

ПЕРСПЕКТИВЫ КОРПУСНЫХ ИССЛЕДОВАНИЙ АНТИЧНЫХ ТЕКСТОВ

Аннотация. В статье рассматриваются перспективы применения современных компьютерных средств работы с текстами значительного объема, известных под общим названием «корпусная лингвистика», к античным текстам. На примере сочинений, подписанных именем Иоанна Богослова, автор рассматривает основные этапы корпусного анализа. Полученный материал может служить основанием для критического анализа текста, изучения системы образов автора и, самое главное, транслатологических исследований.

Ключевые слова: корпусная лингвистика; перевод; Библия; ключевые слова; частотность.

Сведения об авторе: Шитиков Петр Михайлович, кандидат филологических наук, кандидат богословия, доцент кафедры филологического образования Тюменского государственного университета (филиал в г. Тобольске).

Контактная информация: 626152, Тюменская область, г. Тобольск, ул. Аптекарская, д. 4, тел.: 8(912)991-17-91; e-mail: petrkifa@mail.ru.

Введение

Корпусная лингвистика как самостоятельный раздел прикладной лингвистики в настоящее время является самым перспективным направлением развития языкознания. Электронные сборники (корпусы) текстов получили огромное значение как в лингвистических исследованиях, так и в переводоведении, начиная с 90-х гг. XX столетия, когда компьютерные технологии стали общедоступны. Отечественный исследователь данной тематики В.В. Рыков предлагает определять корпус текстов как «некоторое собрание текстов, в основе которого лежит логический замысел, логическая идея, объединяющая эти тексты и воплощенная в правилах организации текстов в корпус, алгоритме и программе анализа корпуса текстов, сопряженной с этим идеологии и методологии» (Рыков 2002). Применение компьютерных технологий открывает перед исследователем новые возможности. Объем текстов, анализируемых с помощью специально разработанных программ, позволяет получать статистические данные о языке, недоступные прежде.

В данной статье рассматриваются основные перспективы применения корпусных методов к античным сакральным текстам на материале небольшого корпуса сочинений Иоанна Богослова.

Методология

Корпусные исследования оперируют значительными объемами текстов в автоматическом режиме. Для получения проверяемого языкового материала были разработаны компьютерные программы, позволяющие максимально быстро обрабатывать огромные объемы текста в соответствии с заданными параметрами. Основные функции, которые призваны выполнять данные программы:

- формирование списка частотности лексических единиц;
- формирование списка ключевых слов;
- создание конкорданса для ключевых слов.

Наиболее популярной в академической среде является программа Oxford WordSmith Tools, которая включает три модуля – *WordList*, *KeyWords*, *Concord*, которые позволяют последовательно выполнить указанные функции.

Посредством модуля *WordList* возможно формирование словарных списков по показателю частотности. Программа позволяет создавать списки лексических единиц с

указанием частотности их употребления в тексте либо наборе текстов. Это необходимый шаг для создания базы ключевых слов.

Модуль *KeyWords* позволяет сформировать список ключевых слов, а также ключи к ключевым словам. Термин «ключевое слово», по определению М. Скотта, означает лексическую единицу, которая появляется в анализируемом тексте с необычной частотой, в сопоставлении со справочным корпусом (Scott 1997: 236). Р. Вильямс использует данный термин для обозначения слов, ключевых для определенной культуры (Williams 1983). На основании ряда экспериментов Дж. Андор пришел к выводу, что ассоциации, порожденные ключевыми словами, в значительной степени определяют тональность текста и подготавливают читателя к его восприятию (Andor 1989: 28–36). Посредством ассоциаций выстраивается связь между ключевым словом и другими лексическими единицами. Фактически устанавливается связь между текстом и мышлением, текстом и культурой.

Модуль *Concord* применяется для изучения ключевых слов в контексте их употребления. Просмотр лексических единиц в линии конкорданса способствует идентификации контекстуального значения, что максимально объективизирует выводы исследователя. Возможность построения линий конкорданса для ключевых слов по алфавиту позволяет удобно определить лексические модели.

Анализ античных библейских текстов в нашем исследовании проводился средствами программы *BibleWorks*, которая включает подлинные тексты, большинство переводов на национальные языки, набор лексиконов, справочников и лингвистических комментариев. Модули программы позволяют провести основные этапы корпусного анализа текста.

Анализ текстов

В качестве основного материала исследования в данной работе выбран так называемый корпус Иоанна – пять библейских книг, подписанных именем Иоанна: Евангелие, три Соборных послания и книга Откровения (Апокалипсис). В качестве материала использован греческий текст Нового Завета в издании *Nestle-Aland*. Данные сочинения отличаются по жанру, времени написания, адресату и кругу рассматриваемых вопросов, в связи с чем неоднократно поднимался вопрос об их авторстве. Рассмотрим кратко основные этапы анализа текста.

На **первом этапе** был сгенерирован список слов по частотности употребления для греческого текста Евангелия от Иоанна. Поскольку служебные части речи, не представляющие интереса для данного исследования, являются наиболее частотными, нами был выставлен фильтр идентификации существительных и глаголов. В таблице 1 представлены наиболее часто употребляемые лексические единицы разбираемого текста.

Таблица 1

Наиболее частотные глаголы и существительные

Номинативные ЛЕ	Ин(ф)	Глагольные ЛЕ	Ин(ф)
ἰησοῦς – Иисус	244	λεῶμι – говорить	480
πατήρ – отец	136	εἶμι – быть	445
θεός – Бог	83	ερχομαι – идти	157
κόσμος – мир	78	ποιεῶ – делать	110
μαθητής – ученик	78	πιστεῦω – верить	98
ἄνθρωπος – человек	59	εχω – иметь	87
υἱός – сын	55	οἶδα – знать	84
κύριος – господин	52	ἀποκρίνομαι – отвечать	78
λόγος – слово	40	δίδωμι – давать	75
ζωή – жизнь	36	οραω – видеть	67

На **втором этапе** определяется список ключевых слов. Проверяемый текст меньшего объема сопоставляется со справочным корпусом, состоящим из текстов, близких по жанру и тематике, для выявления лексических единиц, появляющихся в нем с неожиданной частотностью. Список частотности Евангелия от Иоанна сопоставляется с данными

справочного корпуса, в качестве которого выбраны так называемые синоптические евангелия, сопоставление с ними позволяет выделить характерные именно для Иоанна лексические единицы, которые маркируются как ключевые. Так, существительное πατήρ (отец) встречается в проверяемом корпусе 136 раз, в то время как в справочном корпусе оно встречается 137 раз в совокупности. Лексическая единица ἰησοῦς (Иисус) занимает первое место по показателю частотности у всех евангелистов, однако Иоанн употребляет ее 244 раза, что сопоставимо с показателем всех трех Евангелий (322 раза). Лексическая единица κόσμος (мир) встречается у Иоанна 78 раз, а у синоптиков – всего 15 раз. Для глаголов иллюстративными являются следующие примеры: πιστεύω (верить) у Иоанна – 98 раз, в справочном корпусе – 34, οἶδα (знать) – у Иоанна – 84, у синоптиков – 74, γινώσκω (знать) – у Иоанна – 57, у синоптиков – 60. Очевидно, что данные лексические единицы могут быть помечены как ключевые слова для Евангелия от Иоанна, поскольку отражают важнейшие для автора концептуальные понятия. В таблице 2 приводятся данные о первых десяти глагольных и номинативных ЛЕ, которые могут быть маркированы как ключевые для проверяемого текста. В первом столбце представлена лексическая единица, во втором – частотность употребления у Иоанна, в третьем – частотность употребления у синоптиков.

Таблица 2

Наиболее частотные ключевые слова

ЛЕ (номинативные)	Ин(ф)	Син (ф)	ЛЕ (глагольные)	Ин (ф)	Син (ф)
πατήρ – отец	136	137	πιστεύω – верить	98	34
κόσμος – мир	78	15	οἶδα – знать	84	70
ζωή – жизнь	36	16	μείνω – оставаться	40	12
ἔργον – дело	27	10	ἀγαπάω – любить	37	26
ἀλήθεια – истина	25	7	μαρτυρεῶ – свидетельствовать	33	2
φῶς – свет	23	15	πέμπω – посылать	32	15
ὕδωρ – вода	21	18	ἀποθνήσκω – убивать	28	23
προβάτων – овца	19	15	ἐρωτάω – любить	28	22
ἑορτή – праздник	17	7	θεωρῶ – смотреть	24	16
μαρτυρία – свидетельство	14	4	δοξάζω – прославлять	23	14

Важно отметить, что существуют слова-исключения, которые употребляются автором Евангелия от Иоанна и не встречаются у синоптиков, однако не могут быть маркированы как ключевые, поскольку не отражают концептуальной картины мира автора. Главным образом, это имена собственные (Ναθαναήλ, Νικόδημος), географические наименования (Καπὰ, Τιβερίας) и ЛЕ, употребленные единожды. Всего выявлено 62 номинативных и 25 глагольных слов-исключений.

При сопоставлении Евангелия с другими сочинениями корпуса Иоанна (1, 2, 3 соборные послания и апокалипсис) оказалось, что число слов-исключений весьма значительно: 234 номинативных и 179 глагольных лексических единиц в Евангелии от Иоанна не встречаются в остальном корпусе. Соответственно, в сочинениях, подписанных именем Иоанна, встречается 310 номинативных и 117 глагольных ЛЕ, которые не употребляются в его Евангелии. В таблице 3 представлены наиболее показательные примеры слов-исключений для Евангелия в сравнении с соборными посланиями и книгой Откровения.

Таблица 3

Наиболее частотные слова-исключения

Номинативные ЛЕ	Ин(ф)	Глагольные ЛЕ	Ин(ф)
μαθητής – ученик	78	πορεύομαι – идти	16
ἄρτος – хлеб	24	παράδίδωμι – передавать	15
ἑορτή – праздник	17	ἀγώ – вести	13
μνήμειον – память	16	βαπτίζω – крестить	13
νόμος – закон	15	νίπτω – умывать	13

γραφή – <i>писание</i>	12	ανίστημι – <i>поднимать/воскрешать</i>	8
ῥημα – <i>слово</i>	12	ασθενεω – <i>болеть</i>	8
ιερον – <i>храм</i>	11	δοκεω – <i>казаться</i>	8
πασχα – <i>пасха</i>	10	κραυγαζω – <i>кричать</i>	6
υπηρετης – <i>слуга</i>	9	σωζω – <i>спасать</i>	6

При относительно небольшом лексическом составе анализируемых текстов значительное расхождение в выборе лексических средств может свидетельствовать в пользу теории о разном авторстве книг избранного корпуса. В любом случае, это не позволяет рассматривать корпус Иоанна как единое целое при анализе лексических единиц.

Перспективы исследования

Применение корпусных методов анализа к библейским текстам открывает широкие перспективы для лингвистики античных текстов. Материал, полученный по заданному алгоритму, обладает максимальной объективностью. Доступ к нему осуществляется буквально в один «клик», что значительно сокращает трудоемкость исследования. Данное направление получает широкое распространение. Существует ряд онлайн-проектов, в которых реализуются основные этапы корпусного анализа. Один из пионеров данного направления М.Б. О’Доннел выделил три наиболее перспективные области применения корпусных исследований к Новому Завету (O’Donnell 2005). Во-первых, это критика текста: корпус помогает принять решения относительно лучшего прочтения на основании статистических данных. Во-вторых, это лексикография, поскольку анализ ближайших словосочетаний с помощью корпусных техник позволяет глубже понять связь синонимов. В-третьих, корпусные технологии могут успешно применяться при анализе дискурса. Автор показал, что корпус помогает понять цели и задачи автора.

Применение корпуса открывает новые перспективы перед исследователем Библии и показывает, как много можно сделать в этой области. Переводоведение, семасиология, лексикография, критика текста, стилистика – это лишь немногие направления, которые получают из корпусных штудий фактический языковой материал, что необходимо для верификации теоретических предположений авторов. Например, данные, представленные в настоящей статье, послужили подтверждением для гипотезы о связи лексических единиц с концептуальной основой метафорических высказываний. Нами было показано, что ключевые слова, характерные для автора Евангелия от Иоанна, являются частью системы его метафорического осмысления мира – как физического, так и трансцендентного (Шитиков 2017).

Выводы

Проведенный анализ библейских текстов с помощью инструментов корпусной лингвистики предоставил важнейшие статистические данные об употреблении отдельных лексических единиц в изоляции и контексте. Сопоставление списков частотности Евангелия Иоанна и синоптических евангелий с помощью модуля *KeyWords* позволило выделить наиболее характерные лексические единицы. Уверенно можно сделать вывод об оригинальности языка автора Евангелия. При сравнении ключевых слов в рамках корпуса Иоанна были выявлены значительные вариации в выборе лексических средств. Это позволило оценить степень лексической идентичности различных произведений одного автора.

В целом, следует высоко оценить перспективы применения компьютерных технологий в процессе исследования античных текстов. Несмотря на высокую степень изученности материала, в частности, Евангелия от Иоанна, новые методики позволяют открыть новые грани хорошо знакомых текстов, объективно подтверждая не всегда очевидные закономерности.

Литература

Рыков В. В. 2002. Корпус текстов как реализация объектно-ориентированной парадигмы // Труды международного семинара «Диалог-2002». М.: Наука.

Шитиков П. М. 2017. Опыт комплексного анализа концептуальной метафоры: от оригинала к переводу // Вопросы когнитивной лингвистики 3 (52), 57–62.

Andor J. 1989. Strategies, tactics and realistic methods of text analysis // Heydrich W., Neubauer F., Petöfi J., Sözer E. (eds.). Connexity and coherence: Analysis of text and discourse. Berlin: WalterdeGruyter, 28–36.

Nestle-Aland Novum Testamentum Graece. 2006. Hendrickson Publishers Marketing, LLC.

O'Donnell M. B. 2005. Corpus Linguistics and the Greek of the New Testament. Sheffield Phoenix Press Ltd.

Scott M. 1997. PC analysis of keywords – and key keywords // System. Vol. 25 (2), 236.

Williams R. 1983. Keywords: a vocabulary of culture and society. Oxford university press.

P.M. Shitikov
Tobolsk, Russia

PROSPECTS OF ANTIQUE CORPORATE RESEARCHES

Abstract. This article discusses the prospects for using of modern computer tools for working with large texts, commonly known as corpus linguistics, for ancient texts. Studying the works, signed by the name of John the Theologian, the author examines the main stages of corpus analysis. As a result, it is indicated that that methods can serve as a basis for critical analysis of the text, study of the author's system of the images and, most importantly, translational research.

Keywords: metaphor; corpus linguistic; translation; Bible; cognitive metaphor theory concept.

About the author: Shitikov Peter Mikhailovich, candidate of Philological Studies, candidate of Theology, associate professor of the Department of philological education Tyumen State University (branch in Tobolsk),