

**12 февраля** на очередном заседании семинара с докладом «Как построить корпус текстов, пригодный для нужд авторской лексикографии? (Несколько практических вопросов на примере словаря языка Батюшкова)» выступил д. ф. н. **И. А. Пильщиков** (*Институт мировой культуры МГУ и Таллинский университет*).

**1.** Изучение авторского идиолекта и его эволюции осуществимо только на основе текстологически корректного и хронологически дифференцированного корпуса текстов, отражающего историю каждого авторского произведения путем учета существовавшей в исторической реальности хронологической последовательности текстов — письменных или печатных фиксаций разных редакций и вариантов одного произведения. Для нужд авторской лексикографии тексты, репрезентирующие произведение, должны быть объединены в кластеры по технологии, близкой к технологии корпусов параллельных текстов. В докладе на примере текстов Батюшкова обсуждались проблемы, встающие при создании таких «дифференциальных корпусов» (проект осуществляется при поддержке РГНФ, грант № 11-04-00120).

**2.** В печатных изданиях варианты и редакции произведения обычно воспроизводятся в виде сводки разночтений к тексту, выбранному в качестве основного. Это крайне затрудняет обращение к контексту: при такой подаче материала вообще невозможно дать аутентичную цитату из ранней редакции.

Исключения лишь подтверждают правило и не решают проблемы в целом. Так, в академических собраниях сочинений Пушкина стихотворения лицейского периода печатаются в двух редакциях: ранних, лицейских, и позднейших, созданных при переработке ранних произведений для сборника 1826 года. Аналогичная ситуация известна и другим национальным традициям. Например, стихи Гёте штюрмерского периода также печатаются в двух редакциях — ранней и поздней.

Из всех поэтов пушкинской эпохи поэтом, в наибольшей степени перерабатывавшим свои стихи от одной редакции к другой, был Баратынский. Его тексты в предреволюционном академическом издании, ориентированном на ранние редакции стихотворений, радикально отличаются от изданий советского времени, ориентированных на одну из поздних редакций. Выходящее с 2002 года Полное собрание стихотворений и писем ставит перед собой целью представить каждое произведение в 2-3 редакциях.

Но почему наличие в издании двух разных, значительно отличающихся друг от друга редакций не решает проблемы полного описания авторской лексики и фразеологии? Во-первых, возникает проблема дублей: одинаковые слова и целые строки в произведениях, напечатанных дважды, дважды же попадают в конкорданс и в словарь языка писателя. Между тем фрагменты других текстов подобной «чести» не удостаиваются. Во-вторых, какие-то варианты (и даже целые связные редакции) всегда остаются «за бортом»: в лучшем случае они будут представлены в разделе «других редакций», материал которого учитывать крайне затруднительно.

Возьмем для примера стихотворение Батюшкова «Мечта», которое начиная с юбилейного издания, подготовленного Л. Н. Майковым (1887), печатается в двух редакциях — поздней и ранней. Многие стихи в них совпадают, поэтому, например, в конкордансе к корпусу поэтических текстов Батюшкова, составленном Дж. Т. Шоу (1975), такие строки фигурируют дважды. С другой стороны, первой и последней редакцией «Мечты» не исчерпываются все варианты текста, поскольку в этом произведении можно выделить четыре редакции, каждая из которых представлена, в свою очередь, 2-3 публикациями, не вполне совпадающими по тексту:

Мечта [«О сладостна мечта, дочь ночи молчаливой...»] (1804–1810).

1. Первоначальная редакция.

1) Текст «Любителя словесности» 1806 г. (1804–1806).

2) Текст «Вестника Европы» 1810 г. (начало 1810).

2. Промежуточная редакция (лето 1810).

1) Автограф.

2) Отрывок в «Картинах Финляндии» (1810).

Мечта [«Подруга нежных муз, посланница небес...»] (1811–1817).

3. Новая редакция (1811–1815).

1) Текст «Собрания Русских стихотворений» (1811).

2) Текст рукописного сборника 1812 г.

3) Текст Блудовской тетради (1815).

4. Окончательная редакция.

1) Отрывок в «Опытах в стихах и прозе», ч. I (1816).

2) Текст «Опытов в стихах и прозе», ч. II (1816 или 1817).

Текст промежуточной редакции 1810 г. приведен только в одном издании — «Сочинениях» Батюшкова, вышедших в издательстве «Academia» (1934), а текст новой редакции 1811–1815 гг. приведен в изданиях 1887 и 1934 гг. только в форме разночтений к окончательной редакции.

Каждая редакция стихотворения, помимо элементарных лексических разночтений данной редакции с окончательной, может содержать строки, в которых автором произведены сокращения или перестановки слов по отношению к предыдущей редакции или к предыдущему варианту данной редакции, *не* попавшие в окончательную редакцию (в словаре такие строки не дают новых вокабул, но дают дополнительную цитату на слова, уже попавшие в словник на основании текстов других редакций). Далее, при замене одного слова соседнее слово может изменить свое значение в результате трансформации контекста (в словаре такие строки не дают новых вокабул, но дают дополнительные значения слов, уже попавших в словник на основании текстов других редакций). Кроме того, в каждой редакции стихотворения могут найтись лексические разночтения с предыдущей редакцией или с предыдущим вариантом той же редакции, сохраненные в окончательной редакции (в словаре такие строки не дают новых вокабул, но сдвигают датировку употребления слов и словоформ, уже попавших в словник на основании текста окончательной

редакции). Для корректной датировки тех или иных словоупотреблений необходимо также учитывать промежуточные варианты, отсутствовавшие в первой редакции, исключенные из позднейших редакций, но восстановленные в окончательной редакции.

3. Существующие лингвистические корпуса и электронные библиотеки для решения подобных задач совершенно непригодны, и это неудивительно, поскольку создавались они совсем для других целей. Все эти цифровые собрания характеризуются, в первую очередь, неполнотой представления: ни в одном из них не представлен полный корпус авторских текстов и вариантов текста.

Так, в Поэтическом подкорпусе Национального корпуса русского языка (НКРЯ) сочинения Батюшкова даны по тексту Полного собрания стихотворений, вышедшему во 2-м издании Большой серии «Библиотеки поэта» (1964). Между тем в этом издании вообще нет раздела вариантов: некоторые разночтения приведены в редакторских комментариях (не вошедших в НКРЯ); исключение, как и в майковском собрании, сделано только для стихотворения «Мечта».

Серьезной проблемой остается проблема датировки текстов, унаследованная от печатных источников — научных и научно-массовых изданий, в которых произведения обычно печатаются в поздней редакции, а датируются годом ранней (в лучшем случае в изданиях указываются две даты — дата создания первоначального текста и окончательной редакции). Так подаются тексты в той же «Библиотеке поэта». В результате представленная хронология отличается большой степенью условности.

Например, сделав в Поэтическом подкорпусе НКРЯ запрос по словосочетанию «*прах красноречивый*», мы обнаружим, что применительно к Риму это словосочетание применил Батюшков в элегии «Умиравший Тасс» (1817), а до него Гнедич в послании «К К. Н. Батюшкову» (1807). Можно заключить, что Батюшков цитирует выражение Гнедича, но это не так. В НКРЯ произведения Гнедича представлены и датированы в соответствии со 2-м изданием Большой серии «Библиотеки поэта» (1956), где тексты даны по итоговому собранию стихотворений Гнедича 1832 г., а датированы годом ранней редакции. Однако в ранней редакции послания Гнедича к Батюшкову (1807), опубликованной в «Вестнике Европы» в 1810 г., строки про «*Рима прах красноречивый*» еще не было. Значит, это не Батюшков цитирует Гнедича, а, наоборот, Гнедич — Батюшкова.

В Фундаментальной электронной библиотеке «Русская литература и фольклор» (ФЭБ) несколько классических авторов (Пушкин, Ломоносов, Лермонтов, Грибоедов, Батюшков, Тютчев, Маяковский, Есенин) представлены с достаточной (но тоже, увы, не исчерпывающей) полнотой. Так, в разделе ФЭБ, посвященном Батюшкову, представлены издания «Библиотеки поэта» и «Academia», майковское собрание и академическое издание «Опытов в стихах и прозе», вышедшее в серии «Литературные памятники» (1977). В соответствии с заявленной установкой на аутентичность воспроизведения текстов, эти издания

в электронной библиотеке репродуцируют свои печатные аналоги, от которых наследуют все перечисленные выше проблемы. Подача вариантов в форме разночтений не дает возможности обращаться ко всем контекстам; при этом датировка каждого текста и отдельного его фрагмента не дана пользователю непосредственно, а требует дополнительной работы с текстологическим комментарием и другими сопроводительными материалами. Помимо прочего, не все стихотворения и стихотворные фрагменты включены в этих изданиях в корпус стихотворных текстов: некоторые могут присутствовать в качестве стихотворной вставки в прозаических произведениях или в письмах. Наконец, имеются стихотворные тексты Батюшкова, не учтенные ни в одном собрании сочинений: таков, например, батюшковский перевод начальных стихов 3-й октавы IV песни «Освобожденного Иерусалима» Тассо, процитированные Гнедичем в письме к Батюшкову от 16 октября 1810 г. («Сзывает жителей подземных страны...»).

4. Чтобы построить «дифкорпус», необходимо идентифицировать произведения, составляющие авторский корпус текстов, и тексты, представляющие разные редакции одного произведения. Для этого каждому тексту корпуса присваивается индивидуальный шифр. Каждому стихотворению присваивается номер по общему хронологическому списку стихотворений (например: 3. Послание к стихам моим). Каждой редакции присваивается номер второго порядка (например, в послании «Послание к стихам моим»: 3.1. Печатная редакция; 3.2. Рукописная редакция). Вариантам одной и той же редакции стихотворения присваиваются литеры *a*, *b*, *c* и т. д., выставляемые после номера редакции. Например, стихотворение «Мечта» (№ 4) имеет несколько редакций, из которых первоначальная редакция (4.1) имеет два варианта, представленных текстом «Любителя Словесности» 1806 г. (4.1a) и текстом «Вестника Европы» 1810 г. (4.1b).

Вот как выглядит начало списка стихотворений Батюшкова:

1. «Для нас всё хорошо вдали...» (1802 или 1803).
2. Бог (1804).
3. Послание к стихам моим (1804).
  - 3.1. Текст «Новостей русской литературы» 1805 г.
  - 3.2. Рукописная редакция.
4. Мечта [«О сладостна мечта, дочь ночи молчаливой...»] (1804–1810).
  - 4.1. Первоначальная редакция.
    - 4.1a. Текст «Любителя словесности» 1806 г. (1804–1806).
    - 4.1b. Текст «Вестника Европы» 1810 г. (начало 1810).
  - 4.2. Промежуточная редакция (лето 1810).
    - 4.2a. Автограф.
    - 4.2b. Отрывок в «Картинах Финляндии» (1810).
4. Мечта [«Подруга нежных муз, посланница небес...»] (1811–1817).
  - 4.3. Новая редакция (1811–1815).

- 4.3a. Текст «Собрания Русских стихотворений» (1811).
- 4.3b. Текст рукописного сборника 1812 г.
- 4.3c. Текст Блудовской тетради (1815).
- 4.4. Окончательная редакция.
  - 4.4a. Отрывок в «Опытах в стихах и прозе», ч. I (1816).
  - 4.4b. Текст «Опытов в стихах и прозе», ч. II (1816 или 1817).
- 5. Элегия (Перевод из Парни) [«Как счастье медленно приходит...»] (1804 или 1805).
  - 5.1. Текст «Северного вестника» 1805 г.
  - 5.2. Рукописная редакция.
- 6. Послание к Хлое (Подражание).

Кроме того, необходимо произвести идентификацию фрагментов произведений, позволяющую формально сопоставлять разные редакции одного произведения. Для этой цели предлагается ввести в текстах неопределенных редакций двойную нумерацию строк: помимо номера строки по тексту данной редакции, для этой же строки дается в скобках ее номер в соответствии с окончательной редакцией. Строки, отсутствующие в окончательной редакции, нумеруются по последнему совпадающему стиху, например, 5, 5a, 5b, 5c. Совпадающие или соответствующие друг другу стихи разных редакций, отсутствующие в окончательной редакции, маркируются одинаковыми литерами. Номер строки становится в ее шифре номером третьего порядка.

Вот как, например, выглядит нумерация первоначальной редакции стихотворения «Мечта»:

О сладостна мечта, дщерь ночи молчаливой!	<b>1 (1)</b>
Сойди ко мнѣ съ небесъ въ туманныхъ облакахъ	<b>2 (15)</b>
Иль въ миломъ образѣ супруги боязливой	<b>3 (15a)</b>
Съ слезой блестящею во пламенныхъ очахъ!	<b>4 (15b)</b>
Ты въ душу нѣжную поэта	<b>5 (15c)</b>
Лучемъ проникнувь свѣта	<b>6 (15d)</b>
Горишь какъ огонь зари и красишь пѣснь его.	<b>7 (15e)</b>
Любимца чистыхъ сестръ, любимца твоего	<b>8 (21)</b>
И горестъ сладостна бываетъ:	<b>9 (22)</b>
Онъ въ горести мечтаетъ.	<b>10 (23)</b>

Соответственно, полный номер каждой строки будет выглядеть так: 4.1a.1(1), 4.1a.2(15), 4.1a.3(15a), 4.1a.4(15b) и т. д., при этом стих 4.1a.1 соответствует стиху 4.4b.1 (но не обязательно совпадает с ним!), стих 4.1a.2 соответствует стиху 4.4b.15 и т. д.

Остается открытым вопрос, не нужно ли в особо сложных случаях вводить тройную нумерацию. Иногда это полезно и удобно. Скажем, у Батюшкова в тексте 4.1b (публикация «Мечты» в «Вестнике Европы» 1810 г.) по сравнению с текстом 4.1a пропущены строки 28 и 67–70 и добавлен один стих после стиха

72. В результате стихи 4.1b.59–4.1b.69 будут иметь такую сопоставительную нумерацию:

Увы, но съ юностью исчезнуть и мечтанья!	<b>59/60 (189)</b>
Исчезнуть Грацій лобызанья,	<b>60/61 (190)</b>
Какъ свѣтлые лучи на темныхъ облакахъ.	<b>61/62 (190a)</b>
Веселья на крилахъ	<b>62/63 (190b)</b>
Дни юности стремятся:	<b>63/64 (190c)</b>
Не долго на цвѣтахъ	<b>64/65 (190d)</b>
Любовью упиваться;	<b>65/66 (190e)</b>
Мечты крылатыхъ сновъ	<b>66/71 (191)</b>
Не сыплють тамъ своихъ цвѣтовъ,	<b>67/72 (192)</b>
Гдѣ юность увядаетъ,	<b>68/72a (192a)</b>
Гдѣ тусклый опытность свѣтильникъ зажигаетъ.	<b>69/73 (193)</b>

Иначе говоря,

стих 4.1b.60 соответствует стиху 4.1a.61 и стиху 4.4b.190;

стих 4.1b.61 соответствует стиху 4.1a.62 и не имеет соответствий в тексте 4.4b;

стих 4.1b.67 соответствует стиху 4.1a.72 и стиху 4.4b.192;

стих 4.1b.68 не имеет соответствий в текстах 4.1a и 4.4b;

и т. д.

Третий номер, будучи эвристически удобным, не является необходимым: варианты первоначальной редакции 4.1a и 4.1b могут быть сопоставлены друг с другом через окончательную редакцию 4.4b.

5. Отдельную группу проблем представляет словарная интерпретация хронологически дифференцированного корпуса авторских текстов. Как, например, подавать в словаре лексически идентичные варианты строки? Если полностью — загромождается словарная статья. Если в виде разночтений — трудно или невозможно процитировать аутентичный контекст. Если же не давать вовсе — постоянно будет возникать двусмысленность: то ли текст не имеет вариантов, то ли другие редакции текста дают в этом месте чтение, идентичное представленному.

Другой вопрос: что считать идентичными вариантами. Только ли варианты, дающие новые лексемы и словоформы? Но ведь существуют варианты, отличающиеся друг от друга порядком слов (о чем уже говорилось ранее), а это разные контексты, которые словарь должен учитывать.

Важнейшая проблема — как на основе «дифкорпуса» строить частотный словарь. Действительно, в «дифкорпусе» релятивизируются оба типа частотности — и абсолютная (сколько раз «встречается» в корпусе строка, идентичная в двух разных редакциях текста?), и относительная (от какого общего числа вхождений мы ее считаем?). По-видимому, этот вопрос более-менее удовлетворительно решается только в электронной версии словаря:

пользователь может задавать себе разные подкорпусы (только окончательные редакции; только первоначальные редакции; подкорпус авторских текстов по состоянию на такой-то год; подкорпус окончательных редакций и разночтений, содержащих новые формы и лексемы; etc.) и затем вычислять абсолютную и относительную частотность тех или иных форм в конкретном подкорпусе.