

На правах рукописи

УДК:81'33

Хоу Линьсюе

**ПРОПЕДЕВТИЧЕСКИЙ УЧЕБНЫЙ КОРПУС РУССКОГО
ЯЗЫКА: МЕТОДЫ СОЗДАНИЯ И МЕТОДИЧЕСКИЙ
ПОТЕНЦИАЛ ДЛЯ ОБУЧЕНИЯ НОСИТЕЛЕЙ КИТАЙСКОГО
ЯЗЫКА**

Специальность 5.9.8. Теоретическая, прикладная и сравнительно-
сопоставительная лингвистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата филологических наук

Санкт-Петербург

2024

Работа выполнена на кафедре образовательных технологий в филологии Федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Российский государственный педагогический университет им. А. И. Герцена».

Научный руководитель:

заслуженный деятель науки РФ, доктор филологических наук, профессор, профессор кафедры образовательных технологий в филологии федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Российский государственный педагогический университет им. А. И. Герцена»

Беляева Лариса Николаевна

Официальные оппоненты:

доктор филологических наук, профессор, профессор департамента филологии Санкт-Петербургской школы гуманитарных наук и искусств, заведующий лабораторией языковой конвергенции федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики»

Колмогорова Анастасия Владимировна

кандидат филологических наук, доцент, доцент кафедры математической лингвистики федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный университет»

Митрофанова Ольга Александровна

Ведущая организация:

Федеральное государственное бюджетное образовательное учреждение высшего образования «Иркутский государственный университет»

Защита состоится **21 мая 2024 г. в 16:00** часов на заседании Совета по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук 33.3.018.08, созданного на базе Российского государственного педагогического университета им. А. И. Герцена, по адресу: 199053, г. Санкт-Петербург, 1-я Линия ВО, дом 52, ауд. 22.

С диссертацией можно ознакомиться в фундаментальной библиотеке Российского государственного педагогического университета им. А. И. Герцена (191186, г. Санкт-Петербург, наб. р. Мойки, 48, корп.5) и на сайте университета по адресу: https://disser.herzen.spb.ru/Preview/Vlojenia/000001008_Disser.pdf

Автореферат разослан « ____ » _____ 2024 г.

Ученый секретарь диссертационного совета
Камшилова Ольга Николаевна

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Реферируемая диссертационная работа посвящена исследованию методов создания учебного корпуса русского языка и его методического потенциала для обучения китайских студентов.

По мере всестороннего развития отношений между КНР и РФ в последние годы все больше университетов в Китае открывают специальность «Русский язык» для обучения специалистов в области русского языка, готовых решать технические, научные и методические задачи. Соответственно, активизируются процессы преподавания русского языка в КНР, при этом анализ результатов демонстрирует необходимость повышать качество и эффективность обучения русскому языку.

Особое внимание акцентируется на учебном корпусе, создание которого позволяет оценивать результаты освоения иностранного языка.

Актуальность темы исследования обусловлена тем, что мало исследований, посвященных учебным корпусам текстов, которые отражают результат овладения русским языком носителями китайского языка в процессе его усвоения как вторым языком. Кроме того, остается неясным, каково возможное влияние изучения других иностранных языков, в частности, английского, на эффективность обучения другим языкам, в данном случае русскому.

Степень научной разработанности темы. С ростом доступности информационных технологий в последние годы в современных лингвистических исследованиях все чаще используются репрезентативные корпуса текстов. В разное время исследованиями в сфере корпусной

лингвистики занимались такие российские ученые, как А. В. Зубов, В. В. Рыков, В. А. Плунгян, В. П. Захаров, С. Ю. Богданова, Л. Н. Беляева, О. Н. Камшилова, Е. В. Рахилина и др. Особенное внимание на изучение проблемы в области корпусной лингвистики обращают такие китайские авторы как Лян Маочэн, Сяо Чжунхуа, Сюй Цзяцзинь, Ван Кэфэй, Ван Лифи, Лю Мяо и др. В их работах рассматриваются и разрабатываются классификация корпусов по различным критериям, методы и принципы создания лингвистических корпусов русского и иностранного языков, а также использование корпусов в лингвистических исследованиях.

Как в России, так и в Китае проблема создания и применения учебного корпуса привлекает огромный интерес ученых, подобным исследованиям и разработкам посвящены работы А. И. Ольховской, М. К. Парамоновой, Е. В. Рахилиной, Н. Р. Добрушиной, О. Н. Камшиловой, Гуй Шичун, Вэй Найсин, Вэнь Цюфань, Ян Хуэйчжун и других исследователей.

Однако в этих исследованиях не рассматриваются проблемы создания учебного корпуса русского языка для обучения китайских учащихся, а также система классификации и разметки ошибок, допускаемых китайскими студентами и выявленными на основе учебного корпуса текстов.

Теоретическую основу исследования составили труды отечественных и зарубежных ученых, определяющие следующие аспекты темы:

- лингвистический корпусный подход нацелен на практическое изучение языка в точных ситуациях и материалах, что представляется особо важным для изучения и преподавания языка (В. П. Захаров, Е. П. Соснина, Л. И. Агафонова, Л. К. Раицкая, П. В. Сысоев, С. Ю. Богданова, Т. McEneary, Y. Tono, J. Grigaliūnienė, G. Davies, H. Mortensen, M. Lamy);

- учебный корпус опирается на извлечение, классификацию и разметку ошибок в текстах, он представляет собой массив языковых данных, особым образом размеченный для решения конкретных задач обучения языкам (Е. П. Соснина, О. Н. Камшилова, С. С. Замичева, С. Ю. Богданова, S. Granger, J. Grigaliūnienė, Лян Маочэн, Гуй Шичун, Ян Хуэйчжун);

- ошибки являются неизбежной и неотъемлемой частью процесса изучения иностранного языка, с помощью их анализа на основе учебного корпуса можно выявлять трудности учащихся при овладении иностранными языками (Е. В. Тетерлева, Ю. В. Красиков, Ю. К. Попова, С. Н. Цейтлин, С. James, P. Corder, S. Granger);

- создание учебного корпуса русского языка требует предварительного рассмотрения общего подхода к классификации ошибок, допускаемых китайскими студентами при овладении русским языком (Е. А. Гасконь, Е. В. Жендаренко, М. О. Сорокина, Ю. И. Злобина, Ву Цзюань, Ли Гоучжин, Лю Цянь);

- разработка принципов и методов разметки корпуса является основой для создания корпуса и увеличения ценности его применения (В. П. Захаров, Е. В. Рахилина, Е. С. Елохов, В. А. Плунгян, М. А. Куниловская, С. С. Замичева, G. Leech, M. E. Foster, Чжао Вэйн);

- исследование изучения и преподавания иностранных языков необходимо проводить, учитывая влияние родного и второго иностранного языков на процесс овладения новым иностранным языком (А. А. Щелокова, Н. Д. Гальскова, Н. П. Панжиев, Р. П. Мильруд, С. П. Муродова, A. Wilson, G. Leech, P. Rayson).

Целью работы является разработка принципов и процедур

количественного и качественного анализа ошибок в исследовательском учебном корпусе русского языка китайских студентов для оценки методического потенциала учебного корпуса.

Для достижения поставленной цели необходимо решить следующие **задачи**:

1) проанализировать опыт лингвистических исследований в области создания учебных корпусов;

2) определить состав, структуру, алгоритм разработки учебного корпуса и сферу его применения;

3) собрать и обработать текстовый материал для формирования корпуса;

4) определить понятие ошибки при создании учебного корпуса с целью их идентификации; провести классификацию ошибок и разработать систему их разметки;

5) разработать пилотный вариант учебного корпуса текстов, который можно рассматривать как пропедевтический корпус, в котором определены основные особенности и номенклатура ошибок, и апробировать составленный корпус с точки зрения извлечения ошибок разных типов;

б) провести количественный анализ частотных и типичных ошибок и оценить методический потенциал учебного корпуса.

Объектом исследования является учебный корпус русского языка для обучения китайских студентов как особый вариант исследовательского корпуса текстов.

Предмет исследования составляет методический потенциал учебного корпуса, созданного для выявления типичных ошибок в письменных работах китайских студентов, написанных на русском языке.

Гипотеза исследования: пропедевтический учебный корпус русского языка для обучения китайских студентов, отобранный в соответствии с требованиями корпусной лингвистики, размеченный по определенному принципу и обеспеченный специализированной поисковой программой, является базой данных, достаточной для количественного анализа ошибок и анализа влияния родного и английского языков на изучения русского языка.

Методологическая база исследования включает в себя наблюдение, описание, корпусные методы лингвистического анализа, метод классификации, количественный анализ, сопоставительный анализ, методы компьютерной обработки данных, методы информационных технологий.

Материалом исследования послужили письменные тексты, написанные китайскими студентами, обучающимися по специальности русский язык, поступившими во Второй Пекинский университет иностранных языков и Хэнаньский университет науки и технологий в 2019 году. Общий объем текстов составляет 457 сочинений (82 064 слова).

На защиту выносятся следующие положения:

1) использование корпусных данных и технологий при преподавании иностранных языков становится необходимостью. Применение современных образовательных технологий в обучении русскому языку как иностранному обуславливает создание исследовательского учебного корпуса как вспомогательного инструмента для повышения эффективности и качества обучения русскому языку;

2) эффективность и репрезентативность исследовательского учебного корпуса русского языка для анализа ошибок обеспечивается корректностью установления специальных критериев отбора текстов, написанных

китайскими студентами на русском языке, определением, идентификацией и классификацией ошибок при овладении русским языком и применением методов и принципов разметки на основе системы разметки *TEI*;

3) методика анализа ошибок, допускаемых китайскими студентами при создании текстов на русском языке, основывается на сочетании количественного и сопоставительного подходов. Определение оценок уровня знаний и частоты ошибок с помощью предложенного в работе коэффициента *K*, выделение частотных слов в текстах китайских студентов путем использования корпусного менеджера *AntConc* позволяют установить влияние китайского и английского языков на изучение русского языка и выявить невидимые и типичные ошибки китайских студентов при создании текстов на разные темы;

4) созданный учебный корпус русского языка в качестве базы данных и результаты, полученные в ходе исследования, дают возможность установить методический потенциал учебного корпуса, использование которого позволяет эффективно повышать качество и улучшать методы обучения русскому языку.

Научная новизна исследования определяется тем, что в диссертации впервые:

1) разработана методика создания исследовательского учебного корпуса русского языка для обучения китайских студентов;

2) предложены методы идентификации и классификации ошибок, допускаемых китайскими студентами при овладении русским языком с учетом влияния китайского языка на изучение русского языка;

3) обобщены принципы и методы разметки для создания учебного

корпуса русского языка и анализа ошибок;

4) выявлены и изучены типичные и частотные ошибки, допускаемые китайскими студентами при создании текстов на русском языке; установлено влияние китайского и английского языков на овладение русским языком;

5) выявлен методический потенциал учебного корпуса для преподавания и изучения русского языка.

Теоретическая значимость результатов исследования заключается в том, что разработка принципов и методов классификации и разметки ошибок, допускаемых при овладении русским языком, и методов построения учебного корпуса русских текстов для китайских учащихся вносит вклад в развитие корпусной лингвистики и использования корпусов для обучения иностранным языкам. Выводы, полученные на основе проведенного анализа, обеспечивают теоретическое исследование проблем овладения русским языком в Китае.

Достоверность и научная обоснованность теоретических и практических результатов исследования обеспечивается применением методов использования учебного корпуса в обучении русскому языку, что позволило провести исследование на аутентичном языковом материале, написанном самими китайскими студентами. Достоверность полученных результатов подкреплена объективными количественными данными об анализе ошибок, допускаемых китайскими студентами при овладении русским языком.

Практическая значимость исследования состоит в том, что

- разработанная в диссертации методика может служить методологической основой дальнейших лингвистических исследований учебных корпусов русских текстов;

- выявление типичных и частотных ошибок в письменной речи китайских учащихся предоставляет языковые данные для преподавателей при оптимизации процесса обучения русскому языку;

- разработанный пилотный вариант учебного корпуса русского языка может рассматриваться как пропедевтический корпус (учебный корпус был загружен в систему Интернет на сайт Яндекс-Диск [<https://disk.yandex.ru/d/eNgPH51q4nOMlw>]), что позволяет его использовать как вспомогательный инструмент на занятиях по русскому языку как иностранному;

- созданный учебный корпус русского языка для обучения китайских студентов может быть использован в лексикографических исследованиях, направленных на создание русско-китайского учебного словаря.

Рекомендации по использованию результатов диссертационного исследования. Основные результаты исследования могут быть использованы на теоретических и практических занятиях, при обучении методам создания и применения учебного корпуса, а также при разработке специальных курсов для учителей или преподавателей, которые позволяют им приобрести навыки использования сетевых лингвистических ресурсов с целью решения прикладных лингвистических задач. Созданный исследовательский учебный корпус русского языка может быть использован преподавателями для работы по обучению русскому языку как иностранному и для создания современных дидактических материалов, а также учащимися для самостоятельной работы.

Апробация результатов исследования. Основные положения и результаты исследования были изложены в докладах на следующих научных и практических конференциях: Международная научная конференция «Язык.

Коммуникация. Культура» (Гродно, Беларусь, 28–29 октября 2021 г.); IX (XXIII) Международная научно-практическая конференция молодых ученых «Актуальные проблемы лингвистики и литературоведения» (Томск, 14–16 апреля 2022 г.); VI Международная научно-практическая конференция «Российско-китайское сотрудничество: на пути к глобальному устойчивому развитию» (Новосибирск, 10–11 ноября 2022 г.); 25 Межвузовская студенческая научная конференция «Студент – Исследователь – Учитель» (Санкт-Петербург, 3–17 апреля 2023 г.). По теме диссертации опубликовано 5 работ, из них 3 – в журналах, рекомендованных ВАК РФ.

Структура и объем диссертационной работы определены поставленной целью и задачами. Диссертация состоит из введения, трех глав и выводов к ним, заключения, списка использованной литературы, включающего 182 наименования, из них 124 наименования на русском языке, 29 наименований на английском языке и 29 наименований на китайском языке, и 7 приложений. Общий объем диссертационного исследования составляет 215 страниц печатного текста, из них 177 страницы занимает основное содержание работы, включая 16 таблиц и 17 рисунков, и приложения изложены на 38 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обосновывается тема исследования, ее актуальность и степень научной разработанности, определяются объект, предмет, цель, задачи, гипотеза, теоретическая основа и методы исследования, даются сведения о языковом материале, раскрыта его научная новизна, излагаются положения, выносимые на защиту, аргументируется теоретическая и

практическая значимость, даны рекомендации по использованию результатов исследования, приводятся сведения об апробации результатов работы.

В первой главе диссертации «Корпусная лингвистика в современных исследованиях» описываются этапы развития корпусной лингвистики и базовая типология корпусов; рассматриваются образцы корпусов на различных языках, особое внимание уделяется корпусам, созданным в РФ и КНР, в целях анализа места корпусной лингвистики в сфере образования; исследуются педагогические характеристики учебного корпуса и его возможности при применении в образовательной деятельности; характеризуется обучающий потенциал корпусных технологий в изучении и преподавании русского языка.

По мере развития компьютерных технологий современные лингвистические корпуса демонстрируют огромный потенциал, а корпусные технология и методы все шире внедряются в разные области современных исследований, особенно в сфере образования. Корпусы характеризуются репрезентативностью, сбалансированностью и ориентированностью, и именно эти делает их применение важным в области лексикографии, лингводидактики, преподавания иностранных языков и переводоведения.

Как особый вид корпуса, учебные корпуса основываются на аутентичных данных учащихся в процессе изучения ими иностранного языка и фокусируются на ошибках учащихся. Используя учебные корпуса, преподаватели могут быстро и точно анализировать ошибки, допускаемые учащимися в процессе овладения иностранным языком, чтобы постоянно корректировать свою педагогическую деятельность и более эффективно обучать иностранному языку. Кроме того, учебные корпуса предоставляют

самим учащимся богатые примеры и реальные контексты ошибок, чтобы они могли легко принимать и запоминать языковые правила иностранного языка.

В последнее время активизировалась разработка учебных корпусов в Китае и России. К созданным корпусам относятся Санкт-Петербургский учебный корпус текстов школьников, изучающих английский язык (*SPbEFL LC*), Корпус русских учебных текстов (КРУТ), Хельсинкский аннотированный корпус русских текстов (ХАНКО), Устный учебный корпус русского языка как иностранного, Корпус учебных переводов (*Russian Learner Translator Corpus*), Учебный корпус английского языка при Гонконгском университете науки и технологии, Китайский учебный корпус текстов английского языка и другие. Несмотря на то, что использование корпусных технологий в преподавании иностранного языка привлекает внимание российских и китайских ученых, однако учебные корпусы, которые фокусируются на реальной ситуации обучения китайских студентов при овладении ими русским языком, разрабатываются и используются крайне мало, кроме того, отсутствуют соответствующие исследования по стандартам построения корпуса, принципам и методам разметки и т. д.

Эффективность использования корпусных технологий в педагогической деятельности и отсутствие учебных корпусов русского языка, направленных на обучение китайских студентов, являются основанием, подтверждающим актуальность разработки пилотного варианта учебного корпуса русского языка для обучения китайских студентов и анализа его методического потенциала.

Во второй главе диссертации «Создание учебного корпуса русского языка для обучения китайских студентов» определяются процедура

реализации каждого этапа разработки учебного корпуса и критерии отбора текстов, написанных китайскими студентами на русском языке, для создания учебного корпуса; описываются принципы и методы проведения метаразметки и разметки ошибок в учебном корпусе русского языка; исследуется понятие ‘ошибка’ при обучении русскому языку для идентификации и извлечения ошибок в текстах; предлагаются теги классификации ошибок, отмеченных в текстах для создания учебного корпуса русского языка.

а. Отбор письменных текстов на русском языке, написанных китайскими студентами для создания учебного корпуса

С целью выявления и анализа ошибок, допускаемых китайскими студентами в процессе овладения русским языком, обеспечения аутентичности текстов учебного корпуса русского языка, при создании корпуса были установлены следующие критерии отбора письменных текстов: обычные домашние задания с заданными темами (в учебный корпус включаются только обычные домашние задания с темами, которые близки жизни учащихся); определенный источник текстов (в корпус отбираются тексты, написанные китайскими студентами с разным, но конкретным уровнем владения английским языком); репрезентативность (увеличение корпуса не ведет к увеличению числа типов ошибок и изменения их типологии).

Согласно всем перечисленным выше критериям для создания учебного корпуса русского языка было использовано 457 письменных текстов по 10 темам, написанных 90 студентами из китайских университетов, обучающихся по специальности «Русский язык». Все тексты преобразованы из

формата *.dock* в формат *.txt* для компьютерной обработки. Объем текстов, включенных в учебный корпус русского языка, составляет 82 064 слова. Полные тексты были загружены в систему Интернет на сайт Яндекс-Диск [<https://disk.yandex.ru/d/FIWmfjgIejynBA>].

б. Особенности метаразметки и разметки ошибок в учебном корпусе русского языка

Разметка текстов при создании китайского учебного корпуса русского языка проводилась в соответствии со следующими принципами: разметка должна быть независима от текста и при необходимости должна предусматриваться возможность убирать разметку; теги разметки должны отличаться от представления текстов в корпусе; теги разметки должны быть простыми и индивидуальными для практической применимости; устройство разметки ошибок должно быть иерархическим, обеспечивающим обзорность классов ошибок учащихся.

Опираясь на систему разметки (стандарт) *TEI*, можно разделить виды разметки китайского учебного корпуса русского языка на две части:

1) Первая — метаразметка (экстралингвистическая разметка) включает возраст и пол авторов, тему текста, степень владения английским языком для анализа ошибок. Уровень английского языка определялся в соответствии с баллами по иностранному языку, полученными на ЕГЭ (см. Табл. 1).

Таблица 1

Теги уровня английского языка в учебном корпусе русского языка

Оценка по английскому языку на ЕГЭ	Тег английского языка
150-135	A
134-120	B
119-105	C
104-90	D

Так, например, тег *<4.23.20.F.C>* означает, что тема текста – Мои

увлечения, автор – студент с кодом 23, возраст 20 лет, пол женский, баллы по английскому языку находятся в интервале 119–105. С помощью тегов метаразметки можно позволить пользователям не только найти подходящие языковые материалы в соответствии с их собственными потребностями, но и предоставить данные для анализа причин возникновения ошибок учащихся и особенностей овладения ими русским языком.

2) Вторая — разметка ошибок, которые помечаются строчными латинскими буквами и цифрами, заключенными в угловые скобки <...>, например: *Я люблю путешествовать и <побывал много мест yr3> Китая.* Разметка ошибок проведена автором диссертации, ее результаты выборочно проверялись 3 экспертами — преподавателями русского языка во Втором Пекинском университете иностранных языков и Хэнаньском университете науки и технологий.

в. Идентификация и классификация ошибок, допускаемых китайскими студентами при овладении русским языком

Для того чтобы выявлять и классифицировать ошибки, необходимо определить, что понимается под термином *ошибка*. В учебном корпусе русского языка термин *ошибка* поднимается как нарушение в речи китайских студентов, противоречащее правилам и нормам русского языка и непонятное носителям русского языка. Следует отметить, что идентификация ошибок в учебном корпусе русского языка в основном зависит от параметров грамматичности и правильности.

Разработанная классификация ошибок в основном связана с различием между китайским и русским языками и определяется механизмом разметки ошибки. При проведении предварительного исследования массива текстов

также наследовался опыт классификации ошибок на основе Учебного корпуса английского языка при Гонконгском университете науки и технологии, Китайского учебного корпуса текстов английского языка. В итоге ошибки, допускаемые китайскими студентами при овладении русским языком, разделены на 5 основных типов и 14 подкатегорий, всего выделено 34 варианта (см. Табл. 2).

Таблица 2

Классификация ошибок с тегами в учебном корпусе русского языка

Тип ошибки	Подкатегория ошибки	Название ошибки	Тег	
орфографические ошибки	нарушено правописание	орфографическая ошибка за счет незнания орфограмм	fm1	
		неверное словообразование	fm2	
		неверное использование прописной буквы	fm3	
морфологические ошибки	ошибка в использовании глаголов	неверное использование возвратной формы	vp1	
		неверное использование форм несовершенного или совершенного вида	vp2	
		неверное управление глагола	vp3	
		неверное использование деепричастия	vp4	
		неверный выбор вида причастия	vp5	
	ошибка в использовании имен существительных	ошибка в использовании имен существительных	неверное использование форм числа имен существительных	np1
			отсутствие согласования по роду	np2
			неверное образование форм склонения	np3
			нарушение падежных форм	np4
	ошибка в использовании предлогов	ошибка в использовании предлогов	неверное управление предлога	pp1
			использование неверного предлога	pp2
	ошибка в использовании имен прилагательных	ошибка в использовании имен прилагательных	неверное использование полной или краткой формы	aj1
			неверное использование степеней сравнения	aj2
	ошибка в использовании местоимений	ошибка в использовании местоимений	неверный выбор местоимения	pr1

морфологические ошибки	ошибка в использовании наречий	неверное использование степеней сравнения	ad1
		неверное использование форм долженствования	ad2
	ошибка в использовании числительных	неверное использование форм числительных	nu1
лексические ошибки	ошибка в использовании слова	употребление слова в несвойственном ему значении	wd1
		тавтология и повтор	wd2
		смешение синонимов и паронимов	wd3
	неверное использование устойчивых словосочетаний	неверное использование устойчивых выражений	cc1
		фразеологическая ошибка	cc2
синтаксические ошибки	нарушение структуры предложений	пропуск членов предложения	sn1
		неверный порядок слов	sn2
	нарушение норм формального синтаксиса	нарушение согласования подлежащего и сказуемого	sn3
		нарушение видовойременной соотнесенности однородных сказуемых или сказуемых в главном и придаточном предложениях	sn4
	пунктуационная ошибка	отсутствие знака препинания	pt1
		лишний знак препинания	pt2
		неверное использование знака препинания	pt3
	коммуникационные ошибки	логическое нарушение	нарушение логики при перечислении
отсутствие формальных показателей связи между предложениями			lc2

Полный учебный корпус русского языка был загружен в систему Интернет на сайт Яндекс-Диск [<https://disk.yandex.ru/d/eNgPH51q4nOMlw>].

Третья глава диссертации «Количественный анализ ошибок, зафиксированных в учебном корпусе русского языка, как база методической поддержки обучения» посвящена анализу ошибок, допускаемых китайскими студентами в процессе овладения русским языком. В этой главе предлагается метод оценки уровня знаний и частоты ошибок, для вычисления которого вводится коэффициент K ; рассматриваются частотные ошибки, допускаемые китайскими студентами при создании текстов на

русском языке в аспекте интерференции китайского и английского языков; осуществляется сопоставление частотных слов в текстах китайских студентов и речи носителей русского языка с целью выявления невидимых ошибок китайских учащихся при изучении русского языка; описываются типичные ошибки в текстах на русском языке с разными темами, устанавливается методический потенциал созданного учебного корпуса.

Для оценки уровня знаний и частоты допускаемых ошибок китайскими студентами при овладении русским языком предлагается использовать относительное значение числа ошибок, для вычисления которого вводится коэффициент K :

$$K = F_f / L$$

где F_f – частота ошибки типа f в исследуемом тексте, L – объем текста в словах.

Значения этого коэффициента вычисляются в интервале $0 < K < 1$, при этом, чем ниже значение коэффициента, тем выше уровень знаний. Коэффициент вычисляется для разных авторов, тексты которых вошли в учебный корпус, и для разных типов ошибок, определяемых построенной типологией.

При анализе частоты типов ошибок важно обратить внимание на то, что наибольшее значение коэффициента K наблюдается при ошибках в использовании глаголов (см. Табл. 3). В соответствии с данными, представленными в Таблице 3, китайские студенты сталкиваются с большими трудностями при овладении русским языком, особенно теми грамматическими правилами, которые не соответствуют китайскому языку, такие как управление глаголов, использование глаголов несовершенного или совершенного вида. Значение коэффициента K при неверном использовании

возвратной формы глагола ниже, чем при других ошибках, т. е. студенты допускают меньше ошибок этого типа. Это связано с тем, что возвратная форма глагола в русском языке имеет сходное грамматическое соответствие в китайском языке.

Таблица 3

Значения коэффициента K при ошибках в использовании глаголов

Подкатегория ошибки	Название ошибки	коэффициент K
ошибка в использовании глаголов	неверное использование возвратной формы	0,000195
	неверное использование несовершенного или совершенного вида	0,001401
	неверное управление глагола	0,002267
	неверное использование форм деепричастия	0,000037
	неверный выбор типа причастия	0,000110

Авторы текстов, включенных в учебный корпус русского языка имеют богатый опыт изучения английского языка, именно поэтому английский язык как первый иностранный язык также оказывает влияние на овладение русским языком, особенно на начальных этапах. Основываясь на метаразмечке учебного корпуса русского языка, можно сделать вывод, что в целом студенты, хорошо успевающие по английскому языку, лучше овладевают русским языком. Уже приобретенные студентами знания английского языка, а также привычные методы обучения могут облегчить им изучение русского языка. Однако некоторые грамматические правила в английском языке, не соответствующие русскому языку, все же могут оказывать негативное влияние на изучение русского языка, например, из-за разного управления глаголов и рода имен существительных в русском и английском языках.

В процессе изучения русского языка встречаются невидимые ошибки, которые обнаруживаются при достаточном накоплении языковых материалов.

И в этом случае незаменимую роль играет учебный корпус. С помощью сопоставительного исследования речи китайских студентов и носителей русского языка на основе НКРЯ были обнаружены некоторые невидимые ошибки китайских учащихся, такие как расширение употребления конструкции *у кого-чего* и предлога *для*, и недостаточное использование глагола *быть* и предлога *к* из-за непонимания их точного значения.

Учебный корпус русского языка включает в себя письменные тексты по 10 различным темам, близким к жизни учащихся (см. Табл. 4).

Таблица 4

Темы письменных текстов в учебном корпусе русского языка

Тема письменных текстов	Тема письменных текстов
Мое интересное путешествие	Книга в нашей жизни
Мой родной город/Мой любимый город	Как я провел(а) свой день рождения
Интернет в нашей жизни	Моя мечта
Мои увлечения	Как мы провели праздник Весны
Моя студенческая жизнь	Самый любимый человек

Распределение типичных ошибок варьируется от темы к теме:

- В темах «Мое интересное путешествие» и «Мои увлечения» речь шла о путешествиях и хобби, что заставило студентов использовать больше разных глаголов, поэтому типичной ошибкой в этих двух темах стало неверное управление глаголов, например *Бадминтон, этот самый <быстрый aj1> спорт в мире, <мне очень интересуют vr3 sn3>*. вместо... *меня очень интересуется...*
- В теме «Самый любимый человек» рассказывается о любимом человеке, больше всего ошибок появляется в описании личности и внешности персонажа, см., например, *Я очень люблю <моей pr1 мамы pr3>* вместо ... *люблю свою маму. Я больше люблю его <глазы pr3>*: *<эти pr1> острые*

чёрные <глазы пр3>. вместо глаза.

- В теме «Как мы провели праздник Весны», где используются словосочетания, которые характеризуют такие особенности китайских праздников как, например, 放鞭炮 (запускать петарды), 贴春联 (наклеивать парные надписи с новогодними пожеланиями), студенты часто используют буквальный перевод, что приводит к тому, что такие словосочетания не соответствуют требованиям русского языка, см., например, <срезать красивые цветы сс1> вместо срезать бумажные украшения на окна, <стрелять из бамбука сс1> вместо запускать петарды.

- В теме «Как я провел(а) свой день рождения» китайские студенты допускают пропуск членов предложения из-за буквального перевода с китайского языка на русский, такие как <Мой день рождения в сентябре sn1.> вместо У меня день рождения в сентябре. <Этот день каждого года самым счастливым для меня sn1.> вместо Этот день каждого года является самым счастливым для меня.

С учетом результатов анализа ошибок на основе учебного корпуса преподавателям рекомендуется направлять студентов на изучение различий и сходств между китайским, английским и русским языками, чтобы помочь им легче понимать соответствующие грамматические правила. В связи с выявлением невидимых ошибок появляется возможность составлять учебные материалы, более соответствующие реальным потребностям студентов, так, например, в учебном словаре можно провести специальную разметку для того, чтобы напомнить учащимся о грамматических правилах, в которых они часто допускают ошибки (см. например, играть: во что 玩 (球类, 棋牌 в мяч 打篮球, в карты 打牌, в компьютерные игры 玩电脑游戏), на чем 弹, 演奏 (乐

器 на рояле 弹钢琴, на гитаре 弹吉他) .

Создание пилотного проекта учебного корпуса русского языка дает возможность его использования при решении различных методических задач, при отборе примеров для разработки таких упражнений как задание для определения значения слова или словосочетания с помощью учебного корпуса русского языка (Рис. 1).

Задание: изучите данные конкорданса и определите значение словосочинения «так как»←

евней столицей шести династий». Мне нравится путешествие, **так как** оно предоставляет мне возможность познать многогранные и
↳ поехать за границу с родителями, особенно в Россию, **так как** там я могу получать языковую практику и
поделиться. Я не представляю свою жизнь без движения, **так как** мне нравится активный отдых, осмотр достопримечательности
печатления. Я не представляю свою жизнь без путешествий, **так как** посещение новых мест расширяет мне кругозор, обогащает
путешествовать, и не представляю свою жизнь без движения, **так как** мне нравится активный отдых. Мне очень хотелось
ем могут поделиться. Обожаю путешествовать на велосипеде, **так как** мне нравятся находиться в гармонии с природой,
путешествовать. Я не представляю свою жизнь без движения, **так как** мне нравится активный отдых, осмотр достопримечательности
себя. Именно тогда я подумала о сельском туризме, **так как** считала, что смена обстановки помогает восстановиться. Мо

Рисунок 1. Задание для определения значения словосочетания *так как* с помощью учебного корпуса русского языка

В итоге учебный корпус текстов при обучении русскому языку обладает большим методическим потенциалом, поскольку анализ ошибок учащихся при овладении русским языком на основе учебного корпуса русского языка позволяет расширить методы преподавания русского языка, и также помочь учащимся точно и легко овладеть грамматическими правилами русского языка.

ЗАКЛЮЧЕНИЕ

Исследование проблемы анализа ошибок, допускаемых китайскими студентами в процессе овладения русским языком на основе созданного специализированного исследовательского учебного корпуса русского языка, позволило получить следующие результаты:

1. Корпусная лингвистика является одним из наиболее перспективных направлений в современной теоретической и прикладной лингвистике, активно внедряющимся в различные области науки, особенно в сфере образования. Характеризуясь репрезентативностью, сбалансированностью и ориентированностью, корпуса демонстрируют огромный потенциал в области лексикографии, лингводидактики, переводоведения и обучения языкам. Как особый тип корпуса, учебные корпуса все шире внедряются в обучение иностранным языкам с целью решения практических задач и повышения эффективности в этом процессе. Аутентичные языковые материалы учащихся, систематизированная разметка позволяют учебным корпусам играть незаменимую роль в преподавании и изучении иностранных языков.

2. По мере совершенствования информационных технологий, применяемых в педагогической деятельности, развивается создание и применение учебного корпуса в сфере преподавания русского языка как иностранного. Практика разработки и использования корпусов в этой сфере в Китае только начинает развиваться. При этом большинство исследований по овладению русским языком китайскими учеными используют материалы из НКРЯ, созданном в России. В этом случае отсутствие учебных корпусов, составленных на основе текстов русского языка, написанных китайскими студентами по специальности «Русский язык», подтверждает востребованность разработки пилотного варианта такого корпуса для исследования реальных достижений китайских учащихся при обучении русскому языку.

3. Ориентируясь на цели данной работы, процедура создания учебного корпуса русского языка состояла из трех основных этапов:

1) Отбор письменных текстов на русском языке, написанных китайскими студентами для создания учебного корпуса.

При создании исследовательского учебного корпуса русского языка были установлены специальные критерии отбора письменных текстов: во-первых, в учебный корпус включались только обычные домашние задания с темами, близкими к жизни учащихся; во-вторых, в учебный корпус отбирались тексты, написанные китайскими студентами с разным, но определенным уровнем владения английским языком; в-третьих, учебный корпус, предназначенный для анализа ошибок китайских студентов при овладении русским языком, должен быть репрезентативен.

Согласно всем перечисленным выше критериям для создания учебного корпуса русского языка были использованы 457 письменных текстов по 10 темам, написанные 90 студентами, обучающимися по специальности «Русский язык» в китайских университетах.

2) Проведение разметки ошибок и метаразметки в учебном корпусе.

Разметка ошибок в учебном корпусе русского языка проводилась согласно стандарту *TEI*, в котором был установлен конкретный метод разметки метаинформации и информации о тексте. Опираясь на систему разметки *TEI*, виды разметки учебного корпуса русского языка были разделены на две части: первая — метаразметка (экстралингвистическая разметка), которая включает такую информацию как тема, возраст, пол и уровень владения английским языком авторами текстов. Вторая — собственно разметка ошибок, которые допускают китайские студенты при овладении русским языком.

3) Идентификация и классификация ошибок, допускаемых китайскими

студентами при овладении русским языком.

В данном исследовании термин *ошибка* был определен как нарушение грамматичности, правильности и приемлемости речи с точки зрения носителя языка. То есть ошибка, которая допускается китайскими студентами в процессе овладения русским языком, рассматривается как противоречие правилам и нормам русского языка и воспринимается как неправильность. На основе этого определения и с учетом языковых характеристик китайских учащихся и принятой в Китае типологии ошибок, ошибки в учебном корпусе русского языка разделены на 5 основных типов и 14 подкатегорий, всего выделено 34 варианта.

В целом, созданный исследовательский учебный корпус русского языка может использоваться разными пользователями с целью решения педагогических задач (для составления учебных материалов и словарей), и в лингвистических целях (для исследования развития и изменения языка учащихся при изучении иностранного языка).

4. Использование корпусного менеджера AntConc позволило извлечь из учебного корпуса русского языка 3176 ошибок, в том числе 931 орфографическую ошибку, 1052 морфологические ошибки, 371 лексическую ошибку, 767 синтаксических ошибок, 55 коммуникационных ошибок. Относительное значение числа ошибок, рассчитываемое на основе предложенного коэффициента K , является ведущим методом оценки уровня знаний и частоты допускаемых ошибок китайскими студентами при создании текстов на русском языке.

5. Результат анализа ошибок китайских студентов при создании текстов на русском языке показал, что китайские студенты сталкиваются с большими

трудностями при овладении русским языком, особенно с теми грамматическими правилами, которые не соответствуют китайскому языку, такими как склонение имен существительных, управление глаголов, использование глаголов несовершенного или совершенного вида, использование предлогов и другие. Это связано с большими различиями между китайскими русским языками.

Анализ ошибок китайских студентов на разных уровнях овладения английским языком показал, что английский язык в общем оказывает активное влияние на обучение русскому языку китайскими студентами. Таким образом, преподавателям рекомендуется делать акцент на сопоставительном изучении русского, китайского и английского языков для того, чтобы позволить учащимся быстро овладеть соответствующим грамматическим правилам и повысить эффективность изучения русского языка.

6. Сравнение текстов китайских студентов в учебном корпусе русского языка и носителей русского языка показало, что китайские студенты допускают невидимые ошибки при употреблении глагола *быть*, предлогов *для*, *у* и *к*, что связано с недостаточным запасом слов у учащихся и непониманием точного значения слова. В этом случае учащиеся должны больше читать реальные тексты носителей русского языка для пополнения словарного запаса и четкого понимания значения слова. В то же время необходимо составлять учебные материалы, более соответствующие реальным потребностям китайских студентов.

7. Анализ типичных ошибок китайских студентов в текстах с разными темами, показал, что разные темы также оказывают различное влияние на усвоение русского языка, и китайские студенты сталкиваются с разными

трудностями. В темах, связанных с хобби и путешествиями, китайские учащиеся часто допускают ошибки в управлении глаголов. В таких темах, как проведение дня рождения и праздника Весны, учащиеся сталкиваются с трудностями при выборе вида глаголов и использовании прошедшего времени. При употреблении слов и словосочетаний с китайской спецификой в темах, связанных с праздником Весны, китайские студенты склонны буквально переводить с китайского языка на русский. Именно поэтому при обучении русскому языку необходимо оптимизировать курсы и содержание обучения по разным темам для преподавателей, и для студентов накопить хорошие примеры текстов, особенно написанные носителями русского языка для обогащения своих выражений на русском языке.

В целом, как анализ ошибок, проведенный на основе учебного корпуса русского языка для обучения китайских учащихся, так и аутентичные языковые материалы, представленные в учебном корпусе, подтверждают его методический потенциал для преподавания и изучения русского языка.

Таким образом, проведенное исследование позволило идентифицировать и проанализировать ошибки китайских студентов в процессе усвоения русского языка на основе учебного корпуса русского языка. Созданный учебный корпус русского языка в качестве базы данных, и результаты, полученные в диссертационном исследовании, доступны для применения не только китайским и русским студентам, аспирантам, преподавателям, но и всем заинтересованным лицам.

В целом учебный корпус представляет собой мощный и перспективный инструмент для теоретического и практического обучения иностранным языкам. Технологии корпусов упрощают процесс анализа больших массивов

данных и применяются в целях решения практических педагогических задач, а также подтверждают методический потенциал дальнейшего развития корпусной лингвистики в Китае.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Работы по теме диссертационной работы, опубликованные автором в рецензируемых научных журналах, входящих в перечень ВАКа:

1. Хоу Линьсюе. Классификация ошибок для создания разметки в учебном корпусе русских текстов [Текст] / Линьсюе Хоу // Известия Российского Государственного Педагогического Университета им. А. И. Герцена, 2022. – № 203. – С. 247–254. (0.5 п. л.)
2. Хоу Линьсюе. Методы разметки учебного корпуса русских текстов [Текст] / Линьсюе Хоу // Университетский научный журнал. Серия «Филологические и исторические науки, искусствоведение», 2022. – № 69. – С. 126–132. (0.44 п. л.)
3. Хоу Линьсюе. Анализ ошибок на основе учебного корпуса русских текстов [Текст] / Линьсюе Хоу // Вестник Московского городского педагогического университета. Серия «Филология. Теория языка. Языковое образование», 2023. – № 3(51). – С. 196–207. (0.75 п. л.)

Другие работы, опубликованные автором по теме диссертации:

4. Хоу Линьсюе. Принципы разметки ошибок при создании учебного корпуса русских текстов [Текст] / Линьсюе Хоу // Актуальные проблемы лингвистики и литературоведения: сборник материалов IX (XXIII) Международной научно-практической конференции молодых ученых. – Томск: Издательство Национального исследовательского Томского государственного университета,

2022. – С. 213–218. (0.38 п. л.)

5. Хоу Линьсюе. Возможности применения учебного корпуса в обучении русскому языку в эпоху больших данных [Текст] / Линьсюе Хоу // Российско-китайское сотрудничество: на пути к глобальному устойчивому развитию. Материалы VI международной научно-практической конференции. – Новосибирск: Издательство Новосибирского государственного технического университета, 2022. – С. 91–94. (0.25 п. л.)