще более актуальной она стала в условиях формализации коммуникации и ее насыщения множеством искусственных языков, принципиально нацеленных на максимальную прозрачность и упрощенность метаязыка. Однако логичной интенциональности таких ожиданий и устремлений зачастую не соответствует уже сложившийся *status quo* теоретических концепций. Устранение несовершенства теорий о языке на практике бывает зачастую осложнено не всегда очевидными обстоятельствами, или «нечеткими множествами» данных. Так, для строгой категориальной дифференциации единиц китайского языка существенным препятствием является их генетическая **полифункциональность** – заложенная в самой природе китайского языка вариативность грамматической идентичности языковых единиц. Более того, полифункциональность огромного количества единиц китайского языка осложнена отсутствием выраженных морфологических маркеров того или иного их статуса, а также стилистическими особенностями речевой деятельности (Чжан, 2002–2003). Стилистика использования многих китайских слов часто зависит от того, в прямом или переносном значении они употребляются в каком-то конкретном контексте, что в конечном итоге, влияет на их грамматический статус (Дай, 2007). Вследствие этого, в частности, согласно «Введению в современный китайский язык» Шао Цзинмина полифункциональные слова следует делить на 6 основных групп (Шао, 2008 : 172). Эта данность обуславливает тот факт, что большое количество языковых единиц китайского языка может обладать разной идентичностью – в зависимости от фактически выполняемой ими

семантической или грамматической нагрузки, которую они несут в речи. Это напрямую влияет и на их зачастую ситуативную частеречную принадлежность. Использование единиц китайского языка как той или иной части речи по умолчанию сложно прогнозируемо и зависит от текущей синтаксической позиции (Lan et al., 2023; Wang et al., 2026 и др.). Логика полифункциональности присуща и отдельным единицам русского языка, где она идентифицируется как **антимерия** (Московин, 2013: 25). При антимерии имеет место «... использование слова одной части речи в несвойственной этому слову функции другой части речи или окказиональное образование частей речи» (Культура русской речи, 2011: 55). Однако в русском языке частеречная полифункциональность, или антимерия, не носит массового характера, будучи ассоциируемой, скорее, с неким украшательством речи. Вместе с тем, интерференция многочисленных форм русскоязычной лексики нередко воплощается в омонимии (Ахманова, 1966: 275). Будучи подобными по форме, омонимы в русском языке не меняют свою частеречную принадлежность в зависимости от выполняемой в речи функции и этим они существенно отличаются от полифункциональных единиц китайского языка.

Несмотря на сложность и подвижность частеречной организации языка, в признании ее научной ценности для сопровождения языков различного строя и типологической принадлежности у специалистов имеется полный консенсус. Система частей речи – многогранный лингвистический артефакт, в создание и совершенствование которого внесли вклад многие поколения ученых – от мыслителей античности до технологически оснащенных современных исследователей. Однако вопросы актуализации грамматических категорий современных языков по-прежнему активно обсуждаются (Langacker, 1987; Goldberg, 1995;

Tomasello, 2003 и др.). Даже такой фундаментальный лингвистический концепт как слово поныне является достаточно условным: слова зачастую лишь «... минимальные терминальные элементы синтаксической структуры, [и] воплощение принципов, регулирующих появление морфологического материала» (Anderson, 1992: 17). Характеристика языковых единиц является глобальной проблемой языкознания: «У лингвистов нет надежной базы для идентификации слов в разных языках. ...» (Haspelmath, 2011: 24). Тем не менее, такую базу лингвистика современности обрела с появлением компьютерных технологий. И это уже привело не только к смене лингвистических парадигм, но и к изменению взглядов на допустимость модификации широкого спектра канонов языкознания. Будучи до середины XX в., скорее, гипотетической опцией лингвистической практики, эмпиризм компьютерной эры позволил исследователям заглянуть за горизонт теоретических построений и опереться на фактографию речи. Переоценить значимость компьютеризированной обработки речи сложно до сих пор, но до полной инвентаризации языкового богатства по-прежнему далеко, что можно видеть на примере данного исследования.

Значимость корпусного анализа систем частей речи рассмотренных разноструктурных языков обусловлена близким к полярному позиционированием русского и китайского языков. Задачу такого уровня сложности достаточно сложно было бы выполнить только традиционными методами – без привлечения корпусной методики. Вместе с тем, параллельный анализ речевого функционирования языковых единиц удаленных языков должен основываться на верифицированной как в плане критериев, так и параметров категориальной базе, рациональность и обоснованность которых в значительной степени опирается на систему частей речи.

Подкрепленная статистикой систематизация фундаментальных основ грамматик русского и китайского языков, в свою очередь, позволяет решать широкий круг вопросов корректного описания формальных, содержательных и функциональных параметров речевой практики. Это востребовано уже для элементарной, на первый взгляд, лексикографической репрезентации языковых единиц. Современные требования к обработке языкового материала обуславливают вывод лингвистического описания за рамки интроспекции и субъективных суждений исследователя. Объективная работа с большими массивами текстов даёт возможность фиксировать реальные закономерности употребления языковых единиц на основе их частотных характеристик и дистрибутивных свойств. Это означает возможность верифицировать грамматические категории на материале живой речи, а не только на основе прескриптивных представлений. Верифицированность обработки первичных данных, в свою очередь, обеспечивает надежный фундамент для формирования дальнейших метаязыковых обобщений, что в равной мере актуально для релевантного описания самых разных языков (Tognini-Bonelli, 2001; Hunston, 2002; Levshina, 2022 и др.).

Как можно видеть на примере данного исследования, корпусный подход позволяет, в целом, объективно сопоставить разноструктурные русский и китайский языки – при ощутимой разнице в их грамматическом описании и их лингвокультурном своеобразии. Сопоставительное метаописание русского и китайского языков представляет собой одну из наиболее сложных задач современного языкознания. Типологическая дистанция между ними очевидна: русский язык относится к флективным языкам с развитой морфологией, тогда как китайский является изолирующим языком, в котором

грамматические отношения выражаются преимущественно порядком слов и служебными морфемами. Именно эта асимметрия делает корпусный подход особенно ценным инструментом их сравнения, поскольку он опирается на статистически репрезентативный языковой материал.

При обилии обсуждаемых версий грамматической системы русского языка, за общепризнанный ориентир целесообразно принять так называемый «академический» перечень частей речи: «... наиболее распространенная трактовка выделяет в русском языке 10 частей речи» (Сичинава, 2018: 11). В китайской лингвистике базовый вариант, рекомендуемый для использования в образовательных целях, состоит из 12 частей речи (Учебный план, 1984). Разница достаточно существенная, однако при ближайшем рассмотрении можно заметить, что и русскоязычная, и китайскоязычная системы во многом похожи. В обоих языках противопоставляются знаменательные и служебные части речи, выделяются такие практически равноценные части речи как существительные, глаголы и прилагательные. Вместе с тем, статус наречия и междометия в русском и китайском языках различен. Так, в русской грамматике наречие входит в число знаменательных частей речи, а в китайской лингвистике оно идентифицируется как служебная часть речи. Междометие грамматикой русского языка наделено особым статусом и не является ни знаменательной, ни служебной частью речи, а в китайской грамматике оно отнесено к служебным частям речи. Примечательно, что китайская грамматическая традиция включает в систему частей речи особые классы единиц, не имеющие прямых аналогов в русском языке: счётные слова, обязательно сопровождающие числительные при счёте предметов, и звукоподражательные слова,

также выделяемые в самостоятельный разряд.

Отдельного внимания здесь заслуживает проблема потенциально новых классов языковых единиц, чья частеречная принадлежность остаётся дискуссионной. Для них постепенно в оборот термин «периферийная категория». В русском языке к таким категориям относится предикатив (или категория состояния), в китайском – дифференцирующее слово, обладающие признаками как прилагательного, так и наречия. В данном контексте, по-видимому, именно системный анализ речевой практики может стать основой аргументации для выделения или невыделения подобных единиц в самостоятельные классы, – с учетом реалий их синтаксической нагрузки в контекстах разной распространенности.

Так или иначе, как артефакт, система частей речи при необходимости может быть модифицирована, как и любая грамматическая надстройка естественного языка. Более того, по мере углубления знаний о языке, ее структура может адаптироваться к новым реалиям языкового функционирования или исследовательским приоритетам. Изначально рамки частеречных категорий условны: «Поскольку некоторые слова могут быть более "пограничными", чем другие в пределах класса слов, неудивительно, что границы между двумя классами слов также могут быть нечеткими» (Viber et al., 1999 : 24). Ввиду чрезвычайной вариативности и подвижности речевой практики в языковой системе – подобно статическому электричеству в динамичных физических объектах – в ней перманентно накапливается модификационный потенциал. Это обусловлено целым рядом факторов, в том числе, изначально заложенной в грамматике искусственностью; уникальностью античных языков, грамматические описания для которых часто копировались

учеными для далеко не во всем им подобных языков; эволюцией социокультурной сферы, обслуживаемой тем или иным языком и т. д. Не менее значимыми стимулами для изменения метаописания языка часто выступают и не такие влиятельные, на первый взгляд, факторы как лакунарность или мягкость норм, интерференция взаимодействующих языковых систем, обычаи устного модуса речевой практики, законодательные установления и прочие обстоятельства многообразной функциональности любого живого языка. Даже в случае очевидной насущности перемен в грамматической системе языка как феномена зачастую наблюдается неравномерность ее преобразования во временном измерении. Так, в частности, консенсус относительно применимости частеречной категориальности для китайского языка был достигнут значительно позже ее установления для большинства европейских языков, включая русский.

Несмотря на все сопутствующие трудности, анализ возможных новаций в метаописании того или иного языка вполне научно обоснован и более чем уместен в современных условиях тотальной инвентаризации и метаязыковой формализации естественных языков, участвующих в компьютерно-опосредованной коммуникации. При этом компьютеризация языковой практики не только стимулирует высокую степень аутентичности метаязыка, но и предоставляет для этого считавшиеся относительно недавно фантастические статистические возможности (см., напр., Çöltekin and Rama, 2023; Crosthwaite et al., 2023). На фоне экстенсивной квантитативной информации квалитативные обобщения обладают особой убедительностью и достоверностью. Что немаловажно, квантитативно обеспеченная аргументация в равной мере применима как к близкородственным, так и к таким типологически удаленным языкам как

русский и китайский. При сопоставительном анализе разноструктурных языков корпусная лингвистика, в частности, полностью обеспечивает надёжным метаязыковым инструментарием сбор и презентацию первичных данных. Высокая степень унифицированности подходов к аннотированию содержащихся в корпусе текстов благоприятствует референтной обработке языкового материала как «НКРЯ», так и «Онлайн-корпуса». В целом, корпусный подход даёт объективную картину соответствия грамматических категорий реалиям речи, – в том числе, при сравнении русского и китайского языков.

Разметка китайских корпусов принципиально сложнее, чем для многих других европейских языков по двум причинам. Во-первых, в китайском языке нет морфологического словоизменения, поэтому нельзя ориентироваться на окончания или суффиксы. Во-вторых, текст записывается без пробелов, что делает автоматическое определение границ слов нетривиальной задачей, требующей специальных алгоритмов и последующей ручной проверки. Помимо этого, порядок слов в китайском языке несёт значительную грамматическую нагрузку, что также должна учитывать система разметки. Пожалуй, наиболее существенной проблемой корпусной лингвистики китайского языка является специфика автоматической сегментации текста. В отличие от русского, где границы слов графически очевидны, китайский текст записывается без пробелов между словами, что требует применения сложных алгоритмов разбиения на словоформы и последующей тщательной проверки результатов. Морфологическая разметка китайских корпусов опирается на функциональный принцип: частеречная принадлежность единицы определяется не по её форме, а по синтаксической позиции и контекстуальному окружению. Это принципиально отличается от подхода,

применяемого в тех корпусах, где морфологическая аннотация во многом опирается на словоизменительные парадигмы.

«Онлайн-корпус» считается одним из наиболее развитых инструментов китайской корпусной лингвистики. Его особенность – так называемая функциональная аннотация: частеречная принадлежность слова определяется не по его форме, а по роли в конкретном контексте. Это важно, поскольку в китайском языке одно и то же слово может выступать в разных грамматических функциях. Не менее важно и то, что ключевую роль для восприятия смысла в китайском языке играет порядок слов: «Мандарин нелегко классифицировать с точки зрения порядка слов» (Li and Thompson, 1981: 19). Термин «грамматико-графическая аннотация», используемый в описании данного корпуса, обозначает комплексную систему маркировки, объединяющую информацию о графической форме языковых единиц и их грамматических характеристиках. Данная система предусматривает функциональный подход к определению частеречной принадлежности: категория слова устанавливается не на основании его некоего единообразного лексикографического шаблона, а через анализ его дистрибуции, синтаксической роли и морфологических возможностей в конкретном контексте. Такая конфигурация разметки обусловлена тем, что в китайском языке морфологические маркеры, традиционно служащие надёжными индикаторами частеречной принадлежности во флективных языках, отсутствуют или носят факультативный характер. В этих условиях именно функциональный анализ, учитывающий позиционные и комбинаторные свойства единицы, становится основным инструментом её категоризации. Это позволяет исследователю осуществлять многоаспектный поиск, комбинируя критерии различных уровней. Например,

можно извлечь все глаголы движения в прошедшем времени, употребляющиеся в дейктических конструкциях, или проанализировать частотность элементов определённого семантического поля в текстах конкретного жанра. Что важно для достижения цели данного исследования, в «Онлайн-корпусе» можно собирать и анализировать данные о частеречной идентичности языковых единиц с учетом реализации ими своих грамматических возможностей в контекстах.

Возможности «НКРЯ» специалистам известны достаточно хорошо: на сегодня – это, пожалуй, наиболее репрезентативный и снабженный разнообразной разметкой массив текстов. В нем реализована, в числе других, так называемая морфологическая разметка, актуальная для решения задач нашего исследования. Среди широкого спектра параллельных подкорпусов в нем представлен и русско-китайский подкорпус.

Сегодня корпусная репрезентация частей речи становится одним из наиболее информативных ресурсов, позволяя получить аргументированные и объективные представления о языковой системе. Этот подход продуктивен при сопоставлении, в том числе, разноструктурных языков, таких как русский и китайский, где интуитивные оценки могут быть подвержены интерференции культурных и структурных факторов функционирования языков.

Статистика, получаемая в корпусной среде посредством специализированного программного обеспечения объективна, дискретна и масштабируема. Однако ее лингвистическая ценность полностью зависит от качественной интерпретации, и лишь симбиоз квантитативно-кавалитативных обобщений имеет наивысшую научную ценность. Корпусные данные статистически убедительны и верифицируемы, способствуют достоверной лингвистической интерпретации эмпирического материала, обладают метаязыковой совместимостью с

широким кругом языковых метаописаний, универсальны и релевантны для обработки текстов и фрагментов коммуникации разной кодовой природы, применимы в междисциплинарных исследованиях. Лингвистический потенциал корпусного подхода далек от исчерпания, в том числе, это касается изучения грамматики: «Корпус изначально был задуман как инструмент для изучения грамматики» (Halliday, 1994: 48). Корпусный анализ частеречной системы языка позволяет на высокорепрезентативном материале исследовать специфику современного речевого функционирования и соответствие ему грамматических категорий. Это путь к пониманию закономерностей и динамики языка.

### 5. Выводы

Использование для лингвистического анализа репрезентативных корпусов, таких как «НКРЯ» и «Онлайн-корпус», позволяет выявить функциональные закономерности языка, которые невозможно было бы обнаружить при традиционном подходе. Этот тезис полностью подтвержден результатами проведенного анализа. В данном контексте подтверждено, что частеречная категориальность имеет фундаментальное значение для лингвистики, даже в случае типологической удаленности и разноструктурности описываемых языков. Анализ квалитативных характеристик речи через призму частеречной статистики позволил сделать обоснованные выводы, в частности, о модальности, динамике семантики и особенностях прагматической обусловленности текста. Характерно, что в русском языке большую роль играет морфологическая, а в китайском – синтаксическая адаптивность языковых единиц. В квантитативном аспекте в преимущественно синтетическом русском языке по числу лексем лидируют существительные, а по числу словоупотреблений – местоимения. В аналитическом китайском языке среди языковых единиц – по данным базовой

выборки – также преобладают существительные, но самой востребованной частеречной категорией являются частицы. Вместе с тем, данное соотношение характеризует именно высокочастотный сегмент и не проецируется автоматически на расширенную выборку, в которой представлен более широкий набор частеречных классов. При этом доминирование по числу лексем существительных в обоих рассмотренных языках – русском и китайском – свидетельствует о ключевой роли номинаций в коммуникации. В то же время, различия в частотности других категорий слов (например, местоимений в русском, частиц в китайском языке) указывает на структурные и функциональные особенности глубинного структурирования проанализированных посредством корпусной методики русского и китайского языков. Необходимо отметить, что корпусный анализ продемонстрировал состоятельность предикатива в русском языке, который, хотя и представлен небольшим количеством лексем, имеет высокую частотность словоупотреблений. Аналогично, в китайском языке выявлена метаязыковая значимость такой категории, как дифференцирующее слово. Единицы данных категорий могут быть корректно идентифицированы как элементы определенной совокупности. Их не зависящее друг от друга включение в корпусную аннотацию свидетельствует о том, что частеречные системы являются динамичными и могут инкорпорировать новые разряды.

Квалитативно-квантитативная интерпретация речевой практики, осуществленная благодаря корпусным инструментам практически в режиме реального времени, позволила сделать обоснованные выводы о параметрах возможной модификации частеречных систем русского и китайского языков. В частности, отмечено, что квалитативная

специфика частеречной насыщенности текста определенным образом характеризует его содержательную идентичность, что было подтверждено в аспекте метаописания модальности. Полученные данные квантитативного характера позволили обоснованно проанализировать приоритеты выбора языковых средств для решения коммуникативных задач. Полученные корпусные данные о реальноречевом функционировании единиц того или иного частеречного кластера в разноструктурных языках подтверждают релевантность задействования корпусного инструментария для решения сложных задач. Анализ и интерпретация статистической информации позволяют синтезировать обобщения о такой сложной метаязыковой проблематике, как качество обеспеченности современных языков непротиворечивым грамматическим аппаратом. В частности, предметное исследование частеречных систем русского и китайского языков позволило сделать вывод о насущности выделения в них новых разрядов частей речи и их целесообразном инкорпорировании в лингвистическую практику. Выявленная состоятельность и предикатива, и дифференцирующего слова – как самостоятельных категорий частеречной системы в своих языках – вполне объективна.

Дальнейшие исследования в связи с данной проблематикой перспективны и могут быть реализованы по следующим направлениям. Так, можно расширить эталонный корпус за счёт включения единиц среднечастотного сегмента, что позволит верифицировать выявляемые закономерности на более репрезентативном материале. Актуальным может оказаться жанрово дифференцированный анализ материала: сопоставление частеречного распределения в публицистических, научных, художественных и разговорных текстах способно выявить функционально-

стилистическую специфику обоих языков. Также, полученные данные могут послужить обоснованием практических рекомендаций в области корпусной аннотации – в частности, для уточнения тегсетов, используемых при разметке предикативов в русскоязычных и дифференцирующих слов в китайскоязычных корпусах. Фактически, актуализированный в данном исследовании методологический подход может быть продуктивно развит в широком спектре лингвистической работы и интегрирован в существующие метаязыковые модели. Частеречная категориальность является важным средством идентификации и квалификации языковых единиц в разнообразных метаописаниях, что релевантно, например, для векторного представления языковых данных системами искусственного интеллекта при сентимент-анализе. Прикладная значимость полученных результатов может быть последовательно реализована в лексикографии, лингводидактике, металингвистике и других областях лингвистической практики.

#### Список литературы

- Алпатов В. М. Слово и части речи. М.: Изд. дом ЯСК. 2018. 256 с.
- Баркович А. А., Ван Ц. Лингвистические корпусы китайского языка: функциональный аспект // Вестник МГЛУ. Сер. 1. Филология. 2015. № 5 (78). С. 105–113.
- Баркович А. Метаописание деривационных отношений: специфика системной репрезентации // Mundo Eslavo. 2018. № 17. С. 7–25.
- Виноградов В. В. Русский язык. М.: Рус. яз. 2001. 720 с.
- Захаров В. П., Богданова С. Ю. Корпусная лингвистика. Иркутск: ИГЛУ. 2011. 161 с.
- Культура русской речи: энцикл. слов.-справ. / Под ред. Л. Ю. Иванова. М.: Флинта. 2011. 840 с.
- Ломоносов М. В. Грамматика русского языка, академика М. В. Ломоносова, 1755 года. СПб.: Тип. Имп. Акад. Наук. 1855. 247 с.

Московин В. П. О подходах к определению понятия «троп» // Известия РАН. Серия литературы и языка. 2013. Т. 72, № 2. С. 20–31.

Национальный корпус русского языка и проблемы гуманитарного образования / Отв. ред.-сост. Н. Р. Добрушина. М.: ТЕИС: ГУ-ВШЭ. 2007. 237 с.

Савчук С. О., Архангельский Т. А., Бонч-Осмоловская А. А., Доница О. В., Кузнецова Ю. Н., Ляшевская О. Н., Орехов Б. В., Подрядчикова М. В. Национальный корпус русского языка 2.0: новые возможности и перспективы развития // Вопросы Языкознания. 2024. № 2. С. 7–34.

Сичинава Д. В. Части речи // Материалы к корпусной грамматике русского языка. Выпуск III: Части речи и лексико-грамматические классы. СПб.: Нестор-История. 2018. С. 9–39.

Anderson S. R. A-Morphous Morphology. Cambridge: Cambridge University Press. 1992. 434 p.

Barkovich A. Informational Linguistics: Computer, Internet, Artificial Intelligence and Language // 2019 IEEE 1st International Conference on Artificial Intelligence in Information and Communication (ICAIIIC 2019), Okinawa, Japan, February 11–13. 2019. Pp. 008–013.  
<https://doi.org/10.1109/ICAIIIC.2019.8668989>

Biber D. Longman Grammar of Spoken and Written English / Biber D., Johansson S., Leech G., Conrad S., and Finegan E. Harlow: Pearson Education. 1999. 1204 p.

Bybee J. L. Language, Usage and Cognition. Cambridge: Cambridge University Press. 2010. 262 p.  
<https://doi.org/10.1017/CBO9780511750526>

Çöltekin Ç. and Rama T. What do complexity measures measure? Correlating and validating corpus-based measures of morphological complexity. Linguistics Vanguard. 2023. 9(s1). Pp. 27–43.  
<https://doi.org/10.1515/lingvan-2021-0007>

Croft W. Radical Construction Grammar: Syntactic Theory in Typological Perspective. Oxford: Oxford University Press. 2001. 416 p.  
<https://doi.org/10.1093/acprof:oso/9780198299554.001.0001>

Crosthwaite P. Research trends in corpus linguistics: A bibliometric analysis of two decades of Scopus-indexed corpus linguistics research / Crosthwaite P., Ningrum S., and Schweinberger

M. // International Journal of Corpus Linguistics. 2023. 28 (3). Pp. 344–377.  
<https://doi.org/10.1075/ijcl.21072.cro>

Francis W. N., and Kučera H. Frequency Analysis of English Usage: Lexicon and Grammar. Boston: Houghton Mifflin. 1982. 561 p.

Givón T. Syntax: A Functional-Typological Introduction (Vol. 1). Amsterdam: John Benjamins. 1984. 464 p.

Goldberg A. E. Constructions: A Construction Grammar Approach to Argument Structure. Chicago: University of Chicago Press. 1995. 265 p.

Gries S. T. Quantitative Corpus Linguistics with R: A Practical Introduction. New York: Routledge, 2009. 286 p.  
<https://doi.org/10.4324/9781315746210>

Halliday M. A. K. An Introduction to Functional Grammar (2nd ed.). London: Edward Arnold. 1994. 434 p.

Haspelmath M. The Indeterminacy of Word Segmentation and the Nature of Morphology and Syntax // Folia Linguistica. 2011. 45 (1). Pp. 31–80.  
<https://doi.org/10.1515/flin.2011.002>

Hunston S. Corpora in Applied Linguistics. Cambridge: Cambridge University Press. 2002. 241 p.

Lan, G. Part of speech tagging of grammatical features related to L2 Chinese development: A case analysis of Stanza in the L2 writing context / Lan G., Pan X., Sun Y., and Lu Y. // Frontiers in Psychology, 2023. 14.  
<https://doi.org/10.3389/fpsyg.2023.1139703>

Langacker R. W. Foundations of Cognitive Grammar, Vol. I: Theoretical Prerequisites. Stanford: Stanford University Press, 1987. 540 p.

Levshina N. Corpus-based typology: applications, challenges and some solutions. Linguistic Typology. 2022. 26(1). P. 129–160.  
<https://doi.org/10.1515/lingty-2020-0118>

Li C. N., and Thompson S. A. Mandarin Chinese: A Functional Reference Grammar. Berkeley: University of California Press, 1981. 691 p.

McEnery T., and Hardie A. Corpus Linguistics: Method, Theory and Practice. Cambridge: Cambridge University Press, 2012. 294 p.

Partee B. H. Noun phrase interpretation and type-shifting principles // Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers / Groenendijk J., de Jongh D., and Stokhof M. (Eds.). Dordrecht:

Foris. 1987. P. 115–143.  
<https://doi.org/10.1002/9780470758335.ch15>

Sapir E. Language, an Introduction to the Study of Speech. New York: Harcourt, Brace and World, 1921. 258 p.

Schütze H. Distributional Part-of-Speech Tagging // Proceedings of the Seventh Conference on European Chapter of the Association for Computational Linguistics, 27–31 March 1995. Dublin, Ireland. Pp. 141–148.

Siewierska A. Person. Cambridge: Cambridge University Press, 2004. 327 p.

Sinclair J. Corpus, Concordance, Collocation. Oxford: Oxford University Press. 1991. 170 p.

Tomasello M. Constructing a Language: A Usage-Based Theory of Language Acquisition. Cambridge, MA: Harvard University Press, 2003. 388 p.

Tognini-Bonelli E. Corpus Linguistics at Work. Amsterdam: John Benjamins Publishing Company, 2001. 223 p.

Wang X. Beyond Commonalities: A Quantitative Perspective on Syntactic Features across Mandarin Chinese Varieties / Wang X., Bin S., Chen I., and Hao Y. // Journal of Quantitative Linguistics, 2026. 33(2), P. 81–116.  
<https://doi.org/10.1080/09296174.2025.2567089>

Wang, Z. A review of Chinese sentiment analysis: subjects, methods, and trends / Wang Z., Huang D., Cui J., Zhang X., Ho S.-B., and Cambria E. // Artificial Intelligence Review. 2025. 58, 75. <https://doi.org/10.1007/s10462-024-10988-9>

Xu Z. The Word Status of Chinese Adjective-Noun Combinations // Linguistics. 2018. 56 (1). P. 207–256.  
<https://doi.org/10.1515/ling-2017-0035>

马建忠. 马氏文通. 北京: 商务印书馆. 1983. 447 p. = Ма Ц. Объяснение правил письменного языка господина Ма. Пекин: Шаньу иньшугуань, 1983. 447 с.

黎锦熙. 新著国文语法. 北京: 商务印书馆. 1993. 336 p. = Ли Ц. Новая грамматика национального языка. Пекин: Шаньу иньшугуань. 1993. 336 с.

曹杨波.  
计算机辅助语言学习的发展与前景 // 中国信息科技. 2011. № 3. P. 107–109. = Цао Я. Развитие и перспективы обучения языкам с помощью информационных технологий //

Китайские информационные технологии. 2011. № 3. С. 107–109.

张博. 现代汉语同形同音词与多义词.

北京: 北京语言大学对外汉语研究中心. 2002–2003. = Чжан Б. Омонимия и многозначность современного китайского языка. Пекин: Центр изучения китайс. яз. как второго Пекин. ун-та яз. и культуры. 2002–2003. URL: <http://wenku.baidu.com/view/b3eb771afc4ffe473368ab83.html> (дата обращения: 28.08.2025).

代建桃. 现代汉语同音词研究: 论文 ...

汉语言文字学; 四川师范大学, 成都. 2008. 99 p. = Дай Ц. Изучение омонимов современного китайского языка: дис. ... канд. китайской лингвистики; Сычуаньский педагогический университет. Чэнду, 2008. 99 с.

邵敬敏. 现代汉语通论. 上海:

上海教育出版社. 2007. 335 p. = Шао Ц. Общее введение в современный китайский язык. Шанхай: Изд-во образования Шанхая, 2007. 335 с.

#### Материалы исследования

Ахманова О. С. Словарь лингвистических терминов. М.: Сов. энцикл. 1966. 608 с.

Национальный корпус русского языка. URL: <http://www.ruscorpora.ru> (дата обращения: 28.08.2025).

Ожегов С. И., Шведова Н. Ю. Толковый словарь русского языка: 80 000 слов и фразеологических выражений. М.: А ТЕМП. 2006. 944 с.

Розенталь Д. Э., Голуб И. Б., Теленкова М. А. Современный русский язык. М.: Айрис-пресс. 2010. 448 с.

语料库在线 = Онлайн-корпус. URL: <http://www.cncorpus.org/> (дата обращения: 28.08.2025).

孙浩成 现代汉语语法大全. = Сунь Х. Грамматика современного китайского языка. URL: <http://wenku.baidu.com/view/0ea2c833433239-68011c9253.html> (дата обращения: 28.08.2025).

现代汉语词典 /

中国社会科学院语言研究所. 北京: 商务印书馆. 2016. 1799 p. = Словарь современного китайского языка / Ин-т

языкознания Акад. обществ. наук КНР. Пекин: Шаньу иньшугуань, 2016. 1799 с.

中学教学语法系统提要 (试行) /

人民教育出版社中学语文室. 北京:

人民教育出版社. 1984. 21 p. = Учебный план по грамматической системе средней школы (экспериментальный режим) / Отдел языка и литературы для средней школы Народного образовательного издательства. Пекин: Народ. образоват. изд-во. 1984. 21 с. URL: <https://eol.shzu.edu.cn/meol/analytics/> (дата обращения: 28.08.2025).

### References

- Alpatov, V. M. (2018). *Slovo i chasti rechi* [Word and Parts of Speech], Publishing house YASK, Moscow, Russia. (In Russian)
- Barkovich, A. A. and Wang, Q. (2015). Linguistic Corpora of the Chinese Language: Functional Aspect, *Vestnik MGLU. Ser. I. Filologiya*, 5(78), 105–113. (In Russian)
- Barkovich, A. (2018). Meta-Description of Derivational Relations: Specifics of System Representation, *Mundo Eslavo*, 17, 7–25. (In Russian)
- Vinogradov, V. V. (2001). *Russkiy yazyk* [Russian Language], Rus. Language, Moscow, Russia. (In Russian)
- Zakharov, V. P., Bogdanova, S. Yu. (2011). *Korpusnaya lingvistika* [Corpus Linguistics], IGLU Irkutsk, Russia. (In Russian)
- Ivanov, L. Yu. (ed.) (2011). *Culture of Russian Speech: Encyclopedic Dictionary and Reference Book*, Flinta, Moscow, Russia. (In Russian)
- Lomonosov, M. V. (1855). *Grammatika russkogo yazyka, akademika M. V. Lomonosova, 1755 goda* [Grammar of the Russian language, Academician M. V. Lomonosov, 1755], Imp. Academician Sciences Publ., Saint Petersburg, Russia. (In Russian)
- Moskovin, V. P. (2013). On Approaches to Defining the Concept of "Trope", *Izvestiya RAN. Seriya literatury i yazyka*, 72 (2), 20–31. (In Russian)
- Dobrushina, N. R. (ed.) (2007). *Russian National Corpus and Problems of Humanitarian Education*, TEIS: GU-HSE, Moscow, Russia. (In Russian)
- Savchuk, S. O., Arkhangelsky, T. A., Bonch-Osmolovskaya, A. A., Donina, O. V., Kuznetsova, Yu. N., Lyashevskaya, O. N., Orekhov, B. V., Podryadchikova, M. V. (2024). National Corpus of the Russian Language 2.0: New Possibilities and Development Prospects, *Voprosy Yazykoznaniiya*, 2, 7–34. (In Russian)
- Sichinava, D. V. (2018). Parts of speech, *Materialy k korpusnoy grammatike russkogo yazyka. Vypusk III: Chasti rechi i leksiko-grammaticheskie klassy* [Materials for a Corpus Grammar of the Russian Language. Part III: Parts of Speech and Lexical and Grammatical Classes], St. Petersburg: Nestor-Istoriya, 9–39. (In Russian)
- Anderson, S. R. (1992). *A-Morphous Morphology*, Cambridge University Press, Cambridge, UK. (In English)
- Barkovich, A. (2019). Informational Linguistics: Computer, Internet, Artificial Intelligence and Language, *2019 IEEE 1st International Conference on Artificial Intelligence in Information and Communication (ICAIC 2019)*, Okinawa, Japan, February 11–13, 008–013. <https://doi.org/10.1109/ICAIC.2019.8668989> (In English)
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*, Pearson Education, Harlow. (In English)
- Bybee, J. L. (2010). *Language, Usage and Cognition*, Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511750526> (In English)
- Çöltekin, Ç. and Rama, T. (2023). What do complexity measures measure? Correlating and validating corpus-based measures of morphological complexity, *Linguistics Vanguard*, 9(s1), 27–43. <https://doi.org/10.1515/lingvan-2021-0007> (In English)
- Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*, Oxford University Press, Oxford. <https://doi.org/10.1093/acprof:oso/9780198299554.001.0001> (In English)
- Crosthwaite, P., Ningrum, S., and Schweinberger, M. (2023). Research trends in corpus linguistics: A bibliometric analysis of two decades of Scopus-indexed corpus linguistics research, *International Journal of Corpus Linguistics*, 28 (3), 344–377. <https://doi.org/10.1075/ijcl.21072.cro> (In English)
- Francis, W. N. and Kučera, H. (1982). *Frequency Analysis of English Usage: Lexicon*

and Grammar, Houghton Mifflin, Boston. (In English)

Givón, T. (1984). *Syntax: A Functional-Typological Introduction (Vol. 1)*, John Benjamins, Amsterdam. (In English)

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*, University of Chicago Press, Chicago. (In English)

Gries, S. T. (2009). *Quantitative Corpus Linguistics with R: A Practical Introduction*, Routledge, New York. <https://doi.org/10.4324/9781315746210> (In English)

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar (2nd ed.)*, Edward Arnold, London. (In English)

Haspelmath, M. (2011). The indeterminacy of word segmentation and the nature of morphology and syntax, *Folia Linguistica*, 45(1) 31–80. <https://doi.org/10.1515/flin.2011.002> (In English)

Hunston, S. (2002). *Corpora in Applied Linguistics*, Cambridge University Press, Cambridge. (In English)

Lan, G., Pan, X., Sun, Y. and Lu, Y. (2023). Part of speech tagging of grammatical features related to L2 Chinese development: A case analysis of Stanza in the L2 writing context, *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1139703> (In English)

Langacker, R. W. (1987). *Foundations of Cognitive Grammar, Vol. I: Theoretical Prerequisites*, Stanford University Press, Stanford. (In English)

Levshina, N. (2022). Corpus-based typology: applications, challenges and some solutions, *Linguistic Typology*, 26 (1), 129–160. <https://doi.org/10.1515/lingty-2020-0118> (In English)

Li, C. N. and Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*, University of California Press, Berkeley, CA. (In English)

McEnery, T. and Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*, Cambridge University Press, Cambridge. (In English)

Partee, B. H. (1987). Noun phrase interpretation and type-shifting principles, *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, Groenendijk J., de Jongh D., and Stokhof M.

(Eds.), Foris, Dordrecht, 115–143. <https://doi.org/10.1002/9780470758335.ch15> (In English)

Sapir, E. (1921). *Language, an Introduction to the Study of Speech*, Harcourt, Brace & World, New York. (In English)

Schütze, H. (1995). Distributional part-of-speech tagging, *Proceedings of the Seventh Conference on European Chapter of the Association for Computational Linguistics*, 27–31 March 1995, Dublin, Ireland, 141–148. (In English)

Siewierska, A. (2004). *Person*, Cambridge, University Press Cambridge. (In English)

Sinclair, J. (1991). *Corpus, Concordance, Collocation*, Oxford University Press, Oxford. (In English)

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*, Harvard University Press, Cambridge, MA. (In English)

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*, John Benjamins, Amsterdam. (In English)

Wang, Z., Huang, D., Cui, J., Zhang, X., Ho, S.-B., and Cambria, E. (2025). A review of Chinese sentiment analysis: subjects, methods, and trends, *Artificial Intelligence Review*, 58, 75. <https://doi.org/10.1007/s10462-024-10988-9> (In English)

Wang, X., Bin, S., Chen, I., and Hao, Y. (2026). Beyond Commonalities: A Quantitative Perspective on Syntactic Features across Mandarin Chinese Varieties, *Journal of Quantitative Linguistics*, 33(2), 81–116. <https://doi.org/10.1080/09296174.2025.2567089> (In English)

Xu, Z. (2018). The word status of Chinese adjective-noun combinations, *Linguistics*, 56(1), 207–256. <https://doi.org/10.1515/ling-2017-0035> (In English)

马建忠. 马氏文通. 北京: 商务印书馆. 1983. 447 p. = Ma, C. (1983). *Explanation of Mr. Mas Written Language Rules*, Shanwu Yinshuguan, Beijing. (In Chinese)

黎锦熙. 新著国文语法. 北京: 商务印书馆. 1993. 336 p. = Li, C. (1993). *New Grammar of the National Language*, Shanwu Yinshuguan, Beijing. (In Chinese)

曹杨波.

计算机辅助语言学习的发展与前景 //

中国信息科技. 2011. № 3. P. 107–109. =  
Cao, Ya. (2011). Development and Prospects of  
Language Teaching with the Help of Information  
Technology, *Kitajskie informacionnye tehnologii*  
[Chinese Information Technology], 3, 107–109.  
(*In Chinese*)

张博. 现代汉语同形同音词与多义词.

北京: 北京语言大学对外汉语研究中心. 2002–  
2003. = Zhang, B. (2002–2003). *Homonymy and*  
*Polysemy of Modern Chinese*, Center for the  
Study of Chinese as a Second Language, Beijing  
University of Languages and Culture, Beijing.  
[Online], available at:  
[http://wenku.baidu.com/view/b3eb771afc4ffe4733](http://wenku.baidu.com/view/b3eb771afc4ffe473368ab83.html)  
[68ab83.html](http://wenku.baidu.com/view/b3eb771afc4ffe473368ab83.html) (Accessed 28 August 2025). (*In*  
*Chinese*)

代建桃. 现代汉语同音词研究: 论文 ...

汉语言文字学; 四川师范大学, 成都. 2008. 99 p.  
= Dai, Q. (2008). Study of homonyms in modern  
Chinese, Ph.D. Thesis, Sichuan Normal  
University. Chengdu, China. (*In Chinese*)

邵敬敏. 现代汉语通论. 上海:

上海教育出版社. 2007. 335 p. = Shao, C. (2007).  
*General Introduction to Modern Chinese*,  
Shanghai Education Press, Shanghai. (*In Chinese*)

### Corpus Material

Akhmanova, O. S. (1966). *Slovar*  
*lingvisticheskikh terminov* [Dictionary of  
Linguistic Terms], Sov. Encyclopedia, Moscow,  
Russia. (*In Russian*)

Russian National Corpus [Online],  
available at: <http://www.ruscorpora.ru> (Accessed  
28 August 2025). (*In Russian*)

Ozhegov, S. I., Shvedova, N. Yu. (2006).  
*Tolkovy slovar russkogo yazyka: 80 000 slov i*  
*frazologicheskikh vyrazheniy* [Explanatory  
Dictionary of the Russian Language: 80,000  
Words and Phraseological Expressions], TEMP  
Publ., Moscow, Russia. (*In Russian*)

Rosenthal, D. E., Golub, I. B.,  
Telenkova, M. A. (2010). *Sovremenny russkiy*  
*yazyk* [Modern Russian Language], Iris-press  
Publ., Moscow, Russia. (*In Russian*)

语料库在线 = Online corpus [Online],  
available at: <http://www.cncorpus.org/> (Accessed  
28 August 2025). (*In Chinese*)

孙浩成 现代汉语语法大全. = Sun, H.  
*Grammar of Modern Chinese* [Online], available

at: [http://wenku.baidu.com](http://wenku.baidu.com/view/0ea2c833433239-68011c9253.html)  
[/view/0ea2c833433239-68011c9253.html](http://wenku.baidu.com/view/0ea2c833433239-68011c9253.html)  
(Accessed 28 August 2025). (*In Chinese*)

现代汉语词典 /

中国社会科学院语言研究所. 北京:

商务印书馆. 2016. 1799 p. = Dictionary of  
Modern Chinese / Institute of Linguistics,  
Academy of Social Sciences of the Peoples  
Republic of China (2016), Shanwu Yinshuguan,  
Beijing. (*In Chinese*)

中学教学语法系统提要 (试行) /

人民教育出版社中学语文室. 北京:

人民教育出版社. 1984. 21 p. = Curriculum for  
the Grammar System of Middle Schools  
(Experimental Mode) / Department of Language  
and Literature for Middle Schools, (1984).  
Peoples Educational Publishing House, Beijing.  
[Online], available at:  
<https://eol.shzu.edu.cn/meol/analytics/> (Accessed  
28 August 2025). (*In Chinese*)

*Авторы прочитали и одобрили*  
*окончательный вариант рукописи.*

*The authors have read and approved the*  
*final manuscript.*

*Конфликты интересов: у авторов нет*  
*конфликтов интересов для декларации.*

*Conflicts of interests: the authors have no*  
*conflicts of interest to declare.*

Александр Аркадьевич Баркович, доктор  
филологических наук, профессор, заведующий  
кафедрой теоретического и славянского  
языкознания, Белорусский государственный  
университет, Минск, Республика Беларусь.

Aliaksandr A. Barkovich, Doctor of Philology,  
Professor, Head of the Department of Theoretical  
and Slavic Linguistics, Belarusian State  
University, Minsk, Republic of Belarus.

Цин Ван, кандидат филологических наук,  
независимый исследователь, Минск,  
Республика Беларусь.

Qing Wang, PhD in Philology, independent  
researcher, Minsk, Republic of Belarus.