

Анатомия эссе, которые сработали: анализ 2 282 вступительных эссе

Artyom Konukhov

6 октября 2026 г.

Аннотация

Сильное вступительное эссе отличается не драматичностью темы, а исполнением: смелым первым предложением, сквозным образом и финалом, который возвращает к началу. К такому выводу ведёт анализ 2 282 эссе абитуриентов из 46 открытых источников. Это официальные подборки вузов (Johns Hopkins за 2013–2030 гг., Tufts, Connecticut College, Hamilton и другие), публикации СМИ, библиотеки консультантов и британские UCAS statements. Для 1 433 эссе издатель подтверждает поступление автора.

Каждый текст размечен языковой моделью по коdbуку из 17 полей: тема, структура, тип начала и финала, сюжетная арка, ставки, проявленные качества, тон, приёмы и три оценки качества от 1 до 5 (конкретика, рефлексия, голос). Разметку дополняют стилевые метрики, рассчитанные скриптом. Повторная слепая разметка 140 случайных эссе воспроизводит категории с $\kappa = 0,71–0,86$, а сумму оценок — с ICC = 0,87.

«Сильным» считалось эссе с суммой оценок не меньше 12 из 15; среди 799 основных эссе таких 22 %. Главные результаты:

- **Ставки.** Только 16 % подтверждённых основных эссе построены на тяжёлом опыте. Доля сильных текстов не зависит от ставок: 21–23 %.
- **Начало.** Эссе со смелым утверждением в начале сильны в 32 % случаев, со сценной — в 16 %, с прямым тезисом — в 9 %. Пара «смелое утверждение → возврат к началу» даёт 43 %.
- **Структура.** Структура вокруг одного предмета или метафоры даёт 32 % сильных, хронологический рассказ — 13 %.
- **Голос.** Юмор присутствует в 85 % эссе с высшей оценкой голоса.
- **Разборы.** В 131 разборе от приёмных комиссий чаще всего хвалят характер и ценности (51 %). В 882 разборах консультантов на первом месте конкретные детали (38 %).
- **UCAS. Statements,** где академика занимает больше 70 % текста, сильны почти вчетверо чаще: 35 % против 9 %.
- **Дополнительные эссе.** Рамка, метафора и монтаж дают 24–36 % сильных текстов, «карта ресурсов вуза» в «Why us» — 5 %.
- **Тренды.** В официальных подборках за 2007–2030 годы выпуска юмора стало вдвое меньше (52 → 24 %), эссе вокруг предмета — вдвое больше (15 → 30 %).

Ограничения работы — отсутствие отклонённых эссе и машинная разметка; поэтому результаты описывают признаки сильных эссе, а не причины поступления.

Abstract

We analyse 2,282 college admission essays from 46 public sources, 1,433 with publisher-confirmed admission. An LLM annotated each essay with a 17-field codebook covering topic, structure, opening, ending, arc, stakes, demonstrated qualities, tone, techniques and three 1–5 quality ratings; scripted stylometrics complement the labels. A blind second annotation pass over 140 random essays reproduced the categorical labels (Cohen's $\kappa = 0.71–0.86$) and the rating sum (ICC = 0.87). Strong essays (rating sum $\geq 12/15$; 22% of main essays)

are not distinguished by dramatic subject matter: hardship narratives account for 16% of verified main essays and are no more often strong. They differ in craft: bold-claim openings (32% strong vs. 16% for scene openings), essays organised around a single object or metaphor (32% vs. 13% for chronological narratives) and endings that return to the opening (43% when paired with a bold claim). Supplements follow the same pattern (framed, metaphor-led and montage essays: 24–36% strong; “why us” resource maps: 5%), as do UCAS statements, where those devoting over 70% of the text to the subject are strong almost four times as often (35% vs. 9%). In official collections, humour halved between the classes of 2007–2017 and 2023–2030 (52% → 24%). Admissions officers’ commentaries emphasise character and values (51%), consultants’ commentaries emphasise concrete detail (38%). Findings describe what successful essays look like, not what causes admission.

Ключевые слова: вступительные эссе, Common App, UCAS, контент-анализ, LLM-разметка, нарративная структура, стилометрия.

1. Введение

Вступительное эссе — единственная часть заявки, где абитуриент говорит своим голосом. Оно же окутано самыми противоречивыми советами. В США основное эссе Common App ограничено 650 словами с сезона 2013–2014; до этого ориентиром были 500 слов ([College Essay Advisors, 2013](#)). К нему добавляются десятки коротких дополнительных эссе (supplements). В Великобритании UCAS personal statement даёт 4 000 знаков, а с приёма 2026 года разделён на три вопроса, не менее 350 знаков на каждый ([UCAS, 6. г.](#)).

Абитуриенты учатся писать эссе в основном по примерам «essays that worked», которые публикуют вузы, СМИ и консультанты. Эти примеры разбросаны по сотням страниц, а выводы из них делаются по двум-трём ярким текстам. Научные работы про эссе, наоборот, смотрят на них как на предиктор успеваемости, дохода или личных качеств (раздел 2). Писательское ремесло там почти не измеряется: как устроен текст, чем он начинается и чем заканчивается.

Эта работа заполняет пробел. Мы собрали в один корпус 2 282 опубликованных эссе и разметили каждое по кодбуку, описывающему ремесло, а не демографию. На этой основе мы отвечаем на семь вопросов.

№	Исследовательский вопрос	Раздел
RQ1	О чём пишут поступившие и насколько драматичен их материал?	5
RQ2	Какие структуры, начала, финалы и арки чаще встречаются у высоко оценённых эссе?	6
RQ3	Чем стилистически отличаются эссе с ярким голосом?	7
RQ4	Что хвалят в разборах приёмные комиссии, СМИ и консультанты?	8
RQ5	Какие типы историй повторяются и согласуются ли они с темами, найденными без учителя?	9
RQ6	Чем дополнительные эссе и UCAS statements отличаются от основного эссе?	10
RQ7	Как менялись официально опубликованные эссе за 2007–2030 годы выпуска?	11

Вклад работы:

1. **Корпус.** Корпус из открытых источников с меткой того, подтверждает ли издатель поступление автора, и со слитыми перепечатками одних и тех же текстов.
2. **Кодбук.** Разметка писательского ремесла из 17 полей с валидатором и оценкой надёжности повторной разметки.
3. **Результаты.** Описательные и ассоциативные результаты с доверительными интервалами и регрессией с контрольными переменными.
4. **Практика.** Выводы для абитуриента, проверяемые на конкретных эссе из корпуса.

2. Обзор литературы

Количественные исследования вступительных эссе до сих пор спрашивали, что эссе говорит об авторе: его будущих оценках, доходе семьи, личных качествах. Как устроено само хорошее эссе, они почти не спрашивали.

2.1. Эссе как источник сигналов об абитуриенте

- **Служебные слова и успеваемость.** [Pennebaker et al. \(2014\)](#) проанализировали более 50 000 эссе 25 975 студентов крупного государственного университета. Индекс «категориального ↔ динамического» языка строится по восьми категориям служебных слов. Больше артиклей и предлогов связано с более высоким средним баллом все четыре года учёбы ($r \approx 0,18-0,20$), а больше местоимений и вспомогательных глаголов — с более низким.
- **Социальный класс.** [Alvero et al. \(2021\)](#) изучили 240 000 эссе 60 000 абитуриентов University of California (ноябрь 2016 года). Содержание описывалось тематическими моделями, стиль — словарём LIWC. Оба связаны с доходом семьи сильнее, чем балл SAT, и объясняют заметную долю его дисперсии. Иначе говоря, эссе несёт отпечаток класса не слабее теста.
- **Личные качества.** [Lira et al. \(2023\)](#) дообучили модели RoBERTa на 3 131 эссе, размеченном ассистентами и сотрудниками приёмных комиссий. Модели распознают семь качеств: просоциальную цель, лидерство, командность, стремление учиться, упорство, внутреннюю мотивацию и целеустремлённость. На национальной выборке из более чем 300 000 заявок эти оценки предсказывали выпуск за шесть лет сверх демографии и тестов.

Во всех трёх работах эссе — предиктор исхода или фона, а не объект писательского анализа. Ни одна не размечает структуру, начало, финал или сюжетную арку. И ни одна не работает с теми эссе, которые вузы сами публикуют как образцы.

2.2. Практическое знание о ремесле

Советы по письму эссе живут в пособиях и блогах консультантов. Самая влиятельная рамка принадлежит [Sawyer \(2016\)](#). Он делит эссе на «нарративные», где события связаны причиной и следствием, и «монтажные», где они нанизаны на общую тему. Расхожие правила — «начни со сцены», «показывай, а не рассказывай», «не пиши про трагедию ради трагедии» — повторяются во многих источниках корпуса, но на больших выборках не проверялись.

Особый источник — разборы «почему это эссе сработало». Их публикуют приёмные комиссии (например, Johns Hopkins делает это с выпуска 2013 года), СМИ и консультанты. Это единственный открытый след того, что читатели ценят в тексте. Мы анализируем их в разделе 8.

2.3. Разметка текстов языковыми моделями

[Gilardi et al. \(2023\)](#) показали, что языковая модель без дообучения точнее краудсорсинга в четырёх из пяти задач разметки. Согласие её повторных разметок при этом выше, чем у краудворкеров и обученных аннотаторов. Это делает LLM-разметку тысяч эссе реалистичной, но не отменяет проверку надёжности: повторяемость зависит от модели и формулировки задания ([Reiss, 2023](#)). Её мы измеряем согласием двух независимых проходов (раздел 4.5).

Для анализа мы опираемся на стандартные инструменты:

- интервал Уилсона для долей ([Wilson, 1927](#));
- каппа Коэна ([Cohen, 1960](#)) с шкалой интерпретации [Landis and Koch \(1977\)](#);
- неотрицательная матричная факторизация для поиска тем без учителя ([Lee and Seung, 1999](#));
- взвешенные логарифмы шансов для поиска характерных слов ([Monroe et al., 2008](#)).

Полные ссылки — в списке литературы.

3. Данные

Корпус состоит из 2 282 эссе общим объёмом около 1,06 млн слов. Все они взяты из 46 открытых источников и собраны 5 октября 2026 года. У 1 433 эссе издатель явно подтверждает поступление или оффер.

3.1. Источники и сбор

Тексты скачивались скриптами на Python (requests, BeautifulSoup) и сохранялись дословно, без пересказа. Страницы, удалённые с сайтов, восстанавливались из архива Wayback Machine; так получено 194 эссе, в том числе все эссе Johns Hopkins за 2013–2019 годы. Каждое эссе — отдельный файл с метаданными:

- автор в том виде, как его назвал издатель;
- вуз, куда поступил автор, и год выпуска;
- тип эссе и текст задания, если он опубликован;
- разбор издателя, если он есть;
- ссылка на исходную страницу.

3.2. Уровни верификации и типы эссе

Источники разнесены по тому, насколько надёжно подтверждено поступление автора.

- **Сайты вузов.** Эссе, опубликованные самим вузом.
- **СМИ и консультанты.** Издатель называет вуз, куда поступил автор.
- **UCAS.** Statements с указанным оффером.
- **«Поступление не указано».** Реальные эссе, чей исход издатель не сообщает. В основном это College Essay Guy, CollegeVine и AdmitReport.

Отдельно выделены три типа текста.

- **Основное эссе.** Common App и аналоги.
- **Дополнительное эссе.** Supplements, UC Personal Insight Questions, короткие ответы. Тип определялся по полю «тип эссе» у издателя; если поле не заполнено, по правилу: меньше 250 слов или структура «почему этот вуз».
- **UCAS personal statement.**

Несколько издателей (CollegeVine, CollegeAdvisor, PrepMaven и другие) попадают сразу в два уровня, если часть их эссе подтверждена, а часть нет.

Крупнейшие источники в каждом уровне:

- **Сайты вузов:** Tufts (116), Johns Hopkins (102), Connecticut College (49), St. John's College (41), Hamilton (39).
- **СМИ и консультанты:** CollegeAdvisor (205), EssaysThatWorked.com (199), CollegeVine (80), The Harvard Crimson (78).

Таблица 1. Состав корпуса (число эссе)

Уровень верификации	Источников	Основное	Дополнительное	UCAS	Всего
Сайты вузов	12	272	126	—	398
СМИ и консультанты (поступление указано)	21	220	483	—	703
Великобритания (оффер указан)	11	—	—	332	332
Поступление не указано	7	307	542	—	849
Итого	46	799	1 151	332	2 282

- **Великобритания:** PSE (107), The Uni Guide (84), College Lah (39), UniAdmissions (30), bmed (27), St Anne's College Oxford (18).
- **Исход не указан:** College Essay Guy (550), CollegeVine (227), AdmitReport (59).

3.3. Дедупликация и очистка

Одно и то же эссе часто перепечатают несколько издателей. Дубликаты ищались по пересечению наборов из 8 слов подряд (shingles): два текста считались копиями, если не менее 60 % наборов короткого из них встречалось в длинном. Нашлось 74 такие группы с 79 лишними копиями. При слиянии в корпусе оставалась одна версия, выбранная по трём правилам по порядку:

1. полный текст предпочитался отрывку;
2. сайт вуза — всем остальным источникам;
3. более длинная версия — более короткой.

Разборы из удалённых копий сохранены внутри оставшегося файла.

Ручная очистка:

- разделены три архивные страницы Tufts, где слиплись эссе двух авторов;
- склеено одно эссе CollegeVine, разрезанное скрейпером;
- исправлено 16 склеенных слов вокруг курсива в 11 эссе.

Исключены:

- 26 отрывков Oxbridge Mentors: третий ответ скрыт за формой с email, вуз не назван, тексты шаблонны;
- одно эссе College Essay Guy с пометкой автора сайта «не для публичного использования».

3.4. Описательная статистика

Тип	Медиана, слов	Межквартильный размах	Минимум–максимум
Основное эссе	652	616–664	250–1 865
Дополнительное	255	197–352	28–2 113
UCAS statement	644	624–666	435–753

- **Разборы и задания.** Разбор издателя «почему это сработало» есть у 1 120 эссе, текст задания — у 835, только отрывок опубликован у 55.
- **Год выпуска.** Известен для 414 эссе и охватывает 2007–2030 годы; пик приходится на выпуски 2018–2021.

- **Длина основного эссе.** Половина текстов укладывается в 616–664 слова, то есть авторы практически всегда выбирают лимит в 650 слов. Старые эссе, написанные до жёсткого лимита, доходят до 1 865 слов.

4. Метод

Каждое эссе прочитано целиком и размечено языковой моделью по единому кодбуку. Анализ сочетает описательные доли с доверительными интервалами, тесты ассоциации и регрессию с контрольными переменными.

4.1. Кодбук

Разметка описывает ремесло, а не автора. Полные определения — в приложении [A](#).

Группа полей	Поля (число значений)	Формат
Содержание	тема (26), вторичные темы (0–2), теги	одно значение + список
Построение	структура (11), начало (9), финал (9), арка (10)	одно значение
Масштаб	ставки (3), временной охват (3), доля рефлексии (3)	одно значение
Автор и голос	качества (16, от 2 до 3), тон (10, от 1 до 2), приёмы (17, от 0 до 4)	список
Качество	конкретика, рефлексия, голос	целое 1–5 с якорями
Разбор издателя	что хвалит разбор (20 критериев)	список, только при наличии разбора
UCAS	доля академики (3), типы доказательств (9)	только для британских statements
Текст для читателя	суть, пересказ, главная мысль, сила, слабое место	свободный текст на русском

У трёх оценок качества есть описанные якоря. Например, для голоса: 1 — «обезличенно», 3 — «есть личность», 5 — «незабываемый, узнаваемый голос». Инструкция просит не завышать оценки только потому, что эссе опубликовано, и оставлять 5 для исключительных текстов.

4.2. Процедура разметки

1. **Инструкция.** Одна для всех: определения всех значений и два полностью разобранных примера — американское эссе и UCAS statement.
2. **Пилот.** Первые 70 эссе размечены и проверены вручную до запуска остальных.
3. **Основной прогон.** Корпус перемешан случайно и разбит на 33 пачки по 70 эссе, чтобы каждый аннотатор видел смесь источников. Пачки размечали независимые экземпляры Claude Sonnet, до 20 параллельно. Каждое эссе читалось целиком вместе с заданием и разбором издателя.
4. **Валидатор.** Скрипт проверял допустимые значения, длины полей, наличие русского текста и полноту пачки. Пачка принималась только при нуле ошибок.
5. **Переразметка.** Пять эссе, чей текст менялся при очистке (раздел [3.3](#)), размечены заново.

4.3. Стилиевые метрики

Независимо от модели скрипт считал по каждому эссе:

- длину в словах и абзацах;
- среднюю длину фразы и её разброс;
- долю коротких (до 7 слов) и длинных (от 30 слов) фраз;
- долю слов в кавычках (диалог);
- частоту «I / me / my» и «you» на 100 слов;
- вопросы и числа;
- индексы читаемости Flesch и Flesch–Kincaid;
- лексическое разнообразие MATTR с окном 100 слов;
- наличие 20 частых клише (регулярные выражения, список — в приложении А).

4.4. Исход и выборки

Сильное эссе — сумма трёх оценок не меньше 12 из 15. Таких 175 из 799 основных эссе, или 21,9 % (95 % ДИ 19,2–24,9). Для дополнительных эссе порог снижен до 11, для UCAS — до 10: у коротких и формальных жанров ниже базовый уровень оценок.

Анализ идёт на двух выборках:

- **Распределения** («о чём пишут поступившие») считаются по 492 подтверждённым основным эссе.
- **Связь с качеством** — по всем 799 основным эссе. Уровень верификации входит в регрессию как контрольная переменная.

4.5. Надёжность разметки

Повторная слепая разметка 140 случайных эссе почти полностью воспроизводит первую. Категории совпадают в 83–94 % случаев (κ Коэна 0,71–0,86), оценки — в 84 % (взвешенный κ 0,81–0,86), сумма оценок воспроизводится с ICC = 0,87. Это проверка воспроизводимости, а не правильности. Она показывает, что метки не шум модели, но не говорит, совпадает ли взгляд модели со взглядом приёмной комиссии.

- **Выборка.** Простая случайная выборка 140 эссе из корпуса (генератор с зерном 2026): 44 основных, 64 дополнительных и 32 UCAS.
- **Второй проход.** Два новых экземпляра Claude Sonnet получили по 70 эссе в том же виде и ту же инструкцию, что в основном прогоне, но не видели первой разметки. Журналы вызовов подтверждают, что агенты открывали только инструкцию и свои пачки.
- **Независимость.** Ни одна из 140 «судей» не совпала дословно; медианное сходство строк — 0,51. Это две отдельные разметки, а не копия.
- **Меры.** κ Коэна для номинальных полей; квадратично взвешенный κ для порядковых полей и оценок; ICC(A,1) для суммы оценок (McGraw and Wong, 1996); коэффициент Жаккара для полей с несколькими метками. Доверительные интервалы — бутстреп по эссе, 2 000 повторов. Согласие читается по шкале Landis and Koch (1977): 0,61–0,80 — существенное, выше 0,80 — почти полное.

Где проходы расходятся. Почти всегда между соседними значениями.

- **Структура.** 10 из 19 расхождений касаются структуры вокруг предмета, которую второй проход читал как рассказ в рамке или хронологию, и наоборот. Отдельно κ этой категории — 0,67 (0,44–0,84), у рассказа в рамке — 0,61. Это самые размытые границы среди тех, на которых держатся выводы раздела 6.

Таблица 2. Согласие двух независимых проходов разметки (140 эссе)

Поле	Значений	Совпадение	Согласие (95 % ДИ)	Основные эссе (n = 44)
Тема	26	85,7 %	$\kappa = 0,84$ (0,77–0,90)	0,92
Структура	11	86,4 %	$\kappa = 0,84$ (0,77–0,91)	0,72
Начало	9	88,6 %	$\kappa = 0,86$ (0,80–0,92)	0,80
Финал	9	87,1 %	$\kappa = 0,84$ (0,77–0,91)	0,89
Сюжетная арка	10	84,3 %	$\kappa = 0,81$ (0,74–0,88)	0,87
Архетип (по правилам)	11	84,3 %	$\kappa = 0,80$ (0,73–0,87)	0,71
Ставки	3	94,3 %	$\kappa_w = 0,92$ (0,86–0,97)	0,90
Временной охват	3	91,4 %	$\kappa_w = 0,81$ (0,66–0,92)	0,79
Доля рефлексии	3	82,9 %	$\kappa_w = 0,80$ (0,71–0,88)	0,86
Конкретика	1–5	83,6 %	$\kappa_w = 0,81$ (0,72–0,88)	0,73
Рефлексия	1–5	83,6 %	$\kappa_w = 0,82$ (0,75–0,89)	0,76
Голос	1–5	84,3 %	$\kappa_w = 0,86$ (0,79–0,91)	0,83
Сумма оценок	3–15	91,4 % в пределах ± 1	ICC = 0,87 (0,81–0,91)	0,83
«Сильное эссе»	2	91,4 %	$\kappa = 0,70$ (0,53–0,84)	0,56

κ — каппа Коэна; κ_w — квадратично взвешенная каппа; ICC — внутриклассовая корреляция ICC(A,1). Последний столбец — та же мера на 44 основных эссе выборки. «Сильное эссе» определяется с порогами по жанрам (раздел 4.4). Согласие по каждой метке — в приложении Б.

- **Начало.** 9 из 16 расхождений касаются смелого утверждения. Второй проход ставит его чаще (32 эссе против 25), но 24 из 25 смелых начал первого прохода подтвердил; $\kappa = 0,80$.
- **Финал.** Чаще всего путаются «связь с вузом» и «взгляд в будущее». Возврат к началу воспроизводится с $\kappa = 0,85$, пересказ — с $\kappa = 0,82$.
- **Доля рефлексии.** Все 24 расхождения — на одну ступень: ни разу «низкая» против «высокой».
- **Оценки.** Ни одна оценка не разошлась больше чем на балл. Средние двух проходов одинаковы: конкретика 3,44 и 3,45, рефлексия 2,79 и 2,86, голос 2,86 и 2,86 (тест Уилкоксона, $p \geq 0,11$). Систематического сдвига нет.

Множественные метки. Средний коэффициент Жаккара: приёмы 0,80, качества 0,78, тон 0,75, вторичные темы 0,68; критерии похвалы в 65 эссе с разбором — 0,82. По отдельным меткам медианный κ — 0,84 для приёмов, 0,82 для качеств и 0,72 для тона. Самая шаткая метка — «аналитический» тон ($\kappa = 0,55$). Юмор, на котором держится один из главных выводов, воспроизводится почти идеально: $\kappa = 0,93$.

Метка «сильное эссе». Она воспроизводится хуже остальных: $\kappa = 0,70$ на всей выборке и 0,56 (ДИ 0,00–0,88) на 44 основных эссе. Причина — порог. 15 из 44 основных эссе набрали ровно 11 баллов, на шаг ниже порога, и сдвиг одной оценки на балл меняет метку. Расхождения односторонние: все три эссе, сильные в первом проходе, сильные и во втором; второй проход добавил ещё четыре. Поэтому выводы о доле сильных проверены на непрерывной сумме оценок (раздел 6.5), а она воспроизводится с ICC = 0,87.

Проверка на гало-эффект. Признаки и оценки ставит один аннотатор, поэтому связь между ними может быть завышена: модель, решившая, что эссе сильное, охотнее находит в нём «правильные» приёмы. Два прохода позволяют это проверить. Признак берётся из одного прохода, сумма оценок — из другого. Если гало-эффекта нет, корреляция «между проходами» в среднем равна корреляции «внутри прохода». Так и вышло для всех девяти проверенных признаков:

- число приёмов: $\rho = 0,66$ внутри прохода и 0,66 между проходами;

- юмор: 0,36 и 0,35;
- структура вокруг предмета: 0,27 и 0,25;
- финал-пересказ: -0,20 и -0,19.

Разница нигде не превышает 0,02, а её доверительные интервалы лежат в пределах $\pm 0,06$.

Чего эта проверка не ловит. Оба прохода сделала одна модель с одной инструкцией. Её общие пристрастия, например любовь к метафорам или к юмору, повторяются в каждом проходе, и тест их не видит. Частично их проверяют стилевые метрики, не зависящие от модели: связь доли коротких предложений с оценкой голоса одинакова в обоих проходах ($\rho = 0,45$ и $0,41$). Полная проверка требует экспертов-людей (раздел 13). Повторяемость LLM-разметки нельзя считать данностью: она зависит от модели и формулировки задания (Reiss, 2023), поэтому её нужно мерить на своих данных.

4.6. Статистический анализ

- **Доли.** Доли приведены с 95-процентными доверительными интервалами Уилсона.
- **Ассоциации признаков с «сильным эссе».** Тест хи-квадрат и V Крамера.
- **Многомерная модель.** Логистическая регрессия по 799 основным эссе. Предикторы: начало, структура, финал, ставки, юмор, уровень верификации и логарифм длины. Редкие уровни объединены в «другое»; финал-пересказ тоже отнесён туда, потому что среди таких эссе нет ни одного сильного. Проверка устойчивости — линейная регрессия суммы оценок на те же предикторы.
- **Стиль.** Метрики сопоставлены с оценкой голоса корреляцией Спирмена; группы сравнены тестом Манна–Уитни.
- **Разборы издателей.** Каждый критерий сравнивается точным тестом Фишера (приёмные комиссии против консультантов).
- **Характерные слова.** Взвешенный логарифм отношения шансов с информативным априорным распределением (Monroe et al., 2008). Учитываются слова, встречающиеся не менее 15 раз и не менее чем в 20 эссе.
- **Темы без учителя.** Неотрицательная матричная факторизация (NMF, 14 тем) по TF-IDF 1 950 американских эссе, словарь из 6 617 слов. Согласие с темами модели измерено скорректированной взаимной информацией (AMI).
- **Архетипы.** Назначаются прозрачными правилами поверх разметки (раздел 9). Кластеризация k-means по тем же меткам дала размытые группы (silhouette 0,07–0,09 при $k = 5-12$) и была отвергнута.
- **Динамика по годам.** Логистическая регрессия признака на год выпуска по 268 основным эссе из официальных подборок.

Поправки на множественные сравнения не вводились. Поэтому результаты с p между 0,01 и 0,05 следует читать как предварительные.

5. О чём пишут поступившие (RQ1)

Поступившие чаще всего пишут о повседневном и об идеях, а не о трагедиях. Тяжёлый опыт лежит в основе лишь 16 % подтверждённых основных эссе и не повышает долю сильных текстов.

Какие темы встречаются чаще и реже всего:

- **Самые частые темы:** книги и идеи (10,8 %), необычное хобби или привычка (9,3 %), семья (9,3 %), культура и наследие (8,7 %).

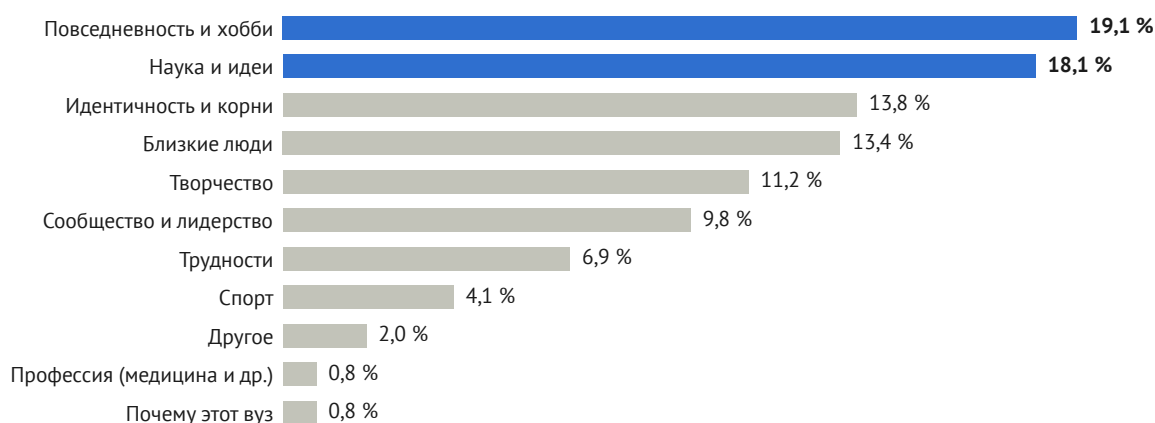


Рис. 1. Повседневность и идеи — самые частые темы поступивших. Доля 492 подтверждённых основных эссе по группам тем, %; две верхние группы вместе дают 37,2 %. Источник: разметка корпуса, 492 подтверждённых основных эссе.

- **Спорт** становится главной темой лишь в 4 % эссе.
- **Мотивация к медицине или профессии** — менее 1 %: эти темы живут в дополнительных эссе и UCAS (раздел 10).

Таблица 3. Ставки и доля сильных эссе

Ставки	Доля подтверждённых основных эссе ($n = 492$)	Доля сильных среди всех основных (95 % ДИ)	n
Повседневное	54,3 %	21,1 % (17,5–25,3)	417
Заметное событие без трагедии	29,5 %	22,7 % (18,0–28,3)	255
Тяжёлый опыт	16,3 %	22,8 % (16,4–30,9)	127

Различия в доле сильных практически нулевые: $\chi^2(2) = 0,3$; $p = 0,85$; V Крамера = 0,02. В регрессии с контролем структуры и стиля (раздел 6) тяжёлый опыт даёт отношение шансов 1,66 (0,95–2,90; $p = 0,07$). Иначе говоря, ни штрафа, ни доказанного бонуса за драму нет.

Важно различать ставки и тему. Эссе, чья тема — сама трудность (болезнь, деньги, неудача), сильны лишь в 11,4 % случаев (5,9–21,0). Это минимум среди крупных групп тем: $\chi^2(10) = 18,9$; $p = 0,041$; $V = 0,15$.

У повседневности и идентичности — по 28 %. При этом тяжёлый опыт часто приходит внутри других тем: 21 из 68 подтверждённых эссе об идентичности и 28 из 66 — о близких людях. Сильные тексты о тяжёлом посвящены не самой беде, а тому, кем автор стал и как изменились его отношения.

Как это выглядит на конкретных эссе. В скобках — источник, год выпуска и сумма трёх оценок из 15. Файлы всех эссе из примеров перечислены в приложении Д.

- **Повседневный материал, сильный текст.** «Fried Rice in One (Not So) Easy Step» (Johns Hopkins, 2024; 13): любительница точных рецептов учится у семейного жареного риса, который готовят без рецепта, и приходит к мысли о жизни без инструкции. «My Bookshelf» (Johns Hopkins, 2013; 13): книжная полка, собранная с отцом, становится портретом семьи и растущих интересов. Эссе о школьной ферме (Tufts, 2021; 13): куры, помидоры и фермерский рынок учат автора видеть скрытые процессы за обычными вещами.

- **Тяжёлый опыт, который сработал.** Эссе о 11 сентября (Connecticut College, 2018; 13): пятилетняя девочка бежит из центра Манхэттена, но эссе о другом — о решении помнить не ужас, а доброту незнакомцев. «Spring Instead of Summer» (Johns Hopkins, 2018; 13): сестра-близнец брата с аутизмом и нарушением зрения строит идентичность вокруг их общей истории, а не вокруг диагноза. «Korean Sticky Notes» (Johns Hopkins, 2028; 13): дочь иммигрантки с небольшим доходом превращает роль маминого «финансового советника» в супермаркете H Mart в изучение корейского и сближение с мамой.
- **Тяжёлый опыт, который не сработал.** Эссе незрячего абитуриента (Tufts, 2018; 7): автор отвергает жалость, но почти не показывает сцен — только утверждения о себе. «Don't Be Sorry» (Johns Hopkins, 2016; 9): жизнь с мамой, у которой рассеянный склероз; бодрый тон и клише оставляют реальные трудности за кадром.

Разница не в тяжести материала, а в том, что делает с ним текст: сильные эссе показывают решение или взгляд автора, слабые пересказывают обстоятельства.

Издатели отбирают драму по-разному.

Источник (основные эссе)	Тяжёлый опыт	Повседневное
Официальные подборки вузов	14 %	60 %
СМИ	17 %	53 %
Подтверждённые подборки консультантов	21 %	44 %

У Johns Hopkins на повседневном материале построено 69 из 102 эссе, на тяжёлом — 8. Вывод: библиотеки консультантов переоценивают трагические истории по сравнению с тем, что выбирают сами вузы.

6. Как построены сильные эссе (RQ2)

Сильные эссе отличаются построением. Они чаще держатся на одном предмете или образе, начинаются со смелого утверждения и используют юмор. Эти различия сохраняются при контроле других признаков.

6.1. Начало

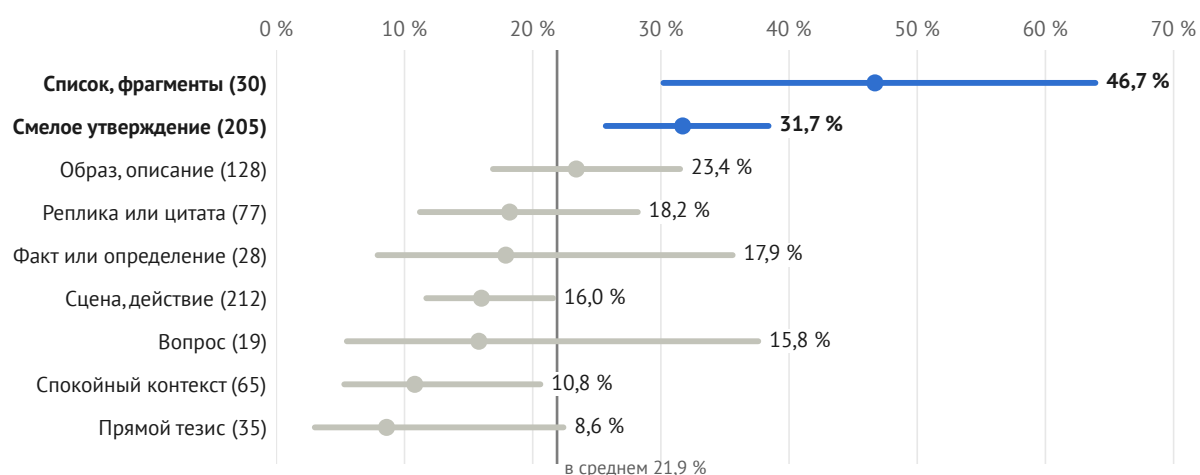


Рис. 2. Смелое первое предложение вдвое чаще даёт сильное эссе, чем сцена. Доля сильных эссе (сумма оценок ≥ 12) по типу начала, % с 95 % ДИ; в скобках — число эссе; синим — интервал выше среднего. Источник: разметка корпуса, 799 основных эссе.

Тип начала связан с качеством сильнее всего из признаков построения: $\chi^2(8) = 36,4$; $p < 0,001$; $V = 0,21$.

- **Смелое утверждение.** Это самое частое начало у подтверждённых эссе (26,8 %) и единственное массовое, чей интервал целиком выше среднего. Это неожиданная, смешная или парадоксальная фраза, которая сразу задаёт характер автора.
- **Сцена.** Начало со сцены — стандартный совет практиков, но сильных эссе среди них вдвое меньше (16,0 %). Его интервал целиком ниже среднего.
- **Контекст и тезис.** Самые слабые начала — спокойный контекст «Я вырос в...» (10,8 %) и прямой тезис (8,6 %). Они тратят первую строку на то, что комиссия знает из анкеты.
- **Список или фрагменты.** Так начинаются лишь 30 эссе, но сильны почти половина из них (46,7 %). Интервал широкий, поэтому это скорее гипотеза, чем вывод.

Примеры из корпуса. Начала пересказаны своими словами; дословно приведена одна фраза.

- **Смелое утверждение.** «I am here because my great-grandfather tied his shoelace» — так начинается эссе Hamilton College (2018; 13 из 15). Фраза задаёт загадку (при чём тут шнурок?) и тему наследия, а последняя фраза возвращается к ботинкам прадеда. Эссе стендап-комика (Connecticut College, 2019; 13, голос 5) открывается признанием: когда автора спрашивают, каково выступать, она говорит неправду. Эссе St. John's College (2026; 13) начинается с «утиного теста», который любимая учительница в пятом классе подала как тревожную истину. Эссе о 11 сентября (Connecticut College, 2018; 13) — с отрицания: в пять лет автор взрослой не стала.
- **Сцена.** Сильный вариант — эссе об ирландских танцах (Hamilton College, 2026; 13): первая строка — возраст рассказчицы, шесть лет, и звуки музыки на празднике; дальше те же сцены в 9, 12, 15 и воображаемых 18 лет. Слабые варианты показывают, что сцена не гарантирует успеха. Эссе о любви к кино (Connecticut College, 2024; 7) открывается взрывом бензовоза на аттракционе, а дальше уходит в общие слова о «магии кино». Эссе Tufts (2017; 7) начинается с зимней простуды, переходит к унижительной «шутке» учителя о терроризме и заканчивается обещанием создать клуб.
- **Список или фрагменты.** «In Pursuit of the Sublime» (Johns Hopkins, 2020; 13, голос 5): первая строка — цель счётчика слов будущего романа, сорок тысяч; меняющиеся показания счётчика ведут от побега в выдуманные истории к честному письму о семье. Эссе Connecticut College (2014; 13, голос 5) открывается однострочным признанием: автору удобнее всего свернувшись калачиком.
- **Спокойный контекст и прямой тезис.** Эссе St. John's College (2021; 8) начинается с того, что автор описывает себя как человека, который всё планирует и много думает, и заканчивается пересказом. «Balancing Life; A Life of Balance» (Johns Hopkins, 2029; 7) открывается заявлением, что жизнью автора управляет идея баланса, а дальше перечисляет роли. Эссе St. John's College (2019; 7) начинается с признания в любви к хорошим книгам.

В слабых началах первая фраза сообщает вывод, в сильных — ставит вопрос, на который придётся отвечать всему тексту.

6.2. Структура

Хронологический рассказ — самая частая структура среди подтверждённых эссе (30,5 %) и одна из самых слабых (13,1 % сильных). Структура вокруг предмета или метафоры

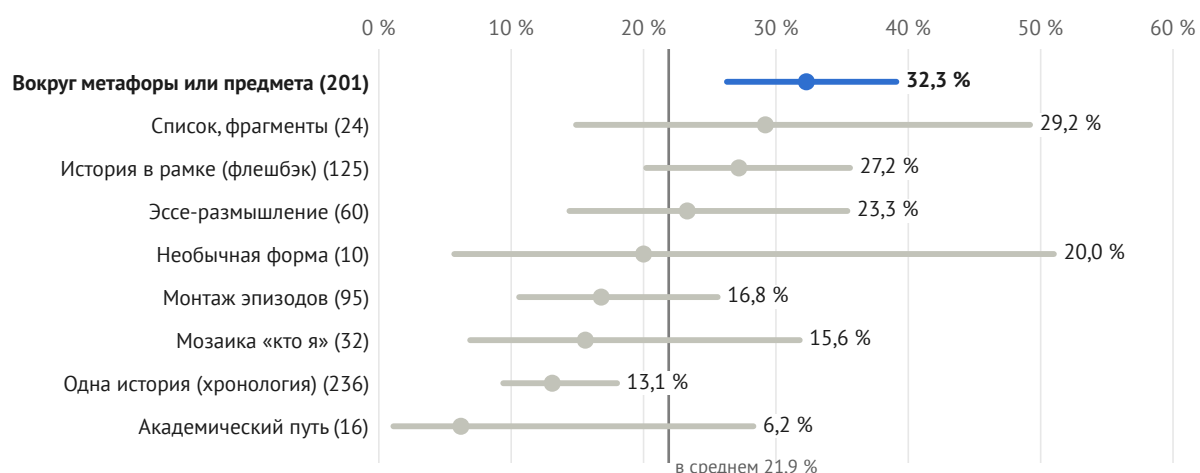


Рис. 3. Эссе вокруг одного предмета сильны в 2,5 раза чаще хронологии. Доля сильных эссе по структуре, % с 95 % ДИ; в скобках — число эссе; синим — интервал выше среднего. Источник: разметка корпуса, 799 основных эссе.

встречается почти так же часто (25,2 %), а сильных среди неё в 2,5 раза больше (32,3 %). Распределение неслучайно: $\chi^2(8) = 30,7$; $p < 0,001$; $V = 0,20$.

Правдоподобное объяснение: хронология подталкивает к пересказу событий, а центральный образ задаёт угол зрения и держит текст в фокусе. История в рамке (начало в моменте → флешбэк → возврат) делает то же средствами композиции: 27,2 % сильных.

Примеры структур.

- **Вокруг предмета или метафоры.** «Corkscrew» (Johns Hopkins, 2015; 14 из 15): очень кудрявые волосы и бесконечные вопросы незнакомцев становятся способом рассказать об открытости миру; поворот в конце — почему автор всегда разрешает трогать волосы. «My Bookshelf» (Johns Hopkins, 2013; 13): полка, построенная с отцом, где каждая книга связана с конкретным человеком. Эссе о Costco (NBC News; автор поступила в Yale, Columbia, Penn, Dartmouth, Cornell и Stanford; 13): двухлетний «конкистадор» в оптовом магазине вырастает в любопытного человека, а магазин — в метафору жажды открытий. «Itch To Understand» (EssaysThatWorked.com, Stanford; 13): «зуд понимать» проходит через весь текст — от детства без наставников до инженерии энергетики.
- **История в рамке.** Эссе о 11 сентября (Connecticut College, 2018; 13): разговор за обедом в настоящем времени запускает флешбек, а точная деталь — порванная рубашка незнакомца — подводит к главной мысли. «Spring Instead of Summer» (Johns Hopkins, 2018; 13): сон о крушении самолёта открывает текст и возвращается в конце.
- **Монтаж эпизодов.** Эссе об ирландских танцах (Hamilton College, 2026; 13): сцены в 6, 9, 12, 15 и 18 лет связаны рефреном о том, что главное — сам танец. «My Best Kept Secret» (Johns Hopkins, 2014; 13): тайный школьный клуб служит нитью для серии странных эпизодов — от споров о трепанации до круговой поэзии.
- **Необычная форма.** «In Pursuit of the Sublime» (Johns Hopkins, 2020; 13) построено как журнал счётчика слов. Эссе Hamilton College (2007; 12) — о том, как трудно написать эссе о себе: отвергнутые темы сами становятся характеристикой автора.
- **Хронология, которая не работала.** «Piece by Piece» (Connecticut College, 2022; 7): от Lego в детстве до керамики в пансионе — пересказ вместо сцен, финал повторяет сказанное. Эссе о переходе от кардиологии к писательству (Connecticut College, 2022; 8): любовь к письму заявлена высокими словами, но не показана ни одной сценой.

6.3. Финал

Таблица 4. Доля сильных эссе по типу финала (799 основных эссе)

Финал	n	Доля сильных (95 % ДИ)
Развязка метафоры	125	28,8 % (21,6–37,3)
Возврат к началу	232	25,9 % (20,7–31,9)
Шутка или твист	43	25,6 % (14,9–40,2)
Открытый финал	14	21,4 % (7,6–47,6)
Финальная сцена или образ	20	20,0 % (8,1–41,6)
Взгляд в будущее	111	19,8 % (13,5–28,2)
Вывод, урок	207	17,9 % (13,3–23,7)
«Поэтому я хочу в ваш вуз»	18	11,1 % (3,1–32,8)
Пересказ	29	0,0 % (0,0–11,7)

Связь финала с качеством слабее, чем у начала: $\chi^2(8) = 17,6$; $p = 0,025$; $V = 0,15$. Единственный однозначный результат — пересказ: среди 29 таких эссе нет ни одного сильного.

Финал сильнее всего работает в паре с началом.

Пара «начало → финал»	n	Доля сильных
Смелое утверждение → возврат к началу	56	42,9 %
Смелое утверждение → взгляд в будущее	31	35,5 %
Смелое утверждение → шутка	23	34,8 %
Образ → вывод	28	32,1 %
Сцена → развязка метафоры	26	30,8 %
Сцена → возврат к началу	73	23,3 %

Как замыкают текст сильные эссе.

- **Кольцо.** Эссе Connecticut College (2014; 13) заканчивается почти той же фразой, с которой начиналось: автору удобнее свернувшись калачиком. Но к финалу эта привычка уже значит «маленький мир», память и одиночество.
- **Зеркальная сцена.** Эссе об ирландских танцах (Hamilton College, 2026; 13) заканчивается воображаемой сценой в 18 лет, которая повторяет первую, в 6 лет.
- **Возврат образа с новым смыслом.** Утки из первой фразы эссе St. John's College (2026; 13) возвращаются в последней: автор надеется стать одной из тех, кто умеет наблюдать за «утками». Ботинки прадеда в эссе Hamilton College (2018; 13) в финале превращаются в открытый вопрос о том, чьё это наследие.
- **Развязка метафоры.** «My Bookshelf» (Johns Hopkins, 2013; 13) заканчивается образом досок полки, которые снова оживают; эссе о школьной ферме (Tufts, 2021; 13) — обещанием искать новую почву для того, чему автор научилась.
- **Пересказ — слабый финал.** «The Story of Us» (Johns Hopkins, 2015; 5 из 15) обещает в финале раскрыть «тайны человечества» через антропологию и историю — после текста, где нет ни одной книги или эпизода. «Piece by Piece» (Connecticut College, 2022; 7) заканчивается повтором тезиса о пользе Lego.

6.4. Сюжетная арка

Арка связана с качеством не слабее начала: $\chi^2(9) = 37,7$; $p < 0,001$; $V = 0,22$. Самая популярная схема «трудность → рост» среди слабейших: её ждут от абитуриента, и в ней труднее сказать что-то неожиданное. Сильнее всего арки без внешнего конфликта, где меняется взгляд автора или виден ход его мысли.

Таблица 5. Доля сильных эссе по арке

Арка	Доля среди подтверждённых	n (все основные)	Доля сильных (95 % ДИ)
Погоня за вопросом	5,1 %	33	33,3 % (19,8–50,4)
Принятие себя	12,8 %	102	30,4 % (22,3–39,9)
Портрет мышления	16,1 %	143	30,1 % (23,2–38,0)
Смена взгляда	18,7 %	116	29,3 % (21,8–38,2)
Урок от человека	8,1 %	53	20,8 % (12,0–33,5)
Ценности в поступках	5,9 %	45	20,0 % (10,9–33,8)
Как родилось увлечение	10,4 %	81	16,0 % (9,6–25,5)
Трудность → рост	18,5 %	173	12,1 % (8,1–17,8)
Что я сделал для других	2,4 %	24	4,2 % (0,7–20,2)
Обоснование выбора	2,0 %	29	3,4 % (0,6–17,2)

Примеры арок.

- **Погоня за вопросом.** Эссе St. John's College (2022; 12): военнотружущий читает «Уловку-22» как точную картину службы в ВВС и защищает сатиру как способ пробить идеологические стены. Эссе St. John's College (2026; 12): разбор «Энеиды» заканчивается шквалом вопросов, которые поэма оставила автору.
- **Смена взгляда.** Эссе Connecticut College (2015; 13): скучающая школьница на семейном походе к дому прабабушки в греческой деревне через старые вещи — карты, передник, фальшивую стену — приходит к чувству дома и родства.
- **«Трудность → рост», слабые версии.** «The Art of Imperfection» (Johns Hopkins, 2028; 8): трудности с рисунком руки и курс data science приводят к выводу о терпении и росте — ровно туда, куда читатель ждёт. Эссе о брейкдансе, выученном по YouTube (Connecticut College, 2023; 8): преодоление описано высокими словами, а кто и что мешало, остаётся неясным.

6.5. Многомерная модель

Логистическая регрессия по 799 основным эссе неслучайна: тест отношения правдоподобия $p < 0,001$, псевдо- R^2 Макфаддена 0,12. При взаимном контроле признаков и длины сохраняются три эффекта.

- **Структура.** По сравнению с хронологией шансы сильного эссе выше у структуры вокруг предмета (ОШ = 2,53; 1,41–4,54), истории в рамке (2,30; 1,25–4,23) и эссе-размышления (2,37; 1,06–5,34).
- **Начало.** Смелое утверждение против сцены: ОШ = 1,87 (1,10–3,16).
- **Юмор.** ОШ = 2,48 (1,64–3,76); без контроля это 32,7 % сильных против 16,7 %.

Различия между финалами после контроля исчезают. Вероятная причина — финал произведен от структуры: развязка метафоры живёт в эссе вокруг предмета, возврат к началу — в историях в рамке.

Два контрольных эффекта тоже значимы.

- **Длина.** Каждые +10 % длины увеличивают шансы в 1,23 раза: короткие основные эссе чаще недоразвиты.
- **Источник.** Эссе без подтверждения поступления слабее официальных: ОШ = 0,60 (0,38–0,92).

Проверка устойчивости — линейная регрессия суммы оценок на те же предикторы. Она объясняет 27 % дисперсии и даёт те же знаки эффектов.

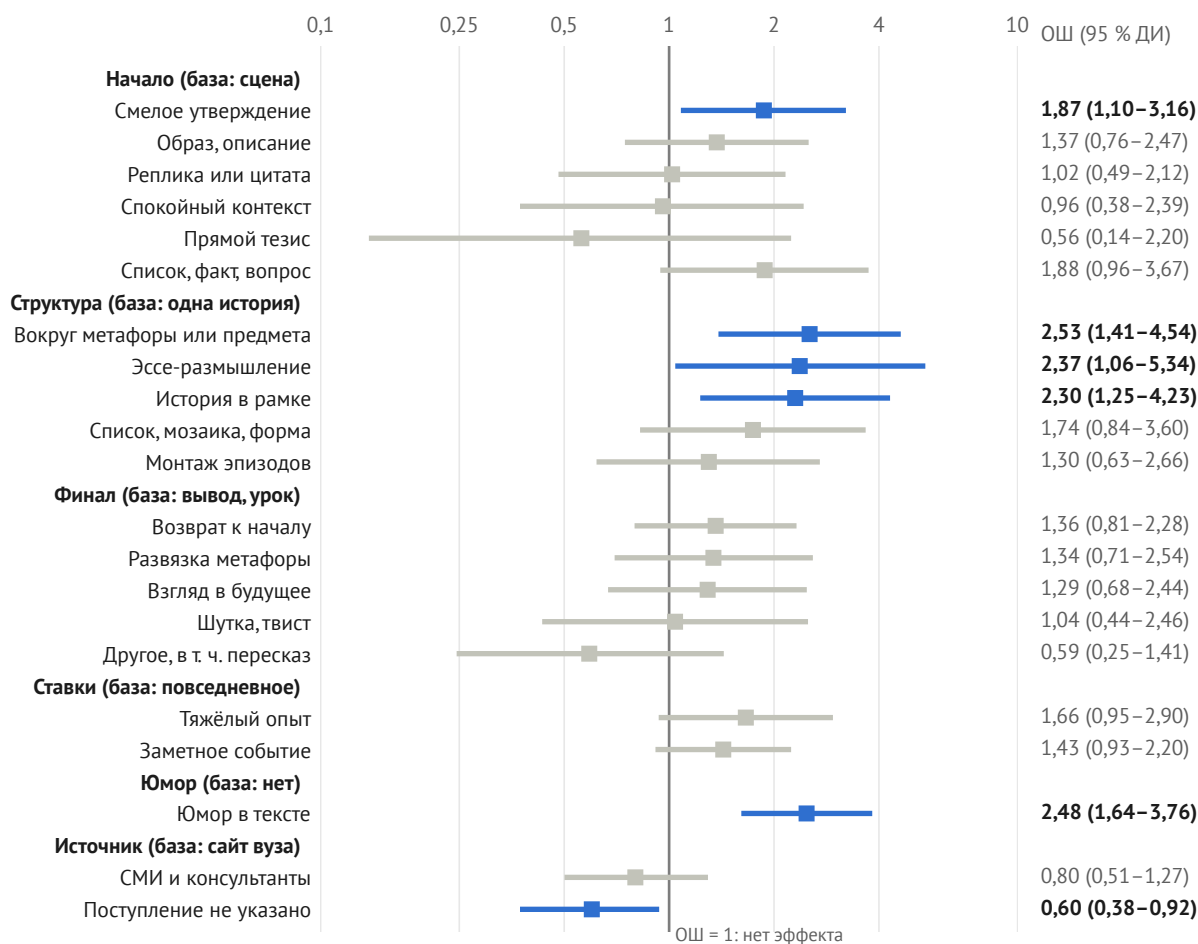


Рис. 4. Структура, смелое начало и юмор повышают шансы на сильное эссе. Отношения шансов с 95 % ДИ, логарифмическая шкала; синим — $p < 0,05$. Логистическая регрессия: сильное эссе ~ признаки + источник + $\log(\text{длина})$; длина не показана (ОШ 8,96 на единицу логарифма). Источник: разметка корпуса, 799 основных эссе.

7. Голос и стиль (RQ3)

Яркий голос измерим. У эссе с высшей оценкой голоса фразы короче, коротких предложений почти втрое больше, чем у слабых, а «я» встречается реже. Почти все такие тексты используют юмор.

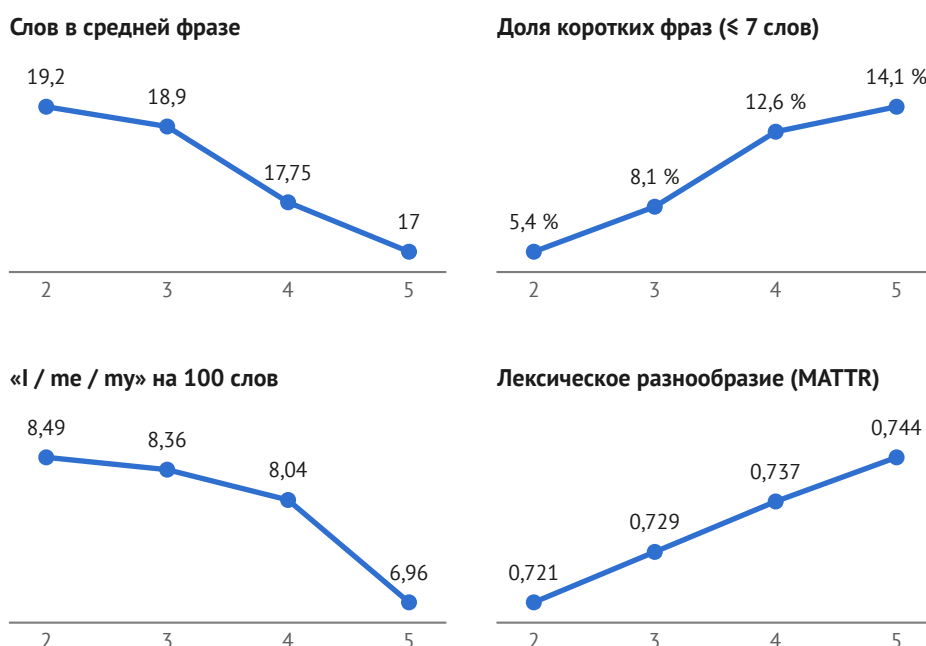
7.1. Стилиевые метрики

Стилиевые метрики считались скриптом, независимо от модели. То, что они монотонно меняются с оценкой голоса, — свидетельство сходящейся валидности: оценка модели опирается на наблюдаемые свойства текста. Сила связи умеренная: даже лучшая метрика объясняет лишь часть того, что читатель слышит как «голос».

Картина однозначна: живой голос строится на ритме (короткие фразы среди обычных), разнообразии слов и простоте синтаксиса. Длинные фразы сами по себе не вредят: решает чередование.

7.2. Приёмы и тон

Всего юмор используют 31 % основных эссе, а среди 20 эссе с голосом 5 — 85 %. Распространённые «ремесленные» приёмы — имена и названия, числа, риторические вопро-



Ось X — оценка голоса. Число эссе: голос 2 — 68, голос 3 — 357, голос 4 — 354, голос 5 — 20

Рис. 5. Чем ярче голос, тем короче фразы и реже «я». Медианы стиливых метрик по оценке голоса (2–5). Метрики посчитаны скриптом, оценка голоса — разметкой. Источник: 799 основных эссе.

Таблица 6. Корреляция Спирмена стиливых метрик с оценкой голоса (799 основных эссе)

Метрика	ρ	p
Доля коротких фраз (≤ 7 слов)	+0,28	< 0,001
Лексическое разнообразие (MATTR)	+0,20	< 0,001
Число абзацев	+0,19	< 0,001
Доля диалога	+0,15	< 0,001
Разброс длины фраз	+0,15	< 0,001
Длина эссе	+0,15	< 0,001
Средняя длина фразы	-0,15	< 0,001
Индекс Flesch–Kincaid (класс)	-0,15	< 0,001
Лёгкость чтения Flesch	+0,13	< 0,001
Вопросы в тексте	+0,13	< 0,001
«I / me / my» на 100 слов	-0,09	0,008
Доля длинных фраз (≥ 30 слов)	-0,02	0,54

сы — с голосом не связаны или связаны отрицательно. Конкретика полезна для достоверности, но голос делает не она.

Самый частый тон — искренний и серьёзный (63 % подтверждённых эссе), и сильных среди него меньше всего. Это не значит, что серьёзность плоха. Лиричный и аналитичный тон тоже серьёзны, но имеют выраженную интонацию. Слабее всего нейтральная интонация без особых примет.

Юмор и голос на примерах.

- «Corkscrew» (Johns Hopkins, 2015; голос 5): гиперболы о кудрях и повторяющийся ответ незнакомцам, что волосы такие от природы, держат 1 200 слов без потери темпа.
- «Temper» (Johns Hopkins, 2018; голос 5): первая фраза с нелепо точной географией скейт-парка в Чикаго сразу задаёт интонацию, а прощение старого обидчика за пиц-

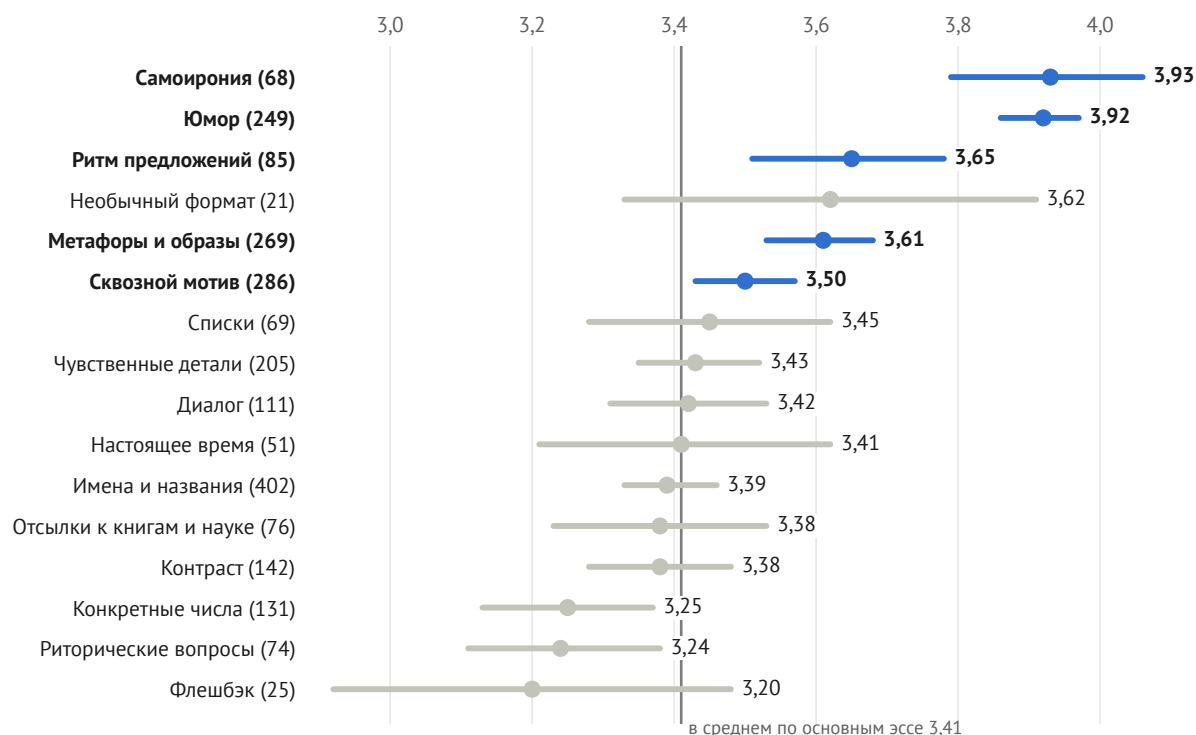


Рис. 6. Юмор и самоирония — самые сильные маркеры голоса. Средняя оценка голоса (1–5) у основных эссе с приёмом, 95 % ДИ; в скобках — число эссе; показаны приёмы, встречающиеся не менее чем в 20 основных эссе. Источник: разметка корпуса.

Таблица 7. Тон и качество (основные эссе)

Тон	Доля среди подтверждённых	Средний голос	Доля сильных
Лиричный	15,0 %	3,66	40,8 %
С юмором	19,9 %	4,01	35,0 %
Игривый	23,2 %	3,84	31,3 %
Аналитичный	9,8 %	3,22	26,9 %
Уязвимый	19,3 %	3,31	24,4 %
Тёплый	29,7 %	3,34	17,4 %
Уверенный	12,8 %	3,38	14,7 %
Искренний, серьёзный	62,6 %	3,18	14,6 %

цей обходится без пафоса.

- «Cinco Reasons Why I am Interested in Pursuing Latin American Studies» (Johns Hopkins, 2014; голос 5): обратный отсчёт пяти причин выбрать специальность и самоироничный финальный поворот.
- Эссе стендап-комика (Connecticut College, 2019; голос 5): сквозной мотив «самовозгорания» на сцене и финальная шутка о смехе из жалости превращают страх в материал.
- **Ритм.** В эссе об ирландских танцах (Hamilton College, 2026) короткий рефрен повторяется при каждом новом возрасте; эссе Connecticut College (2014) держится на коротких фразах-признаниях.

7.3. Слова

Таблица 8 сравнивает категории служебных слов, как в исследовании Pennebaker et al. (2014).

Таблица 8. Служебные слова на 100 слов, медианы (основные эссе)

Категория	Сильные	Остальные	p (Манн–Уитни)
«I / me / my»	8,02	8,48	0,003
Предлоги	12,35	12,56	< 0,001
Вспомогательные глаголы	4,83	5,02	0,07
Отрицания	0,75	0,61	0,04
Артикли	7,30	7,13	0,31

Сильные эссе реже говорят «я», но не становятся от этого «категориальнее». Предлогов в них даже немного меньше, а разница в артиклях незначима. Стилль, который у Pennebaker предсказывает оценки в университете, и стилль, который читатель называет живым, — разные вещи.

Взвешенные логарифмы шансов (Monroe et al., 2008) показывают слова, характерные для каждой группы.

Характерно для сильных эссе	Характерно для остальных
she, her — другие люди в кадре	team, students, others, help — обобщённое «сообщество»
is, are — настоящее время perhaps, like, then — сомнение, сравнение	had, was, got — отчёт о прошлом skills, work, research, education, college — словарь резюме
you, it — разговорная интонация philosophy, language, name, clock, tree — конкретные предметы и идеи	my, i, life — абстрактное «я и моя жизнь» able, continue — канцелярит

Слабое эссе звучит как отчёт о прошлом со словарём резюме. Сильное — как разговор в настоящем времени, с другими людьми и конкретными предметами в кадре.

7.4. Клише

Хотя бы одно из 20 частых клише есть в 21 % эссе корпуса. Чаше других встречаются «I realized that» (173 раза) и «my passion for» (132). Доля сильных среди эссе с клише и без них значимо не различается: 19,4 % против 22,8 %, $p = 0,36$. Отдельное клише текст не убивает, но и ничего не добавляет.

Пример: «I realized that» есть и в эссе стендап-комика (Connecticut College, 2019), и в эссе о 11 сентября (Connecticut College, 2018), и оба получили 13 из 15. Клише теряется, когда вокруг него конкретные сцены.

8. Что хвалят те, кто читает (RQ4)

Приёмные комиссии описывают сильное эссе как портрет человека, а консультанты — как хорошо сделанный текст. Разборы издателей составляют самый прямой публичный след того, что читатель ценит в эссе. Мы разметили 1 094 таких разбора по 20 критериям.

Приёмные комиссии вдвое чаще консультантов называют три качества человека:

- **характер и ценности** — 51 % против 24 %, $p < 0,001$;
- **рост** — 31 % против 12 %;
- **любопытность** — 28 % против 14 %.

Вклад в сообщество комиссии упоминают в 2,5 раза чаще (23 % против 9 %). Это вопрос «каким студентом он будет у нас», который в разборах Johns Hopkins звучит постоянно.

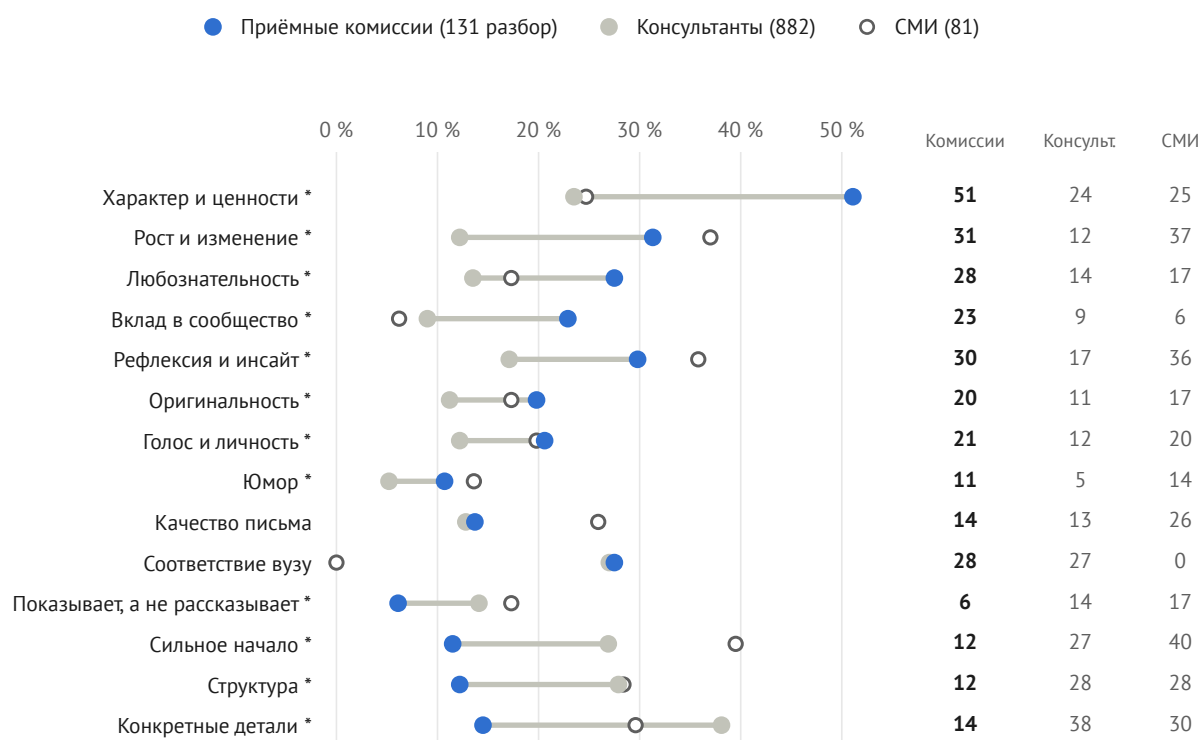


Рис. 7. Комиссии хвалят характер, консультанты — детали и структуру. Доля разборов «почему это сработало», где хвалят критерий, %; * — разница комиссий и консультантов значима (точный тест Фишера, $p < 0,05$). Источник: 131 разбор приёмных комиссий, 882 разбора консультантов и 81 разбор СМИ, размеченные по 20 критериям.

Консультанты, наоборот, вдвое–втрое чаще хвалят ремесло:

- **конкретные детали** — 38 % против 15 %;
- **структуру** — 28 % против 12 %;
- **сильное начало** — 27 % против 12 %.

СМИ — прежде всего The Harvard Crimson (76 из 81 разбора) — ближе к консультантам по вниманию к началу (40 %). При этом они чаще всех пишут о росте и рефлексии.

Расхождение не значит, что комиссии равнодушны к ремеслу: сильные структуры и начала встречаются в их подборках не реже (раздел 6). Ремесло для них — средство, а не предмет оценки: техника, которая ничего не доказывает об авторе, в разборах комиссий не упоминается.

Как это звучит в самих разборах (пересказ).

- **Johns Hopkins о «My Best Kept Secret» (2014).** Разбор хвалит не приём, а человека: одна нелепая тема — тайный клуб — показывает разные грани характера. Это любовь к идеям ради удовольствия, готовность поддержать одноклассников и отстаивать свои взгляды. Отдельно отмечено, что эссе странное в хорошем смысле.
- **Johns Hopkins о «Speak» (2014).** Эссе о жизни с заиканием ценят за открытое окно в эмоции и борьбу автора. Сцены в кофейне, в Мюнхене и на первой экскурсии по кампусу названы деталями, которые раскрывают силу характера.
- **EssaysThatWorked.com об «Itch To Understand».** Главное достоинство по разбору — центральная метафора «зуда понимать». Совет: брать метафору, которая действительно выражает мысль, а не просто звучит оригинально.
- **EssaysThatWorked.com о «Baking With Buni».** Разбор хвалит обманутое ожидание. Эссе ставит задачу — говорить по-румынски без акцента, — но решает её неожиданно.

но, через выпечку, и финал оказывается связан с началом.

Комиссия описывает, каким студентом автор окажется в кампусе; консультант — каким приёмом это сделано.

Соответствие вузу важно обеим сторонам одинаково (27,5 % и 27,0 %, $p = 0,92$). СМИ его не упоминают вовсе, потому что разбирают только основные эссе.

Ограничение. 102 из 131 разбора приёмных комиссий написаны Johns Hopkins. Остальные — Connecticut College (9), Olin (8), Emory (7) и University of Georgia (5). Вывод описывает прежде всего язык одной комиссии, но остальные четыре вуза следуют той же логике.

9. Типы историй и темы без учителя (RQ5)

Из 11 типов историй один одновременно самый частый и самый сильный — «маленький предмет — большой смысл». Самые слабые типы строятся на списке заслуг или на обосновании выбора.

9.1. Архетипы

Архетип — тип истории, который выбирает автор. Он назначается по разметке правилами в порядке приоритета:

1. UCAS или обоснование выбора;
2. тяжёлые ставки;
3. идентичность;
4. структура вокруг предмета;
5. монтаж или список;
6. эссе-размышление;
7. урок от человека;
8. дело для других;
9. короткий эпизод;
10. увлечение;
11. остальное — «портрет характера».

Правила предпочтены кластеризации (раздел 4.6): их легко применить к своему черновику.

Распределение доли сильных по архетипам неслучайно: $\chi^2(10) = 25,0$; $p = 0,005$; $V = 0,18$. Архетипы делятся на три группы.

- **Частые и сильные.** Сюда относится «маленький предмет»: 17 % эссе и 32 % сильных — единственный тип, чей интервал целиком выше среднего. Близки к нему «корни и идентичность» (26 %).
- **Редкие, но сильные.** «Один момент» (26 %) и «ум в действии» (25 %). Эти типы требуют редкого материала: одной поворотной сцены или настоящего интеллектуального интереса.
- **Слабые.** «Портрет характера» (10 %), «дело для других» (14 %), «история увлечения» (13 %) и «обоснование выбора» (5 %). Все они легко скатываются в пересказ резюме.

«Испытание» и «мозаика» — популярные типы со средней или ниже средней долей сильных. Они работают, но не дают форы сами по себе.

Архетипы на примерах. В скобках — источник, год выпуска и сумма оценок из 15.

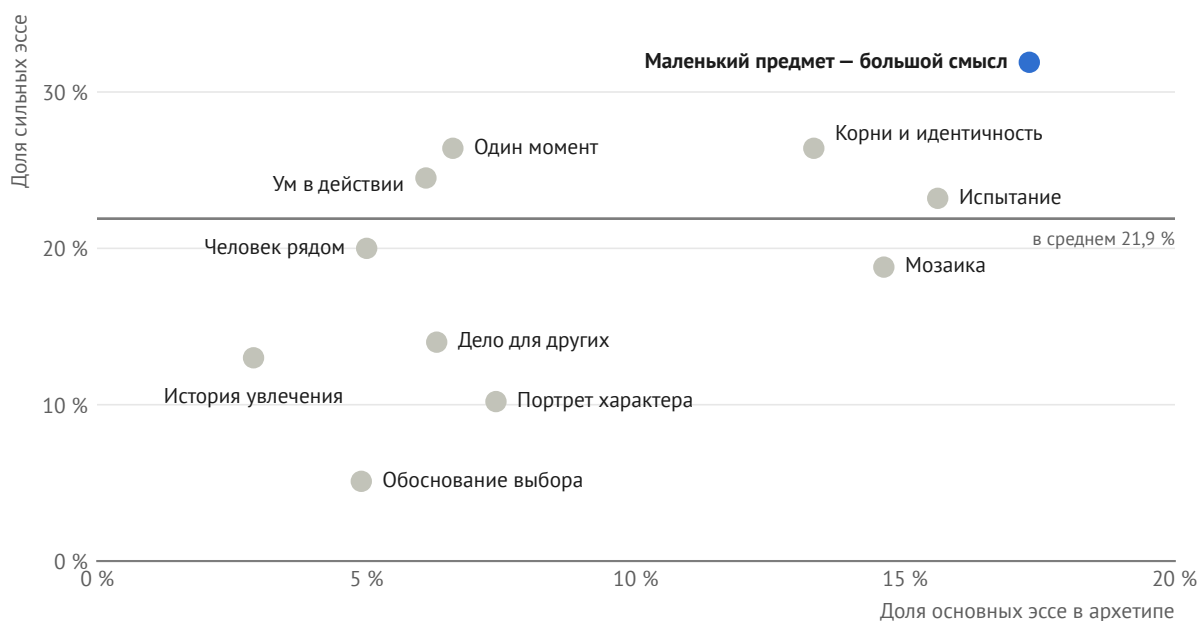


Рис. 8. «Маленький предмет» — самый частый и самый сильный тип истории. Архетипы по правилам из приложения А: по горизонтали — доля основных эссе, по вертикали — доля сильных. Источник: разметка корпуса, 799 основных эссе.

Архетип	Примеры из корпуса
Маленький предмет — большой смысл	«Fried Rice in One (Not So) Easy Step» (Johns Hopkins, 2024; 13): семейный рис без рецепта как урок импровизации · эссе о принтере этикеток P-Touch (The Harvard Crimson; 13): людей нельзя свести к ярлыкам
Испытание	11 сентября глазами пятилетней (Connecticut College, 2018; 13) · «Korean Sticky Notes» (Johns Hopkins, 2028; 13): мамин «финансовый советник» учит корейский
Корни и идентичность	«In our house, English is not English» (The Tab; 13): «неправильный» английский мамы из Малайзии — от стыда к дому из слов · «Corkscrew» (Johns Hopkins, 2015; 14)
Мозаика	ирландские танцы в 6, 9, 12, 15 и 18 лет (Hamilton College, 2026; 13) · «My Best Kept Secret» (Johns Hopkins, 2014; 13)
Ум в действии	«Уловка-22» глазами военнослужащего (St. John's College, 2022; 12) · сознательный отказ от телевизора (Connecticut College, 2014; 12)
Человек рядом	дом, ставший маникюрным салоном мамы-иммигрантки (Hamilton College, 2022; 13) · «Temper» (Johns Hopkins, 2018; 13): скейт-парк и прощение обидчика
Дело для других	«Hope Lives In That Laughter» (EssaysThatWorked.com, Grinnell; 12): дебатёру запретили речь об однополых браках, и он учится спорить с несогласными
Один момент	поход к дому прабабушки в греческой деревне (Connecticut College, 2015; 13) · первый стендап (Connecticut College, 2019; 13)
История увлечения	закусочная отца и съёмка хот-догов как американской иконографии (Connecticut College, 2019; 12) · нахлыст и ручей Fall Creek (Hamilton College, 2022; 12)
Портрет характера	одинадцать лет в одной и той же «униформе» и её связь с искусством (College Confidential, Tufts; 12)
Обоснование выбора	Economics, Cambridge (College Lah; 12): спор с автором книги и собственная регрессия · «Dissonance» (Johns Hopkins, 2018; 12): «Why us» на метафоре диссонанса

9.2. Темы без учителя

Чтобы проверить темы разметки независимым методом, NMF нашла 14 тем прямо в словах 1 950 американских эссе. Согласие с темами модели умеренное: AMI = 0,25.

Таблица 9. Темы NMF, их слова и доля сильных основных эссе

Тема NMF	Ключевые слова	Доминирующая тема разметки	Эссе	Доля сильных (95 % ДИ)
Чувственные образы	eyes, water, air, light, red	хобби и странности (17 %)	235	31,9 % (25,2–39,4)
Музыка	music, piano, song, notes	музыка (45 %)	97	27,3 % (16,3–41,8)
Разговорная рефлексия	know, think, things, feel	хобби, идеи	272	26,9 % (20,1–34,9)
Культура и язык	culture, language, identity, chinese	идентичность (45 %)	132	25,4 % (16,1–37,8)
Книги и письмо	book, reading, story, writing	книги и идеи (39 %)	130	23,2 % (14,8–34,4)
Инженерия и IT	engineering, computer, design, math	наука (34 %)	164	23,1 % (12,6–38,3)
Политика	political, policy, global, climate	активизм (34 %)	108	22,7 % (10,1–43,4)
Повествование в прошлом	didn, did, felt, wanted, knew	здоровье, неудачи	90	18,9 % (10,6–31,4)
Семья	family, parents, father, mom	семья (44 %)	123	13,4 % (7,2–23,6)
Сообщество и служение	community, club, service, help	волонтерство (32 %)	184	9,8 % (3,9–22,5)
Медицина	medical, biology, patients, brain	медицина, вуз	130	7,5 % (2,6–19,9)
Командный спорт	team, game, coach, season	спорт (35 %)	116	7,5 % (3,2–16,3)
Вуз и программы (две темы)	research, courses; tufts, campus	почему этот вуз (78–95 %)	169	основных эссе почти нет

Одиннадцать тем NMF совпадают с темами разметки, три описывают не содержание, а регистр: чувственные образы, разговорную рефлексия и повествование в прошлом. Отсюда умеренный AMI.

Независимый метод воспроизводит главное. Сильнее всего эссе с сенсорной и разговорной лексикой. Слабее всего — на словаре командного спорта, служения и медицины, то есть тех сфер, где абитуриенты чаще всего пересказывают достижения.

10. Дополнительные эссе и UCAS как отдельные жанры (RQ6)

У дополнительных эссе и британских statements свои правила, но принцип тот же, что у основного эссе: выигрывает текст, который показывает ход мысли автора, а не перечень. В supplements это рамка, метафора или монтаж вместо «карты ресурсов вуза». В UCAS — академика больше чем на 70 % текста и прочитанное, а не опыт стажировок.

10.1. Дополнительные эссе

В корпусе 1 151 дополнительное эссе: «Why us» и «Why major», UC Personal Insight Questions, эссе о сообществе и активностях, короткие ответы. Медиана — 255 слов, исход подтверждён у 52,9 %. Сильными считались тексты с суммой оценок от 11 (раздел 4.4); таких 10,9 % (9,3–12,9).

Оценки ниже, чем у основных эссе: конкретика 3,27 против 3,76, рефлексия 2,46 против 3,14, голос 2,78 против 3,41. Сильнее всего проседают рефлексия и голос — на 0,6–0,7 балла: в 250 словах мало места для мысли.

Частые структуры:

- «Why us», то есть карта «мои интересы ↔ ресурсы вуза», — 24,7 %;
- хронологический рассказ — 23,0 %;
- короткий ответ — 15,2 %;
- эссе-размышление — 7,6 %;
- остальные семь структур — от 2 до 6 % каждая.

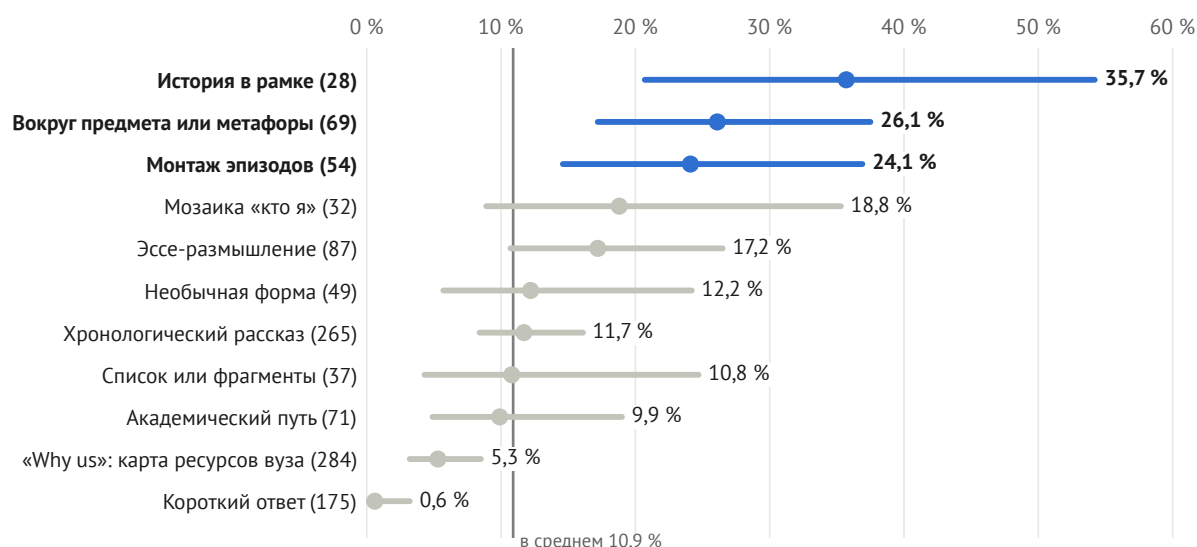


Рис. 9. В дополнительных эссе выигрывают рамка, метафора и монтаж, а не «карта ресурсов». Доля сильных (сумма оценок ≥ 11) по структуре, % с 95 % ДИ; в скобках — число эссе. Источник: разметка корпуса, 1 151 дополнительное эссе.

Связь структуры с качеством здесь сильнее, чем у основных эссе: $\chi^2(10) = 77,9$; $p < 0,001$; $V = 0,26$ (у основных $V = 0,20$). Рамка, метафора и монтаж дают 24–36 % сильных — в 2–3 раза больше среднего. «Карта ресурсов» даёт 5,3 %, короткий ответ — 0,6 %.

Длина. До 150 слов сильных почти нет (0,5 %), при 150–499 словах их 12 %, от 500 слов — 22,6 %. Но структура работает и при равной длине. В логистической регрессии с контролем длины, типа начала и уровня верификации (псевдо- $R^2 = 0,18$) шансы выше, чем у хронологии:

- у рамки — ОШ = 3,08 (1,24–7,61);
- у метафоры — 2,51 (1,24–5,07);
- у монтажа — 2,14 (1,00–4,57).

У «Why us» шансы ниже: ОШ = 0,36 (0,18–0,70). Начало действует так же, как в основных эссе: смелое утверждение повышает шансы (ОШ = 1,88; 1,05–3,36), спокойный контекст снижает (0,32; 0,13–0,77).

«Why us». В 284 таких эссе конкретика вуза — норма: 91,5 % называют курсы, профессоров или традиции. Поэтому она не отличает сильные тексты от слабых: 5,4 % сильных с названиями и 4,2 % без них. Отличает начало. Из 42 «Why us», начинающихся смелым утверждением, сильны 19,0 % (10,0–33,3). Из 91, начинающегося прямым тезисом вроде «I want to attend X because...», — 1,1 % (0,2–6,0); точный тест Фишера $p < 0,001$.

Сильные «Why us» подают вуз как ответ на личный вопрос автора. Пример — спор об «Илиаде» и «Одиссее» по дороге на урок физики: из него автор выводит, почему UChicago

для него — «погоня за вопросами».

Ещё примеры.

- **«Why us» со смелым началом.** «Dissonance» (Johns Hopkins, 2018; 12): первая фраза объявляет, что мозг автора звучит нестройно. Метафора диссонанса связывает Магритта, спутник Европу, свободу воли и феминизм с конкретными ресурсами JHU — вплоть до идеи возродить студенческий журнал. Эссе для Penn (CollegeAdvisor; 11): субботний семейный Scrabble как серьёзное дело и закон убывающей полезности у будильника ведут к экономике и программе Wharton.
- **«Why us» с прямым тезисом.** «Pursuing My Passion in Research» (Johns Hopkins, 2025; 7): первая фраза — о мечте всей жизни помогать человечеству, дальше общие формулы и средний балл 4,0. Фрагмент для Olin (2029; 7): связь принципа «тикун олам» с миссией вуза заявлена, но не подкреплена ни одним проектом.
- **Структура в 250–650 словах.** «The Illuminati changed my life» (PrepScholar, UChicago; 13): амбиграммы, открытые после романа Дэна Брауна, становятся метафорой двойственности самого автора. Эссе о книге (CollegeVine, Stanford; 13): затёртый экземпляр «1984» с пометками на двух языках — путь иммигранта от детских книжек к взрослому чтению по-английски.

10.2. UCAS personal statements

В корпусе 332 statements с подтверждёнными офферами: Oxford, Cambridge, LSE, UCL, Imperial, медицинские школы и другие. Медиана — 644 слова. Сильными считались statements с суммой оценок от 10; таких 25,6 % (21,2–30,6). 90 statements разбиты на три вопроса формата 2026 года, но изначально в нём написаны лишь 6. Остальные 84 переложила в эту форму сама The Uni Guide (раздел 13).

Жанр почти без вариаций.

- 97,9 % — академический путь: интерес → углубление → доказательства → цель.
- Половина начинается прямым тезисом (50,9 %), ещё пятая часть — биографией (20,2 %).
- Финал — взгляд в будущее (55,7 %) или пересказ (30,7 %).

Язык плотнее, чем у американских эссе. Медианная фраза — 23,6 слова против 18,4 у основных эссе. Индекс удобочитаемости Флеша — 40,6 против 62,9, уровень Флеша–Кинкейда — 13,5 против 9,1. «Я» встречается реже: 6,2 раза на 100 слов против 8,2. Клише есть в 29 % statements, чаще других «my passion for» (37 текстов). Из трёх оценок слабее всего голос: 2,24 против 3,41 у основных эссе.

Что отличает сильные statements. Главный признак — доля академики. Statements, где предмету отдано больше 70 % текста, сильны в 35,2 % случаев, остальные — в 9,0 %, почти вчетверо реже ($p < 0,001$). Помогает и начало не прямым тезисом: 31,3 % против 20,1 % ($p = 0,02$).

Доказательства интереса делятся на две группы.

- **Повышают долю сильных:** чтение (29,7 % против 9,1 %), исследование или ERQ (33,9 % против 21,4 %), олимпиады (31,1 % против 22,5 %), лекции и онлайн-курсы (30,5 % против 22,4 %).
- **Понижают:** школьные предметы (22,1 % против 33,7 %), work experience (20,6 % против 33,6 %), наблюдение за врачом (14,3 % против 29,0 %).

Внеучебные навыки есть почти везде (95 %) и ничего не различают. Несколько доказательств не лучше одного: число типов почти не связано с оценкой ($\rho = 0,10$; $p = 0,06$).

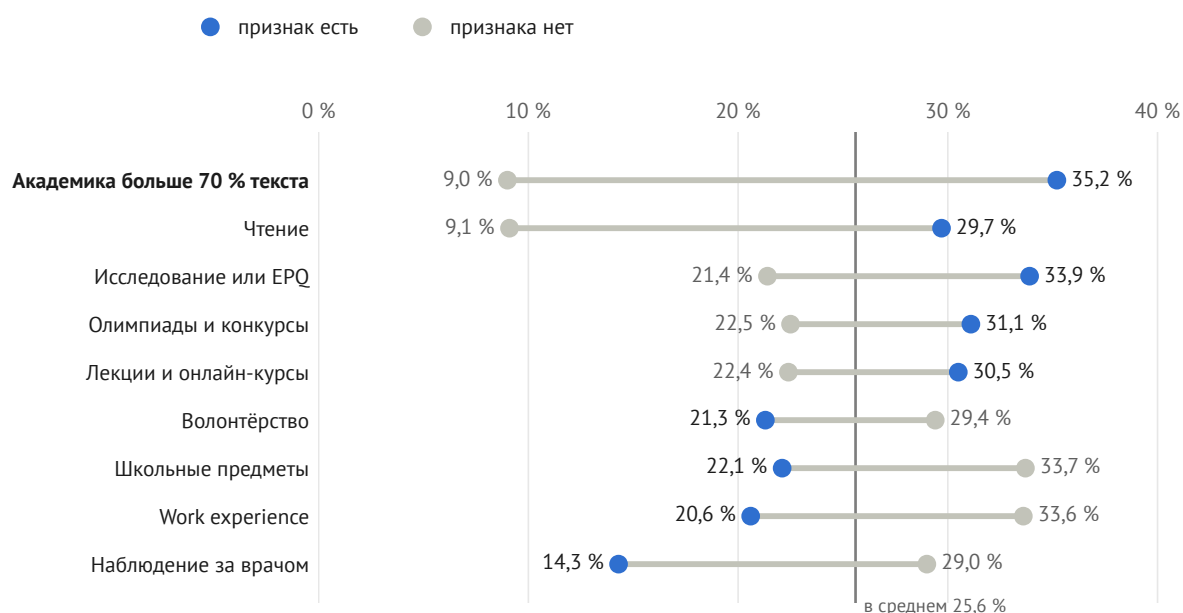


Рис. 10. В UCAS выигрывают академика и чтение, а не work experience. Доля сильных statements (сумма оценок ≥ 10) с признаком и без него. Источник: разметка корпуса, 332 UCAS statements.

Таблица 10. Признаки сильных UCAS statements: логистическая регрессия (332 statements)

Признак	Есть у	ОШ (95 % ДИ)	ОШ с контролем предмета
Академика больше 70 % текста	63 %	3,55 (1,69–7,47)	4,11 (1,90–8,88)
Начало не прямым тезисом	49 %	2,01 (1,15–3,51)	1,92 (1,07–3,44)
Финал не пересказом	69 %	1,52 (0,80–2,90)	1,43 (0,72–2,82)
Чтение	80 %	3,11 (1,21–8,02)	2,43 (0,90–6,58)
Исследование или EPQ	34 %	1,45 (0,79–2,65)	1,42 (0,76–2,66)
Олимпиады и конкурсы	36 %	1,23 (0,69–2,20)	1,79 (0,94–3,38)
Лекции и онлайн-курсы	40 %	0,84 (0,47–1,50)	1,00 (0,55–1,83)
Волонтерство	47 %	0,99 (0,56–1,75)	1,06 (0,58–1,95)
Школьные предметы	70 %	0,50 (0,25–0,99)	0,48 (0,24–0,97)
Work experience	61 %	0,52 (0,29–0,92)	0,62 (0,34–1,13)
Наблюдение за врачом	23 %	0,35 (0,15–0,82)	0,50 (0,19–1,27)

Обе модели включают все признаки таблицы, внеучебные навыки и признак разбивки на три вопроса (по сути — источник The Uni Guide). Первая добавляет логарифм длины, вторая — группу предметов (гуманитарные, социальные, STEM, медицина, прочие). Псевдо- R^2 — 0,17 и 0,21. Жирным выделены ОШ, чей 95 % ДИ не включает 1.

Медицина. Медицинские statements слабее остальных: сильных 11,5 % против 17,1 % в STEM, 37,8 % в социальных науках и 42,9 % в гуманитарных. Дело не в предмете: с контролем доказательств и доли академики медицина от STEM не отличается (ОШ = 1,29; 0,40–4,10). Дело в устройстве текста.

- В медицинских statements меньше академики: больше 70 % текста она занимает у 49 % против 62–74 % в остальных крупных группах.
- В них намного чаще есть наблюдение за врачами: 72 % против 7–14 %.

Поэтому с контролем предмета отрицательный эффект work experience и наблюдения за врачом слабее и теряет значимость. Но и без медицины statements с work experience слабее: 22,3 % сильных против 39,0 % ($p = 0,004$). Они отдают меньше места самому предмету: академика преобладает у 59 % из них, а среди statements с исследованием — у 75 %.

Устойчивость. Линейная регрессия суммы оценок с контролем предмета ($R^2 = 0,39$)

подтверждает пять эффектов:

- академика больше 70 % — +1,12 балла ($p < 0,001$);
- исследование или EPQ — +0,41 ($p = 0,005$);
- начало не прямым тезисом — +0,32 ($p = 0,01$);
- олимпиады — +0,31 ($p = 0,03$);
- школьные предметы — -0,42 ($p = 0,008$).

Эффект нового формата по этим данным оценить нельзя: из 90 statements с тремя вопросами 84 переформатировал издатель, а настоящих новых statements всего 6.

Итог для UCAS: выше оценены statements, где интерес доказан работой с предметом: что прочитано, с чем автор спорит, что проверил сам. Пример — statement по экономике в Cambridge: спор с автором книги и собственная регрессия на данных World Bank. Напомним, что «сильное» — оценка модели, а не решение вуза: оффер получили все 332 автора.

Примеры statements.

- **Сильные.** Earth Sciences, Oxford (St Anne's College; 12): каждый абзац — конкретная находка с выводом, от красного и зелёного серпентинита в Коверраке до критики того, как МакГуайр подбирает данные. English, Cambridge (UniAdmissions; 12): Леви, Гоголь и Кафка разобраны по приёмам, а личная деталь — симпатия к герою Леви из-за любви к химическим практикумам — оживляет анализ. Chinese Studies, Oxford (The Uni Guide; 12): автор спорит и с Жаком, и с Хаттоном и честно оставляет вопрос о Мао открытым.
- **Слабые.** Medicine, UCL (6med; 7): предпоследний абзац — перечень достижений и качеств без примеров, вывод о призвании уместается в одну фразу. Graduate-entry medicine, Sunderland (6med; 7): наблюдение за врачами, работа в пандемию и исследование перечислены без рефлексии и без объяснения, почему выбран именно этот путь.

11. Как менялись эссе в официальных подборках (RQ7)

За 2007–2030 годы выпуска вкус официальных подборок заметно сместился. Юмора стало вдвое меньше, эссе вокруг одного предмета — вдвое больше, а длина уперлась в лимит Common App. Доля тяжёлых историй снижалась, но этот тренд статистически не подтверждается.

Данные. 268 основных эссе из официальных подборок с указанным годом выпуска: Johns Hopkins — 95, Connecticut College — 46, Hamilton — 38, St. John's — 34, Tufts — 26, ещё 29 из пяти вузов. Эссе сгруппированы в три периода:

- до 2017 года выпуска (2007–2017, 66 эссе);
- 2018–2022 (99 эссе) — первые выпуски, писавшие при лимите Common App в 650 слов;
- 2023–2030 (103 эссе).

Тренд каждого признака оценён логистической регрессией на год выпуска. Проверка устойчивости — та же модель с фиксированными эффектами источника: вузы присоединялись к корпусу в разные годы. Например, American University и Emory есть только в последнем периоде, St. John's — с 2018 года.

Юмор. Самый надёжный тренд: шансы встретить юмор падают примерно на 10 % в год. Он держится и с контролем источника, и внутри одного JHU (ОШ = 0,88; $p = 0,009$). При этом юмор по-прежнему связан с сильной оценкой: в каждом периоде эссе с юмором

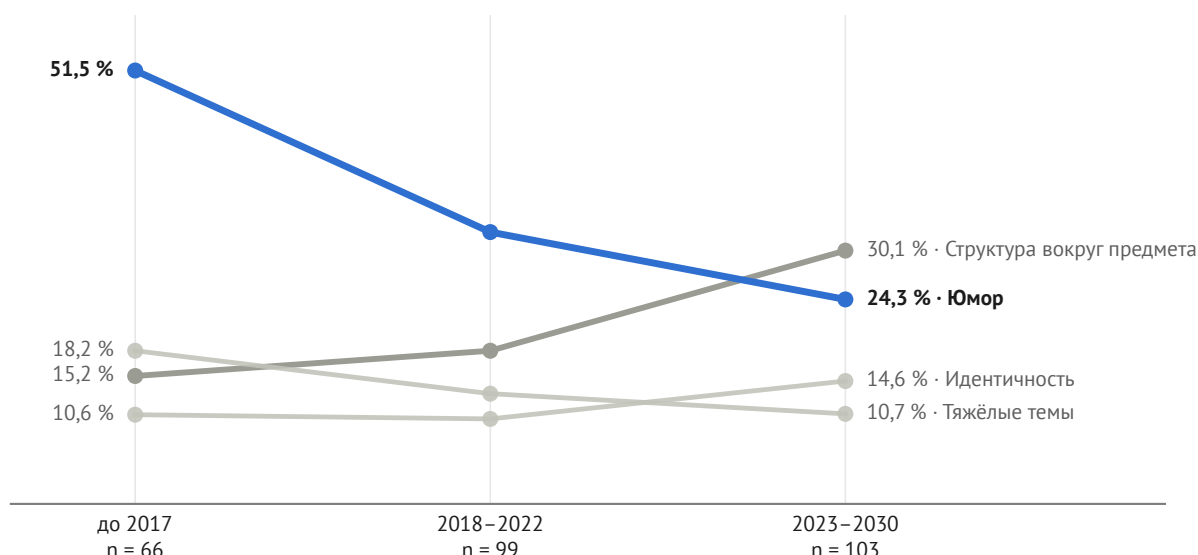


Рис. 11. Юмора в официальных подборках стало вдвое меньше: с 52 до 24 %. Доля основных эссе с признаком по году выпуска. Источник: разметка корпуса, 268 основных эссе из официальных подборок с указанным годом выпуска, 2007–2030.

Таблица 11. Тренды по году выпуска: отношение шансов за год (268 эссе)

Признак	ОШ за год (95 % ДИ)	p	С контролем источника: ОШ (p)
Юмор	0,906 (0,863–0,952)	< 0,001	0,901 (< 0,001)
Структура вокруг предмета	1,067 (1,008–1,129)	0,025	1,068 (0,032)
Тяжёлые темы	0,954 (0,896–1,016)	0,15	0,956 (0,18)
Идентичность	1,042 (0,971–1,118)	0,25	1,054 (0,19)
Хронологический рассказ	0,966 (0,920–1,014)	0,16	0,972 (0,29)
Смелое начало	0,984 (0,938–1,031)	0,50	0,978 (0,42)
Сильное эссе (оценка модели)	0,961 (0,915–1,009)	0,11	0,977 (0,41)

ОШ меньше 1 — признак с каждым годом встречается реже. Контроль источника: пять крупнейших вузов отдельно, остальные вместе.

сильны чаще на 10–14 процентных пунктов. Вузы стали реже показывать смешные эссе, а не перестали ценить те, что показывают.

Голос. Средняя оценка голоса снижается вслед за юмором: 3,80 → 3,57 → 3,40, или –0,03 балла в год ($p < 0,001$; с контролем источника $p = 0,001$). Часть снижения объясняет сам юмор: если его учесть, наклон уменьшается до –0,02 балла в год, но остаётся значимым ($p = 0,007$).

Структура вокруг предмета. Доля выросла с 15,2 до 30,1 %; тренд устойчив к составу источников. Внутри одного JHU он того же знака, но незначим (ОШ = 1,04; $p = 0,44$): 95 эссе одного вуза не хватает, чтобы его увидеть.

Тяжёлые темы и идентичность. Доля тяжёлых историй снизилась с 18,2 до 10,7 %, но в пределах шума ($p = 0,15$). У идентичности тренда нет. Популярное убеждение, что после 2020 года вузы ждут больше историй о трудностях и идентичности, на официальных подборках не подтверждается.

Длина. Медиана выросла с 577 до 650–655 слов, но сдвиг на грани значимости (тест Манна–Уитни, $p = 0,07$). Сильнее изменился разброс. До 2018 года в коридор 600–700 слов попадали 13,6 % эссе, после — 70–74 %. Стандартное отклонение упало с 230 до 164 слов (тест Левене, $p < 0,001$). Тексты не стали длиннее — они стали одинаковой длины и упираются в лимит.

Что это значит. Тренды описывают вкус редакторов подборок, а не обязательно приёмных комиссий: вуз сам решает, какие эссе показать (раздел 13). Но именно эти подборки абитуриенты читают как образец. Поэтому сдвиг к «предметным» эссе сам себя усиливает: когда все пишут вокруг предмета, выделиться приёмом труднее. Юмор идёт в обратную сторону: его стало меньше, а связь с сильной оценкой осталась.

Как выглядят оба полюса. Типичные юмористические тексты ранних подборок — «Cinco Reasons...» (Johns Hopkins, 2014) и «Corkscrew» (Johns Hopkins, 2015). Типичное «предметное» эссе последних лет — «Fried Rice in One (Not So) Easy Step» (Johns Hopkins, 2024): один семейный рецепт держит весь текст.

12. Обсуждение

Результаты складываются в одну картину: сильное эссе отличается не материалом, а то, как автор им распоряжается. Драма, престижная активность или редкая тема сами по себе оценку не повышают. Повышают её угол зрения, напряжение в начале, живой голос и финал, который замыкает текст. Ниже — как это объяснить, как это соотносится с прежними работами и что из этого следует для абитуриента.

12.1. Почему исполнение важнее материала

Три независимых наблюдения указывают в одну сторону.

- **Ставки не решают.** Доля сильных одинакова у повседневных, заметных и тяжёлых историй (раздел 5). Слабее всего эссе, где темой стала сама трудность: они чаще пересказывают беду, чем показывают автора.
- **Форма решает при равных прочих.** Структура вокруг предмета, рамка и эссе-размышление повышают шансы в 2,3–2,5 раза по сравнению с хронологией, смелое начало — в 1,9 раза, юмор — в 2,5 раза (раздел 6.5). Те же структуры выигрывают и в дополнительных эссе (раздел 10.1).
- **Словарь слабых эссе — словарь резюме.** В слабых текстах чаще встречаются team, help, skills, research и прошедшее время отчёта (раздел 7.3). Темы NMF с самой низкой долей сильных — командный спорт, служение и медицина (раздел 9.2): там проще всего перечислить заслуги.

Эти наблюдения связывает одно объяснение: хорошая форма заставляет выбирать. Центральный предмет или рамка задают угол зрения, и всё, что в него не попадает, отбрасывается. Хронология такого фильтра не даёт и тянет к пересказу: «сначала было, потом стало». Смелое утверждение в начале ставит вопрос, на который текст обязан ответить. Возврат к началу в финале показывает, что ответ найден. Эта пара даёт самую высокую долю сильных среди комбинаций начала и финала — 43 % (раздел 6.3).

12.2. Что стоит за «живым голосом»

Голос — самая субъективная из трёх оценок, но у неё есть наблюдаемые корреляты (раздел 7):

- короткие фразы среди обычных;
- разнообразие слов;
- настоящее время;
- другие люди в кадре;
- юмор.

Это близко к тому, что практики называют «показывай, а не рассказывай»: сцена и реплика вместо обобщения.

Результат расходится с работой [Pennebaker et al. \(2014\)](#), но не противоречит ей. Там «категориальный» стиль — больше артиклей и предлогов — предсказывал университетские оценки. У нас сильные эссе не категориальнее остальных, а предлогов в них даже меньше (раздел 7.3). Эссе и учебная работа вознаграждают разное. В эссе читатель ищет человека, в учебной работе — аналитическую дисциплину. Стиль, полезный для успеваемости, не обязательно делает текст живым. Осторожность: у нас исход — оценка модели, а не будущие оценки авторов, поэтому напрямую сравнить два результата нельзя.

12.3. Два читателя: комиссия и консультант

Разборы показывают, что приёмные комиссии и консультанты читают эссе по-разному (раздел 8). Комиссии называют качества человека: характер, рост, любознательность, вклад в сообщество. Консультанты называют приёмы: детали, структуру, начало. Соответствие вузу обе стороны ценят одинаково.

Это не противоречие, а разделение ролей. Для комиссии приём — средство: он ценен, если что-то показывает об авторе. Консультант учит самому приёму, потому что именно его можно передать. Для абитуриента вывод прост: приём, который ничего не говорит о вас, не работает, как бы хорошо он ни был выполнен. Оговорка: 102 из 131 разбора комиссий написаны Johns Hopkins.

12.4. Ремесло и неравенство

[Alvero et al. \(2021\)](#) показали, что содержание и стиль эссе связаны с доходом семьи сильнее, чем балл SAT. Наши результаты подсказывают один из возможных механизмов этой связи. Сильное эссе — вопрос ремесла: композиции, начала, ритма. Ремеслу учат, а доступ к такой учёбе неравен. К тому же консультационные библиотеки в корпусе переоценивают трагические истории: 21 % против 14 % в подборках вузов (раздел 5). Абитуриент, который учится на них, получает искажённый образец.

Открытая разметка ремесла — один из способов выровнять доступ: приёмы, которые раньше передавались через платных консультантов, становятся видны любому. [Lira et al. \(2023\)](#) показали, что модели распознают личные качества в эссе и эти оценки предсказывают выпуск. Та же технология может давать абитуриенту обратную связь о ремесле.

У этого есть цена, и её видно уже в разделе 11: когда приём становится нормой, он перестаёт выделять. Доля эссе вокруг предмета в официальных подборках удвоилась, и «маленький предмет» рискует стать новым шаблоном. Образцы полезны как инструмент думания о своём материале, а не как форма для заполнения.

12.5. Практические выводы для абитуриента

Выводы ниже следуют из ассоциаций, а не из экспериментов. Но каждый опирается на несколько наблюдений в корпусе.

1. **Не ищите трагедию.** Повседневный материал работает не хуже тяжёлого; решает угол зрения, а не масштаб события (разделы 5 и 9).
2. **Найдите «ключ».** Предмет, привычка или образ, через который видно, как вы думаете, — самый частый и самый сильный тип истории. Держите его до последнего абзаца (разделы 6.2 и 9.1).
3. **Начните с утверждения, которое нужно доказать.** Не «Я вырос в...» и не прямой тезис: первая фраза должна ставить вопрос, а не пересказывать анкету (раздел 6.1).

4. **Закончите возвратом.** Вернитесь к образу или фразе из начала, но с новым смыслом. Пересказ — единственный тип финала без единого сильного эссе (раздел 6.3).
5. **Пишите о мысли, а не о заслугах.** Слова вроде team, skills, help, research — признак отчёта. Покажите сцену, другого человека, конкретный предмет (раздел 7.3).
6. **Юмор — если он ваш.** Это самый сильный стилиевой признак, а в официальных подборках его становится всё меньше (разделы 6.5 и 11). Ритм важнее длины фраз: делайте короткую фразу среди обычных (раздел 7.1).
7. **Используйте лимит.** Короткие основные эссе чаще недоразвиты: каждые +10 % длины увеличивают шансы в 1,23 раза (раздел 6.5).
8. **Supplement — маленькое эссе, а не анкета.** Рамка, метафора или монтаж работают и в 250 словах. «Why us» начинайте с личного вопроса, на который отвечает вуз, а не с «I want to attend X because...» (раздел 10.1).
9. **UCAS — про предмет.** Больше 70 % текста — о предмете: что вы прочитали, с чем спорите, что проверили сами. Work experience и школьная программа работают, только если из них следует вывод о предмете (раздел 10.2).
10. **Вычёркивайте клише, но не бойтесь их.** «I realized that» и «my passion for» оценку не снижают, но и ничего не добавляют (раздел 7.4).
11. **Проверьте черновик по архетипу.** Если он выходит «портретом характера» или «делом для других», ищите сцену или предмет, который соберёт текст (раздел 9.1).

13. Ограничения и угрозы валидности

Работа описывает, как устроены опубликованные эссе поступивших и какие из них модель оценивает выше. Она не показывает, что именно решает поступление. Ниже — ограничения, сгруппированные по тому, какую часть выводов они затрагивают.

13.1. Отбор: кого мы видим

- **Только принятые и опубликованные.** В корпусе нет отклонённых эссе, поэтому работа не может сказать, чем поступившие отличаются от непоступивших. Сравнение «сильные против остальных» идёт внутри успешных текстов и по оценке модели. Эссе — лишь часть заявки: оценки, тесты, рекомендации и приоритеты вуза нам не видны.
- **Отбор издателей.** Вузы и консультанты сами решают, что публиковать. Официальные подборки отражают вкус редакторов (раздел 11), библиотеки консультантов — их методику. Консультант может публиковать те эссе, которые иллюстрируют его же советы, и тогда логика замыкается: «работает то, что советуют». Частичная защита — контроль уровня верификации в регрессии (раздел 6.5).
- **Перекося источников.** Johns Hopkins даёт 102 из 131 разбора приёмных комиссий и 95 из 268 эссе в анализе трендов. Выводы о «языке комиссий» описывают прежде всего один вуз.
- **Исход не подтверждён у 849 эссе.** Они входят в анализ качества с контролем уровня, но не в распределения тем. Происхождение 59 эссе AdmitReport не документировано. Сайт не называет ни авторов, ни исходов и лишь утверждает, что примеры взяты из реальных заявок; часть текстов может быть синтетической.
- **География и язык.** Только англоязычные эссе для вузов США и Великобритании. Переносимость на другие системы приёма не проверялась.

13.2. Измерение: что значит «сильное»

- **Модель вместо людей.** Все метки и оценки поставила одна модель по одной инструкции. Повторная разметка показывает, что метки воспроизводимы (раздел 4.5), но не что они верны. Сравнения с экспертами-людьми нет.
- **Общий источник признаков и исхода.** Признаки и оценки ставит один аннотатор. Проверка между проходами исключает гало-эффект внутри прогона, но не общие пристрастия модели (раздел 4.5). Если модель, вслед за пособиями, считает метафору или юмор признаком хорошего текста, она переносит это убеждение и в оценки. Частичная защита — стилевые метрики, посчитанные без модели, и темы NMF: оба независимых метода сходятся с разметкой (разделы 7.1 и 9.2).
- **Рубрика определяет исход.** «Сильное» — это сумма оценок конкретики, рефлексии и голоса. Другая рубрика — например, с акцентом на достижения или соответствие вузу — дала бы другой исход. Наш исход ближе к качеству письма, чем к решению комиссии.
- **Порог.** Бинарная метка «сильное» воспроизводится хуже непрерывной суммы: $\kappa = 0,56$ на основных эссе против $ICC = 0,83$. Поэтому ключевые модели продублированы линейной регрессией суммы оценок (разделы 6.5 и 10.2).

13.3. Анализ

- **Ассоциации, а не причины.** Все результаты наблюдательные. Эссе со смелым началом могут быть сильнее не из-за начала, а потому, что их авторы вообще лучше пишут. Регрессия контролирует наблюдаемые признаки, но не общее умение автора.
- **Множественные сравнения.** Поправок не вводилось (раздел 4.6). Предварительными стоит считать результаты с p между 0,01 и 0,05: связь финала с качеством ($p = 0,025$), слабость эссе о самой трудности ($p = 0,041$), эффект смелого начала с контролем ($p = 0,02$), рост структуры вокруг предмета в подборках ($p = 0,025$) и часть эффектов в дополнительных эссе и UCAS.
- **Малые группы.** Многие категории насчитывают 10–30 эссе, и их интервалы широки. Группы меньше 10 эссе в таблицы не включены.
- **Годы.** Год выпуска известен лишь для 414 эссе. Тренды оценены на 268 эссе официальных подборок, состав которых меняется со временем (раздел 11).

13.4. Тексты

- **Переформатирование.** The Uni Guide переложила свои statements в формат из трёх вопросов и местами дублирует или перефразирует текст. 18 явно испорченных и 3 дубликата исключены, но оставшиеся 84 могут немного отличаться от поданных. По той же причине нельзя оценить эффект нового формата UCAS (раздел 10.2).
- **Архив.** 194 эссе восстановлены из Wayback Machine, и их форматирование (курсив, абзацы) могло пострадать. Найденные дефекты исправлены вручную (раздел 3.3), но проверка не была сплошной.
- **Дедупликация.** Порог в 60 % общих наборов из 8 слов ловит перепечатки, но не пересказы. Редактированные версии одного эссе могли остаться отдельными файлами.

13.5. Этика

Эссе опубликованы открыто, но написаны школьниками, и многие из них — о болезни, семье и идентичности. Мы не перепубликуем тексты и не пытаемся установить авторов: корпус хранится локально для учебного анализа, а в статье приводятся только краткие пересказы. Эссе с пометкой «не для публичного использования» исключено.

14. Заключение

Мы собрали 2 282 опубликованных вступительных эссе, разметили их по кодбуку писательского ремесла и проверили, что разметка воспроизводима. Главный вывод: сильные эссе отличает исполнение, а не материал. Тяжёлый опыт лежит в основе шестой части подтверждённых эссе и оценку не повышает. Повышают её структура вокруг одного предмета или в рамке, смелое первое предложение, юмор и финал, который возвращается к началу. Тот же принцип — показывать ход мысли, а не перечень — работает в дополнительных эссе и в UCAS statements. В британском жанре сильнее всего доля академике и работа с прочитанным.

Работа добавляет три вещи: корпус с уровнями верификации, кодбук ремесла с оценкой надёжности и проверку популярных советов на большой выборке. Часть советов подтверждается: показывать, а не рассказывать, и не писать о трагедии ради трагедии. Часть — нет: начало со сцены, самый расхожий совет, даёт вдвое меньше сильных эссе, чем смелое утверждение.

Следующие шаги вытекают из ограничений (раздел 13):

- оценка подвыборки экспертами-людьми, чтобы проверить оценки модели;
- эксперимент с читателями: одно и то же эссе с разными началами и финалами, чтобы отделить эффект приёма от умения автора;
- сравнение с отклонёнными эссе, если такой корпус станет доступен;
- разборы других приёмных комиссий, чтобы снять перекос в сторону Johns Hopkins;
- наблюдение за тем, не превращается ли «маленький предмет» в новый шаблон.

Для абитуриента вывод уместается в одну фразу: не ищите историю больше, чем у других, — ищите угол, под которым видно, как вы думаете.

Список литературы

Выходные данные и ссылки проверены 6 октября 2026 года. Источники самих эссе перечислены в приложении Г.

- A. J. Alvero, S. Giebel, B. Gebre-Medhin, A. L. Antonio, M. L. Stevens, and B. W. Domingue. Essay content and style are strongly related to household income and SAT scores: Evidence from 60,000 undergraduate applications. *Science Advances*, 7(42):eabi9031, 2021. doi: [10.1126/sciadv.abi9031](https://doi.org/10.1126/sciadv.abi9031).
- J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960. doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104).
- College Essay Advisors. The new 2013–14 Common Application. Блог College Essay Advisors, 2013. URL <https://www.collegeessayadvisors.com/the-new-2013-14-common-application/>.
- F. Gilardi, M. Alizadeh, and M. Kubli. ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30):e2305016120, 2023. doi: [10.1073/pnas.2305016120](https://doi.org/10.1073/pnas.2305016120).

- J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977. doi: [10.2307/2529310](https://doi.org/10.2307/2529310).
- D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999. doi: [10.1038/44565](https://doi.org/10.1038/44565).
- B. Lira, M. Gardner, A. Quirk, C. Stone, A. Rao, L. Ungar, S. Hutt, L. Hickman, S. K. D’Mello, and A. L. Duckworth. Using artificial intelligence to assess personal qualities in college admissions. *Science Advances*, 9(41):eadg9405, 2023. doi: [10.1126/sciadv.adg9405](https://doi.org/10.1126/sciadv.adg9405).
- K. O. McGraw and S. P. Wong. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1):30–46, 1996. doi: [10.1037/1082-989X.1.1.30](https://doi.org/10.1037/1082-989X.1.1.30).
- B. L. Monroe, M. P. Colaresi, and K. M. Quinn. Fightin’ words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4):372–403, 2008. doi: [10.1093/pan/mpn018](https://doi.org/10.1093/pan/mpn018).
- J. W. Pennebaker, C. K. Chung, J. Frazee, G. M. Lavergne, and D. I. Beaver. When small words foretell academic success: The case of college admissions essays. *PLoS ONE*, 9(12):e115844, 2014. doi: [10.1371/journal.pone.0115844](https://doi.org/10.1371/journal.pone.0115844).
- M. V. Reiss. Testing the reliability of ChatGPT for text annotation and classification: A cautionary remark. arXiv:2304.11085, 2023. URL <https://arxiv.org/abs/2304.11085>.
- E. Sawyer. *College Essay Essentials: A Step-by-Step Guide to Writing a Successful College Admissions Essay*. Sourcebooks, Naperville, IL, 2016.
- UCAS. How to write your personal statement: 2026 entry onwards. Сайт UCAS, 6. г. URL <https://www.ucas.com/applying/applying-to-university/writing-your-personal-statement/how-to-write-your-personal-statement-for-2026-entry-onwards>. Дата обращения: 6 октября 2026 г.
- E. B. Wilson. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212, 1927. doi: [10.1080/01621459.1927.10502953](https://doi.org/10.1080/01621459.1927.10502953).

Приложение А. Кодбук, правила архетипов и клише

А.1. Значения полей. Ниже — все допустимые значения в порядке инструкции с русскими названиями из статьи; для полей построения в скобках дан код. Полный текст инструкции с двумя разобранными примерами — в файле `analysis/scripts/annotation_guide.md`.

- **Тема (26).** Семья · идентичность и культура · переезд и язык · религия и вера · гендер и сексуальность · здоровье и болезнь (свои или близких, включая психическое здоровье) · утрата · деньги и работа · спорт · музыка, театр, танец · искусство, тексты, медиа · наука и технологии · книги и идеи (нетехнические интеллектуальные интересы) · хобби и странности (необычное хобби, предмет, привычка, коллекция) · еда и кулинария · место, дом, природа · путешествия · волонтерство и сообщество · активизм и политика · клубы и лидерство · дружба и отношения · неудача и ошибка (не связанная со здоровьем) · медицина и профессия (мотивация к профессии) · учебный предмет («why major») · почему этот вуз · другое.
- **Группы тем (рис. 1).** Близкие люди — семья, утрата, дружба · идентичность и корни — идентичность и культура, переезд и язык, религия, гендер · трудности — здоровье, деньги и работа, неудача · спорт · творчество — музыка и сцена, искусство и тексты · наука и идеи — наука, книги и идеи, учебный предмет · повседневность и хобби — хобби, еда, место, путешествия · сообщество и лидерство — волонтерство, активизм, клубы · профессия · почему этот вуз · другое.

- **Структура (11).** Хронологический рассказ (single_narrative) — одна история примерно по порядку · история в рамке (framed_narrative) — начало в моменте, флешбэк, возврат · монтаж эпизодов (montage) — несколько эпизодов с общей нитью · вокруг предмета или метафоры (central_metaphor) · список или фрагменты (segmented_list) — явный список, нумерованные или озаглавленные части · эссе-размышление (idea_exploration) — вопрос или идея, мало сюжета · необычная форма (unconventional_form) — письмо, рецепт, сценарий, инструкция, словарная статья · мозаика «кто я» (identity_mosaic) — портрет через грани, роли и противоречия · академический путь (academic_journey) — интерес → углубление → выводы → будущее · «Why us» (why_us_mapping) — мои интересы ↔ ресурсы вуза · короткий ответ (short_answer) — меньше ~120 слов без развёрнутой структуры.
- **Начало (9), первые одно-два предложения.** Сцена, действие (scene_action) · образ, описание (vivid_description) · реплика или цитата (dialogue_quote) · смелое утверждение (bold_claim) — неожиданное, провокационное или смешное · вопрос (question) · список, фрагменты (list_fragments) · факт или определение (fact_definition) · спокойный контекст (background_context) — «Я вырос в...» · прямой тезис (direct_thesis) — «Медицина привлекает меня, потому что...».
- **Финал (9), последний абзац.** Возврат к началу (callback) — к образу или фразе из начала с новым смыслом · взгляд в будущее (forward_looking) · вывод, урок (insight) · финальная сцена или образ без явной морали (image_scene) · развязка метафоры (metaphor_payoff) · связь с вузом (college_link) · пересказ (summary_restatement) · открытый финал (open_ended) · шутка, твист (humor_twist).
- **Сюжетная арка (10).** Трудность → рост (challenge_growth) · как родилось увлечение (discovery_passion) · смена взгляда (perspective_shift) · принятие себя (identity_reconciliation) · ценности в поступках (values_in_action) · портрет мышления (portrait_of_mind) — какой я, без центрального конфликта · что я сделал для других (impact_service) · урок от человека (relationship_lesson) · погоня за вопросом (intellectual_inquiry) · обоснование выбора (motivation_case) — аргументы в пользу курса или профессии.
- **Ставки (3).** Повседневное (everyday) — обычная жизнь и увлечения · заметное событие (significant) — серьёзный вызов без травмы · тяжёлый опыт (hardship) — болезнь, смерть, бедность, насилие, дискриминация, вынужденный переезд, война.
- **Временной охват (3).** Один момент — сцена или день (moment) · недели-месяцы (period) · годы (years).
- **Доля рефлексии (3).** Мало — меньше 20 % текста (low) · баланс — 20–50 % (balanced) · много — больше 50 % (high).
- **Качества (16; 2–3 на эссе, только показанные, а не заявленные).** Любознательность · стойкость · эмпатия · инициатива и лидерство · креативность · юмор · самоосознание · ответственность · открытость · командность · упорство · интеллектуальная энергия · мост между культурами · самостоятельность · принципиальность · служение другим.
- **Тон (10; 1–2 на эссе).** Искренний · с юмором · лиричный · аналитичный · уязвимый · игривый · напряжённый · тёплый · уверенный · академичный.
- **Приёмы (17; 0–4 на эссе, только заметные).** Сквозной мотив · чувственные детали · диалог · юмор · контраст · конкретные числа · настоящее время · обращение на «ты» · ритм предложений (намеренное чередование коротких и длинных, фрагменты) · списки · флешбэк · самоирония · имена и названия · отсылки к книгам и науке · риторические вопросы · необычный формат · метафоры и образы.
- **Критерии разбора (20; только если у эссе есть разбор издателя).** Голос и личность · конкретные детали · рефлексия и инсайт · рост и изменение · любознатель-

ность · характер и ценности · вклад в сообщество · соответствие вузу · юмор · оригинальность · структура · сильное начало · сильный финал · показывает, а не рассказывает · честность и уязвимость · фокус на авторе (например, когда эссе о другом человеке) · связь с будущим · академическая глубина · необычная тема · качество письма.

- **Только UCAS.** Доля академички: мало, средне или больше 70 % текста. Типы доказательств (9): чтение · лекции и онлайн-курсы · олимпиады и конкурсы · исследование или EPQ · work experience · наблюдение за врачом или профессионалом · волонтерство · школьные предметы · внеучебные навыки.
- **Свободный текст на русском.** Суть — одна строка до 22 слов · пересказ — 3–5 предложений · главная мысль — до 12 слов · теги — 2–5 · сила · слабое место.

Таблица А1. Якоря оценок качества

Балл	Конкретика	Рефлексия	Голос
1	общие утверждения	нет или мораль-клише	обезличенный, формальный
2	мало деталей	заявлена, но поверхностна	грамотный, но безличный
3	несколько конкретных деталей	ясная, но ожидаемая	есть личность
4	конкретика на протяжении всего текста	вдумчивая, с нюансами	выраженная личность
5	живые точные детали, которые мог дать только этот автор	неожиданная, самокритичная, сложная	безошибочно узнаваемый, запоминающийся

Инструкция предупреждает: большинство эссе получает 3–4, пятёрка — только для исключительных текстов. Короткие ответы меньше 120 слов оцениваются относительно их длины.

А.2. Правила архетипов. Правила применяются по порядку; эссе получает первый подошедший архетип.

Таблица А2. Правила назначения архетипа

№	Архетип	Условие
1	Обоснование выбора	UCAS; или арка «обоснование выбора»; или структура «академический путь» или «Why us»
2	Испытание	ставки «тяжёлый опыт»
3	Корни и идентичность	арка «принятие себя» или группа тем «идентичность и корни»
4	Маленький предмет — большой смысл	структура «вокруг предмета или метафоры»
5	Мозаика	структура «монтаж», «мозаика кто я», «список» или «необычная форма»
6	Ум в действии	структура «эссе-размышление» или арка «погоня за вопросом»
7	Человек рядом	арка «урок от человека» или группа тем «близкие люди»
8	Дело для других	арка «что я сделал для других» или «ценности в поступках»; или группа тем «сообщество и лидерство»
9	Один момент	хронология или рамка при охвате «один момент» или «недели–месяцы»
10	История увлечения	арка «как родилось увлечение»
11	Портрет характера	всё остальное

А.3. Клише. Двадцать шаблонов искались регулярными выражениями без учёта регистра: «ever since I was a child / young / little» · «since / from a young age» · «for as long as I can remember» · «make a (positive) difference» · «(step) out of my comfort zone» · «changed

my life» · «broaden(ed) my horizons» · «hard work pays off» · «never give up» · «follow(ing) my dreams / passion» · «think(ing) outside the box» · «at the end of the day» · «Webster / dictionary defines» · «the rest is history» · «taught me the value / importance of» · «I have always been passionate / fascinated / interested» · «my passion for» · «a valuable lesson» · «opened my eyes» · «I realized that».

Приложение Б. Надёжность по отдельным меткам

Таблица Б1. Согласие двух проходов по категориям, на которых держатся главные выводы (140 эссе)

Категория	Эссе в 1-м / 2-м проходе	В обоих	κ (95 % ДИ)
Смелое утверждение (начало)	25 / 32	24	0,80 (0,67–0,92)
Сцена (начало)	26 / 24	23	0,90 (0,79–0,98)
Прямой тезис (начало)	34 / 31	29	0,86 (0,74–0,95)
Вокруг предмета или метафоры	17 / 17	12	0,67 (0,44–0,84)
Хронологический рассказ	29 / 31	27	0,87 (0,76–0,96)
История в рамке	12 / 7	6	0,61 (0,28–0,83)
Возврат к началу (финал)	15 / 15	13	0,85 (0,67–0,97)
Пересказ (финал)	10 / 14	10	0,82 (0,61–0,96)
Развязка метафоры (финал)	10 / 11	9	0,85 (0,64–1,00)
Тяжёлый опыт (ставки)	10 / 10	9	0,89 (0,69–1,00)
Юмор (приём или тон)	24 / 27	24	0,93 (0,83–1,00)

Поля с несколькими метками. В скобках — средний коэффициент Жаккара и медиана κ ; далее — κ по каждой метке, которая встречается хотя бы в 10 % эссе одного из проходов.

- **Качества (0,78; 0,82).** Инициатива и лидерство 0,91 · мост между культурами 0,88 · интеллектуальная энергия 0,86 · юмор 0,85 · креативность 0,83 · стойкость 0,82 · любознательность 0,81 · упорство 0,80 · эмпатия 0,78 · самоосознание 0,75 · служение другим 0,72.
- **Тон (0,75; 0,72).** Уязвимый 0,80 · академичный 0,76 · игривый 0,76 · искренний 0,72 · тёплый 0,69 · уверенный 0,64 · аналитичный 0,55.
- **Приёмы (0,80; 0,84).** Юмор 0,90 · конкретные числа 0,89 · отсылки к книгам и науке 0,87 · метафоры и образы 0,84 · сквозной мотив 0,84 · чувственные детали 0,84 · имена и названия 0,74 · контраст 0,71.
- **Вторичные темы (0,68; 0,76).** Наука и технологии 0,83 · волонтерство и сообщество 0,80 · книги и идеи 0,77 · активизм и политика 0,75 · идентичность и культура 0,66 · семья 0,66.
- **Критерии разбора, 65 эссе с разбором (0,82; 0,80).** Голос и личность 1,00 · соответствие вузу 1,00 · сильное начало 0,90 · характер и ценности 0,88 · конкретные детали 0,87 · показывает, а не рассказывает 0,83 · структура 0,80 · качество письма 0,80 · рост 0,77 · фокус на авторе 0,77 · честность 0,77 · оригинальность 0,69 · рефлексия 0,67.

Приложение В. Воспроизводимость

- **Сбор (5 октября 2026 года).** Python 3.12, requests и BeautifulSoup. Удалённые страницы взяты из Wayback Machine через CDX API. Тексты сохранены дословно, мета-

данные — в шапке каждого файла.

- **Дедупликация.** Множества наборов из 8 слов подряд; копия — если не менее 60 % наборов короткого текста встречаются в длинном.
- **Разметка.** Claude Sonnet, инструкция `annotation_guide.md`, 33 пачки по 70 эссе в случайном порядке, валидатор `validate_ann.py`.
- **Повторный проход.** Выборка — `random.Random(2026).sample(..., 140)`; бутстреп — `numpy.random.default_rng(2026)`, 2 000 повторов.
- **Стилевые метрики.** Предложения делятся по знакам . ! ? ...; короткая фраза — до 7 слов, длинная — от 30. MATTR считается окном в 100 слов с шагом 10. «Я» — это I, me, my, mine, myself на 100 слов. Индексы Флеша и Флеша–Кинкейда считаются по стандартным формулам, слоги — по эвристике гласных.
- **Модели.** Логистическая и линейная регрессия — `statsmodels`; редкие уровни объединены в «другое». NMF — `scikit-learn`: $K = 14$, инициализация NNDSVDA, зерно 1. TF-IDF: слова из трёх и более букв, не менее чем в 8 и не более чем в 40 % эссе, без английских стоп-слов, сублинейный `tf`. Взвешенный логарифм отношения шансов (Monroe et al., 2008): априорное распределение — частоты слов во всех 799 основных эссе с суммарным весом 1 000.
- **Версии.** Python 3.12, pandas 3.0, numpy 2.5, SciPy 1.18, statsmodels 0.15, scikit-learn 1.9; рисунки — matplotlib 3.11.
- **Файлы.** Разметка — `analysis/data/annotations.jsonl`; статистика статьи и второй проход разметки — `analysis/data/paper/`; скрипты и инструкция по запуску — `analysis/scripts/`; исходники этой статьи и скрипт рисунков — `paper/`. Тексты эссе хранятся локально и не распространяются.

Приложение Г. Источники корпуса

Число эссе после дедупликации. Пять издателей встречаются в двух уровнях, если часть их эссе подтверждена, а часть нет; всего источников 46.

- **Сайты вузов (12 источников, 398 эссе).** Tufts University — 116 · Johns Hopkins University — 102 · Connecticut College — 49 · St. John's College — 41 · Hamilton College — 39 · MIT Admissions blog — 11 · American University — 11 · University of Georgia — 9 · Olin College of Engineering — 8 · Emory University — 7 · Siena College — 3 · Oberlin College — 2.
- **СМИ и консультанты, поступление указано (21 источник, 703 эссе).** CollegeAdvisor — 205 · EssaysThatWorked.com — 199 · CollegeVine — 80 · The Harvard Crimson — 78 · PrepMaven — 39 · Shemmassian Academic Consulting — 22 · PrepScholar — 20 · College Confidential — 14 · Veridian College Prep — 13 · College Essay Advisors — 9 · Philadelphia magazine — 6 · The Stanford Daily — 5 · Ocean Educational Consulting — 3 · The Admissions Guru — 2 · Her Campus — 2 · U.S. News & World Report, The Tab, NBC News, HuffPost, Forbes, BuzzFeed — по 1.
- **Великобритания, оффер указан (11 источников, 332 statements).** Personal Statement Examples (PSE) — 107 · The Uni Guide — 84 · College Lah — 39 · UniAdmissions — 30 · 6med — 27 · St Anne's College, Oxford — 18 · Studential — 9 · Young Economist Initiative — 6 · The Medic Portal — 6 · Blackstone Tutors — 5 · Oxbridge Interview Questions — 1.
- **Поступление не указано (7 источников, 849 эссе).** College Essay Guy — 550 · CollegeVine — 227 · AdmitReport — 59 · CollegeAdvisor — 6 · College Essay Advisors — 3 · PrepScholar — 2 · PrepMaven — 2.

Приложение Д. Эссе из примеров

Все эссе, упомянутые в примерах, есть в библиотеке сайта essays.archik.tech: ссылка ведёт на страницу эссе с разбором и ссылкой на оригинал.

- Сайты вузов: отказ от телевизора — essays.archik.tech/library/?e=b5fddb2a4e
- Сайты вузов: свернувшись калачиком — essays.archik.tech/library/?e=c3d3c21ad8
- Сайты вузов: дом прабабушки в Греции — essays.archik.tech/library/?e=da8ca032c0
- Сайты вузов: 11 сентября — essays.archik.tech/library/?e=2b358261ad
- Сайты вузов: закусочная отца — essays.archik.tech/library/?e=7c3edd477d
- Сайты вузов: стендап — essays.archik.tech/library/?e=f8a0482f4e
- Сайты вузов: от кардиологии к писательству — essays.archik.tech/library/?e=6be3ad329a
- Сайты вузов: «Piece by Piece» — essays.archik.tech/library/?e=025ca8bed4
- Сайты вузов: брейкданс по YouTube — essays.archik.tech/library/?e=b9220ac788
- Сайты вузов: любовь к кино — essays.archik.tech/library/?e=d83d66a8e0
- Сайты вузов: эссе о том, как трудно написать эссе — essays.archik.tech/library/?e=1b3e5d6cea
- Сайты вузов: шнурок прадеда — essays.archik.tech/library/?e=3e53e63534
- Сайты вузов: нахлыст — essays.archik.tech/library/?e=63829e0db7
- Сайты вузов: мамин маникюрный салон — essays.archik.tech/library/?e=df430a5f85
- Сайты вузов: ирландские танцы — essays.archik.tech/library/?e=c76722cfa0
- Сайты вузов: «My Bookshelf» — essays.archik.tech/library/?e=aa3104cf81
- Сайты вузов: «Cinco Reasons...» — essays.archik.tech/library/?e=c8ecd5ed93
- Сайты вузов: «My Best Kept Secret» — essays.archik.tech/library/?e=70a7772f9b
- Сайты вузов: «Speak» — essays.archik.tech/library/?e=78f26a83ce
- Сайты вузов: «Corkscrew» — essays.archik.tech/library/?e=c6ce705082
- Сайты вузов: «The Story of Us» — essays.archik.tech/library/?e=154fbf9055
- Сайты вузов: «Don't Be Sorry» — essays.archik.tech/library/?e=34e4ea629c
- Сайты вузов: «Dissonance» — essays.archik.tech/library/?e=ac4f291197
- Сайты вузов: «Spring Instead of Summer» — essays.archik.tech/library/?e=a786998a0a
- Сайты вузов: «Temper» — essays.archik.tech/library/?e=fade2d17fc
- Сайты вузов: «In Pursuit of the Sublime» — essays.archik.tech/library/?e=89f6d5a7c1
- Сайты вузов: «Fried Rice in One (Not So) Easy Step» — essays.archik.tech/library/?e=289eb5f01f
- Сайты вузов: «Pursuing My Passion in Research» — essays.archik.tech/library/?e=be03de22fa
- Сайты вузов: «Korean Sticky Notes» — essays.archik.tech/library/?e=14f47a8f3a
- Сайты вузов: «The Art of Imperfection» — essays.archik.tech/library/?e=17c72cf8af
- Сайты вузов: «Balancing Life; A Life of Balance» — essays.archik.tech/library/?e=ec42cbcd31
- Сайты вузов: Olin, «тикун олам» — essays.archik.tech/library/?e=65e19a981c
- Сайты вузов: любовь к книгам — essays.archik.tech/library/?e=d47b3807ea
- Сайты вузов: самоописание вместо начала — essays.archik.tech/library/?e=3c24fdb77c
- Сайты вузов: «Уловка-22» — essays.archik.tech/library/?e=40152dffce
- Сайты вузов: «Энеида» — essays.archik.tech/library/?e=c55f9187de

- Сайты вузов: «утиный тест» — essays.archik.tech/library/?e=557fc143e8
- Сайты вузов: «шутка» учителя о терроризме — essays.archik.tech/library/?e=85ffaae231
- Сайты вузов: незрячий абитуриент — essays.archik.tech/library/?e=870fa8e2e1
- Сайты вузов: школьная ферма — essays.archik.tech/library/?e=a5d39dc30c
- СМИ и консультанты: одна и та же «униформа» — essays.archik.tech/library/?e=c0bca23ffc
- СМИ и консультанты: UChicago, «Илиада» и «Одиссея» — essays.archik.tech/library/?e=34965cf88c
- СМИ и консультанты: Penn и Scrabble — essays.archik.tech/library/?e=d2ff957837
- СМИ и консультанты: «Baking With Buni» — essays.archik.tech/library/?e=30daac5940
- СМИ и консультанты: «Hope Lives In That Laughter» — essays.archik.tech/library/?e=8c6a4ae487
- СМИ и консультанты: «Itch To Understand» — essays.archik.tech/library/?e=0d7592a4fd
- СМИ и консультанты: принтер этикеток P-Touch — essays.archik.tech/library/?e=9243460019
- СМИ и консультанты: Costco — essays.archik.tech/library/?e=1baaac14cd
- СМИ и консультанты: «The Illuminati changed my life» — essays.archik.tech/library/?e=ad36d4fe15
- СМИ и консультанты: «In our house, English is not English» — essays.archik.tech/library/?e=4adf32be03
- UCAS: Graduate-entry medicine, Sunderland — essays.archik.tech/library/?e=e08232a38c
- UCAS: Medicine, UCL — essays.archik.tech/library/?e=2f882e15b5
- UCAS: Economics, Cambridge — essays.archik.tech/library/?e=b377040f9e
- UCAS: Earth Sciences, Oxford — essays.archik.tech/library/?e=9c38859098
- UCAS: Chinese Studies, Oxford — essays.archik.tech/library/?e=18b379cbc4
- UCAS: English, Cambridge — essays.archik.tech/library/?e=ff7437489d
- Поступление не указано: «1984» на двух языках — essays.archik.tech/library/?e=b5cf3f3b1c