

Гиздулин Эльдар Фатихович, стажер-исследователь, Ферганский государственный университет



КОРПУСНАЯ ЛИНГВИСТИКА КАК МЕТОД АНАЛИЗА ХУДОЖЕСТВЕННОГО ТЕКСТА

<https://zenodo.org/records/18767155>

Аннотация: Статья посвящена анализу корпусной лингвистики как метода изучения художественного текста. Рассматриваются ключевые определения, принципы построения литературных корпусов и основные аналитические процедуры — частотный анализ, исследование коллокаций, стилистическая идиография. Особое внимание уделяется применению Национального корпуса русского языка для решения филологических задач, а также потенциалу корпусных методов в выявлении авторского идиостя и эволюции художественной речи.

Ключевые слова и выражения: корпусная лингвистика, художественный текст, Национальный корпус русского языка, коллокации, авторский идиостиль, частотный анализ, стилистическая идиография.

KORPUS LINGVISTIKASI BADIY MATNNI TAHLIL QILISH USLUBI SIFATIDA

Annotatsiya: Maqola korpus lingvistikasini badiiy matnni o'rganish usuli sifatida tahlil qilishga bag'ishlangan. Asosiy ta'riflar, adabiy korpuslarni tuzish tamoyillari va asosiy tahliliy jarayonlar - chastotali tahlil, kollokatsiyalarni o'rganish, uslubiy idiografiya ko'rib chiqiladi. Rus tili milliy korpusini filologik muammolarni hal qilishda qo'llashga, shuningdek, muallif idiositili va badiiy nutq evolyutsiyasini aniqlashda korpus usullarining imkoniyatlariga alohida e'tibor qaratilmoqda.

Kalit so'z va iboralar: korpus lingvistikasi, badiiy matn, rus tili milliy korpusi, kollokatsiyalar, muallif idiositil, chastotali tahlil, uslubiy idiografiya.

CORPUS LINGUISTICS AS A METHOD OF LITERARY TEXT ANALYSIS

Annotation: The article is devoted to the analysis of corpus linguistics as a method of studying literary text. The key definitions, principles of literary corpus construction, and basic analytical procedures — frequency analysis, collocation research, and stylistic idiography - are considered. Special attention is paid to the use of the National Corpus of the Russian language to solve philological problems, as well as the potential of corpus methods in identifying the author's idiosyncrasy and the evolution of artistic speech.

Key words and phrases: corpus linguistics, literary text, National corpus of the Russian language, collocations, author's idiosyncrasy, frequency analysis, stylistic idiography.

Вводная часть.

Цифровая трансформация гуманитарного знания в XXI веке привела к формированию нового научного направления — цифровой филологии, центральным инструментом которой стала корпусная лингвистика. Эта методология, основанная на эмпирическом анализе больших массивов текстовых данных с использованием компьютерных технологий, предлагает радикальный сдвиг от интуитивно-интерпретативных подходов к объективному, количественно верифицируемому исследованию языковых явлений. В контексте литературоведения корпусный подход позволяет выявлять скрытые структуры художественной речи, недоступные при традиционном чтении, и реконструировать авторский идиостиль на основе статистически значимых закономерностей.

Актуальность применения корпусных методов к анализу художественного текста обусловлена не только технологическими возможностями, но и теоретическими вызовами: необходимостью преодоления субъективности интерпретации, желанием систематизировать наблюдения за языковой индивидуальностью писателя и стремлением связать микроструктуру текста с макроконтекстом эпохи. Особенно продуктивным оказался этот подход в исследованиях классической русской литературы XIX–XX веков, где богатство языкового материала сочетается с высокой степенью стилистической самосознательности авторов. Цель данной статьи — проанализировать теоретические основы, методические процедуры и практические результаты применения корпусной лингвистики к изучению художественного текста, с акцентом на русскоязычную филологическую традицию.

Анализ литературы и методология исследования.

В работе использован комплексный методологический подход, включающий теоретический анализ, сравнительно-описательный метод, корпусное лингвистическое исследование и обобщение эмпирических данных. В качестве основного источника текстовых данных выступал Национальный корпус русского языка (НКРЯ), а также специализированные литературные корпуса, созданные в рамках Уральского федерального университета и других научных центров.

Теоретическую базу составили работы В. А. Райскиной, Ю. А. Волосновой, М. Ю. Мухина, А. Д. Матюшкина, Л. Г. Бабенко и А. Г. Гачевой. Критерии отбора источников: индексация в ВАК/Scopus, наличие открытого доступа, соответствие тематике корпусной лингвистики и анализа художественного текста. Особое внимание уделено исследованиям, в которых количественные данные последовательно переводятся в качественную филологическую интерпретацию.

Результаты и обсуждение.

Корпусная лингвистика определяется как «наука о языке, основанная на анализе репрезентативных, структурированных и размеченных массивов текстов (корпусов) с использованием компьютерных методов» [2, 52]. Литературный корпус — это не просто собрание текстов, а строго отобранный массив, репрезентативный для решения конкретной исследовательской задачи: изучения идиостиля одного автора, сравнения стилей нескольких писателей или анализа эволюции художественного языка в определённый исторический период. Ключевым преимуществом корпусного подхода является его эмпирическая обоснованность: вместо единичных примеров,

подбираемых для иллюстрации гипотезы, исследователь оперирует полным массивом употреблений языковой единицы, что позволяет делать статистически валидные выводы [3, 87].

Основные аналитические процедуры корпусной лингвистики позволяют решать широкий спектр филологических задач. Частотный анализ выявляет не только абсолютную встречаемость слова, но и его относительную частотность в сравнении с базовым корпусом (например, НКРЯ). Это особенно важно для диагностики авторской лексической специфики: слова, редкие в общем языке, могут оказаться ключевыми для идиостиля писателя [4, 119]. Например, у Ф. М. Достоевского значительно выше среднего по корпусу частотность слов, связанных с внутренним напряжением («мука», «страдание», «тревога»), тогда как у И. С. Тургенева доминируют лексемы, описывающие природу и внешние состояния («ветер», «тень», «шёпот»).

Конкордансный анализ предоставляет все контексты употребления искомой единицы, что даёт возможность проследить её семантические вариации, функциональные особенности и стилистическую окраску. Так, анализ контекстов слова «сердце» у Л. Н. Толстого показывает, что оно чаще всего сочетается с глаголами движения («забилось», «дрогнуло», «сжалось»), что свидетельствует о физиологизации внутреннего переживания — характерной черте его психологизма [5, 2031].

Особую ценность представляет коллокационный анализ — исследование слов, устойчиво сочетающихся друг с другом. С помощью статистических мер (t-score, MI-score) можно выявить так называемые «индивидуально-авторские биграммы» — пары слов, характерные для одного писателя и статистически незначимые у других. В рамках Уральского стилиометрического проекта был создан корпус избранных текстов Л. Н. Толстого, Ф. М. Достоевского, И. С. Тургенева, И. А. Гончарова и А. П. Чехова объёмом около 4 млн текстоформ [5, 2027]. Сопоставительный анализ выявил, что выражение «шаги персонажей» (в формах «его шаги», «лёгкие шаги», «шаги замерли») является статистически значимым маркером стиля Л. Н. Толстого. Это позволяет сделать вывод о его особом внимании к физиологии движения как проявлению внутреннего состояния героя — черте, ранее отмечавшейся в литературоведении, но теперь подтверждённой количественно [5, 2030].

Национальный корпус русского языка (НКРЯ) стал центральным инструментом для подобных исследований в России. Он предоставляет не только огромный массив текстов (более 1 млрд слов), но и мощный программный интерфейс для поиска, анализа и визуализации данных. Учебное пособие А. Д. Матюшкина «Вслед за словом. Анализ художественного текста с опорой на Национальный корпус русского языка» содержит практические алгоритмы работы с ключевыми словами, коллокациями и метафорическими моделями, позволяя студентам и преподавателям проверять свои читательские гипотезы на эмпирической основе [6]. Это способствует формированию новой филологической компетенции, сочетающей традиционное понимание текста с навыками работы с большими данными.

Однако корпусный анализ не является самоцелью и не заменяет традиционный литературный анализ. Его главная функция — служить инструментом для генерации и верификации гипотез. Корпусные данные предоставляют объективную основу, но их интерпретация требует глубокого филологического чутья и знания историко-культурного контекста. Например, выявление высокой частотности мотива смерти у Достоевского может быть связано не только с его личными предпочтениями, но и с доминирующими идеями эпохи (нигилизм, экзистенциальный кризис), жанровыми конвенциями (роман-полемика) или биографическими обстоятельствами (смертный

приговор, ссылка). Таким образом, корпусная лингвистика наиболее продуктивна в синтезе с другими методами — стилистическим, историко-литературным, культурологическим и психолингвистическим анализом.

Более того, корпусный подход позволяет решать задачи, недоступные другим методам. Например, с его помощью можно проследить эволюцию авторского идиостиля на протяжении творческого пути. Анализ ранних и поздних произведений Л. Н. Толстого показывает снижение частотности абстрактной философской лексики и рост доли конкретных, бытовых деталей — что коррелирует с его переходом к «народнической» эстетике [4, 122]. Аналогично, у А. П. Чехова в поздних рассказах наблюдается увеличение числа незавершённых синтаксических конструкций и пауз, что объективно подтверждает тезис о развитии у него «подтекстовой» манеры письма.

Важным направлением является применение корпусных методов к анализу перевода. Сравнение оригинала и перевода в параллельных корпусах позволяет выявить стратегии передачи авторского стиля, а также типичные искажения, вносимые переводчиками. Например, исследования показывают, что в английских переводах Достоевского часто сглаживается его характерная для русского языка полисиндетическая структура (множество союзов «и»), что ведёт к упрощению ритма и потери экспрессивности [8, 252].

Тем не менее, существуют и ограничения корпусного подхода. Во-первых, он лучше всего работает с поверхностными уровнями языка (лексика, морфология, простая синтаксическая структура), тогда как анализ сложных метафор, иронии или межтекстуальных отсылок остаётся в сфере традиционной герменевтики. Во-вторых, качество анализа напрямую зависит от качества корпуса: наличие ошибок в разметке, неполнота текстов или несбалансированность жанрового состава могут исказить результаты. В-третьих, существует риск иллюзии объективности: статистическая значимость не всегда равна смысловой важности. Слово может быть редким, но ключевым для понимания текста (например, «катаклизм» в одном из рассказов Чехова), и наоборот — частое, но функционально нейтральное.

Несмотря на эти ограничения, корпусная лингвистика открывает новые горизонты для изучения художественного текста. Она позволяет систематизировать наблюдения, выявлять скрытые структуры авторского идиостиля, проследивать динамику развития художественного языка и объективно сравнивать стили разных эпох и культур. Будущее филологии, несомненно, связано с дальнейшей интеграцией цифровых методов, которые, будучи правильно применяемыми, не подменяют, а обогащают глубину гуманитарного понимания.

Вывод.

Корпусная лингвистика утвердилась как значимый методологический инструмент в современной филологии, обеспечивающий эмпирическую основу для анализа художественного текста. Её ценность заключается не в замене традиционных интерпретативных подходов, а в их дополнении объективными, количественно верифицируемыми данными, что позволяет выявлять скрытые структуры авторского идиостиля и системные закономерности художественной речи. Наибольшую продуктивность демонстрируют такие процедуры, как частотный, конкордансный и коллокационный анализ, особенно в сопоставительном режиме — при изучении стилей разных авторов, жанров или периодов творчества одного писателя. Центральную роль в развитии этого направления в русскоязычной науке играет Национальный корпус русского языка, который сочетает масштаб текстовой базы с возможностями точного поиска и статистической обработки, делая корпусные методы доступными как для

исследователей, так и для образовательных целей. При этом эффективность корпусного подхода достигается лишь в синтезе с глубоким филологическим анализом: статистическая значимость должна интерпретироваться в контексте историко-литературных, культурных и биографических реалий. Перспективы дальнейшего развития метода связаны с интеграцией технологий искусственного интеллекта для анализа сложных семантических и нарративных структур, расширением мультязычных параллельных корпусов и внедрением корпусных данных в преподавание литературы, стилистики и сравнительно-исторического языкознания.

Список использованной литературы:

1. Райскина, В. А., Дубнякова, О. А. Современные методы корпусной лингвистики при анализе текста (на примере корпуса BFM) / В. А. Райскина, О. А. Дубнякова // Актуальные вопросы современной науки. — 2015. — № 2. — С. 85–89.
2. Волоснова, Ю. А. Корпусная лингвистика: проблемы и перспективы / Ю. А. Волоснова // Вестник Московского университета. Серия 19: Лингвистика и межкультурная коммуникация. — 2006. — № 1. — С. 52–58.
3. Пиотровский, Р. Г. Математическая лингвистика / Р. Г. Пиотровский. — М.: Высшая школа, 1977. — 384 с.
4. Мухин, М. Ю. Лексическая статистика и стиль авторов классической прозы XIX в. / М. Ю. Мухин // Теоретическая семантика и идеографическая лексикография: Словарь. Дискурс. Корпус. — Екатеринбург: Изд-во УрФУ, 2019. — С. 119–126.
5. Mukhin, M. Yu., Mukhin, N. Yu. Idiostyle characteristics of lexical compatibility in the 19th-century prose: Ural stylometric project / M. Yu. Mukhin, N. Yu. Mukhin // Journal of Siberian Federal University. Humanities & Social Sciences. — 2020. — Vol. 13, № 12. — P. 2027–2034.
6. Матюшкин, А. Д. Вслед за словом. Анализ художественного текста с опорой на Национальный корпус русского языка / А. Д. Матюшкин. — М.: Studiorum, 2021. — 128 с.
7. Чередилина, М. Ю. Корпусный анализ в исследовании авторского стиля / М. Ю. Чередилина // Бизнес. Образование. Право. — 2016. — № 3 (36). — С. 114–118.
8. Гачева, А. Г. Цифровая филология: новые методы и старые вопросы / А. Г. Гачева // Новое литературное обозрение. — 2018. — № 150. — С. 245–258.
9. Бабенко, Л. Г. Идеографическая семантика и корпусные исследования / Л. Г. Бабенко // Вопросы языкознания. — 2020. — № 4. — С. 3–18.