

# ПОЭТИЧЕСКИЙ ПОДКОРПУС НАЦИОНАЛЬНОГО КОРПУСА РУССКОГО ЯЗЫКА КАК АКЦЕНТОЛОГИЧЕСКИЙ ИСТОЧНИК

## POETRY SUB-CORPUS OF RUSSIAN NATIONAL CORPUS AS AN ACCENTOLOGICAL SOURCE

*Корчагин К.М. ([stivendedal@gmail.com](mailto:stivendedal@gmail.com))*

Московский государственный институт радиотехники, электроники и автоматики (технический университет)

Рассматривается применение поэтического подкорпуса НКРЯ для акцентологических исследований. Дается краткое описание элементов метаразметки поэтического текста, важных для его акцентологической интерпретации, рассматриваются основные типы стихотворных метров и пригодность их использования в качестве готовых акцентологических схем для выбранных словоформ.

### 1. Поэтический корпус

Поэтический (под)корпус – часть НКРЯ<sup>1</sup> со специфической *метаразметкой*, в которой отражены основные жанровые и формальные параметры поэтического текста.

Присутствие данной метаразметки позволяет программными средствами восстановить акцентную схему каждой входящей в рассматриваемый поэтический текст словоформы с определенной точностью (см. ниже). Несмотря на то, что диахрония и синхрония русского ударения более чем подробно рассмотрены в работе [Зализняк 1985], использования поэтических текстов позволяет существенно уточнить некоторые отдельные факты. В историческом аспекте это особенно важно из-за скудости акцентуированного языкового материала XVIII-XIX вв.; в аспекте современного состояния русской акцентной системы это позволяет зафиксировать факты, характерные уже не только для современного разговорного языка, но являющиеся частью формирующейся литературной нормы. Также становится возможным проследить судьбу некоторых относительно распространенных акцентных архаизмов или варваризмов.

### 2. Языковой материал в поэтическом корпусе

Рассмотрим представление языковой информации в поэтическом корпусе.

Интерфейс поиска, в целом, одинаков для поэтического и основного корпусов НКРЯ. Однако в поэтическом корпусе существует дополнительный набор метатекстовых атрибутов, позволяющих осуществлять поиск по характерным параметрам поэтического текста. Кратко рассмотрим те параметры, которые принципиальны для нашей работы:

- *Метр*. Традиционные силлабо-тонические метры: ямб, хорей, дактиль, амфибрахий, анапест.

Силлабические метры обозначаются, напр., *С12ж* в строке запроса или указанием параметра «силлабический» в меню.

Тонические метры представлены строчными логэдами различного вида, дольником, тактовиком и акцентным стихом; также возможно задание обобщенной характеристики метра — «тонический».

Кроме того параметр «метр» может принимать значения «*свободный стих*» и «*гетерометрия*». Специально отмечаются гекзаметр и пентаметр как наиболее

---

<sup>1</sup>См. <http://www.ruscorpora.ru>; поэтический корпус доступен по адресу <http://www.ruscorpora.ru/search-poet.html>.

традиционные для русской поэзии типы дольника (ср. [Гаспаров 2001 : 141], где проясняется почему гекзаметр и его дериваты принято относить к тоническим метрам<sup>2</sup>).

- *Стопность*. В меню представлены наиболее распространенные типы стопности: однородная (2, 3, 4-ст. и т.д.), вольная (3,4; 4,6-ст. и т.д.), урегулированная (4+2, 4+3-ст. и т.д.). В строке запроса допускается ввод значений типа 4(5), по которым отбираются 4-ст. стихотворения со спорадически возникающими 5-ст. строками. Этот параметр действует и для тонических (означает количество иктов) и для силлабических стихов (означает количество слогов); не применяется к верлибрам.
- *Клаузула*. В меню указаны наиболее частотные типы клаузул, однако (особенно это актуально для сложной строфики) пользователь также может указать нужное чередование в строке запроса. Также возможен поиск по обобщенному типу клаузулы — вольной или регулярной — без дополнительных уточнений.

Более подробно особенности этих параметров в контексте нашей темы будут рассмотрены ниже.

Силлабо-тонические и тонические тексты поэтического корпуса снабжены обозначениями сильных мест стопы (иктов) в соответствии с типом метра. Для силлабо-тонических текстов это (за редкими исключениями<sup>3</sup>) позволяет указать одну из возможных позиций ударения в словоформе. Для тонических текстов, однако, надо иметь ввиду возможную вариативность ударения<sup>4</sup>, ограниченную, однако, требованиями соблюдения метрической схемы (особенно строгой в дольнике). Для силлабических текстов ударение фиксируется только в зоне клаузулы, т.к. убедительно судить об акцентуации форм из этих текстов нельзя<sup>5</sup>.

Метрическая разметка дополняется морфологической разметкой, данные которой позволяют сравнить реально наблюдающуюся акцентуацию словоформы (в метрической схеме) и теоретическую, основанную на данных русской грамматики. Наиболее любопытны, конечно, случаи расхождения этих двух параллельных источников, т.е. случаи, когда, пользуясь техническими терминами стиховедения, ударение падает не на сильное, а на слабое место в строке. Примеры такого расхождения мы приведем в последнем разделе нашей работы.

---

<sup>2</sup> Следует заметить, что речь идет о формах *гекзаметра*, не совпадающих с Дбж. Последние в поэтическом подкорпусе НКРЯ размечаются как регулярные дактили.

<sup>3</sup> Например, в случае нерегулярных *перебоев*, т.е. наличия ударения на слабом месте стопы одновременно с его отсутствием на сильном. Ср. в [Гаспаров 2001 : 109] пример из М. Цветаевой. Впрочем, особо «вопиющие» случаи перебоев получает в поэтическом подкорпусе специальную разметку, позволяющую рассматривать их как дольники или логазды. Ср. стихотворение А. Сумарокова с перебоем (правда, регулярным) на первой стопе трехстопного ямба:

Бла`гополу`чны дни`  
На`шими вре`мена`ми;  
Ве`селя мы` одни`...

Регулярность перебоев позволяет постулировать тут логазд типа 0\*2\*1\*. В случае частотных нерегулярных перебоев будет считаться, что текст содержит вкрапления *дольника* (или др. тонического метра).

<sup>4</sup> Ср. пример, приведенный М. Гаспаровым: *радио vs радио́* (на французский манер) в тексте В. Брюсова «Тетрадь» [Гаспаров 2001 : 103].

<sup>5</sup> Тем более, что основная часть русских силлабических текстов написана в XVII – нач. XVIII вв., причем часто выходцами из Украины, Белоруссии, Польши, привыкшими либо к известной им церковнославянской норме либо к фиксированному ударению польского образца. Для современных образцов зачастую удобнее трактовать некоторые изосиллабические тексты как равносложные дольники (в корпусе планируется ввести специальную дополнительную помету — *изосиллабичная тоника*).

### 3. Системы стихосложения и акцентуация словоформ

**3.1.** Наибольшей предсказательной силой обладает силлабо-тоническая система стихосложения<sup>6</sup>. Обращения к использующим ее текстам позволяют определить положение ударения с точностью до стопы. Тем самым, гипотетическая акцентная схема конкретной словоформы представляет собой регулярно повторяющуюся последовательность сильных и слабых мест. Точность предсказания расположения ударения меняется в зависимости от длины словоформы и от силлабического «веса» стопы.

Как известно, русская силлабо-тоника различает стопы с силлабическими «весами» 2 и 3:

|                 |                  |
|-----------------|------------------|
| <i>вес = 2:</i> | <i>вес = 3:</i>  |
| 01 — Ямб        | 001 — Анапест    |
| 10 — Хорей      | 010 — Амфибрахий |
|                 | 100 — Дактиль,   |

где «1» означает сильное (гипотетически ударное) место стопы.

Метрическая схема (в традиционных описаниях = метр), таким образом, представляет из себя «праформу» чередования сильных и слабых мест стиха, подвергающуюся определенным акцентным преобразованиям, конечным результатом которых выступает конкретная акцентная реализация этой схемы (= ритм). Разметка поэтического подкорпуса НКРЯ воспроизводит метрическую схему в соответствии со значениями определенных параметров. Тем самым, строка Я4<sup>7</sup> в корпусе может выглядеть так (1):

(1) *Над вы́сью пла́менной Си́ня*<sup>8</sup>.

Точность реальной акцентной схемы этой строки, как мы видим, зависит от длины словоформ: так, для двусложной формы *вы́сью* ударение определяется однозначно<sup>9</sup>, а реальная акцентная схема формы *пла́менной* предполагает два варианта; форма же *Си́ня* также допускает только один вариант ударения, т.к. находится в финальной позиции строки и реальное ударение фиксируется в ней требованиями размера (в данном случае женской клаузулой). Разумеется, эта проблема не стоит для форм, не предполагающих разнообразия акцентных вариантов, но, например, различие книжной устаревшей формы *библиотека* и современной *библиоте́ка* попадает ровно в этот класс случаев.

Для стоп, силлабический «вес» которых равен трём, предсказание может быть осуществлено более точно. Так, Аф3ж выглядит следующим образом (2):

(2) *Однóй мы живём и мечтóю,*

где акцентуация всех форм может быть вычислена с абсолютной точностью.

**3.2.** Тексты, использующие силлабическую систему стихосложения, в последние два с половиной столетия появляются крайне редко и сами по себе содержат только одну в большинстве случаев закреплённую позицию ударения — *клаузулу*<sup>10</sup>. Это сильно сужает множество рассматриваемого материала, особенно, если принять во внимание, что в «классической» силлабике XVII-XVIII вв. в финальной позиции строки стоят грамматиче-

<sup>6</sup> Именно она лежит в основе большинства русских стихотворений.

<sup>7</sup> Праформа: 01010101x, где x = {0x, 1, ø}.

<sup>8</sup> Символ “” означает сильное место праформы, а не реальное ударение.

<sup>9</sup> Для русского стихосложения крайне нехарактерны переносы ударения на «слабую» позицию праформы вместо реально наблюдающейся сильной.

<sup>10</sup> Иногда фиксируется также *предцезурное* ударение, но в большинстве текстов *русской* силлабике оно располагается вольно (хотя, например, в зрелых текстах В.К. Тредиаковского полностью урегулировано).

ски предсказуемые словоформы (обычно формы прошедшего времени глаголов<sup>11</sup>). Такие исследования, конечно, имеют наибольшую ценность для истории языка.

**3.3.** Среди тонических метров наибольший интерес для акцентологии представляют *логаэды* и *дольник*. Тактовик и акцентный стих не позволяют предсказывать позицию ударения — по причине расшатанности своей акцентной структуры<sup>12</sup>. Логаэды в этом отношении дают наибольшую четкость интерпретации реальных акцентных схем (на втором месте остаются трехсложные силлабо-тонические метры). Они представляют из себя регулярно повторяющиеся последовательности сильных и слабых мест, не укладывающиеся в схему силлабо-тонических метров и отличающиеся от др. чисто тонических метров регулярной повторяемостью акцентной схемы<sup>13</sup>. Эта регулярность имеет крайне «пуристический» характер и реальная акцентная схема логаэдической строки (= ритм) почти всегда совпадает с диктуемой праформой. Напр., логаэд 2\*1\*2\* (из Ф. Сологуба) на всем своем протяжении сохраняет единообразную акцентную схему:

(3) *И к кому́ воззвать: помоги...* и т.д.

Точность предсказания здесь, понятным образом, повышается до 100%. Но подобных текстов, в принципе, немного в русской поэтической традиции.

*Дольник* является менее предсказуемым метром. Самый распространенный тип дольника представляет собой, как бы, трехсложный силлабо-тонический метр, который может подвергаться или нет операции удаления отдельных слогов (ударных или безударных); это усечение обычно называют *синкопой*, а саму операцию *синкопированием*. Заметим, что в стиховедческой литературе часто принимается другая модель этого метра (ср. [Гаспаров 2002]), хотя круг явлений, описываемый указанными моделями, полностью совпадает.

Тем не менее, в «зрелом» русском дольнике<sup>14</sup> наблюдается аналогичное наблюдаемому в логаэдах отсутствие «сверхсхемных» ударений, поэтому можно говорить о значительной предсказуемости реальной акцентной схемы словоформы. В раннем же дольнике (начала XX века) дело обстоит не так, потому что дольник часто смешивается с др. типами тонических метров.

**3.4.** В корпусе текстов поэта первой половины XVIII в. А. Кантемира наблюдается значительное расхождение между современной и исторической акцентуацией. Тексты этого периода (в особенности, стихотворные) не используются в таких обобщающих источниках сведений о диахронии русской акцентной системы, как [Воронцова 1979]<sup>15</sup> и [Зализняк 1985]. А. Кантемир писал «стандартным» для своего времени силлабическим стихом (таким же, как, например, «отец русской силлабики» С. Полоцкий), почти все размеры которого требовали «женского» окончания стиха (клаузулы), т.е. ударения на предпоследнем

<sup>11</sup> Необходимо помнить, что силлабика XVII в. существует фактически не на русском, а на церковнославянском языке. Хотя при этом наблюдается сильная диффузность ц.-сл. акцентной системы, стремящейся к совпадению с синхронной ей русской, принимать на веру все акцентные факты не представляется целесообразным.

<sup>12</sup> Регулярных образцов тактовика в русской традиции немного. В народной эпической традиции образцы этого метра часто сопровождаются не всегда могущими быть правильно интерпретированными акцентными эффектами. У «современных» авторов (начиная с второй четверти XX в.) тактовик часто смешивается с акцентными стихом или, наоборот, присутствует спорадическими вкраплениями в «дольниковом» тексте.

<sup>13</sup> Связям логаэдов и дольников посвящена обширная литература. Классический обзор вопрос можно найти в [Гаспаров 2002].

<sup>14</sup> Примерную генеалогию этого метра можно найти в той же классической работе [Гаспаров 2002]. Однако детально эта проблема еще не нашла своей разработки.

<sup>15</sup> Притом, что в указанной монографии материалом для описания русской акцентной системы служат именно поэтические тексты.

месте строки<sup>16</sup>. В корпусе сатир Кантемира это правило спорадически нарушается, если исходить из современного состояния языка. Однако для большинства случаев, опираясь, например, на данные [Зализняк 1985], можно восстановить акцентуацию, характерную для идиолекта Кантемира<sup>17</sup>, во многом позволяющую восполнить «сожженные страницы» истории русского литературного ударения<sup>18</sup>.

Так, в корпусе сатир встречается рифменная пара *дру́же* (Voc.) ~ *у́же* (ADV.), второй член которой отражает исконное древнерусского сочетание ортотонической формы *у* (акцентная маркировка ↓ [Зализняк 1985 : 35]) и энклитики *же* (также ↓). Правило, позволяющее получить акцентуированную словоформу [Зализняк 1985 : 123], предполагает, что ударение ставится на самом левом морфе (с маркировкой ↓ или →); именно это и имеет место в данном примере, что совершенно совпадает с наблюдаемым в раннедревнерусской акцентной системе.

Другой пример должен звучать более дико. Это рифменная пара *во́йдет* ~ *подо́йдет*, для которой, казалось бы, ничто не мешает принять ударность последнего слога (в нарушение требований метра). Однако можно сделать другое предположение, не нарушающее требование метра. Дело в том, что А. Кантемир, как и многие другие русские силлабисты, прошел украинскую выучку<sup>19</sup>. Прямым подтверждением наличию некоторой доли украинизмов в идиолекте Кантемира являются частотные рифменные пары на *-e* ~ *-и-*, где *-e* совпадает с этимологическим *ь*, читавшемся в украинском варианте церковнославянского языка как */i/*<sup>20</sup>. В собственно русском языке того времени история функционирования буквы *ь* и скрывающегося за ней звука принципиально иная (ср. [Панов 2002]). Вернемся к рассматриваемой рифменной паре. В украинском языке наблюдается перемещение ударения в префиксальных глаголах на префикс в презенсе с неслоговым корнем (*во́-йде́т, подо́-йде́т*; ср. [Зализняк 1985 : 363]), что совпадает с требованием метра сохранять постоянное ударение на предпоследнем слоге строки. Наше положение не может быть признано однозначно доказуемым, но оно, кажется, может хорошо продемонстрировать возможности изучения исторической акцентологии при помощи поэтического подкорпуса НКРЯ.

### Литература:

- Воронцова В.Л. Русское литературное ударение XVIII-XX вв. Формы словоизменения. — М. «Наука», 1979
- Гаспаров М.Л. Русский стих начала XX века в комментариях. Изд. 2-е (дополненное). — М.: «Фортуна Лимитед», 2001
- Гаспаров М.Л. Очерк истории русского стиха. — М.: «Фортуна Лимитед», 2002
- Зализняк А.А. От праславянской акцентуации к русской. — М.: «Наука», 1985
- Панов М.В. История русского литературного ударения XVIII-XX вв. Изд. 2-е, стереотипное. — М.: Едиториал УРСС, 2002

<sup>16</sup> Кантемир спорадически нарушал это правило, например, в своих переводах из Анакреонта. Ср., напр., стихотворение «О весне» («Смотри, как с весны возрастом...»), где это правило нарушается уже в первой строке [Кантемир 1956].

<sup>17</sup> Вопрос о литературных идиолектах принципиален для письменной традиции, существующей в условиях отсутствия нормированного литературного языка.

<sup>18</sup> Однако необходимо отметить, что вопрос о качестве «ударной константы» для русского силлабического стиха — дискуссионный, и вряд ли сможет быть решен когда-нибудь однозначно. Обзор мнений можно найти, например, в [Гаспаров 2002 : 52].

<sup>19</sup> Об этом можно справиться, например, в биографии Кантемира в [Кантемир 1956].

<sup>20</sup> Такие рифмы нередки и у С. Полоцкого, который являлся носителем схожего, по крайней мере, по этому параметру идиолекта.