

Модель фактуры русских поэтических текстов

В. Б. БАРАХНИН*, О. Ю. КОЖЕМЯКИНА, И. В. КУЗНЕЦОВА,
В. А. КАРПОВА (МИРОНОВА)

Федеральный исследовательский центр информационных и вычислительных технологий,
630090, Новосибирск, Россия

*Контактный автор: Барахнин Владимир Борисович, e-mail: bar@ict.nsc.ru

Поступила 5 апреля 2021 г., доработана 24 мая 2021 г., принята в печать 31 мая 2021 г.

При исследовании зависимости смыслового содержания поэтических текстов от их метроритмических характеристик и строфики возникает необходимость в описании фактур — совокупностей названных характеристик поэтических текстов. В статье сформулировано определение фактуры — структурной модели поэтического текста, которая однозначно определяет метроритмический и строфический шаблон стихотворного текста; представлены алгоритмы определения метра и стопности, а также строфики стихотворения, позволяющие определить его фактуру. В ходе работы создана таблица всех фактур, встречающихся в стихотворениях А.С. Пушкина, с приведением примеров строф для каждого типа фактуры.

Ключевые слова: анализ поэтических текстов, определение метра, метроритмический анализ, идентификация рифмы, фактура.

Цитирование: Барахнин В.Б., Кожемякина О.Ю., Кузнецова И.В., Карпова (Миронова) В.А. Модель фактуры русских поэтических текстов. Вычислительные технологии. 2021; 26(3):107–117. DOI:10.25743/ICT.2021.26.3.007.

Введение

Уровни структуры стиха представляют собой определенную иерархию: метр, ритм, фонетика, лексика, грамматика, речевой жанр (композиционно-речевое целое), мотивика, что хорошо коррелируется с многоуровневой моделью информации, изложенной, например, в работе немецкого исследователя В. Гитта [1]. При этом процесс анализа стиха предусматривает первоначальное рассмотрение каждого уровня как самостоятельной смысловой единицы с последующим связыванием этих наблюдений с другими элементами структуры. Так как исследователи-литературоведы имеют дело с корпусами поэтических текстов большого объема, возникает необходимость автоматизации анализа различных уровней структуры стиха, что позволит освободить исследователей от рутинной работы и при этом резко расширить круг анализируемых авторов.

Базисной единицей анализа характеристик фонометрического уровня поэтического текста являются так называемые фактуры, учитывающие строфику и метроритмику. Согласно определению “Поэтического словаря” [2] фактура стиха — это “совокупность технологических приемов поэта, которые наряду с идейно-художественным замыслом определяют степень его поэтической культуры и мастерства”, которая “проявляется главным образом в лексике, фонике, ритме стиха”.

В настоящем исследовании мы будем пользоваться “минимальным” определением фактуры, в котором будем учитывать только метрические и строфические характеристики. Однако более тонкий анализ поэтических текстов предполагает, что в рассмотрение должны приниматься также ритмические характеристики, учитывающие, в дополнение к “идеальной” метрике, еще и пиррихии, спондеи и т. п., а также тонкие фонетические характеристики в рифме: точность, открытость, исследуемые фоносемантикой звуковые характеристики и т. д.

Итак, фактурой мы будем называть совокупность метроритмических и строфических характеристик поэтического текста, таких как:

- 1) стихотворный размер — стопность и вид размера (метр);
- 2) схема строфы (строфика) — перечисление групп рифмующихся строк с учетом типа клаузул.

Метроритмические характеристики учитывают размер строфы, стопность каждой строки, при этом полнота/неполнота строк по умолчанию входит в строфические характеристики. Под строфическими характеристиками мы понимаем вид взаимного расположения рифмических цепей (смежный, перекрестный, охватный и т. д.) и количество слогов, объединенных ударением (что определяет рифму мужскую, женскую, дактилическую и т. д.).

Фактура стихотворения в целом — это фактура входящих в него строф в предположении, что оно равнострофическое.

Таким образом, фактура является структурной моделью поэтического текста и однозначно определяет метроритмический и строфический шаблон стихотворного текста.

1. Стихотворный размер

Методика определения метра и стопности стиха достаточно строго изложена в работе [3]. Однако, поскольку она предполагает идеальную акцентуацию текста, не учитывая реальные проблемы в анализе текстов, связанные с наличием омографов, проклитик и т. п., нами был предложен подход к модификации данной методики [4], который был успешно применен в разрабатываемой авторами статьи системе автоматического анализа поэтических текстов [5].

В статье [3] также приведен алгоритм определения строфики поэтического текста, который заключается в поиске рифмующихся строк с использованием веб-приложения “Большой словарь рифм” [6]. Это приложение принимает слово и возвращает множество слов, рифмующихся с ним. Поскольку отправка запросов веб-приложению для каждого интересующего слова — длительный процесс, а извлечение слов и множеств рифмующихся к ним слов, а также последующее сохранение их в базе данных требуют больших ресурсов, в настоящей работе использован альтернативный способ поиска рифмованных строк (см. разд. 2).

В данном разделе описаны математическая модель и алгоритм определения метра и стопности с учетом проблем, изложенных в [4].

Приведем основные понятия из теории стихосложения, необходимые для построения математической модели определения метра и стопности.

Стихотворный размер (метр) определяется числом и порядком чередования ударных и безударных слогов в стопах силлабо-тонического стихосложения.

Стопа — это последовательность одного или нескольких слабых (безударных) и одного сильного (ударного) слогов, чередующихся в определенном порядке.

Анакруза — слабые (безударные) слоги, предшествующие первому метрическому ударению в строке.

Клаузула — последний ударный слог в строке и все, что находится после него.

Мужская клаузула — окончание строки стихотворения ударным слогом.

Женская клаузула — окончание строки стихотворения одним безударным слогом.

Дактилическая клаузула — окончание строки стихотворения двумя безударными слогами.

Спондей — сверхсхемное ударение на слабом месте — утяжеленная стопа.

Пиррихий — пропуск схемного ударения — отсутствие ударений на сильных местах.

Математическая модель метра и стопности представляет собой ритмическую схему ударных и безударных символов.

Слово — это структурная единица языка, в том числе и поэтического текста. Его ритмическая схема представляет собой вектор символов, обозначающих ударные и безударные слоги; число слогов определяется по числу букв, обозначающих гласные звуки, $vowels = \{a, y, o, э, ы, и, e, ё, ю, я\}$. Исключение составляют дифтонги в записанных кириллицей заимствованных словах, читаемые так, как это принято на языке оригинала. Такой прием характерен, в частности, для творчества А.А. Блока: например, исходя из ритмического рисунка, в стихотворении “В кабаках, в переулках, в извивах . . .”:

*И мелькала за парюю пара . . .
Ждал я светлого ангела к нам,
Чтобы здесь, в ликованьи троттуара,
Он одну приобщицл небесам.*

Сочетание букв “уа” в слове “троттуара” (от франц. *trottoir*) произносится как один гласный звук [wa]. В поэме “Возмездие” сочетание букв “ау” в слове “Фауст” (нем. *Faust*) произносится как один гласный звук [au]:

*Так, с жизнью счет сводя печальный,
Презревши молодости пыл,
Сей Фауст, когда-то радикальный,
“Правел”, слабел. . . и всё забыл.*

Забегая вперед, опишем алгоритм обработки таких ситуаций: если автоматический анализ показывает, что в строке есть лишний с точки зрения ритмического рисунка слог и при этом в ней встречается сочетание букв “уа” или “ау”, то такое сочетание букв рассматривается как один гласный звук. Заметим, что строки с буквенными символами, не входящими в русский алфавит, мы не рассматриваем.

Разрабатываемая модель метра и стопности учитывает случаи, когда в слове больше одной ударной позиции (имеется пиррихий) и неоднозначной ударной позиции (омонимы). Неоднозначные ударные позиции в слове обозначаются как x . Для разрешения неоднозначности в расстановке ударений применяется метод “по аналогии” [4].

Ритмическая схема слова w определяется как вектор:

$$w = [c_0, \dots, c_i, \dots, c_n], \quad c_i \in \{0, 1, x\}, \quad i \in [0, n - 1].$$

Тогда ритмическая схема строки s стихотворения представляет собой объединение слов (упорядоченное множество векторов):

$$s = \cup w_i, \quad i \in [0, k - 1],$$

а ритмическая схема текста стихотворения — вектор, элементами которого являются строки:

$$t = [s_0, \dots, s_i, \dots, s_m], \quad i \in [0, m - 1].$$

Далее опишем алгоритм определения метра и стопности, который заключается в сопоставлении ритмических вариантов стиха изучаемого поэтического текста (ПТ) с набором ритмических шаблонов из определенного репертуара метроритмических вариантов стиха, в основе которого лежит модель, описанная выше.

Согласно алгоритму необходимо получить ритмическую схему каждого слова, владея информацией об акцентуации. Для этого необходимо слово разбить на слоги; определить, какие из них ударные/безударные; записать вектор, состоящий из 0 и 1, обозначающих безударные и ударные символы соответственно.

В общем виде ритмический вариант строки представляет собой объединение ритмических схем слов, он получается в результате действия оператора $strMask$ на строку s по правилу

$$strMask(s) = \cup w_i, \quad i \in [0, k - 1],$$

а слоговая схема всего стихотворения — в результате действия оператора $textMask$ на ритмическую схему стихотворения t по правилу

$$textMask(t) = \cup strMask(s_i), \quad i \in [0, m - 1].$$

Посредством действия оператора S на ритмическую схему текста стихотворения t восстанавливается значение слога на позиции x , где была неоднозначность в расстановке ударения: внутри каждой строки s_i определяется элемент $c_i = x$ из элементов с индексом $i \in [0, k - 1]$, создается вектор p размерности $\dim(p) = \dim(t)$, после чего вычисляется x как максимально часто встречаемый элемент в векторе p . Таким образом, ритмическая схема стихотворения с однозначной акцентуацией t' определяется как $t' = S(t)$.

Далее для каждой слоговой схемы строки $strMask(s)$ необходимо получить параметр классификации текстов по метру и стопности — число слогов в строке $R(s) = \dim(strMask(s))$.

Следующий параметр — вектор межакцентных интервалов $\mathbf{r}(s)$, он означает количество безударных слогов между i -м и $i + 1$ -м ударными слогами. Посредством действия оператора $accInterval$ образуются проекции вектора слоговой схемы строки $strMask(s)$ на подпространства, соответствующие координатам идущих подряд “безударных” символов с последующей операцией взятия размерности m_k каждой из проекций:

$$\begin{aligned} \mathbf{r}(s) &= accInterval(strMask(s)) = \\ &= (\dim(strMask(s)_1), \dim(strMask(s)_2), \dots, \dim(strMask(s)_k), \dots, \dim(strMask(s)_{n-1})), \\ strMask(s)_k &= [c_{k,0}, \dots, c_{k,m_{k-1}}], \quad c_{k,i} = 0, \quad i = 0, \dots, m_{k-1}. \end{aligned}$$

Кроме того, $anacrusa(s) = \dim(strMask(s)_0)$ — число безударных слогов в начале строки s (анакруза), $clausula(s) = \dim(strMask(s)_n)$ — число безударных слогов в конце строки s (клаузула), параметр $k = \sum_{i=1}^n c_i$ — число ударных слогов в строке.

Далее тексты классифицируются по пяти условиям из работы [4], которые определены с использованием функции $F(s) = R(s) - clausula(s)$.

1. Если существуют $i, j \in [0, \dim(t)]$ такие, что $F(s_i) \neq F(s_j)$, $k(s_i) \neq k(s_j)$ и при этом выполняется одно из условий:
 - $|F(s_i) - F(s_j)|$ кратны только 2,
 - $|F(s_i) - F(s_j)|$ кратны только 3,
 - $|F(s_i) - F(s_j)|$ кратны только 5,
 то ритмическая схема t представляет собой схему неравностоппного метрического стиха. Если же $|F(s_i) - F(s_j)|$ лежат в интервале натурального ряда, но не удовлетворяют приведенным выше условиям кратности, то ритмическая схема t представляет собой схему дисметрического стиха.
2. Если существуют $i, j \in [0, \dim(t)]$ такие, что $F(s_i) \neq F(s_j)$, $k(s_i) = k(s_j)$, то ритмическая схема t представляет собой схему k -акцентного стиха — неравностоппного или равностоппного.
3. Если существуют $i, j \in [0, \dim(t)]$ такие, что $F(s_i) \neq F(s_j)$, $k(s_i) \neq k(s_j)$, а $r(i) = r(j)$ — постоянный межакцентный интервал, то ритмическая схема t представляет собой схему неравностоппного k -стоппного силлабо-тонического стиха.
4. Если существуют $i, j \in [0, \dim(t)]$ такие, что $F(s_i) = F(s_j)$, $k(s_i) = k(s_j)$, а $r(i) = r(j)$ — постоянный межакцентный интервал, то ритмическая схема t представляет собой схему строгого равностоппного k -стоппного силлабо-тонического стиха.
5. Если существуют $i, j \in [0, \dim(t)]$ такие, что $F(s_i) = F(s_j)$, $k(s_i) \neq k(s_j)$, а $r(i) = r(j)$ — постоянный межакцентный интервал, то ритмическая схема стихотворения t представляет собой схему равностоппного изосиллабического стиха.

2. Строфика

Для нахождения рифмы использовалось правило, основанное на положениях из работы В.М. Жирмунского [7]: строки рифмуются, если у последних слов в строке одинаковая позиция ударного слога и окончания этих слов либо совпадают, либо являются парами созвучий, перечисленными в [7]. Разумеется, перечисленные пары созвучий отражают понимание рифмы в работах отечественных литературоведов начала XX в., но мы применяем наши алгоритмы именно к анализу либо классической русской поэзии, либо более современных авторов, работающих в русле классической традиции.

Множество, состоящее из номеров строк с одинаковыми акцентными позициями, обозначим как $A_j = [a_{j1}, a_{j2}, \dots, a_{jm}]$, а множество номеров строк с фонетически совпадающими окончаниями — как $E_i = [e_{i1}, e_{i2}, \dots, e_{im}]$. Тогда множество строк, рифмующихся друг с другом, есть пересечение множеств A_j и E_i :

$$R = A_j \cap E_i,$$

где $i < N$, $j < N$, N — число строк.

Алгоритм нахождения строфик заключается в сопоставлении шаблонов строфик (собранных из метрического справочника к стихотворениям А.С. Пушкина [8]) изучаемому поэтическому тексту.

Промежуточные результаты работы алгоритма определения строфики, проверяемого на романе в стихах А.С. Пушкина “Евгений Онегин”
The intermediate results of the algorithm for determining the stanza, which is tested on the novel in verses “Eugene Onegin” by A.S. Pushkin

Пример шаблона строфик	Рифмующиеся строки, [[жен.], [муж.]]	Точность
<i>ababccddehhekk</i>	[[[0, 2], [4, 5], [8, 11]], [[1, 3], ...]]	(0.96, “Onegin”)
<i>bb</i>	[[], [[0, 1]]]	(0.43, “bb”)
<i>abba</i>	[[[0, 3]], [[1, 2]]]	(0.22, “abba”)
<i>ababccb</i>	[[[0, 2], [4, 5]], [[1, 3, 6]]]	(0.48, “ababccb”)

Ниже представлены шаги алгоритма нахождения строфики ПТ.

- Шаблоны строфик из справочника переводятся из буквенного представления в множество индексов рифмующихся строк (например, строфика вида *abab* переводится в множество индексов рифмующихся строк $[[0, 2], [1, 3]]$). Для описания строфики используются последовательности букв латинского и греческого алфавитов согласно правилам, принятым в справочнике [8]: латинскими буквами обозначены клаузулы рифмованных стихов, греческими буквами — клаузулы белых стихов. Буквы обоих алфавитов, выдающиеся над строкой (например, *b, d, δ, θ*), обозначают мужские клаузулы; буквы, не выдающиеся над строкой (например, *a, c, e, α*), — женские клаузулы; буквы, спускающиеся под строку (например, *g, j, γ*), — дактилические клаузулы; *f* и *ψ* — гипердактилические клаузулы.
- Фиксируется информация о том, какие из строк данного ПТ имеют мужские/женские клаузулы (чтобы различать, например, случаи *abab* и *baba*):

$$R = [[M_0, M_1, \dots, M_i, \dots, M_n], [F_0, F_1, \dots, F_i, \dots, F_m]],$$

т.е. R — множество, состоящее из множеств номеров строк с мужскими ($M_i, i \in [0, n - 1]$) и женскими ($F_i, i \in [0, m - 1]$) клаузулами.

- Из множества строк с мужской клаузулой образуется подмножество всевозможных пар строк, мощность которого равна C_n^2 :

$$[M_0, M_1, \dots, M_i, \dots, M_{n-1}] = [[M_0, M_1], [M_0, M_2], [M_0, M_3], \dots, [M_0, M_{n-1}], [M_v, M_{n-1}]],$$

где M_{i1} — последняя строка с мужской клаузулой. Это сделано с целью проверки, удовлетворяет ли данная пара строк правилам образования рифмы.

- Для каждого шаблона строфики ищется точность определения. В качестве строфики выбирается шаблон с максимальной точностью:

$$precision = good/count,$$

где *good* — индикатор того, что строки рифмуются с друг другом, *count* — счетчик всех проверяемых пар строк.

- Аналогичным образом шаги 3 и 4 выполняются для строк с женской клаузулой.

Для наглядности реализации алгоритма поиска строфики ПТ приведем некоторые из проверяемых шаблонов и их процент точности, с которым они были обнаружены в данном стихотворении (см. таблицу).

3. Фактуры

Как было сказано выше, фактура определяется как совокупность метроритмических характеристик, она однозначно задает шаблон стихотворного текста. Приведем пример получаемого шаблона для стихотворения А.С. Пушкина “К Наташе”:

*Вянет, вянет лето красно;
Улетают ясны дни;
Стелется туман ненастной
Ночи в дремлющей тени;
Опустели злачны нивы;
Хладен ручеек извивый,
Лес кудрявый поседел;
Свод небесный побледнел.*

Это стихотворение соответствует фактуре “*ababccdd 4x*”.

Таким образом, математической моделью фактуры является вектор характеристик **Facture** = [*k*, *M*, *S*]:

- 1) *k* — стопность (целочисленное значение или особые символы для вольного стиха (В), долownika (дол) и т. п.);
- 2) *M* — метр (элемент из множества {хорей (х), ямб (я), дактиль (д), амфибрахий (ам), анапест (ан), должник (дол) });
- 3) *St* — строфика (слово из символов латинского или греческого алфавитов в соответствии с правилами из справочника [8], описанными в разд. 2).

Этот вектор учитывает и отражает все рассматриваемые метроритмические характеристики текста, перечисленные выше. Еще раз подчеркнем: по этой схеме описываются только равнострофические стихотворения; стихотворения разностопные, со свободной строфикой и т. п., требуют для описания их составных фактур использования более сложных структур.

Приведем поэтапное описание работы, связанной с определением фактур.

1. На основании данных “Метрического справочника” [8] составлена полная таблица примеров фактур (от одного до трех примеров на каждую фактуру), используемых Пушкиным в равнострофических монометрических поэтических текстах. Таблица состоит из следующих столбцов:
 - уникальный индекс фактуры;
 - метр;
 - стопность;
 - строфика;
 - пример строфы с такой фактурой;
 - номер примера в издании “Красной Нивы” [9];
 - год написания стихотворения с соответствующей фактурой.
2. Проведен анализ полученных результатов, выявлены и учтены неточности, допущенные в справочнике [8].
3. Сведения, представленные в таблице, занесены в базу данных с присвоением каждой фактуре уникального индекса.
4. Разработан алгоритм определения фактуры, на его основе реализована программа на языке Python, определяющая фактуру входного стихотворения.
5. Результаты работы программы представляются в следующих форматах:

- текст в интерфейсе среды разработки;
- файл в формате JSON [10];
- файл в формате CSV [11], такое представление удобно использовать при анализе больших корпусов текстов, объединенных общим свойством (например, произведений одного автора).

Всего в собрании сочинений А.С. Пушкина [9] найдено 82 типа фактур (ввиду ограниченности объема статьи полный список не приводим). Классификация фактур по размеру имеет вид:

- 3х — 3 типа;
- 4х — 8 типов;
- 3я — 3 типа;
- 4я — 26 типов;
- 5я — 13 типов;
- 6я — 10 типов;
- Вя — 8 типов;
- 2д — 2 типа;
- 2ам — 2 типа;
- 4ам — 4 типа;
- 2ан — 2 типа;
- дол — 1 тип.

Приведем перечень всех встречающихся видов строф различной длины (правила буквенного обозначения строфики подробно описаны в разд. 2):

- 2 — aa, bb ;
- 3 — baa ;
- 4 — $abab, baba, \alpha\beta\alpha\beta, \beta\alpha\beta\alpha, aabb, bbaa, baab, \delta b\delta b, abba, bfbf$;
- 5 — $abaab, ababa, ababb, babba, abbba, bbaba, aabab$;
- 6 — $ababab, bababa, abbaab, baabab, aabccb, babadd, bbadad, abbadd, baabcc, bbaadd, aabccb$;
- 7 — $ababaab$;
- 8 — $ababccdd, ababccdd, ababcdcd, babadccd, abababcc, ababcdcd, baabcdcd, baabcdde, bababadd$;
- 9 — $bbabaddcb$;
- 10 — $bbaddcchh, ababccdeed, babaddchch$;
- 14 — $ababccdehhekk, ababababccbebe, baabbabaddckck$.

Для составления таблицы были использованы материалы “Метрического справочника” [8]: табл. XVIII и алфавитный индекс стихотворений. При этом в “Метрическом справочнике” были обнаружены некоторые неточности: в табл. XVIII указана только 51 фактура (сам термин в справочнике не употребляется — в аналогичном смысле используется термин “строфический вид”), а также 25 видов сложных строф, встречающихся только в однострофных стихотворениях. Еще 31 фактура либо не упоминается (хотя указаны в алфавитном индексе, например “ $ababa$ 5я” или “ $baab$ 6я”), либо отнесена к другому типу строфических видов, но мы, исходя из нашего определения фактуры, считаем нужным разделять подобные виды: например “ aa 4я”, “ $aabb$ 4я”, “ $bbaa$ 4я”.

Подчеркнем, что в данном исследовании рассматриваются только равнострофические монометрические стихотворения, статистика которых приведена в табл. XVIII “Метрического справочника”. Таким образом, в процессе работы были проверены и

разобраны все произведения из алфавитного индекса. Фактуры всех найденных равнострофических стихотворений занесены в нашу таблицу.

Однако отметим, что в стихотворениях со свободной строфикой также могут встречаться уникальные фактуры — см. табл. III “Строфический индекс” “Метрического справочника”. Большое количество различных строфик, указанных в данной таблице, объясняется тем, что при анализе учитываются не только равнострофические стихотворения, но также и произведения со свободной строфикой, при этом произведение может состоять из нескольких строф, которые относятся к разным типам фактур. Например, строфы стихотворения “Пророк” А.С. Пушкина таковы: “4я *baba*”, “4я *bbaa*”, “4я *bbadda*”, “4я *bbadda*”, “4я *baba*”, “4я *bb*”, “4я *abab*” (совпадение обозначений в разных строфах не означает совпадения соответствующих рифм). В таком случае произведение нельзя отнести к одному типу фактуры: для его различных строф определяются соответствующие типы фактур.

Заключение

Построена математическая модель метра и стопности стихотворения на основе методики И.А. Пильщикова из работы [3], модифицированной нами в работе [4] посредством учета возможности неоднозначной расстановки ударений, связанной с наличием омографов, проклитик, пиррихий и т. п. На основании дополненной модели был разработан и программно реализован модифицированный алгоритм определения метра и стопности.

Построена и программно реализована математическая модель автоматического определения строфики стихотворения.

Сформулировано определение фактуры — структурной модели поэтического текста, которая однозначно задает метроритмический и строфический шаблон стихотворного текста; изложена математическая модель фактуры. На основании данных “Метрического справочника” [8] составлена полная таблица примеров фактур, используемых А.С. Пушкиным в равнострофических монометрических поэтических текстах. Разработан алгоритм определения фактуры, на его основе реализована программа на языке Python, определяющая фактуру анализируемого стихотворения.

Благодарности. Работа выполнена при финансовой поддержке РФФ (проект № 19-18-00466).

Список литературы

- [1] Gitt W. Ordnung und information in technik und natur. Gitt W. (Hrsg.): Am Anfang war die Information. Gräfeling: Resch KG; 1982: 171–211.
- [2] Квятковский А.П. Поэтический словарь. М.: Советская энциклопедия; 1966: 276.
- [3] Бойков В.Н., Каряева М.С., Соколов В.А., Пильщиков И.А. Об автоматической спецификации стиха в информационно-аналитической системе. Аналитика и управление данными в областях с интенсивным использованием данных: Труды Международной научной конференции DAMDID/RCDL-2015 (Обнинск, 15–16 октября 2015 г.); 144–151.
- [4] Varakhnin V.B., Kozhemyakina O.Yu., Kuznetsova I.V. Development and implementation of the algorithm for automatic analysis of metrorhythmic characteristics of Russian poetic texts. CEUR Workshop Proceedings. 2019; (2523):290–298.

- [5] Система анализа поэтических текстов ФИЦ ИВТ. Адрес доступа: <http://поем.ict.nsc.ru/meterrhyme> (дата обращения: 05.04.2021)
- [6] Большой словарь рифм. Адрес доступа: <https://rifmovnik.ru/find.htm> (дата обращения: 05.04.2021)
- [7] **Жирмунский В.М.** Рифма, ее история и теория. Петроград: Academia; 1923: 339.
- [8] **Лапшина Н.В., Романович И.К., Ярхо Б.И.** Метрический справочник к стихотворениям А.С. Пушкина. М.; Л.: Academia; 1934: 144.
- [9] **Пушкин А.С.** Полное собрание сочинений: В 6 т. М.; Л.: Гослитиздат; 1930–1931. (Приложение к журналу “Красная нива”).
- [10] **Crockford D.** The application/JSON media type for JavaScript Object Notation (JSON). Internet Engineering Task Force. 2006: 10. DOI:10.17487/RFC4627.
- [11] Common format and MIME type for Comma-Separated Values (CSV) files. Адрес доступа: <https://www.rfc-editor.org/rfc/rfc4180.txt> (дата обращения: 05.04.2021).

The model of facture of Russian poetic texts

BARAKHNIN VLADIMIR B.*, KOZHEMYAKINA OLGA YU., KUZNETSOVA IRINA V.,
KARPOVA (MIRONOVA) VERA A.

Federal Research Center for Information and Computational Technologies, 630090, Novosibirsk, Russia

*Corresponding author: Barakhnin Vladimir B., e-mail: bar@ict.nsc.ru

Received April 5, 2021, revised May 24, 2021, accepted May 31, 2021.

Abstract

The dependence of the semantic content of poetic texts on their metrorhythmic characteristics and the structure of stanzas requires describing the factures — the combination of these characteristics of poetic texts.

The purpose of this article is to formulate the definition of facture as a structural model of a poetic text, which uniquely determines the metrorhythmic and strophic pattern of a poetic text, and also to develop the algorithms for automatic determination of facture.

To achieve this goal, an algorithm for automatic determination of meter and number of feet, as well as the algorithm for automatic determination of the structure of stanzas of poems, has been developed. A distinctive feature of the algorithm for automatic determination of meter and number of feet of poems is to take into account the real problems of accentuation, which are associated with the presence of homographs, proclitics, etc., moreover, in the case of detection of a lexeme in a line that cannot be unambiguously accentuated using the automatic accentuation methods, it is used to establish the possible accentuation “by analogy” based on the accentuation of unambiguously accentuated lines. The algorithm for automatic determination of the structure of stanzas is based on the identification of the rhyming lines on the basis of their consonance.

The testing of the algorithm was carried out on the corpus of poems by A.S. Pushkin, for which the "Metric Reference Book" was manually created in the 1930s with a description of metrorhythmics and stanzas. The developed algorithm showed the almost complete compliance of the results of the

automatic determination of the factures with the metrorhythmic and strophic characteristics given in the reference book.

Thus, the algorithm for automatic facture determination which is developed and programmatically implemented by the authors of the paper, can be a reliable tool for the complex analysis of metrorhythmic and strophic characteristics of Russian poetic texts.

Keywords: analysis of poetic texts, determination of meter, metrorhythmic analysis, identification of rhyme, facture.

Citation: Barakhnin V.B., Kozhemyakina O.Yu., Kuznetsova I.V., Karpova (Mironova) V.A. The model of facture of Russian poetic texts. Computational Technologies. 2021; 26(3):107–117. DOI:10.25743/ICT.2021.26.3.007. (In Russ.)

Acknowledgements. This research was supported by RSF (project No. 19-18-00466).

References

1. **Gitt W.** Ordnung und information in technik und natur. Gitt W. (Hrsg.): Am Anfang war die Information. Gräfeling; Resch KG; 1982: 171–211.
2. **Kvyatkovskiy A.P.** Poeticheskiy slovar' [The poetry dictionary]. Moscow: Sovetskaya Entsiklopediya; 1966: 276. (In Russ.)
3. **Boikov V.N., Karyaeva M.S., Sokolov V.A., Pilshchikov I.A.** Ob avtomaticheskoy spetsifikatsii stikha v informatsionno-analiticheskoy sisteme [About the automatic verse specification in the information analytical system]. Data Analytics and Management in Data-Intensive Areas Subjects. Proceedings of the International Scientific Conference DAMDID/RCDL2015. Obninsk, October 15–16, 2015: 144–151. (In Russ.)
4. **Barakhnin V.B., Kozhemyakina O.Yu., Kuznetsova I.V.** Development and implementation of the algorithm for automatic analysis of metrorhythmic characteristics of Russian poetic texts. CEUR Workshop Proceedings. 2019; (2523):290–298.
5. The system of the analysis of poetic texts FRC ICT. Available at: <http://poem.ict.nsc.ru/meterrhyme> (accessed 05.04.2021) (In Russ.)
6. Great rhyme dictionary. Available at: <https://rifmovnik.ru/find.htm> (accessed 05.04.2021) (In Russ.)
7. **Zhirmunskiy V.M.** Rifma, ee istoriya i teoriya [The rhyme, its history and theory]. Petrograd: Academia; 1923: 339. (In Russ.)
8. **Lapshina N.V., Romanovich I.K., Yarkho B.I.** Metricheskiy spravochnik k stikhotvoreniyam A.S. Pushkina [The metrical reference book for the poems of A.S. Pushkin]. Moscow; Leningrad: Academia; 1934: 144. (In Russ.)
9. **Pushkin A.S.** Polnoe sobranie sochineniy. V 6 tomakh [Full collection of writings: In 6 volumes]. Moscow; Leningrad: Goslitizdat; 1930–1931. (Appendix to the magazine “Krasnaya Niva”). (In Russ.)
10. **Crockford D.** The application/JSON media type for JavaScript Object Notation (JSON). Internet Engineering Task Force. 2006: 10. DOI:10.17487/RFC4627.
11. Common format and MIME type for Comma-Separated Values (CSV) files. Available at: <https://www.rfc-editor.org/rfc/rfc4180.txt> (accessed 05.04.2021).