DOI: 10.48647/ICCA.2026.41.73.031

В.В. Башкеев
ИКСА РАН

Опыт применения современных технических средств для анализа всего корпуса текстов «Хань шу»

Аннотация. В представленных тезисах обобщается первый опыт сквозного статистического анализа полного корпуса «Хань шу» (漢書) как единой, но внутренне неоднородной конструкции, включающей разделы «Ди цзи» (帝紀), «Бяо» (表), «Чжи» (志) и «Лечжуань» (列傳). Показано, как современные технические средства радикально снижают временные издержки рутинно-подготовительной работы и позволяют быстро перейти к аналитической части исследования общего корпуса текстов в рамках комплексного исторического памятника от его структурного скелета (раздела «Ди цзи» 帝紀) и сжатого изложения паспортных данных в таблицах «Бяо» (表) до обобщающих сочинений в форме трактатов Чжи» (志) и нарративного массива в каталоге преданий «Лечжуань» (列傳). Предложенный подход дает воспроизводимую модель для дальнейших историко-филологических исследований широкого круга письменных памятников имперского Китая.

Ключевые слова: Хань шу; статистический анализ; структурно-жанровый анализ.

Victor V. Bashkeev
ICCA RAS

Experience in using modern technical methods to analyze the entire corpus of the «Han Shu»

Abstract. The following text summarize a first experience of an end-to-end analysis of the complete Hanshu (漢書) corpus as a single but internally heterogeneous construct, including the sections Di ji (帝紀), Biao (表), Zhi (志), and Liezhuan (列傳). It is shown how modern technical tools radically reduce the time costs of routine preparatory work and make it possible to move quickly to the analytical stage of studying the corpus of texts as a whole within a complex historical monument: from its structural backbone (the Di ji 帝紀 section) and the compressed presentation of record data in the Biao (表) tables to generalized writings in the form of Zhi (志) treatises and the narrative mass of biographies in the Liezhuan (列傳). The proposed approach provides a reproducible model for further historical-philological research on a broad range of written monuments of imperial China.

Keywords: Han Shu; statistical analysis; structural-genre analysis.

Бурное развитие в последние 1,5—2 года потребительских версий больших языковых моделей (БЯМ) и взрывной рост качества их работы в конце 2025 — начале 2026 гг. в условиях глобальной конкуренции американских и китайских IT конгломератов создали доступные массовому потребителю исследовательские инструменты, важность которых уже невозможно игнорировать. Подчеркнем сразу, автор далек от идеи написания текстов научных работ при помощи так называемых нейросетей и предостерегает аудиторию, в особенности научную, от попадания в маркетинговую ловушку «необходимости внедрения искусственного интеллекта», ясно осознавая исходную манипулятивность и ложность этого понятия. Однако технические возможности, открывшие доступ к программированию широкого круга задач на естественном языке минуя специальные программные среды посредством пользовательского интерфейса БЯМ, столь сильно меняют рабочий процесс, что на это невозможно не обратить внимание.

Так в рамках научной работы в период с января по февраль—март 2026 года средствами различных БЯМ нами были осуществлены:

1. Полный технический перевод[1] на русский язык всего объема текста Хань шу (разделы «Таблицы», «Трактаты», «Предания»). Общий объем полученного русского текста составил порядка 160 а.л., точное число переведенных иероглифов: 736 338

2. Написание средствами БЯМ и Python специального ПО для создания, индексации и анализа всего китайского и русского корпуса с функциями полнотекстового поиска и индексирования по различным параметрам (формирование баз данных персоналий, топонимов и структурных элементов текста) и выдачи базовой статистики по числу иероглифов в каждом разделе.

3. Точные статистические подсчеты объема иероглифов по разделам: 54 290 в Дицзи (~7% от общего объема), 75 312 (9,5% от общего объема) в Бяо, 147149 (~18,5% от общего объема) в Чжи и 513 877 (65% от общего объема) в Лечжуань. Отметим, что теперь в точности известен объем всего памятника в иероглифах: 790 628, что впредь позволяет избегать распространенной обтекаемой формулировки «около 800 тыс».

4. Также было подсчитано точное количество и иероглифический объем императорских эдиктов, явно введенных формулой *чжаоюэ* 詔曰: 277 эдиктов, общим объемом 21 688. Из них в Дицзи — 176 (отсутствуют только во 2 главе, общий объем 14627 иероглифов), в трактатах — 19 (встречаются в половине глав раздела, чаще всего — в главе 23, общий объем 961 иероглиф) и в Лечжуань — 82 (встречаются в 28 из 70 глав,

[1] Данный тренд представлен и в англоязычной среде: https://ctext.org/instructions/translation

общий объем 6080 иероглифов, наибольшие объемы в жизнеописаниях Чжай Фанцзиня и Ван Мана — 636 и 1590 иероглифов соответственно) .

Дальнейшая разработка приложения позволит добавить более точную аналитическую настройку для структурно-жанрового и контент-анализа, однако уже доступный базовый функционал является важным подспорьем для работы над публикацией 3—10 томов русского перевода и дальнейших исследований империй Западная Хань и Синь.

Библиографический список

Бань Гу. Хань шу («История Хань») : в 8 т. / пер. с китайского В.В. Башкеева под ред. М.Ю. Ульянова; коммент., вступит. ст. и приложения В.В. Башкеева, М.Ю. Ульянова. М.: ИДВ РАН; Вост. лит., 2021 (Памятники письменности Востока. CLVI, 1). ISBN 978-5-8381-0381-9.

漢書（維基文庫）[Hanshu, Wikisource]. URL: https://zh.wikisource.org/wiki/漢書 (accessed: 04.04.2026).

漢書（Chinese Text Project）[Hanshu, CText]. URL: https://ctext.org/han-shu (accessed: 04.04.2026).