

Метод анализа текстов на золотое сечение

Сергей Логичев, 08 февраля 2026

Реферат

В работе рассмотрен метод автоматического анализа текстов на предмет использования пропорций золотого сечения. Предложены количественные метрики оценки внутренней структуры текстов, необходимые для выполнения такого анализа. Приведены количественные результаты анализа и графические визуализации для текстов А. С. Пушкина в составе 10-томного собрания сочинений. Анализ выполнялся с помощью скриптов авторской программы xMarkup (xmarkup.sourceforge.net).

Введение

Интуитивно представляется, что любой текст является завершённым информационным объектом, обладающим широким набором различных характеристик - структурных, ритмических, семантических, эстетических и т.д. Как и в созданной художником картине, тут присутствуют свои «цвета», «игра теней», сочетание «красок», композиция и прочая. Однако нам нужны специальные волшебные «очки» или «текстоскоп», чтобы увидеть всё это. Но мы могли бы использовать для этого набор специальных метрик, визуализируемых на экране компьютера. Точно так же, как делается в стилометрии. Но в тоже время наша цель далека от целей стилометрии.

Проверим гипотезу: **можно ли оценить художественный текст на наличие свойств, которые соответствует пропорции золотого сечения?**

Напомним, что классическая пропорция золотого сечения равна

$$\Phi = (1 + \sqrt{5})/2 \approx 1.61803.$$

Например, квадрат отношения высоты пирамиды Хеопса к половине длины её основания равен ≈ 1.63 . Пропорции человеческого тела также подчиняются зависимости золотого сечения, что показал еще Леонардо да Винчи. То же наблюдается для объектов живой и не живой природы, включая рукава спиральных галактик. Со времен античности золотое сечение использовалось при создании гармонии в произведениях архитектуры, искусства, литературы, поэзии и музыки. Нас в данном контексте интересуют только литературные произведения. Можно ли применить к тексту некую формулу (метод анализа), чтобы судить о наличии в нём пропорций золотого сечения?

Следовательно, требуется найти такие количественные метрики, которые можно проверить на кратность числу Φ или его обратного значения $\phi = 1/\Phi = 0.618$. Предположим, что такие метрики связаны со структурой (внутренней «геометрией») текста.

Используемые метрики

Рассмотрим следующие метрики:

L - средняя длина синтагмы (фразы) как число входящих в неё слов,

Q - сумма квадратов относительных длин слов в синтагме.

Можно предложить следующую абстракцию, чтобы понять смысл этих метрик. Представим себе синтагму или фразу как набор векторов слов, исходящих из единого центра. Тогда отдельную фразу можно воспринимать как цветок одуванчика. «Правильный» одуванчик должен иметь определенное число семян на ножках нужной длины. Семена на ножках — это вектора слов, нормированные по суммарной длине фразы в буквах (значимых символах). Ориентация векторов нам безразлична, т. к. мы для простоты игнорируем семантические и статистические связи между словами и учитываем только их длину. Тогда Q даст оценку «формы» одуванчика как поверхности, натянутой на частокол векторов слов. Так как нам интересна оценка всего текста, а не отдельной фразы, то суммируем Q по всем фразам. Формулы для вычисления наших метрик будут следующие:

$$L = N / n$$

где N – суммарная длина текста в словах, а n – число синтагм или фраз в нем.

$$Q = \sqrt{\sum(\sum(W_{ij} / S_i)^2)}$$

где W_{ij} – число букв в составе j -го слова i -ой фразы ($j = 1 \dots k_i$; $i = 1 \dots n$),

k_i – число слов в i -ой синтагме,

S_i – длина i -ой синтагмы в буквах ($i = 1 \dots n$).

Чтобы значение Q было в строго ограниченном и наиболее компактном интервале, нормируем его по числу синтагм (n) и выполним логарифмическое преобразование:

$$Q = \ln(1 / (Q/n))$$

тогда в любом случае будем иметь $0 \leq Q < M$, где $M \neq \infty$ (т. е. M - конечная величина).

Нужно отметить, что величина Q зависит от средней длины слов и их количества в синтагме. Чем больше в тексте встречается длинных слов, тем больше будет Q .

Для небольших текстов, состоящих из очень коротких фраз, значение L будет приближаться к 1, а Q - к нулю. Для вырожденных случаев, когда текст состоит из единственного слова или междометия, например: «А!» или «Вечерело.» - Q будет равно 0.

Для небольших текстов, состоящих из очень длинных фраз, значение L будет достаточно большим, а Q - стремиться к нулю. Для вырожденных случаев, когда текст состоит из единственного (сколь угодно длинного) слова, например: «Абракадабра!» - Q будет равно 0.

Значения L тоже лежат в ограниченном интервале $[1, k]$, где k обычно < 10 . В редких случаях для текстов, состоящих из нескольких (или одной) очень длинных фраз, L будет приближаться к числу слов в тексте.

Необходимо сделать предупреждение: *предлагаемый подход окажется мало применим для текстов, в которых заведомо отсутствуют знаки препинания, по которым определяются границы отдельных синтагм.* Например, такое наблюдается в поэзии Александра Введенского. В этом случае потребуется искусственное деление текста, например, по абзацам, строфам или строкам.

Примеры исследования золотого сечения в текстах

Золотое сечение активно используется для анализа гармонии в художественных текстах и музыке. Считается, что произведения, структура которых следует этой пропорции, воспринимаются как наиболее «правильные» и эстетически завершенные.

В книге «Золотое сечение» [1] анализируется присутствие последовательности Фибоначчи (5, 8, 13, 21, 34) для числа строк в стихотворениях позднего Пушкина. В частности отмечается, что кульминация «Пиковой дамы» (сцена Германа и графини) приходится на 535-ю строку из 853, что соответствует пропорции 1.6.

В работе «Архитектоника художественной формы» [2] исследуется воплощение закона золотого сечения в поэмах Пушкина, таких как «Медный всадник», «Моцарт и Сальери» и восьмая глава «Евгения Онегина».

В работе «Когнитивный принцип "золотое сечение" в поэтической антропологии А. С. Пушкина» [3] изучается связь математических пропорций с поэтикой и стилистикой.

В публикации «Golden Ratio Poetry: Mathematical Poetry According to ϕ » [6] приведены примеры использования золотого сечения при написании художественных текстов и поэзии.

Полученные результаты

В качестве анализируемых текстов были взяты произведения из 10-томного собрания сочинений А.С. Пушкина, опубликованного в электронном виде на сайте Русской Виртуальной Библиотеки (<https://rvb.ru/pushkin/toc.htm>).

В качестве контрольных примеров дополнительно были сгенерированы два корпуса случайных текстов. Для большей интриги в одном из них длина слов в синтагмах определялась последовательностью Фибоначчи: 1,2,3,5,8. Например, что-то типа: «Я не тот люблю аперитив, а уж мой ковер красивый!» (нам тут важны только длины слов, без учета их смысла).

Обработка выполнялась с помощью специальных скриптов программы [xMarkup](#), а визуализация результатов с помощью библиотеки pandas на Python. В отдельных случаях пришлось немного исправлять разметку оригинальных HTML-текстов, чтобы получить правильные и значимые результаты.

Визуализации полученных результатов представлены на Рис. 1-12. По оси X отложены значения метрики L, а по оси Y – значения Q для отдельных произведений.

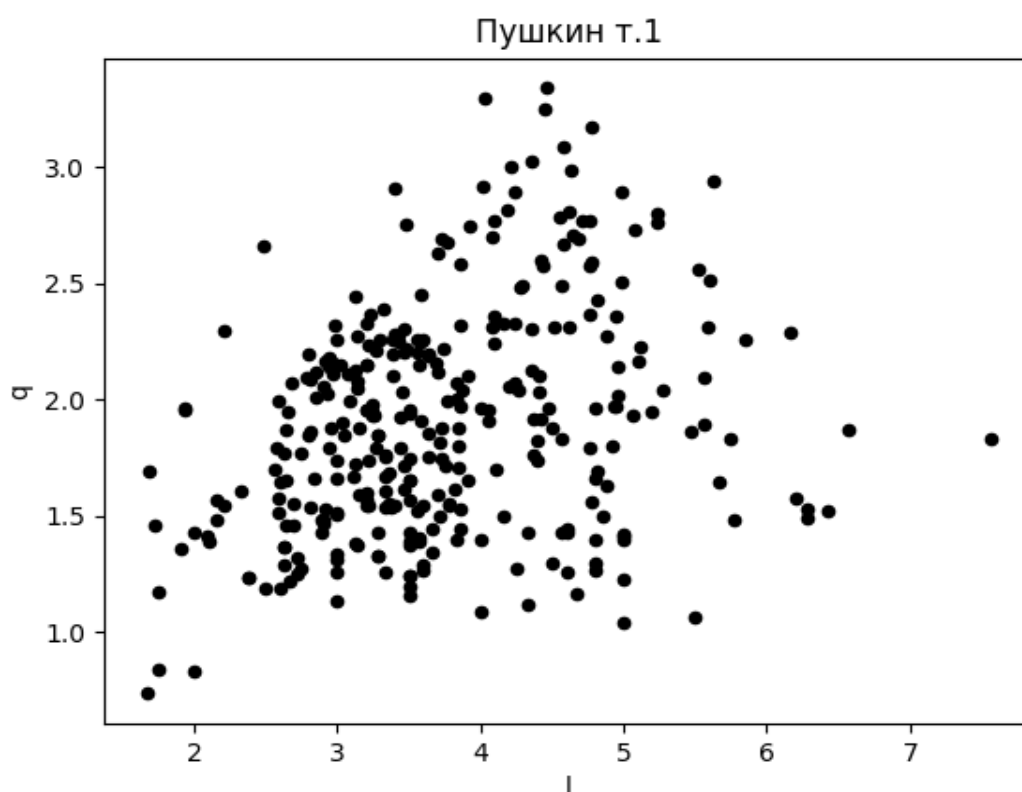


Рис. 1. Т. 1 Стихотворения 1814–1822

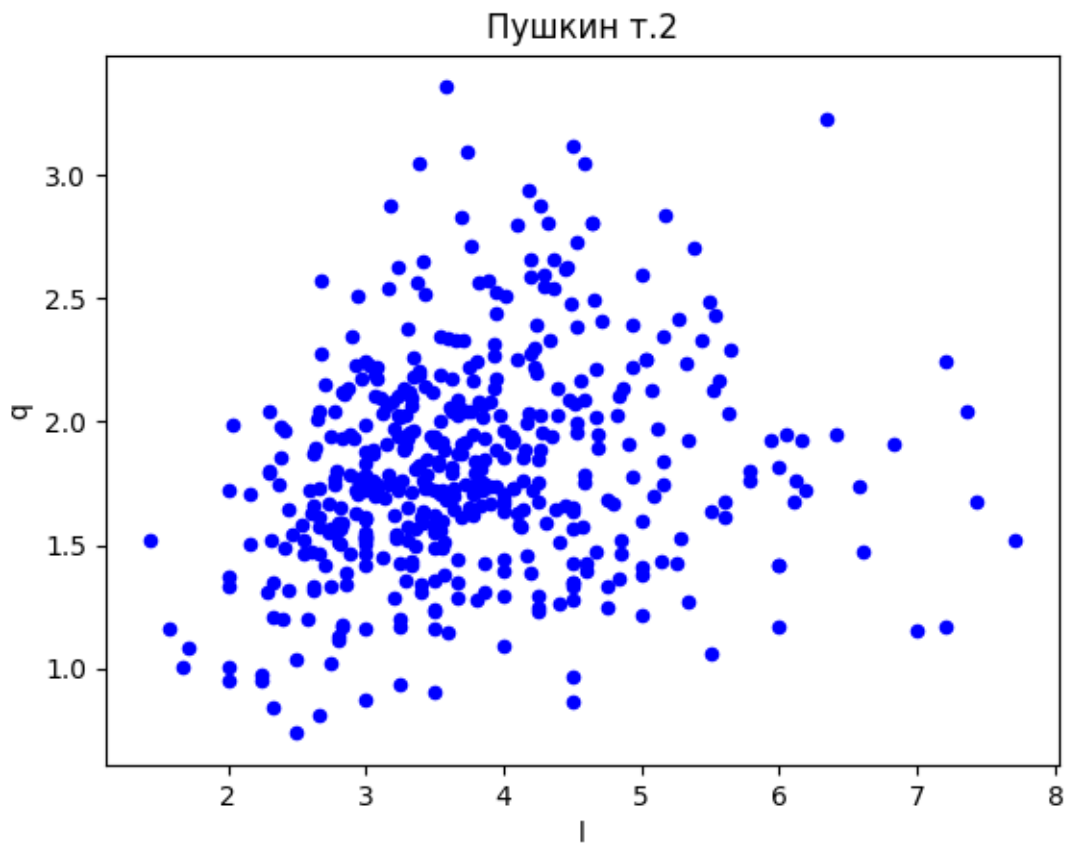


Рис. 2. Т. 2 Стихотворения 1823–1836

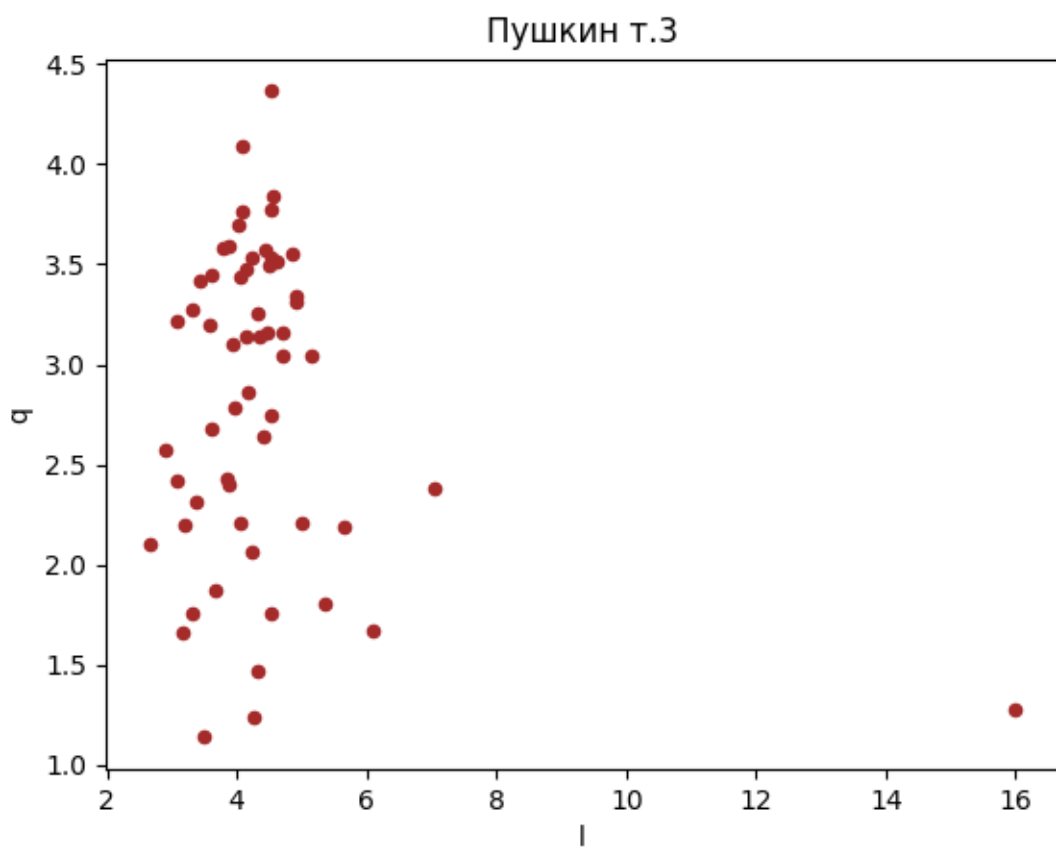


Рис. 3. Т. 3 Поэмы. Сказки

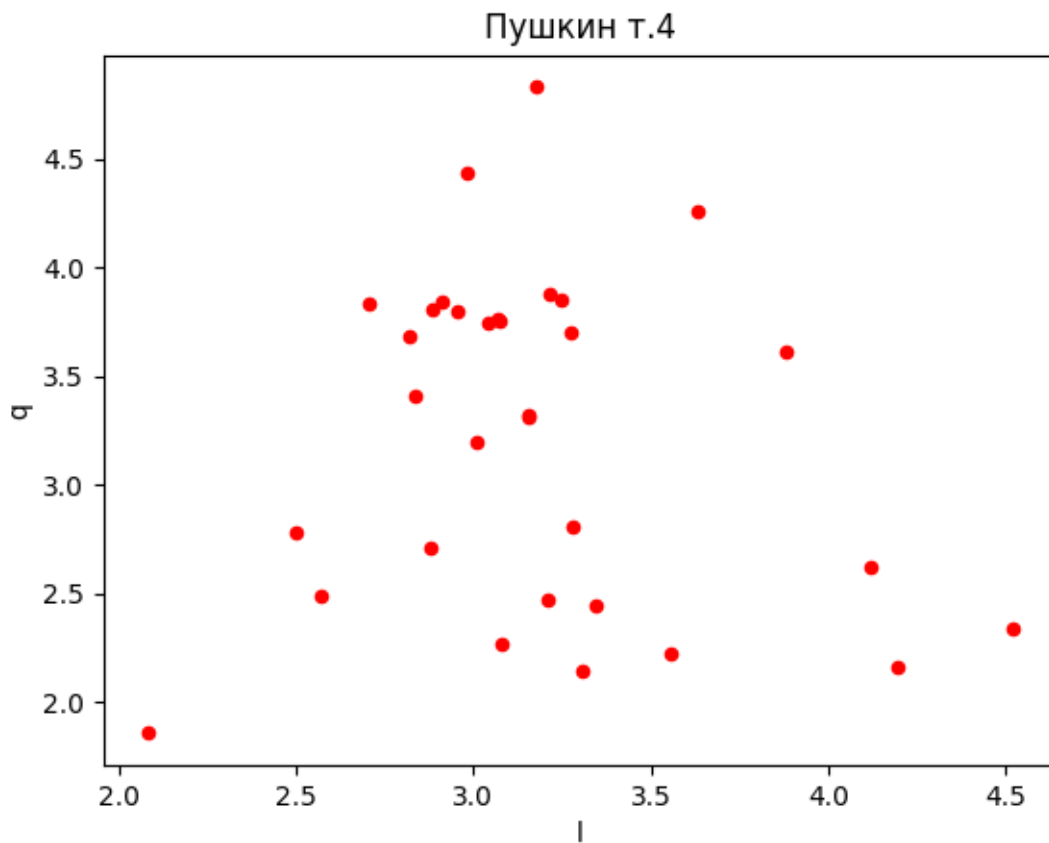


Рис. 4. Т. 4 Евгений Онегин, Драматические произведения

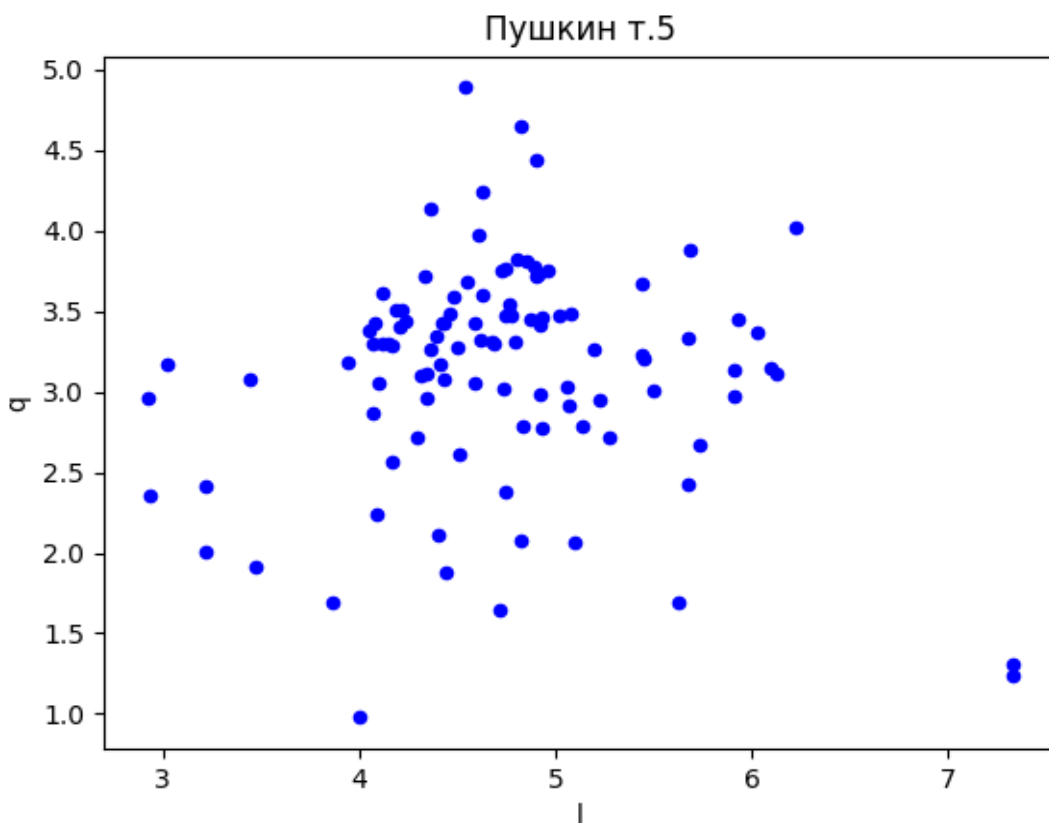


Рис. 5. Т. 5 Романы. Повести

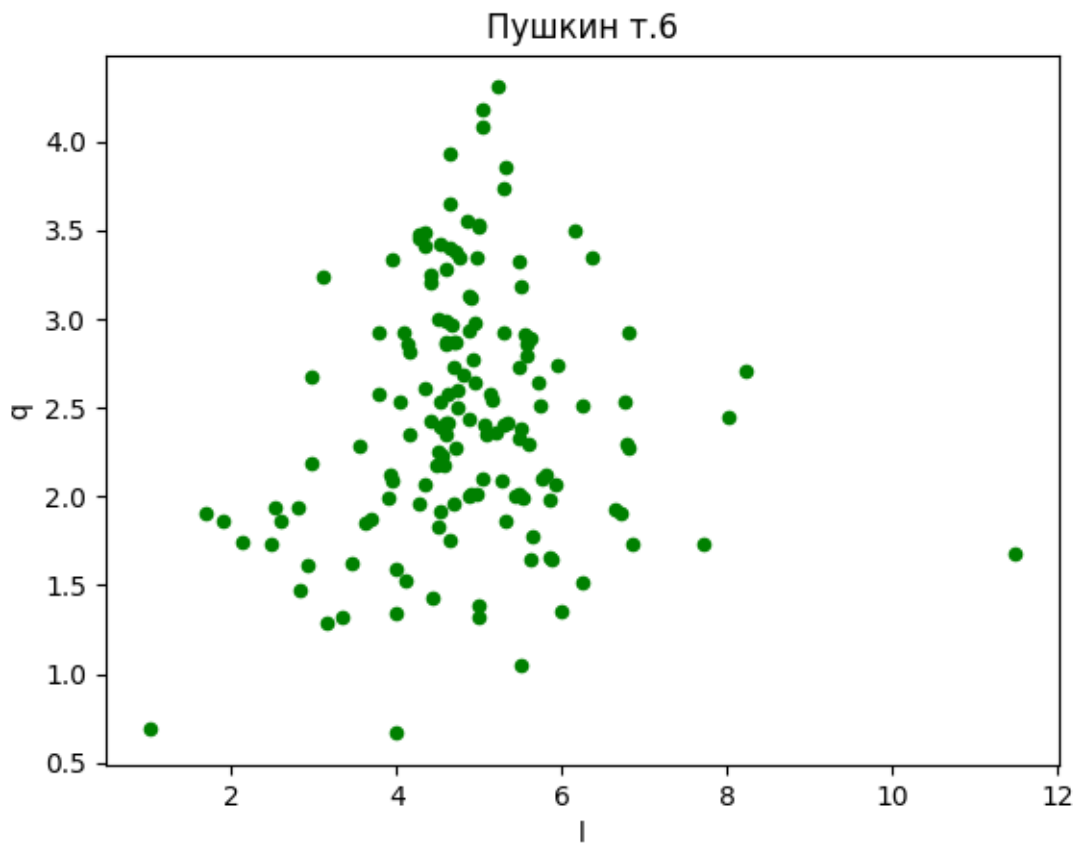


Рис. 6. Т. 6 Критика и публицистика

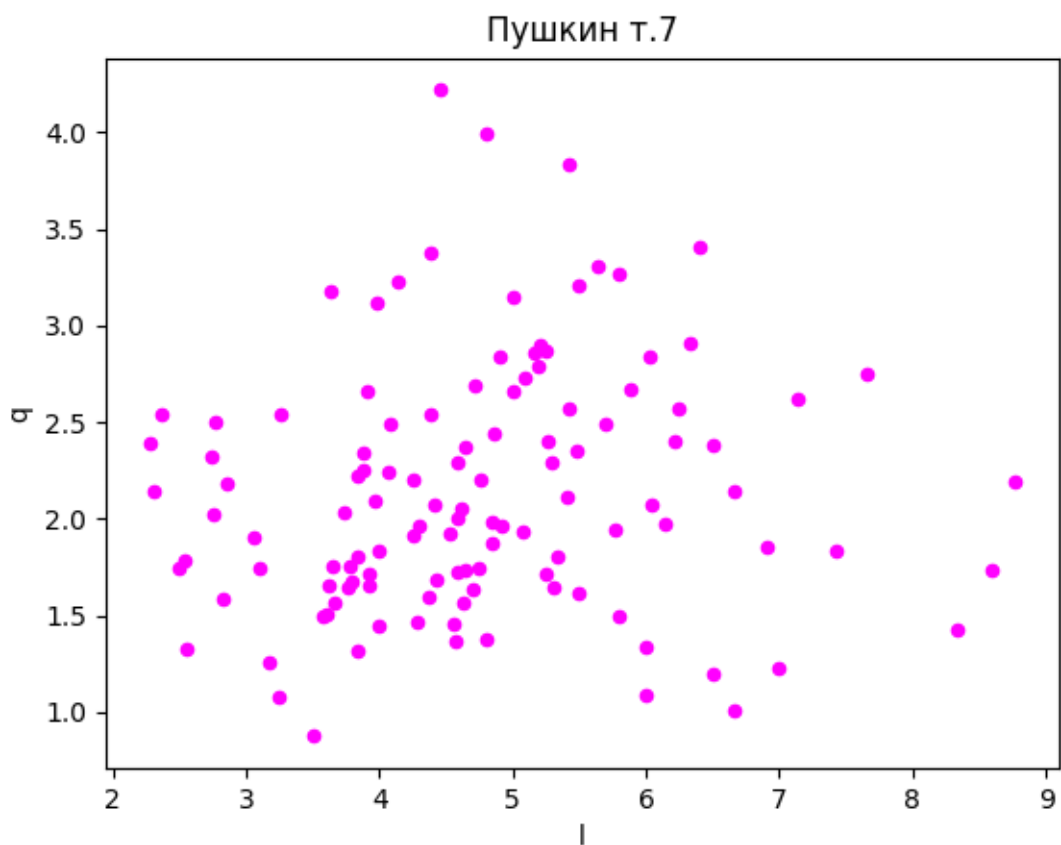


Рис. 7. Т. 7 История Пугачева, Исторические статьи и материалы, Воспоминания и дневники

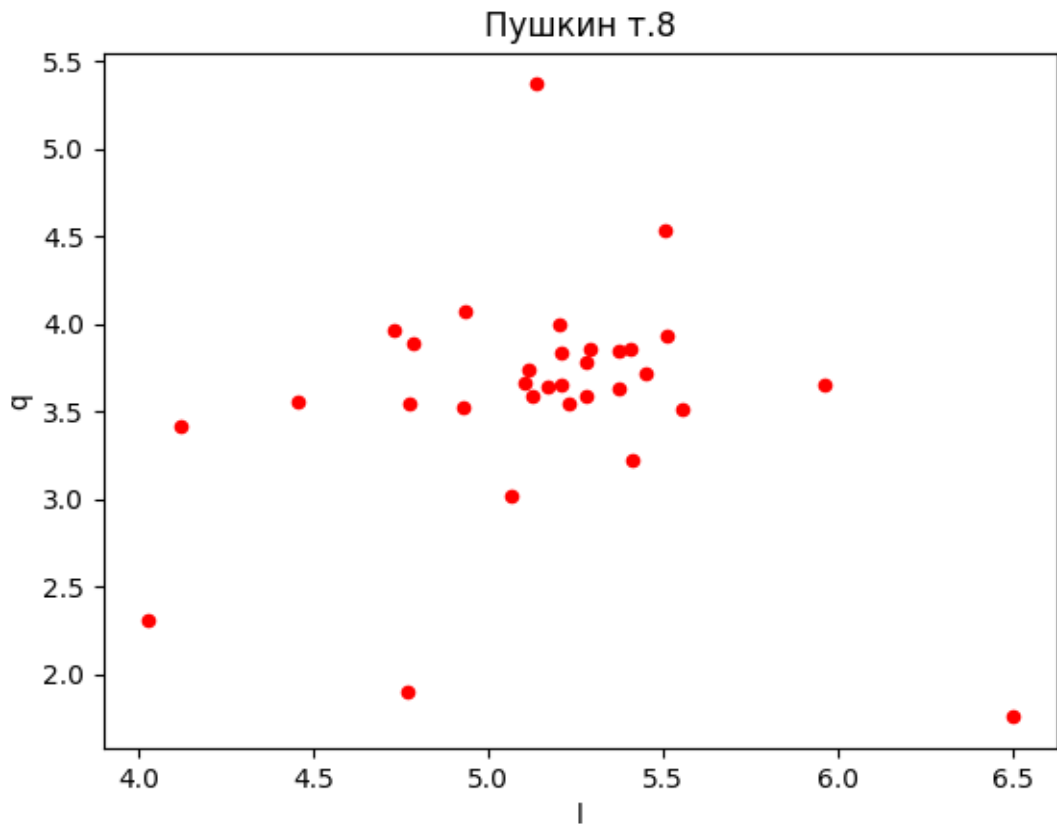


Рис. 8. Т. 8 История Петра I, Записки бригадира Моро-де-Бразе

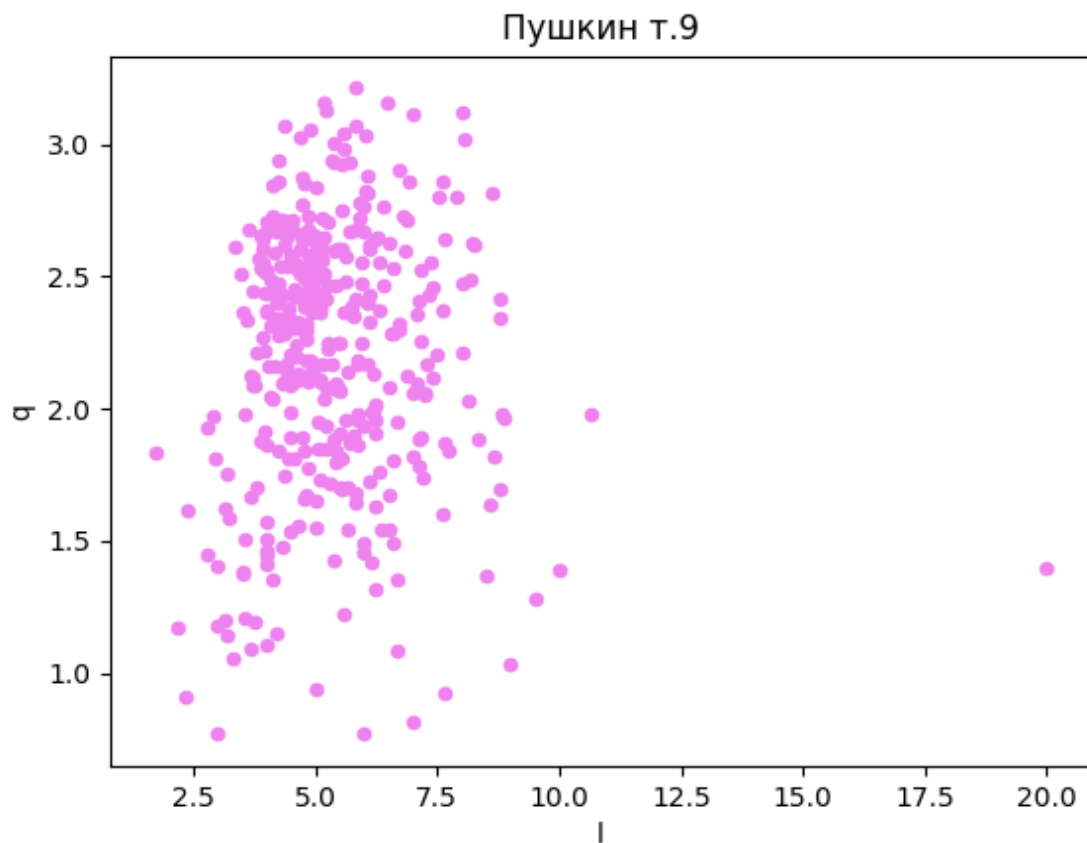


Рис. 9. Т. 9 Письма 1815-1830

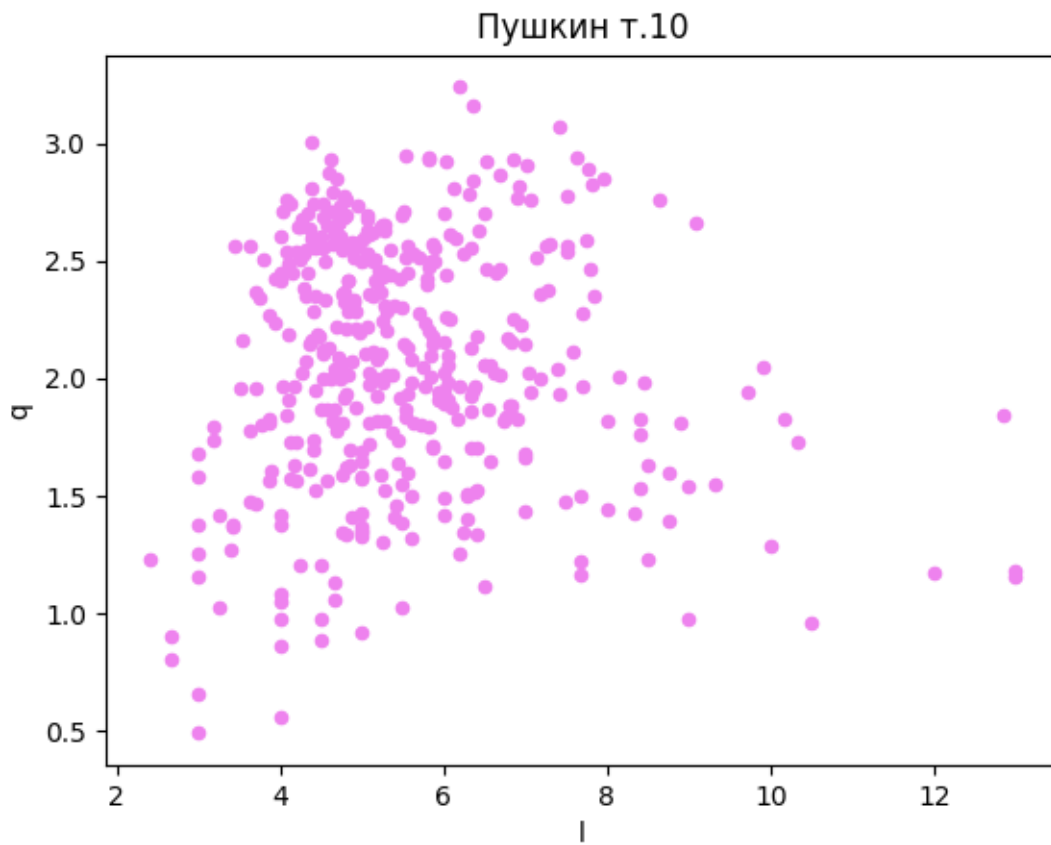


Рис. 10. Т. 10 Письма 1831-1837



Рис. 11. Случайные тексты I (длина слов в числах Фибоначчи)

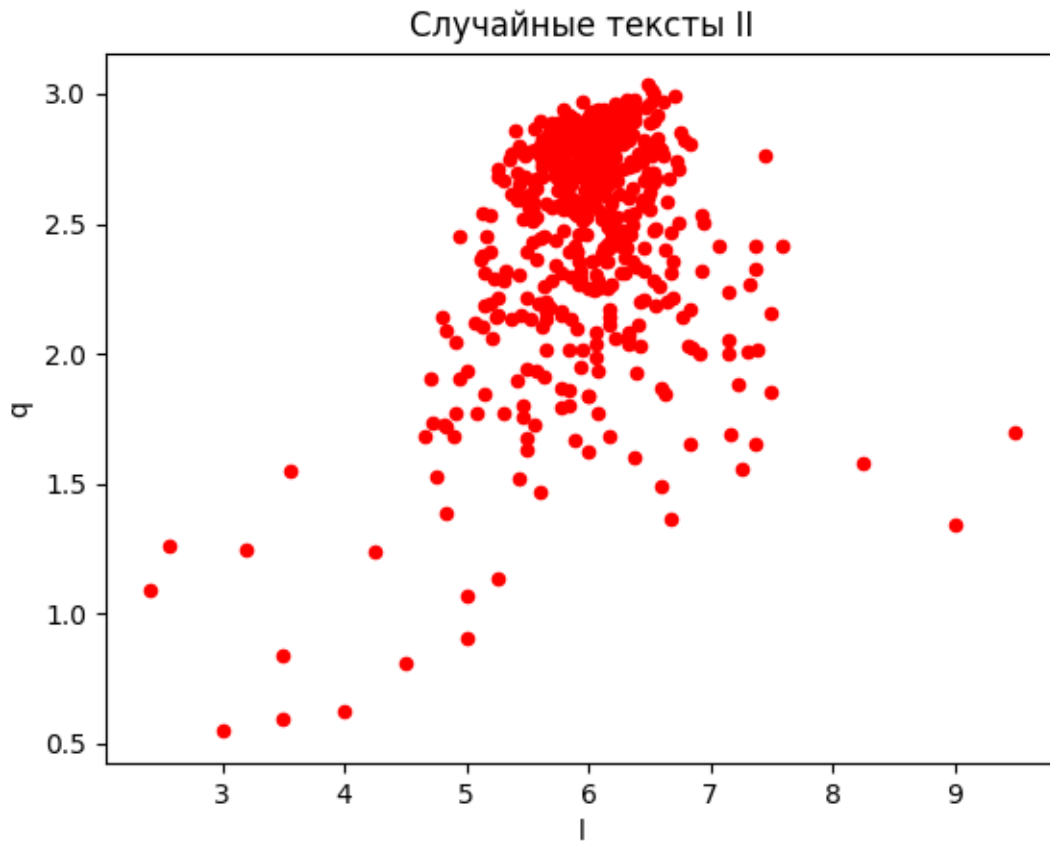


Рис. 12. Случайные тексты II

Анализ полученных результатов

Теперь проверим пропорции золотого сечения в произведениях каждого тома и контрольных случайных текстах.

Приведем разброс значений и среднее (mean) для метрик L и Q.

Покажем в таблице топ-N произведений, для которых наблюдаются признаки золотого сечения. Т.е. когда значение Q/L пропорционально числу Φ с заданной погрешностью $\epsilon \leq 0.05$. Это можно записать более строго:

$$|Q/L - \epsilon| = \begin{cases} a * \Phi, & \text{где } a = 1/2, 1, 2, 3, 4, 5 \\ \varphi = 1 / \Phi \end{cases}$$

В колонках таблицы приведем следующие значения:

- **N** – число слов в тексте;
- **n** – число синтагм (фраз);
- **L** – средняя длина синтагмы в словах;
- **Q** – метрика условного «качества» текста;
- **Φ^*** – вычисленное значение пропорции Q/L (как предполагаемого «кандидата» на золотое сечение).

Под каждой таблицей укажем общее число произведений, имеющих признаки золотого сечения.

Стихотворения А. С. Пушкина (тт. 1-2)

$1.43 < L < 7.71$, $L_{\text{mean}} = 3.76$; $0.74 < Q < 3.36$; $Q_{\text{mean}} = 1.85$.

Топ-5 произведений, имеющих признаки золотого сечения:

##	наименование	N	n	L	Q	Φ*
1	Наполеон на Эльбе (1815)	504	109	4.624	2.809	1.646
2	Странник	525	113	4.646	2.806	1.656
3	Песни западных славян. 8. Марко Якубович	469	101	4.643	2.804	1.660
4	Наполеон	469	103	4.553	2.787	1.634
5	Песни западных славян. 1. Видение короля	367	81	4.53	2.729	1.660

Всего текстов с признаками золотого сечения: 71 из 796 (~9%).

Поэмы и сказки А. С. Пушкина (тт. 3-4)

$2.08 < L < 16$, $L_{\text{mean}} = 3.965$; $1.14 < Q < 2.97$, $Q_{\text{mean}} = 4.83$.

Топ-10 произведений, имеющих признаки золотого сечения:

##	наименование	N	n	L	Q	Φ*
1.	Евгений Онегин — Полный текст	23044	7256	3.176	4.83	0.657
2.	Евгений Онегин. Ранние редакции	7813	2151	3.63	4.263	0.852
3.	Евгений Онегин. Глава восьмая	3430	1066	3.22	3.88	0.829
4.	Евгений Онегин. Глава первая	3199	985	3.248	3.853	0.843
5.	Евгений Онегин. Глава третья	2834	959	2.955	3.797	0.778
6.	Евгений Онегин. Глава шестая	2611	851	3.068	3.759	0.816
7.	Сцены из рыцарских времен	3292	1071	3.074	3.753	0.819
8.	Евгений Онегин. Глава четвертая	2529	831	3.043	3.745	0.813
9.	Скупой рыцарь	2299	816	2.817	3.682	0.765
10.	Моцарт и Сальери	1349	476	2.834	3.407	0.832

Всего текстов с признаками золотого сечения: 15 из 90 (16.7%).

Тексты поэмы «Руслан и Людмила» и романа в стихах «Евгений Онегин» (ЕО) обрабатывались с разбивкой по главам. Для ЕО из обработки были исключены авторские комментарии и примечания.

Интересно отметить, что полный текст ЕО согласуется с обратной пропорцией золотого сечения $1/\Phi$, тогда как тексты отдельных глав с величиной $1/2\Phi$ (как и тексты поэм). Однако, невозможно с какой-либо достоверностью судить о том, что поэт в этих произведениях явно использовал золотое сечение. Для этого нужен отдельный анализ каждого произведения, с разбивкой его на более мелкие части – строки или строфы. Это выходит за рамки нашего исследования.

Романы и повести А. С. Пушкина (т. 5)

$2.92 < L < 7.33$, $L_{\text{mean}} = 4.72$; $0.98 < Q < 4.89$, $Q_{\text{mean}} = 3.13$.

Произведения, имеющих признаки золотого сечения:

##	наименование	N	n	L	Q	Φ*
1.	Дубровский. Глава III	1402	270	5.193	3.259	1.593

2.	Дубровский. Глава XIX	748	158	4.734	3.012	1.572
3.	Пиковая дама. V	669	136	4.919	2.988	1.646
4.	Гости съезжались на дачу. Ранние редакции	442	103	4.291	2.721	1.577
5.	Дубровский. Глава XIV	329	79	4.165	2.565	1.624
6.	Повесть из римской жизни. Ранние редакции	74	23	3.217	2.009	1.602

Всего текстов с признаками золотого сечения: 6 из 105 (15.5%).

Тексты «Арап Петра Великого», «Дубровский», «Капитанская дочка», «Пиковая дама» обрабатывались с разбивкой по главам.

Критика и публицистика А. С. Пушкина (т. 6)

$1.0 < Q < 11.5$, $Q = 4.83$; $0.67 < Q < 4.308$, $Q = 2.458$.

Топ-5 произведений, имеющих признаки золотого сечения:

##	наименование	N	n	L	Q	Ф*
1.	Отрывок из литературных летописей	1334	243	5.49	3.322	1.652
2.	Последний из свойственников Иоанны д'Арк	908	185	4.909	3.117	1.575
3.	О втором томе «Истории русского народа» Полевого	787	159	4.95	2.983	1.659
4.	О г-же Сталь и о г. А. Муханове	625	134	4.664	2.967	1.572
5.	Баратынский принадлежит к числу отличных наших поэтов...	625	128	4.883	2.937	1.662

Всего текстов с признаками золотого сечения: 12 из 151 (~8%).

История Пугачева, Исторические статьи и материалы, Воспоминания и дневники (т. 7)

$2.27 < L < 8.77$, $L_{\text{mean}} = 4.744$; $0.876 < Q < 4.22$, $Q_{\text{mean}} = 2.139$.

Топ-5 произведений с признаками золотого сечения:

##	наименование	N	n	L	Q	Ф*
1.	Из дневника 1831 года	950	190	5.0	3.146	1.589
2.	Державин	233	57	4.088	2.494	1.639
3.	Старинные пословицы и поговорки	171	44	3.886	2.344	1.658
4.	Удельные князья...	58	19	3.053	1.901	1.606
5.	Table-talk. У Крылова над диваном...	53	10	5.3	1.642	3.228

Всего текстов с признаками золотого сечения: 5 из 119 (4.2%).

История Петра I, Записки бригадира Моро-де-Бразе (т. 8)

$4.024 < L < 6.5$, $L_{\text{mean}} = 5.155$; $1.755 < Q < 5.37$, $Q_{\text{mean}} = 3.6$.

Произведения с признаками золотого сечения:

##	наименование	N	n	L	Q	Ф*
1.	История Петра I. 1709 (вторая половина)	3016	506	5.96	3.651	1.633
2.	История Петра I. 1709 (до Полтавского)	2033	366	5.555	3.51	1.583

	сражения)					
--	-----------	--	--	--	--	--

Всего текстов с признаками золотого сечения: 2 из 32 (6.2%).

Письма (тт. 9-10)

$1.714 < L < 20$, $L_{\text{mean}} = 5.457$; $0.49 < Q < 3.24$, $Q_{\text{mean}} = 2.134$.

Топ-5 писем:

##	наименование	N	n	L	Q	Φ^*
1.	Письма. Керн А. П., 13 и 14 августа 1825 г.	1133	218	5.197	3.154	1.648
2.	Письма. Вульф А. Н., 21 июля 1825 г.	1056	203	5.202	3.13	1.662
3.	Письма. Пушкину Л. С., 24 сентября 1820 г.	827	169	4.893	3.054	1.602
4.	Письма. Нащокину П. В., середина марта 1834 г.	627	136	4.610	2.932	1.572
5.	Письма. Пушкиной Н. Н., 2 сентября 1833 г.	612	133	4.602	2.873	1.602

Всего текстов с признаками золотого сечения: 55 из 806 (6.8%).

Корпус случайных текстов I (по послед-ти Фибоначчи)

$2.75 < L < 8.0$, $L_{\text{mean}} = 4.51$; $0.47 < Q < 2.85$, $Q_{\text{mean}} = 2.32$.

Топ-5 текстов, имеющих признаки золотого сечения:

##	наименование	N	n	L	Q	Φ^*
1.	Случайный текст типа I	468	100	4.68	2.845	1.645
2.	Случайный текст типа I	466	101	4.64	2.829	1.631
3.	Случайный текст типа I	447	97	4.61	2.808	1.641
4.	Случайный текст типа I	462	99	4.67	2.806	1.663
5.	Случайный текст типа I	444	99	4.48	2.797	1.604

Всего текстов с признаками золотого сечения: 72 из 512 (14.1%).

Как мы уже упоминали, в этом корпусе длина слов в синтагмах бралась согласно последовательности Фибоначчи 1, 2, 3, 5, 8. Длина синтагмы была ограничена 10.

Корпус случайных текстов II

$2.0 < L < 9.25$, $L_{\text{mean}} = 6.03$; $0.24 < Q < 3.02$, $Q_{\text{mean}} = 2.487$.

Топ-5 текстов, имеющих признаки золотого сечения:

##	наименование	N	n	L	Q	Φ^*
1.	Случайный текст типа II	144	20	7.2	2.21	3.255
2.	Случайный текст типа II	109	16	6.81	2.13	3.202
3.	Случайный текст типа II	88	13	6.77	2.08	3.251
4.	Случайный текст типа II	92	14	6.57	2.04	3.226
5.	Случайный текст типа II	86	14	6.14	1.92	3.191

Всего текстов с признаками золотого сечения: 12 из 512 (2.3%).

Как мы видим, для текстов с чисто случайной структурой признаки золотого сечения минимальные.

Заключение

Наибольшие признаки золотого сечения имеют стихотворения, поэмы и романы А. С. Пушкина. Так, поэмы и сказки показывают признаки золотого сечения в 16.7% случаев, романы и повести - в 15.5% случаев, а стихотворения - в 9% случаев.

Примечательно, что роман в стихах «Евгений Онегин» показал максимальные признаки золотого сечения в сравнении с другими произведениями Пушкина - как целый текст, так и отдельные главы (1, 3, 4, 6, 8). Однако для совокупности произведений в рамках каждого отдельного тома нельзя со значимым уровнем доверия даже предполагать, что поэт предпочитал использовать пропорции золотого сечения. Например, чтобы принять гипотезу об использовании золотого сечения в поэмах и сказках (тт. 3-4), нужно чтобы число текстов с признаками золотого сечения было не менее 55 из 90 при уровне доверия 95%. А у нас их всего 15 из 90, т.е. преваляют тексты без признаков золотого сечения. Смотри [9] на предмет биномиальных испытаний при оценке гипотез, аналогичных выше рассмотренной.

Наш грубый метод оценки, учитывающий лишь распределение длин слов внутри текста, может выявить только признаки или «следы» предполагаемого золотого сечения, но на этом основании нельзя сделать никаких выводов. Т.е. гипотеза, заявленная в начале: **можно ли оценить художественный текст на наличие свойств, которые соответствует пропорции золотого сечения** пока остается ни отвергнутой, ни принимаемой. Вероятно, нужно рассмотреть уточненный метод оценки, основанный на делении текстов не на синтагмы, а на абзацы, предложения или строфы для стихотворных текстов.

Любопытно отметить, что текст «Евгения Онегина», ранее анализируемый другими исследователями на наличие пропорций золотого сечения, показал в нашем анализе максимальные результаты! Было бы интересно обработать текст «Евгения Онегина» с разбивкой на строки или строфы, чтобы принять или отвергнуть гипотезу преднамеренного использования золотого сечения. Но это уже тема другого исследования.

Стоит также отметить, что полученные результаты дают качественный отпечаток (условную «дактилоскопию») авторских текстов с помощью метрик L и Q. Причем даже в зависимости от характера произведений, - так для стихотворений характерны более компактные значения L и Q, в сравнении с прозой или публицистикой.

Библиография

1. Н.А. Васютинский — Золотая пропорция; Москва, «Молодая гвардия», 1990; ISBN 5-235-00806-5.
2. А. В. Пустовит, Архитектоника художественной формы, 2019.
3. О. Н. Гринбаум, Когнитивный принцип "золотое сечение" в поэтической антропологии А. С. Пушкина.
4. Гринбаум О.Н., Золотое сечение» и метафизика пушкинского стиха <https://www.trinitas.ru/rus/doc/0232/004a/02320032.htm>
5. М. А. Абрамов, диссертация «Структурно-статистический анализ паттернов золотого сечения и зеркальной симметрии в стихотворениях А. С. Пушкина», 1999.
6. Radoslav Rochallyi, Golden Ratio Poetry: Mathematical Poetry According to ϕ , <https://maa.org/math-values/golden-ratio-poetry-mathematical-poetry-according-to/>
7. Sarah Glaz, Poems structured by integer sequences, *Journal of Mathematics and the Arts*, 2016, 10(1-4): 44-52, doi.org/10.1080/17513472.2016.1231574

8. О.Н. Гринбаум. Третья глава романа А. С. Пушкина «Евгений Онегин»: письмо Татьяны Лариной в ритмико-смысловом освещении. В сборнике *Respectus Philologicus*, 2010, Nr. 10 (22), стр. 43-54.
9. Биноминальный тест. <https://habr.com/ru/articles/783816/>
10. В. Феллер. Введение в теорию вероятностей и ее приложения. Т. 1. Москва, Мир, 1984; см. § 2, Биноминальное распределение.