



Технология определения цветовой характеристики текста художественного произведения (на материале немецкого языка)

А. И. Горожанов¹, В. А. Шевцова²

¹Московский государственный лингвистический университет, Москва, Россия, agorozhanov@linguanet.ru

²Минский государственный лингвистический университет, Минск, Республика Беларусь, schewzowa_w@mail.ru

Аннотация. В статье предлагается технология определения цветовой характеристики произведения художественной литературы на основе корпусного подхода, предусматривающего применение автоматизированного анализа. В качестве материала исследования привлекаются оригинальные тексты романов Франца Кафки «Замок», «Процесс», «Америка». При их анализе применяется вышеозначенная технология; она дает точное представление об общем «цвете» текста при условии достаточной полноты банка цветowych данных конкретного произведения.

Ключевые слова: цветовая характеристика, корпусная лингвистика, сбалансированный корпус, немецкий язык, художественная литература, Франц Кафка, Python

Для цитирования: Горожанов А. И., Шевцова В. А. Технология определения цветовой характеристики текста художественного произведения (на материале немецкого языка) // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2023. Вып. 6 (874), С. 63–68. DOI 10.52070/2542-2197_2023_6_874_63

Original article

Technology for Determining the Colour Characteristics of the Text of a Fiction Work (on the Material of the German Language)

Alexey I. Gorozhanov¹, Valentina A. Shevtsova²

¹Moscow State Linguistic University, Moscow, Russia, agorozhanov@linguanet.ru,

²Minsk State Linguistic University, Minsk, Republic of Belarus, schewzowa_w@mail.ru

Abstract. A technology for determining the colour characteristics of a fiction work based on the corpus approach, which involves the use of automatized analysis, is described. The original texts of Franz Kafka's novels "The Castle", "The Trial" and "America" are used as research material. It is concluded that the shown technology gives an accurate idea of the general "colour" of the text, if the colour data bank for a particular fiction work is sufficiently complete.

Keywords: colour characteristics, corpus linguistics, balanced corpus, German language, fiction, Franz Kafka, Python

For citation: Gorozhanov, A. I., Shevtsova, V. A. (2023). Technology for Determining the Colour Characteristics of the Text of a Fiction Work (on the Material of the German Language). Vestnik of Moscow State Linguistic University. Humanities, 6(874), 63–68. 10.52070/2542-2197_2023_6_874_63

ВВЕДЕНИЕ

В природе цвет – это данность, независимая от людей, однако, в разных языках и лингвокультурах представление цвета происходит своеобразно. Наименования красок составляют сложную систему, которая обнаруживает показательные расхождения как на уровне литературного языка, так и на уровне идиостиля писателя. При этом цветообозначения рассматриваются как своеобразный аспект видения «картины мира» писателя, поэта, отраженной в его конкретном произведении через восприятие этого мира автором. Еще в древности Платон и Плутарх использовали представление о цвете для характеристики душевного состояния человека, его голоса и других проявлений внутренних свойств. «Это было своеобразным переводом информации с одного языка в другую семиотическую систему, непонятного через понятное» [Шевцова, 2004, с. 87].

В настоящее время проблема определения цветовой (в некоторых работах – колористической) характеристики текста является многоплановой и изучается с различных позиций. Достаточно широко представлены исследования, в которых проводится сопоставительный анализ восприятия тех или иных цветов в различных лингвокультурах [Дудникова, Медведева, 2022; Ким, Шаповалова, Минеев-Ли, 2022; Левицкий, Никульшина, 2022].

Другим направлением является определение функции цветообозначений в художественных текстах [Божедонова, 2022; Солопова, 2022; Мавсевич, Захаров, 2021, Голованова, 2020]. Не менее актуальными являются исследования цветообозначений в текстах литературных произведений методами компьютерной лингвистики. Так, в частности благодаря данной методике стало известно, что достаточно популярным для русских классиков считается черный цвет.

В сфере переводоведения ведутся дискуссии о том, какие переводческие решения подходят для правильной передачи терминов цветообозначения [Тимко, 2020; Егорова, Сизова, 2018]. В этой же связи Я. Е. Матросова выделяет «коэффициент плотности терминов цветообозначения», который определяется с помощью авторской программы для ЭВМ [Матросова, 2021], а также консолидирует усилия исследователей в области морфологической классификации цветообозначений [Матросова, 2016].

Авторы настоящего исследования ставят перед собой двоякую цель. *Во-первых*, получить общую цветовую характеристику художественного текста с наименьшими трудозатратами. *Во-вторых*, рассчитать некую «среднюю» цветовую характеристику, которую можно было бы определить как «цвет художественного произведения». Указанное

выше целеполагание позволит построить технологию определения цветовой характеристики художественного текста как совокупность процессов обработки текстовых данных. Таким образом, в центре нашего исследования стоит технологическая составляющая, тогда как языковой материал используется в иллюстративном качестве.

Для достижения поставленных задач привлекаются программные средства, главным из которых является язык программирования Python совместно с библиотекой `sraCy` [Горожанов, Гусейнова, Степанова, 2022]. Какие-либо сторонние программы для ЭВМ нами не используются. Языковой материал исследования представлен аутентичными текстами романов Франца Кафки «Замок», «Процесс» и «Америка».

ПОДГОТОВКА ТЕКСТА И ГЕНЕРАЦИЯ ЛИНГВИСТИЧЕСКОГО КОРПУСА

В качестве первого из романов для цветового анализа был выбран «Замок». Начальной стадией анализа указанного романа как аутентичного текста явилась его нормализация, т. е. превращение в такой формат, который был бы удобен для автоматической обработки специализированным программным обеспечением. В нашем случае этот процесс сводится к тому, что содержимое файла в формате TXT освобождается от всех данных, не имеющих прямого отношения к содержанию художественного произведения (иллюстрации, обложка, библиографическое описание, номера страниц и т. п.). Далее извлекаются все заголовки частей, а абзацные переносы заменяются одним пробелом. Все эти трансформации возможно осуществить с минимальным привлечением ручного труда.

Полученный нормализованный файл проходит автоматическую обработку с помощью авторской программы для ЭВМ – *генератора корпуса*, которая написана на Python с применением `sraCy`. В качестве средства построения графического интерфейса пользователя служит библиотека PyQt (рис. 1):

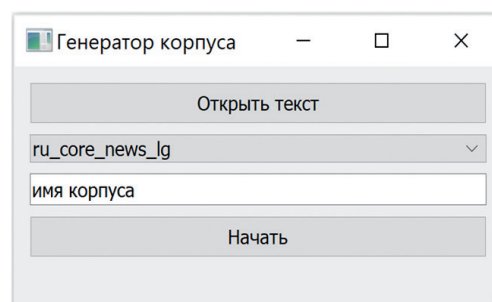


Рис. 1. Графический интерфейс генератора корпуса

Суть этой небольшой программы (ок. 190 строк основного кода и ок. 70 строк графического интерфейса пользователя) состоит в том, что она создает пустую базу данных SQLite и заполняет ее по особым правилам, заданным текстом и метаданными об этом тексте [Горожанов, 2022]. Тем самым создается сбалансированный корпус одного художественного произведения, т. е. такой корпус, который дает представление не о языке в целом, а о его частичной реализации – о романе.

Текст хранится в базе данных в виде пронумерованных предложений (для таблицы предложений) и пронумерованных токенов (для таблицы токенов). Метаданными о тексте являются:

- начальная / инфинитивная форма токена (лемма);
- часть речи, к которой принадлежит токен;
- морфологические признаки токена (свой набор для каждой части речи);
- соотношенность токена с его предложением (с идентификатором в таблице предложений).

Перед генерацией корпуса пользователь выбирает исходный файл, наименование файла корпуса и язык текста (на данный момент предусмотрены русский, английский и немецкий языки).

Примечательно, что не только токенизация (разбиение на токены) и разбиение на предложения, но также и получение всех метаданных происходит в полностью автоматическом режиме, благодаря возможностям библиотеки spaCy.

ОПРЕДЕЛЕНИЕ ЦВЕТОВЫХ ХАРАКТЕРИСТИК РОМАНОВ

На предыдущем этапе мы получили лингвистический корпус романа «Замок», который насчитывает 5404 предложения и 131943 токена. Далее программными средствами нами был получен частотный список лемм прилагательных, имеющих в романе, а из него вручную выбраны цветообозначения.

Оптимизации времени поиска послужило то, что необходимо было сделать ручную выборку не из всего объема словоупотреблений (131943)

и не из общего списка лемм (7666), а только из лемм-прилагательных (1190). Таким образом, объем поиска был сокращен более чем в 110 раз.

В результате был получен следующий список, который был сразу расширен некоторыми распространенными, но пока не зафиксированными в исследуемых текстах прилагательными: schwarz, blau, hellblau, dunkelblau, braun, hellbraun, dunkelbraun, grün, hellgrün, dunkelgrün, orange, lila, rot, hellrot, dunkelrot, rosa, gelb, hellgelb, dunkelgelb, grau, hellgrau, dunkelgrau, gold, silbern, golden, beige, weiß, cremefarben, goldfarbig, gelblich.

Этот список был взят за основу как некий *банк цветowych данных*, с помощью которого должен был быть осуществлен поиск эксплицитных цветowych характеристик в тексте.

Список был использован для поиска по леммам-прилагательным, который позволил посчитать количество употреблений всех возможных словоформ (например, «grau», «grauen», «grauem» для «grau»), а также для компонентного поиска по формулам «*lemma» и «lemma*», который позволил выявить сложные слова с участием компонентов из списка (например, «blauäugig» или «grauschmutzig»).

В результате в романе были выявлены 36 эксплицитных словоупотреблений цветowych характеристик (превалировали *серый* и *голубой*), каждая из которых была представлена в виде трехкомпонентной записи по правилам палитры RGB. Простым подсчетом среднего арифметического по каждому из трех компонентов был определен «средний» или «общий» цвет романа: (156, 107, 130), что примерно соответствует цвету под кодовым наименованием «Penelope Pink».

Сходным образом были получены цветowych характеристики романа «Процесс». Самым часто употребляемым цветом оказался черный, а «общим» (142, 128, 93) – цвет, наиболее близкий к цвету под кодовым наименованием «Wood Chi».

При анализе романа «Америка» на лидирующих позициях выступили белый и черный цвета, а «средним» результатом оказался (162, 108, 93), или «Honey Maple» (табл. 1):

Таблица 1

ОБРАЗЦЫ ПОЛУЧЕННЫХ ЦВЕТОВ



АНАЛИЗ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Сопоставим полученные результаты с «мысленной» интерпретацией романов. В «Замке» действие происходит зимой, следовательно, может показаться, что главным цветом должен быть белый или серый, что не сочетается с полученным «Penelope Pink», который можно трактовать как темно-розовый или грязно-розовый.

Однако такой цвет очень часто используется для колеровки оштукатуренных каменных зданий, например, строений, которые могут составлять комплекс замка. В этой связи «общий» цвет приобретает вполне понятный смысл – инстинктивно главный герой стремится в замок, постоянно представляет его себе как цель.

Цвет «Процесса» близок к грязно-зеленовато-коричневому. Если читатель романа представит себе исторический центр Праги, который незначительно изменился со времен Ф. Кафки, то такой цвет будет доминировать как очень близкий к цвету камня мостовых и ограждений набережной, некоторых зданий – особенно в пасмурную погоду и в сумерки.

Цвет «Америки» можно трактовать как близкий к коричневому; как и в предыдущем случае, он настраивает на атмосферу города, «каменных джунглей» мегаполисов Америки времени повествования. Превалирующий черный дает ассоциацию с многочисленными автомобилями.

В смысловых параметрах приведенной выше интерпретации цвета у Кафки невозможно не упомянуть и о некоторых ограничениях, которые накладываются на полученный результат. Прежде всего, мы провели анализ только эксплицитных цветовых характеристик, тогда как на восприятие читателя могут оказывать влияние и имплицитные. Например, лемма «Schnee» (снег), которая употребляется в романе «Замок» 35 раз, может приводить «средний» цвет произведения к более светлому тону, равно как и лексические единицы с семой «деревянный» внесут скорее серые и коричневые оттенки.

Далее для использования предлагаемой технологии на материале других литературных текстов необходимо значительно расширить банк цветовых данных, в который могут войти, например, фитонимы (белочник, белошка, белужник, белуха, белушка и т. д.) [Кузнецова, 2020] или цвета пищевых продуктов (абрикосовый, апельсиновый, арбузный, баклажанный, брусничный и т. д.) [Юрина, Боровкова, 2014].

При генерации корпуса и распределении токенов по частям речи, spaCy может допускать небольшую погрешность, которая, впрочем, редко превышает несколько процентов.

ЗАКЛЮЧЕНИЕ

Нами была предложена и апробирована технология определения цветовых характеристик художественного текста, которая заключается в последовательности автоматизированных процессов за счет применения специализированного программного обеспечения, разработанного в лаборатории фундаментальных и прикладных проблем виртуального образования Московского государственного лингвистического университета.

Практическая ценность работы заключается в том, что предложенная технология может быть использована, например, при оформлении художественных произведений, а также при их экранизации, например, выбора цвета фасадов зданий, одежды персонажей, мебели, светофильтров и пр.

Обозначим два направления дальнейшего исследования.

Во-первых, совершенствование банка цветовых данных, а также его расширение на другие языки. *Во-вторых*, использование получаемых цветовых характеристик для более глубокой интерпретации художественного произведения с привлечением методов и знаний смежных с лингвистикой научных дисциплин.

СПИСОК ИСТОЧНИКОВ

1. Шевцова В. А. Символика цвета в русской фразеологии // Русский язык и литература. 2004. № 6 (57). С. 87–93.
2. Дудникова Л. В., Медведева М. С. К вопросу о системе цветообозначений в русском, английском и французском языках (на материале национального корпуса русского языка) // Гуманитарные и социальные науки. 2022. Т. 90. № 1. С. 56–61. DOI 10.18522/2070-1403-2022-90-1-56-61.
3. Ким А. А., Шаповалова Ю. Ю., Минеев-Ли В. Е. Лингвокультурологические особенности символики некоторых цветообозначений у британцев и славян // Вестник Томского государственного педагогического университета. 2022. № 3 (221). С. 34–45. DOI 10.23951/1609-624X-2022-3-34-45.
4. Левицкий А. Э., Никульшина Т. Н. Чёрный в русском и английском сказочных дискурсах (проблемы семантики и лингвокультурологии) // Когнитивные исследования языка. 2022. № 2 (49). С. 699–704.
5. Божедонова А. Е. Функциональная характеристика цветообозначения мастей лошади в эпических текстах в сравнительном аспекте // Восток. Афро-Азиатские общества: история и современность. 2022. № 1. С. 230–244. DOI 10.31857/S086919080016064-7.

6. Солопова А. А. Художественные функции цвета при описании превращений в «Метаморфозах» Овидия // Индоевропейское языкознание и классическая филология. 2022. № 26–2. С. 1098–1111. DOI 10.30842/ielcp230690152671.
7. Масевич А. Ц., Захаров В. П. Диахронические изменения частотности цветообозначений в русских поэтических текстах // Вестник Новосибирского государственного университета. Серия: Лингвистика и межкультурная коммуникация. 2021. Т. 19. № 4. С. 5–35. DOI 10.25205/1818-7935-2021-19-4-5-35.
8. Голованова Н. Ю. Репрезентанты цветовой палитры в повести Л. Н. Андреева «В тумане» // Актуальные вопросы современной филологии и журналистики. 2020. № 4 (39). С. 101–107. DOI 10.36622/AQMPJ.2020.39.4.014.
9. Тимко Н. В. Специфика перевода цветообозначений в политическом тексте (на материале заголовков статей медиаресурса BBC) // Известия Волгоградского государственного педагогического университета. 2020. № 1 (144). С. 225–229.
10. Егорова О. А., Сизова В. В. Цветовая полифония в поэзии В. Высоцкого (в сравнении с переводами на английский язык) // Вестник Тверского государственного университета. Серия: Филология. 2018. № 2. С. 220–224.
11. Матросова Я. Е. Каждый охотник желает автоматизировать вычисление коэффициента плотности терминов цветообозначения в тексте // Мосты. 2021. № 3 (71). С. 21–29.
12. Матросова Я. Е. Морфологические классификации русских цветообозначений, представленных в языке художественной литературы // Вестник Нижегородского государственного лингвистического университета им. Н. А. Добролюбова. 2016. № 36. С. 61–74.
13. Горожанов А. И., Гусейнова И. А., Степанова Д. В. Инструментарий автоматизированного анализа перевода художественного произведения // Вопросы прикладной лингвистики. 2022. № 45. С. 62–89. DOI 10.25076/vpl.45.03.
14. Горожанов А. И. Экспериментальное моделирование базы данных сбалансированного лингвистического корпуса // Филологические науки. Вопросы теории и практики. 2022. Т. 15. № 10. С. 3382–3386. DOI 10.30853/phil20220563.
15. Кузнецова Е. Л. Закономерности отражения колористической характеристики во внутренней форме русских диалектных номинаций лекарственных трав // Журнал Белорусского государственного университета. Филология. 2020. № 3. С. 50–57.
16. Юрина Е. А., Боровкова А. В. Способы толкования образных цветообозначений, мотивированных наименованиями пищевых продуктов в «Словаре русской кулинарной метафоры» // Вопросы лексикографии. 2014. № 2 (6). С. 106–121.

REFERENCES

1. Shevtsova, V. A. (2004). Simvolika tsveta v russkoi frazeologii = Symbolism of colour in Russian phraseology. *Russkij jazyk i literatura*, 6(57), 87–93. (In Russ.)
2. Dudnikova, L. V., Medvedeva, M. S. (2020). To the question about the system of colour terms in Russian, English and French (on the material of the Russian national corpus). *The Humanities and Social sciences*, 90(1), 56–61. DOI 10.18522/2070-1403-2022-90-1-56-61. (In Russ.)
3. Kim, A. A., Shapovalova, Yu. Yu., Mineev-Li, V. E. (2022). Linguocultural features of the symbolism of some color designations among the British and Slavs. *Tomsk State Pedagogical University Bulletin*, 3(221), 34–45. DOI 10.23951/1609-624X-2022-3-34-45. (In Russ.)
4. Levitskiy, A. E., Nikulshina, T. N. (2022). Black in Russian and English fairy tale discourse: semantic and linguocultural perspectives. *Kognitivnye issledovaniya jazyka*, 2(49), 699–704. (In Russ.)
5. Bozhedonova, A. E. (2022). Functional characteristic of the designation of the equine coat color in epic texts in a comparative aspect. *Vostok (Oriens) societies: history and modernity*, 1, 230–244. DOI 10.31857/S086919080016064-7. (In Russ.)
6. Solopova, A. A. (2022). Hudozhestvennye funktsii cveta pri opisani prevrashhenij v «Metamorfozah» Ovidija = Functions of colour terms in Ovid's Metamorphoseon. *Indo-European Linguistics and Classical Philology*, 26(2), 1098–1111. DOI 10.30842/ielcp230690152671. (In Russ.)
7. Masevich, A. Ts., Zakharov, V. P. (2021). Diahronicheskie izmeneniya chastotnosti cvetooboznachenij v russkikh poeticheskikh tekstah = Diachronic changes in the frequency of colour designation in Russian poetic texts. *Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Lingvistika i mezhkul'turnaja kommunikacija*, 19(4), 5–35. DOI 10.25205/1818-7935-2021-19-4-5-35. (In Russ.)
8. Golovanova, N. Yu. (2020). Reprezentanty cvetovoj palitry v povesti L. N. Andreeva «V tumane» = The representatives of colour palette in the novel "In the fog" by L. N. Andreev. *Aktual'nye voprosy sovremennoj filologii i zhurnalistiki*, 4(39), 101–107. DOI 10.36622/AQMPJ.2020.39.4.014. (In Russ.)
9. Timko, N. V. (2020). Specifika perevoda cvetooboznachenij v politicheskom tekste (na materiale zagolovkov statej mediareursa BBC) = The specifics of the translation of color names in political texts (based on the articles' headlines of the media resource BBC). *Izvestia of the Volgograd State Pedagogical University*, 1(144), 225–229. (In Russ.)

10. Egorova, O.A., Sizova, V.V. (2018). Cvetovaja polifonija v poezii V. Vysockogo (v sravnenii s perevodami na anglijskij jazyk) = Color polyphony in poetry of V. Vysotsky and its translations into English. Vestnik TvGU serija Filologija, 2, 220–224. (In Russ.)
11. Matrosova, Ya. E. (2021). Kazhdyi okhotnik zhelaet avtomatizirovat' vychislenie koeffitsienta plotnosti terminov tsветоoboznacheniya v tekste = Every hunter wants to automate the calculation of the density coefficient of colour terms in the text. Mosty. Zhurnal perevodchikov, 3(71), 21–29. (In Russ.)
12. Matrosova, Ya. E. (2016). Morphological Classifications of Colour Terms in Russian Literature. Nizhny Novgorod Linguistics University Bulletin, 36, 61–74. (In Russ.)
13. Gorozhanov, A. I., Guseynova, I. A., Stepanova, D. V. (2022). Instrumentarij avtomatizirovannogo analiza perevoda hudozhestvennogo proizvedeniya = Tools for automated analysis of fiction work translation. Issues of applied linguistics, 45, 62–89. DOI 10.25076/vpl.45.03. (In Russ.)
14. Gorozhanov, A. I. (2022). Eksperimental'noe modelirovanie bazy dannykh sbalansirovannogo lingvisticheskogo korpusa = Experimental database modelling of a balanced linguistic corpus. Philology. Theory&Practice, 15(10), 3382–3386. DOI 10.30853/phil20220563. (In Russ.)
15. Kuznetsova, E. L. (2020). Regularities of reflecting color characteristic in an inner form of Russian dialectal nominations of medicinal herbs. Journal of the Belarusian State University. Philology, 3, 50–57. (In Russ.)
16. Yurina, E. A., Borovkova, A. V. (2014). The ways of interpreting figurative color terms motivated by food names in the dictionary of Russian culinary metaphors. Russian journal of lexicography, 2(6), 106–121. (In Russ.)

ИНФОРМАЦИЯ ОБ АВТОРАХ

Горожанов Алексей Иванович

доктор филологических наук, доцент
профессор кафедры грамматики и истории немецкого языка
факультета немецкого языка
Московского государственного лингвистического университета

Шевцова Валентина Анатольевна

кандидат филологических наук, доцент,
декан факультета немецкого языка
Минского государственного лингвистического университета

INFORMATION ABOUT THE AUTHORS

Gorozhanov Alexey Ivanovich

Doctor of Philology, Associate Professor
Professor in the Department of German Language Grammar and History
Faculty for German Language
Moscow State Linguistic University

Shevtsova Valentina Anatolievna

PhD (Philology), Associate Professor
Dean of the Faculty for German Language
Minsk State Linguistic University

Статья поступила в редакцию
одобрена после рецензирования
принята к публикации

14.03.2023
18.03.2023
20.04.2023

The article was submitted
approved after reviewing
accepted for publication