

Почему ошибался Жуковский: о внутритекстовых причинах метрических сбоев в «Одиссее»¹

В 1849 году В. А. Жуковский завершил перевод «Одиссеи», который, несмотря на критику отдельных неточностей² и эстетического облика в целом³, остается для массового русского читателя главным текстом о путешествиях древнегреческого героя.

Благодаря своему месту в культуре этот перевод многократно становился предметом специальных исследований⁴, среди которых, однако, трудно обнаружить такие, которые были бы посвящены его метрике. Это не случайно: со стиховедческой точки зрения «Одиссея» довольно однообразна. Жуковский почти не пользуется выразительными возможностями русского гекзаметра и пишет почти исключительно ровным шестистопным дактилем, которым описываются 11915 из 12106 стихов поэмы. Строки с единственным стяжением на первой стопе допускаются Жуковским всего в 96 случаях, на четвертой — в 14 случаях, по 11 строк приходится на одно стяжение третьей и второй стоп. Комбинации замен дактилических стоп на хореические в одной строке и вовсе единичны. Здесь на метрическом уровне проявляет себя общая «сглаженность углов» перевода Жуковского, отраженная в полемическом высказывании последнего по времени переводчика Гомера М. А. Амелина: «“Одиссее” на русской почве не повезло. <...> В. А. Жуковский <...> разрушил весь Гомеров эпический строй, убрав повторы и длинноты, попытался превратить эпос в роман, а из вынужденного скитальца Одиссея сделал туриста а-ля Чайльд-Гарольд»⁵. Особенно явным стиховедческое однообразие становится в свете данной в том же тексте емкой характеристики гекзаметра: «Гекзаметр по-русски я рассматриваю как дольник с мужской или женской цезурой на шестиударной основе, в котором любое ударение, кроме предцезурного и последнего в строке, может быть опущено и стать мнимым или подразумеваемым, как это обычно происходит в русских дольниках. Основной ритмический закон оригинала: в каждых трех соседствующих стихах ритмический рисунок повторяться не должен. Он создает необходимую вариативность, мешающую быстрому засыпанию при чтении или прослушивании»⁶. Так как текст Жуковского «необходимой вариативности» лишен, на первый взгляд, исследовательского интереса со стиховедческой точки зрения он не представляет.

Однако в представленном Жуковском варианте гомеровской поэмы есть 49 случаев особого рода, которые мы будем далее называть метрическими аномалиями. Речь идет в основном о тех же дактилических строках, длина которых, однако, отличается от гекзаметра на одну стопу как в сторону увеличения, так и в сторону уменьшения.

Многих он в помощь себе приведёт из песчаного Пилоса, многих (II, 326)
Слёзы я пролил, увидя его; состраданье мне душу проникло. (XI, 55)
Якорный камень, надёжным канатом корабль утвердили у берега; (XV, 498)

1 Автор благодарит В. В. Файера, Е. А. Кузьменко и Ю. Б. Орлицкого за неоценимую помощь в проведении настоящего исследования.

2 См. Егунов А. Н. Гомер в русских переводах XVIII–XIX веков. Москва—Ленинград: Наука, 1964; Ярхо В. Н. В. А. Жуковский — переводчик «Одиссеи» // Жуковский В. А. Полное собрание сочинений и писем: В 20 т. Т. 6. М.: Яз. рус. культуры, 2010. С. 460—481.

3 См. цитируемое далее высказывание М. А. Амелина.

4 Одно из ярких исследований новейшего времени: Виноцкий И. Теодиссея Жуковского: Гомеровский эпос и революция 1848—1849 годов // Новое литературное обозрение. 2003. № 60. С. 171—193.

5 Гомер Одиссея. Песнь первая. Перевод с древнегреческого и предисловие Максима Амелина // Новый мир. 2013. № 2. С. 138.

6 Там же.

Как видно, в этих примерах Жуковский использует семистопный дактиль вместо привычного (и требуемого) шестистопного. Таких строк во всем переводе насчитывается 30.

В этой земле не знаком мне; скажи, где дорога (VI, 177)
Сонный в спокойном безветрии плыть и достигнешь (VII, 319)
Путь, сокрушивший колена столь многих мужей (XIV, 236)

Такого рода случаев, когда строка оказывается на стопу короче, чем остальные, меньше, они исчерпываются 19 стихами.

Архивные документы свидетельствуют, что Жуковский боролся с такими аномалиями, находил и исправлял именно их в своем переводе: «[В бумагах Фарнгагена сохранился листок руки Жуковского (июнь, 1849 г.), следующего содержания:]

Главные поправки, которые нужно сделать в первом томе Одиссеи. Прочие поправки напечатаны в конце тома»⁷.

Эти поправки разделены на два списка: «Лишние стопы» и «Недостаёт стопы», в каждом из которых указаны песнь, номер стиха и исправленный вариант.

Однако 49 метрически аномальных строк в финальном тексте все же сохранилось. Мы не знаем, произошло это благодаря авторскому заданию или в силу случайности. Но если рассматривать эти стихи как маркированные элементы художественной структуры, то можно попытаться ответить на вопрос о внутритекстовых причинах их появления. Возможно, если такие причины отыщутся, это будет свидетельствовать в пользу того, что метрические аномалии следует рассматривать, скорее, как прием, нежели как недосмотр переводчика.

В то же время известно, что человеку в принципе свойственно отыскивать структурные закономерности даже там, где их нет; к этому его побуждает базовое когнитивное свойство сознания, которое описывается как «эпигенетическое» (то есть такое, к которому человек предрасположен генетически) правило «ищи структуру в хаосе»⁸. Чтобы по возможности оградить себя от легковесных предположений, мы используем не только собственные наблюдения, но и инструменты автоматической обработки текста и анализа данных; точнее, попытаемся сравнить результаты традиционного филологического подхода и применения количественных методов.

Можно ли в принципе решить проблему внутритекстовых причин метрических аномалий, используя методы анализа данных? Даже не перечисляя конкретные методы, следует сказать, что это серьезный вызов. Анализ данных служит для обнаружения тенденций в больших массивах информации. В нашем случае ситуация обратная: мы пытаемся выявить механизм появления не тенденций, а исключений: 49 случаев на 12 тысяч строк — это слишком малая величина. Аналитику проще пренебречь ею и сказать, что Жуковский не ошибался, чем рассматривать такие данные всерьез. Но поскольку филологический подход требует как раз внимания к деталям, попробуем извлечь из существующих методов все, что они смогут нам предложить, и проанализировать, чем строки с метрическими аномалиями (и их текстовое окружение) отличаются от остального текста.

Сразу оговоримся, что, поскольку Жуковский не знал языка оригинала и пользовался немецким подстрочником, мы также опустим отношения с античным исходником и сосредоточимся на русскоязычном тексте перевода.

В своем стремлении обнаружить закономерности в появлении метрических аномалий мы будем постепенно сужать рассматриваемый контекст.

Самый широкий (после масштаба всей поэмы) случай текстового окружения аномальной строки — это песнь. Для самого Жуковского деление на песни было значимым, именно песнями поэт мерил осуществленный объем перевода: «он создает специальные таблицы,

⁷ Летопись работы В. А. Жуковского над переводом «Одиссеи» (по материалам его эпистолярия) // Жуковский В. А. Полное собрание сочинений и писем: В 20 т. Т. 6. М.: Яз. рус. культуры, 2010. С. 659.

⁸ «“epigenetic rule” that says “Seek structure amid chaos”» Calvin W.H., Bickerton D. *Lingua ex machina: Reconciling Darwin and Chomsky with the human brain*. Cambridge, 2000. P. 33.

фиксирующие процесс работы над «Одиссеей», с указанием точной датировки перевода отдельных песен»⁹. С песен, рассмотренных как отдельные тексты, мы и начнем наш анализ.

Метрические аномалии распределены по песням неравномерно. Больше всего их (по 6) в IX и XII песнях (обе переведены в 1844 году). В песнях I, III, XVI, XIX—XXIV аномалий нет совсем. При этом песни с пиковым числом аномалий представляют нам аномалии разного типа. В упоминавшихся IX и XII песнях число в 6 аномалий складывается из 3 семистопных и 3 пятистопных строк. В то же время V песнь, насчитывающая 5 аномалий, имеет совсем другое соотношение типов аномалий, так как здесь 4 семистопные строки и только 1 пятистопная. Подробно распределение аномалий по песням представлено в таблице 1.

Песнь	Год	Аномалий 7-стопные	5-стопные	Всего строк
I	1842	-	-	440
II	1843	3	3	434
III	1843	-	-	497
IV	1844	1	1	847
V	1844	5	4	493
VI	1844	4	3	331
VII	1844	4	1	347
VIII	1844	3	2	586
IX	1844	6	3	566
X	1844	5	3	574
XI	1844	2	1	640
XII	1844	6	3	453
XIII	1848	4	2	440
XIV	1848	2	1	533
XV	1848	2	2	557
XVI	1848	-	-	481
XVII	1849	1	-	606
XVIII	1849	1	1	428
XIX	1849	-	-	604
XX	1849	-	-	394
XXI	1849	-	-	434
XXII	1849	-	-	501
XXIII	1849	-	-	372
XXIV	1849	-	-	548
Всего		49	30	12106

Таблица 1. Распределение метрических аномалий по песням в «Одиссее»

Наличие аномалий в песнях можно представить в виде графика. См. рис. 1.

⁹ Янушкевич А. С., Киселев В., Никонова Н. Примечания к текстам // Жуковский В. А. Полное собрание сочинений и писем: В 20 т. Т. 6. М.: Яз. рус. культуры, 2010. С. 409.

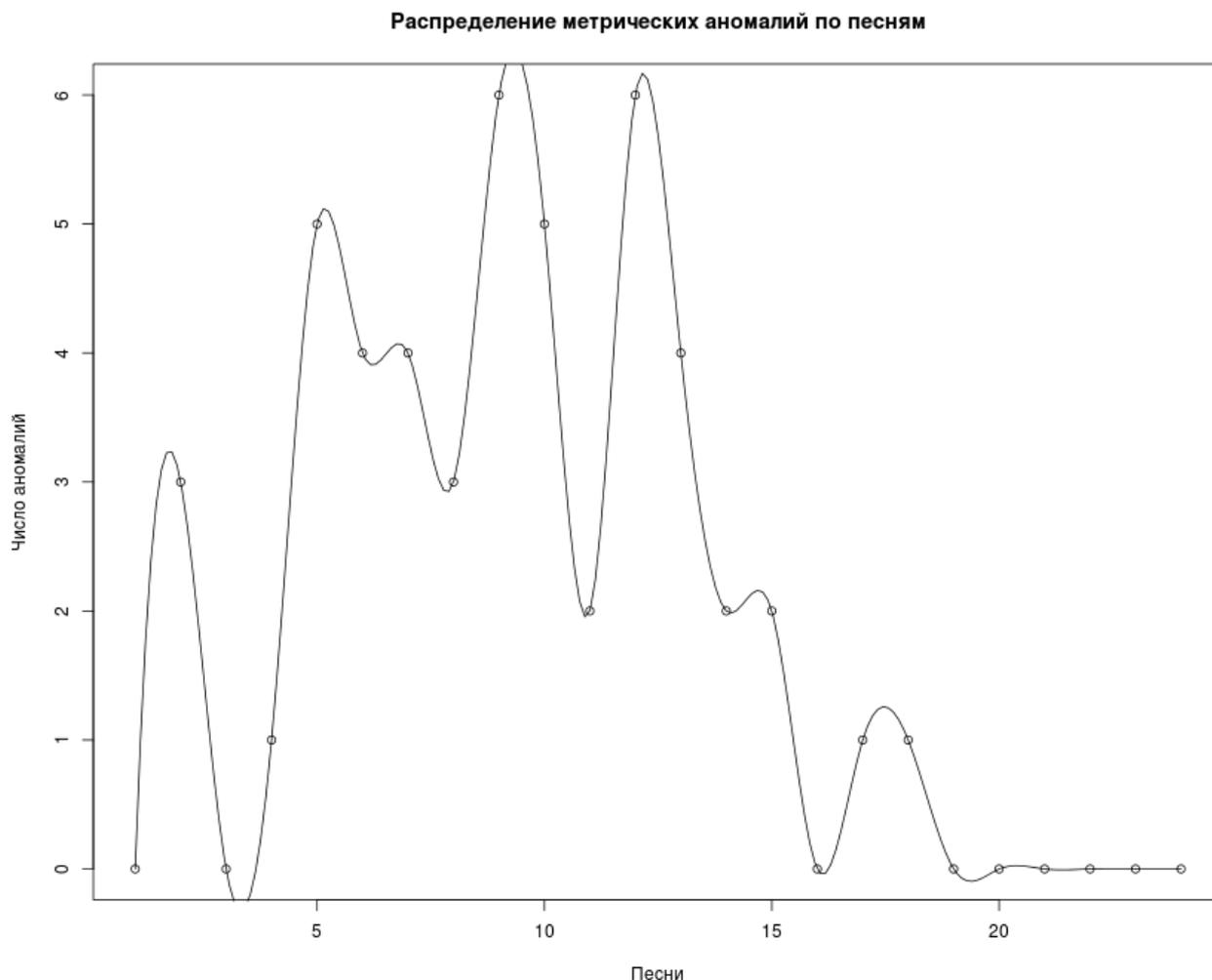


Рис. 1. Распределение метрических аномалий по песням в «Одиссее»

Как кажется, оперирование контекстом песни предполагает уместность подхода, при котором мы разделим все песни на две неравные группы. В первую попадут те, в которых встречаются метрически аномальные строки, во второй окажутся те, в которых таких строк не наблюдается. Таким образом, в первой группе будут II, IV—XV, XVII, XVIII песни; во второй — I, III, XVI, XIX—XXIV песни. Легко заметить, что первая группа захватывает, в основном, середину поэмы, а во второй оказываются песни «сюжетной рамки» начала и финала гомеровского текста.

Сначала попытаемся обойтись традиционным герменевтическим инструментарием. Существуют ли видимые глазу отличия песен первой группы от песен второй группы на каком-либо из возможных уровней (прежде всего, конечно, сюжетном)?

В принципе, наиболее значимый сюжетный фактор, который делит весь текст «Одиссеи» на две части, это состояние главного героя: половину повествования он находится на суше (Огигия, Итака), а вторую половину — в морском путешествии (реальном или данном через нарратив). При этом рассказ о путешествиях Одиссея приходится именно на середину поэмы. Однако Одиссей покидает Огигию в середине V песни (метрические аномалии появляются уже во II и в IV) и прибывает на Итаку в XIII в то время как метрические аномалии встречаются и в XIV, XV, XVII и XVIII песнях (хотя и единичны в двух последних). Таким образом, 6 песен из 15 нарушают предложенную модель распределения «аномальных» и «неаномальных» песен поэмы.

В то же время в «Одиссее» есть и побочный сюжет, связанный с фигурой Телемаха, который также переживает морское путешествие. Его отплытие состоялось гораздо раньше,

во II песне, причем одна из строк, описывающих спуск корабля на воду, содержит метрическую аномалию:

с ней рядом
Сел Телемах, и гребцы, отвязавши поспешно канаты,
Также взошли на корабль и сели на лавках у весел.
Тут светлоокая Зевсова дочь даровала им ветер попутный,
Свежий повеял зефир, ошумляющий тёмное море.

(II, 417—421)

Выделенная курсивом строка содержит семь дактилических стоп против ожидаемых шести.

Песнь III не содержит метрических аномалий, но и Телемах в ней не путешествует. Почти все содержание этой части составляет его разговор с Нестором в Пилосе. Сын возвращается на Итаку позже Одиссея, в конце XV песни, в которой мы последний раз в поэме наблюдаем неединичный случай появления метрических аномалий. Если бы мы согласились пренебречь XVII и XVIII песнями как маргинальными примерами, то отсутствие Телемаха на Итаке (или ситуация путешествия Телемаха) могло бы претендовать на объяснительный фактор, отделяющий «аномальные» песни от «неаномальных».

Теперь обратимся к автоматической обработке текста. В ее терминах наша задача может быть сформулирована как оценка тематической близости текстов между собой (каждую песнь мы считаем отдельным текстом) и последующая классификация текстов на основе этой близости. Если автоматическая классификация совпадет с распределением песен на «аномальные» и «неаномальные», это будет означать, что тексты, содержащие аномалии, тематически значимо отличаются от тех, в которых мы аномалий не наблюдаем.

Существует несколько метрик, позволяющих оценить тематическую близость текстов. В этом исследовании мы будем использовать самую распространенную, построенную на основе вычисления TF-IDF (от англ. TF — term frequency 'частотность терминов', IDF — inverse document frequency 'обратная частотность документов'). Принцип этой методики в том, что если некоторый термин встречается часто в одном тексте из коллекции и при этом редко в остальных текстах той же коллекции, то он должен быть признан значимым для определения темы этого текста. После того, как мы вычислили TF-IDF для всех слов имеющихся в нашей коллекции текстов, мы можем представить получившийся результат в виде вектора (по одному вектору на текст) и найти косинусное расстояние между этими векторами, и тем самым определить, насколько тексты похожи друг на друга с точки зрения встречаемости в них слов (с учетом частотности этих слов в других текстах коллекции). В рамках компьютерной лингвистики такой метод действительно позволяет с достаточно хорошей точностью отделить друг от друга тексты разной тематической направленности.

Даже несмотря на то, что вычисление и TF-IDF, и косинусного расстояния много раз производились, считать их можно разными способами. Так, исходный текст можно взять в его изначальном виде («Муза, скажи мне о том многоопытном муже»), а можно преобразовать, приведя каждую словоформу к ее нормализованному виду, словарной форме или т. н. «лемме» («Муза сказать я о тот многоопытный муж»). Кроме того, в компьютерной лингвистике и смежных технических дисциплинах принято при предобработке исключать из текстов так называемые стоп-слова, то есть частотные служебные слова (предлоги, частицы, союзы, некоторые местоимения), которые обязательно встречаются в каждом тексте и никак не влияют на его тему. Для поэтического текста необходимость исключения стоп-слов из анализа неочевидна: служебное слово в условиях тесноты стихового ряда может иметь поэтический (семантический, тематический) вес, актуализируя не работающие в прозаическом (или просто нехудожественном) тексте значения. Тем не менее, к перечисленным двум вариантам подсчета TF-IDF (на нормализованном и ненормализованном материале) можно добавить еще два: с исключением стоп-слов из ненормализованного текста

(«Муза, скажи мне многоопытному муже») и с исключением стоп-слов из нормализованного текста («Муза сказать я многоопытный муж»). Мы решили попробовать все четыре варианта подсчетов и оценить пользу каждого из них для тематической классификации песен «Одиссеи».

После того, как мы произвели все нужные подсчеты, можно было приступить к собственно автоматической классификации, инструментом которой выступил самый распространенный алгоритм кластеризации, то есть разбиения имеющегося множества объектов на группы, а именно алгоритм k -средних¹⁰. Этот метод принимает на вход результаты подсчета косинусного расстояния между векторами и число кластеров, на которое следует разбить имеющееся множество. В нашем исследовании это число равно 2, так как мы ожидаем, что алгоритм разделит все песни «Одиссеи» на две группы, которые в идеальном случае совпадут с группами «аномальных» и «неаномальных» песен.

Итак, нами было получено четыре варианта кластеризации песен «Одиссеи» на основе подсчетов TF-IDF и косинусного расстояния. В первый кластер попали песни, в которых мы ожидали бы увидеть метрические аномалии, во второй — песни, в которых аномалий в идеальном случае не было бы. Кроме состава кластера мы приводим также термины, которые в наибольшей степени повлияли именно на такое распределение песен по кластерам. Довольно хорошо видно, что набор таких слов для первого кластера отражает морскую и приключенческую тематику, а набор для второго кластера — сухопутную и бытовую.

1. Словоформы, стоп-слова исключены.

Кластер 1: III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII

Термины 1-го кластера: море, моря, феакийских, царь, земли, земле, берёг, спутники, немедля, так, феакиян, остров, я

Кластер 2: I, II, XIV, XV, XVI, XVII, XVIII, XIX, XX, XXII, XXI, XXIII, XXIV

Термины 2-го кластера: свинопас, лук, женихи, сын, женихов, женихам, кончил, хитроумный, сказал, чем, сына

2. Словоформы, стоп-слова включены.

Кластер 1: XII, VII, IX, XI, XIII, VI, X, V, VIII

Термины 1-го кластера: море, феакийских, царь, земли, моря, спутники, земле, берёг, феакиян, немедля, девы, остров, богиня, стадо, вёслами, отвечивал

Кластер 2: I, II, III, IV, XIV, XV, XVI, XVII, XVIII, XIX, XX, XXI, XXII, XXIII, XXIV

Термины 2-го кластера: женихи, свинопас, сын, лук, женихов, женихам, кончил, сына, хитроумный, сказал, многих

3. Леммы, стоп-слова исключены.

Кластер 1: V, VI, VII, VIII, IX, X, XI, XII, XIII

Термины 1-го кластера: алкиной, феакийский, царь, цирцея, пещера, дева, спутник, посейдон, товарищ, ветер, баран, брег, феакиянин, река, циклоп, весло, стадо, странник, утес

Кластер 2: I, II, III, IV, XIV, XV, XVI, XVII, XVIII, XIX, XX, XXI, XXII, XXIII, XXIV

Термины 2-го кластера: жених, пенелопа, телемах, евмей, свинопас, лук, сын, еврикля, старец, менелай, антиной, нестор, атрид, телемах, странник, рассказывать, стрела, царица, копье

4. Леммы, стоп-слова включены.

Кластер 1: III, IV, V, VI, VII, VIII, IX, X, XI, XII, XIII, XV

Термины 1-го кластера: алкиной, феакийский, царь, менелай, цирцея, ветер, спутник, конь, нестор, брег, посейдон, пещера, дева, товарищ, град, колесница, стадо, странник, весло, баран

Кластер 2: I, II, XIV, XVI, XVII, XVIII, XIX, XX, XXI, XXII, XXIII, XXIV

Термины 2-го кластера: жених, пенелопа, телемах, евмей, свинопас, лук, еврикля, антиной, старец, сын, копье, стрела, странник, служанка, одиссей, старушка, хитроумный, господин, сказать

¹⁰ Все получившийся модели доступны в бинарном виде по адресу:
<https://github.com/nevmenandr/Zhukovsky/tree/master/clusterization>

На первый взгляд, все эти автоматически полученные классификации пересекаются с заранее заданным делением на песни относительно присутствия в них метрических аномалий, но все же совпадают с ним не вполне. Как узнать, какой из результатов кластеризации лучше согласуется с нашим делением, и согласуется он с ним значимо или не более чем случайно? Для этого в статистике есть специальная методика вычисления κ («каппы») Коэна, то есть меры согласия¹¹. Чем больше κ (в диапазоне от 0 до 1), тем больше мера согласия между разными классификациями. Если посчитать κ для деления на «аномальные» и «неаномальные» песни с одной стороны и для полученных результатов кластеризации с другой, то получатся следующие значения:

Словоформы, стоп-слова исключены: $\kappa = 0.5102$

Словоформы, стоп-слова включены: $\kappa = 0.5294$

Леммы, стоп-слова исключены: $\kappa = 0.5294$

Леммы, стоп-слова включены: $\kappa = 0.5833$

Таким образом, с нашей предварительной классификацией лучше всего согласуется та модель, которая построена на лемматизированном тексте, из которого не были исключены стоп-слова. Значима эта согласованность или она могла получиться и в результате случайного распределения? На этот вопрос отвечает параметр p -value, который вычисляется одновременно с κ . Считается, что если он меньше, чем 0.05, то согласованность значима и вероятность случайного совпадения минимальна. Для той модели, где подсчеты строились на леммах и включали стоп-слова, p -value = 0.00316, то есть согласие двух классификаций оказывается значимо.

Мы визуализировали нашу кластеризацию, ее можно увидеть на рис. 2. График отражает тематическое сходство между песнями, подсчитанное как косинусное расстояние между векторами их TF-IDF. Цветами обозначена принадлежность песни к одному из двух кластеров.

¹¹ См. Cohen J. A coefficient of agreement for nominal scales // Educ. Psychol. Measurement. 1960. V. 20. P. 37—46; Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit // Psychol. Bull. 1968. V. 70, P. 213—220.

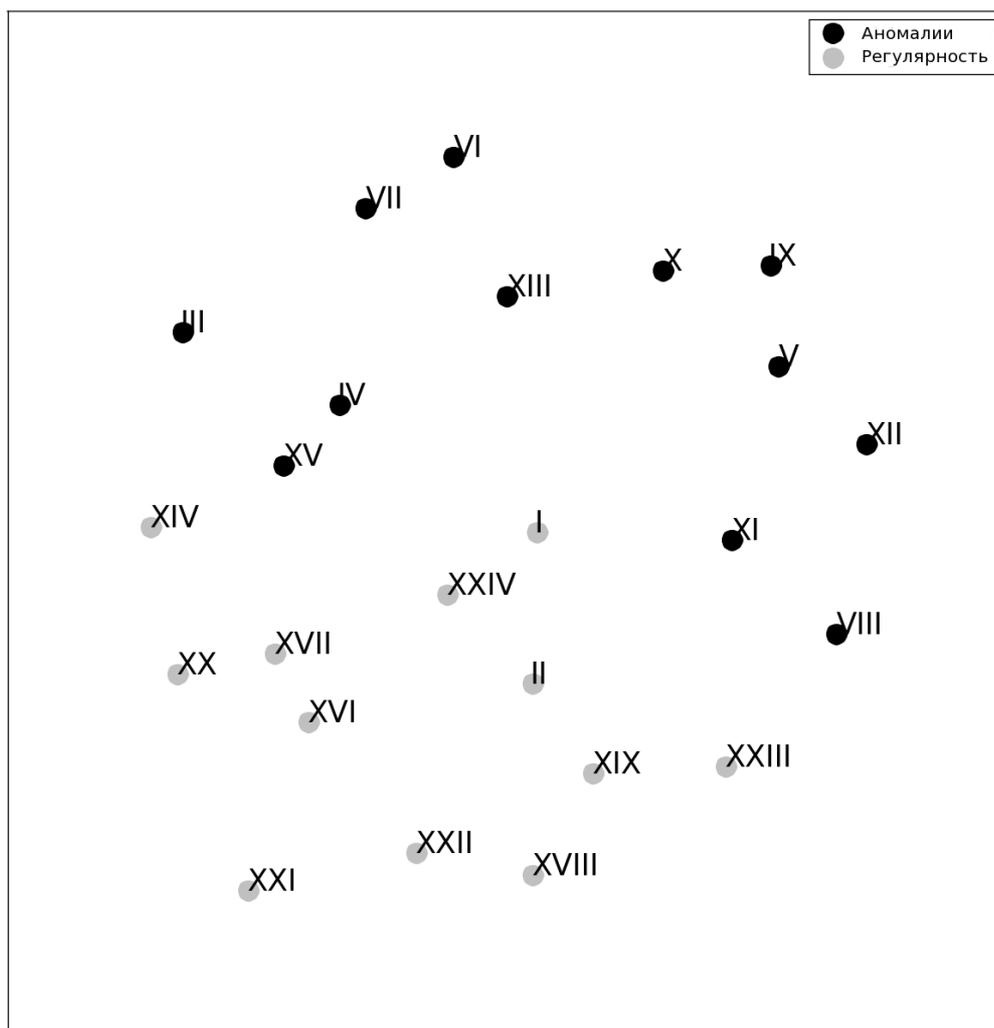


Рис. 2. Кластеризация песен «Одиссеи» по TF-IDF методом k -средних.

Следует отметить, что разрежённость точек-песен и их почти равномерное распределение на графике свидетельствуют об отсутствии ярко выраженных тенденций к объединению песен в группы-кластеры и тематической противопоставленности этих групп между собой. Тем не менее, для наблюдателя информативна удалённость некоторых песен друг от друга. Например, соседство VI и VII песен нельзя назвать случайным, обе они рассказывают о том, что произошло с Одиссеем на острове феаков; а IX и XXI песни находятся в разных концах пространства кластеризации, поскольку тематически не пересекаются: в первой речь идет о начале возвращения главного героя из под стен Трои, а во второй события развиваются в сугубо сухопутной обстановке и приводят в конце концов к избиению женихов Пенелопы.

Итак, опыт кластеризации на основе распределения частотности слов в песнях (которое в конечном счете имеет в виду автоматическое определение тематики текста) показал, что кластеры хотя и не вполне строго, но в целом соответствуют делению на песни с метрическими аномалиями и без них.

В целом не удивительно, что формально проанализированное содержание середины поэмы отличается от обрамляющего её контура, это подтверждает наше интуитивное представление о тематической разнице между двумя группами песен, одну из которых можно условно назвать «морской», а вторую «сухопутной», при этом метрические аномалии появляются главным образом в морских песнях, а в сухопутных себя практически не проявляют.

Разумеется, выявленная сюжетная закономерность может оказаться простым совпадением: нельзя исключать, что Жуковский «ошибался» в середине работы над «Одиссеей» (и, что важно, не исправил свои ошибки) по внешним причинам, никак не связанным с планом содержания. Но упоминавшееся эпигенетическое правило заставляет видеть в этом систему.

Однако контекст песни слишком широкий, с его помощью мы не сможем объяснить, почему аномалии появляются в той или иной конкретной строке. Тем не менее, такой исследовательский вопрос должен быть поставлен: есть ли в аномальных стихах или в их ближайшем окружении такие факторы, которые приводят к появлению отступлений от метрической схемы именно в этом месте поэмы, а не в другом?

Ответ на этот вопрос с помощью традиционного герменевтического подхода вызывает серьезные затруднения: перед глазами исследователя оказывается слишком большой по объему материал. Одновременно держать в уме сразу 49 случаев со всеми их контекстами чрезвычайно непросто; еще сложнее определить, насколько найденная закономерность (если представить себе, что мы ее найдем) значима именно для анализируемого контекста и не распределяется ли она равномерно по всему тексту поэмы. Хотя именно здесь на помощь могут прийти методы анализа данных, остается и другая проблема: не ясно, какого рода факторы мы должны искать: к аномалиям на уровне метрики приводит какое-то явление из области фоники, лексики, грамматики или той же метрики?

Без привлечения компьютерных аналитических инструментов нам не удалось обнаружить никаких значимых факторов, которые бы предсказывали появление аномалий в конкретных стихах «Одиссеи». Все мотивы, которые оказываются более-менее устойчиво связаны с аномальными строками (море, Афина, растение, ветер), равномерно распределены по всему остальному тексту поэмы, и не могут считаться оказывающими влияние на появление аномалий.

В некоторых стихах появляются имена собственные, которые могут по необходимости удлинять строку перевода:

Рядом они. Той порою Паллада Афина по улицам града (VIII, 7)

В то же время имена собственные весьма частое явление в поэме, где они встречаются в среднем один раз на 8 строк. Таким образом, необходимость их введения в перевод недостаточно для объяснения аномалий.

Конкретно в приведенной строке аномалии удалось бы избежать, убрав одно из имен богини. Такого рода случаи одиночного появления слова «Паллада» без «Афины» в тексте встречаются:

Паче ж других светлоокой богине, великой Палладе (II, 433)

О, когда б возлюбить светлоокая дева Паллада (III, 218)

Причина того, что стих VIII, 7 дошел до нас именно в этом виде, лежит либо в области авторского задания, либо авторского недосмотра, но никаких строго запрещающих правильный гекзаметр внутритекстовых причин за этим увидеть нельзя.

Алгоритм поиска внутритекстовых факторов появления метрических аномалий на уровне строк с помощью инструментов анализа данных выглядит так. Сначала нужно составить список параметров стиха, которые в принципе могли бы создавать предпосылки для

аномалий, хотя мы и не знаем, действительно ли они сами или в какой-то комбинации влияют на метрику строки. Затем следует исчислить эти параметры во всех строках и составить таблицу атрибутов, в которой каждому стиху были бы сопоставлены значения выделенных параметров. Наконец, используя эту таблицу и один из подходящих алгоритмов анализа данных, мы должны построить модель, которая бы показала, насколько хорошо наш набор параметров (и какие именно параметры из всего набора) объясняет 49 случаев аномалий среди 12106 строк всей поэмы.

Мы взяли для проверки следующие параметры: 1) позиция в песне (влияет ли то, в начале, в середине или в конце песни находится аномальная строка; все песни разделены на 5 частей); 2) количество слов в строке; 3) набор ударных гласных в строке; 4) частотность (TF-IDF) употребленных в строке слов; 5) набор частей речи в строке (и по одному атрибуту в таблице на каждую часть речи); 6) число имен собственных в строке; 7) знак конца предложения в середине строки; 8) является ли строка вопросительным предложением; 9) есть ли в строке отрицание; 10) является ли строка законченным предложением; 11) есть ли в строке противительные союзные конструкции; 12) метрическая схема строки.

Мы сознаем, что предложенные параметры во многом подобраны случайно, но это вынужденный шаг: мы имеем дело с явлением неизвестной внутритекстовой природы и заранее не можем предположить, какой именно фактор оказал влияние на ситуацию с аномалиями. Мы исходили из того, что какие-то из этих параметров, возможно, не являются важными сами по себе, но способны косвенно указать хотя бы на текстовый уровень, в котором кроется истинная причина «ошибок» Жуковского (например, значимость числа предлогов или союзов могла бы говорить о влиянии синтаксиса). Кроме того, значения всех этих параметров для каждой строки легко высчитывается автоматически благодаря простым операциям обработки текста. В таблице 2 можно увидеть пример конечного представления получившихся данных для первых двух строк «Одиссеи»; разумеется, для реальных строк число столбцов в такой таблице гораздо больше.

Наличие аномалии (0/1)	Позиция в песне (от 1 до 5)	Число слов в строке	Набор ударных гласных	Наличие редких слов (TF-IDF)	Набор частей речи	Число имен собственных
0	1	8	уиооуо	1	SVMPMAS M	1
0	1	9	аоаооу	0	VBPSCAS MV	1

Таблица 2. Фрагмент таблицы атрибутов строк по выбранным параметрам

Мы создали 4 таблицы атрибутов. В первой отражены значения перечисленных выше параметров для отдельных стихов за исключением метрической схемы строки, которая, очевидным образом, отличалась у аномальных и неаномальных строк, то есть являлась не фактором, а результатом появления метрической «ошибки». Эта таблица (именно ее фрагмент отражен в таблице 2) содержала 21 столбец-атрибут. Вторая таблица включала в себя все атрибуты первой, но к ней были добавлены подсчеты тех же самых атрибутов для предыдущих и последующих трех строк (с добавлением атрибутов метрических схем этих строк), таким образом, общее число атрибутов достигло 59. Эта таблица появилась благодаря гипотезе, что факторы, приводящие к появлению аномалий, кроются не в самих аномальных строках, а в их текстовом окружении. Третья таблица повторяла принцип построения второй, но включала в себя значения для 5 предыдущих и 5 последующих строк. Число атрибутов также равнялось 59. Наконец, четвертая таблица была сформирована так же, как и две предыдущие, но в ней окружающий контекст расширился до 7 строк.

Для анализа получившегося набора данных нами была выбрана комбинация CfsSubsetEval¹² как метода оценки значимости атрибута и разбросанного поиска (scatter

¹² Подробно о методе: Hall M. A. Correlation-based Feature Subset Selection for Machine Learning. Hamilton, New

search¹³) как метода задания множества потенциально значимых атрибутов. С их помощью были проанализированы все четыре сформированные на предыдущем этапе таблицы.

Лучший результат, который показал анализ атрибутов, был продемонстрирован на наборе данных для стиха с контекстом в 5 предыдущих и 5 последующих строк, а как значимые были отмечены сразу 16 атрибутов. Однако этот результат нельзя назвать удовлетворительным, его мера качества оценивается алгоритмом в 0.05481. Эта цифра означает, что если потенциально и существует какой-то фактор, на 100 % объясняющий появление метрических аномалий, то выявленный набор атрибутов справляется с этой задачей только на 5.5 %. Фактически таким результатом можно пренебречь.

Хуже всего справляется с объяснением аномалий таблица атрибутов, не учитывающая окружающий строку контекст: 0.03564, результат анализа таблицы, включающей контекст в 3 строки, очень похож на приведенный выше 0.05478, а обработка таблицы с 7-строчным контекстом, показала меру качества только 0.05335.

Итог рассмотрения контекста строки в том, что традиционное филологическое наблюдение не обнаруживает в стихе ничего специфического, а анализ данных подтверждает предположение об отсутствии очевидных причин появления метрических аномалий.

Общий вывод, учитывающий рассмотрение контекстов разного уровня, по всей видимости, в том, что на небольших наборах данных, которыми человек способен оперировать одновременно, традиционное герменевтическое погружение в текст даёт более интересные результаты, а количественно-статистические методы в целом пока грубы и на тонком поэтическом материале рисуют лишь приблизительную картину. Хуже всего для анализа данных подходит то, что описывается в терминах контингентности, то есть не необходимое, но и не случайное. С такого рода нечеткими закономерностями точные методы пока справляются плохо. Однако ценность количественного анализа возрастает с расширением материала, и даже если такой инструмент не подсказывает нам решения, то, по крайней мере, он способен подтвердить, что решение на данном материале и в рамках текущего набора параметров поиска найти невозможно.

Zealand, 1998.

13 См. Glover F., Laguna M., Martí R. Fundamentals of Scatter Search and Path Relinking // Control and Cybernetics. 2000. N 29 (3). P. 653—684