

RU

Формальная модель оценки образа персонажа художественного произведения (на материале романа Дж. Оруэлла «1984»)

Горожанов А. И., Красикова Е. А.

Аннотация. Цель предлагаемого исследования – разработать и апробировать формальную модель оценки образа персонажа произведения художественной литературы на основе корпусного подхода. Научная новизна заключается в том, что впервые такого рода анализ героя произведения художественной литературы проводится на основе корпусного подхода. В итоге был построен сбалансированный корпус оригинального текста романа Дж. Оруэлла «1984», включающий разметку иерархических связей между токенами. Осуществлен ряд поисковых запросов к корпусу с помощью авторского программного обеспечения «Генератор сбалансированного лингвистического корпуса и корпусный менеджер», в результате которых был получен частотный список имен собственных романа и выведены списки токенов, характеризующих его главных действующих лиц. Полученные данные извлекаются в ходе работы встроенного в корпусный менеджер специального алгоритма, который выбирает из текста значимые для характеристики персонажа и связанные с ним лексические единицы. Тем самым строится модель оценки характерных черт персонажа как формальное описание системы связанных между собой иерархическими связями элементов текста. Результаты исследования были проанализированы и сопоставлены с результатами мысленной интерпретации. В итоге были составлены характеристики персонажей с точки зрения формального описания их действий и тех объектов, на которые были направлены их действия.

EN

Formal model for evaluating the image of a character in a fictional work (based on G. Orwell's novel "1984")

A. I. Gorozhanov, E. A. Krasikova

Abstract. The aim of the research is to develop and test a formal model for evaluating the image of a character in a fictional work based on the corpus approach. The scientific originality lies in the fact that for the first time such an analysis of the character of a fictional work is carried out on the basis of the corpus approach. In the course of the work, a balanced corpus of the original text of the novel "1984" by G. Orwell was formed, which includes the marking of hierarchical relationships between the tokens. A number of search queries to the corpus were carried out using the author's software "Balanced Linguistic Corpus Generator and Corpus Manager". With the help of this tool, a frequency list of proper names from the novel was obtained and lists of tokens characterizing its main characters were output. The data obtained is extracted during the operation of a special algorithm built into the corpus manager, which selects from the text significant character traits and related lexical units. Thus, a model for evaluating the image of a character is built as a formal description of a system of interconnected hierarchical text elements. The results of the study were analyzed and compared with the results of mental interpretation. Thus, the characters' profiles were compiled in terms of a formal description of their actions and those objects to which their actions were directed.

Введение

Процесс интерпретации произведений художественной литературы традиционно носит субъективный характер, поскольку неизбежно преломляется через жизненный опыт читателя, его мировосприятие и тот набор ожиданий, который складывается у него еще до прочтения текста.

Персонажи художественных произведений также представляются читателям по-разному, о чем свидетельствуют часто диаметрально противоположные образы одних и тех же героев в экранизациях различных режиссеров и воплощенные различными киноактерами. В нашей работе мы предпринимаем шаги к составлению

объективного «портрета» персонажа художественного произведения, опираясь исключительно на лингвистические данные, выбрав в качестве материала оригинальный текст романа Дж. Оруэлла «1984».

Актуальность настоящего исследования подтверждается значительным количеством публикаций российских и зарубежных ученых в высокорейтинговых научных изданиях по исследуемому вопросу, которые можно сгруппировать следующим образом.

Во-первых, произведение Дж. Оруэлла «1984» анализировалось ранее в различных лингвистических аспектах: в данном романе-антиутопии исследовались коммуникативные стратегии формирования имиджа (Гусейнова, Горожанов, 2023), лингвостилистическое описание окказионализмов (Темкина, Сафина, 2021), семантические характеристики образных метафор (Дедова, 2018), выявлялись лингвокультурологические особенности романа (Долженко, 2014) и др.

Во-вторых, актуальной для современных исследователей является проблема интерпретации текста, в частности художественного произведения. Одним из популярных направлений в анализе текста является когнитивный подход. Например, при помощи такого подхода стало изучаться концептуальное основание текста сквозь призму бинарных концептуальных оппозиций, которые лежат в основе фреймов и определяют сюжет речевого произведения (Беляевская, 2023). Художественный текст как объект исследования рассматривался в прагматическом аспекте (Suleymanova, Utkirova, 2022), проводился анализ стратегий конструирования коннотированного образа в художественно-публицистическом жанре (Гусейнова, Горожанов, 2023), на примере романа Дж. Оруэлла «1984» выявлялась взаимосвязь между крупнейшими историческими событиями XX века и литературным творчеством (Литвяк, Каменчук, 2021), анализировались особенности языковой игры (Игнатенко, 2022). Упомянем также текстологический анализ для конкретных целей, а именно – для улучшения понимания при синхронном переводе (Garzone, 2000), интерпретацию текста в лингводидактических целях (Deane, 2020).

В-третьих, отметим исследования, затрагивающие проблему идиостиля (Бойчук, Воронцова, Шляхтина и др., 2019; Evans, 2023). Так, в этой области рассматриваются особенности функционирования языковой личности автора в художественном тексте (Катермина, 2016), исследуется идиостиль Ф. М. Достоевского (Баранов, Добровольский, Фатеева, 2021), анализу подвергается кинематографичность идиостиля Ю. Н. Тынянова (Размашкин, 2001), рассматривается индивидуальный авторский стиль Юй Хуа на материале романа «Братья» (кит. «兄弟») (Новиков, 2024), анализируются особенности функционирования понятия «новояз» (созданное Дж. Оруэллом) в политическом дискурсе как одного из средств манипулирования (Шестерина, 2023), изучаются лингвокогнитивные механизмы индивидуально-авторского стиля (Яновская, Чижикова, Золотых, 2019) и др.

Наконец, неоспоримый интерес вызывают исследования в области корпусной лингвистики с привлечением технологий обработки естественного языка (англ. natural language processing – NLP). Упомянем ряд научных трудов, посвященных данной тематике: анализ семантически связанных слов в медицинском корпусе на материале арабского языка (Almahmoud, Hammo, 2024), поиск фразеологизмов в лингвистическом корпусе на материале индоевропейских языков (Gorozhanov, Guseynova, Stepanova, 2024), поиск вербальных маркеров идентичности в научном дискурсе (Гончарова, Заврумов, Халеева, 2024), анализ текстов немецкоязычных СМИ (Степанова, 2023) и т. д.

Отдельно обратим внимание на то, что ранее нами были предприняты некоторые попытки провести комплексное исследование, направленное на интерпретацию художественного текста, которое бы совмещало не только «традиционную» мысленную интерпретацию, но и «точный» корпусный подход (Горожанов, Степанова, 2022).

Сказанное выше позволяет заключить, что заявленная тема исследования рассматривается в разных аспектах и является востребованной в рамках современной научной парадигмы.

Мы ставим перед собой следующие задачи:

- 1) составить специальный сбалансированный корпус романа Дж. Оруэлла «1984», предусматривающий разметку иерархических связей токенов в предложении;
- 2) осуществить ряд поисковых запросов к ключевым персонажам романа, используя специально разработанный алгоритм анализа окружения заданной единицы корпуса;
- 3) провести анализ полученных результатов, сопоставив их с оценкой в ходе мысленной интерпретации.

При решении первой задачи используется метод генерации сбалансированного лингвистического корпуса по правилам библиотеки spaCy (Gorozhanov, Guseynova, Stepanova, 2024). Вторая задача решается с применением метода моделирования, поскольку алгоритм поиска функционирует согласно модели, которую мы определяем как *формальное описание системы связанных между собой иерархическими связями элементов текста*. Третья задача требует привлечения сопоставительного анализа и метода интерпретации.

Уточним, что «образ персонажа» мы трактуем в самом широком смысле как совокупность лексических единиц (глаголов, существительных, прилагательных и наречий), формально связанных с именем собственным – наименованием персонажа.

Практическая ценность работы заключается в том, что ее результаты могут быть использованы для объективного комплексного анализа художественных произведений с целью лучшего понимания интенции автора и более глубокого изучения механизмов построения образа персонажей.

Обсуждение и результаты

Для составления сбалансированного корпуса романа Дж. Оруэлла «1984» была задействована специальная функция программы-генератора упомянутого выше комплекса «Генератор сбалансированного лингвистического корпуса и корпусный менеджер» (Свидетельство Роспатент № 2023683209), разработанного в лаборатории

фундаментальных и прикладных проблем виртуального образования Московского государственного лингвистического университета.

При этом создаваемая база данных получает дополнительное поле, в которое помещается номер главного токена для текущего токена. Эти данные формируются с помощью внутренних инструментов sраСу, но тем не менее правила выстраивания подобной иерархии достаточно понятны.

Приведем несколько примеров. В Таблице 1 дано первое предложение романа «1984».

Таблица 1. Разметка первого предложения романа «1984»

id	№ токена	№ предложения	токен	часть речи	лемма	№ главного токена
1	0	1	It	PRON	it	1
2	1	1	was	AUX	be	1
3	2	1	a	DET	a	5
4	3	1	bright	ADJ	bright	5
5	4	1	cold	ADJ	cold	5
6	5	1	day	NOUN	day	1
7	6	1	in	ADP	in	5
8	7	1	April	PROPN	April	6
9	8	1	,	PUNCT	,	1
10	9	1	and	CCONJ	and	1
11	10	1	the	DET	the	11
12	11	1	clocks	NOUN	clock	13
13	12	1	were	AUX	be	13
14	13	1	striking	VERB	strike	1
15	14	1	thirteen	NUM	thirteen	13
16	15	1	.	PUNCT	.	13

Наименования частей речи указаны по международной классификации, принятой в библиотеках обработки естественного языка, и отличаются от традиционных наименований, принятых в языкознании. Главным словом в предложении является изменяемая часть сказуемого (№ 1), ему непосредственно подчиняются подлежащее (№ 2), именная часть сказуемого (№ 5), запятая (№ 8), сочинительный союз (№ 9) и глагол во второй части предложения (№ 13). Артикль и прилагательные в общем случае всегда подчиняются главному существительному (например, № 5), предлоги являются главными словами для существительных (например, № 6).

Рассмотрим еще один пример, в котором встречается имя собственное (см. Таблицу 2).

Таблица 2. Разметка второго предложения романа «1984»

id	№ токена	№ предложения	токен	часть речи	лемма	№ главного токена
17	16	2	Winston	PROPN	Winston	17
18	17	2	Smith	PROPN	Smith	21
19	18	2	,	PUNCT	,	17
20	19	2	his	PRON	his	20
21	20	2	chin	NOUN	chin	21
22	21	2	nuzzled	VERB	nuzzle	21
23	22	2	into	ADP	into	21
24	23	2	his	PRON	his	24
25	24	2	breast	NOUN	breast	22
26	25	2	in	ADP	in	21
27	26	2	an	DET	an	27
28	27	2	effort	NOUN	effort	25
29	28	2	to	PART	to	29
30	29	2	escape	VERB	escape	27
31	30	2	the	DET	the	32
32	31	2	vile	ADJ	vile	32
33	32	2	wind	NOUN	wind	29
34	33	2	,	PUNCT	,	21
35	34	2	slipped	VERB	slip	21
36	35	2	quickly	ADV	quickly	34
37	36	2	through	ADP	through	34
38	37	2	the	DET	the	39
39	38	2	glass	NOUN	glass	39
40	39	2	doors	NOUN	door	36
41	40	2	of	ADP	of	39
42	41	2	Victory	PROPN	Victory	42
43	42	2	Mansions	PROPN	Mansions	40
44	43	2	,	PUNCT	,	34
45	44	2	though	SCONJ	though	47
46	45	2	not	PART	not	46

id	№ токена	№ предложения	токен	часть речи	лемма	№ главного токена
47	46	2	quickly	ADV	quickly	47
48	47	2	enough	ADJ	enough	34
49	48	2	to	PART	to	49
50	49	2	prevent	VERB	prevent	47
51	50	2	a	DET	a	51
52	51	2	swirl	NOUN	swirl	49
53	52	2	of	ADP	of	51
54	53	2	gritty	ADJ	gritty	54
55	54	2	dust	NOUN	dust	52
56	55	2	from	ADP	from	49
57	56	2	entering	VERB	enter	55
58	57	2	along	ADP	along	56
59	58	2	with	ADP	with	56
60	59	2	him	PRON	he	58
61	60	2	.	PUNCT	.	21

Здесь важно заметить, что имя и фамилия главного персонажа являются двумя отдельными токенами, причем токен со старшим номером (№ 17) является главным для токена с младшим номером (№ 16). Внутри вводной конструкции (№№ 19-32) главным словом является глагол (№ 21). Собственно сказуемое предложения (№ 34) подчиняется глаголу вводной конструкции (т. е. зависимости строится линейно). От него в прямой зависимости находится токен № 47, которому далее подчиняется инфинитивная конструкция (№№ 48-59).

Предложение № 2 является показательным, так как демонстрирует сложные случаи выстраивания иерархии в предложении. В нем впервые упоминается персонаж Winston Smith, образ которого, по нашему мнению, формируют токены №№ 20, 21, 34, 35, 38, 39, 41, 42.

На этом этапе исследования необходимо было уточнить некоторые основополагающие моменты, связанные с «объективным» и «субъективным» в процессе построения образа персонажей литературного произведения, что является важным для нашей работы, поскольку мы стремимся выделить именно «объективное».

Во-первых, автор (рассказчик) строит художественную реальность произведения, включая описание и поведение персонажей, руководствуясь своим творческим замыслом и принимая самостоятельно решения о развитии сюжетной линии, что мы будем считать «объективным». Хотя, справедливости ради, вспомним цитату А. С. Пушкина, передаваемую Л. Н. Толстым: «Представь, какую штуку удрала со мной Татьяна! Она – замуж вышла. Этого я никак не ожидал от нее» (Л. Н. Толстой в воспоминаниях современников: в 2 т. / подгот. текста и коммент. Н. Н. Гусева, В. С. Мишина; предисл. К. Н. Ломунова. М., 1955. Т. 1, с. 231-232). Во-вторых, персонаж может что-то рассказывать о себе, что мы примем как «субъективное». В-третьих, о персонаже что-то говорят другие персонажи, что также является, по нашему мнению, «субъективным».

Таким образом, разрабатывая формальную модель оценки образа персонажа, мы будем придерживаться слов автора (рассказчика), по возможности исключая из поля зрения реплики персонажей о себе, в то время как упоминания персонажа в речи других действующих лиц полностью исключить из формальной модели на данном этапе исследования еще не представляется возможным.

Мы решили первую задачу исследования, выполнив разметку текста романа и дополнительно «расшифровать» правила построения иерархических связей встроенными инструментами библиотеки spaCy. Объем корпуса составил 7108 предложений или 121 437 токенов.

Решение второй, основной, задачи исследования потребовало не только создания алгоритма поиска токенов в корпусе, но, прежде всего, описания формальной модели, согласно которой этот поиск будет осуществляться.

Исходя из сформулированного выше тезиса мы сфокусируемся на описании действий персонажа автором (рассказчиком). Отправной точкой был выбран токен – наименование персонажа. Сбор данных происходит по очереди отдельно в каждом предложении, в котором встретится указанный оператор исходный токен (например, Winston или Julia). Далее реализуется поиск всех зависимых от исходного токена слов, если они являются существительными, глаголами, прилагательными или наречиями. У каждого экземпляра исходного токена есть свое одно главное слово (как правило, глагол); номера последних фиксируются алгоритмом, и далее работа производится с ними. При этом ограничения по частям речи не устанавливаются. На следующей ступени поиска программа отбирает все зависимые от главного слова исходного токена лексические единицы. Если таковыми являются глаголы, то алгоритм повторяется для каждого такого случая, чтобы не пропустить предложения с однородными сказуемыми или цепочкой глагольных форм, связанных с исходным токеном (как в предложении Таблицы 2).

Разработанную формальную модель можно представить следующим образом (стрелки направлены от главного слова к зависимым словам) (см. Схему 1).

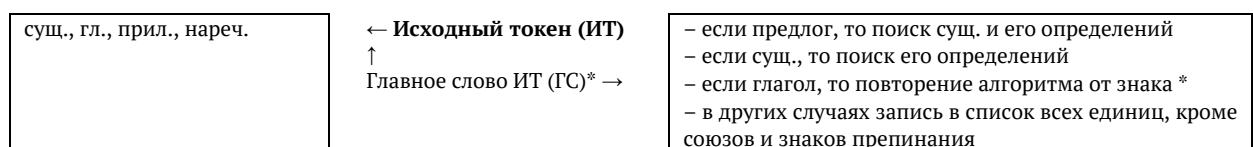


Схема 1. Упрощенное изображение формальной модели

Затем был сформирован частотный список имен собственных романа (54 единицы), в котором оказались не только персонажи, но и географические наименования, а также несколько ошибочно взятых токенов, образующих небольшую допустимую погрешность. Приведем ниже некоторые пункты этого списка, где первое число – это порядковый номер в частотном списке, а второе – количество упоминаний в корпусе:

1. Winston : 455
3. Julia : 89
5. O'Brien : 83
20. Goldstein : 29
42. Charrington : 13

Из пяти выбранных персонажей первый является главным действующим лицом, второй является протагонистом, третий и пятый скорее антагонисты, а четвертый упоминается, но не фигурирует в романе как активное действующее лицо.

Для каждого персонажа был выведен частотный список согласно формальной модели. Исключим из полученных результатов служебные слова и токены, встречающиеся менее трех раз (если список объемлен):

Исходный токен – Winston (получены 782 уникальные единицы):

9. think (думать) : 21
11. know (знать) : 16
17. moment (момент) : 12
18. remember (вспоминать) : 12
19. time (время) : 11
20. look (смотреть) : 11
21. sit (сидеть) : 10
23. back (назад) : 9
26. face (лицо) : 9
36. speak (говорить) : 7
39. finger (палец) : 7
41. voice (голос) : 6
43. work (работа) : 6
44. room (комната) : 6
45. suddenly (внезапно) : 6
46. head (голова) : 6
50. aware (~осознавать) : 6
51. glance (взгляд) : 6
54. way (путь) : 5
56. never (никогда) : 5
58. find (находить) : 5
59. follow (следовать) : 5
60. hand (рука) : 5
61. feeling (чувство) : 5
65. watch (наблюдать) : 5
66. believe (верить) : 5
68. notice (замечать) : 5
69. table (стол) : 5
70. hear (слушать) : 5
71. Julia (имя собственное) : 5
72. eye (глаз) : 5
75. telescreen (телескрин) : 4
76. place (место) : 4
77. small (маленький) : 4
78. raise (поднимать) : 4
79. strike (ударять) : 4
80. wake (будить) : 4
81. long (долгий) : 4
82. side (сторона) : 4
85. Syme (имя собственное) : 4
86. tray (поднос) : 4
88. catch (хватать) : 4
89. hold (держат) : 4
91. gaze (пристально глядеть) : 4
92. hour (час) : 4
93. lie (лгать) : 4
97. control (контролировать) : 4
98. tell (рассказать) : 4

- 99. word (слово) : 3
- 103. write (писать) : 3
- 105. first (первый) : 3
- 106. effort (усилие) : 3
- 109. examine (оценивать) : 3
- 110. good (хороший) : 3
- 112. mean (иметь значение) : 3
- 113. mother (мать) : 3
- 114. attention (внимание) : 3
- 115. well (хорошо) : 3
- 116. arm (рука) : 3
- 117. dial (набирать номер) : 3
- 118. claim (заявлять) : 3
- 119. read (читать) : 3
- 120. wonder (чудо) : 3
- 121. promptly (немедленно) : 3
- 123. vaporize (испарить) : 3
- 124. remark (заметить) : 3
- 125. seem (казаться) : 3
- 126. feel (чувствовать) : 3
- 128. carry (нести) : 3
- 129. vaguely (неясно) : 3
- 130. pay (платить) : 3
- 131. leave (покидать) : 3
- 132. constantly (постоянно) : 3
- 133. heart (сердце) : 3
- 134. murmur (шептать) : 3
- 135. intermittently (прерывисто) : 3
- 136. part (часть) : 3
- 138. slow (медленный) : 3
- 139. ear (ухо) : 3
- 140. little (маленький) : 3
- 141. sleep (спать) : 3
- 142. forget (забывать) : 3
- 143. listen (слушать) : 3
- 144. glass (стекло) : 3

Исходный токен – Julia (получена 181 уникальная единица):

- 7. Winston (имя собственное) : 5
- 8. seem (казаться) : 4
- 12. side (сторона) : 4
- 16. room (комната) : 3
- 17. wake (будить) : 3
- 20. betray (обманывать) : 3
- 26. good (хороший) : 2
- 27. tell (рассказать) : 2
- 28. dead (мертвый) : 2
- 31. sleep (спать) : 2
- 33. indifferently (равнодушно) : 2
- 34. lie (лгать) : 2
- 36. hand (рука) : 2
- 37. point (точка) : 2
- 38. way (путь) : 2
- 40. show (показывать) : 2
- 41. sound (звук) : 2
- 44. love (любовь) : 2

Исходный токен – O'Brien (получено 249 уникальных единиц):

- 2. say (сказать) : 28
- 6. look (смотреть) : 9
- 8. hand (рука) : 8
- 9. stand (стоять) : 7
- 14. seem (казаться) : 5

15. smile (улыбаться) : 5
20. face (лицо) : 4
21. lever (рычаг) : 4
23. side (сторона) : 3
25. Winston (имя собственное) : 3
26. nod (кивать) : 3
28. hold (держаться) : 3
30. know (знать) : 3
32. cage (клетка) : 3
34. speak (говорить) : 2
35. friend (друг) : 2
37. meet (встречать) : 2
39. intently (намеренно) : 2
40. gesture (жест) : 2
42. little (мало) : 2
43. air (воздух) : 2
44. push (толкать) : 2
47. watch (наблюдать) : 2
49. small (маленький) : 2
50. white (белый) : 2
53. direct (прямой) : 2
57. remember (вспоминать) : 2
58. lie (лгать) : 2
59. control (контролировать) : 2
60. head (голова) : 2
61. lay (лежать) : 2
62. kindly (любезно) : 2
66. answer (отвечать) : 2
68. mean (иметь в виду) : 2
69. shoulder (плечо) : 2
70. thing (вещь) : 2
71. bad (плохой) : 2
72. happen (случаться) : 2
73. rat (крыса) : 2
74. pick (собирать) : 2

Исходный токен – Goldstein (получено 111 уникальных единиц):

4. Emmanuel (имя собственное) : 3
12. Enemy (Враг) : 1
13. renegade (рenegат) : 1
15. engage (вовлекать) : 1
16. activity (деятельность) : 1
17. counter (счетчик) : 1
19. revolutionary (революционно) : 1
20. condemn (проклясть) : 1
21. death (смерть) : 1
22. escape (побег) : 1
23. figure (персона) : 1
24. principal (главный) : 1
26. attack (атака) : 1
28. venomous (ядовитый) : 1
29. plausible (правдоподобный) : 1
30. hate (ненавидеть) : 1
31. despise (презирать) : 1
32. author (автор) : 1
33. circulate (распространять) : 1
35. clandestinely (тайно) : 1
40. invincible (непобедимый) : 1
41. fearless (бесстрашный) : 1
48. denounce (осуждать; доносить) : 1
68. trial (суд) : 1
87. bomb (бомба) : 1

Исходный токен – Charrington (получено 39 уникальных единиц):

1. Mr (мистер) : 11
6. old (старый) : 2
8. week (неделя) : 1
9. gather (собирать) : 1
12. inscription (надпись) : 1
15. widower (вдовец) : 1
16. inhabit (обитать) : 1
17. shop (магазин) : 1
18. year (год) : 1
21. foresee (предвидеть) : 1
22. difficulty (трудность) : 1
23. teach (учить) : 1
24. think (думать) : 1
25. Winston (имя собственное) : 1
26. animal (животное) : 1
27. extinct (потухший) : 1
32. room (комната) : 1
35. jacket (пиджак) : 1
39. Julia (имя собственное) : 1

Если предварительно оценить полученный результат только с точки зрения количественных показателей, то бросается в глаза отношение количества уникальных единиц к общему количеству полученных токенов – списки в основном заполнены единичными употреблениями, – что говорит, кроме прочего, о разнообразии языковых средств, используемых Дж. Оруэллом. Что касается технического исполнения алгоритма, то процедуры сработали штатно, не выявив сбоев программного обеспечения.

В рамках решения третьей задачи исследования проведем анализ полученных результатов, исходя из наполнения списков и владея, с точки зрения читателя, содержанием романа, т. е. совмещая автоматический анализ с мысленной интерпретацией.

Изучение списков для всех персонажей, на первый взгляд, не дает четкого описания образа действующего лица. Тем не менее самым показательным является список для действующего лица Goldstein, в котором преобладают отрицательные характеристики. Отсюда мы можем заключить, что писатель намеренно очернил этого персонажа, создавая языковыми средствами гротескный образ абсолютного зла.

Главный персонаж Winston, в отличие от остальных, отобранных для нашего анализа, «читает» и «пишет», что является его ключевым отличием от остальных героев романа. Это заключение стало возможным именно после изучения списков, составленных согласно формальной модели, тогда как после простого прочтения романа с позиции читателя этот аспект образа персонажа не выделяется настолько очевидно.

Удивительным можно считать еще и тот факт, что персонаж Julia, который является очень ярким при «мысленном» анализе, уступает в количестве связанных элементов персонажу O'Brien, хотя последний является представителем негативных сил согласно сюжету книги. Это наталкивает на мысль о том, что фокус повествования направлен на взаимодействие главного героя (Winston) и представителя системы (O'Brien), в то время как Julia является скорее не действующим субъектом, но средством, позволяющим глубже показать взаимодействие оси "Winston – O'Brien".

Что касается персонажа Charrington, то элементы модели описывают его достаточно ясно: мужчина, стар, вдовец, имеет отношение к магазину, связан с персонажами Winston и Julia, что полностью соответствует действительности.

Заключение

Итак, мы можем заключить, что поставленные в исследовании задачи были выполнены. Во-первых, был изготовлен сбалансированный корпус текста романа Дж. Оруэлла «1984», учитывающий разметку иерархических связей токенов в предложении.

Во-вторых, была разработана и визуализирована формальная модель оценки образа персонажа, которая, главным образом, характеризует выбранных персонажей с позиции автора (рассказчика). Другими словами, в рамках предложенной модели производится сбор данных о том, *какой* персонаж, *что* он делает, на *какие* объекты художественной реальности направлены его действия и *каковы* их свойства. Согласно модели, был запрограммирован специальный алгоритм, который позволил осуществить ряд запросов к сбалансированному корпусу, результатом работы которого явились частотные списки токенов (лемм), связанных с исходными словами поиска. Ограничения модели сводятся к тому, что она не учитывает упоминание о персонаже через референты (личные местоимения и имена нарицательные) и реплики самого персонажа (прямая речь).

В-третьих, был произведен анализ полученных автоматическим путем данных, которые подверглись сопоставлению с данными «мысленного» анализа с позиции восприятия содержания романа читателем. Результаты позволили выявить некоторые моменты, не очевидные при простом прочтении, в том числе степень важности персонажей и отличительные особенности главного героя от других действующих лиц.

В перспективе необходимо установить величину погрешности при составлении характеристики персонажа, апробировать полученную модель на материале художественных произведений на других языках.

В целом работа по данной теме позволила нам прийти к следующим общим умозаключениям. При интерпретации (анализе) текста с помощью технологий искусственного интеллекта, к которым также относятся технологии обработки естественного языка, рискованно опираться на предварительно «натренированные» языковые модели, тем более если характер этих моделей не до конца ясен или не ясен вовсе. В этом случае получаемая погрешность становится неизвестной и переменной величиной. Заметим, что здесь мы понимаем под «моделью» сокращение от англоязычного технического термина «большая языковая модель» (англ. large language model).

Обучая, или «тренируя», языковые модели на конкретных, пусть и больших по объему, текстах (т. е. индуктивно), гарантировать высокое качество интерпретации (анализа) можно только для самих этих текстов и текстов, в некоторой степени сходных с ними. Универсальный характер и высокая точность интерпретации (анализа) возможны, по нашему мнению, только при помощи моделей, построенных на правилах языка как системы (т. е. дедуктивно).

В настоящей статье была описана попытка именно такого подхода. Заметим тут же, что мы ни в коем случае не отказываемся от преимуществ современных «натренированных» языковых моделей, появление которых стало возможным благодаря изобретению и распространению мощных и быстрых ЭВМ, оперирующих огромными объемами данных. До этого, на заре компьютерной лингвистики, исследователям приходилось исходить только из структуры языка в силу того, что компьютеры, несмотря на их внушительные размеры, работали чрезвычайно медленно – см., например, Джоржтаунский эксперимент (Hutchins, 1997). По всей видимости, будущее за гибридным подходом, совмещающим в себе оба указанных способа.

Источники | References

1. Баранов А. Н., Добровольский Д. О., Фатеева Н. А. Идиостиль Ф. М. Достоевского: направления изучения // Вестник Российского университета дружбы народов. Серия «Теория языка. Семиотика. Семантика». 2021. Т. 12. № 2. <https://doi.org/10.22363/2313-2299-2021-12-2-374-389>
2. Беляевская Е. Г. О бинарной структуре когнитивных оснований текстоформирования // Когнитивные исследования языка. 2023. № 2 (53).
3. Бойчук Е. И., Воронцова И. А., Шляхтина Е. В., Беляева О. В. Идиостиль и ритм текста: монография. Ярославль: Ярославский государственный педагогический университет им. К. Д. Ушинского, 2019.
4. Гончарова О. В., Заврумов З. А., Халева С. А. Алгоритмы поиска вербальных маркеров идентичности в современном научном дискурсе // Актуальные проблемы филологии и педагогической лингвистики. 2024. № 2. <https://doi.org/10.29025/2079-6021-2024-2-18-29>
5. Горожанов А. И., Степанова Д. В. Интерпретация художественного произведения: корпусный подход // Филологические науки. Вопросы теории и практики. 2022. Т. 15. Вып. 1. <https://doi.org/10.30853/phil20220020>
6. Гусейнова И. А., Горожанов А. И. Коммуникативные стратегии формирования имиджа в романе-антиутопии Дж. Оруэлла «1984» // От слова к дискурсу: взаимодействие форм и (не)предсказуемость смыслов: тезисы международной научной конференции (12-13 мая 2023 г.). Мн.: Минский государственный лингвистический университет, 2023.
7. Дедова Н. А. Сравнительный анализ семантических характеристик образных метафор текста-оригинала и текста-перевода романа Дж. Оруэлла «1984» // Известия Смоленского государственного университета. 2018. № 2 (42).
8. Долженко С. Г. Диалог культур в романах Олдоса Хаксли «О дивный новый мир» и Джорджа Оруэлла «1984» // Филологические науки. Вопросы теории и практики. 2014. № 8-1 (38).
9. Игнатенко А. В. Особенности языковой игры в прозе Лю Чжэньюня на примере романа «Я не Пань Цзиньянь» // Вестник Санкт-Петербургского университета. Востоковедение и африканистика. 2022. Т. 14. № 3. <https://doi.org/10.21638/spbu13.2022.308>
10. Катермина В. В. Языковая личность автора в художественном тексте // Человек. Культура. Образование. 2016. № 2 (20).
11. Литвяк О. В., Каменчук А. С. Тема вины в романе «Процесс» Франца Кафки и в романе «1984» Джорджа Оруэлла // Успехи гуманитарных наук. 2021. № 9.
12. Новиков Е. А. Идиостиль в творчестве Юй Хуа на примере романа «Братья» (2005) // Восточный калейдоскоп: материалы докладов и сообщений IV всероссийской научно-практической конференции молодых ученых-востоковедов (13 декабря 2023 г.). М.: РУДН им. Патриса Лумумбы, 2024.
13. Размашкин И. Ю. Кинематографичность идиостиля Ю. Н. Тынянова: композиционно-синтаксический аспект: автореф. дисс. ... к. филол. н. СПб., 2001.
14. Степанова Д. В. Программный комплекс для генерации динамического корпуса текстов СМИ // Вестник Минского государственного лингвистического университета. Серия «Филология». 2023. № 6 (127).
15. Темкина В. Л., Сафина А. С. Индивидуально-авторские новообразования в романе Джорджа Оруэлла «1984» // Современные исследования социальных проблем. 2021. Т. 13. № 1-2. <https://doi.org/10.12731/2077-1770-2021-13-1-2-213-218>

16. Шестерина Е. А. «Новояз» в современном немецком политическом дискурсе // Дискурс профессиональной коммуникации. 2023. Т. 5. № 3. <https://doi.org/10.24833/2687-0126-2023-5-3-93-106>
17. Яновская И. В., Чижикова О. В., Золотых Н. В. Лингвокогнитивные механизмы индивидуально-авторского начала в поэтическом дискурсе // Вестник Волгоградского государственного университета. Серия «Языкознание». 2019. Т. 18. № 3. <https://doi.org/10.15688/jvolsu2.2019.3.21>
18. Almahmoud R. H., Hammo B. H. SEWAR: A Corpus-based N-gram Approach for Extracting Semantically-related Words from Arabic Medical Corpus // Expert Systems with Applications. 2024. Vol. 238. <https://doi.org/10.1016/j.eswa.2023.121767>
19. Deane P. Building and Justifying Interpretations of Texts: A Key Practice in the English Language Arts // ETS Research Report Series. 2020. Vol. 2020. <https://doi.org/10.1002/ets2.12304>
20. Evans M. Interjections and Individual Style: A Study of Restoration Dramatic Language. Language and Literature. 2023. Vol. 32 (3). <https://doi.org/10.1177/09639470231158695>
21. Garzone G. Textual Analysis and Interpreting Research // The Interpreters' Newsletter. 2000. Vol. 10.
22. Gorozhanov A. I., Guseynova I. A., Stepanova D. V. Natural Language Processing and Fiction Text: Basis for Corpus Research // RUDN Journal of Language Studies, Semiotics and Semantics. 2024. Vol. 15. No. 1. <https://doi.org/10.22363/2313-2299-2024-15-1-195-210>
23. Hutchins J. From First Conception to First Demonstration: The Nascent Years of Machine Translation, 1947-1954. A Chronology // Machine Translation. 1997. Vol. 12. <https://doi.org/10.1023/A:1007969630568>
24. Suleymanova N. M., Utkirova Sh. U. Some Features of the Meaning "Literary Text" in the Pragmalinguistic Aspect // Ученый XXI века. 2022. № 10 (91).

Информация об авторах | Author information



Горожанов Алексей Иванович¹, д. филол. н., доц.

Красикова Елизавета Александровна², к. филол. н.

^{1, 2} Московский государственный лингвистический университет



Alexey Ivanovich Gorozhanov¹, Dr

Elizaveta Alexandrovna Krasikova², PhD

^{1, 2} Moscow State Linguistic University

¹ a_gorozhanov@mail.ru, ² krasikova.liza@mail.ru

Информация о статье | About this article

Дата поступления рукописи (received): 06.08.2024; опубликовано online (published online): 17.09.2024.

Ключевые слова (keywords): корпусная лингвистика; формальная модель; интерпретация художественного произведения; образ персонажа; роман Дж. Оруэлла «1984»; английский язык; corpus linguistics; formal model; interpretation of fictional works; character image; G. Orwell's novel "1984"; English language.