

## РАЗДЕЛ III. ФИЛОЛОГИЯ И ЕЕ ПРОБЛЕМНОЕ ПОЛЕ, ПЕРСПЕКТИВЫ РАЗВИТИЯ

*ББК 81я21*

*УДК 82.0*

*Я. М. Марголис<sup>1</sup>*

*Групп университет, Германия*

*[jakob\\_margolis@gmx.de](mailto:jakob_margolis@gmx.de)*

### КОЛИЧЕСТВЕННЫЙ АНАЛИЗ СЛОВАРЕЙ И ТРАНЗИТА РУССКОЯЗЫЧНОЙ ПОЭЗИИ

*«...необходима статистика; необходим словарь слов: Баратынского, Пушкина, Тютчева. В руках чуткого критика словари – ключи к тайнам духа поэтов». Андрей Белый «Поэзия слова»*

Цель исследования – количественный анализ словарей и транзита (перехода) русского поэтического языка во времени от поэта к поэту и в пространстве от страны к стране.

На первом этапе исследования был проведён статистический анализ для оценки количества словоупотреблений, словоформ и лексем в словарях русских поэтов XVIII – XX веков. В результате анализа «Словаря языка русской поэзии XX века» Института русского языка РАН и словарей поэтов XIX века найдены сильные возрастающие зависимости величины словарного запаса от количества словоупотреблений в стихах поэтов. Найденные зависимости применены для оценки словарных запасов поэтов, не вошедших в словари. Показано, что количество словоупотреблений и словарные запасы поэтов растут во времени от XVIII до первой половины XX века. Границы поэтического языка расширяются в указанный период и определены пороговые значения их транзита.

Для поэтов XVIII – XIX веков на основе новой версии поэтического корпуса русского языка найдены слабые возрастающие зависимости количества словоупотреблений от возраста поэтов.

---

<sup>1</sup> Марголис Яков Моисеевич, математик, статистик, Dr.-Ing., литератор, фрилансер, Штутгарт, Германия.

Показано как можно проводить сравнения словарных запасов поэтов количественно (статистически) и качественно (филологически) для оценки транзита поэтического языка во времени.

На втором этапе исследования независимыми жюри были отобраны четыре группы русскоязычных поэтов XXI века из разных стран, собраны их стихи за последние годы и намечены пути для определения транзита их поэтического языка на фоне языка русских поэтов XVIII – XX веков. В дальнейшем по результатам этих исследований будет проведён сравнительный анализ с языком русскоязычных поэтов XXI века, живущих в России, Украине, Беларуси, Латвии, Грузии, Казахстане и Узбекистане. Исследование выполнено в рамках подготовки ко второму этапу филологического проекта «Русскоязычная поэзия в транзите», университет Трира, Германия.

**Ключевые слова:** квантитативный анализ, словарь, транзит, поэтический язык, лингвостатистика, словарный запас, словоупотребления, словоформа, лексема, современные русские поэты XXI века

## 1. Словари новой версии поэтического корпуса русского языка

Национальный корпус русского языка (НКРЯ) и его новая версия поэтического корпуса (ПКРЯ) позволяют делать количественные анализы и поиски в словарях поэтов [9]. Посмотрим, например, результаты поиска в новой версии поэтического корпуса русского языка в словарях Марины Цветаевой (МЦ) и Анны Ахматовой (АА). Объём всего ПКРЯ около 90 тысяч документов (не всегда совпадает с количеством стихов поэта), состоит из 12,4 миллиона слов и включает стихи 951 русского поэта. Пример: поиск ведётся по поэтическому корпусу стихов Марины Цветаевой объёмом 1468 док., 168711 словоупотреблений. В запросе задано: найти в поэтическом корпусе словоупотребления «любовь и ненависть». Найдено: 162 документа, 248 вхождений. В ПКРЯ есть возможность посмотреть, в каком контексте и в каких стихах у МЦ встречаются заданные лексемы «любовь и ненависть», например:

*Как гроза меняет местность –  
Так **любовь** меняет лик.  
Две змеи его снедают,  
Две – высасывают кровь:  
Злая **ненависть** к счастливцу,  
К дивной девушке **любовь**.*

М. И. Цветаева, Этери (1940)

Могут быть так же подсчитаны словарные запасы и частоты словоформ и лексем, найденные на каждой странице поэтического корпуса, см. таблицу 1.

Табл. 1: Пример одной страницы словарных запасов Марины Цветаевой в словоформах и в лексемах поэтического корпуса.

| Словарный запас СЗ в словоформах. Частота |            |    |
|---|------------|----|
| 1   | любовь     | 15 |
| 2   | любви      | 7  |
| 3   | любовью    | 4  |
| 4   | ненависть  | 3  |
| 5   | ненавистью | 1  |
| Словарный запас СЗ в лексемах             |            |    |
| 1   | любовь     | 26 |
| 2   | ненависть  | 4  |

Естественно, что словарный запас СЗ в словоформах (5 словоформ) может быть в несколько раз больше словарного запаса СЗ в лексемах (2 лексемы). Сумма частот словоформ равны частоте лексем «любовь»  $15+7+4=26$  и «ненависть»  $3+1=4$ .

Приведём ещё один интересный пример: частота лексемы «любовь» в словарях М. Цветаевой и А. Ахматовой, В. Жуковского и А. Пушкина.

Для поэтического корпуса МЦ найдено: 1468 док., 168711 словоупотреблений, а поиск лексемы «любовь» в поэтическом корпусе МЦ даёт 152 док. и 231 словоформу.

Для поэтического корпуса АА найдено: 945 док., 56403 словоупотреблений, а поиск лексемы «любовь» в поэтическом корпусе АА даёт 76 док. и 80 словоформ.

В абсолютных числах частота лексемы «любовь» у М. Цветаевой (231) больше, чем у А. Ахматовой (80), потому что у МЦ больше общее количество словоупотреблений (168711), чем у АА (56403).

Однако важна не абсолютная, а относительная частота (ОЧ) на 10 тысяч словоупотреблений поэтического текста.

У МЦ:  $ОЧ\_МЦ = 231/(168711/10000) = 13,69$  раз в среднем встречается лексема «любовь» в 10 тысячах словоупотреблениях МЦ.

У АА:  $ОЧ\_АА = 80/(56403/10000) = 14,18$  раз в среднем встречается лексема «любовь» в 10 тысячах словоупотреблениях АА – это немного выше, чем у МЦ.

Для сравнения посмотрим значимость словоформы «любовь» в поэтических словарях мужчин-поэтов: В. Жуковского (ВЖ) и А. Пушкина (АП) по данным НКРЯ.

Для поэтического корпуса Жуковского найдено: 701 док., 347067 словоупотреблений, включая все переводы ВЖ (один только перевод «Одиссеи» Гомера насчитывает около 100 тысяч словоупотреблений). Поиск лексемы «любовь» в поэтическом корпусе Жуковского даёт 191 документ и 458 словоформ.

Для поэтического корпуса Пушкина найдено: 927 док., 195572 словоупотреблений, а поиск лексемы «любовь» в поэтическом корпусе АП даёт 205 документов и 493 словоформы.

В абсолютных числах частота лексемы «любовь» у Жуковского (458) не очень сильно отличается от частоты Пушкина (493). Однако, как мы видели выше, важна не абсолютная, а относительная частота (ОЧ) на одинаковой длине текста, например, на 10 тысяч слов текста.

У Жуковского:  $ОЧ\_ВЖ = 458/(347067/10000) = 13,20$  раз в среднем встречается лексема «любовь» в 10 тысячах словоупотреблениях, что сопоставимо с относительной частотой лексемы «любовь»  $ОЧ\_МЦ$  (13,69) и  $ОЧ\_АА$  (14,18).

У Пушкина:  $ОЧ\_АП = 493/(195572/10000) = 25,22$  раз встречается лексема «любовь» в 10 тысячах словоупотреблениях – это почти в два раза больше, чем у Цветаевой, Ахматовой, Жуковского.

Для сравнения лексема «любовь» у М. Лермонтова:  $ОЧ\_МЛ = 322/(132819/10000) = 24,25$  – на уровне А. Пушкина.

## **2. Анализ словарей русских поэтов конца XIX – первой половины XX века**

Институт русского языка (ИРЯ) Российской академии наук (РАН) публикует с 2001 года «Словарь языка русской поэзии XX века», в 2019 году вышел 8-й том словаря, составитель Л. Шестакова [18]. Источниками словаря послужили издания 10 русских поэтов конца XIX – первой половины XX века, содержащие достаточно представительный корпус их стихотворных произведений (включая поэмы и драматические произведения в стихах).

Иннокентий Анненский. Стихотворения и трагедии. 1990.

Анна Ахматова. Сочинения: в 2 томах. 1995.

**Я. М. Марголис**

Александр Блок. Собрание сочинений: в 8 томах. 1960.

Сергей Есенин. Собрание сочинений: в 5 томах. 1961.

Михаил Кузмин. Избранные произведения. 1990.

Осип Мандельштам. Сочинения: в 2 томах. 1990.

Владимир Маяковский. Стихи, поэмы. 1969.

Борис Пастернак. Сочинения: в 5 томах. 1989.

Велимир Хлебников. Творения. 1986.

Марина Цветаева. Собрание сочинений: в 7 томах. 1994.

В словаре ИРЯ представлена таблица 2: авторы, количество словоупотреблений (КСУ), количество различных словоформ (сф) или словарный запас различных словоформ (СЗ сф) в стихотворных произведениях каждого из указанных выше поэтов. В таблицу 2 был добавлен возраст каждого поэта, который понадобится нам в дальнейшем для анализа зависимости количества, написанного поэтом, от количества прожитых поэтом лет.

В таблице 2 для сравнения с поэтами XX века приведены данные поэтов XIX века – Александра Пушкина (АП) и Михаила Лермонтова (МЛ), взятые из «Национального корпуса русского языка» [9] и словарей языка А. Пушкина и М. Лермонтова [3; 4].

При этом словарные запасы поэтов XX века представлены ИРЯ [18] в словоформах, а словарные запасы Пушкина и Лермонтова представлены в [17, 20] в лексемах. Первое число в таблице 2 указывает СЗ лексем во всём творчестве АП и МЛ, а в скобках указан СЗ только в поэтическом творчестве. Данные [9; 17; 18; 20] сведены в таблицу 2.

Таблица 2. Данные ИРЯ РАН для 10 поэтов конца XIX и первой половины XX века, включая серебряный век русской поэзии [18].

| Автор, возраст   | КСУ<br>ИРЯ | СЗ различных<br>словоформ |
|------------------|------------|---------------------------|
| И. Анненский, 55 | 20053      | 7802                      |
| А. Ахматова, 77  | 47572      | 14215                     |
| А. Блок, 40      | 104357     | 30154                     |
| С. Есенин, 30    | 46849      | 14684                     |
| М. Кузмин, 63    | 50140      | 16671                     |

|                      |        |             |
|----------------------|--------|-------------|
| О. Мандельштам, 47   | 47714  | 16060       |
| В. Маяковский, 37    | 50801  | 20007       |
| Б. Пастернак, 70     | 89121  | 29599       |
| В. Хлебников, 36     | 68123  | 21851       |
| М. Цветаева, 49      | 140257 | 43593       |
| <b>Для сравнения</b> | НКРЯ   | СЗ лексем   |
| А. Пушкин, 37        | 195572 | 21290 (13К) |
| М. Лермонтов, 27     | 132819 | 14939 (10К) |

По данным словаря ИРЯ в поэтическом творчестве поэтов первой половины XX века наибольшее количество словоупотреблений (КСУ) у Марины Цветаевой – 140 тысяч, а затем у Александра Блока – 104 тысячи, словарные запасы словоформ у Цветаевой и Блока также наибольшие. По данным поэтического корпуса НКРЯ у Александра Пушкина в поэтическом творчестве 195 тысяч словоупотреблений, а у Михаила Лермонтова 132 тысячи. Всего словарный запас А. Пушкина насчитывает более 21 тысячи лексем, а М. Лермонтова около 15 тысяч лексем. В поэзии А. Пушкина около 13 тысяч (13К) лексем, а в поэзии М. Лермонтова около 10 тысяч (10К) лексем [17, 20].

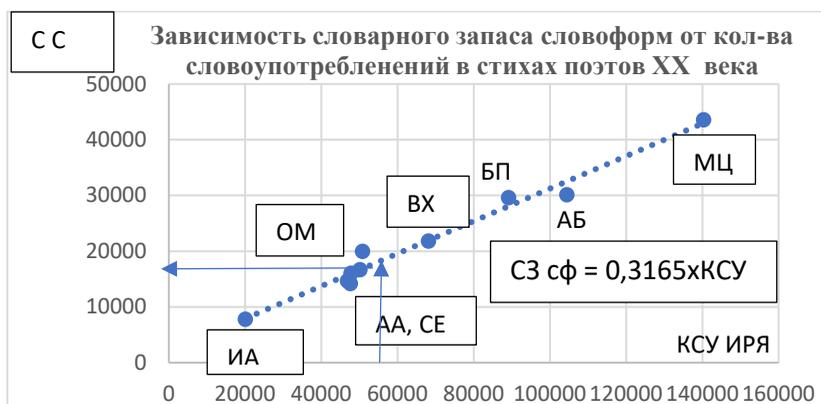
Как уже было отмечено, словарные запасы поэтов первой половины XX века указаны в словаре ИРЯ [18] в словоформах, а словарные запасы Пушкина [17] и Лермонтова [20] – в лексемах. Из таблицы 1 понятно, что словарные запасы поэтов в словоформах могут в несколько раз превышать словарные запасы в лексемах.

Обилие чисел в табл. 2 не позволяет представить в целом картину поэтического ландшафта первой половины XX века; «за деревьями леса не видно», поэтому проведём визуализацию данных.

Отложим в системе координат по горизонтали на оси абсцисс количество словоупотреблений (КСУ) у поэтов первой половины XX века, а по вертикали на оси ординат словарные запасы словоформ (СЗ сф). Каждый поэт, таким образом, представлен точкой в системе координат, и поэты XX века выстроились почти по одной восходящей прямой, как звёзды на небе в поясе созвездия Ориона, см. график 1.

Точки на графике 1 лежат очень близко к прямой линии, что говорит о сильной корреляции (зависимости) словарного запаса поэта от общего количества написанных им слов в произведениях: чем больше написал поэт, тем больше его словарный запас. Такая эмпирическая зависимость соответствует нашим интуитивным представлениям о структуре поэтического текста. Легко найти формулу как считать СЗ поэта по значению его КСУ, но проще пользоваться графиком для оценки словарного запаса СЗ и ниже мы покажем, как это делать.

График 1. Данные ИРЯ: КСУ и СЗ сф для поэтов XX века.



Сокращения: ИА – И. Анненский; СЕ – С. Есенин;  
АА – А. Ахматова; ОМ – О. Мандельштам; МК – М. Кузмин;  
ВМ – В. Маяковский; ВХ – В. Хлебников; БП – Б. Пастернак;  
АБ – А. Блок; МЦ – М. Цветаева. Стрелками показано как определить словарный запас поэта, не входящего в словарь, например, Игоря Северянина (ИС).

Из графика 1 мы видим, что по данным ИРЯ РАН существует сильная линейная возрастающая зависимость между КСУ и СЗ словоформ для 10 поэтов конца XIX – первой половины XX века. Полученная модель (график) и методика может использоваться для оценки СЗ других поэтов, не входящих в словарь. Оценка другого поэта по графику 1 возможна, начиная с КСУ 20 тысяч

словоупотреблений, а по формуле с нуля, что важно для поэтов с малым количеством словоупотреблений.

КСУ показывает языковую энергетику, словесный потенциал, дар письменной речи поэта. Например, у Игоря Северянина по данным ПКРЯ количество словоупотреблений КСУ равно 56982 слова и по графику 1 находим, что его словарный запас должен быть около 18 тысяч словоформ:  $C3\text{ сф} = 0,3165 \times \text{КСУ}$  (на графике 1 показано стрелками).

По количеству словоупотреблений (КСУ) и словарным запасам словоформ (СЗ сф) поэты делятся в системе координат по словесному потенциалу на пять групп-кластеров:

1) ИА; 2) СЕ, АА, ОМ, МК; 3) ВМ, ВХ; 4) БП, АБ; 5) МЦ.

Из графика 1 видно, что по словесному потенциалу Игорь Северянин входит в кластер 2) СЕ, АА, ОМ, МК.

Кроме данных ИРЯ РАН2001 года, существует более полная по количеству словоупотреблений новая версия Поэтического корпуса русского языка, которая постоянно пополняется как новыми поэтами разных эпох, так и расширением произведений уже имеющегося в поэтическом корпусе списка 951 поэта. Поэтому в новой версии ПКРЯ [9] могут быть более полные данные КСУ поэтов несколько большие, чем данные ИРЯ РАН [18]. Надо заметить, что и в НКРЯ далеко не у всех поэтов представлен полный корпус их стихотворных произведений.

### **3. Анализ словарей русских поэтов XIX века**

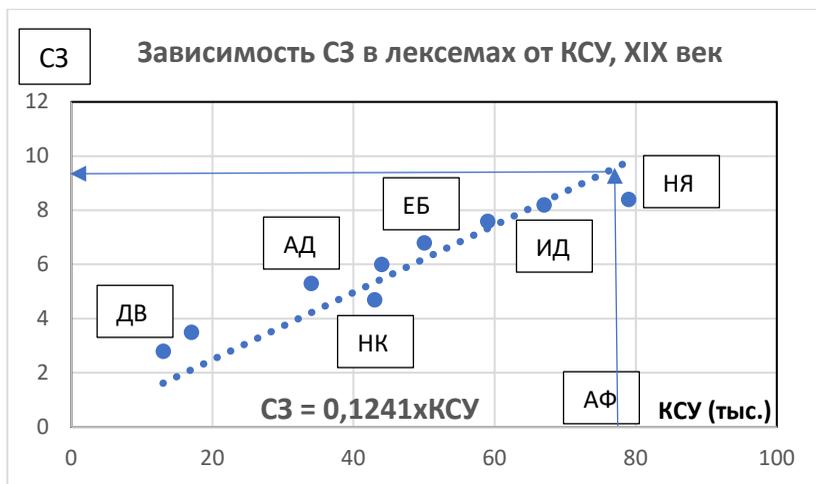
По данным словарей русских поэтов XIX века (составители Н. Л. Васильев и Д. Н. Жатков [4]) представлена сводная таблица количества словоупотреблений (КСУ) в поэтическом творчестве поэтов и количества *лексем* в их словарях.

Таблица 3: данные поэтов XIX века по словарям В. Н. Васильева и Д. Н. Жаткова (В. Ж.)

| Поэт, сокращение    | КСУ В.Ж. (тыс.) | Кол-во лексем<br>(тыс.) |
|---------------------|-----------------|-------------------------|
| А. Дельви́г (АД)    | 34              | 5,3                     |
| Н. Языков (НЯ)      | 79              | 8,4                     |
| К. Рылеев (КР)      | 44              | 6                       |
| А. Полежаев (АП)    | 67              | 8,2                     |
| И. Дмитриев (ИД)    | 59              | 7,6                     |
| Е. Баратынский (ЕБ) | 50              | 6,8                     |
| Д. Веневитинов (ДВ) | 13              | 2,8                     |
| Д.В. Давыдов (ДВВ)  | 17              | 3,5                     |
| Н. Карамзин (НК)    | 43              | 4,7                     |

По данным табл. 3 построен график 2 зависимости количества лексем в словарных запасах поэтов от количества словоупотреблений в их поэтическом творчестве.

График 2: зависимость количества лексем в словарях поэтов от количества словоупотреблений в их поэтическом творчестве по данным В. Ж. для поэтов XIX века.



По данным [4] существует сильная линейная возрастающая зависимость между КСУ и величиной словарных запасов в *лексемах* для поэтов XIX века:  $СЗ \text{ лексем (тыс.)} = 0,1241 \times КСУ \text{ (тыс.)}$ .

Таким образом, по двум независимым источникам [18] и [4], в разных эпохах (XIX и XX веков) и в разных единицах измерения словарных запасов (в словоформах и в лексемах) существуют сильные возрастающие зависимости между количеством словоупотреблений в стихах поэтов и величиной их словарных запасов. По этим моделям и графикам можно оценивать СЗ поэтов, не вошедших в словари. Например, у Афанасия Фета (АФ) по данным ПКРЯ количество словоупотреблений КСУ равно 79822 слова и по графику 2 находим, что его словарный запас должен быть около 9,9 тысячи лексем (показано стрелками на графике 2).

Разные источники дают несколько отличающиеся количеством словоупотреблений данные, например, у Ивана Дмитриева в НКРЯ КСУ – 56 тысяч, а КСУ В.Ж. – 59 тысяч, но между данными НКРЯ и В. Ж. существует сильная корреляция, и ей можно пользоваться для уточнения КСУ. Количество словоупотреблений у поэтов в словарях В. Ж. больше, чем в НКРЯ, однако в НКРЯ намного больше список поэтов (951 поэт разных веков), поэтому надо использовать преимущества двух независимых источников.

#### **4. Словари поэтов XVIII – XIX веков, словесные потенциалы и возраст поэтов**

В синтаксических словарях собраны поэты XVIII и XIX веков, составитель Н. В. Патроева [15; 16]. Для каждого поэта из указанных словарей с помощью новой версии поэтического корпуса русского

*Я. М. Марголис*

языка было найдено количество словоупотреблений (КСУ) и возраст, результаты представлены в таблицах 4 и 5.

Лидером в XVIII веке по количеству словоупотреблений КСУ в поэтическом творчестве является Гавриил Державин, проживший 73 года, – 93426 словоупотреблений. Меньше всех прожил Антиох Кантемир – 32 года, но его КСУ значителен – 49935, а меньше всего КСУ – 20111 – у Александра Радищева, прожившего 53 года.

Табл. 4. Количество словоупотреблений КСУ в словарях поэтов XVIII века [16].

| Поэты 18 века     | Сокращ. | Возраст | КСУ НКРЯ |
|-------------------|---------|---------|----------|
| Г. Державин       | ГД      | 73      | 93426    |
| А. Кантемир       | АК      | 35      | 49935    |
| Н. Карамзин       | НК      | 59      | 40719    |
| М. Ломоносов      | МЛ      | 53      | 53303    |
| А. Радищев        | АР      | 53      | 20111    |
| В. Третьяковский  | ВТ      | 66      | 68500    |
| А. Сумароков      | АС      | 59      | 76598    |
| М. Херасков       | МХ      | 73      | 56702    |
| И. Дмитриев       | ИД      | 77      | 55564    |
| Среднее значение: |         | 61      | 57206    |

Средний возраст поэтов в XVIII веке по словарю [16] составлял 61 год, а среднее КСУ на одного поэта XVIII века – 57 тысяч слов.

Как зависит количество словоупотреблений в стихах поэтов XVIII века от возраста, видно из графика 3.

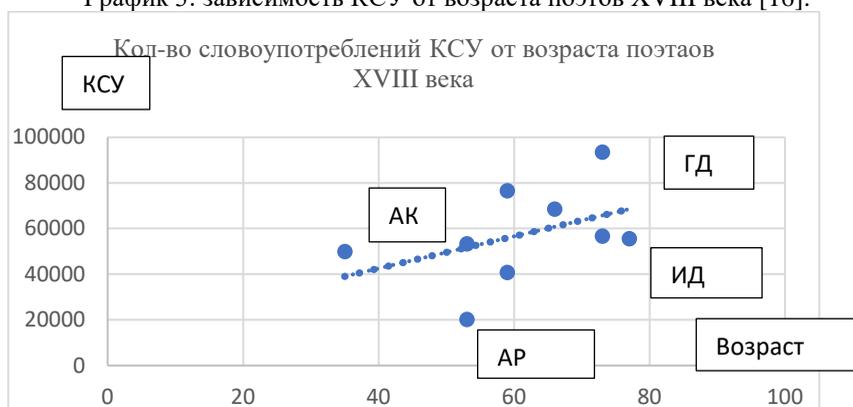
Для поэтов XVIII века в целом наблюдается слабая зависимость между возрастом и КСУ с большим разбросом данных (низкий коэффициент корреляции). Как мы увидим ниже, слабая зависимость между количеством прожитых лет и количеством написанного поэтом сохранится и в XIX веке, но эта зависимость отсутствует в первой

половине XX века. Для сравнения с XVIII веком: средний возраст поэтов в XIX веке [15] снизился и составил 55 лет, а среднее КСУ увеличилось и составило 68 тысяч слов. Средний возраст для поэтов первой половины XX века [18] составил 50 лет, а КСУ по новой, более полной версии НКРЯ составляет 109 тысяч слов.

Таким образом, мы видим, что количество словоупотреблений, соответственно графикам 1 и 2, и словарные запасы поэтов растут со временем от XVIII до первой половины XX века включительно, и границы поэтического языка расширяются в указанный период.

Оценка среднего КСУ для второй половины XX века не проводилась, поскольку пока нет единого мнения специалистов о том, каких поэтов следует включить в выборку для исследования.

График 3: зависимость КСУ от возраста поэтов XVIII века [16].



Слабая зависимость КСУ от возраста поэтов XVIII века, так как большой разброс точек от графика.

Табл. 4. Количество словоупотреблений в словарях поэтов XIX века [9, 15]

| Имя поэта      | Сокращ. | Возраст | КСУ НКРЯ |
|----------------|---------|---------|----------|
| А. Пушкин      | АП      | 37      | 195572   |
| М. Лермонтов   | МЛ      | 27      | 132819   |
| Е. Баратынский | ЕБ      | 44      | 39641    |

|                  |    |    |        |
|------------------|----|----|--------|
| В. Жуковский     | ВЖ | 69 | 347067 |
| П. Вяземский     | ПВ | 86 | 68076  |
| К. Батюшков      | КБ | 68 | 34879  |
| И. Крылов        | ИК | 75 | 42615  |
| А. Дельвиг       | АД | 32 | 28428  |
| Н. Языков        | НЯ | 43 | 60538  |
| Ф. Тютчев        | ФТ | 69 | 33818  |
| Среднее значение |    | 55 | 68039  |

Слабая линейная зависимость между количеством прожитых лет и количеством написанного поэтом в XIX веке объясняется низким коэффициентом корреляции. КСУ В. Жуковского указано в НКРЯ вместе с его переводами «Одиссеи» Гомера (около 100 тысяч КСУ) и другими переводами, которые не учитывались. С появлением нового, более точного словаря поэта [4] можно будет уточнить среднее значение словоупотреблений для поэтов XIX века. Ф. Тютчев включён в табл. 4, где дана характеристика авторов с использованием данных НКРЯ.

Средний возраст поэтов в первой половине XX века [18] ещё более снизился и составил 50 лет, однако среднее КСУ по данным новой версии НКРЯ составляет 109 тысяч слов. Среднее КСУ по новой версии НКРЯ и границы поэтического языка расширяются: в 18 веке – 57 тысяч, в 19 веке – 68 тысяч, в первой половине 20 века – 109 тысяч слов.

График 4. Слабая возрастающая зависимость между количеством прожитых лет и количеством написанного поэтом в XIX веке.

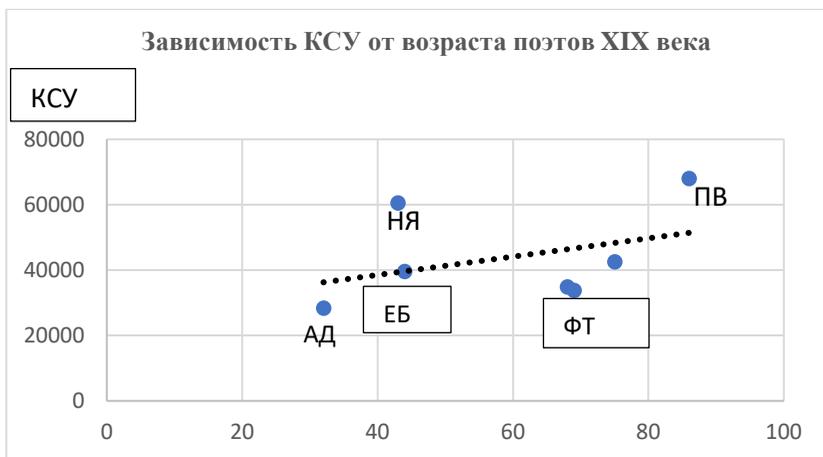
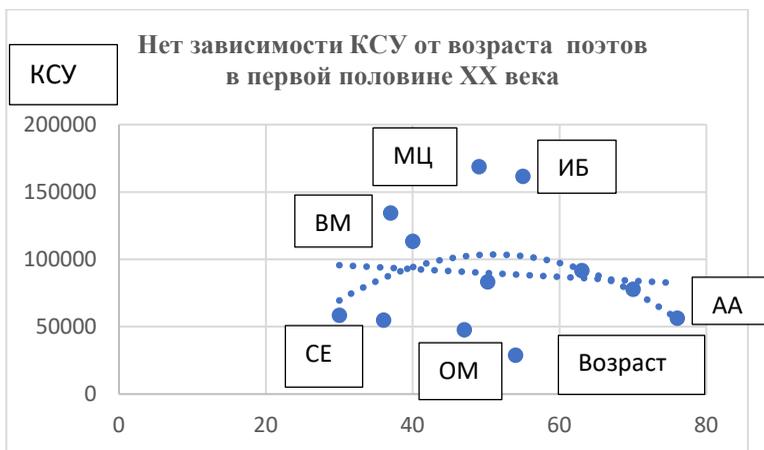


График 5. Для десяти поэтов первой половины XX века [18] по данным НКРЯ и одного поэта из второй половины XX века (И. Бродский) по его конкордансу [12] данные представлены в координатах возраст – КСУ, но зависимости ни линейной, ни параболической не обнаружено.



В первой половине XX века в русской поэзии не наблюдается зависимости: чем больше лет прожил поэт, тем больше он написал. Как заметил поэт А. Кушнер, *«Времена не выбирают, в них живут и умирают»*. За каждой точкой на графике 5 стоят сломанные судьбы выдающихся русских поэтов XX века: средний возраст поэтов в первой

половине XX века самый низкий – всего 50 лет. Мы видим, что «распалась связь времён»: в первой половине XX века в русской поэзии четыре трагедии – три самоубийства и одна гибель в ГУЛАГе.

*Век мой, зверь мой, кто сумеет  
заглянуть в твои зрачки  
и своею кровью склеит  
двух столетий позвонки?*

.....  
*Мне на плечи кидается век-волкодав,  
Но не волк я по крови своей:  
Запихай меня лучше, как шапку, в рукав  
Жаркой шубы сибирских степеней...*

Осип Мандельштам, конец 1935 года

### **5. Сравнение словарных запасов и транзит поэтического языка**

Сравнение словарных запасов позволяет найти слова и предпочтения, которые есть в лексиконе одного поэта, например, А. Пушкина (АП): «пиит» (19), «воитель» (19), «сводня» (17), «пастушка» (15), клюка (8) и подобные им архаичные слова из лексики прошлых веков, но этих слов нет в лексиконе другого поэта, например, М. Лермонтова (МЛ). Частота таких слов у него равна нулю, МЛ (0), см. словари АП и МЛ [17; 20].

Таким образом, можно узнать, какая часть словарного запаса МЛ общая со словарным запасом АП, а какая часть словарного запаса МЛ отличается от него, переходит в транзит. Поясним эту задачу графически по данным словарей АП [17], МЛ [20] и [3; 4].

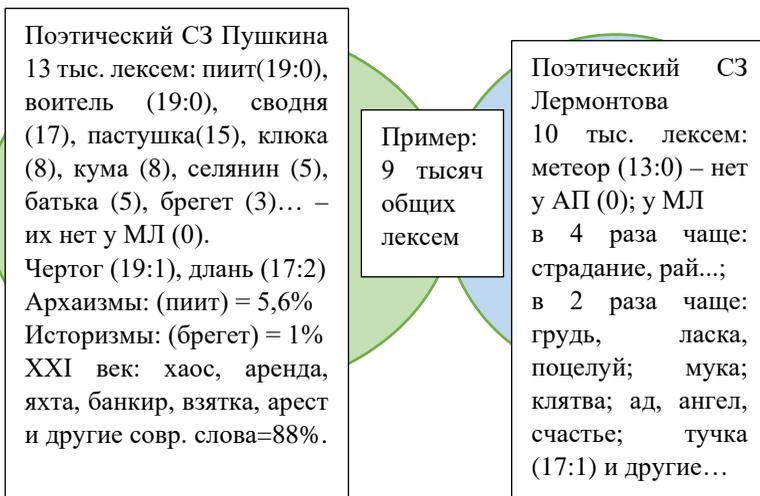


График 6. Сравнение словарных запасов двух поэтов (АП – зелёный, МЛ – синий) позволяет определить, какие слова, в каком контексте и в каком количестве появляются в словаре другого поэта намного чаще и количественно оценить транзит поэтического языка от поэта к поэту. 9 тысяч общих лексем взяты для примера, остальные числа – фактические данные [3; 4; 17; 20].

У АП чаще, чем у МЛ встречаются слова: «поэт» (230), «стих» (193), «вино» (154), «муза» (148), «певец» (141), «лира» (129), «пир» (119).

У Лермонтова меньше архаизмов и чаще, чем у Пушкина, встречаются слова: метеор (13:0) у АП (0); у МЛ в 4 раза чаще: страдание, рай; в 2 раза чаще: грудь, ласка, поцелуй; мука; клятва; ад, ангел; тучка (17:1). Слова поэтов говорят о лейтмотивах их творчества. А. Пушкин писал: «*На свете счастья нет, но есть покой и воля*». Относительная частота лексемы «счастье» встречается у АП в два раза реже, чем у МЛ, а лексема «покой» чаще.

Слова-архаизмы А. Пушкина («пиит», «воитель» и др.) в современном русском языке заменены синонимами («поэт», «воин» и др.), а слова-историзмы, например, «брегет» (т. е. часы) ушли в прошлое навсегда, и этого слова уже нет в словаре М. Лермонтова:

*Надев широкий боливар,  
Онегин едет на бульвар  
И там гуляет на просторе,  
Пока недремлющий брегет  
Не прозвонит ему обед.*

.....

*Еще бокалов жажда просит  
Залить горячий жир котлет,  
Но звон брегета им доносит,  
Что новый начался балет.*

А. С. Пушкин. Евгений Онегин

Один из ведущих современных исследователей языка Пушкина профессор Н. Л. Васильев пишет в своей монографии «О Пушкине: язык классика»: «Анализ лексики, зафиксированной в словаре писателя, показал следующее. Общее число устаревших лексем составляет в нем чуть более тысячи единиц. Среди них преобладают архаизмы, т. е. слова, замененные сейчас синонимами; в меньшей мере присутствуют историзмы, т. е. слова, которые означают реалии, ушедшие в прошлое. Доля первых в пушкинском лексиконе составляет 5,6%, вторых – 1%.

Среди историзмов присутствуют слова абшид («свидетельство об отставке»), брегет («часы особой конструкции»), ванька («извозчик из крестьян»), капот («вид женской одежды»), камердинер («слуга»), картель («письменный вызов на дуэль»), ливрея («парадная одежда слуг»), мушка («искусственная родинка на лице») и другие... Проведенные исследования показали, что доля стилистически нейтральных слов в языке Пушкина составляет 88%... Следовательно, в подавляющем большинстве слова, употребленные писателем, по-прежнему нужны в современной речи» [2].

Посмотрим на примере архаизмов «пиит» и «длань» их словоупотребления во времени в языке поэтов разных времён и поколений: у Г. Державина, вторая половина XVIII века, 14 словоупотреблений «пиит» и 41 «длань»:

*Так, ты, греко-российска Харита!  
Вблизи как меня восседая,  
Коснулась ко мне дланью пиита,  
Со мной однодушно дыхая...*

Г. Р. Державин «Полигимнии» (1816)

У А. Пушкина «пиит» (19) и «длань» (17), у М. Лермонтова «пиит» (0) и «длань» (2), что мы уже видели при сравнении со словарным запасом А. Пушкина. Для поэтов XX века с наибольшим количеством словоупотреблений в поэзии: у М. Цветаевой, первая половина XX века, «пиит» (2), «длань» (15), у И. Бродского, вторая половина XX века, «пиит» (0), «длань» (0). При этом заметим, что словоупотребление «длань», но уже в современном виде «ладонь, рука» наиболее часто встречается в словарных запасах поэтов XX века: М. Цветаевой, О. Мандельштама, А. Ахматовой, И. Бродского.

В начале 60-х годов прошлого века первые результаты работы по статистическому анализу лексики Пушкина группой «Словарь языка Пушкина» Института русского языка и сектора прикладного языкознания Института языкознания дали интересные результаты. Специалист группы «Словарь языка Пушкина» Р. М. Фрумкина поставила вопрос о сравнении наиболее частых слов в словарных запасах поэтов, например, Карамзина и Пушкина: «Полученные результаты позволяют поставить следующий вопрос. Несомненно, что, например, язык Карамзина является по отношению к современному состоянию русского литературного языка более ранней ступенью, чем язык Пушкина. Вместе с тем в сочинениях Карамзина, несомненно, встречаются все или почти все те слова, которые входят в 500 самых частых слов у Пушкина. Будут ли эти слова у Карамзина входить в ту же

частотную группу, что и у Пушкина? В качестве рабочей гипотезы можно высказать мысль о том, что список «500 самых частых слов у Пушкина» и «500 самых частых слов у Карамзина» совпадут лишь частично: часть тех слов, которые у Карамзина принадлежат к 500 самым частым словам, у Пушкина окажутся употребляемыми реже и войдут, например, лишь в список «2000 самых частых слов», и т. д. Представляется, вообще говоря, возможным, располагая лексику каждого из двух или более авторов в порядке убывания частот и рассматривая с разных точек зрения степень коррелированности таких списков, найти некоторые объективные характеристики словаря именно данного писателя. Можно полагать также, что исследование статистических характеристик словарей различных писателей даст возможность отделить индивидуальные особенности лексики писателя от фона, характерного для эпохи в целом» [19].

Предложенное сравнение словарных запасов поэтов отвечает на вопрос, поставленный 60 лет назад, когда ещё не было словарей других поэтов, кроме Пушкина, а компьютеры были большой редкостью. Подобная методика позволяет сегодня, в век компьютерных технологий и при наличии множества словарей, количественно оценить транзит языка одного поэта по отношению к языку другого поэта.

Покажем это на условном примере (график 6): допустим, что 9 тысяч лексем являются *пересечением*, общей частью поэтических словарных запасов Пушкина и Лермонтова. Такие пересечения словарных запасов поэтов можно определить, например, с помощью современных компьютерных программ Statistical Analysis System (SAS) [10].

Тогда у Лермонтова будет: 10 тысяч – 9 тысяч лексем = 1 тысяча лексем, переходящих в транзит, не входящих в СЗ Пушкина или значительно превышающих по частоте употребления лексем у Пушкина, что составляет  $1/10=10\%$  от поэтического словарного запаса Лермонтова.

Если мы берём для сравнения словарей выборки, а не все словарные запасы поэтов, то выборки должны быть *репрезентативными* (представительными для всего СЗ) как по количеству, так и по методу отбора лексем в выборки.

Можно взять словарь Пушкина за эталон, и от него отсчитать транзит словарей поэтов других поколений, как предыдущих (Державин, Кантемир, Ломоносов), так и будущих, например, Цветаевой – первая половина XX века, Бродского – вторая половина XX века и современных русскоязычных поэтов XXI века.

## **6. Критерии транзита поэтического языка от Пушкина до наших дней**

Оценить такой сложный процесс, как изменение, транзит поэтического языка от поэта к поэту или из одной эпохи в другую эпоху одним количественным, статистическим критерием (отличием словарных запасов поэтов) невозможно. Поэтому введём ещё качественный, филологический критерий, показывающий, за счёт чего происходит разность словарных запасов поэтов и её филологическую направленность. Качественное отличие поэтов разных эпох необходимо объяснять ещё диалектикой семантики, синонимией, различиями терминологии и разными уровнями языка: лексикой, фонетикой, морфологией, словообразованием, синтаксисом, в этих деталях скрыт смысл транзита языка.

Идея сравнения словарей Пушкина и Карамзина была впервые высказана сразу после появления «Словаря языка Пушкина» Р. М. Фрумкиной [19] и нашла своё продолжение в работах Н. Л. Васильева и Д. Н. Жаткина [3; 4], где сравнивался словарь Пушкина со словарями поэтов XIX века. Важно продолжить эти исследования дальше и сравнить словарный запас (СЗ) А. Пушкина (АП) со словарями поэтов XX века: Цветаевой, Блока, Мандельштама, Ахматовой, Пастернака, Есенина, Хлебникова, Маяковского, Бродского и молодых русскоязычных поэтов XXI века.

Таким образом, можно количественно (статистически) и качественно (филологически) оценить транзит поэтического языка от Г. Державина и А. Пушкина до М. Цветаевой и И. Бродского и до наших дней через изменения словарных запасов и лексики поэтов следующих эпох. Говоря метафорически, поэтический язык – это почтовый поезд, проходящий транзитом через станции поэтов и забирающий на каждой из них письма-стихи. Поэты пишут письма-стихи и отправляют их с почтовым поездом времени на следующие станции – в будущее.

Использование двух критериев (количественного – статистического и качественного – филологического) позволяет с двух сторон оценить транзит поэтического языка от поэта к поэту и даёт филологам более объективные инструменты и критерии для дальнейших исследований.

Табл. 6. Отличия словарей поэтов XIX – XXI веков от словарного запаса (СЗ) А. Пушкина

| Поэты XIX века  | Количественные отличия от СЗ АП   | Качественные отличия от СЗ АП  |
|---|---|--|
| М. Лермонтов  | 12,6% [3, 4]<br>                 | Лексика: меньше архаизмов, историзмов, меньше проявлений поэтики классицизма и «высокой» поэтической лексики. Стилистика: меньшее разнообразие стилей и другие отличия |
| П. Вяземский<br>К. Рылеев   | 20,2%;<br>11,7%;  | Пример архаизма: лексема «длань»   |
| А. Дельвиг<br>Н. Языков и др.   | 9,4%;<br>14,9% [3, 4]   | 18 в.: Г. Державин<br>(41 раз)   |
| Поэты XX века   |                                  | 19 в.: А. Пушкин<br>(17 раз)   |
| А. Ахматова,<br>А. Блок,<br>С. Есенин,<br>О. Мандельштам,<br>В. Маяковский,<br>Б. Пастернак,<br>М. Цветаева,<br>И. Бродский | М. Цветаева<br><br>И. Бродский | Первая половина XX века: М. Цветаева<br>(15 раз)<br><br>Вторая половина XX века: И. Бродский (0)   |

| Русские поэты XXI века  | ↓                  | Русскоязычные поэты XXI века  |
|---|--------------------|---|
| В. Бородин,<br>Р. Амелин,<br>В. Банников,<br>А. Глазова,<br>Л. Юсупова,<br>К. Корчагин,<br>и другие поэты РФ. | Начало XXI<br>века | М. Бородин, Укр.;<br>П. Банников, Казахс.;<br>Д. Стронец, Беларус.;<br>С. Тимофеев, Латвия;<br>И. Кулишова, Груз.;<br>Ш. Абдуллаев, Узб.;<br>и поэты др. стран. |

В табл. 6 словарный запас А. Пушкина взят за эталон, за точку отсчёта. Интересно посчитать отличия от СЗ А. Пушкина СЗ поэтов XX века.

### 7. Профиль поэта: слова, наиболее часто используемые поэтами XX века

*«Составив словарь того или иного стихотворения, мы получаем – пусть грубые и приблизительные – контуры того, что составляет мир, с точки зрения этого поэта»,* – писал Ю. М. Лотман в XX веке [6].

Говоря языком XXI века, века информационных технологий, можно сказать, что поэтические словари – это банки данных, где поэты хранят свои духовные капиталы.

Профессор славистики из Канады Т. Патера составила конкордансы А. Ахматовой и И. Бродского [11; 12], и подсчитала их словарные запасы по одной и той же методике. Поэтому можно сравнить, например, наиболее часто встречающиеся существительные у двух поэтов, увидеть в них общие лексемы в контексте стихов, их комбинации и то, какая часть словарного запаса Иосифа Бродского пересекается со словарным запасом его духовной наставницы Анны Ахматовой, а какая часть отлична от него, переходит в транзит.

Конкорданс И. Бродского в 6 томах [12] основан на 587 стихотворениях и поэмах поэта, содержит 153860 словоупотреблений и 19650 лексем (в скобках указаны частоты лексем по мере их убывания).

Существительные: **жизнь (385)**, глаз (302), время (299), день (261), окно (254), человек (249), рука (229), земля (228), год (227), ночь (219), слово (215), лицо (214), свет (206), душа (204), **любовь (198)**,

небо (193), мир (191), снег (191), вещь (188), **смерть (177)**, тело (163), память (63), язык (речь, 38).

Глаголы: **быть** (1148), знать (227), сказать (226), видеть (212), смотреть (178), говорить (176), мочь (176), идти (175), спать (147), стоять (137), жить (135), глядеть (102), думать (101), лететь (98).

Другие части речи: как (1938), что (1675), нет/нету (369), да (306); будто (119), ибо (108), словно (93). Цвета: чёрный (120), белый (107), жёлтый (49), красный (46), подробнее о цвете в словаре И. Бродского см. книгу проф. В. Полухиной [13].

Посчитаем относительную частоту (ОЧ) лексемы «любовь» у И. Бродского (ИБ) и сравним её с ОЧ на 10 тысяч словоупотреблений у А. Ахматовой и М. Цветаевой, у В. Жуковского и А. Пушкина, как мы это сделали выше. В конкордансе к поэзии И. Бродского [12] находим:  $OЧ_{ИБ} = 198 / (153860 / 10000) = 12,87$  раз лексема «любовь» встречается у ИБ в 10000 словоупотреблений vs. 14,18 у Ахматовой; 13,69 у Цветаевой; 13,20 у Жуковского; 24,25 у Лермонтова и 25,22 у Пушкина.

Посмотрим конкорданс Ахматовой [11].

Существительные: рука (142), глаз (126), голос (124), день (122), дом (111), сердце (106), ночь (102), слово (96), ветер (95), небо (89), земля (86), жизнь (79), смерть (79), город (77), любовь (77), тень (74).

Глаголы: видеть (85), сказать (83), говорить (78), петь (76), жить (74), глядеть (73). Цвета: чёрный (110), белый (100).

Откроем конкордансы Мандельштама [5, 7].

Существительные: вода (77), земля (72), небо (69), ночь (68), рука (68), воздух (67), человек (60), жизнь (57), мир (57), день (54), глаз (53), голова (51), звезда (51), век (48), слово (46), кровь (41), лес (41), солнце (41), дом (40), улица (39), сердце (39).

Глаголы: быть (215), сказать (60), любить (56), петь (56), жить (55), идти (52), хотеть (49), говорить (47), мочь (45). Мандельштам – поэт прилагательных и самые частые у него: нежный (48), лёгкий (47), какой (47), тёмный (43). Цвета: чёрный (82), белый (50).

Частые существительные у Пушкина – «поэт» и «стих», у Ахматовой – «рука», «глаз», «голос», «сердце», у Мандельштама – «вода», «земля», «небо», у Цветаевой – «руки», «сердце», «день», «бог», «глаза», у Бродского – «жизнь», «глаз», «время», «день».

У Лермонтова слово «смерть» встречается чаще, чем у Пушкина, и как писал Тютчев: *«Нам не дано предугадать, как слово наше отзовется...»*.

У Ахматовой самые часто встречающиеся существительные: рука (142), глаз (126), голос (124), сердце (106) связаны с телом,

а существительные «жизнь» и «смерть» имеют значительно меньшую, но одинаковую частоту 79.

У Мандельштама три самых часто встречающихся существительных – вода (77), земля (72), небо (69) связаны с природой и сотворением мира: «*В начале сотворил Бог небо и землю*» (Библия, Бытие, гл.1.1.).

У Цветаевой – руки (169), сердце (148), день (144), бог (143), глаза (138), относительная частота лексемы «любовь» – близка к частоте Ахматовой.

У Бродского, который так много писал о смерти, начиная с юности, как это ни парадоксально, самым часто встречающимся в стихах существительным является лексема «жизнь» (385 раз!), а лексема «смерть» (177 раз) встречается в два раза реже.

Наиболее часто встречающиеся у поэтов цвета – чёрный и белый. У ИБ: чёрный (120), белый (107); у АА: чёрный (110), белый (100); у ОМ: чёрный (82), белый (50), у МЦ: чёрный (29), белый (29).

Границы поэтического языка у поэтов первой половины XX века (среднее количество словоупотреблений на одного поэта по данным НКРЯ 109 тысяч) почти в два раза шире, чем у поэтов XVIII века (57 тысяч словоупотреблений на одного поэта). Пороговые значения границ поэтического языка определены по данным новой версии поэтического корпуса русского языка и могут уточняться по мере пополнения поэтического корпуса и появления новых словарей поэтов. Поэтический язык не стоит на месте, он всё время развивается, архаизмы уходят, а новая лексика пополняет его. Пушкин не понял бы выражение: «Сегодня читали поэму Цветаевой «Автобус» («автобус» – термин из XX века) в зуме («зум» – термин из XXI века).

Ещё одна метафора: поэтический язык – это бассейн слов, из одной трубы в него вливаются новые слова, а в другую трубу выливаются старые. Граница поэтического языка – уровень слов в поэтическом бассейне, и, как было показано выше, границы русского поэтического языка расширяются от XVIII века до первой половины XX века.

Для второй половины XX века такие исследования пока не проводились, поскольку современные исследователи не пришли к единому мнению: кого из поэтов включать в выборку, чтобы она была репрезентативной?

## **8. Каким глаголом «жечь сердца людей» и какими существительными?**

Самым часто встречающимся глаголом в стихах поэтов является глагол «*быть*»: у ИБ 1148 раз, у МЦ он повторяется 436 раз, у ОМ 215 раз. Как тут не вспомнить Шекспира: «Быть или не быть». Кстати, о Шекспире: словарный запас его пьес по разным методикам от 18 до 29

тысяч лексем, порядка как у Пушкина во всём творчестве – более 21 тысячи лексем.

Мы видим, что оценка словарных запасов – дело сложное, зависит от поэтов, которых включили в выборку, от полноты выбранных стихотворений и поэм, от количества взятых в выборку словоупотреблений, от единиц измерения словарных запасов, методики их подсчёта и других факторов. Поэтому важно учитывать все источники текстов, все словари и конкордансы поэтов и сравнивать их между собой. Всё познаётся в сравнении!

Из 20 самых частых существительных в стихах АА, ОМ, МЦ, ИБ у всех поэтов чаще всего встречаются: «рука», «день», «жизнь», «глаз», «слово» – эти и другие лексемы составляют *ядро слов*, пересечение *всех* словарных запасов 4 поэтов, их общую часть. Интересно, какое ядро слов будет у молодых поэтов XXI века и будет ли оно пересекаться с ядром слов поэтов XX века? Понятно, что надо рассматривать слова в контексте, то, в каком смысле они употребляются: «рукой» можно бить, но «рукой» можно ласкать.

*Поэзия* – чудесная цепочка чувств, мыслей, *слов*, созвучий, образов и чего-то неведомого, растворённого между ними. Контекст – это уже следующее, *смысловое звено* поэтической цепочки: чувство – *смысл* – слово. Современные программы искусственного интеллекта позволяют анализировать поэтические тексты не только по словам, но и по смыслу, создавать файлы «*смысловых запасов*» поэтов и сравнивать их между собой, так же как мы сравнивали «словарные запасы» поэтов.

Рассматривая далее словари АА, ОМ, МЦ, ИБ, мы видим, что существительное «человек» есть в первых 20 словах у трёх поэтов, но нет у АА; у АА и МЦ слово «любовь» встречается чаще, чем у Бродского, а у Пушкина слово «любовь» в два раза чаще, чем у АА и МЦ. Слово «небо» в первых 20 словах есть у троих, но нет у МЦ, однако Цветаева – единственный поэт, у кого чаще, чем у других, встречается слово «Бог».

## 9. Перспектива: сравнительный анализ языка молодых поэтов XXI века

Первая треть XIX века, эпоха Пушкина и Лермонтова, считается «золотым веком» русской поэзии. Начало XX века – это «серебряный век» русской поэзии. Назовём условно и авансом начало XXI века «бронзовым веком» русской поэзии. Возьмём стихи молодых поэтов XXI века и посмотрим, действительно ли начало XXI века является «бронзовым веком» русской поэзии, или вся бронза уже ушла на памятники великим предшественникам?

Четыре независимых жюри отобрало молодых русскоязычных поэтов XXI века, живущих как в России, так и за рубежом, для сравнения и оценки транзита их поэтического языка в XXI веке на фоне языка ведущих русских поэтов XVIII–XX веков.

Настоящее исследование было выполнено в рамках подготовки ко второму этапу филологического проекта «Русскоязычная поэзия в транзите» университета Трира, Германия [8, 14] и лежит в русле цифровых методов современной филологии XXI века [1].

Жюри издательского проекта «Центрифуга» Центра Вознесенского в Москве (Мария Степанова, Илья Данишевский и Лев Оборин) отобрало из множества современных поэтов 4 кандидатов: Ростислава Амелина, Анну Глазову, Василия Бородина, Лидию Юсупову (два молодых поэта уже являются лауреатами литературной премии имени Андрея Белого).

Известный литературный критик Дмитрий Кузьмин, Латвия, предложил из множества кандидатов ещё пятерых молодых поэтов для проведения исследования: Кирилла Корчагина (лауреата премии имени Андрея Белого), Оксану Васякину (лауреата премии «Лицей»), Вадима Банникова (лонг-лист премии «Дебют»), Галину Рымбу и Кузьму Коблова.

Другой литературный критик современной русской поэзии Юрий Орлицкий, РГГУ, Москва, выбрал для исследования ещё одну группу поэтов, куда вошли Екатерина Соколова, Антон Азаренков, Алла Горбунова, Анна Орлицкая, Станислава Могилева.

Автором статьи были предложены для исследования транзита поэтического языка в постсоветском пространстве русскоязычные поэты, живущие в Украине (Максим Бородин), Беларуси (Дмитрий Стронец, Татьяна Скарынкина, Ольга Злотникова, Мария Малиновская), Латвии (Сергей Тимофеев), Грузии (Инна Кулишова), Казахстане (Павел Банников), Узбекистане (Шамшад Абдуллаев).

Собраны стихи поэтов за последние годы, намечены пути для оценки транзита языка поэтов XXI века во времени относительно языка русских поэтов XVIII–XX веков и в постсоветском пространстве от страны к стране.

## **10. Заключение и основные выводы исследования**

В фокусе исследования было только одно звено поэтической цепочки – *слова*, количественный анализ текстов и словарей русских поэтов XVIII–XX веков для последующего сравнения с ними языка современных поэтов XXI века и оценки транзита их поэтического языка во времени и в пространстве. Подведём основные итоги исследования.

1. Найдены сильные возрастающие зависимости между количеством словоупотреблений и словарными запасами

как в словоформах, так и в лексемах для поэтов XIX – первой половины XX веков на основе данных «Словаря языка русской поэзии XX века» Института русского языка РАН и словарей поэтов XIX века. На основе найденных зависимостей проведена оценка словарных запасов поэтов, не вошедших в словари.

2. Показано, что количество словоупотреблений и словарные запасы поэтов растут во времени от XVIII до первой половины XX века. Границы поэтического языка расширяются в указанный период; определены оценки их пороговых значений. Пороговые значения границ поэтического языка могут уточняться по мере пополнения существующих и появления новых словарей поэтов.

3. Найденны слабые возрастающие зависимости между количеством словоупотреблений и возрастом поэтов XVIII–XIX веков. Для поэтов первой половины XX века такой зависимости не обнаружено.

4. Показано, как можно проводить сравнения словарных запасов поэтов количественно (статистически) и качественно (филологически) для оценки транзита поэтического языка во времени от поэта к поэту и в пространстве от страны к стране.

5. Независимыми жюри отобраны четыре группы поэтов XXI века из России и других стран, собраны их стихи. Намечены пути для сравнения и оценки транзита их поэтического языка во времени на фоне языка поэтов XVIII–XX веков и в пространстве относительно поэтов из других постсоветских республик: Украины, Беларуси, Латвии, Грузии, Казахстана, Узбекистана.

6. Перспективы: в 2021 году будет проведен сравнительный анализ поэтического языка русскоязычных поэтов XXI века из семи постсоветских республик, результаты будут опубликованы.

### **Литература**

1. Абрамзон Т. Е. Цифровые методы современной филологии: возможности и ограничения // Мировая литература глазами современной молодежи. Цифровая эпоха: сборник материалов IV международной. молодежной научно-практической конференции, 18–20 сентября 2018 г. / науч. ред. С. В. Рудакова. Магнитогорск: Изд-во Магнитогорск. гос. техн. ун-та им. Г. И. Носова, 2018. С. 9–11.

2. Васильев Н. Л. О Пушкине: язык классика, поэтика романа «Евгений Онегин», писатель и его современники: монография / Музей заповедник А. С. Пушкина «Болдино»; Нижегородский университет. Нижний Новгород, Саранск: Изд-во Нижегород. ун-та, 2013. 387 с.

3. Васильев Н. Л., Жаткин Д. Н. Опыт сравнения поэтических лексиконов А. С. Пушкина и М. Ю. Лермонтова // Балтийский гуманитарный журнал. 2017. Т. 6. № 1 (18). С. 21–24.
4. Васильев Н. Л., Жаткин Д. Н. Словари поэтического языка поэтов XIX века. М.: Флинта; Наука, 2013–2018.
5. Koubourlis D. J. A Concordance to the Poems of Osip Mandelstam. Ithaca NY: Cornell University Press, 1974 (словарь словоформ).
6. Лотман Ю. М. О поэтах и поэзии. Анализ поэтического текста. Санкт-Петербург: Искусство – СПб, 1996. 846 с.
7. Митюшин Л. Конкорданс к стихам Осипа Мандельштама на основе собрания сочинений ОМ: в 4 т. М.: Арт-Бизнес-Центр, 1993–1997 (словарь лексем).
8. Margolis J. Количественный анализ словарей и транзита русскоязычной поэзии, тезисы доклада, университет Трир. URL: <https://bit.ly/2GEgxsK> (дата обращения: 30.11.2020).
9. Новая версия поэтического корпуса русского языка в национальном корпусе русского языка (см. в сети НКРЯ, 2018). URL: <https://ruscorpora.ru/new/> (дата обращения: 30.10.2020).
10. Oerthel F., Tuschl S. Statistische Datenanalyse mit dem Programmpaket SAS. Universität Passau, R. Oldenburg Verlag, München, Wien, 1995.
11. Patera T. A Concordance to the Poetry of Anna Akhmatova, ARDIS, 1995. 339 p.
12. Patera T. A Concordance to the Poetry of Joseph Brodsky. Books 1-6 (1962–1996). The Edwin Mellen Press, 2002 (Slavistik, McGill University, Montreal, Canada).
13. Полухина В. Словарь цвета поэзии Иосифа Бродского. Москва: НЛО, 2016. 300 с.
14. Русскоязычная поэзия в транзите, исследовательский проект университета Трира, Германия, рук. Х. Шталь, 2017–2021.
15. Синтаксический словарь русской поэзии XIX века / под ред. Н. В. Патроевой. От Пушкина до Языкова (готов к печати).
16. Синтаксический словарь русской поэзии XVIII века: в 4 т. / под ред. Н. В. Патроевой. Т. 1, 2: Кантемир, Тредиаковский, Ломоносов. Санкт-Петербург: Изд-во Дмитрий Буланин, 2018.
17. Словарь языка Пушкина: в 4 т. / отв. ред. акад. В. В. Виноградов, 1956–1961 (1-е изд.), 2000 (2-е изд.), Москва: «Азбуковник», 2000.
18. Словарь языка русской поэзии XX века. Т. I–VIII / сост.: Л. Л. Шестакова (отв. ред.), А. С. Кулева (ред.), А. В. Гик. Москва:

Изд. Дом ЯСК, 2001–2019. (Studia philologica). Ин-т русского языка (ИРЯ) РАН.

19. Фрумкина Р. М. Статистическая структура лексики Пушкина // Вопросы языкознания. 1960. № 3. С. 78–81.

20. Частотный словарь языка М. Ю. Лермонтова / Под ред. В. В. Бородина, А. Я. Шайкевича; Сост. Авдеева А. А., Бородин В. В., Быкова Н. Я., Козокина С. М., Гордеева Н. А., Макарова Л. А., Шайкевич А. Я. // Лермонтовская энциклопедия / АН СССР. Ин-т рус. лит. (Пушкин. Дом); Науч.-ред. советизд-ва «Сов. энцикл.». Москва: Сов. энцикл., 1981. С. 717–774.

### **REFERENCES**

1. Abramson T. E. Digital methods of modern philology: opportunities and limitations//World literature through the eyes of modern youth. Digital age: a collection of materials of the IV international. youth scientific and practical conference, September 18–20, 2018 / scientific. ed. S. V. Rudakova. Magnitogorsk: Publishing House of Magnitogorsk. State Technician. un. G. I. Nosova, 2018. Pp. 9–11. 2.

2. Vasiliev N. L. "About Pushkin: the language of the classic, the poetics of the novel" Eugene Onegin, "the writer and his contemporaries": monograph/ Museum reserve A.S. Pushkin "Boldino"; Nizhny Novgorod University, Nizhny Novgorod, Saransk, 2013. 387 p.

3. Vasiliev N. L., Zhatkin D. N. Experience of comparing the poetic lexicons of A. S. Pushkin and M. Yu. Lermontov// Baltic Humanitarian Journal. 2017. Vol. 6. № 1 (18). Pp. 21–24.

4. Vasiliev N. L., Zhatkin D. N. Dictionaries of the poetic language of poets of the XIX century. Moscow: Flint; Science, 2013–2018.

5. Koubourlis D. J. A Concordance to the Poems of Osip Mandelstam. Ithaca NY: Cornell University Press, 1974 (dictionary of dictionary forms).

6. Lotman Y. M. On poets and poetry. Analysis of the poetic text. St. Petersburg, 1996.

7. Mityushin L. Concordance to the poems of Osip Mandelstam based on the collected works of OM in 4 vols. M.: Art Business Center, 1993–1997 (lexeme dictionary).

8. Margolis J. Quantitative analysis of dictionaries and transit of Russian-language poetry, abstracts of the report, University of Trier, 2.12.2020, URL: <https://bit.ly/2GEgxsK>.

9. A new version of the poetic corps of the Russian language in the national corps of the Russian language, see the NKR network, 2018. URL: <https://ruscorporu.ru/new/>.

10. Oerthel F., Tuschl S. Statistische Datenanalyse mit dem Programmpaket SAS. Universität Passau, R. Oldenburg Verlag, München, Wien, 1995.

11. Patera T. A Concordance to the Poetry of Anna Akhmatova. ARDIS, 339 pages, 1995.

12. Patera T. A Concordance to the Poetry of Joseph Brodsky. Books 1-6 (1962–1996), The Edwin Mellen Press, 2002 (Slavistik, McGill University, Montreal, Canada).

13. Polukhina V. Dictionary of the Color of Poetry of Joseph Brodsky. 2016.

14. Russian-language poetry in transit, research project of the University of Trier, Germany, hands. H. Stahl, 2017–2021.

15. Syntax dictionary of Russian poetry of the XIX century, [ed. N. V. Patroeva]. From Pushkin to Yazykov (ready for printing).

16. Syntax dictionary of Russian poetry of the 18th century: in 4 vols. / [Ed. N.V. Patroeva]. Volume 1, 2: Kantemir, Trediakovsky, Lomonosov. St. Petersburg: publishing house Dmitry Bulanin, 2018.

17. Pushkin language dictionary in 4 volumes, ed. ed. Acad. V. V. Vinogradov, 1956 – 1961 (1st ed.), 2000 (2nd ed.), Moscow, "Azbukovnik," 2000.

18. Dictionary of the language of Russian poetry of the 20th century. Volume I-VIII, comp: L. L. Shestakova (ed. ed.), A. S. Kuleva (ed.), A. V. Gik. M.: Publishing House of YASK, 2001–2019. (Studia philologica). In-t of the Russian language (IRA) RAS.

19. Frumkina R. M. "The statistical structure of Pushkin's vocabulary" // Questions of linguistics. 1960. № 3. P. 78–81.

20. Frequency dictionary of the language M. Yu. Lermontov / Ed. V. V. Borodin, A. Ya. Shaikevich; Sost. Avdeeva A. A., Borodin V. V., Bykova N. Ya., Kozokina S. M.

## QUANTITATIVE ANALYSIS OF DICTIONARIES AND TRANSIT OF RUSSIAN POETIC LANGUAGE

*Jakob M. Margolis*

Dr.-Ing.

(Trier University, Germany)

### **Abstract**

The goal of this research is the quantitative analysis of dictionaries and transit of the Russian poetic language in time from poet to poet and in space from country to country.

In a first step statistical analysis is applied to quantify the number of all tokens, single tokens and lexemes in the dictionaries of Russian poets of the 18th-20th centuries. From the «Dictionary of Russian poetic language in the 20th century» (Institute of the Russian Language of the Russian Academy of Science) and the dictionaries of poets of the 19th century dependencies between the size of vocabulary and number of all tokens are found.

For poets from the 18th and 19th centuries, based on a new version of the Russian poetic body, dependencies between the number of all tokens and the age of a poet are established.

It is shown how the transit of vocabulary from poet to poet can be observed quantitatively (statistically) and qualitatively (philologically).

In a second step for four groups of contemporary Russian language poets from different countries the works of the last years are collected and the transit of their language is sketched based on the foregoing results. These four groups have been chosen by an independent jury.

A comparative analysis of the language usage of 21th century Russian language poets living in Russia, Ukraine, Belarus, Latvia, Georgia, Kazakhstan and Uzbekistan is foreseen.

**Keywords:** quantitative analysis, dictionary, language, transit, lexicometry, statistical linguistics, vocabulary, all tokens, token, type, lexeme, modern Russian poets of the XXI century

**Для цитирования:** Марголис Я. М. Количественный анализ словарей и транзита русскоязычной поэзии // Libri Magistri. 2020. № 4 (14). С. 46–75.

*Поступила 08.11.2020*