

## ОБСТОЯТЕЛЬСТВА КОРПУСНОГО ИССЛЕДОВАНИЯ РУССКОГО ЯЗЫКА (В КРУГАХ КИТАЙСКИХ РУСИСТОВ)

Ли Дань

К.филол.н., Хэйлунцзянский университет; Цзилиньский  
университет иностранных языков,  
lid120306@126.com

### CIRCUMSTANCES OF THE CORPUS RESEARCH RUSSIAN LANGUAGE (IN THE CIRCLES OF CHINESE RUSISTS)

Li Dan

*Summary:* According to the Chinese National Knowledge Infrastructure (China National Knowledge Infrastructure, hereinafter - CNKI), the Chinese Russian organizations did not pay due attention to the study of corpus linguistics, which is still at the initial stage. at the same time, Chinese Russian scholars need to learn more from Chinese English linguists to improve the experience for corpus studies, fully clarify the role and place of corpus linguistics in the study of the Russian language, and promote the intensive development of corpus linguistics of the Russian language in China.

*Keywords:* the modern situation; corpus study; Chinese Russian; Russian language

*Аннотация:* В статье исследуется вопрос современного состояния (По Китайской национальной инфраструктуре знаний, China National Knowledge Infrastructure, далее – CNKI) корпусного исследования русского языка. В Китае долгое время не уделялось должного внимания исследованию корпусной лингвистики, поэтому данная сфера исследования языка находится там на начальном этапе. Китайским русистам необходимо перенять опыт исследований у китайских лингвистов английского языка, чтобы улучшить качество исследований, выяснить роль и место корпусной лингвистики в исследовании русского языка, а также способствовать интенсивному развитию корпусного исследования русского языка в Китае.

*Ключевые слова:* современная ситуация; корпусное исследование; китайские русисты; русский язык.

### Исследование цифровых показателей

В 60-ые годы XX века под влиянием языковой идеологии Н. Хомского большинство лингвистов перешли к применению интроспективного метод изучения языка, но Квирк, Фрэнсис, Лич и другие лингвисты, переживая сильное давление, продолжали исследования в области корпусной лингвистики. После того, как Квирк создал SEU, в 60-ые и 70-ые годы XX века появилось три больших компьютерных корпуса: BROWN, LOB, LLC, которые являются лингвистическими корпусами английского языка, содержащими миллион слов. Появление этих электронных корпусов является началом современной корпусной лингвистики.

С 90-х годов XX века технологии электронного распространения и хранения информации в компьютерах развиваются быстрыми темпами, что обусловило быстрое расширение масштаба корпусного строительства. В 90-е годы XX века BNC создали онлайн-корпус объёмом в 100 миллионов слов, а в начале XXI века частота поиска в онлайн-корпусе (WEBCORP) составила один миллиард раз - пять миллиардов раз, каждый день страница новостей обновлялась более 2 миллионов раз [Хэ Аньпин 2004: 3]. Бурное развитие лингвистического корпуса оказало большое влияние на области лингвистики и прикладной лингвистики. Для исследования на сайте "www.cnki.net" определяем объект поиска как "название изда-

ния", поисковые слова как "корпус языка", проводим поиск в области основных научных периодических изданий с 1990 по 2017 годы и получаем 1125 статей. Проводим классификацию данных статей по требованиям разных стандартов, которые наглядно отражают состояние корпусного исследования русского языка в Китае. Результаты представлены далее (схема 1, схема 2, таблица 1).

Как можно заметить, в Китае корпусная лингвистика зародилась недавно, в 1992 году, в 90-е годы XX века переживала смутное время, с 2008 года корпусная лингвистика вступила в век ускоренного развития, в 2012 году достигла вершины корпусного исследования, когда было опубликовано 107 статей на тему корпуса языка, что демонстрирует рост внимания китайских лингвистов к исследуемой проблеме. В настоящее время китайские исследователи также оказывают большое внимание корпусной лингвистике, каждый год в основных изданиях публикуется не менее 100 статей о корпусной лингвистике.

Статистические данные отражают, что корпусные исследования выявили особенности диверсификации: в содержание исследования включаются фундаментальные исследования, лингвистические исследования, прикладные исследования, политические исследования, художественная литература и тому подобное. В Китае корпусная лингвистика перешла от ознакомления с со-

ответствующими теориями к конкретным прикладным исследованиям, в дальнейшем корпусная лингвистика

может повлиять на данную тенденцию. Непрерывные поиски и развитие технологий в последние двадцать лет

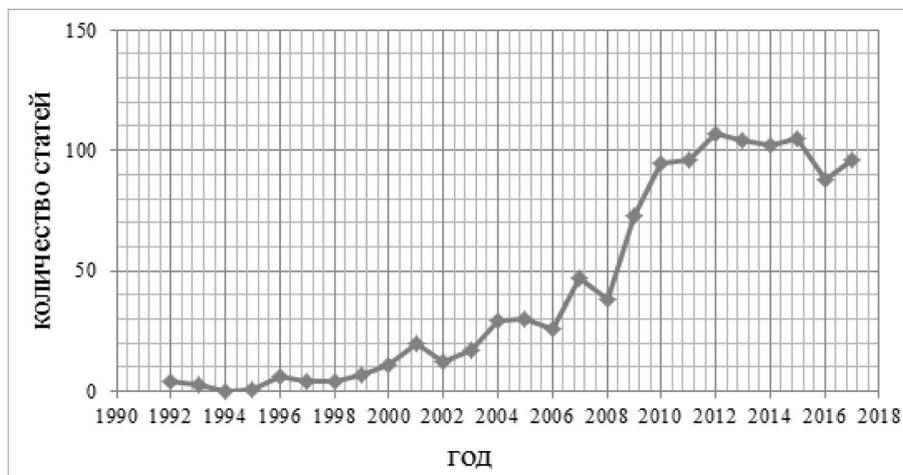


Схема 1. Статистика опубликованных работ о корпусном исследовании русского языка в Китае по годам

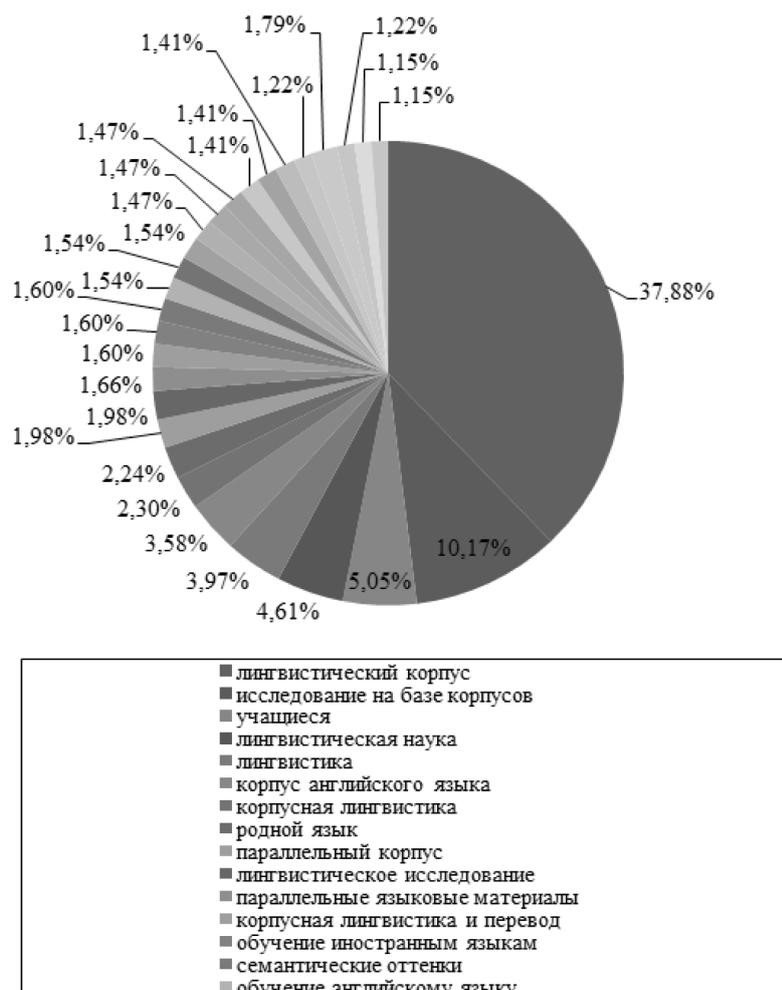


Схема 2. Распределения темы работ о корпусных исследованиях в Китае

Таблица 1.

Корпусные исследования (по языкам)

язык число	английский язык	китайский язык	русский язык	японский язык	французский язык	вьетнамский язык	итого
количество	662	441	11	9	1	1	1125
пропорция (%)	58.84	39.20	0.9	0.8	0.08	0.08	100

способствовали развитию корпусной лингвистики, ее локализации.

Корпусная лингвистика по-прежнему является актуальной в исследованиях учёных английского языка в Китае, однако все больше исследователей других языков также сместили фокус в сторону корпусной лингвистики, в том числе, исследователи китайского языка, которые непрерывно берут пример с исследователей английского языка, достигли больших успехов: корпус языка Пекинского университета и корпус языка Китайского университета СМИ. Создание и развитие корпусов языка Китая, соответствуя тенденции исследования и развитию корпусов мира, идёт таким путём: от мелкой обработки к глубокой, от одноязычных текстов к многоязычным, от мелкому масштабу к крупному, от этапа развития следования зарубежным исследователям к самоисследованию, и, наконец, корпусное исследование в Китае достигло мирового уровня [Чэнь Хун 2017:40]. По сравнению с корпусом английского и китайского языка, корпусное исследование русского языка в Китае развивается медленно. В данной статье в основном анализируется положение корпусной лингвистики русского языка и её тенденции развития.

**Особенности корпусного исследования русского языка в кругах китайских русистов**

Несмотря на то, что в 90-е годы Фу Синшан, Сюэ Энькуй и другие китайские русисты уже последовательно изучали компьютерную лингвистику и достигли несомненных успехов, однако нельзя не признать тот факт, что корпусные фундаментальные исследования и корпусные прикладные исследования для китайских русистов всё-таки являются слабым звеном.

Проведя поиск данных корпусов полных текстов на сайте "www.cnki.net", мы можем в общем описать особенности корпусных исследований китайских русистов. Войдя в сеть "www.cnki.net" мы определили классификационные каталоги литератур как "русский язык", источники изданий как "основные издания", поисковые слова как "корпус русского языка", достаточно долго производили поиск, и в итоге нашли 11 статей с 2005 по 2017 годы. Данная статистическая цифра более реально показала особенности корпусного исследования в кругах китай-

ских русистов, в настоящее время было опубликовано 11 статей о корпусных исследованиях, хотя количество работ не соответствует развитию корпусов русского языка в России.

Обобщение корпусных исследовательских аспектов китайских русистов

1. У Сюй Ханьчэна (2005), Чэнь Хун (2012) и других китайских исследователей много исследований посвящено ознакомлению с корпусами русского языка. Они не только всесторонне исследовали особенности развития корпусов русского языка в России, но и разработали в некоторой степени инструкцию использования корпусов русского языка: Сюй Ханьчэн рассматривает основные понятия корпусной лингвистики, приводит типологию корпусов, выясняет общие правила и методы классификации; Чэнь Хун приводит типы корпусной разметки русского языка: разметка исходных источников информации, разметка части речи, разметка исходной формы, синтаксическая разметка и семантическая разметка. Лингвистическая разметка - одно из основных понятий корпусной лингвистики. Разметка даёт возможность идентифицировать тексты по различным параметрам, позволяя осуществлять осмысленный поиск по корпусу. Данная статья подчёркивает то, что корпусные разметки русского языка максимально удобны для пользователей корпуса.
2. Корпусное исследование. За свои преимущества и широкое распространение направление параллельных корпусов текстов (корпусное исследование русского языка) стало одним из интенсивно развивающихся направлений современной корпусной лингвистики. Теперь уже создано несколько важных параллельных корпусов: Русско-китайский переводческий корпус, Китайско-русский параллельный корпус официально-деловых текстов, русско-китайский корпус переводов литературы на базе рассказов А.П. Чехова, Китайско-русские параллельные корпуса отдельных произведений (Китайско-русский параллельный корпус романа «Страна вина», Параллельный корпус «Русские переводы трактата «Дао Дэ Цзин» и другие), переводческий параллельный корпус. Далее кратко представим особенности двух главных параллельных корпусов. Лю Мяо и Шао Цин

(2016) создали русско-китайский корпус переводов литературы на базе рассказов А.П. Чехова, это двуязычный параллельный корпус, в основу которого ложатся классические литературные произведения на русском языке и их переводы. Китайский ученый Цуй Вэй, сотрудник «Института иностранных языков НОАК», работает с коллегами над созданием переводческого параллельного корпуса русско-китайских и китайско-русских текстов, который должен включать в себя следующие подкорпуса:

- 1) подкорпус официально-деловых текстов,
- 2) подкорпус художественной литературы,
- 3) подкорпус новостных текстов,
- 4) подкорпус текстов военной тематики,
- 5) подкорпус текстов по экономике и торговле [崔卫, 张岚 2014].

Тем не менее, можно сказать, что работа над параллельным корпусом русских и китайских текстов только началась. С ее результатами как русские, так и китайские пользователи еще мало знакомы, в силу их малой доступности и одноязычного интерфейса.

### 3. Лингвистическое исследование на базе корпусов.

В настоящее время корпусное исследование уже стало тенденцией в кругах исследователей английского языка, а в кругах изучающих русский язык корпусное исследование пока не вызвало большого внимания, за исключением русистов Сунь Миньцин и Тао Юань. Сунь Миньцин (2013) исследовала проблему синтаксической квантизации глагольной семантики в русском языке на базе национального корпуса русского языка, на примере глагола «смотреть», с помощью поисковой программы «Ant Conc 3. 2. 1» исследована частота значений, признаки сочетаемости и синтаксическая модель этого глагола; Тао Юань (2015) изучили правила перевода типичных сложных предложений с союзом «чтобы» на базе русско-китайского параллельного корпуса, куда внесены гуманитарные научные тексты. Ее статья подчёркивает три особенности сложных предложений с союзом «чтобы»: проявление-скрытие, простота-сложность, обиходные выражения - устранение обиходных выражений, а также исследуются оттенки трёх типов данных построений, в итоге автор делает вывод: в переводе сложных предложений с союзом «чтобы» скрытие сильнее проявления, простота сильнее сложности, обиходные выражения сильнее устранения обиходных выражений.

#### Перспективы корпусного исследования китайских русистов

Анализируя статистику работ по корпусной лингвистике в основных изданиях, мы заметили, что корпусные исследования медленно развиваются в Китае,

то есть корпусная лингвистика ещё не стала одним из основных направлений лингвистики. С созданием и развитием нескольких представительных корпусов русского языка появилась возможность корпусного исследования, в то же время исследователям этой области стоит больше учиться у исследователей английского языка, изучить опыт, знакомиться с теориями корпусного исследования, лингвистического исследования на базе корпуса и обучения на базе корпуса. А также открыть новую научную исследовательскую область, что обуславливает перспективы корпусного исследования русского языка.

#### Ознакомление с теориями корпусного исследования

Стоит отметить, что ознакомление с теориями корпусного исследования произошло относительно успешно, чем мы обязаны работам Сюй Ханьчэну, Ван Ужэнь и других исследователей, которые внесли вклад в представление теории корпуса, а также работам исследователей английского языка за их первоначальный вклад, создание определенного фундамента для нас, но мы должны признать, что в то же время по мере создания всё большего количества корпусов и развития различных технологий обработки текстовой информации произошли значительные изменения: выработаны некоторые универсальные принципы и языки разметки лингвистической информации в корпусе, созданы специальные инструменты для работы с корпусом на любом языке, так называемые корпусные менеджеры. Выработка универсальных стандартов и технологий сделала возможным создание больших представительных корпусов за очень короткий срок, поэтому китайские русисты продолжают представлять последние достижения русских корпусов, особенно на основе состояния корпусных исследований в России: прикладные области корпусов русского языка, создание параллельных корпусов, поисковые программы и статистические методы русских корпусов.

#### Лингвистическое исследование на базе корпуса языка

Исследовательские результаты показывают, что корпус языков играет важную роль в исследовательских областях лексики, грамматики, семантики, прагматики, анализа дискурса и других областях. Если говорить более конкретно, то на базе корпуса проводят глубокое изучение и научное описание таких языковых явлений, как лексическая сочетаемость, типы сложных предложений, особенности предложных конструкций, признаки частей речи [王臻 2007]. Лексика является одной из основных тем корпусной лингвистики, с помощью корпусных методов исследовательская область и методы традиционной лексикологии расширяются: диахронические

исследования лексики и её значений, употребление лексики, дистрибуция грамматических категорий, модель управления слова и его сочетаемостные признаки. Кроме того, мы можем изучить употребление русских пространственных предлогов на базе национального корпуса русского языка, мы должны вывести необходимые данные о русских пространственных предлогах из корпуса русского языка, использовать соответствующую программу для анализа собранных данных, обратить внимание на сочетающиеся семантические свойства предлога и семантические модификации предлогов в динамической сочетаемости с другими словами, рассмотреть значение предложно-падежной конструкции, иначе невозможно всестороннее, полностью проанализировать семантику предлогов. Конечной целью работы будет: получение убедительных научных выводов на базе корпуса языка для создания семантической сети русских пространственных предлогов. В данной статье делается заключение, что лингвистические исследования на базе корпуса постепенно становятся тенденцией в лингвистических исследованиях, которыми интересуются всё большее количество исследователей из-за их преимуществ. Корпусные исследования дают возможность проверить теоретические результаты на основе многочисленных полученных данных, что открывает новые аспекты изучения лексикологии.

#### Корпусное исследование

Учитывая все особенности развития корпусного исследования в кругах китайских русистов, для достижения больших результатов в сфере корпусной лингвистике, нам необходимо следовать вплотную за развитием передовой технологии русских корпусов. Перед нами стоят следующие задачи:

1. Многие классические произведения на русском языке, переведенные в разные годы, очень важны для исследования языковой общности, переводческого стиля переводчика, различий в языке и культуре, поэтому в Китае необходимо создать крупный русско-китайский переводческий корпус литературы.
2. За разработку русско-китайского переводческого корпуса литературы принялись недавно, более того, исследование данной темы ещё находится на начальном этапе, объём текста относительно небольшой, а стиль и языковые данные относительно однообразны, в будущем возможно увеличение количества анализируемых писателей и объёма произведений; для пополнения базы данных могут быть использованы стихи, сценарии и произведения разных жанров.
3. Положение искусственной разметки в русско-китайском корпусе литературных переводов довольно затруднительно: с одной стороны, ис-

кусственная разметка является обязательным условием для литературы и текстов других стилей, она ложится в основу классификации свойств языковых данных; с другой стороны, эффективность искусственной разметки в некоторых моментах недостаточна: не представляется возможным выполнить разметку многочисленных текстов, особенно с развитием глубокого исследования лингвистической семантики и стремительным прогрессом в области компьютерной техники. В перспективе следует осуществить полуавтоматическую разметку или автоматическую разметку компьютера, кроме того, следует уделять больше внимания разметкам грамматических свойств, поэтому необходимо совершенствовать классификацию и рамку свойственных разметок, т.е. добавить свойственные семантике разметки, тексты и тому подобное, а также ввести разные семантические словари в корпус, усилить поисковую функцию.

4. Прикладная ценность переводческого корпуса в основном проявляется в анализе языковых данных и статистике цифровых данных, с помощью корпуса мы можем исследовать переводческие тактики и методы переводчиков, переводческие нормы и лексические отношения между двумя языками, можем выбрать примеры в корпусе для составления учебных материалов, улучшить переводческое качество машины с помощью корпуса.
5. Создание корпуса русских учебных текстов. Данный корпус создан на платформе материалов внутри страны, например, тест четвёртого уровня по специальности «Русский язык» в Китае, тест восьмого уровня по этой же специальности в Китае, тест шестого уровня по русскому языку в Китае, тест для поступления в аспирантуру и тому подобное. Корпус русских учебных текстов — это коллекция текстов на русском языке, написанных студентами разных вузов. Тексты сопровождаются несколькими типами разметки (метатекстовой, морфологической разметкой и разметкой по ошибкам), что позволяет осуществлять поиск по корпусу. Мы можем провести системный анализ проблем обучения лексике, грамматике и синтаксису русского языка как иностранного, также можем исследовать привычное мышление или концептуальную систему, оформленные в изучении русского языка под влиянием родного языка, влияющие на выражения на русском языке.

#### Учебное исследование на базе корпуса

В последние двадцать лет в Китае оформилась новая тенденция: корпусное исследование, которое способствует обучению иностранным языкам, однако, колле-

ги-русисты должны продолжать исследования, чтобы добиться больших успехов в области корпусной лингвистики. Учебное исследование на базе корпуса имеет актуальное значение для помощи в обучении. В настоящее время способы активизации учебной деятельности учащихся (DDL) уже стали одной из основных научно-исследовательских тем, то есть оформленная ситуация в корпусе помогает учащимся самим изыскивать языковые нормы. Отличие этого способа от традиционных методов обучения состоит в том, что он поощряет самих учащихся самостоятельно заниматься, стимулирует собственные изыскания и открытия. Данный корпус гарантирует правильность выбранных нами примеров, а именно, русских примеров для выражения китайской мысли. Анализируя частоту слов в корпусе, мы можем получить популярные (наиболее частотные, употребительные) 2000 слов или 3000 слов, а затем составить словарь с этими популярными словами. Соединение практического обучения русскому языку и корпусного исследования русского языка — это наиболее эффективный способ изучения русского языка, возможность привлечения студентов к научно-исследовательской работе. С одной стороны, это усилит связь обучения и научно-исследовательской работы, а с

другой, достижения научно-исследовательской работы найдут применение в обучении русскому языку, что принесёт пользу учащимся, а это является настоящей конечной целью нашей научно-исследовательской работы.

### Заключение

Корпусная лингвистика объективно характеризует язык на базе корпуса. Она анализирует использование языка - речи, чтобы адекватно описать языковые конструкции, способы употребления слов и их форм. Поэтому корпусная лингвистика – это не название новой дисциплины, а новый исследовательский метод [谢元花2002]. Корпусная лингвистика достаточно долго будет основной тенденцией развития лингвистики, благодаря своим преимуществам, она окажет благотворное влияние на лингвистику в целом. Китайские русисты отстали от англоязычных лингвистов в корпусных исследованиях, поэтому необходимо, чтобы всё большее количество ученых сместили фокус в область корпусных исследований. Надеемся, что расцвет корпусных исследований русского языка в Китае не за горами.

### ЛИТЕРАТУРА

1. 陶源. 基于俄汉平行语料库的 чтобы 从句翻译操作规范研究 // 解放军外国语学院学报. 2015. №5. P. 117-125.
2. 刘淼, 邵青. 俄汉文学翻译语料库的创建——基于契诃夫小说平行语料库的设计与建构 // 外语学刊. 2016. №1. P. 154-158.
3. 谢元花. 基于语料库的词汇研究与外语教学 // 广东外语外贸大学学报. 2002. №2. P. 6-12.
4. 许汉成. 俄语语料库的新发展, 中国俄语教学. 2005. №1. P. 21-26.
5. 崔卫, 张岚. 俄汉翻译平行语料库及其应用研究 // 解放军外国语学院学报. 2014. №1. P. 84.
6. 孙敏庆. 基于俄语国家语料库的俄语动词语义句法量化研究—以视觉动词смотреть为例 // 解放军外国语学院学报. 2013. №6. P. 37-42.
7. 王臻. 俄语语料库语言学研究现状与展望 // 中国俄语教学. 2007. №2. P. 44-47.
8. 陈虹. 俄语语料库的标注 // 中国俄语教学. 2012. №2. P. 38-44.
9. 何安平. 语料库语言学与英语教学[M]. 北京: 外语教学与研究出版社, 2004:3.
10. 陈虹. 基于语料库的俄语语篇研究[M]. 北京: 经济科学出版社, 2017:40.

© Ли Дань (lid120306@126.com).

Журнал «Современная наука: актуальные проблемы теории и практики»