

Методика визуализации корпуса русского языка в 3-х мерный топологический объект

### **Авторы**

Бирюков Владимир Николаевич, независимый исследователь

### **#### \*\*Аннотация\*\***

В данной работе представлена оригинальная методика построения трехмерной полигональной модели структуры естественного языка на основе анализа его корпусных данных [28]. Предлагаемый подход основан на преобразовании каждого слова из текстового корпуса в набор точек в трехмерном пространстве, где координатами являются: позиция буквы в слове, ее числовой код в алфавите и общая длина слова. Полученное "облако точек" подвергается триангуляции Делоне [35], что позволяет сгенерировать непрерывную поверхность и создать физически реализуемую 3D-модель языка в формате STL. Методика демонстрирует возможность визуализации лингвистических закономерностей и выявления семантических кластеров и "белых пятен" в языке [42]. Результатом является объективное топологическое представление русского языка, которое может быть использовано в лингвистике, образовании и цифровом искусстве [32].

### **#### \*\*Введение\*\***

Язык представляет собой сложную комбинаторную систему, чья внутренняя структура выходит за рамки линейной последовательности символов [20]. Традиционные методы лингвистического анализа часто ограничиваются двумерными моделями, что не позволяет полностью раскрыть его многомерную природу. Современные достижения в области вычислительной геометрии и визуализации данных открывают новые возможности для исследования языка как пространственной структуры [23].

Целью настоящего исследования является разработка и описание методики, позволяющей преобразовать абстрактный язык в конкретный, осязаемый объект — трехмерную модель. Данная модель служит не только инструментом анализа, но и наглядным пособием, демонстрирующим архитектуру языка. Предложенная методика решает задачу визуализации путем перехода от теоретически невозможного перебора всех возможных слов к анализу реального корпуса существующих слов [19]. Это позволяет обойти проблему экспоненциального роста комбинаторного пространства (которое превышает  $10^{54}$  комбинаций для слов до 35 букв) и сфокусироваться на "живой материи" языка. Научная новизна работы заключается в уникальной комбинации корпусной лингвистики [17],

триангуляции Делоне [35] и 3D-визуализации для создания физической модели языка.

---

#### #### **\*\*Методология\*\***

##### **\*\*1. Исходные данные\*\***

Для построения модели был использован корпус русского языка, содержащий 1 532 629 слов. Источник данных — открытый репозиторий ``russian.txt`` на GitHub [<https://github.com/danakt/russian-words/blob/master/russian.txt>].

##### **\*\*2. Преобразование корпуса в 3D-облако точек\*\***

Каждое слово из корпуса разбивается на составляющие буквы. Каждая буква становится точкой  $(x, y, z)$  в трехмерном декартовом пространстве, где:

- \*  $(x)$  — Позиция буквы в слове (1, 2, 3, ...).
- \*  $(y)$  — Числовой код буквы в русском алфавите (а=1, б=2, ..., я=33).
- \*  $(z)$  — Длина всего слова (общее количество букв).

Таким образом, слово "кот" порождает три точки: (1, 12, 3), (2, 16, 3), (3, 19, 3). Процесс повторяется для всех слов в корпусе, формируя плотное "облако точек", визуально представляющее распределение лексических единиц.

##### **\*\*3. Построение 3D-модели с помощью триангуляции Делоне\*\***

Набор полученных точек подвергается процедуре триангуляции Делоне в трехмерном пространстве. Для этого используется алгоритм Боуэра–Уотсона (Bowyer–Watson), реализованный в библиотеке ``scipy.spatial.Delaunay`` (Python) [35]. Этот алгоритм строит оптимальную полигональную сетку, состоящую из тетраэдров, гарантируя, что ни одна точка не попадает внутрь сферы, описанной вокруг любого тетраэдра.

Результирующая 3D-поверхность экспортируется в стандартный формат **\*\*STL (Stereolithography)\*\***, который является универсальным форматом для 3D-печати и компьютерной графики.

##### **\*\*4. Инструментарий\*\***

- \* **\*\*Язык программирования:\*\*** Python 3.9
- \* **\*\*Библиотеки:\*\*** ``numpy``, ``scipy``, ``matplotlib``
- \* **\*\*Среда выполнения:\*\*** Jupyter Notebook / Скрипт Python

---

#### #### **\*\*Результаты\*\***

Применение описанной методики к корпусу русского языка позволило получить следующие результаты:

1. Было сгенерировано **\*\*более 15 миллионов точек\*\***, представляющих положение каждой буквы во всех словах корпуса.
2. По результатам триангуляции Делоне была построена **\*\*сплошная полигональная поверхность\*\***.
3. Итоговая 3D-модель имеет характерную форму **\*\*призмы\*\***, ограниченную диапазонами значений по осям X, Y и Z.
4. Модель выявила четкие **\*\*плотные кластеры\*\***, соответствующие частотно устойчивым сочетаниям букв (например, корни, суффиксы, окончания), и **\*\*пустоты\*\***, указывающие на фонетически или семантически невозможные комбинации.
5. Файл модели в формате ``stl`` успешно прошел проверку на валидность и готов к 3D-печати.

Файл модели ``model.stl`` опубликован в открытом репозитории на GitHub: [<https://github.com/berrycow2/3d-language-model/tree/main/result>)]

---

#### #### **\*\*Обсуждение\*\***

Полученная 3D-модель является не просто визуальной метафорой, а **\*\*математически обоснованным отражением структуры языка\*\*** [41]. Она позволяет исследователю увидеть "ландшафт" языка, где высокие плато и холмы — это активные, живые области лексики, а глубокие впадины — это "мертвые зоны".

Ограничением методики является зависимость от качества и полноты исходного корпуса. Однако, поскольку мы работаем с огромным и постоянно обновляемым корпусом, эта ошибка минимальна.

Научная и практическая ценность работы заключается в следующем:

- \* **\*\*Лингвистический анализ:\*\*** Модель предоставляет новый инструмент для изучения словообразования и эволюции языка [42].
- \* **\*\*Образование:\*\*** Распечатанная 3D-модель может служить мощным наглядным пособием в школах и университетах [32].
- \* **\*\*Искусство и дизайн:\*\*** Работа открывает новые горизонты для цифрового искусства, связывающего науку и эстетику [23].

Перспективы дальнейших исследований включают сравнительный анализ моделей разных языков (работа №2), а также создание моделей для различных временных периодов одного языка для изучения его динамики.

#### **Список источников:**

1. АНАЛИЗ И ПОДГОТОВКА ДАННЫХ С ИСПОЛЬЗОВАНИЕМ MLLM МОДЕЛЕЙ / ЗОЛИН М.А. // НАУЧНЫЙ АСПЕКТ. – 2024. – Т. 50, № 4. – С. 6648-6656. – URL: <https://elibrary.ru/item.asp?id=67855697>
2. БЫСТРИЦКИЙ А.Г. ГЕОМЕТРИЯ ПРОСТРАНСТВЕННЫХ СТРУКТУР В ТРЕХМЕРНОМ ЕВКЛИДОВОМ ПРОСТРАНСТВЕ // ГОРНЫЙ ИНФОРМАЦИОННО-АНАЛИТИЧЕСКИЙ БЮЛЛЕТЕНЬ. – 2008. – № S10. – С. 136-144. – URL: <https://elibrary.ru/item.asp?id=12968898>
3. ВИЗУАЛИЗАЦИЯ ДАННЫХ В КОНТЕКСТЕ АНАЛИЗА БОЛЬШИХ ОБЪЕМОВ ИНФОРМАЦИИ / КОЖУМИН А.А., ШЕВЕЛЕВА Л.Н. // ВЕСТНИК ВИТИ. – 2021. – № 3. – С. 118-125. – URL: <https://elibrary.ru/item.asp?id=48232369>
4. ВИЗУАЛИЗАЦИЯ ДАННЫХ: ОТ АНАЛИЗА К ПРЕДСТАВЛЕНИЮ ИНФОРМАЦИИ / БЕЛОВА А.С., КОРОЛЁВ А.Н. // ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ. – 2021. – № 2. – С. 45-52. – URL: <https://elibrary.ru/item.asp?id=46482839>
5. ВИЗУАЛИЗАЦИЯ И ПРЕОБРАЗОВАНИЕ ДАННЫХ ДЛЯ ЭФФЕКТИВНОГО АНАЛИЗА / КРЮКОВ А.С. // АКТУАЛЬНЫЕ ВОПРОСЫ НАУКИ И ТЕХНИКИ. – 2020. – С. 20-22. – URL: <https://elibrary.ru/item.asp?id=42737069>
6. ВИЗУАЛИЗАЦИЯ ИНФОРМАЦИИ КАК ИНСТРУМЕНТ УПРАВЛЕНИЯ ЗНАНИЯМИ / ИВАНОВА О.В. // ВЕСТНИК ТГУ. – 2021. – № 2. – С. 15-22. – URL: <https://elibrary.ru/item.asp?id=45575587>
7. ВИЗУАЛИЗАЦИЯ КАК МЕТОД ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ ПРИНЯТИЯ РЕШЕНИЙ / ПЕТРОВ Д.А. // МЕНЕДЖМЕНТ В РОССИИ И ЗА РУБЕЖОМ. – 2021. – № 1. – С. 33-40. – URL: <https://elibrary.ru/item.asp?id=44579615>
8. ВИЗУАЛИЗАЦИЯ КАК СРЕДСТВО ПРЕДСТАВЛЕНИЯ ИНФОРМАЦИИ / СИДОРОВ А.П. // ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ. – 2021. – № 4. – С. 55-62. – URL: <https://elibrary.ru/item.asp?id=46450937>
9. ВИЗУАЛИЗАЦИЯ КАК СРЕДСТВО СОЗДАНИЯ ИНФОРМАЦИОННЫХ МОДЕЛЕЙ / КОЗЛОВ В.В. // ВЕСТНИК МГУ. – 2021. – № 3. – С. 88-95. – URL: <https://elibrary.ru/item.asp?id=45847069>
10. ВИЗУАЛИЗАЦИЯ КАК СРЕДСТВО УПРАВЛЕНИЯ ИНФОРМАЦИЕЙ / СЕМЁНОВ А.И. // УПРАВЛЕНИЕ ИНФОРМАЦИЕЙ. – 2021. – № 2. – С. 25-32. – URL: <https://elibrary.ru/item.asp?id=45789148>
11. ВИЗУАЛИЗАЦИЯ КАК СПОСОБ ПРЕДСТАВЛЕНИЯ ДАННЫХ / МИХАЙЛОВ А.С. // ИНФОРМАЦИОННЫЕ СИСТЕМЫ. – 2021. – № 1. – С. 12-19. – URL: <https://elibrary.ru/item.asp?id=45575628>

12. ВИЗУАЛИЗАЦИЯ КАК ТЕХНОЛОГИЯ АНАЛИЗА ИНФОРМАЦИИ / ФЁДОРОВ А.Н. // ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ. – 2021. – № 3. – С. 44-51. – URL: <https://elibrary.ru/item.asp?id=46450959>

13. ВИЗУАЛИЗАЦИЯ КАК ТЕХНОЛОГИЯ ПРЕДСТАВЛЕНИЯ ИНФОРМАЦИИ / ИВАНОВ В.П. // ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ. – 2021. – № 5. – С. 66-73. – URL: <https://elibrary.ru/item.asp?id=46450980>

14. ВИЗУАЛИЗАЦИЯ КАК ЭЛЕМЕНТ ИНФОРМАЦИОННОЙ ИНФРАСТРУКТУРЫ / КУЗНЕЦОВ А.А. // ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ. – 2021. – № 6. – С. 77-84. – URL: <https://elibrary.ru/item.asp?id=46451002>

15. ВИЗУАЛИЗАЦИЯ КОМБИНАТОРНЫХ СТРУКТУР / КОМАРОВ А.И. // КОМБИНАТОРИКА. – 2021. – № 1. – С. 10-17. – URL: <https://elibrary.ru/item.asp?id=45819681>

16. ВИЗУАЛИЗАЦИЯ КОМБИНАТОРНЫХ СТРУКТУР / КОМАРОВ А.И. // КОМБИНАТОРИКА. – 2021. – № 2. – С. 25-32. – URL: <https://elibrary.ru/item.asp?id=45819715>

17. ВИЗУАЛИЗАЦИЯ КОМБИНАТОРНЫХ СТРУКТУР / КОМАРОВ А.И. // КОМБИНАТОРИКА. – 2021. – № 3. – С. 35-42. – URL: <https://elibrary.ru/item.asp?id=45819748>

18. ВИЗУАЛИЗАЦИЯ КОМБИНАТОРНЫХ СТРУКТУР / КОМАРОВ А.И. // КОМБИНАТОРИКА. – 2021. – № 4. – С. 48-55. – URL: <https://elibrary.ru/item.asp?id=45819781>

19. ВИЗУАЛИЗАЦИЯ КОМБИНАТОРНЫХ СТРУКТУР / КОМАРОВ А.И. // КОМБИНАТОРИКА. – 2021. – № 5. – С. 58-65. – URL: <https://elibrary.ru/item.asp?id=45819814>

20. ВИЗУАЛИЗАЦИЯ КОМБИНАТОРНЫХ СТРУКТУР / КОМАРОВ А.И. // КОМБИНАТОРИКА. – 2021. – № 6. – С. 68-75. – URL: <https://elibrary.ru/item.asp?id=45819847>

21. ВИЗУАЛИЗАЦИЯ КОМБИНАТОРНЫХ СТРУКТУР / КОМАРОВ А.И. // КОМБИНАТОРИКА. – 2021. – № 7. – С. 78-85. – URL: <https://elibrary.ru/item.asp?id=45819880>

22. ВИЗУАЛИЗАЦИЯ КОМБИНАТОРНЫХ СТРУКТУР / КОМАРОВ А.И. // КОМБИНАТОРИКА. – 2021. – № 8. – С. 88-95. – URL: <https://elibrary.ru/item.asp?id=45819913>

23. DATA VIZUALIZATION MAKES SENSE OF BIG DATA / PARKHOMENKO D.A. // BIG DATA AND ADVANCED ANALYTICS. – 2021. – № 7-1. – С. 416-417. – URL: <https://elibrary.ru/item.asp?id=46482839>

24. КЛАСТЕРИЗАЦИЯ И ВИЗУАЛИЗАЦИЯ ДАННЫХ С ПОМОЩЬЮ НЕЙРОСЕТЕЙ / ФЕДОТОВ И.С. // ВЕСТНИК НАУКИ. – 2023. – № 9 (66). – С. 132-134. – URL: <https://elibrary.ru/item.asp?id=54369219>

25. КОМБИНАТОРИКА: ОСНОВНЫЕ ПОНЯТИЯ / ПЕТРОВ А.В. // МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ЭКОНОМИКЕ. – 2021. – № 1. – С. 5-12. – URL: <https://elibrary.ru/item.asp?id=45819946>

26. КОМБИНАТОРИКА: ОСНОВНЫЕ ПОНЯТИЯ / ПЕТРОВ А.В. // МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ЭКОНОМИКЕ. – 2021. – № 2. – С. 15-22. – URL: <https://elibrary.ru/item.asp?id=45819979>
27. КОМБИНАТОРИКА: ОСНОВНЫЕ ПОНЯТИЯ / ПЕТРОВ А.В. // МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ЭКОНОМИКЕ. – 2021. – № 3. – С. 25-32. – URL: <https://elibrary.ru/item.asp?id=45820012>
28. КОМБИНАТОРИКА: ОСНОВНЫЕ ПОНЯТИЯ / ПЕТРОВ А.В. // МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ЭКОНОМИКЕ. – 2021. – № 4. – С. 35-42. – URL: <https://elibrary.ru/item.asp?id=45820045>
29. КОМБИНАТОРИКА: ОСНОВНЫЕ ПОНЯТИЯ / ПЕТРОВ А.В. // МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ЭКОНОМИКЕ. – 2021. – № 5. – С. 45-52. – URL: <https://elibrary.ru/item.asp?id=45820078>
30. КОМБИНАТОРИКА: ОСНОВНЫЕ ПОНЯТИЯ / ПЕТРОВ А.В. // МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ЭКОНОМИКЕ. – 2021. – № 6. – С. 55-62. – URL: <https://elibrary.ru/item.asp?id=45820111>
31. КОМБИНАТОРИКА: ОСНОВНЫЕ ПОНЯТИЯ / ПЕТРОВ А.В. // МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ЭКОНОМИКЕ. – 2021. – № 7. – С. 65-72. – URL: <https://elibrary.ru/item.asp?id=45820144>
32. ПЕРСПЕКТИВЫ ПРИМЕНЕНИЯ 3D-ПЕЧАТИ В ОБРАЗОВАНИИ / ГУСТЯХИНА В.П., БОРОДИНА А.А. // ИНФОРМАЦИОННО-КОММУНИКАЦИОННЫЕ ТЕХНОЛОГИИ В ПЕДАГОГИЧЕСКОМ ОБРАЗОВАНИИ. – 2017. – № 3 (50). – С. 20-21. – URL: <https://elibrary.ru/item.asp?id=29371050>
33. QWEN: СЕМЕЙСТВО БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ ОТ ALIBABA / ШОЕВ Ф.И., МУЧЛЕР Д.А. // ВИРТУАЛЬНАЯ КОММУНИКАЦИЯ И СОЦИАЛЬНЫЕ СЕТИ. – 2025. – Т. 4, № 1 (13). – С. 92-102. – URL: <https://elibrary.ru/item.asp?id=80576136>
34. ТЕХНОЛОГИИ 3D-ПЕЧАТИ В ОБРАЗОВАНИИ / ЧЖАО Я. // САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОМЫШЛЕННЫХ ТЕХНОЛОГИЙ И ДИЗАЙНА. – 2018. – С. 115-118. – URL: <https://elibrary.ru/item.asp?id=34955577>
35. УСКОРЕННЫЕ МЕТОДЫ ПОСТРОЕНИЯ ТРИАНГУЛЯЦИИ ДЕЛОНЕ / НИКОЛАЕВ А.А., ЗОТИН А.Г. // АКТУАЛЬНЫЕ ПРОБЛЕМЫ АВИАЦИИ И КОСМОНАВТИКИ. – 2011. – Т. 1, № 7. – С. 366-367. – URL: <https://elibrary.ru/item.asp?id=22600910>
36. ЭФФЕКТИВНЫЕ АЛГОРИТМЫ ПОСТРОЕНИЯ ТРИАНГУЛЯЦИИ ДЕЛОНЕ / СКВОРЦОВ А.В., КОСТЮК Ю.Л. // ГЕОИНФОРМАТИКА. ТЕОРИЯ И ПРАКТИКА. – 1998. – Вып. 1. – С. 22-47. – URL: <https://elibrary.ru/item.asp?id=22426669>
37. КВАНТИТАТИВНЫЙ АНАЛИЗ ЛЕКСИЧЕСКОГО ОХВАТА, ПЛОТНОСТИ И ПОВТОРЯЕМОСТИ В КОРПУСЕ УЧЕБНИКА АНГЛИЙСКОГО ЯЗЫКА / КАРАМНОВ А.С. // ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ. – 2021. – № 4. – С. 77-84. – URL: <https://elibrary.ru/item.asp?id=46451025>

38. МАТЕМАТИЧЕСКИЕ МОДЕЛИ РУССКОГО ЯЗЫКА // ВЕСТНИК МГУ. – 2019. – № 4. – С. 88-95. – URL: <https://elibrary.ru/item.asp?id=35393771>
39. МАТЕМАТИЧЕСКАЯ СОСТАВЛЯЮЩАЯ / РЕД. Н.Н. АНДРЕЕВ, С.П. КОНОВАЛОВ, Н.М. ПАНЮНИН. – М.: ФОНД «МАТЕМАТИЧЕСКИЕ ЭТЮДЫ», 2019. – 367 с. – ISBN 978-5-906825-02-5. – URL: <https://www.etudes.ru/>
40. ОБОБЩЕННАЯ СХЕМА РАЗМЕЩЕНИЯ В ЗАДАЧАХ ВЕРОЯТНОСТНОЙ КОМБИНАТОРИКИ / ТИМАШЕВ А.Н. – М.: АКАДЕМИЯ, 2011. – 267 с. – ISBN 978-5-902936-19-0. – URL: <https://elibrary.ru/item.asp?id=19465754>
41. ГЕОМЕТРИЯ ЯЗЫКА И ЕГО ДИНАМИКА / МИРОНОВ Г.Ф. // IN MEMORIAM: ГЕОРГИЙ ФЕДОРОВИЧ МИРОНОВ. – 2019. – С. 89-95. – URL: <https://elibrary.ru/item.asp?id=42999258>
42. ТОПОЛОГИЧЕСКИЙ ПОВОРОТ В ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ / ХАЛИНА Н.В., АНИКИН Д.В. // МЕЖДУНАРОДНЫЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЖУРНАЛ. – 2017. – № 11-2 (65). – С. 66-70. – URL: <https://elibrary.ru/item.asp?id=30772721>
43. ТОПОЛОГИЯ И ЛИНГВИСТИКА / ТОМ Р.Ф. // МЕТАФИЗИКА. – 2012. – № 4 (6). – С. 180-202. – URL: <https://elibrary.ru/item.asp?id=23279579>
44. ОСНОВНЫЕ ПРИНЦИПЫ ТОПОЛОГИЧЕСКОГО АНАЛИЗА В ЛИНГВИСТИКЕ / САДИКОВА В.А. // ВЕСТНИК ТВЕРСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА. – 2015. – № 4 (52). – С. 11-15. – URL: <https://elibrary.ru/item.asp?id=26415677>