

В XXI веке возрастает интерес к проблемам корпусной лингвистики и актуальность корпусных методов. Вероятно, что это объясняется тем, что «корпусная методология – инструментарий, позволяющий комплексно исследовать языковую проблематику практически без изъятий; уникальная основа для решения сложных синкретичных проблем лингвистики [1, с. 11]. Следственно, в целях решения таких проблем во всех уголках мира создаются и развиваются корпуса текстов. Не стала исключением и Китайская Народная Республика.

В последние десятилетия в Китае, как и в других странах, идет активная разработка лингвистических корпусов. **Корпус текстов** – большой, представленный в электронном виде, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач [2, с. 3]. Под терминов корпус текстов также понимается вид корпуса данных, единицами которого являются тексты или их достаточно значительные фрагменты, включающие, например, какие-то полные фрагменты макроструктуры текстов данной проблемной области [3, с. 113].

Со времен создания первого корпуса текстов на китайском языке непрерывно ведется работа по их разработке не только в университетах, институтах и научно-исследовательских центрах на территории Китая, но и за пределами КНР. Следствием этой активной деятельности является создание немалого количества разнообразных корпусов текстов как на китайском языке, так и двуязычных лингвистических корпусов.

По результатам первых попыток создания корпусов было получено собрание китайских текстов для исследования частотности «Сборник текстов для изучения единиц разговорного стиля языка» (《语体文应用字汇》). Данный сборник можно назвать прообразом современных корпусов текстов: в нем собраны как самостоятельные языковые знаки, так и служебные, причем он демонстрирует, что по частотности служебные элементы языка превалируют над самостоятельными. Впоследствии на основании данного сборника был разработан ряд учебных пособий [4].

Позднее в общем доступе появляются и другие корпуса текстов китайского языка. Одним из первых корпусов был такой, как Корпус современного китайского языка (The Modern Chinese Language Corpus) Центра китайской лингвистики при Пекинском университете¹. В нем предусмотрен поиск по расстоянию между иероглифами, однако присутствует лишь метаразметка. Отсутствие морфологической и синтаксической разметки делают этот корпус текстов менее эффективным при проведении грамматических исследований.

¹ 北京大学中国语言学研究中心 [Электронный ресурс]. – Режим доступа: http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp. – Дата доступа: 25.11.2019.

В отличие от предыдущих лингвистических баз данных, наиболее полным и развитым является Корпус общего современного китайского языка¹, который содержит 2 субкорпуса: 现代汉语语料 (Корпус текстов современного китайского языка) и 古代汉语语料库 (Корпус текстов древнекитайского языка). В первом субкорпусе пользователям открываются возможности поиска по морфологическим признакам словоформ: морфологическая разметка содержит не только частеречное деление, но и признаки грамматических категорий определенного грамматического класса. Субкорпус древнего китайского языка чуть менее функционален: не содержит разметки, однако позволяет искать по словам и словосочетаниям [5].

Многие корпусы текстов создаются на территории Китайской Народной Республики на базе высших учебных заведений и крупнейших образовательных центров. Обобщая опыт и знания, создатели таких корпусов работают в этом направлении для реализации не только исследовательских, но и учебных целей. Примерами таких лингвистических корпусов служат следующие корпусы текстов:

1. Речевой корпус современного пекинского диалекта (北京地区现场即席话语语料库)²;

2. Корпус Центра исследований китайского языка Пекинского университета (北京大学中国语言研究中心语料库)³;

3. Корпус Пекинского университета языка и культуры (BCC语料库)⁴ [4, 5].

По причине высокой развитости средств массовой информации в довольно широкое распространение получили также корпусы публицистического характера, например:

1. Аннотированный корпус газеты «Жэньминь Жибао» (《人民日报》标注语料库)⁵;

2. Корпус прессы LIVAC Гонконгского педагогического института⁶ [6].

Важное место среди корпусов текстов китайского языка занимают параллельные корпусы. Такие корпусы текстов создаются и используются для сравнительных исследований языков (в области лексикологии, грамматики, стилистики, переводоведения и т.д.), для улучшения качества и эффективности перевода, а также для обучения переводу и иностранным языкам. Параллельные корпусы обеспечивают нас информацией, которую двуязычные словари обычно не содержат. Они предлагают эквиваленты не только на уровне слова, но и на уровне словосочетаний и фразеологизмов, а также переводы безэквивалентной лексики, терминов и терминологических словосочетаний.

¹ 语料库在线 [Электронный ресурс]. – Режим доступа: www.cncorpus.org. – Дата доступа: 25.11.2019.

² 北京地区现场即席话语语料库 [Электронный ресурс]. – Режим доступа: <http://ling.cass.cn/dangdai/corpus.htm>. – Дата доступа: 25.11.2019.

³ 北京大学中国语言研究中心 [Электронный ресурс]. – Режим доступа: <http://ccl.pku.edu.cn/corpus.asp>. – Дата доступа: 25.11.2019.

⁴ BCC 语料库 [Электронный ресурс]. – Режим доступа: <http://bcc.blcu.edu.cn/>. – Дата доступа: 25.11.2019.

⁵ 《人民日报》标注语料库 [Электронный ресурс]. – Режим доступа: www.icl.pku.edu.cn. – Дата доступа: 25.11.2019.

⁶ LiVac Basic Online [Электронный ресурс]. – Режим доступа: <http://www.livac.org/>. – Дата доступа: 25.11.2019.

Параллельные корпуса текстов также широко используются в процессе обучения иностранному языку. В задачах обучения переводу они могут рассматриваться как реферативная информация и предоставлять образцы профессионального перевода при изучении приемов и способов перевода [7, с. 114].

Яркими представителями таких корпусов текстов считаются:

1. Переводческий корпус русского и китайского языков (俄汉翻译语料库检索系统)¹;
2. Параллельный корпус китайского и английского языков (汉英双语语料库), Субкорпус Корпуса Центра исследований китайского языка Пекинского университета) [4].

Несмотря на широкое разнообразие и всю функциональность корпусов текстов китайского языка, на сегодняшний день существует ряд проблем, возникающих не только в процессе создания лингвистических корпусов, но и при их непосредственном использовании.

В целом принцип создания корпусов китайского языка не отличается от общепринятых в корпусной лингвистике принципов, однако сама специфика китайского языка усложняет процесс разработки корпуса текстов, накладывая различного рода ограничения. Как известно, китайский язык характеризуется низкой степенью выраженности морфологических признаков. Это влечет за собой проблемы в ходе аннотирования текстов, входящих в корпус. Поскольку один языковой знак может принадлежать сразу нескольким классам слов, то закономерно возникновение трудностей при создании морфологической разметки [5]. Сложность создания синтаксической и морфологической разметки обуславливает потребность задействования ручной обработки и больших энергозатрат.

Еще одной проблемой, возникающей в ходе создания лингвистических корпусов китайского языка, является «монолитность» китайских текстов: на письме китайские иероглифы не разделяются пробелами, а представлены единым пластом непрерывного текста. Отсутствие формального деления текстового потока с помощью пробелов безусловно усложняет процесс вычленения отдельных лексических единиц. Однако для создания разметки все же необходимо разделение потока иероглифов на отдельные значимые единицы – слова и словосочетания. В этих целях в процессе создания текста производится процедура сегментации, что является весьма трудоёмкой задачей. В сложных словах между иероглифами существуют разнообразные связи и отношения, затрудняющие сегментацию слов при создании корпуса китайского языка.

Несмотря на высокий уровень развития информационных технологий и востребованность лингвистических корпусов среди исследователей различных научных областей, специфика китайского языка обуславливает сложность создания морфологической, синтаксической и семантической разметки и влечет за собой препятствия на пути создания лингвистических корпусов китайского языка. Тем не менее, на сегодняшний день корпуса текстов китайского языка снабжены всеми необходимыми функциями, которые позволяют не

¹ 俄汉翻译语料库检索系统 [Электронный ресурс]. – Режим доступа: <http://www.rucorpus.cn/>. – Дата доступа: 25.11.2019.

только провести достоверный анализ употребления тех или иных единиц языка, но и отобрать ценный эмпирический материал, служащий базой для решения трудоёмких лингвистических задач.

Таким образом, доступность, высокая функциональность и возможность быстрого поиска позволила лингвистическим корпусам стать неотъемлемой частью работы как исследователей-лингвистов, так и специалистов других областей. Лингвистические корпусы китайского языка обладают широкими возможностями лексико-грамматического поиска, внушительным и разнообразным по содержанию объемом данных, что делает каждый из них уникальной лингвистической базой данных, которая не только позволяет решать задачи, возникающие в ходе лингвистических исследований, но также оказывает существенную помощь переводчикам и способствует улучшению качества и эффективности преподавания китайского языка как иностранного.

ЛИТЕРАТУРА

1. *Баркович, А. А.* Корпусная лингвистика: специфика современных метаописаний языка / А. А. Баркович // Вестник Томского государственного университета. – 2016. – № 406. – Томск : ООО «Издательство ТГУ». – С. 5–13.
2. *Захаров, В. П.* Корпусная лингвистика: учебно-метод. пособие / В. П. Захаров. – СПб. : СПбГУ, 2005. – 48 с.
3. *Баранов, А. Н.* Введение в прикладную лингвистику : Учебное пособие / А. Н. Баранов. – М.: Эдиториал УРСС, 2001. – 360 с.
4. *Баркович, А. А.* Лингвистические корпусы китайского языка : функциональный аспект / А. А. Баркович, Ван Цин // Вестник МГЛУ. – Минск, 2015. – № 5. – С. 105–113. – (Серия 1, Филология).
5. *Колпачкова, Е. Н.* Корпусы китайского языка : современное состояние и основные проблемы [Электронный ресурс] / Е. Н. Колпачкова. – Режим доступа: <http://orient.spbu.ru>. – Дата доступа: 20.11.2019.
6. *Лу, Исинь.* Принципы создания корпусов китайского языка [Электронный ресурс] / Лу Исинь. – Режим доступа: <http://cyberleninka.ru>. – Дата доступа: 25.11.2019.
7. *Ушакова, Н. В.* Применение корпусов текстов в обучении иностранному языку / Н. В. Ушакова, Е. П. Курчева // Актуальные проблемы современного образования : материалы VII Международной научно-методической конференции, Воронеж, 20 апреля 2015 г. – Воронеж : ВИЭСУ, 2015. – С. 111–114.