

**О. В. Лавриненко**  
Днепропетровск

**ИНТЕРНЕТ-ТЕХНОЛОГИИ В СФЕРЕ  
ФИЛОЛОГИЧЕСКОГО ОБРАЗОВАНИЯ  
(поэтический корпус русского языка как источник  
литературоведческой информации)**

*Статтю присвячено устрою та особливостям поетичного корпусу Національного корпусу російської мови. Поетичний корпус має особливу розмітку, що складається з чотирьох загальних параметрів (авторство, назва, дата написання і жанр), а також восьми власне віршознавчих параметрів (метр, строфіка, клаузула, рима, міра вірша, формула, зона римування, ікти), частина з яких може мати і подальше дроблення. Можливості російського поетичного корпусу можуть бути задіяні у викладанні теорії літератури, історії російської літератури, текстології.*

**Ключові слова:** корпусна лінгвістика, Національний корпус російської мови, поетичний корпус, Інтернет.

*Статья посвящена устройству и особенностям поэтического корпуса Национального корпуса русского языка. Поэтический корпус имеет особую разметку, состоящую из четырех общих параметров (авторство, название, дата написания и жанр), а также восьми собственно стиховедческих параметров (метр, строфика, клаузула, рифма, мера стиха, формула, зона рифмовки, икты), часть из которых может иметь и дальнейшее дробление. Возможности русского поэтического корпуса могут быть задействованы в преподавании теории литературы, истории русской литературы, текстологии.*

**Ключевые слова:** корпусная лингвистика, Национальный корпус русского языка, поэтический корпус, Интернет.

*The rapid growth in popularity of the Internet and its active use at home, in the domain of education and in professional activity require an overview of the World Wide Web and its possibilities in the context of philological education and research process. One of the most notable achievements of philological technologies on the Internet is the emergence of corpus linguistics, which was originally formed on the material of English corpora. The largest of all Russian projects of this kind is the Russian National Corpus, created in 2004. In 2006, it affixed to its structure a poetry corpus, designed to solve poetics issues. By November 2014 the poetry corpus consists of more than 72 thousand texts (about 10.3 million word usages) and covers Russian poetic texts up from the XVIII century. It has a special marking including four general parameters (authorship, title, date of writing and genre) and eight special parameters (metre, stanzaic prosody, clausula (rhythmic figure used to add finesse and finality to the end of a sentence or phrase), rhyme (repetition of similar sounds in two or more words, most often in the final syllables), verse measure, formula (an index combining scansion, numbers of verse measures and clausulae), zone of rhyming (clausula, extended to the nearest word boundary), ictus (rhythmic stress in verse, retaining its unique structure)), some of which may have further fragmentation. You can search using one of the parameters or any combination thereof, and combine them with general search possibilities of the Russian National corpus. Opportunities of the poetic corpus may be involved in teaching of subjects related to literary theory, history of Russian literature and textology.*

**Keywords:** corpus linguistics, Russian National Corpus, poetry corpus, Internet.

Интернет с 90-х гг. XX в. стремительно и прочно входит в повседневную жизнь людей. И в бытовых потребностях, и в учебе, и в профессиональной деятельности современный человек редко полностью обходится без Интернета. С одной стороны, Интернет – место механического перенесения в электронную среду процессов и операций, функционирующих и в реальном мире (например, электронные версии печатных текстов, фильмов, аудиозаписей), с другой – в Интернете постепенно формируются возможные только в электронной среде явления (онлайн-чаты, компьютерные онлайн-игры, электронный перевод и др.). Филологические же науки были обогащены полноценным новым методом исследований, сформировавшимся только благодаря возможностям Интернета. Это корпусный метод, который нацелен на решение прежде всего лингвистических задач, но может быть использован и в литературоведении в определенных случаях. Под языковым корпусом понимают «информационно-справочную систему, основанную на собрании текстов на некотором языке в электронной форме» [4]. Чтобы корпус был приспособлен для решения филологических задач, его необходимо особым образом ограничить (в плане круга составляющих его текстов) и разметить структурные фрагменты корпуса (тексты, предложения, словоупотребления), то есть наделить их набором лингвистических, текстологических и других параметров, по которым будет осуществляться поиск. В этой связи сформировалась специальная дисциплина – корпусная лингвистика, которая занимается «разработкой общих принципов построения и использования лингвистических корпусов (корпусов текстов) с применением компьютерных технологий» [2, с. 7].

Лингвистические корпуса могут быть довольно разными и противопоставляться по цели (общие, специальные), литературности (литературные, диалектные, терминологические и т. д.), жанру (публицистические, драматургические, фольклорные и т. д.), характеру разметки (морфологические, семантические, просодические и т. д.) и другим критериям (подробнее о типологии корпусов см. [2, с. 20–32]). Отдельно стоит сказать о национальных корпусах. Это специально разработанный корпус (или совокупность корпусов), посвященный какому-либо конкретному языку и нацеленный на представление «данного языка на определенном этапе (или этапах) его существования и во всем многообразии жанров, стилей, территориальных и социальных вариантов» [4]. Поэтический корпус является частью Национального корпуса русского языка (далее – НКРЯ), который был открыт в апреле 2004 г. по адресу <http://ruscorpora.ru> (подробнее о процессе подготовки проекта [см. 3]). С того времени начинается активное развитие русской корпусной лингвистики и стремительное расширение и пополнение НКРЯ. Поэтический корпус НКРЯ был открыт в декабре 2006 года. В него входят русские поэтические произведения (за исключением стихотворных драматических сочинений, которые пока еще не интегрированы в этот корпус), начиная с XVIII века.

Сейчас продолжается работа по наполнению корпуса новыми текстами. 28 октября 2014 года поэтический корпус пополнен произведениями некоторых поэтов серебряного века и поэтов 1940–1960-х гг. объемом около 700 тыс. словоупотреблений. Таким образом, к ноябрю 2014 г. поэтический корпус НКРЯ насчитывает свыше 72 тыс. текстов (это около 10,3 млн. словоупотреблений). Поиск по поэтическому корпусу НКРЯ доступен по адресу <http://ruscorpora.ru/search-poetic.html>. Главной особенностью поэтического корпуса является специальная разметка, структурирующая и описывающая различные параметры поэтического текста, которая функционирует в поэтическом корпусе наряду с основной разметкой НКРЯ. Разрабатывая концепцию специальной «поэтической» разметки, лингвисты, работавшие над поэтическим корпусом, поставили перед собой цель обеспечить «как потребности исследователей русского языка, так и потребности исследователей русской поэзии, в том числе стиховедов, заинтересованных в изучении формальных особенностей русского стиха – метрики, ритмики, каталектики, рифмы, строфики» [1, с. 72], что подразумевало «создание некоторого предварительного варианта универсального электронного метрического справочника по всей русской поэзии» [1, с. 72].

Параметры специальной разметки текстов поэтического корпуса НКРЯ (подробнее см. [1, с. 75–98]):

#### 1. Автор и сопутствующие параметры

Для текстов с неочевидным авторством предусмотрены различные решения. У народных произведений автор помечен как *обобщенный*. При полной анонимности автора ненародного произведения соответствующий параметр определен пометой *неизвестен*. При сомнительном авторстве (например, некоторые тексты из собраний сочинений М. Лермонтова, Ап. Григорьева и др.) тексту присваивается помета *dubium*. В редких случаях, когда при коллективном авторстве известна лишь часть имен, неизвестные авторы обозначаются индивидуальной обобщенной пометой. Примером такого текста в поэтическом корпусе НКРЯ является стихотворение «Гауншильд и Энгельгард...», авторство которого обозначено как «Пушкин А.С. | лицеисты» [1, с. 75].

#### 2. Название произведения

Помимо авторского названия (если оно есть), в поэтическом корпусе НКРЯ всегда указывается первая строка произведения. В некоторых случаях дополнительно определяются параметры *книга* и *цикл*.

#### 3. Дата написания произведения

Случаи с произведениями с сомнительной датировкой и с несколькими хронологически обозначенными редакциями фигурируют по техническим причинам как недифференцированно неточные датировки, то есть никак не противопоставляются по хронологическим параметрам при запросах.

#### 4. Жанр

Очевидно, что корпусная жанровая разметка должна базироваться на несколько иных посылах, чем система жанров, выстраиваемая

литературоведением и теорией речевых жанров. Для успешного использования корпуса необходима формализация и определенность в системе размечаемых жанров. Проблема жанровой размытости, неоднозначности, комбинированности в корпусе обычно игнорируется. В поэтическом корпусе НКРЯ выделяется шесть основных жанров: *стихотворение, поэма, пьеса, роман в стихах, кинофильм*. Как *кинофильм* помечаются фильмы, в основе которых стихотворные драмы (например, «Собака на сене», «Гусарская баллада» и др.) [1, с. 78]. Стихотворения и пьесы далее делятся на ряд поджанров.

## 5. Собственно стиховая разметка

### 5.1. Метр

Кроме пяти стандартных силлабо-тонических метров (*ямба, хорей, анапеста, амфибрахия, дактиля*), в корпусе размечены и особые четырех- (*неон*), пяти- (*пентон*), шестисложные метры и более (*гиперпентон*). Также размечены тонические метры (*дольник, верлибр, гекзаметр* и др.).

### 5.2. Строфика и графическая строфика

По этому параметру тексты имеют числовой показатель длины строфы, а также соответствующе ему традиционное обозначение.

### 5.3. Клаузула

Для клаузулы в поэтическом корпусе НКРЯ приняты следующие обозначения:

- м – мужская (сирень; 0 послеударных слогов);
- ж – женская (сирени; 1 послеударный слог);
- д – дактилическая (сиреневый; 2 послеударных слога);
- г – гипердактилическая (сиреневая; 3 и более послеударных слога) [1, с. 92].

### 5.4. Рифма

В поэтическом корпусе использована стандартная схема обозначения рифм с использованием букв русского алфавита, например:

- охватная | абба;
- нечетная | ахах;
- зятянутая | абааб или аббаб или аабаб, и т. п. (т. е. любая пятистрочная строфа на две рифмы);
- тройная | ааа и др. [1, с. 93].

В некоторых случаях дополнительно указывается качество рифмы (например, *монотонная рифма, тавторифма, диссонансная рифма* и др.).

### 5.5. Мера стиха

Концепция корпусной разметки поэтических текстов в НКРЯ предполагает, что тексты разных систем стихосложения имеют соответственно разные стихотворные меры строк. Таким образом, тексты силлабической системы в корпусе исчисляются в слогах, тонической – в иктах, силлабо-тонической – в стопах.

### 5.6. Формула

Это компактное объединение трех основных параметров поэтического текста (*метр, число мер, клаузула*), которое позволяет упростить поиск

нужного типа стихотворения. Формула *Я4ж*, например, будет означать четырехстопный ямб с женской клаузулой. При регулярном чередовании разнотипных строк формулы объединяются знаком +, например: *Я5м+Я3ж*, *Аф3ж+Ан2м* и т. д.

### 5.7. Зона рифмовки

Зона рифмовки представляет собой клаузулу, расширенную до ближайшего словораздела, то есть правая граница зоны рифмовки – это конец строки, а левая находится в зоне последнего словораздела, предшествующего последнему икту.

### 5.8. Икты

В текстах силлабо-тонической системы с помощью специального знака грависа ( ` ) размечаются автоматически все сильные места.

Таким образом, тексты в поэтическом корпусе имеют стандартную общую разметку НКРЯ, четыре общих параметра (авторство, название, дата написания и жанр), а также восемь собственно стиховедческих параметров, часть из которых может иметь и дальнейшее дробление. Все эти параметры, по отдельности или в различных комбинациях, могут использоваться не только как фильтр поискового запроса, но и в качестве самостоятельного запроса. Это открывает интересные возможности при изучении литературоведческих дисциплин. Прежде всего, это прекрасная иллюстративная база для стиховедческих категорий, изучаемых в рамках курсов «Введение в литературоведение» и «Теория литературы». В контексте курса «Истории русской литературы» можно довольно нетрудоемкими методами проследить эволюцию различных жанров русской поэзии, эволюцию поэтических приемов отдельных русских поэтов. В еще большей степени поэтический корпус НКРЯ облегчает сбор информации в исследовательской деятельности. Тем, чья сфера исследовательских интересов связана со стихосложением, особенно с категориями, последовательно размеченными в корпусе, использование этого ресурса может сэкономить дни монотонной работы с неразмеченными текстовыми файлами, либо даже месяцы и годы работы с рукописными картотеками – в докомпьютерную эпоху единственно возможным способом сбора подобной информации.

### Библиографические ссылки

1. Гришина Е.А. Поэтический корпус в рамках НКРЯ: общая структура и перспективы использования / Е.А. Гришина, К.М. Корчагин, В.А. Плунгян, Д.В. Сичинава // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. – СПб.: Нестор-История, 2009. – С. 71–113.
2. Захаров В.П. Корпусная лингвистика: учебник для студентов направления «Лингвистика». 2-е изд., перераб. и доп. / В.П. Захаров, С.Ю. Богданова– Иркутск: ИГЛУ, 2011. – 169 с.
3. Сичинава Д.В. Национальный корпус русского языка: очерк предыстории / Д.В. Сичинава // Национальный корпус русского языка: 2003–2005. – М.: Индрик, 2005. – С. 21–30.
4. Что такое Корпус? – Режим доступа: <http://ruscorpora.ru/corpora-intro.html>

*Надійшла до редколегії 31 жовтня 2014 р.*