

ПРОБЛЕМА РЕПРЕЗЕНТАЦИИ В АВТОРСКИХ КОРПУСАХ

*Джуманиязова Интизор Давлатназар кизи
преподаватель университета Маъмун, город Хива.*

Аннотация

Статья посвящена анализу проблемы репрезентативности в авторских корпусах как одного из ключевых критериев корпусной лингвистики. Рассматриваются сущность и основные подходы к определению репрезентативности, её роль в создании национальных и авторских корпусов, а также влияние на достоверность лингвистических исследований. Особое внимание уделено различию между объёмом корпуса и его соразмерностью изучаемому языковому материалу. На основе трудов В. Рыкова, И.Ф. Ганиевой и А.М. Лаврентьева раскрываются теоретические аспекты репрезентативности, подчёркивается отличие корпуса от простой коллекции текстов. Авторские корпуса рассматриваются как полноценная модель языка писателя, включающая жанровое и стилевое разнообразие его произведений. Описываются практические принципы формирования и расширения авторского корпуса, а также влияние жанровой структуры и выбора единицы текста на его репрезентативность. Делается вывод о том, что репрезентативность является необходимым условием создания качественного авторского корпуса и основой для проведения всестороннего лингвистического анализа творчества автора.

Ключевые слова: корпусная лингвистика, репрезентативность, авторский корпус, сбалансированность, жанровая структура, лингвистическая аннотация, частотный анализ, текстовая база, языковая модель, цифровые корпуса.

Одним из ключевых понятий корпусной лингвистики является **репрезентативность**. Под репрезентативностью понимают необходимое и соразмерное описание текстов корпуса, позволяющее считать корпус адекватным отражением языка или речевой деятельности, которую он моделирует.¹ Первым условием создания любого корпуса является определение его репрезентативности в широком смысле. Основная задача авторских корпусов — обеспечить репрезентацию языка писателя, его творческого стиля и художественного мира. Существуют различные подходы к определению репрезентативности. Нельзя рассчитать её строго математически в отношении общенациональных корпусов, однако концепция репрезентативности может и должна использоваться при проектировании национальных,

¹ Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005. – С. 5.

специализированных и авторских корпусов. Корпус способен охватывать различные аспекты языка в цифровом виде, что делает возможной автоматическую обработку текста с учётом его лингвистических особенностей.² Создание полностью репрезентативного корпуса — задача сложная, однако чрезвычайно важная. Корпуса, максимально соответствующие требованиям репрезентативности и сбалансированности, рассматриваются как идеальные и служат мощным инструментом для лингвистов и специалистов, изучающих язык как форму культуры.³

Корпусная лингвистика предоставляет теоретические основы и практические механизмы создания репрезентативных массивов данных для научных исследований.⁴ Широкий спектр статистических и лингвистических возможностей корпусов позволяет пользователю получать информацию о частотности словоформ, лексем, грамматических категорий, динамике их употребления, а также о совместной встречаемости единиц (униграмм, биграмм, триграмм). Репрезентативность корпуса за определённый период делает возможным изучение динамики лексического состава, его жанровых, стилистических и авторских особенностей. Репрезентативность корпуса за определённый период делает возможным изучение динамики лексического состава, его жанровых, стилистических и авторских особенностей. Таким образом, основным аспектом, который следует учитывать при создании корпуса, является его репрезентативность. Она включает в себя определение размера корпуса, то есть количества и типов текстов, включаемых в корпус, количества слов в одном тексте и общего количества слов в корпусе, а также типов данных.⁵ При создании корпуса необходимо учитывать размер корпуса, количество и типы включаемых текстов, объём каждого текста, общую совокупность словоупотреблений, а также тип данных. Для авторских корпусов важно обеспечивать репрезентативность творчества конкретного автора, отражая особенности его стиля и жанровое разнообразие произведений.

Любой корпус отражает речевую деятельность человека — сложный семиотический объект. Достоверность информации, полученной из корпуса, зависит от репрезентативности и сбалансированности.⁶ Авторские корпуса должны предоставлять максимально полную и непротиворечивую информацию

² Валиев Т. К проблемам кибернетической лексикографии и языкового корпуса // Научный вестник. - Самарканд, 2016. №2 (96). – С. 68.

³ Лаврентьев А.М. Корпусная лингвистика: идеология, методы, технологии. file:///C:/Users/User/Desktop/KORPUS.pdf

⁴ Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005. – С. 5.

⁵ Koder – A multi-register corpus for investigating register variation in contemporary German. ISSN 2243-4712. Asociación Española de Lingüística de Corpus (AELINCO) DOI 10.32714/ricl.07.04. – pp. 69-83.

⁶ Богоявленская Ю.В. Сопоставительный объектно-ориентированный корпус: определение понятия и принципы формирования. Уральский государственный педагогический университет, Екатеринбург, Россия УДК 81'42(045) – С. 13.

об авторе и его произведениях. Задача создателей корпусов — собрать, структурировать и обработать максимально релевантный текстовый материал. Репрезентативность корпуса определяется не только количеством текстов, но и их связностью, жанровым и стилевым разнообразием.

Репрезентативность и пропорциональность корпуса отражают степень достаточности и соразмерности описания языка в текстах разных эпох, жанров, стилей и авторов. Эти критерии обеспечивают достоверность результатов, получаемых при работе с корпусом. Авторские корпуса позволяют быстро и точно анализировать язык и стиль произведений конкретного автора. Исходя из принципа репрезентативности, тексты автора должны быть равномерно распределены по жанрам и включены в базу корпуса в полном объёме.

В условиях современных информационных технологий авторские корпуса становятся важнейшим инструментом исследования творчества писателя. Они являются «выверенной компьютерной моделью» языка автора, обеспечивая доступность и понимание его произведений. При создании корпуса необходимо тщательно анализировать тексты, определять их структуру и жанровую принадлежность. В. Рыков подчёркивает, что корпуса определяют достоверность информации, поэтому проблему репрезентативности можно рассматривать как «проблему адекватного отражения, адаптации или комбинирования больших массивов текстов или других фрагментов речевой деятельности».⁷ И.Ф. Ганиева отмечает, что репрезентативность — это не объём материала, а соразмерность представленного фрагмента языка. В то же время одной из особенностей корпусов является их значительный объём (миллионы словоупотреблений) и наличие лингвистически аннотированного материала. В статье «Корпусная лингвистика: идеология, методы, технологии» А.М. Лаврентьев пишет, что корпус отличается от текстовой коллекции (библиотеки) своей репрезентативностью, а также языковой разметкой. Он подчёркивает, что объём корпуса должен соответствовать задачам исследования.⁸ Авторские корпуса представляют собой массив текстов, охватывающий основные жанры творчества автора (прозу, поэзию, драму) и отличаются от электронной библиотеки рядом признаков. Поскольку авторские корпуса предназначены для лингвистов, в них необходимо обеспечивать пропорциональное включение текстов по жанрам и типам.

Создание авторского корпуса, его расширение, доработка и завершение являются важнейшими этапами формирования полноценной базы данных.

⁷ Рыков В.В. Корпус текстов как реализация объектно-ориентированной парадигмы. 2002. <https://www.dialog-21.ru/digest/2002/articles/rykov/>

⁸ Лаврентьев А.М. Корпусная лингвистика: идеология, методы, технологии. file:///C:/Users/U ser/Desktop/KORPUS.pdf

Интернет играет значительную роль в сборе материалов и обеспечении открытого доступа к корпусам. Благодаря Интернету можно включать в корпуса не только художественные произведения, но и публицистику, переписку и другие малые жанры. С течением времени корпус может расширяться, однако изменения не должны нарушать его репрезентативность.⁹ Разнообразие, но и точность информации об авторе имеют большое значение для репрезентативности корпусов авторства. По мнению В.П. Захарова и С.Ю. Богдановой, при определении жанрово-тематической структуры корпуса важно учитывать, какую единицу считать текстом, что напрямую влияет на репрезентативность корпуса.¹⁰

Список использованной литературы:

1. Абжалова М. Языковые модули программы редактирования и анализа: монография. - Ташкент: Надирабегим, 2021. – С. 75.
2. Богоявленская Ю.В. Сопоставительный объектно-ориентированный корпус: определение понятия и принципы формирования. Уральский государственный педагогический университет, Екатеринбург, Россия УДК 81'42(045) – С. 13.
3. Валиев Т. К проблемам кибернетической лексикографии и языкового корпуса // Научный вестник. - Самарканд, 2016. №2 (96). – С. 68.
4. Джуманиязова Интизор Давлатназар кизи. Использование авторского корпуса в преподавании. Материалы международной научно-практической конференции «Проблемы и решения при использовании синхронного перевода и общность языка и культуры» - Ургенч 2025.
5. Захаров В., Богданова С. Корпусная лингвистика. – СПб. Санкт-Петербургский ГУ, 2020. – С. 23.
6. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011. – С. 36.
7. Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005. – С. 5.
8. Koder – A multi-register corpus for investigating register variation in contemporary German. ISSN 2243-4712. Asociación Española de Lingüística de Corpus (AELINCO) DOI 10.32714/ricl.07.04. – стр. 69-83.
9. Лаврентьев А.М. Корпусная лингвистика: идеология, методы, технологии. file:///C:/Users/User/Desktop/ KORPUS.pdf
- 10.Рыков В.В. Корпус текстов как реализация объектно-ориентированной парадигмы. 2002.
- 11.<https://www.dialog-21.ru/digest/2002/articles/rykov/>
- 12.<https://www.twirpx.com/file/483985/>

⁹ Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005. – С. 5.

¹⁰ Захаров В.П., Богданова С.Ю. Корпусная лингвистика. – Иркутск: ИГЛУ, 2011. – С. 36.