

УДК 81 '33

Николаева Марина Николаевна

кандидат филологических наук,
доцент кафедры английской
филологии
Московский городской
педагогический университет
г. Москва, Россия

Marina Nikolaeva

PhD in Philology,
Associate professor
Moscow City University
Moscow, Russia
nikolaevam@mgpu.ru

Овсянникова Марина Анатольевна

кандидат филологических наук,
доцент кафедры английской
филологии
Московский городской
педагогический университет
г. Москва, Россия

Marina Ovsyannikova

PhD in Philology,
Associate professor
Moscow City University
Moscow, Russia
ovsyannikovama@mgpu.ru

ИНСТРУМЕНТЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА
ДЛЯ КОРПУСНОГО АНАЛИЗА ИДИОМ С КОЛОРОНИМОМ GREEN:
СРАВНИТЕЛЬНЫЙ АСПЕКТ

Данная статья представляет результаты сравнительного анализа применения корпусного анализа на ресурсе english-corpora.org (корпус NOW) и инструментов искусственного интеллекта (Mistral, Deepseek, Perplexity, GigaChat) при отборе лексического материала, в частности, фразеологических единиц с компонентом колоронимом *green* в английском языке. Основное внимание уделено смешанной идиоме *green behind the ears*, которая была выделена с помощью ИИ и верифицирована в корпусе NOW. Авторы анализируют ее происхождение и контексты употребления. Также рассматривается фразеологический бленд *green around the gills with envy*, обнаруженный в корпусе, и его интерпретация с помощью ИИ. Исследование демонстрирует как преимущества, так и ограничения доступных в России ИИ-платформ: недостоверность источников у GigaChat и ограниченность временного охвата данных у Mistral и Deepseek. Однако авторы подчеркивают ценность инструментов ИИ для расширения возможностей лингвистического анализа, в особенности в сочетании с традиционными корпусными методами.

Ключевые слова: искусственный интеллект; корпусный анализ; фразеология; колороним; идиома; бленд.

AI TOOLS FOR CORPUS ANALYSIS
OF GREEN-COLOUR IDIOMS: COMPARATIVE STUDY

The article presents the results of a comparative study of corpus data analysis on the resource english-corpora.org (the NOW corpus) and artificial intelligence tools (Mistral, Deepseek, Perplexity, GigaChat) when selecting lexical material, in particular, phraseological units with the colouronym component *green* in English. The main focus is on the mixed idiom

green behind the ears which was identified by AI and verified in the NOW corpus. The authors analyze its origin and usage contexts. The phraseological blend *green around the gills with envy* found in the corpus and its interpretation using AI are also considered. The study demonstrates both the advantages and limitations of the AI platforms available in Russia: the unreliability of sources from GigaChat and the limited time coverage of data from Mistral and Deepseek. However, the authors emphasize AI value for expanding the possibilities of linguistic analysis, especially in combination with traditional corpus methods.

Key words: artificial intelligence; corpus analysis; phraseology; colouronym; idiom; blend.

На территории России доступен агрегатор корпусных текстов English Corpora (<https://www.english-corpora.org>). Наиболее полный корпус на данном сайте – это корпус веб-изданий, который обновляется в режиме реального времени, корпус NOW [1]. Однако с данным ресурсом часто возникают технические проблемы: зависает страница поиска, поиск осуществляется не сразу, часто приходится перезагружать сайт. В настоящий момент с развитием платформ искусственного интеллекта (ИИ) исследователи могут применять новые технологии для лингвистического анализа наряду с традиционными методами [2, 3]. В рамках данного исследования была предпринята попытка воспользоваться доступными в России инструментами искусственного интеллекта для поиска фразеологизмов с колоронимом *green* на английском языке в корпусе NOW.

Колоронимы являются универсальными языковыми единицами, отражающими восприятие окружающего мира человеком, и, более того, влияющими «на изменение восприятия мира и различных человеческих ценностей» [4, с. 15]. Колороним зеленого цвета в современное время привлекает значительное внимание лингвистов. Как отмечает Мишель Пастуро, зеленый цвет, «некогда презираемый, отвергнутый, нелюбимый, сегодня превратился в мессианский цвет» [5, с. 139]. Обладая амбивалентной семантикой, он вызывает как положительные, так и отрицательные ассоциации. К отрицательным чаще всего относят зависть, скупость, ревность, к положительным – свежесть, юность, гармонию, естественность, симпатию и доверие. Интерпретируя смысл колоронимов, «человек опирается на прошлый опыт, составляющий часть индивидуального и коллективного знания» [6, с. 395], представленного, в частности, в фразеологических выражениях, или идиомах.

Как уже было сказано выше, корпусный менеджер English Corpora не всегда работает корректно, поэтому поиск совокупности фразеологизмов по ключевому слову *green* проблематичен. В связи с этим на первом этапе для выделения идиом с колоронимом *green* была использована платформа Perplexity. Платформа предложила 16 идиом, из которых внимание авторов статьи привлекла идиома *green behind the ears*.

На следующем этапе сама идиома была верифицирована через корпус NOW и найдено 47 случаев ее употребления с 2010 года (начало работы корпуса) по 2025 год. Приведем некоторые примеры:

*He could see we were a bit **green behind the ears** so he realised we needed help* (2013).

*I was fairly new in the job, and about as **green behind the ears** as it gets* (2018).

*I was a young researcher back then, still **green behind the ears*** (2025).

Данная идиома представляет собой смешанную идиому (idiom blend), которая могла возникнуть либо как слияние семантически близких *greenhorn* и *be wet behind the ears*, либо оказалась в английском языке как калька от немецкого фразеологизма *grün hinter den Ohren*. Знание схожих фразеологических оборотов и контексты, найденные в корпусе, делает значение очевидным: молодой, неопытный, желторотый, еще зеленый.

Сайт English-Corpora.org обещает внедрить ИИ в работу сайта уже в августе 2025 года [7], однако на момент написания статьи функции анализа коллокаций с помощью нейросетей не было. Авторы статьи приняли решение проверить, могут ли доступные в России нейросети (Mistral AI, Deepseek AI, Perplexity и GigaChat) найти контексты использования определенного выражения за конкретный временной период и провести дальнейший анализ. Сначала был задан следующий промпт: *найди контексты использования идиомы green behind the ears за 2025 год*. Так, удалось выяснить, что у моделей ограничена база знаний: у Mistral – до 1 ноября 2024 года, у Deepseek – до июля 2024 года: *Поиск примеров именно за текущий год требует анализа огромного массива данных, генерируемых ежесекундно, что недоступно в моей текущей модели* [8, 9]. Запрос был изменен на 2023 год, который входил во временной период базы данных, указанный моделями, но контексты употребления получить не удалось.

GigaChat дал на первый взгляд полный ответ с источниками, однако при проверке 10 источников ни один не был настоящим: одни страницы не открывались, на других отсутствовал цитируемый текст [10].

Необходимо отметить положительную сторону использования нейросетей в данном исследовании. Все ИИ платформы дали отсылку к самому известному случаю публичного использования этой идиомы: речь Барака Обамы в 2008 году, когда он сказал, что его оппонент Джон Маккейн считает Обаму молодым и неопытным политиком: *Now, Senator McCain suggests that somehow, you know, I'm **green behind the ears** and, you know, I'm just spouting off, and he's somber and responsible* [11].

Благодаря проверке частотности лексических единиц с помощью Google Books Ngram Viewer, удалось выяснить, что идиома *green behind the ears* приведена в книге «The Synonym Finder» как синоним для *unaccustomed* в значении *not used to, not ready for, not up to; uninitiated, unpracticed, inexperienced; green, raw, wet behind the ears* уже в 1978 году [12, p. 1271].

Интересно, что корпусный поиск по отдельным идиомам с колоронимом *green* наряду с множественными употреблениями идиом *green around the gills*

и *green with envy* предложил единичное употребление окказионального фразеологического бленда: *green around the gills envy*. В статье жанра *Agony Aunt* американского онлайн-издания *Slate Magazine – Politics, Business, Technology and the Arts* 2025 года приводится отрывок из письма, адресованного лицу, представленного псевдонимом *Prudence* (рус. благоразумие, осторожность, расчетливость): *What do you do with **green around the gills envy**? My partner inherited a five-bedroom house, and we plan to renovate it with grand dreams* [13]. При проверке данной идиомы в *Google Books Ngram Viewer* результаты получены не были.

На запрос с анализом данной идиомы нейросети дали очевидный ответ: идиомы *green around the gills* и *green with envy* имеют разные значения и контексты употребления. Их объединяет только слово *green*, но семантически они не связаны [8, 9, 10]. Однако, интересно отметить, что *Perplexity*, дав подробный разбор идиом, из которых состоит бленд, указал на возможность использования такого фразеологизма: *Совместное употребление (редко, но возможно). Иногда можно встретить оба выражения в одном тексте, когда описывают человека, который выглядит плохо из-за зависти или сильных эмоций: He was **green around the gills with envy** when he saw his rival's promotion (Он выглядел болезненно и бледно от зависти, когда увидел повышение своего соперника)* [14]. На дополнительный запрос о конкретных примерах в литературе или фильмах *Perplexity* дал еще несколько примеров, однако указал, что они «стилизированные», объяснив это так: *Реальных цитат с одновременным употреблением идиомы **green around the gills** и слова **envy** в классической литературе или известных произведениях найти очень сложно, так как эти выражения обычно используются отдельно и в разных контекстах: **Green around the gills** – идиома, связанная с физическим состоянием (тошнота, бледность); **Envy** – эмоциональное состояние, чаще описываемое через метафоры (**green-eyed monster**) или отдельные выражения* [14]. В указанном выше примере из американского онлайн-издания происходит обыгрывание значения колоронима *green*: колороним *green* в составе идиомы *green around the gills* служит стилистическим фразовым эпитетом, находящимся в акцентированной препозиции к определяемому слову *envy*, и выполняет функцию передачи эмоционального состояния сильной зависти, приводящей человека к болезненному недомоганию. Интерпретация смысла употребления идиомы с колоронимом в таком ключе согласуется с восприятием зависти в западноевропейской и американской культуре как ассоциируемой с зеленым цветом, в отличие от русской культуры, в которой зависть может характеризоваться как белая или черная.

Таким образом, следует отметить, что применение корпусного анализа и инструментов ИИ, взаимодополняющих друг друга при рассмотрении фразеологизмов, в нашем случае с колоронимом *green*, значительно расширяет экспериментальные возможности и представляется перспективным в дальнейших лингвистических исследованиях.

ЛИТЕРАТУРА

1. NOW Corpus (News on the Web) : официальный сайт. URL: <https://www.english-corpora.org/now/> (дата обращения: 10.07.2025).
2. Овсянникова М. А., Николаева М. Н. Нейросети как инструмент анализа текста: сравнительный аспект // Язык в эпоху цифровых трансформаций и развития искусственного интеллекта : сборник научных статей по итогам международной научной конференции, Минск, 23–24 октября 2024 года. Минск : Минский государственный лингвистический университет, 2024. С. 570–575.
3. Петрова И. М., Иванова А. М., Никитина В. В. Современные цифровые технологии в лингвистических исследованиях : Учебное пособие для обучающихся по направлению «Лингвистика»; Московский городской педагогический университет, Институт иностранных языков. Москва : Языки Народов Мира, 2022. 259 с.
4. Слестникова Т. В., Черкашина Е. И. Цвет и цветообозначение в лингвистических исследованиях. Москва : Языки народов мира, 2021. 240 с.
5. Пастуро М. Зеленый. История цвета; пер. с фр. Н. Кулиш. 4 изд. Москва : Новое литературное обозрение, 2024. 168 с.
6. Николаева М. Н. Концептуализация колоронимов синего цвета в итальянском языке // Когнитивные исследования языка. 2025. № 1–1(62). С. 394–398.
7. English Corpora : официальный сайт. URL: <https://www.english-corpora.org> (дата обращения: 10.07.2025).
8. Deepseek : официальный сайт. URL: <https://chat.deepseek.com> (дата обращения: 10.07.2025).
9. Mistral AI : официальный сайт. URL: <https://chat.mistral.ai> (дата обращения: 10.07.2025).
10. GigaChat : официальный сайт. URL: <https://giga.chat> (дата обращения: 10.07.2025).
11. The Second Presidential Debate: Obama vs. McCain. URL: <https://archive.nytimes.com/www.nytimes.com/elections/2008/president/debates/transcripts/second-presidential-debate.html> (дата обращения: 10.07.2025).
12. Urdang, Laurence., LaRoche, Nancy. The Synonym Finder. Rodale Books, 1978. 1361 p.
13. Help! My Fiancé Took a DNA Test. URL: <https://slate.com/advice/2025/06/family-advice-fiance-wedding-dna-test.html> (дата обращения: 10.07.2025).
14. Perplexity : официальный сайт. URL: <https://www.perplexity.ai> (дата обращения: 10.07.2025).