

Гибкий Павел Валерьевич

Минский государственный лингвистический университет

e-mail: pavel.gibkiy@bk.ru

**ИСПОЛЬЗОВАНИЕ КОРПУСОВ ТЕКСТОВ
ПРИ ИССЛЕДОВАНИИ ЭФФЕКТА ВРЕМЕННЫХ ОТНОШЕНИЙ
В КИТАЙСКОМ ЯЗЫКЕ**

Аннотация. В статье рассматриваются возможности использования корпусов текстов для поиска языковых единиц, являющихся синтаксическими средствами реализации временных отношений в китайском языке (полусуффиксов 了(le) и 过(guò).

Ключевые слова: время, корпусы текстов, частотность, временные отношения, синтаксис, полуаффиксы.

Pavel Gibkij

Minsk State Linguistic University

e-mail: pavel.gibkiy@bk.ru

**THE USE OF TEXT CORPORA IN THE STUDY
OF THE TEMPORARY RELATIONS EFFECT
IN THE CHINESE LANGUAGE**

Abstract. The article discusses the possibilities of using text corpora to find language units, that are syntax implementations of temporary relations in Chinese (semi-suffixes 了(le) and 过(guò).

Keywords: time, text corpora, temporal relations, syntax, semisuffixes.

Актуальность исследования заключается в двух аспектах: востребованностью и высокой эффективностью использования корпусов текстов для лингвистических исследований, недостаточной исследованностью вопросов китайской грамматики, и, в частности, функционирования временных отношений.

По словам А. А. Барковича, «современные лингвоинформационные технологии открывают широкие перспективы для исследований с использованием корпусных методик и на материале корпусных данных» [1, с. 105].

Востребованность и эффективность корпусов текстов обусловлена достаточно высокой точностью «метаязыковой обработки эмпирических данных и лингвистической релевантности доступных в данном контексте обобщений» [1, с. 105]. Первые корпусы текстов китайского языка относятся к периоду 1920-х годов» [11, с. 160].

Разработка современных корпусов китайского языка «привела в октябре 1990 г. к разработке и опубликованию госстандарта GB-1371 – «Стандарт сегментации

слов современного китайского языка при обработке информации» [14, с. 132]. В настоящее время китайскоязычные корпуса текстов в большинстве своем используют те же принципы обработки текста, что и европейские аналоги, «основные принципы организации корпуса китайского языка в целом не расходятся с принятыми в корпусной лингвистике» [5, с. 7], но в силу особенностей грамматического строя и лексических особенностей китайского языка возникают некоторые различия.

Корпусы текстов китайского языка в большинстве своем используют те же принципы обработки и лемматизации текста, что и европейские аналоги, но в силу особенностей грамматического строя и лексических особенностей китайского языка возникают некоторые различия. Во-первых, в ряде корпусов китайского языка, в том числе и в корпусе современного китайского языка «...имеется только разметка, отсутствуют... морфологическая и синтаксическая разметки ...» [5, с. 2], что «... делает использование данного корпуса в области грамматических исследований... затруднительным» [5, с. 2].

Так как в данной работе исследуются возможности использования корпусов текстов для поиска языковых единиц, являющихся синтаксическими средствами реализации временных отношений в китайском языке, целесообразно рассмотреть средства реализации значения времени в китайском языке. Во флективных языках категория времени считается системой «противопоставленных друг другу рядов грамматических форм...» [6]. В китайском языке «отсутствует морфология, если под морфологией понимать словоизменение» [4, с. 9], по причине чего нет и не может быть «... классификаций категорий глаголов» [3, с. 67], «склонения существительных ... и т. п.» [8, с. 37]. Именно поэтому значение времени может выражаться только на уровне синтаксиса с помощью полусуффиксов 了(le) и 过(guò). Китаисты по-разному определяют вышеупомянутые форманты, содержащие «дополнительную информацию о характере процесса с точки зрения его завершенности...» [4, с. 39]. Мы рассматриваем случаи комбинаций 了(le) и 过(guò), посредством которых создается эффект временных отношений. К примеру, в предложении 最后一箱弹药在半路上不见了, ... 然后来了几个雪怪 (‘Последний ящик с боеприпасами пропал по дороге, ... там на него напали вендиго’) [9] первое действие (‘пропал’) предшествует второму (‘напали’), занимает позицию раньше [13, с. 457]: эффект временных отношений создан комбинацией полусуффиксов 了(le). Во фрагменте 很多...以前小时候吃过做过的 [11] (‘в детстве очень много чего было съедено и сделано’) первое действие (‘было съедено’) находится в неуточненной временной локализации по отношению ко второму (‘было сделано’), неизвестно, занимает оно позицию Раньше или позицию Позже [13, с. 457]: эффект временных отношений создан комбинацией полусуффиксов 过(guò).

В таблице 1 представлен анализ характеристик, преимуществ и недостатков национального корпуса русского языка [8] и лингвистического корпуса китайского языка Пекинского университета языка и культуры [12] (см. табл. 1).

Таблица 1. – Обзор лингвистического корпуса китайского языка Пекинского университета языка и культуры и русско-китайского параллельного корпуса НКРЯ

| Название | Характеристики | Преимущества и недостатки |
|--|--|---|
| Лингвистический корпус китайского языка Пекинского университета языка и культуры (ВСС) | Общее количество материала корпусов ВСС составляет около 15 миллиардов слов, включая газеты (2 миллиарда), литературу (3 миллиарда), Weibo (3 миллиарда), науку и технику (3 миллиарда), комплексный (1 миллиард) и древний китайский (2 миллиарда) [2, с. 133] | Что касается преимуществ, во-первых, ВСС характеризуется большим объемом и включением различных языковых регистров. Во-вторых, «языковые материалы ... прошли процесс сегментации и разметки по характеру и синтаксису» [2, с. 134]. Что касается недостатков, в первую очередь, это отсутствие параллельного русско-китайского полкорпуса, перевода на русский язык и пиньиня. Более того, отсутствует частеречная разметка, что затрудняет обработку больших массивов текстов |
| Русско-китайский параллельный корпус НКРЯ | Объем корпуса составляет «...более 3,5 миллионов слов» и около «... 55 тысяч словоупотреблений» [13, с. 173], присутствует семантическая (словарные толкования); грамматическая и фонетическая (пиньинь) разметка, поиск языковых единиц можно осуществлять по разным параметрам | Что касается преимуществ, во-первых, присутствуют перевод с китайского на русский, что облегчает восприятие текстов корпуса русскоязычными людьми и сокращает время, затрачиваемое на обработку информации. Во-вторых, присутствуют ссылки на фамилии переводчиков. В-третьих, разработана удобная поисковая система. В-пятых, корпус «размечен грамматически» [10]. Что касается недостатков, в первую очередь, это малый объем корпуса. Кроме того, в корпусе представлены «тексты с неснятой омонимией, ... в них не разрешена неоднозначность» [9] слов, соответствующих одному и тому же иероглифу, но произносящихся по-разному (说, shuì/shuō) |

В таблице 2 представлен рейтинг частотности иероглифов, являющихся синтаксическими средствами реализации временных отношений в китайском языке, составленный на материале Лингвистического корпуса китайского языка Пекинского университета языка и культуры (ВСС) [12] и русско-китайского параллельного корпуса Национального корпуса русского языка (НКРЯ) [8].

Таблица 2. – Рейтинг частотности полусуффиксов 过 (guò), 了 (le) и их комбинаций в лингвистическом корпусе китайского языка Пекинского университета языка и культуры (ВСС) и в русско-китайском параллельном корпусе национального корпуса русского языка

| Иероглиф | Основное значение иероглифа | Основная грамматическая реализация | Абсолютная частотность иероглифа в ВСС | Абсолютная частотность иероглифа в русско-китайском параллельном корпусе |
|------------|---|------------------------------------|--|--|
| 1. 过 (guò) | «содержит дополнительную информацию о характере процесса с точки зрения его завершенности ...» [8, с. 39] | полусуффикс | 2831532 документа и вхождения [11] | 30 документов, 3 465 вхождений [9] |
| 2. 了 (le) | «содержит дополнительную информацию о характере процесса с точки зрения его завершенности...» [6, с. 39]. | полусуффикс | 17935880 документов и вхождений [11] | 30 документов, 29 509 вхождений [9] |
| 3. 过 过 | создание эффекта временных отношений | комбинация двух полусуффиксов | 3164 документа и вхождения [11] | 3 документа, 5 вхождений [9] |
| 4. 了 了 | создание эффекта временных отношений | комбинация двух полусуффиксов | 10544 документов и вхождений [11] | 1 документ, 1 вхождение [9] |

Таким образом, в данной работе была описана история развития корпусов текстов китайского языка, охарактеризованы их преимущества и недостатки, особенности, проанализированы основные возможности использования корпусов текстов для поиска языковых единиц китайского языка, охарактеризован эффект временных отношений, создаваемый полусуффиксами 了 (le) и 过 (guò), выявлена частотность употребления иероглифов, обозначающих синтаксические средства реализации временных отношений в китайском языке.

СПИСОК ЛИТЕРАТУРЫ

1. Баркович, А. А. Лингвистические корпуса китайского языка: функциональный аспект / А. А. Баркович, Ван Цин // Вестник МГЛУ. – 2015. – № 5 (78). – С. 105–113.
2. Ванифань, Цай. Развитие исследования корпуса в Китае – от частотного словаря к лингвистическому корпусу / Цай Ванифань // Вестник Московского университета. – 2019. – № 2. –

- Режим доступа: <https://cyberleninka.ru/article/n/razvitie-issledovaniya-korpusa-v-kitae-ot-chastotnogo-slovarya-k-lingvisticheskomu-korpusu>. – Дата доступа: 13.12.2021.
3. Гибкий, П. В. Сохранение эквивалентности при передаче значения пассивного залога [Электронный ресурс] / П. В. Гибкий, Н. В. Супрунчук // Вестник Омского государственного педагогического университета. Гуманитарные исследования. – 2021. – № 4 (33). – С. 66–69. – Режим доступа: <https://cyberleninka.ru/article/n/sohranenie-ekvivalentnosti-pri-peredache-znacheniya-kategorii-passivnogo-zaloga-s-russkogo-yazyka-na-kitayskiy-na-materiale-saytov>. – Дата доступа: 23.01.2022.
 4. Гордей, А. Н. Принципы исчисления семантики предметных областей / А. Н. Гордей. – Минск : Белгосуниверситет, 1998. – 156 с.
 5. Колпачкова, Е. Н. Корпусы китайского языка: современное состояние и основные проблемы / Е. Н. Колпачкова // Труды международной конференции «Корпусная лингвистика. – 2015». – СПб.: Издательство Санкт-Петербургского университета. – 2015. – С. 278–286.
 6. Ли, Ван. Части речи / Ван Ли // Новое в зарубежной лингвистике. – М. : Прогресс. – 1989. – № 22. – С. 37–38.
 7. Лингвистический энциклопедический словарь. Грамматическая категория [Электронный ресурс]. – Режим доступа: <http://tapemark.narod.ru/les/115a.html>. – Дата доступа: 01.10.2021.
 8. Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: <http://ruscorpora.ru/>. – Дата доступа: 28.04.2018.
 9. Русско-китайский параллельный корпус НКРЯ [Электронный ресурс]. – Режим доступа: <http://ruzhcorp.ruscorpora.ru/>. – Дата доступа: 19.12. 2021.
 10. Сяохуэй, Чэнь. О параллельных корпусах русских и китайских текстов / Чэнь Сяохуэй, О. В. Кукушкина // Вестник Московского университета. – 2018. – № 2. – Режим доступа: <https://cyberleninka.ru/article/n/o-parallelnyh-korpusah-russkih-i-kitayskih-tekstov>. – Дата доступа: 17.02.2022.
 11. Юэ, Фэн. Специфика корпусных исследований в современном китайском языкознании [Электронный ресурс] / Фэн Юэ // Вестник Московского государственного лингвистического университета. Гуманитарные науки. – 2020. – № 3 (832). – Режим доступа: <https://cyberleninka.ru/article/n/spetsifika-korpusnyh-issledovaniy-v-sovremennom-kitayskom-yazykoznanii>. – Дата доступа: 05.12.2021.
 12. ВСС 语料库 [电子资源]. – 文档存取方式: <http://bcc.blcu.edu.cn/>. – 文档存取日: 18.01.2022 = ВСС. Лингвистический корпус [Электронный ресурс]. – Режим доступа: <http://bcc.blcu.edu.cn/>. – Дата доступа: 18.01. 2022.
 13. McTaggart, J. E. The Unreality of Time / J. E. McTaggart // Mind: A Quarterly Review of Psychology and Philosophy. – 1908. – P. 456–458.
 14. 冯志伟. 中国语料库研究的历史与现状 // Journal of Chinese Language and Computing. – 2002. – № 2 (11). – С. 127–136 = Фэн Чживэй. История и состояние исследования корпусов в Китае // Журнал китайского языка и вычислительной техники. – 2002. – № 2 (11). – С. 127–136.