

Корпус стихотворений А. С. Пушкина

Е. И. Вожик, Е. О. Казакова, Р. А. Лисюков

Аннотация

Содержание корпуса составляют стихотворные тексты А. С. Пушкина, которым в общем индексе произведений автора (Вожик Е. И., Балакин А. Ю., Беляк Г. Н., Казакова Е. О., Маслинский К. А., Димяненко А. А., Колотовкина А. С., Лисюков Р. А. Индекс произведений и писем А. С. Пушкина // Репозиторий открытых данных по русской литературе и фольклору. 2023. V2. DOI: <https://doi.org/10.31860/openlit-2023.6-B015>) присвоена категория «А» (стихотворение). Наполнение корпуса осуществлено на основании двух наиболее авторитетных академических собраний сочинений Пушкина: 1) Пушкин А. С. Полное собрание сочинений, 1837–1937: В 16 т. М.; Л.: Изд-во АН СССР, 1937–1959; 2) Пушкин А. С. Полное собрание сочинений: В 20 т. СПб.: Наука, 1999–наст. вр. Таким образом, настоящий корпус представляет наиболее актуальную, соответствующую современному развитию пушкинистики подборку текстов поэта. Стихотворные тексты объединены в zip-архиве texts.zip, где каждое стихотворение представлено в виде отдельного .txt файла (кодировка UTF-8); имя файла содержит уникальный номер (uid) стихотворения, позволяющий однозначно идентифицировать его в корпусе пушкинских сочинений. Метаданные к текстам находятся в таблице metadata.tsv. Тексты в корпусе соотносятся с таблицей метаданных согласно uid-номерам. Принципы выборки текстов в корпус, особенности его содержания и описание структуры таблицы метаданных приведены в файле readme.pdf. Список сокращений, использованных в таблице метаданных, и соответствующих им полных библиографических описаний содержится в таблице bibliography.tsv.

Формат цитирования датасета: *Вожик Е. И., Казакова Е. О., Лисюков Р. А. Корпус стихотворений А. С. Пушкина // Репозиторий открытых данных по русской литературе и фольклору. 2023. V2. DOI: <https://doi.org/10.31860/openlit-2023.8-C005>*

Общее описание данных

В настоящий корпус вошли основные редакции стихотворных текстов А. С. Пушкина, описанных в статьях «Пушкинской энциклопедии» (СПб., 2009–2024) и включенных в категорию «А» «Индекса произведений и писем А. С. Пушкина» (*Вожик Е. И., Балакин А. Ю., Беляк Г. Н., Казакова Е. О., Маслинский К. А., Димяненко А. А., Колотовкина А. С., Лисюков Р. А. Индекс произве-*

дений и писем А. С. Пушкина // Репозиторий открытых данных по русской литературе и фольклору. 2023. V2. DOI: <https://doi.org/10.31860/openlit-2023.6-B015>; подробнее о принципах присвоения категорий см. в файле *readme.pdf* «Индекса...»), — всего 913 стихотворных текстов, в числе которых сказки, одиночные стихотворные строки, стихотворные переводы фрагментов драматических сочинений (например, перевод монолога Изабеллы, героини трагедии Альфьери «Филипп» («Сомненье, страх, порочную надежду...», 1827)), переводы стихотворных повестей (например, перевод повести Вольтера «То, что нравится женщинам» («Короче дни, а ночи доле...», 1825)), стихотворные посвящения на книгах, имеющие самостоятельный характер (например, послание Д. В. Давыдову («Тебе, певцу, тебе, герою!..», 1834), а также стихотворные отрывки и наброски.

Тексты произведений приведены по т. н. «новому» академическому собранию сочинений Пушкина, работа над которым ведется в Институте русской литературы (Пушкинском Доме) Российской Академии наук с 1999 г. (*Пушкин А. С. Полное собрание сочинений: В 20 т. СПб.: Наука, 1999–наст. вр.;* далее — АПСС), и отдельным томам «старого» академического собрания сочинений 1937–1959 гг. (*Пушкин А. С. Полное собрание сочинений, 1837–1937: В 16 т. М.; Л.: Изд-во АН СССР, 1937–1959;* далее — Акад.). Новейшее издание стихотворений Пушкина не завершено: на момент публикации настоящего корпуса в свет вышли три тома АПСС, охватывающие стихотворное творчество поэта с 1813 по 1826 гг. (лицейский, петербургский, южный и михайловский периоды), — представление же о более поздних стихотворениях Пушкина продолжает в значительной степени формироваться на основе сведений, содержащихся в Акад. Таким образом, в работе над корпусом использовались следующие тома указанных изданий:

Акад.:

- отдельные стихотворения 1817–1825 гг., не вошедшие в АПСС: том 2, книга 1 (М.; Л., 1947);
- стихотворения 1826–1836 гг., сказки: том 3, книги 1 и 2 (М.; Л., 1948; 1949));
- стихотворный набросок «Полюби меня, девица...» (1829), объединенный С. М. Бонди с планом поэмы «<Русская девушка и черкес>»: том 5 (М.; Л., 1948);
- стихотворные наброски «“Женись”. — “На ком?” — “На Вере Чацкой...”» (1829–1830) и «Жуковский <> святой...» (1823): том 17 (справочный) (М.; Л., 1959);

АПСС:

- стихотворения 1813–1817 гг.: том 1 (СПб., 1999);
- стихотворения 1817–1820 (Петербург) и 1820–1824 гг. (Юг): том 2, книги 1 и 2 (СПб., 2004; 2016);

- стихотворения 1824–1826 гг. (Михайловское): том 3, книга 1 (СПб., 2019);
- «Сцена из Фауста» (1825): том 7 (СПб., 2009).

При выборе источника текста, представленного в обоих собраниях сочинений, приоритетным считалось решение составителей АПСС: в нем «заново осуществлена проверка, а иногда и пересмотр текстов <...>, учтены вновь обнаруженные источники (автографы и копии), дано обоснование и в ряде случаев уточнение датировки» (От редакции // Пушкин А. С. Полное собрание сочинений: В 20 т. СПб., 1999. Т. 1. С. 9).

В тех случаях, когда стихотворный текст по решению редколлегии АПСС был выведен из состава лирических стихотворений и отнесен к числу произведений других жанров (например, поэм) или опубликован в разделе «Другие редакции и варианты», содержание которого не позволяет установить окончательный, т. н. «канонический» текст произведения (черновые рукописи Пушкина в АПСС не транскрибированы, но осмыслены в динамике с тем, чтобы дать представление о процессе работы над текстом), — в корпус включался текст, установленный по Акад. (см., например, случай стихотворения «[Мое] беспечное незнанье...», 1823, текст которого был реконструирован И. Н. Медведевой в 1938 г.). Помещение подобных текстов в состав корпуса обусловлено стремлением учесть не только новейшее знание о творческой истории пушкинских произведений, но и историю складывания исследовательской традиции, нашедшую отражение в «Пушкинской энциклопедии».

В ряде случаев, тем не менее, в корпус включался текст, опубликованный в разделе «Другие редакции и варианты» АПСС. Речь идет о стихотворениях, имеющих самостоятельный идентификатор: см. стихотворение «Ты сердцу непонятный мрак...» (1822), опубликованное в АПСС в качестве другой редакции стихотворения «Люблю ваш сумрак неизвестный...» (1822–1824), но имеющее идентификатор, отличный от идентификатора основной редакции. Если произведение не имеет самостоятельного идентификатора и рассматривается современными исследователями в качестве варианта другого стихотворения, то в корпус помещался текст основной редакции: см., например, стихотворение «Дай, Никита, мне одеться...» (1821), публикуемое составителями АПСС в разделе других редакций и вариантов к стихотворению «Раззевавшись от обедни...» (1821).

Части стихотворных циклов «Песни западных славян», «Песни о Стеньке Разине» и «Подражания Корану», а также наброски к ненаписанным «Подражаниям...» рассматриваются как самостоятельные тексты и публикуются в корпусе под собственными заголовками. Связь между произведениями цикла устанавливается с помощью его идентификатора (колонка *cycle*). Авторские примечания к стихотворениям, входящим в состав цикла, приобщаются к их текстам. Примечания, относящиеся ко всему циклу, помещены в корпусе под общим заголовком цикла (см. файлы A550 и

A569, соответствующие циклам «Песни западных славян» и «Подражания Корану»).

Полнота данных

Тексты стихотворений (архив `texts.zip`)

В корпус входят все основные редакции стихотворных текстов Пушкина, описанных в статьях «Пушкинской энциклопедии» и отнесенных к категории «А» в «Индексе произведений и писем А. С. Пушкина», а также другие редакции ряда стихотворений (раздел «Другие редакции и варианты» АПСС) при наличии у таких редакций самостоятельного идентификатора (см. выше). В корпус не включались:

- тексты, отнесенные составителями академических собраний сочинений к разделу «Dubia» (за исключением тех, которым, в силу их широкой известности и/или изученности, посвящены отдельные статьи «Пушкинской энциклопедии» (например, эпиграмма «Мы добрых граждан позабавим...», конец 1810-х – начало 1820-х));
- тексты стихотворений, входящие в состав писем и неотделимые от их прозаической части (например, стихотворение «Тебе, о Нестор Арзамаса...», являющееся частью послания В. Л. Пушкину от 28 (?) декабря 1816 г.);
- тексты, написанные Пушкиным в соавторстве;
- тексты поэм (в т. ч. лицейских, таких как «Монах», 1813; «Бова», 1814), а также романа в стихах «Евгений Онегин».

Тексты стихотворений приведены без заголовков, данные о которых содержатся в таблице метаданных к корпусу. Неотъемлемой частью стихотворного текста признаются *и перитекстовые элементы*: внутренние подзаголовки, эпиграфы, авторские примечания и т. д., размещаемые в одном файле со стихотворением, к которому относятся, и выделенные специальными тегами (см. ниже). Иноязычные тексты, отдельные строки и слова оставлены в корпусе без перевода.

Таблица метаданных `metadata.tsv`

Таблица метаданных к стихотворениям корпуса (`metadata.tsv`) включает в себя сведения, основной набор которых воспроизводится по «Пушкинской энциклопедии».

Формат и структура данных

Тексты стихотворений (архив texts.zip)

Тексты стихотворений представлены в виде отдельных .txt файлов (кодировка UTF-8), объединенных в zip-архив texts.zip.

Названия файлов содержат уникальный идентификатор произведения (uid) (например, A20), а также, при необходимости, указание на номер его редакции, отделенный от основного идентификатора нижним подчеркиванием (например, A19 – ранняя редакция стихотворения «Безверие», A19_2 – его поздняя редакция). Нижнее подчеркивание используется также для наименования файла A509_2, содержащего стихотворение «Тебя зову на томной лире...» (1831), которое объединено составителями «Пушкинской энциклопедии» с наброском «Одна стихи ему читала...» (1831), однако в Акад. представлено как самостоятельное произведение.

Частям стихотворных циклов присвоены uid'ы, включающие идентификатор цикла и порядковый номер текста в цикле (порядок алфавитный), отделенный от основного идентификатора точкой (например, A550.1, A550.2); соответствующие названия имеют и файлы, содержащие текст стихотворений, которые образуют цикл.

Для выделения из состава текстовых файлов перитечстовых элементов в корпусе используется семантическая разметка по стандарту TEI. Размечены элементы, которые маркируют (1) обстоятельства создания стихотворения, а также (2) особенности его формального устройства или функциональное значение отдельных его составляющих. Таким образом, в корпусе использованы следующие теги:

(1)

- date – время написания стихотворения;
- note – автocomментарии к стихотворению; тег также используется для разметки стихотворения «Кто на снегах возрастил Феокритовы нежные розы?..» (1829), предваряемого указанием на обстоятельства адресации текста «(При посыпке бронзового Сфинкса)»;
- source – источник стихотворения (для переводных текстов; см. «Gonzago» в стихотворении «С португальского» (1825–1827));
- place – место написания стихотворения;

(2)

- epigraph – эпиграф, предваряющий стихотворение;
- head – внутренние подзаголовки стихотворения;
- speaker – указание на персонажа, которому принадлежит реплика (для стихотворений, написанных в драматической форме);

- **stage** — указание на место, где происходят события стихотворения, или на действия, совершаемые его персонажами (см. «(Исчезает.)» в «Сцене из “Фауста”» (1825)).

Прозаические фрагменты произведений, не отделимые от их стихотворной части, в корпусе не размечены (см., например, «Кольна (Подражание Оссиану)» (1814)).

Служебные символы (прямые скобки [], куда помещены слова, зачеркнутые Пушкиным, и угловые < >, заключающие редакторские дополнения) воспроизводятся в текстах стихотворений строго по Акад. и АПСС.

Таблица метаданных `metadata.tsv`

Метаданные к стихотворным текстам Пушкина представлены в виде таблицы в формате tab-separated values (значения, разделенные символом табуляции) в файле `metadata.tsv`. Единицей данных, соответствующей отдельной строке в таблице, является стихотворение или одна из его редакций (при наличии).

Графы таблицы оставлены пустыми для стихотворений, сведения о которых отсутствуют в «Пушкинской энциклопедии» отсутствуют.

Столбцы:

- **volume** — номер тома «Пушкинской энциклопедии», в котором опубликована статья о стихотворении, или литера А (для произведений, не имеющих в энциклопедии особой статьи);
- **uid** — уникальный идентификатор стихотворения; принципы присвоения идентификаторов см. в файле `readme.pdf` «Индекса произведений и писем А. С. Пушкина»;
- **cycle** — уникальный идентификатор, общий для стихотворений, входящих в тот или иной цикл; столбец оставлен пустым для произведений, не образующих цикла;
- **title** — основное название стихотворения, принятое в качестве такового в «Пушкинской энциклопедии»; если у стихотворения есть поздняя редакция, ее название приводится по АПСС;
- **version** — сведения об имеющихся редакциях стихотворений (при наличии);
- **incipit** — первая строка стихотворения; данные об инципитах отсутствуют для циклов стихотворений;
- **link** — название энциклопедической статьи, на которую ссылается статья о стихотворении; в ряде случаев в столбце фигурируют названия т. н. «редакторских» циклов, отсутствующие в колонке `title`, но

имеющиеся в таблице «Индекса произведений и писем А. С. Пушкина»;

- `link_uid` — уникальный идентификатор произведения, к статье о котором ведет отсылка; в ряде случаев такие произведения относятся к категориям, отличным от А (см. стихотворную строку «Он далече, далече плывет в печальном тумане...», 1821, упоминание которой содержится в статье о замысле «»);
- `terminus_post_quem` — нижняя граница датировки стихотворения с указанием года;
- `terminus_ante_quem` — верхняя граница датировки стихотворения с указанием года;
- `dated` — развернутая датировка стихотворения, а также основания для нее (приводится по «Пушкинской энциклопедии»; если сведения о датировке здесь отсутствуют, столбец оставлен пустым);
- `autographs` — сведения об автографах стихотворения, сопровождающиеся их краткой характеристикой, а также указанием места хранения и шифра (приводятся по «Пушкинской энциклопедии»; если сведения об автографах здесь отсутствуют, столбец оставлен пустым); архивные шифры пушкинских автографов, хранящихся в Рукописном отделе ИРЛИ РАН (ф. 244, оп. 1), даются сокращенно: ПД, с указанием номера единицы хранения и, в случае необходимости, листа рукописи; если автограф стихотворения неизвестен, в графе значится *Неизвестен*.
- `lifetime_publication` — сведения о том, было ли стихотворение опубликовано при жизни Пушкина; допустимые значения: *да/нет*;
- `first_publication` — библиографические данные о первой публикации стихотворения (приводятся по «Пушкинской энциклопедии»; если сведения о первой публикации здесь отсутствуют, столбец оставлен пустым);
- `bibliography` — перечень избранной литературы о стихотворении (приводится по «Пушкинской энциклопедии»; если этот перечень здесь отсутствуют, столбец оставлен пустым);
- `citation` — библиографическая ссылка на текст стихотворения в собрании сочинений Пушкина; в ссылках используются условные сокращения: Акад. и АПСС — далее следует указание на номер тома, книги (при наличии) и страницы;
- `filename` — название файла, содержащего текст стихотворения; файл, соотносящийся с циклом «Песни о Стеньке Разине», в архиве `texts.zip` отсутствует — вместо него в корпусе представлены отдельные стихотворения, входящие в этот цикл;

- `comments` — мотивированное обоснование выбора источника, по которому устанавливается текст стихотворения (для стихотворных произведений, установление текстов которых представляет трудность);
- `n_lines` — количество строк в стихотворении (без учета перитекстовых элементов стихотворения, выделенных тегами);
- `n_words` — количество слов в стихотворении (без учета перитекстовых элементов стихотворения, выделенных тегами).

Таблица метаданных `bibliography.tsv`

Сокращения, принятые в таблице метаданных, соответствуют тем, которые используются составителями «Пушкинской энциклопедии». Перечень условных сокращений и их расшифровку см. в таблице `bibliography.tsv`.

Столбцы:

- `abbreviation` — сокращенное наименование источника;
- `bibliographic_description` — расшифровка сокращения, приведенного в столбце `abbreviation`.

Благодарности

Составители корпуса выражают благодарность рецензенту Антонине Мартыненко, высказавшей ценные замечания, которые позволили усовершенствовать способ представления данных в корпусе.

Обновления во второй версии корпуса (V2)

- Исправлен ряд ошибок.
- Дополнены сведения о стихотворениях, описанных в пятом томе «Пушкинской энциклопедии».
- В соответствии с обновленным словарником пятого тома актуализировано количество строк в таблице `metadata.tsv`.
- Дополнены сведения о редакциях стихотворений.
- Дополнены сведения о наличии или отсутствии прижизненной публикации стихотворения.
- Из корпуса удалено стихотворение «Что-то грезит Баратынский...» (A773), которому редакторы «Пушкинской энциклопедии» приняли решение не посвящать отдельной статьи.
- Идентификатор A773 присвоен стихотворению «Уединение», ранее имевшему тот же идентификатор, что и у стихотворения «Деревня». Таким образом, файл с текстом «Уединения» был переименован из A143_2 в A773.