Л.Г. Митюшин

Конкорданс к стихам Осипа Мандельштама

Так вот кому летать и петь
И слова пламенная ковкость, –
Чтоб прирожденную неловкость
Врожденным ритмом одолеть!

О. Мандельштам, "Автопортрет".

1. Введение

Предлагаемый конкорданс (указатель контекстов) отражает оригинальные стихотворения Осипа Мандельштама, содержащиеся в четырехтомном "Собрании сочинений" [1]. Его главное отличие от известного конкорданса Д. Кубурлиса [2] состоит в том, что у Кубурлиса входами являются словоформы, а в настоящей работе – лексемы.

В конкордансе представлены тексты из следующих разделов:

Стихотворения;
Стихи для детей;
Шуточные стихи;
Стихотворения (ранние редакции и варианты);
Строки из уничтоженных или утерянных стихотворений.

Учтены также "Дополнения", приведенные в т. 4 (стр. 205, 363–364). Совокупность этих текстов далее называется "полным корпусом"; под "основным корпусом" понимаются тексты из всех перечисленных разделов, кроме раздела "Ранние редакции и варианты". Тексты из раздела "Переводы" в конкорданс не включены.

В цитируемых контекстах сделаны исправления по списку в т. 4 (с. 597, "Замеченные опечатки и неточности"). Внесены и другие изменения по сравнению с печатным оригиналом: восстановлено несколько ошибочно замененных слов, ряд слов приведен к современной орфографии. Список этих изменений см. в Приложении 2. В словах, написанных латиницей, опущены диакритические знаки, так как они не представимы в используемой кодировке символов.

Информация представлена в виде 57 файлов, не считая данного. Она включает собственно конкорданс, а также алфавитные и частотные списки лексем. Далее следует подробное описание содержания этих массивов, порядка, в котором в них расположена информация, и процедур вычисления статистических данных.

2. Файлы

Все файлы, о которых говорится ниже, имеют тип txt с кодировкой Windows-1251, и расширение txt в их именах будет опускаться. Полный список файлов приведен в Приложении 1.

2.1. Конкорданс

Конкорданс содержится в 29 главных файлах Concord01_a, Concord02_b, … Concord32_ya, соответствующих буквам русского алфавита (кроме *ё, ъ, ы, ь*), и двух дополнительных файлах Concord33_nerus и Concord34_sluzh. Входами главных файлов являются слова, встречающиеся в текстах стихотворений и начинающиеся на соответствующую букву; для каждого слова приведены все его контексты – строки стихотворений, содержащие это слово. Под словом здесь понимается не отдельная словоформа (например, *мысль*), а лексема, т.е. словарная единица, относящаяся к определенной части речи и объединяющая всю совокупность соответствующих ей словоформ. В данном случае это будет лексема **мысль** (существительное), которой соответствуют словоформы *мысль, мысли, мыслью, мыслей* и т.д. Приведем фрагмент файла Concord13_m, содержащий контексты для лексемы **мысль**:

	_мысль 7 10 сущ.	
	Верные, четкие мысли –	1:41
	В сердце жар, и в мыслях пустота, –	1:63
	Мысли живой стрела?	1:86
	И мысль бесплотная в чертог теней вернется.	1:207 1:207б 1:207в
	Мыслям и чувствам моим по старинному праву?	3:22 3:22а
	Недобрым образом иль мыслей недобором,	3:146
	Мыслью пенится, сам себе снится, –	3:164 3:164г 3:164д 3:164е 3:164ж
*	Настроил мысль мою на величавый лад.	1:108а
*	Прямизна нашей мысли не только пугач для детей:	3:83а(7)
*	И европейской мысли разветвленье	3:(84-87)а(1) 3:(84-87)в(1) 3:(84-87)д

В правой колонке указаны номера стихотворений: перед двоеточием – номер тома, после двоеточия – номер стихотворения в томе. Имена лексем начинаются с символа подчеркивания. Для каждой лексемы приводятся два значения частоты, первое из них показывает число употреблений в основном, второе – в полном корпусе. Контексты из раздела "Ранние редакции и варианты" приведены последними и помечены знаком *.

Если для строки указано несколько номеров стихотворений, это означает, что во втором и следующих стихотворениях, всегда относящихся к разделу "Ранние редакции и варианты", данная строка повторяется без изменений. В таких случаях при подсчете частот слово учитывается только один раз (см. ниже раздел 5).

Файл Concord33_nerus содержит контексты для слов, написанных латиницей, и чисел в цифровой записи. В файл Concord34_sluzh вынесены из

главных файлов контексты для 12 служебных слов: **а**, **в**, **за**, **и**, **из**, **к**, **на**, **не**, **от**, **по**, **с**, **у**.

2.2. Алфавитные и частотные списки

По файлам конкорданса построены списки лексем, не содержащие контекстов. Они делятся на два вида: списки, упорядоченные по алфавиту и по частоте. Имеется 13 списков каждого вида: полный список лексем, собственные имена (т.е. пишущиеся с заглавной буквы), и списки отдельных частей речи – глаголы, существительные, прилагательные, наречия, предикативы, вводные слова, числительные, предлоги, союзы, частицы, междометия.

Каждая строка списка содержит имя лексемы и два значения частоты, как в файлах конкорданса. В полном списке лексем указывается также часть речи. В алфавитных списках порядок строк точно воспроизводит порядок лексем в конкордансе. В частотных списках лексемы упорядочены по убыванию первой частоты (число употреблений в основном корпусе), а при одинаковой первой частоте – по алфавиту. В частотных списках для каждой лексемы указывается ее ранг – номер в списке.

В конце каждого списка приводится статистика: число различных лексем рассматриваемой категории и число словоупотреблений в основном и полном корпусе. Сводная статистика по всем частям речи приведена в Приложении 3.

В Приложении 4 показано начало полного частотного списка лексем (первые 100 слов).

3. Лексемы

3.1. Имена лексем

Имя лексемы в главных файлах конкорданса и в файле Concord34_sluzh – это основная, словарная форма слова. В качестве лексемы может также выступать цепочка слов, функционирующая как одно слово: **всё ещё**, **едва ли**, **как будто**, **несмотря на**, **так себе** и т.п. (список таких лексем см. в Приложении 5).

В этих файлах в именах лексем, кроме заглавных и строчных русских букв, могут встречаться следующие символы:

пробел	**несмотря на**
дефис	**по-прежнему**
точка	**И.Т.Р.**
апостроф	**д'Оревильи**
знак ударения	**за'мок**

В именах лексем, в отличие от текстов стихотворений, буква *ё* не заменена на *е*.

В файле Concord33_nerus имена лексем – слова, написанные латинскими буквами (фактически это не лексемы, а словоформы), и числа в цифровой записи. Слова могут включать заглавные и строчные латинские буквы и апостроф, числа – десятичные цифры.

В имена лексем, представляющих инициалы и символы букв, включены соответствующие примечания в скобках: **А (буква)**, **Б. (инициал)**.

Знак ударения (апостроф после ударной гласной) используется только как средство различения лексем, принадлежащих к одной части речи: **а'тлас** – **атла'с**, **за'мок** – **замо'к**, **осы'паться** – **осыпа'ться**, **сре'зать** – **среза'ть** и т.п. Ударение всегда ставится, если оба конкурирующих варианта присутствуют в конкордансе. Если же представлен только один вариант, то ударение может быть отмечено в том случае, когда альтернативный вариант достаточно вероятен, например, в словах **ки'рка**, **крате'р**, **му'ка**, **наре'зать**, **пересыпа'ть**, чтобы сразу было ясно, что это не **кирка'**, **кра'тер**, **мука'**, **нареза'ть**, **пересы'пать**.

3.2. Части речи

В конкордансе используются следующие части речи:

ЧАСТЬ РЕЧИ	СОКРАЩЕНИЕ
глагол	–
существительное	сущ.
прилагательное	прил.
наречие	–
предикатив	предик.
вводное слово	вводн.
числительное	числит.
предлог	–
союз	–
частица	–
междометие	межд.
префикс	–
иноязычное слово	ин.
число	–

Распределение лексем по частям речи в значительной степени следует "Словарю русского языка в 4 тт." [3], с тенденцией к более эксплицитной дифференциации частей речи, свойственной словарю А.А. Зализняка [4]. Имеется в виду следующее: в четырехтомном словаре часто встречаются пометы "в значении существительного" (в статьях прилагательных), "в значении прилагательного" (в статьях причастий) и т.п. В словаре Зализняка таким значениям слов во многих случаях явным образом приписываются другие части речи (существительное, прилагательное и т.п.).

Отметим, что в данной работе местоимения не считаются самостоятельной частью речи; слова, традиционно относимые к местоимениям, рассматриваются как существительные (**я**, **ты**, **он**, **себя**, **кто**, **что**, **все**, **всё** и т.п.) и прилагательные (**мой**, **твой**, **свой**, **какой**, **такой**, **чей**, **тот**, **этот**, **весь** и т.п.). **Который** считается существительным (*листы, которые умрут*) или прилагательным (*который раз*).

Часть речи "префикс" приписана единственному слову **полу**, которое встречается четыре раза:

_полу 4 4 префикс
Полу-явь и полу-сон, [2 вх.] 1:67
Полу-город, полу-берег конный, [2 вх.] 3:167

Здесь **полу** считается отдельным словом по двум причинам: во-первых, оно пишется неслитно, и во-вторых, будучи эквивалентно наречию **наполовину**, оно может относиться к группе слов (в последнем употреблении – к сочетанию *берег конный*).

Всем словам, написанным латиницей, присвоена часть речи "иноязычное слово". К ней отнесены и некоторые слова, написанные русскими буквами, но воспринимаемые как запись иноязычного слова: **американ**, **дзенкую**, **пепайдевкос** и т.п.

Часть речи "число" приписана числам в цифровой записи, которые встречаются в заглавиях стихотворений, подзаголовках и примечаниях.

3.3. Состав лексем

В этом разделе описываются особенности принятого в конкордансе членения совокупности словоформ на наборы словоформ, соответствующие отдельным лексемам (парадигмы).

По аналогии со словарем Зализняка [4], глагольные видовые пары считаются разными лексемами. Исключение составляют случаи, когда основные формы членов пары совпадают (как у глаголов **ранить**, **сочетать** и т.п.).

Варианты написания служебных слов не считаются отдельными лексемами. Это относится как к вариантам, обусловленным сочетаемостью (*в – во, к – ко, о – об – обо*), так и к случаям произвольного варьирования, где один из вариантов, как правило, бывает стилистически окрашенным (*вокруг – вкруг, же – ж, или – иль, чтобы – чтоб*).

В случае знаменательных слов также возможно варьирование в рамках одной лексемы. Наиболее массовый пример – существительные среднего рода, оканчивающиеся на *ие* (**дыхание**, **пение**, **видение**), формы которых очень часто пишутся с мягким знаком (*дыханье, пенье, виденья*). С другой стороны, в таких словах, как **счастье**, **веселье**, **новоселье**, *ь* может заменяться на *и*: *счастия, веселием, новоселием*. Одной лексемой считаются варианты *более – боле, вслед*

– *вослед*, *заранее – заране*, *уже – уж* (наречия), *далеко – далёко* (существительное, наречие, предикатив), *что ли – что ль* (вводное слово).

Некоторые варианты не объединяются в одну лексему, например, **восьмой – осьмой**, **миллион – мильон**, **ложь – лжа**, **судья – судия**. В подобных случаях обычно даются взаимные ссылки – так, в строке лексемы **миллион** написано "[см. также _мильон]", и аналогично для **мильон**. Иногда в текстах встречается только более редкий вариант, например, **осьмигранный** (а **восьмигранный** отсутствует); тогда на месте, где должен был бы быть основной вариант, помещается ссылочная строка вида

_восьмигранный [см. _осьмигранный]

В конкордансе не различаются предлоги, управляющие разными падежами (**в** с винительным/предложным, **за** с винительным/творительным, **с** с родительным/винительным/творительным и т.п.).

Устойчивые словосочетания чаще всего интерпретируются как комбинации отдельных слов. Аргументами, оправдывающими такую интерпретацию, являются (а) возможность представить значение сочетания в виде суммы значений компонентов, и (б) возможность разрывного употребления сочетания. Например, *что за* (*Что за двоевластье там?*) рассматривается как сочетание существительного **что** и частицы **за**, поскольку возможно разрывное употребление: *Что это у тебя за книга?* Тем не менее, в конкордансе присутствуют лексемы, состоящие более чем из одного слова, но "эквивалентные слову" (их список приведен в Приложении 5).

Лексемы в конкордансе понимаются как слова с определенной основной формой, принадлежащие к определенной части речи. Систематическое различение значений внутри лексем не проводится. Однако в ряде случаев, когда совокупность употреблений лексемы в текстах стихотворений естественно разбивается на несколько четко ограниченных значений или пучков значений, ее контексты для удобства чтения делятся на соответствующие группы. Имя лексемы повторяется в начале каждой группы и сопровождается примечанием, характеризующим данное значение (пучок значений). Следует подчеркнуть, что такие примечания не должны рассматриваться как толкования, их задача – дифференцировать выделенные значения. Частоты лексемы приводятся для каждой группы в отдельности.

Приведем примеры для лексем **зреть** и **парк**:

_зреть 3 3 глагол [созревать]
Плод нарывал. Зрел виноград. 2:13 2:13в
И под музыку зреют плоды. 3:175
Зрела, маялась, шла через рвы. 3:179

_зреть 2 2 глагол [видеть]
Зренья нет - ты зришь в последний раз. 3:47
Мы сами видим, засыпая, зрим, 3:143 3:143а

_парк 5 6 сущ. [большой сад]
Казармы, парки и дворцы, 1:95 1:95а
Казармы, парки и дворцы... 1:95 1:95а
Направился к каштанам парка, 1:155
Огромный парк. Вокзала шар стеклянный. 2:2
От гремучего парка глазам? 3:181
* А в парке путаницы сито, 1:(122,123)а

_парк 2 2 сущ. [место стоянки транспорта]
Жили в парке два трамвая: 2:38
А Трам не хочет в парк, 2:38

В алфавитных и частотных списках разбиение на группы не учитывается, и приводятся суммарные частоты лексем по всем группам. Таким образом, там будет лексема **зреть** (глагол) с частотами 5, 5 и лексема **парк** (существительное) с частотами 7, 8.

3.4. Порядок лексем

Дадим точное описание порядка, в котором лексемы расположены в конкордансе. При упорядочении апострофы, точки и знаки ударения в именах лексем не учитываются (можно представлять себе, что они вычеркиваются). Имена упорядочиваются лексикографически со следующим отношением порядка между символами: (а) пробел; (б) дефис; (в) русские буквы по алфавиту без различения верхнего и нижнего регистра и букв *е* и *ё*; (г) латинские буквы по алфавиту без различения регистра; (д) цифры в порядке возрастания. Если отношение порядка между двумя словами остается неопределенным, т.е. если они различаются только регистром и буквами *е* и *ё*, то возможны два случая:

(1) Регистры некоторых букв не совпадают. Тогда рассматривается самое левое несовпадение, и более ранним считается слово, у которого на этом месте стоит строчная буква. В частности, нарицательные имена идут раньше соответствующих собственных (**август** < **Август**, **бах** < **Бах** и т.п.).

(2) Все регистры совпадают, и слова различаются только буквами *е* и *ё*. Тогда более ранним считается слово, у которого на месте самого левого несовпадения стоит *е* (**все** < **всё**).

Если имена двух лексем без учета апострофов, точек и знаков ударения полностью совпадают, но эти лексемы относятся к разным частям речи, то их порядок определяется порядком частей речи в списке, приведенном в разделе 3.2. Если у двух лексем совпадают имена и части речи, и в них по-разному поставлены знаки ударения, то более ранней считается лексема, у которой знак ударения стоит левее (**а'тлас** < **атла'с**, **за'мок** < **замо'к**).

Буква как элемент алфавита предшествует такой же букве с точкой, представляющей инициал (**А (буква)** < **А. (инициал)**).

Лексемы, состоящие из одних цифр, т.е. являющиеся десятичными числами, следуют после всех остальных лексем и упорядочиваются по возрастанию чисел.

4. Контексты

Как правило, в качестве контекста приводится одна строка стихотворения, содержащая форму рассматриваемой лексемы. Исключение составляют очень короткие строки: они могут дополняться одной или несколькими предыдущими или последующими строками, записанными через косую черту (/). Например, один из контекстов для лексемы **волна** выглядит так:

Греки сбондили Елену / По волнам, 3:27

Такое объединение делается достаточно редко, поэтому для простоты вместо "контекст" будем далее говорить "строка".

Справа от цитируемой строки помещается ссылка на содержащее эту строку стихотворение в виде номера тома и номера стихотворения в томе. Если приведенная строка целиком повторяется в стихотворениях из раздела "Ранние редакции и варианты", то после первой ссылки указываются номера этих стихотворений. Здесь повторение строки означает точное совпадение цепочки словоформ; совпадение знаков препинания не требуется.

Если некоторая строка встречается в стихотворении более одного раза, она приводится только один раз, но слева от номера стихотворения помещается примечание [2 ст.], [3 ст.] и т.п. В этом случае требуется точное совпадение строк, включая знаки препинания. Заметим, что если для строки указано более одного стихотворения, примечание [2 ст.] и т.п. относится только к первому из них; для остальных оно может не выполняться. Например, запись

Я больше не ревную, [2 ст.] 1:215 1:215а

указывает на то, что в стихотворении 215 из т. 1 строка "*Я больше не ревную,*" присутствует два раза; в то же время в другой редакции этого стихотворения 215а она встречается только один раз.

Если цитируемая строка содержит более одного вхождения рассматриваемой лексемы, она также приводится один раз, и слева от номера стихотворения помещается примечание [2 вх.], [3 вх.] и т.п. Возможно совмещение обоих типов примечаний – например, для лексемы **ни** (союз) приводится строка

Ни поволоки искусства, ни красок пространства веселого! [2 вх.] [2 ст.] 3:35

Строки – контексты данной лексемы упорядочены следующим образом. Сначала приводятся строки из основного корпуса, затем строки из раздела "Ранние редакции и варианты" (эти строки помечены знаком * в левой

колонке). В каждой из этих групп строки расположены в порядке возрастания номеров томов и стихотворений.

В поле номеров записываются пометы "заглавие", "эпиграф", "посвящение", "подзаголовок", "примечание", "заглавие цикла", "название книги", отражающие функциональную роль данной строки в стихотворении или группе стихотворений. Строки основного текста помет не имеют.

5. Частоты

Два значения частоты, указанные после имени каждой лексемы, равны суммарному числу употреблений лексемы в приведенных для нее строках (при этом вклады строк с примечанием [2 ст.] умножаются на 2, строк с примечанием [3 ст.] – на 3, и т.п.). Первое значение частоты равно числу употреблений в строках, не имеющих метки * (т.е. в основном корпусе), второе – числу употреблений во всех приведенных строках (т.е. в полном корпусе).

Эти две частоты принципиально отличаются друг от друга. Первая учитывает *все* вхождения лексемы в текстах основного корпуса; вторая дополнительно учитывает вхождения в текстах из раздела "Ранние редакции и варианты", однако только в таких строках, которые не являются точными копиями (без учета знаков препинания) строк стихотворений основного корпуса или строк предыдущих версий из раздела "Ранние редакции и варианты". Поэтому вторая частота в общем случае учитывает *не все* вхождения лексемы в текстах полного корпуса.

Повторяющеся строки исключены из рассмотрения по той причине, что стихотворения в разделе "Ранние редакции и варианты" и в основном корпусе имеют значительные совпадающие части. Например, стихотворение 209 из т. 1 *"Когда Психея-жизнь спускается к теням"* отличается от другой редакции 209а только последней, четвертой строфой; первые 12 строк полностью совпадают. Другой пример – стихотворения 64 и 64а из т. 3 *"Мы живем, под собою не чуя страны"*, где не совпадают только 3 строки из 16. Принятый способ подсчета приводит к тому, что значения частот в полном корпусе в большей степени отражают независимые, "оригинальные" употребления лексем.

Проводя эту идею до конца, следовало бы исключить из рассмотрения не только повторяющиеся строки, но и повторяющиеся части строк. Например, предпоследняя строка в упомянутом выше стихотворении 209 выглядит так: *"Дохнет на зеркало и медлит передать"*, а в 209а так: *"Дохнет на зеркало, – и медлит уплатить"*. Очевидно, все слова, кроме *уплатить*, не являются здесь "новыми" употреблениями; тем не менее, при принятом способе вычисления частот они учитываются как самостоятельные. Исключение при подсчете частот таких повторений потребовало бы более детального и содержательного анализа текстов стихотворений, что выходит за рамки настоящей работы.

Приложение 1. Список файлов.

ИМЯ ФАЙЛА	СОДЕРЖАНИЕ
_Description.doc	описание конкорданса
Concord01_a.txt	конкорданс для слов на *а*
Concord02_b.txt	. . . . . *б*
Concord03_v.txt	. . . . . *в*
Concord04_g.txt	. . . . . *г*
Concord05_d.txt	. . . . . *д*
Concord06_e.txt	. . . . . *е, ё*
Concord07_zh.txt	. . . . . *ж*
Concord08_z.txt	. . . . . *з*
Concord09_i.txt	. . . . . *и*
Concord10_j.txt	. . . . . *й*
Concord11_k.txt	. . . . . *к*
Concord12_l.txt	. . . . . *л*
Concord13_m.txt	. . . . . *м*
Concord14_n.txt	. . . . . *н*
Concord15_o.txt	. . . . . *о*
Concord16_p.txt	. . . . . *п*
Concord17_r.txt	. . . . . *р*
Concord18_s.txt	. . . . . *с*
Concord19_t.txt	. . . . . *т*
Concord20_u.txt	. . . . . *у*
Concord21_f.txt	. . . . . *ф*
Concord22_kh.txt	. . . . . *х*
Concord23_ts.txt	. . . . . *ц*
Concord24_ch.txt	. . . . . *ч*
Concord25_sh.txt	. . . . . *ш*
Concord26_shch.txt	. . . . . *щ*
Concord30_ae.txt	. . . . . *э*
Concord31_yu.txt	. . . . . *ю*
Concord32_ya.txt	. . . . . *я*
Concord33_nerus.txt	конкорданс для латиницы и чисел
Concord34_sluzh.txt	конкорданс для **а**, **в**, **за**, **и**, **из**, **к**, **на**, **не**, **от**, **по**, **с**, **у**

ListAlph00_total.txt	алфавитный список всех лексем
ListAlph01_glagol.txt	. . . . . глаголов
ListAlph02_sushch.txt	. . . . . существительных
ListAlph03_pril.txt	. . . . . прилагательных
ListAlph04_narechie.txt	. . . . . наречий
ListAlph05_predik.txt	. . . . . предикативов
ListAlph06_vvodn.txt	. . . . . вводных слов
ListAlph07_chislit.txt	. . . . . числительных
ListAlph08_predlog.txt	. . . . . предлогов
ListAlph09_soyuz.txt	. . . . . союзов
ListAlph10_chastitsa.txt	. . . . . частиц
ListAlph11_mezhd.txt	. . . . . междометий
ListAlph12_sobst.txt	. . . . . собственных имен

ListFreq00_total.txt	частотный список всех лексем
ListFreq01_glagol.txt	. . . . . глаголов
ListFreq02_sushch.txt	. . . . . существительных
ListFreq03_pril.txt	. . . . . прилагательных
ListFreq04_narechie.txt	. . . . . наречий
ListFreq05_predik.txt	. . . . . предикативов
ListFreq06_vvodn.txt	. . . . . вводных слов
ListFreq07_chislit.txt	. . . . . числительных
ListFreq08_predlog.txt	. . . . . предлогов
ListFreq09_soyuz.txt	. . . . . союзов
ListFreq10_chastitsa.txt	. . . . . частиц
ListFreq11_mezhd.txt	. . . . . междометий
ListFreq12_sobst.txt	. . . . . собственных имен

Приложение 2. Изменения в текстах стихотворений.

(2а) Восстановление ошибочно замененных слов.

Основанием для восстановления является сравнение с изданиями [5–8].

НАПЕЧАТАНО В "СОБРАНИИ СОЧИНЕНИЙ"		ВОССТАНОВЛЕНО
И в лицо нам **берег** влажный Бьет соленым покрывалом.	(стих. 1:55)	берег → ветер
Чернеет на **скале** гранитной Самоубийца молодой.	(стих. 1:116)	скале → скамье
Кинематограф. Три скамейки. **Сентиментальная** горячка.	(стих. 1:119)	Сентиментальная → Сантиментальная
Воздух твой граненый. В **спальнях** тают горы Голубого дряхлого стекла.	(стих. 1:206)	спальнях → спальне
И **как** хорошо мне и тяжко, Когда приближается миг,	(стих. 3:68-78(2))	как → так

(2б) Приведение к современной орфографии.

НАПИСАНИЕ В "СОБРАНИИ СОЧИНЕНИЙ"	НОВОЕ НАПИСАНИЕ
Верлэн	Верлен
лэди	леди
пэан	пеан
спортсмэн	спортсмен
чорт	черт
чортов	чертов
шопот	шепот
шопотом	шепотом
шопотком	шепотком
шопоть	шепоть
Ян-цзы	Янцзы

Приложение 3. Статистические данные.

ЧАСТЬ РЕЧИ	РАЗЛИЧНЫХ СЛОВ		СЛОВ В ТЕКСТАХ	
	в основном корпусе	в полном корпусе	в основном корпусе	в полном корпусе
глагол	2204	2333	5834	6463
существительное	4489	4658	15173	16863
прилагательное	1853	1941	5901	6548
наречие	446	461	1610	1772
предикатив	76	81	222	255
вводное слово	21	21	49	52
числительное	46	46	220	236
предлог	54	54	3582	3945
союз	47	48	3260	3627
частица	46	46	1167	1285
междометие	23	23	108	119
префикс (_полу)	1	1	4	4
иноязычное слово	83	84	100	101
число	8	9	8	13
все части речи	9397	9806	37238	41283

ПРИМЕЧАНИЯ.

1. При подсчете числа словоупотреблений в полном корпусе (4-й столбец) не учитываются строки, совпадающие со строками стихотворений основного корпуса или строками предыдущих версий из раздела "Ранние редакции и варианты" (см. раздел 5 основного текста).

2. Из-за неоднословных лексем число графических слов в основном корпусе может несколько превышать цифры в 3-м столбце. В частности, общее число графических слов в основном корпусе равно 37327. (Слова с дефисом считаются одним графическим словом, если в конкордансе им соответствует одна лексема.)

Приложение 4. Начало полного частотного списка лексем.

1	и	1935	2161	союз
2	в	1236	1362	предлог
3	я	813	903	сущ.
4	на	643	709	предлог
5	не	619	682	частица
6	с	406	431	предлог
7	как	358	392	союз
8	он	310	336	сущ.
9	ты	242	265	сущ.
10	быть	215	229	глагол
11	мы	215	225	сущ.
12	а	193	208	союз
13	мой	158	171	прил.
14	что	152	170	сущ.
15	она	143	153	сущ.
16	к	140	152	предлог
17	но	131	146	союз
18	весь	126	138	прил.
19	за	125	135	предлог
20	они	121	141	сущ.
21	из	119	138	предлог
22	от	119	133	предлог
23	у	113	124	предлог
24	по	106	118	предлог
25	где	103	118	наречие
26	свой	100	115	прил.
27	твой	100	110	прил.
28	кто	96	102	сущ.
29	ещё	89	104	наречие
30	для	85	92	предлог
31	этот	84	89	прил.
32	чёрный	82	96	прил.
33	что	82	90	союз
34	как	81	85	наречие
35	вы	80	86	сущ.
36	вода	77	83	сущ.
37	когда	77	88	союз
38	или	73	85	союз
39	земля	72	90	сущ.
40	же	71	80	частица
41	так	71	79	наречие
42	бы	70	75	частица
43	о	70	86	предлог
44	чтобы	70	77	союз
45	небо	69	77	сущ.
46	до	68	73	предлог
47	ночь	68	82	сущ.
48	рука	68	69	сущ.
49	воздух	67	75	сущ.
50	то	63	70	сущ.
51	без	62	76	предлог

52	о	62	70	межд.
53	под	62	68	предлог
54	над	60	62	предлог
55	сказать	60	63	глагол
56	человек	60	63	сущ.
57	себя	58	62	сущ.
58	жизнь	57	66	сущ.
59	мир	57	63	сущ.
60	любить	56	56	глагол
61	петь	56	60	глагол
62	жить	55	65	глагол
63	день	54	58	сущ.
64	глаз	53	57	сущ.
65	идти	52	60	глагол
66	голова	51	54	сущ.
67	звезда	51	57	сущ.
68	там	51	61	наречие
69	всё	50	56	сущ.
70	хотеть	49	54	глагол
71	век	48	50	сущ.
72	говорить	47	52	глагол
73	тот	47	60	прил.
74	это	47	54	сущ.
75	губа	46	49	сущ.
76	ли	46	50	частица
77	ни	46	56	союз
78	слово	46	50	сущ.
79	белый	45	50	прил.
80	душа	45	53	сущ.
81	мочь	45	48	глагол
82	один	45	49	прил.
83	только	45	54	частица
84	лёгкий	44	47	прил.
85	стоять	44	55	глагол
86	здесь	43	46	наречие
87	если	42	49	союз
88	кровь	41	44	сущ.
89	лес	41	45	сущ.
90	солнце	41	48	сущ.
91	дом	40	45	сущ.
92	какой	40	47	прил.
93	нежный	40	48	прил.
94	сам	40	46	прил.
95	сон	40	44	сущ.
96	два	39	42	числит.
97	сердце	39	46	сущ.
98	такой	39	42	прил.
99	тёмный	39	43	прил.
100	улица	39	42	сущ.

Приложение 5. Неоднословные лексемы в конкордансе.

ЛЕКСЕМА	ЧАСТЬ РЕЧИ
а не то	союз
а то	союз
быть может	вводн.
вот ещё	межд.
вряд ли	наречие
всё ещё	наречие
всё же	частица
всё равно	наречие
всё равно	предик.
должно быть	вводн.
едва ли	наречие
как будто	союз
как будто	частица
как бы	союз
как бы	частица
куда как	наречие
между тем	наречие
Микель Анджело	сущ. (принятое в 1930-е гг. написание имени *Микеланджело*)
может быть	вводн.
несмотря на	предлог
несмотря что	союз
ой ли	межд.
потому что	союз
почти что	наречие
так себе	предик.
так сказать	вводн.
то бишь	союз
то есть	союз
только что	наречие
хотя бы	союз
хотя бы	частица
что ли	вводн.

Литература

1. О. Мандельштам. Собрание сочинений в 4 тт. М.: Арт-Бизнес-Центр, 1993–97. Сост. П. Нерлер, А. Никитаев (тт. 1–3); П. Нерлер, А. Никитаев, Ю. Фрейдин, С. Василенко (т. 4).

2. D.J. Koubourlis, ed. A Concordance to the Poems of Osip Mandelstam. Ithaca NY: Cornell University Press, 1974.

3. Словарь русского языка в 4 тт. Под ред. А.П. Евгеньевой. М.: Русский язык, 1985–88.

4. А.А. Зализняк. Грамматический словарь русского языка: Словоизменение. М.: Русский язык, 1980.

5. О. Мандельштам. Сочинения в 2 тт. М.: Художественная литература, 1990. Сост. П. Нерлер (т. 1); С. Аверинцев, П. Нерлер (т. 2).

6. О. Мандельштам. Полное собрание стихотворений. СПб.: Академический проект, 1995. Сост. А. Мец.

7. О. Мандельштам. Камень. Л.: Наука, 1990. Изд. подг. Л. Гинзбург, А. Мец, С. Василенко, Ю. Фрейдин.

8. О. Мандельштам. Стихотворения. Л.: Советский писатель, 1973. Сост. Н. Харджиев.