



А. Я. ШАЙКЕВИЧ,  
В. М. АНДРЮЩЕНКО, Н. А. РЕБЕЦКАЯ

СТАТИСТИЧЕСКИЙ  
СЛОВАРЬ  
ЯЗЫКА ДОСТОЕВСКОГО



STUDIA PHILOGICA

**Рецензенты:**

доктор филол. наук *В. П. Григорьев*,  
доктор филол. наук *Н. Н. Перцова*

**А. Я. Шайкевич, В. М. Андрущенко, Н. А. Ребецкая**

Ш 12      Статистический словарь языка Достоевского / Рос. акад. наук. Ин-т русского языка им. В. В. Виноградова. – М.: Языки славянской культуры, 2003. – 880 с., разд. паг. (XLVIII, 832 с.). — (Studia philologica).

ISSN 1726-135X  
ISBN 5-94457-145-4

«Статистический словарь языка Достоевского» включает всю лексику трех основных жанров писателя — художественной литературы, публицистики и писем (более 43 тысяч разных слов). Словарь построен на корпусе текстов в 2,9 млн. словоупотреблений и значительно превосходит по объему любые другие частотные словари русского языка. По степени лингвистической дифференциации Словарь уникален и в мировом масштабе. В таблицах Словаря лексика Ф. М. Достоевского представлена в распределении по основным жанрам и по периодам творчества. Словарь адресован филологам и всем любителям творчества Ф. М. Достоевского.

**ББК 83**

*Анатолий Янович Шайкевич*

*Владислав Митрофанович Андрущенко*

*Наталья Александровна Ребецкая*

**СТАТИСТИЧЕСКИЙ СЛОВАРЬ ЯЗЫКА ДОСТОЕВСКОГО**

Издатель А. Кошелев

Корректор В. В. Мачкова

Художник-консультант Л. М. Панфилова

Подписано в печать 16.08.2003. Формат 84×108 1/16.

Бумага офсетная № 1, печать офсетная

Усл. печ. л. 89,1. Заказ № 1998. Тираж 550 экз.

Издательство «Языки славянской культуры»

ЛР № 02745 от 04.10.2000.

Тел.: 207-86-93. Факс: (095) 246-20-20 (для аб. М153).

E-mail: lrc-kozlov@mmi-net.ru

Каталог в ИНТЕРНЕТ <http://www.lrc-press.ru>; <http://www.lrc-mik.narod.ru>

ГУП Московская типография № 2  
Министерства Российской Федерации  
по делам печати, телерадиовещания  
и средств массовых коммуникаций (МПТР России).  
Тел.: 282-24-91. 129085, Москва, пр. Мира, 105.

ISBN 5-94457-145-4



9 785944 571458 >

## Оглавление

Введение

VI-XLVIII

### Таблицы

|   |     |
|---|-----|
| 1. Распределение лексем по основным жанрам                                  | 1   |
| 2. Относительная частота лексем в основных жанрах                           | 479 |
| 3. 100 самых частых лемм в текстах Достоевского                             | 515 |
| 4. 100 самых частых лемм в художественных произведениях Достоевского        | 516 |
| 5. 100 самых частых лемм в критике и публицистике Достоевского              | 517 |
| 6. 100 самых частых лемм в письмах Достоевского                             | 518 |
| 7. 40 самых частых существительных в основных жанрах                        | 519 |
| 8. 40 самых частых глаголов в основных жанрах                               | 520 |
| 9. 40 самых частых полных прилагательных в основных жанрах                  | 521 |
| 10. Частотный спектр рангового словаря лемм всего корпуса текстов           | 522 |
| 11. Частотный спектр рангового словаря лемм беллетристики                   | 523 |
| 12. Частотный спектр рангового словаря лемм критики и публицистики          | 524 |
| 13. Частотный спектр рангового словаря лемм писем                           | 525 |
| 14. Лексические маркеры беллетристики                                       | 526 |
| 15. Лексические маркеры критики и публицистики                              | 537 |
| 16. Лексические маркеры писем   | 550 |
| 17. Лексические маркеры беллетристики с максимальными значениями S          | 561 |
| 18. Лексические маркеры критики и публицистики с максимальными значениями S | 562 |
| 19. Лексические маркеры писем с максимальными значениями S                  | 563 |
| 20. Грамматические классы лемм и аффиксы                                    | 564 |
| 21. Относительная частота грамматических классов лемм и аффиксов            | 572 |
| 22. Распределение лексем в беллетристике по периодам творчества             | 580 |
| 23. Распределение лексем критики и публицистики по периодам творчества      | 684 |
| 24. Распределение лексики писем по периодам жизни Достоевского              | 757 |

## Введение

Настоящий Словарь подготовлен в Отделе машинного фонда Института русского языка РАН. Работа над Словарем начиналась в рамках проекта «Словарь языка Достоевского», (руководитель – Ю. Н. Караулов), поддержанного РГНФ, а затем выделилась в самостоятельное направление. На этом этапе авторы использовали также финансовую поддержку РГНФ, оказанную более широкому проекту «Дистрибутивно-статистическое описание языка русской прозы 1855–1880 гг.» (01-04-00247а).

Следует с самого начала подчеркнуть, что цели обоих словарей не совпадают. Цель «Словаря языка Достоевского» – показать лексику писателя во всем ее богатстве (с детальной семантической разработкой, с собранием иллюстративных примеров, с исчерпывающим словоуказателем и т. п.). Итогом является лексикографическая серия, намного превосходящая по объему лучшие образцы авторской лексикографии, такие как первый опыт на русской почве – «Словарь языка Пушкина» [Словарь Пушкина] или замечательный «Словарь языка Мицкевича» [Słownik]. Первый выпуск этой серии уже вышел в свет [Словарь Достоевского].

Задача «Статистического словаря языка Достоевского» скромнее, он должен представить лексику Достоевского в количественном виде, повторив и обогатив опыт уникального конкорданса к Шекспиру [Sprevack]. Однако и при таком ограничении результат оказался бы слишком объемным для бумажного издания (речь идет о многих сотнях авторских листов), а потому было принято решение издать Словарь в гибридном виде – как однотомную книгу, показывающую лишь часть таблиц, и как сопровождающий ее компакт-диск, содержащий информацию в полном объеме. Конечно, в первом опыте такого рода нас подстерегают многие технические трудности издания, а также психологические предубеждения читателей, но именно на этом пути нам видится дальнейший прогресс академической лексикографии.

Предваряя описание структуры Словаря, выскажем одно замечание относительно развития статистической лексикографии. В 1960–1970-х гг. наблюдалось широко распространенное увлечение частотными словарями, особенно в связи с педагогическими и информационными приложениями. От очень скромных по объему изданий (400 тыс. словоупотреблений) лексикография шагнула к рубежу в 1 млн. словоупотреблений, а затем и к новым рекордам – максимальные дифференцированный словарь американских текстов для школы содержит более 5 млн. словоупотреблений [Carroll], а словарь, созданный Институтом французского языка [Dictionnaire], опирается на корпус литературных текстов объемом более 70 млн. словоупотреблений. Затем наступает кризис: электронные корпуса текстов продолжают множиться и увеличиваться по объему (в некоторых из них счет идет уже на сотни миллионов словоупотреблений), но не видно новых частотных словарей, которые были бы созданы на основе этих корпусов. В чем же дело? Причин может быть много, назовем некоторые из них.

1) Программными средствами можно легко и просто получить статистику графических слов. Именно такая информация представлена в вышеупомянутом словаре Керрола [Carroll]. Но читателю обычно нужно большее – графические слова должны быть сведены в осмысленные лингвистические единицы, они должны быть лемматизированы. Процесс же лемматизации не поддается алгоритмам на сто процентов. Доля ручного вмешательства хотя и уменьшается относительно, но продолжает расти абсолютно. При росте объема текстового корпуса в 100 раз объем ручного труда при постредактировании возрастет, скажем, в 10 раз.

2) До сих пор не разработаны хорошие автоматизированные процедуры формирования выборки на большом корпусе текстов. Впрочем, эта трудность не

существует при обработке замкнутого корпуса целиком (как, например, в случае текстов Достоевского).

3) Наконец, существует и психологический фактор. Лингвостатистика, как она складывалась в середине XX в., в какой-то степени была во власти математического фетишизма: открытие «закона» Ципфа создавало иллюзию новой области статистических исследований, возникла новая дисциплина, все более терявшая связи с лингвистикой, филологией, информатикой.

Предлагаемый Словарь должен сделать шаг в обратном направлении.

### 1. Корпус текстов Достоевского и его членение

Настоящий Словарь опирается на 30-томное академическое издание Ф. М. Достоевского и в основном следует принципам классификации текстов, принятым в этом издании, т. е. включает три основных жанра — «Художественная литература», «Критика и публицистика» и «Письма». Эти три жанра в совокупности и составляют корпус текстов Достоевского, послуживший базой для всех статистических таблиц «Статистического словаря языка Достоевского». Общий объем корпуса — 2889 тыс. графических слов<sup>1</sup> (145980 разных графических слов), в том числе: «Художественная литература» — 1835 тыс. слов (110744 разных графических слова), «Критика и публицистика» — 524 тыс. слов (59446 разных графических слов), «Письма» — 531 тыс. слов (43689 разных графических слов). Не вошли в наш корпус текстов ранние редакции и варианты, подготовительные материалы и тексты записных книжек. Применение статистических методов к подобным текстам было бы почти невозможным. Не вошли в корпус и деловые бумаги, где индивидуальность автора почти не проявляется. Разумеется, эти группы текстов должны учитываться при составлении исчерпывающего словника Достоевского.

Ряд текстов из «Дневника писателя» отнесен к художественной литературе: «Бобок», «Кроткая», «Мальчик у Христа на елке», «Мужик Марей», «Сон смешного человека», «Столетняя».

### 2. Лингвистические единицы, отраженные в статистических таблицах

В настоящем Словаре представлены как исходные графические слова (только в электронной части Словаря), так и результаты всевозможных процедур над графическими словами (слияние разных грамматических форм слова, слияние вариантов, расщепление, объединение в одну единицу двух и более графических слов, следующих друг за другом). Прежде всего речь идет об орфографических вариантах (*адрес* и *адресс*, *прощание* и *прощанье*), в которых могли проявляться орфографические нормы времени или пристрастия издателей. Подобные варианты объединяются в одну единицу. С другой стороны, сохранена статистическая информация о таких вариантах, как *бриллиант* и *брильянт*, *Авдотья Сергеевна* и *Авдотья Сергевна*, *вести* и *весть*.

Некоторые графические слова разделяются на две или даже три леммы. Речь идет о частицах вроде *-де*, *-ка*, *-с*, *-таки*, *-то*. Однако сохраняются нерасчлененными слова с «неопределенным» *-то*, присоединяемым к основам вопросительных (и некоторых указательных) местоимений (*где-то*, *какой-то*, *откуда-то*, *такой-то*).

Что касается грамматических форм изменяемых слов, то здесь доминирует традиционное представление о частях речи (например, графические слова на *-о*, вроде *абсурдно*, *бездарно*, *безобразно*, *вековечно*, расщепляются на наречия и прилагательные). Однако, вслед за Словарем Пушкина, компаративы сохраняются как отдельные грамматические единицы (при этом формы на *-ее* и *-ей* сливаются воедино). Отдельно фигурируют и суперлативы.

<sup>1</sup> Термин «графическое слово» представляется более правильным, чем общепринятый термин «словоформа». Один раз встретившееся у Достоевского слово *взяточка-то-с* заслуживает названия «графическое слово», но вряд ли будет идентифицировано лингвистами как особая «словоформа». Точно так же графическое слово *ви-но-ват*, встретившееся три раза, едва ли кем-либо будет объявлено особой словоформой. С другой стороны, встретившаяся последовательность *по...за...буду...* («Белые ночи») в словаре графических слов будет отражена как три слова, в словаре лемм — прибавит единицу к частоте слова *позабыть*.

Последовательное системное разделение грамматических форм по частям речи в зависимости от синтаксической функции было бы слишком трудоемким. Здесь был принят неформальный принцип: если есть основания предполагать, что синтаксические функции форм отягощены еще и семантическими или стилистическими различиями, они должны быть разведены в статистических таблицах. Сохранена статистическая информация о формах числа многих существительных (*око и очи, ухо и уши, вода и воды, брат и братья* и т. п.), о формах императива многих глаголов (*не беспокойтесь, ступай* и т. п.). Часто различаются субстантивированные прилагательные (*артельный, блаженный, ближний, больной, большие, бывшее, дворовый, знакомый, рабочий* и т. п.) и омонимичные исходные прилагательные. Выделены и многие адъективированные причастия, например *благоухающий, верующий, воинствующий, волнующий, заплывший, исхудавший, минувший, обрусевший* и т. п. В максимальной степени грамматическая информация дана для глагола *быть*, для которого указана совокупная частота форм прошедшего времени (представлена формой *был*) и форм будущего времени (представлена формой *будет*).

Выше упомянутый неформальный принцип распространяется и на случаи семантического расщепления, и на случаи объединения последовательностей слов в особые единицы. Так разведены: *а* (союз), *а* (междометие), *а* (вопросительное слово) и *а* (буква); *батюшка* (отец), *батюшка* (обращение), *батюшка* (священник); *благо* (сущ.) и *благо* (союз); *брак* (супружество) и *брак* (дефект) и т. п. Сохранена статистическая информация о словах вроде *акт* (церемония), *банк* (игра), *брат* (обращение), *будет* (достаточно) и т. п.

Довольно часто в статистических таблицах даются сочетания слов, например: *так и быть, была не была, как есть, так и есть, что ни есть, все равно, прежде всего, вещественные доказательства, порядок вещей, в порядке вещей, взад да вперед, взад и вперед, на взгляд, на первый взгляд, по первому взгляду, с первого взгляда, быть молодцу не укор*. Отдельно показаны все имена собственные, в том числе имена с отчествами.

Любую строчку в статистических таблицах настоящего Словаря будем называть **лексемой**. Следовательно, такие строки, как *порядок вещей, в порядке вещей, Аглая, Аглая Ивановна, Адрианополь (город), «Адрианополь» (гостиница), благо (сущ.), благо (союз)*, — все это лексемы. Те лексемы, чьи частоты не входят в частоту других лексем, будем называть **леммами**. Лексемы, не являющиеся леммами, печатаются в таблицах с отступом. Обращаясь, например, к таблице 1, мы найдем там:

|              | Всего | X    | K   | П   |
|--------------|-------|------|-----|-----|
| Бог          | 1730  | 1081 | 173 | 476 |
| Бог в помощь | 1     | 1    |     |     |
| Бог ведаёт   | 2     | 1    |     | 1   |
| Бог весть    | 1     | 1    |     |     |
| Бог знает    | 322   | 207  | 34  | 81  |
| «Бог»        | 2     |      | 2   |     |
| боги         | 36    | 25   | 9   | 2   |
| ради Бога    | 459   | 137  | 2   | 320 |
| слава Богу   | 137   | 90   | 11  | 36  |

Частоты сочетаний *Бог в помощь, Бог ведаёт, Бог весть, Бог знает* уже учтены в строке *Бог*, но в ней не учтены частоты лемм *«Бог»*, *боги*, *ради Бога*, *слава Богу*. Если читатель не согласится с такой лемматизацией и захочет получить частоту слова *Бог* в рамках лексикографической традиции, он сможет суммировать частоты этих четырех лемм и получит строку:

|     |      |      |     |     |
|-----|------|------|-----|-----|
| Бог | 2364 | 1333 | 197 | 834 |
|-----|------|------|-----|-----|

Если же, напротив, читатель захочет повысить статус словосочетания *Бог знает*, превратив его в отдельную лемму, ему надо вычесть частоту словосочетания из частоты леммы, получая для леммы *Бог* строку:

|     |      |     |     |     |
|-----|------|-----|-----|-----|
| Бог | 1408 | 874 | 139 | 395 |
|-----|------|-----|-----|-----|

Таким образом, различие лексем (вообще) и лемм (в частности) не принципиально – при любом решении статистическая информация сохранена.

### 3. Типы статистических таблиц, представленных в Словаре

Преобладающий тип статистической таблицы (примером может служить таблица 1) содержит текстовую часть, включающую лингвистические объекты: графические слова, лексемы (как в таблице 001)<sup>2</sup>, леммы (как в таблице 003), и цифровую часть, состоящую из одного или нескольких столбцов (в таблице 1 – четыре столбца). Как правило, строки лемм в таблице упорядочены по обычному алфавитному принципу.

Исключениями являются обратные частотные словари, в которых единицы упорядочены по алфавиту, как если бы они читались справа налево. Примером может служить таблица 001.

Таблица 001

#### Фрагмент обратного частотного словаря графических слов

|     |            |     |             |    |             |     |           |
|-----|------------|-----|-------------|----|-------------|-----|-----------|
| 52  | ба         | 5   | пошиба      | 7  | проба       | 53  | свадьба   |
| 2   | б-ба       | 19  | лба         | 1  | утроба      | 2   | усадьба   |
| 111 | баба       | 6   | столба      | 42 | особа       | 5   | ходьба    |
| 1   | бой-баба   | 1   | Памба       | 8  | способа     | 181 | судьба    |
| 24  | слаба      | 1   | дифирамба   | 3  | герба       | 1   | ворона-   |
| 36  | раба       | 7   | бомба       | 1  | серба       |     | судьба    |
| 4   | штаба      | 4   | Комба       | 4  | ущерба      | 1   | похвальба |
| 4   | деба       | 1   | апломба     | 2  | корба       | 1   | стрельба  |
| 154 | хлеба      | 2   | Колумба     | 9  | губа        | 1   | мольба    |
| 57  | неба       | 1   | тумба       | 1  | пагуба      | 2   | гульба    |
| 7   | погреба    | 560 | оба         | 1  | Соллогуба   | 1   | гоньба    |
| 4   | Феба       | 2   | худоба      | 2  | Гекуба      | 49  | борьба    |
| 37  | служба     | 8   | жалоба      | 11 | клуба       | 138 | просьба   |
| 30  | дружба     | 52  | злоба       | 3  | груба       | 1   | письмо-   |
| 5   | тяжба      | 2   | озноба      | 2  | сруба       |     | просьба   |
| 13  | изба       | 29  | гроба       | 1  | труба       | 9   | женитьба  |
| 2   | биба       | 1   | гардероба   | 10 | шуба        | 31  | люба      |
| 1   | Скриба     | 5   | короба      | 20 | рыба        |     |           |
|     |            |     | ....        |    |             |     |           |
| 73  | бить       | 7   | избить      | 1  | дрррробить  | 1   | нарубить  |
| 2   | забить     | 8   | прибить     | 1  | пособить    | 8   | грубить   |
| 4   | ослабить   | 1   | зашибить    | 52 | оскорбить   | 1   | нагрубить |
| 1   | набить     | 1   | пришибить   | 15 | сбить       | 1   | сгрубить  |
| 8   | грабить    | 1   | ошибить     | 8  | отбить      | 1   | срубить   |
| 1   | заграбить  | 2   | долбить     | 97 | убить       | 3   | трубить   |
| 17  | ограбить   | 1   | добить      | 20 | губить      | 2   | выбить    |
| 1   | пограбить  | 1   | разжалобить | 5  | загубить    | 387 | любить    |
| 1   | вбить      | 2   | озлобить    | 39 | погубить    | 2   | залюбить  |
| 1   | подбить    | 1   | знобить     | 5  | стгубить    | 1   | влюбить   |
| 8   | перебить   | 2   | побить      | 1  | усугубить   | 5   | разлюбить |
| 4   | теребить   | 5   | дробить     | 1  | голубить    | 4   | возлюбить |
| 62  | употребить | 3   | раздробить  | 1  | приголубить | 2   | долюбить  |
| 25  | истребить  | 2   | коробить    | 6  | рубить      | 67  | полюбить  |
| 23  | разбить    | 4   | пробить     | 1  | зарубить    |     |           |

Не приходится разъяснять эвристическую ценность обратного словаря. Наряду с обратным словарем графических слов в электронной части Словаря представлен и обратный словарь лемм (43577 разных лемм). Весьма объемные обратные словари отражают весь корпус текстов; что касается отдельных частей корпуса, то для них будут даны сведения о частоте словообразовательных

<sup>2</sup> Все таблицы «Введения» нумеруются с начальным нулем.

элементов (префиксов и суффиксов). Во всех остальных таблицах принят прямой алфавитный порядок.

В существующих частотных словарях до половины общего объема приходится на ранговые словари, т. е. на таблицы, в которых единицы расположены в порядке уменьшения их частоты ( $f$ ) и соответствующего возрастания их ранга ( $r$ )<sup>3</sup>. Примером может служить таблица 3.

В описываемом Словаре ранговые словари занимают очень скромное место — в электронной части Словаря даются четыре списка по 500 самых частых графических слов для всего корпуса текстов, для совокупности художественных текстов, для публицистики и для писем; аналогичным образом будут включены таблицы слов (лемм). Такое решение объясняется просто: ранговыми словарями практически нельзя пользоваться. В них можно ответить на такие экзотические вопросы, как «какие именно слова имеют частоту 15?» или «какое слово занимает 305-е место в ранговом словаре?», но нельзя найти конкретные слова средней и низкой частоты. Если же читателю все-таки понадобится перейти от частоты к соответствующему рангу, это можно будет сделать при помощи таблицы, уместящейся на одной-двух страницах (см. таблицу 10). Структура этих кратких таблиц описана в следующем параграфе.

Все примеры таблиц, представленные до сих пор, содержат абсолютные частоты лингвистических единиц. Их преимущество — представление полного объема информации, их недостаток — сложность непосредственного сравнения данных, входящих в разные столбцы. Как правило, столбцы отражают данные разных подкорпусов, каждый из которых не совпадает по объему с другими. Например, в таблице 1 общий объем «Критики и писем» примерно совпадает, но «Художественная литература» превышает их в три с половиной раза. Конечно, рассматривая строки с небольшой совокупной частотой, читатель мысленно учтет это обстоятельство и сделает правильный вывод. Вот три примера из таблицы 1:

|          | Всего | X  | K  | П |
|----------|-------|----|----|---|
| бритва   | 36    | 23 | 12 | 1 |
| брошюра  | 25    | 3  | 13 | 9 |
| брюнетка | 15    | 14 |    | 1 |

Без каких бы то ни было сложных вычислений читателю ясно, что слово *бритва* сосредоточено в критике, что слово *брошюра* крайне редко появляется в художественной литературе, а слово *брюнетка* именно в этом жанре и сосредоточено. Если же совокупная частота велика, то требуется проводить некоторые арифметические операции, что без калькулятора делать трудно.

Разрешить данную трудность можно при помощи таблиц относительных частот, где частоты приведены к общему знаменателю (скажем, на 100 тыс. словоупотреблений). Именно этот принцип характеризует таблицу 2.<sup>4</sup>

Представление результатов в виде относительных частот имеет одно ограничение — оно бессмысленно в приложении к редким явлениям. В связи с этим в Словаре вводится еще и специальная мера оценки статистической значимости реальных частот:

$$S = (f - m - 1) / \sqrt{m},$$

где  $f$  — наблюдаемая частота данного события,

$m$  — математическое ожидание этого события, подсчитанное на основе какой-то нулевой гипотезы.

Эта величина нашла в Словаре самое широкое применение. Важно, что при этом в круг анализа вовлекаются также хотя и редкие, но значимые события, иногда даже двукратное появление слова или словосочетания. Предположим нам

<sup>3</sup> У многих лингвостатистиков именно ранговый словарь именуется «частотным словарем», для второго основного варианта частотного словаря они используют термин «алфавитно-частотный словарь».

<sup>4</sup> Как это принято в статистике, в таблицах относительных частот численные значения меньше 0,5 показаны многоточием.



надо оценить статистическую значимость слова *деньги* в макрожанре «Письма». Частота слова *деньги* в текстах Достоевского равна 2097, в письмах оно встретилось 822 раза. Предположим, что слово *деньги* не зависит от специфики макрожанра, тогда, зная долю писем во всем корпусе (0,1833), мы можем подсчитать математическое ожидание появления этого слова в письмах:

$$2097 \times 0,1833 = 284.$$

Подставляя величины 822 и 284 в нашу формулу, получаем  $S=31$  (величину исключительно высокую), отсюда вывод — слово *деньги* очень характерно для писем Достоевского. Будем называть такие единицы **лексическими маркерами** подкорпусов (макрожанров, микрожанров, периодов творчества, отдельных текстов, отдельных персонажей и т. п.). Соответствующие списки целиком входят в Словарь. В этих списках частота (f) подчинена мере статистической значимости (S). Примером могут служить таблицы 14-16.

Формула оценки статистической значимости может быть использована для выявления текстуальных связей слов. Весь текст механическим образом членится на фрагменты равной длины (скажем, 40 слов), а затем подсчитывается число фрагментов, в которых одновременно встретились слово x и слово y. Если реальная частота совместной встречаемости статистически значима, делается вывод о текстуальной связи двух слов. Таким образом, в Словаре найдет отражение еще один лингвистический объект — **текстуальные связи** слов.

Так, редкое слово *агония* встретилось в жанре критики всего 4 раза, но показало текстуальные связи с пятью словами: *актер* (S=18), *естественный* (S=6), *зритель* (S=10), *правда* (сущ.) (S=2), *умирать* (S=6). Из этих пяти связей одна (со словом *умирать*) может считаться общезначимой для русского языка, остальные — обусловлены конкретным текстом, где ведется речь об изображении агонии на сцене.

Более сбалансированное соотношение общезначимых и текстуальных связей обнаружим в жанре «Критика и публицистика» у слова *Гоголь*.

Таблица 002

**Текстуальные связи слова Гоголь (f=47)**

| S  | f |                | S | f |               | S | f |                |
|----|---|----------------|---|---|---------------|---|---|----------------|
| 18 | 3 | «Женитьба»     | 4 | 3 | исчезнуть     | 2 | 2 | выражаться     |
| 16 | 5 | перевод        | 4 | 2 | по-французски | 2 | 3 | господин       |
| 15 | 8 | Тургенев       | 4 | 4 | повесть       | 2 | 2 | комический     |
| 13 | 3 | «Мертвые души» | 4 | 2 | Поприщин      | 2 | 5 | литература     |
| 12 | 5 | Островский     | 4 | 4 | произведение  | 2 | 4 | начинать       |
| 11 | 2 | Виардо         | 4 | 3 | сверх того    | 2 | 3 | писатель       |
| 11 | 2 | непереводимый  | 4 | 5 | язык          | 2 | 2 | следовать      |
| 7  | 3 | жанр           | 3 | 2 | где-то        | 2 | 2 | сочинение      |
| 7  | 3 | комедия        | 3 | 2 | драма         | 2 | 7 | тогда          |
| 7  | 3 | перевести      | 3 | 2 | Лермонтов     | 2 | 7 | хотя           |
| 7  | 2 | Репин          | 3 | 2 | портрет       | 2 | 3 | художественный |
| 5  | 2 | Писемский      | 3 | 2 | правый        | 2 | 4 | что-то         |
| 5  | 4 | смех           | 3 | 5 | Пушкин        | 2 | 2 | Щедрин         |
| 4  | 2 | Диккенс        | 3 | 4 | французский   |   |   |                |

Здесь мы обнаруживаем текстуальные связи, которые могли бы ожидать от имени *Гоголь* даже не проводя специальных исследований, с другой же стороны, встречаемся со связями, обусловленными конкретным текстом, в котором речь идет о непереводимости Гоголя (*перевод, непереводимый, перевести, по-французски, французский, Виардо, Тургенев*). Подробнее об анализе текстуальных связей см. ниже.

**4. Описание отдельных таблиц книги**

Таблица 1 «Распределение лексем по основным жанрам» содержит 43577 лемм (и 3445 других лексем). Основные жанры (или макрожанры) сокращенно записаны как X — Художественная литература, К — Критика и публицистика, П — письма. Для экономии места слова с частотой 1 напечатаны в подбор в конце каждой страницы, два последних жанра записаны здесь кодами К и П, слова, встретившиеся в художественной литературе, не сопровождаются каким-либо символом.

Таблица 2 «Относительная частота лексем в основных жанрах» содержит 3000 лексем. Относительные частоты округлены до целого числа, лексемы,

встретившиеся в данном жанре с частотой менее 0,000005, показаны многоточием. Следует помнить, что Средняя, указанная в первом столбце, это – средняя взвешенная, т. е. получена с учетом неравного объема трех макрожанров.

Таблица 3 «100 самых частых лемм в текстах Достоевского» представляет собой верхнюю часть рангового словаря. Символы *r* и *f* читаются как «ранг» и «частота» соответственно. Аналогично устроены таблица 4 «100 самых частых лемм в художественных произведениях Достоевского», таблица 5 «100 самых частых лемм в критике и публицистике Достоевского», таблица 6 «100 самых частых лемм в письмах Достоевского».

Большой наглядностью для читателя обладают таблицы, в которых верхушка рангового словаря представлена в распределении по знаменательным частям речи. Это таблица 7 «40 самых частых существительных», таблица 8 «40 самых частых глаголов», таблица 9 «40 самых частых прилагательных». В этих таблицах каждый столбец посвящен своему основному жанру, внутри столбца слова упорядочены по убыванию частоты.

Информация о ранговом словаре (полностью представленном в электронной части) завершается четырьмя таблицами, в которых фигурируют только числа. Это таблица 10 «Частотный спектр рангового словаря лемм всего корпуса текстов», таблица 11 «Частотный спектр рангового словаря лемм беллетристики», таблица 12 «Частотный спектр рангового словаря лемм критики и публицистики», таблица 13 «Частотный спектр рангового словаря лемм писем». Четыре столбца этих таблиц интерпретируются следующим образом: в первом столбце показан ранг леммы в ранговом словаре. Цифры в последней строке первого столбца указывают на число разных лемм – 43577 во всем корпусе, 34257 в художественной литературе, 21448 в критике и публицистике, 17367 в письмах. Во втором столбце показана частота слова данного ранга, в третьем – накопленная частота всех слов с данным рангом и ниже. Последняя строка этого столбца – совокупная частота лемм данного корпуса. Заметим, что эта цифра несколько меньше числа графических слов, поскольку среди лемм довольно много словосочетаний. Наконец, в четвертом столбце дается накопленная относительная частота, т. е. результат деления накопленной частоты данного ранга на совокупную частоту лемм.

Следующие три таблицы показывают лексические маркеры основных жанров. Это таблица 14 «Лексические маркеры беллетристики», таблица 15 «Лексические маркеры критики и публицистики», таблица 16 «Лексические маркеры писем». Лексические маркеры поданы здесь в алфавитном порядке. Таблицы 17, 18 и 19 содержат информацию о лексических маркерах, упорядоченных по убыванию *S* и тем отличающихся от таблиц 14–16. В кратком виде эта информация представлена в таблице 003.

В таблицах 20 и 21 представлена грамматическая информация – редкий гость в частотных словарях. В таблице 20 даются абсолютные частоты во всем корпусе и по трем основным жанрам; в таблице 21 та же информация подана в виде относительных частот.

Грамматические классы включают традиционные части речи и некоторые подклассы внутри частей речи. Особо выделены деадъективные наречия, изменившие свои синтаксические функции: примером могут служить лексемы *довольно*, *следовательно*, *действительно*, *точно*, *давно*, *конечно*. Русские суффиксы соотнесены с частями речи и с грамматическими подклассами внутри частей речи. Так они и сгруппированы в данных таблицах. Что касается префиксов, то они хорошо коррелируют с мотивирующими основами в процессе словопроизводства, но результирующие производные слова (часто осложненные суффиксами) уже прямо не соотнесены с частями речи. Префиксы и первые компоненты сложных слов поданы в общем алфавитном порядке.

Поскольку в префиксальных образованиях часто наблюдается процесс морфологического опрошения, особо выделяются группы слов с этимологическим префиксом, чья семантическая мотивация частично или полностью затуманена. Примером может служить группа слов с начальным *v-*: *вкус*, *влияние*, *вменять*, *вместо*, *вовсе*, *внедрить*, *внезапно*, *вникнуть*, *внимание*, *вонзить*, *вперить*, *впечатление*, *впитаться*, *вплоть*, *впросак*, *вряд*, *всадник*. Такие группы слов даются в таблицах как отдельные строки, причем префикс в таком случае сопровождается знаком звездочки (\*).

## Важнейшие лексические маркеры основных жанров

| Художественная литература |                   | Критика и публицистика |                | Письма |                   |
|---------------------------|-------------------|------------------------|----------------|--------|-------------------|
| S                         |                   | S                      |                | S      |                   |
| 44                        | он                | 74                     | всё            | 72     | письмо            |
| 28                        | она               | 62                     | народ          | 70     | писать            |
| 25                        | -с                | 50                     | мы             | 59     | Достоевский       |
| 24                        | князь             | 50                     | наш            | 58     | мой               |
| 22                        | вдруг             | 50                     | русский        | 53     | роман             |
| 21                        | был               | 49                     | Европа         | 52     | написать          |
| 20                        | я                 | 37                     | народный       | 48     | рубль             |
| 17                        | Алеша             | 36                     | Франция        | 47     | Ф.                |
| 16                        | Лиза              | 35                     | Россия         | 46     | получить          |
| 16                        | лицо              | 34                     | война          | 45     | я                 |
| 16                        | рука              | 31                     | европейский    | 42     | ты                |
| 15                        | глаз              | 30                     | лишь           | 41     | Аня               |
| 15                        | так               | 29                     | русский (сущ.) | 38     | Петербург         |
| 14                        | дверь             | 29                     | славянин       | 38     | твой              |
| 14                        | комната           | 23                     | идея           | 37     | будет             |
| 14                        | проговорить       | 23                     | литература     | 37     | Паша              |
| 13                        | вскричать         | 23                     | новый          | 35     | выслать           |
| 13                        | да (утверждение)  | 23                     | общество       | 34     | прислать          |
| 13                        | Митя              | 23                     | политический   | 33     | ваш               |
| 13                        | смотреть          | 23                     | статья         | 33     | лист              |
| 13                        | что-то            | 23                     | турок          | 33     | многоуважаемый    |
| 12                        | будто             | 22                     | восточный      | 32     | Катков            |
| 12                        | весь              | 22                     | господин       | 31     | год               |
| 12                        | как               | 22                     | искусство      | 31     | деньги            |
| 12                        | как будто         | 22                     | он             | 30     | если              |
| 12                        | какой-то          | 21                     | большинство    | 30     | Старая Русса      |
| 12                        | нет (отрицание)   | 21                     | настоящий      | 29     | целовать          |
| 12                        | Раскольников      | 21                     | нация          | 28     | до свидания       |
| 12                        | с                 | 21                     | свой           | 28     | журнал            |
| 12                        | стоять            | 20                     | в              | 28     | Николай           |
| 12                        | так и             | 20                     | Германия       | 28     | просьба           |
| 12                        | что (местоимение) | 20                     | идеал          | 28     | уведомить         |
| 11                        | Голядкин          | 20                     | именно         | 27     | жить              |
| 11                        | Грушенька         | 20                     | исторический   | 27     | редакция          |
| 11                        | давеча            | 20                     | правительство  | 27     | «Русский вестник» |
| 11                        | опять             | 20                     | развитие       | 27     | Стелловский       |
| 11                        | пред              | 20                     | столь          | 27     | Федор             |
| 11                        | старик            | 20                     | цивилизация    | 27     | Федя              |
| 10                        | арестант          | 20                     | этот           | 26     | август            |
| 10                        | было (частица)    | 19                     | век            | 26     | адрес             |
| 10                        | быстро            | 19                     | великий        | 26     | «Заря»            |
| 10                        | глядеть           | 19                     | вера           | 26     | о                 |
| 10                        | говорить          | 19                     | высший         | 26     | обнимать          |
| 10                        | голова            | 19                     | люди           | 26     | Эмилия            |
| 10                        | заметить          | 19                     | маршал         | 25     | брат              |
| 10                        | как бы            | 19                     | обвинение      | 25     | июль              |
| 10                        | спросить          | 19                     | республика     | 25     | просить           |
| 10                        | стол              | 19                     | факт           | 25     | ради Бога         |
| 10                        | убить             | 19                     | явление        |        |                   |

Книга завершается тремя большими таблицами, где для основных жанров показана лексическая информация в связи с отдельными периодами жизни и творчества Достоевского. По своей структуре эти таблицы аналогичны таблице 2, отличаясь от нее содержательной интерпретацией столбцов.

В таблице 22 «Распределение лексем в беллетристике по периодам творчества» представлены пять столбцов — суммарный и четыре частных по периодам: 1844–1849, 1856–1865, 1866–1872, 1873–1880. Обычно выделяются три периода, но в данной таблице третий период представлен в виде двух столбцов, что позволит проверить устойчивость тех или иных хронологических тенденций.

Таблица 23 посвящена критике и публицистике по периодам творчества. Три выделенных периода (1845–1848, 1860–1865, 1873–1881) разделены длительными периодами публицистического молчания и в особой аргументации не нуждаются.

Заметим, что здесь исключены из рассмотрения обширные цитаты (всего 32 тыс. словоупотреблений).

Наибольшее число столбцов представлено в таблице 24, посвященной письмам Достоевского: 1832–1843 — годы учения, 1844–1849 — литературная деятельность вплоть до ареста, 1854–1859 от выхода из каторги до возвращения в Петербург, 1860–1866 литературная деятельность в Петербурге, 1867–1871 женитьба на А. Г. Сниткиной и пребывание за границей, 1872–1881 — от возвращения в Россию до конца жизни.

Объем хронологических подкорпусов отражен в таблице 004.

Таблица 004

**Объем хронологических и жанровых подкорпусов**  
(в тысячах графических слов)

| Художественная литература |      | Критика и публицистика |     | Письма    |     |
|---------------------------|------|------------------------|-----|-----------|-----|
|                           |      |                        |     | 1832–1843 | 18  |
| 1844–1849                 | 247  | 1845–1848              | 12  | 1844–1849 | 28  |
| 1856–1865                 | 406  |                        |     | 1854–1859 | 67  |
|                           |      | 1860–1865              | 140 | 1860–1866 | 79  |
| 1866–1872                 | 668  |                        |     | 1867–1871 | 144 |
| 1873–1880                 | 515  | 1873–1881              | 373 | 1872–1881 | 194 |
| Итого                     | 1835 |                        | 524 |           | 531 |

В следующем параграфе будут отмечены и другие таблицы, представленные только в электронной части Словаря.

### 5. Перспективы использования Словаря

Уже из предыдущего изложения должно быть ясно, что основной путь получения все более детальной информации заключается в последовательной дифференциации совокупного корпуса текстов Достоевского. В таблицах книжной части Словаря более или менее полно нашли отражение два верхних уровня получающейся иерархии подкорпусов: основные жанры (или макрожанры) и хронологические подкорпусы внутри макрожанров. На каждом таком уровне можно исследовать статистически значимые расхождения между подкорпусами в попытке выявить характерные черты каждого из подкорпусов и определить их взаимоотношения. Попробуем самым кратким образом показать возможности этого пути.

Обратимся для начала к характеристике основных жанров при помощи таблиц 1–21. Некоторые из этих таблиц (например, таблицы 10–13) могут показаться совершенно эзотерическими, сама интерпретация их потребовала бы слишком много места для нашего «Введения». Но даже такие таблицы могут дать полезные результаты. Например, сравнение таблиц 12 и 13 ясно показывает, что при равном объеме корпус критики и публицистики и корпус писем сильно расходятся по числу лемм: 21448 разных лемм в критике и публицистике,

17367 — в письмах. Еще больше расхождения в числе разных графических слов (соответственно 59446 и 43689)<sup>5</sup>.

Обращение к ранговым словарям (таблицы 4–9) кажется и убедительным, и наглядным. Действительно, слово человек возглавляет список существительных в художественных произведениях, оно занимает 4-е место в критике и публицистике и лишь 19-е в письмах. Однако следует помнить, что небольшие изменения в частоте могут сильно сказываться на месте в ранговом списке. Таблица 2 показывает следующие цифры относительной частоты (на 100 000 слов) — 219 в художественной литературе, 196 в критике и 108 в письмах. Противопоставленность писем двум остальным жанрам (в отношении слова человек) можно считать доказанным.

Для взвешенного суждения о противопоставлении подкорпусов должны использоваться таблицы всех трех типов: таблицы относительных частот, ранговые словари и таблицы лексических маркеров. При этом надо учитывать, что отдельные слова и целые классы слов обладают разной дифференцирующей силой. Так, в таблице 7 мы находим всего 10 существительных, присутствующих во всех трех списках (*человек, дело, время, раз, день, сердце, жизнь, случай, мысль, год*). Пять слов (*минута, люди, лета, вопрос, душа*) объединяют списки художественной литературы и критики; семь слов представлены в первом и третьем списках (*письмо, Бог, ночь, брат, вид, друг, час*). Всего два слова (*журнал и статья*) объединяют второй и третий списки. Таблица 8 содержит 17 глаголов, общих всем трем спискам (*быть, знать, говорить, мочь, сказать, хотеть, стать, видеть, думать, любить, пойти, взять, прийти, сделать, иметь, жить, дать*). Отсюда следует вывод — в сравнении с существительными глаголы в меньшей степени различают жанры.

Таблица 21 особенно интересна при поиске межжанровых различий.

Художественные произведения характеризуются повышением доли глаголов (особенно глаголов на -ся), деадъективных наречий, местоимений-существительных, местоимений-наречий, местоимений 3-го лица, междометий, слов *да* и *нет*, частиц при местоимениях (типа *какой-то, какой-нибудь*). Среди аффиксов отметим уменьшительные суффиксы, особенно слов женского рода (*-к-а, -очк-а, -шк-а*), суффикс прилагательных *-ив-ый*, глагольные суффиксы *-ну-ть* и *-ива-ть*, продуктивные префиксы *воз-, за-, по-, под-, раз-* и *у-*. Ярких отрицательных маркеров у этого макрожанра нет.

Жанр критики и публицистики очень выразительно характеризуется грамматическими показателями. Яркими положительными маркерами выступают имена нарицательные (особенно среднего рода), прилагательные, суперлативы, местоимения-прилагательные, союзы, вводные слова. Два последних класса указывают на синтаксическую сложность. В отличие от художественной литературы с ее преобладанием глаголов, данный жанр носит именной характер. Глаголы оказываются отрицательными маркерами. Тяга к обобщениям и универсальности косвенно проявляется в деперсонализации, местоимения 1-го и 2-го лица становятся отрицательными маркерами жанра. Среди аффиксов характерны суффиксы абстрактных существительных: *-ость, -ство, -ние, -тие*; префиксы *без-* и *не-*; сложные слова с первыми компонентами *едино-, обще-, против-, само-*. Очень характерны заимствованные суффиксы: *-аж, -изм, -ент, -мент, -ика, -ема, -ура, -ия*.

Положительных маркеров жанра писем очень немного, это — числительные, имена собственные, местоимения 1-го и 2-го лица, модальные слова. Яркие отрицательные маркеры — возвратные местоимения и местоимения 3-го лица. Обращение к таблицам лексем<sup>6</sup> дополнит эту общую характеристику жанра писем еще

<sup>5</sup> Адепты лингвостатистики обычно интерпретируют подобные различия в терминах богатства и бедности словаря. Лингвисты должны быть осторожнее: им нужно сначала договориться о значении терминов «богатство» и «бедность». Действительно ли наличие трех разных лемм *анализ, анализировать, анализирование* в жанре критики и публицистики свидетельствует о большем богатстве в сравнении с жанром писем, где представлено только слово *анализ*? Примем ли мы как контраргумент три разные леммы *безденежно, безденежный, безденежье*, встретившиеся в корпусе писем?

<sup>6</sup> Лексические маркеры берутся из таблиц 17 (художественная литература), 18 (критика и публицистика) и 19 (письма). В квадратных скобках приводятся лексические маркеры с меньшими значениями S, взятые соответственно из таблиц 14, 15 и 16.

одной любопытной деталью: среди важнейших лексических маркеров мы находим *будет, если, надеяться, во всяком случае [авось, будущий, в случае, дай Бог, ежели, кабы, мочь, на всякий случай, наверно]*.

Письма обращены в будущее, а значит, максимально усиливают условность и неопределенность. Публицистика с ее всеобщностью опирается предпочтительно на формы настоящего времени – среди лексических маркеров находим и слово *есть*. Художественное повествование преимущественно выбирает прошедшее время (лексический маркер – *был*).

Любопытны межжанровые различия в подаче времени. Художественные произведения (таблица 17) показывают такие лексические маркеры: *вдруг, давеча, опять, было, быстро, минута, мгновение, в эту минуту, снова, тотчас*. Здесь перед нами выступает «малое время». Среди лексических маркеров писем (таблица 19) находим *год, месяц (и названия месяцев), неделя (и дни недели), день, лето [весна, вчера, завтра, зима, осень, послезавтра, сегодня]* – одним словом – перед нами «среднее время». В критике и публицистике (таблица 18) мы находим «большое время»: *всё, новый, настоящий, исторический, век, современный [будущее, вековечный, грядущий, древний, нынешний, столетие, теперешний, эпоха]*.

Наши три жанра могут характеризоваться некоторыми общими чертами модальности, которые откроются перед нами через лексические маркеры. В письмах это обобщенная просьба – *просьба, просить, ради Бога [благодарить, нужно, отказать, похлопотать, ради Христа, умолять, хлопоты]*; в публицистике – настойчивая категоричность, выстраиваемая как будто бы на силлогизмах, – *лишь, именно, факт, явление, убеждение, вера, принцип, смысл, значение, признавать, формула [абсурд, без сомнения, бесспорно, в сущности, в том-то и дело, важнейший, вывод, выводить, доказывать, истина, конечно, логика, настоящий (реальный), нельзя не, непонимание, несомненно, отрицать, ошибка, понимание, правда, следственно, стало быть, суть, тем самым, фальшь]*; в художественной литературе – сравнение (будто, как, как будто, как бы, казалось, точно).

Лексические маркеры самым явным образом демонстрируют нам структурные особенности текстов данного жанра. Для художественной литературы характерно массовое введение прямой речи и сопровождающих ее ремарок, отсюда обилие глаголов говорения – *проговорить, вскричать, говорить, заметить, спросить, крикнуть, сказать, воскликнуть, вскрикнуть, закричать, прибавить, пробормотать, продолжать, прошептать, кричать, подхватить, рассказать, рассказывать и т. д.* Своя структура хорошо видна и в письмах – *многоуважаемый, голубчик, любезнейший, милый [дорогой, кланяться, бесценный, глубокоуважаемый, добрый, милостивый государь]* открывают письмо; закрывают его слова до свидания, *обнимать, принять уверения, P. S., благословлять [всегдашний слуга, жать, иметь честь, привет, прощай]*. Впрочем, в критике и публицистике нет никакого намека на структуру текста, которая проявилась бы в лексических маркерах.

Конечно, лексические маркеры дают исследователю максимум информации об основных темах жанра. Вот характерные группы маркеров жанра писем:

1) *письмо, писать, написать, получить, выслать, прислать, адрес, почта, послать, посылать [доставить, конверт, ответ, отправить, переписка, телеграмма, черкнуть]*;

2) *роман, лист, журнал, редакция, «Русский вестник», «Заря», напечатать, корректура, работа, печатать, издание, печатный лист, повесть, статья, работать [«Бедные люди», «Бесы», «Гражданин», «Двойник», день и ночь, «Детская сказка», «Дневник», «Дядюшкин сон», задержать, запоздать, «Идиот», книга, издать, «Карамазовы», «Мертвый дом», номер, «Отчетственные записки», перевод, подписчик, «Преступление и наказание», сочинение, срок, типография, цензор, цензура, «Эпоха»]*;

3) *рубли, деньги, долги, серебро, заплатить, контракт, счет (деньги), уплата [акция, ассигнация, банкир, вексель, доверенность, задолжать, занять (деньги), копейка, кредитор, курс, марка, неустойка, продажа, расписка, талер, франк]*;

4) *здоровье, здоров, лечение [анфизема, болезнь, болен, ванна, воды, диета, желудок, кашель, нездоровье, нервы, падухая, припадок, простудиться, хворать]*;

- 5) *Петербург, Старая Русса, Эмс, Дрезден, Семипалатинск, здесь, выехать, ехать, Женева, [Барнаул, Вевей, возвращение, гостиница, квартира, Кузнецк, Лоскутная (гостиница), Милан, Омск, отель, отъезд, Ревель, Тверь, Флоренция];*  
 6) *Лиля, Федя [детишки, детки, нянька, Сонечка];*  
 7) *[дождь, климат, погода].*

Очень четко членится на темы жанр критики и публицистики:

- 1) *мы, наш, русский, Европа, Россия, европейский, славянин, турок, цивилизация, еврей, европеец, славянский, человечество, католический [бескорыстие, болгарин, всемирный, грек, добровольцы, допетровский, дораста, достижение, зверство, истязание, Лефорт, мир, немец, освобождение, отечество, райа, Русь, серб, татарин, угнетенный, черногорец, Черняев];*  
 2) *народ, народный, идея, общество, нация, идеал, развитие, великий, вышедший, почва, вопрос, начало (принцип), национальный, общественный, принцип, смысл, интеллигентный, Петр, значение, мир, реформа, сознание, христианство, язык [братство, всеединение, выжить (добыть жизнью), глумиться, гражданин, дух, единение, жизнь, западник, земля, земство, знамя, крестьянин, либеральный, материализм, нигилист, обличать, обособление, племя, православие, прогресс, прогрессист, просвещение, публика, ретроградство, реформа, самодеятельность, святыня, славянофил, соединение, стремление, теория];*  
 3) *Франция, война, политический, большинство, Германия, правительство, маршал, республика, Восточный вопрос, папа (римский), партия, страна, армия, Константинополь, Англия, Австрия, Мак-Магон, республиканец, Бисмарк, войско, политика [Азия, атака, Биконсфильд, буржуазия, бюджет, дипломат, клерикал, коммунизм, король, левый (полит.), легитимист, меч, мир (покой), Национальное собрание, парламент, победа, пролетарий, революция, римский, сословие, социализм, султан];*  
 4) *литература, статья, автор, Пушкин, тип, поэт, «Современник», -бов, интерес, читатель [бездарный, Белинский, гений, Гоголь, действительность, Добролюбов, Жорж Занд, занимательность, критика, культура, Лермонтов, Онегин, писатель, произведение, творчество, фельетон, Чернышевский, читатель];*  
 5) *искусство, художник [Айвазовский, академия, архитектура, выставка, жанр, живопись, картина, талант, художественный];*  
 6) *обвинение [беременность, девочка, дело (юрид.), защитник, Невский проспект, подсудимая, присяжный, ребенок, самоубийство, среда, суд].*

Второй уровень дифференциации – хронологические подкорпусы внутри основных жанров (таблицы 22–24). Обратимся для иллюстрации к таблице 005, представляющей собой сжатый вариант таблицы 22, в котором третий и четвертый столбцы исходной таблицы слиты в один столбец, отражающий третий период творчества Достоевского.

Эволюция лексики Достоевского практически не была объектом лингвистического анализа, поэтому публикуемые таблицы могут дать начало самым разным исследованиям. Однако необходимы сугубая осторожность и неспешность в интерпретации материала. Ясно, что тривиальным окажется случай влияния отдельного текста, в котором данное слово обладает крайне высокой частотой. Примером может служить слово *арестант* из «Записок из мертвого дома». Особенно характерно такое поведение для имен нарицательных, обозначающих героев (*бабушка* в «Неточке Незвановой» и в «Игроке», *дядя* и *дядюшка* в «Дядюшкином сне», *наш герой* в «Двойнике», *князь* во многих текстах, *жена*, *муж* и т. п.). Большую приуроченность к конкретным текстам обнаруживают и многие другие существительные: не только *музыка* в «Неточке Незвановой», но и *топор*, *шинель* и т. п.

На противоположном полюсе оказываются слова, более или менее равномерно распределенные по трем периодам, – прежде всего слова с относительной частотой более чем 1 на 1000: *весь, время, другой, думать, за, к, какой, минута, ни, об, она, первый, по, раз, с, сам, слово, теперь, хотеть* и т. п.

**Распределение некоторых частых лексем художественной литературы  
по периодам творчества**

Относительная частота на 100 000 графических слов

|                | 1844 | 1856 | 1866 |                  | 1844 | 1856 | 1866 |
|----------------|------|------|------|------------------|------|------|------|
|                | -49  | -65  | -80  |                  | -49  | -65  | -80  |
| а (союз)       | 654  | 893  | 924  | вино             | 5    | 19   | 12   |
| али            | 3    | 9    | 16   | возле            | 20   | 7    | 2    |
| арестант       | ...  | 128  | 1    | волнение         | 39   | 25   | 20   |
| ах             | 63   | 48   | 47   | вообще           | 11   | 43   | 22   |
| барин          | 23   | 19   | 8    | вопрос           | 22   | 36   | 72   |
| бедный         | 76   | 49   | 28   | восторг          | 36   | 28   | 20   |
| без            | 110  | 132  | 131  | вот              | 358  | 271  | 307  |
| беспокойство   | 13   | 15   | 12   | вроде            | 2    | 13   | 12   |
| бледный        | 34   | 22   | 17   | все-таки         | 19   | 61   | 61   |
| Бог            | 63   | 45   | 63   | вскрикнуть       | 21   | 30   | 12   |
| Бог знает      | 21   | 10   | 10   | вскричать        | 7    | 66   | 48   |
| Боже           | 61   | 38   | 17   | вчера            | 37   | 27   | 50   |
| более          | 126  | 95   | 88   | вы               | 1151 | 854  | 1121 |
| больше         | 41   | 81   | 68   | высший           | 6    | 23   | 33   |
| бояться        | 58   | 56   | 60   | газета           | 2    | 8    | 8    |
| будто          | 225  | 152  | 75   | где              | 72   | 72   | 68   |
| бумага         | 31   | 12   | 14   | главное (вв.сл.) | 1    | 24   | 35   |
| бутылка        | 1    | 9    | 6    | глаз             | 197  | 117  | 120  |
| бы             | 378  | 406  | 559  | глубокий         | 25   | 10   | 11   |
| бывало         | 26   | 34   | 9    | говорить         | 402  | 380  | 342  |
| было (частица) | 53   | 38   | 53   | голова           | 141  | 94   | 87   |
| быть           | 1488 | 1518 | 1337 | гораздо          | 10   | 13   | 19   |
| будет          | 242  | 229  | 184  | горе             | 30   | 13   | 9    |
| был            | 1063 | 1072 | 929  | гореть           | 20   | 7    | 7    |
| есть           | 97   | 102  | 109  | грудь            | 48   | 16   | 16   |
| в              | 2241 | 2338 | 2465 | грустно          | 24   | 5    | 7    |
| вагон          |      | 4    | 7    | да (союз)        | 443  | 386  | 333  |
| вдруг          | 173  | 154  | 352  | да (утверждение) | 112  | 101  | 87   |
| ведь           | 157  | 248  | 213  | давеча           | 1    | 29   | 49   |
| великий        | 7    | 15   | 33   | давешний         | ...  | 4    | 11   |
| вера           | 2    | 6    | 13   | даже             | 183  | 371  | 398  |
| верить         | 22   | 45   | 55   | дама             | 36   | 30   | 31   |
| верно (вв.сл.) | 25   | 24   | 8    | дверь            | 61   | 45   | 73   |
| веровать       |      | 2    | 17   | девочка          | 19   | 18   | 16   |
| весь           | 1162 | 1268 | 1194 | действительно    | 40   | 55   | 59   |
| весьма         | 64   | 12   | 39   | делать           | 77   | 83   | 46   |
| взглянуть      | 58   | 34   | 17   | деньги           | 60   | 91   | 105  |
| видеть         | 191  | 132  | 158  | дескать          | 78   | 14   | 19   |



|              | 1844 | 1856 | 1866 |                     | 1844 | 1856 | 1866 |
|--------------|------|------|------|---------------------|------|------|------|
|              | -49  | -65  | -80  |                     | -49  | -65  | -80  |
| дети         | 28   | 28   | 38   | кредитка            | ...  | ...  | 5    |
| диван        | 11   | 14   | 27   | крикнуть            | 6    | 9    | 39   |
| дитя         | 23   | 15   | 3    | кричать             | 26   | 48   | 37   |
| для          | 145  | 202  | 184  | кровать             | 29   | 7    | 10   |
| до           | 186  | 282  | 215  | кроме               | 17   | 36   | 42   |
| до того, что | 11   | 30   | 22   | кругом              | 48   | 18   | 21   |
| добрый       | 69   | 47   | 28   | лестница            | 43   | 14   | 27   |
| довольно     | 46   | 45   | 62   | лишь                | 13   | 3    | 79   |
| досада       | 19   | 10   | 11   | маленький           | 41   | 36   | 38   |
| дочь         | 11   | 43   | 24   | может (=вв.сл.)     | 6    | 16   | 34   |
| дрожать      | 50   | 19   | 21   | может быть (вв.сл.) | 93   | 117  | 130  |
| дух          | 46   | 26   | 22   | мочь                | 284  | 309  | 350  |
| едва         | 44   | 15   | 13   | музыка              | 18   | 6    | 4    |
| ее           | 207  | 174  | 157  | мы                  | 335  | 421  | 292  |
| если         | 155  | 220  | 254  | мысль               | 55   | 50   | 100  |
| еще          | 292  | 380  | 375  | надо                | 17   | 115  | 124  |
| же           | 636  | 747  | 747  | наконец             | 240  | 134  | 106  |
| женщина      | 32   | 29   | 57   | напротив            | 17   | 26   | 37   |
| забыть       | 28   | 40   | 44   | не                  | 2097 | 2242 | 2362 |
| завопить     | ...  | 1    | 8    | нежный              | 13   | 10   | 4    |
| задрожать    | 14   | 4    | 6    | некоторый           | 38   | 61   | 63   |
| закричать    | 60   | 33   | 19   | неужели             | 19   | 39   | 30   |
| затем        | 23   | 13   | 32   | неужто              |      | 1    | 12   |
| зачем        | 53   | 41   | 53   | но                  | 619  | 795  | 866  |
| земля        | 17   | 23   | 38   | ну                  | 289  | 247  | 184  |
| злой         | 24   | 25   | 13   | нужно               | 97   | 20   | 12   |
| значит       | 4    | 29   | 24   | о! (междометие)     | 30   | 30   | 80   |
| и            | 4145 | 4379 | 4698 | образ               | 9    | 3    | 4    |
| ибо          | 22   | 10   | 42   | образом             | 45   | 27   | 16   |
| идея         | 18   | 20   | 41   | однако              | 30   | 31   | 64   |
| из           | 279  | 387  | 323  | однако же           | 23   | 22   | 29   |
| или          | 125  | 136  | 165  | около               | 17   | 11   | 18   |
| именно       | 47   | 85   | 97   | он                  | 2075 | 1976 | 2265 |
| иметь        | 47   | 71   | 92   | они                 | 330  | 475  | 405  |
| к            | 571  | 582  | 612  | оно (местоимение)   | 61   | 39   | 39   |
| -ка          | 21   | 14   | 10   | оно (частица)       | 52   | 9    | 6    |
| как          | 1020 | 865  | 877  | от                  | 404  | 339  | 322  |
| как будто    | 200  | 143  | 53   | ответить            | 8    | 11   | 37   |
| как бы       | 14   | 23   | 118  | отвечать            | 100  | 115  | 42   |
| как раз      | 1    | ...  | 12   | отчего              | 47   | 22   | 8    |
| какой-то     | 185  | 159  | 107  | очень               | 123  | 210  | 225  |
| каρέта       | 29   | 12   | 8    | перед               | 90   | 146  | 44   |
| конец        | 18   | 24   | 42   | передать            | 5    | 6    | 25   |
| конечно      | 49   | 59   | 92   | пистолет            | ...  | 3    | 8    |
| который      | 385  | 325  | 295  | плакать             | 45   | 39   | 31   |
| краснеть     | 15   | 4    | 7    | по                  | 349  | 356  | 358  |
| красота      | 9    | 9    | 10   | подле               | 13   | 20   | 24   |

|                | 1844 | 1856 | 1866 |                   | 1844 | 1856 | 1866 |
|----------------|------|------|------|-------------------|------|------|------|
|                | -49  | -65  | -80  |                   | -49  | -65  | -80  |
| пока           | 4    | 7    | 33   | слишком           | 32   | 47   | 71   |
| покамест       | 22   | 17   | 7    | словно            | 52   | 5    | 4    |
| покраснеть     | 28   | 12   | 14   | смущение          | 19   | 5    | 7    |
| полно (те)     | 41   | 12   | 8    | столь             |      | 2    | 20   |
| получить       | 16   | 18   | 29   | страдание         | 5    | 11   | 13   |
| поминутно      | 19   | 12   | 9    | стыд              | 25   | 8    | 12   |
| постель        | 47   | 14   | 16   | счастливый        | 45   | 18   | 18   |
| потом          | 164  | 100  | 100  | счастье           | 45   | 34   | 19   |
| потому         | 36   | 79   | 79   | так               | 710  | 595  | 559  |
| потому что     | 94   | 84   | 93   | таким образом     | 35   | 13   | 12   |
| почему         | 27   | 47   | 62   | такой             | 400  | 333  | 296  |
| почти          | 92   | 107  | 137  | темный            | 23   | 8    | 10   |
| правда (сущ.)  | 29   | 40   | 54   | теперь            | 334  | 310  | 309  |
| право (вв.сл.) | 40   | 20   | 15   | то (союз)         | 170  | 183  | 263  |
| право (сущ.)   | 13   | 17   | 26   | -то               | 251  | 268  | 233  |
| пред           | 11   | 3    | 73   | только            | 326  | 377  | 398  |
| про            | 93   | 110  | 122  | тоска             | 54   | 25   | 11   |
| пробормотать   | 5    | 6    | 23   | тот               | 545  | 613  | 732  |
| проговорить    | 48   | 48   | 67   | точно             | 27   | 59   | 65   |
| продолжать     | 39   | 60   | 55   | тяжело            | 27   | 26   | 9    |
| произнести     | 6    | 10   | 25   | убеждение         | 4    | 9    | 16   |
| просить        | 52   | 53   | 52   | убить             | 9    | 15   | 59   |
| прощай (-те)   | 42   | 19   | 15   | ужас              | 32   | 19   | 16   |
| пусть          | 17   | 29   | 39   | ужасно            | 21   | 45   | 69   |
| пьяный         | 9    | 29   | 28   | уже               | 223  | 239  | 332  |
| работа         | 24   | 52   | 8    | ум                | 20   | 31   | 48   |
| работать       | 16   | 16   | 6    | усмехнуться       | 4    | 2    | 19   |
| рад            | 34   | 30   | 27   | факт              | 2    | 13   | 25   |
| ради Бога      | 26   | 6    | 4    | хороший           | 44   | 34   | 28   |
| разумеется     | 25   | 82   | 43   | хорошо            | 93   | 67   | 50   |
| револьвер      |      |      | 10   | человек           | 262  | 218  | 209  |
| родной         | 47   | 14   | 10   | человечество      | 2    | 10   | 12   |
| Россия         | 2    | 8    | 19   | чем               | 37   | 73   | 62   |
| рука           | 241  | 159  | 187  | через             | 51   | 62   | 50   |
| -с             | 173  | 130  | 211  | черный            | 28   | 13   | 14   |
| самый          | 176  | 235  | 267  | черт              | 19   | 18   | 36   |
| свой           | 647  | 535  | 420  | чрез              | 3    | 4    | 20   |
| сей            | 30   | 20   | 48   | чрезвычайно       | 22   | 54   | 41   |
| сейчас         | 35   | 52   | 96   | чрезвычайный      | 2    | 13   | 29   |
| сердце         | 205  | 95   | 82   | что (местоимение) | 704  | 568  | 521  |
| серебро        | 14   | 9    | 1    | чтобы             | 183  | 224  | 231  |
| сказать        | 401  | 332  | 295  | чувствовать       | 77   | 40   | 25   |
| слеза          | 101  | 46   | 29   | этот              | 1090 | 1550 | 1571 |

В следующем частотном интервале (более 1 на 10000) очень много служебных слов, местоименных наречий, модальных слов: *вероятно, вовсе, всякий, где, друг друга, завтра, коли, нельзя, непременно, никогда, потому что*; очень широко представлены глаголы: *бояться, бросить, встретить, глядеть, желать, идти, лежать, посмотреть, сидеть, случиться, слушать, слышать*. Уникален глагол *просить*, вообще не показывающий расхождений по периодам. Не вызывает удивления стабильность частот слов *полный, последний, взгляд, год, жизнь, стол*; в некоторых других случаях исследователь начнет искать связь с индивидуальной авторской семантикой; таковы существительные *беспокойство, дама, девочка, дорога, красота, недоумение, честь*; среди прилагательных отметим *маленький* и *глупый*.

Подавляющее большинство частых слов обнаруживает расхождения между периодами. В некоторых случаях аномальное положение второго периода (резкое повышение или понижение частоты) может объясняться влиянием жанровой композиции этого периода (бессюжетность «Записок из мертвого дома» и «Зимних заметок о летних впечатлениях»). Так можно объяснить увеличение частоты у *мы* и *они* и уменьшение у *его, он*; большую употребительность слов *ведь, значит, рассказ(ыва)ть* и, наоборот, относительное уменьшение таких маркеров сюжетного повествования, как *вы, здесь, тут, вдруг, вот, видеть, заметить, сесть, стоять, спать, лестница, квартира, комната, окно*.

Однако в большинстве случаев тенденция изменения частот монотонна: либо последовательный рост, либо сокращение. Например, частота слова *который* сокращается, употребительность многих союзов (*а, но, однако, или, али, хоть, хотя, все-таки, если, чтобы*) нарастает (суммарная относительная частота 1922 – 2609 – 2802 соответственно). Отмечены значительный рост местоимений *тот* и *этот* (1635 – 2163 – 2303) и сокращение местоимений *оно* и *что* (765 – 607 – 560).

Изредка изменения частоты отражают реальные изменения в обществе. Ясно, что обозначения уходящих реалий (*карета, писарь, серебро*) будут сокращать частоту, обозначения новых – ее увеличивать (*вагон, газета, кредитка, револьвер, застрелиться*). В других случаях эти изменения связаны со сменой тем писателя (*бумага, бутылка, деньги*).

На протяжении трех периодов внутри семантических групп постоянно разыгрывается соперничество синонимов. Доля отдельного слова в группе – яркое тому свидетельство: *возле* (61%-26%-8%) и *подле* (39%-74%-92%); *отчего* (63%-32%-11%) и *почему* (37%-68%-89%); *нужно* (85%-15%-15%) и *надо* (15%-85%-85%); *однако же* (86%-74%-45%) и *однако* (14%-26%-55%); *покамест* (82%-71%-18%) и *пока* (18%-29%-82%), *отвечать* (93%-91%-47%) и *ответить* (7%-9%-53%). Вот как выглядит доля пяти «соперников» в группе слов сравнения: *как будто* (68%-59%-21%), *словно* (17%-2%-2%), *точно* (9%-25%-26%), *как бы* (5%-10%-47%), *вроде* (1%-5%-5%). По-видимому совпадает с общеязыковой тенденцией уменьшение частоты четырех вводных слов: *право* (41 – 21 – 16), *знать* (13 – 3 – 1), *знаете ли* (11 – 2 – 2), *верно* (26 – 25 – 8). На этом фоне удивительным кажется некоторый сдвиг к архаике в третьем периоде – в паре *перед* и *пред* доля второго слова составляет 10%-2%-63%, в паре *через* и *чрез* такие цифры – 6%-6%-29%; странным кажется и рост относительной частоты у слов: *сей* (24 – 17 – 47), *столь* (0 – 2 – 20) и у междометия *о!* (31 – 31 – 82).

Иногда наблюдается рост целой семантической группы. Показательна в этом отношении группа слов обозначения степени (*весьма, вполне, высший, гораздо, даже, до того, что, едва, именно, крайне, несколько, особенно, очень, по крайней мере, почти, слишком, совершенно, совсем, в высшей степени, только, ужасно, чуть (ли) не, чрезвычайно, чрезвычайный*). Их совокупная относительная частота увеличилась в полтора раза (1344 – 1837 – 2066).

Может меняться и полнота разработки каких-то фрагментов текста. С течением времени все реже появляются слова общей положительной оценки (*добрый, хороший, хорошо* 204 – 148 – 100). Ко второму периоду резко сокращается частота слов, описывающих портрет персонажа: *глаза, голова, губы, рот, зубы, щеки, грудь, темный, черный* (общая частота 533 – 299 – 297).

Особенно динамично протекает конкуренция в группах слов, характерных для ремарок, а именно слов говорения (и шире – производства звуков), слов-

обозначений эмоций. Вот некоторые группы с их относительными частотами по трем периодам:

*говорить, сказать, закричать, крик, вскрикнуть, шептать* (922-788-695);

*вскричать, кричать, обращаться* (47 + 150 - 103);

*вопрос, проговорить, крикнуть, обратиться, воскликнуть, произнести, пробормотать, сообщить, высказать, заявить, прокричать, восклицать, завопить* (106 - 167 - 362);

*сердце, чувство, чувствовать, впечатление, волнение* (419 - 222 - 195);

*восторг, счастливый, счастье* (127 - 84 - 58);

*довольный, наслаждение, рад* (55 - 76 - 49);

*слезы, плакать, тоска, дрожать, бледный, страх, ужас, горе, задрожать, тяжело, отчаяние, грустно, злой, досада, замирать, рыдание* (554 - 299 - 228);

*покраснеть, краснеть, стыд, смущение* (88 - 29 - 40);

*страдание, ненавидеть, гнев, злобно* (16 - 38 - 46).

Складывается впечатление, что этот участок лексики, сформировавшись в первый период, затем частично обновлялся при стабильности слов говорения (1075 - 1105 - 1160) и общем сокращении совокупной частоты группы эмоций (1259 - 748 - 616).

Заметим, что основная лексическая граница проходит между первым и вторым периодами. На стыке же второго и третьего периода заметна лишь та семантическая группа, которая связана с переходом к «идейным» романам: *мысль, верить, ум, ибо, идея, великий, Россия, веровать, убеждение, вера, человечество* (166 - 214 - 404).

В электронной части Словаря даются дальнейшие уровни дифференциации корпуса: подкорпусы, соответствующие периодам творчества, членятся на отдельные произведения. Используются и другие принципы классификации, например группировка писем по адресатам.

В художественной литературе естественным оказывается выделение микрожанров: авторская речь, речь персонажей, ремарки, письма и т. п. Подробнее см. [Шайкевич 1995, 1996]. В меньшей мере микрожанры могут быть выделены в публицистике. В качестве примера рассмотрим микрожанры художественной литературы (таблицы 006 и 007).

Таблицы микрожанров откроют широкие возможности изучения как самих этих типов текстов, так и стилистической характеристики лексем. В качестве примера рассмотрим микрожанр ремарок. Косвенным свидетельством естественности выделенного микрожанра служит легкость складывания лексических маркеров в некоторую цельную картину. В микрожанре ремарок такая естественность очевидна: ни в одном другом микрожанре группы маркеров не выделяются столь явно, ни в одном другом — не выступает столь явно функциональная нагрузка каждой группы.

Прежде всего в ремарках должен идентифицироваться говорящий, иногда есть указание и на адресат (характерный маркер — предлог к (886 - 17)<sup>1</sup>). Кроме очевидных местоимений — *он, она, я* и т. п. писатель использует для этой цели либо фамилию (*Мармеладов, Раскольников, Разуმიхин, Свидригайлов, Лебедев* и т. п.), либо имя и отчество (*Наталья Дмитриевна, Мария Александровна, Степан Трофимович* и т. п.), либо только имя (*Настенька, Сашенька, Фома, Алеша, Нелли, Наташа, Соня, Аглая, Ипполит, Коля, Алеша, Грушенька, Илюша, Лиза, Митя* и т. п.). Реже наблюдается альтернативное название — *Порфирий и Порфирий Петрович, Авдотья Романовна и Дуня, Лужин и Петр Петрович, Ганя и Гаврила Ардалионович, Николай Всеволодович и Ставрогин, Иван и Иван Федорович, Миусов и Петр Александрович, Катя и Катерина Ивановна, Лиза и Lise.*

<sup>1</sup> Первое число, приводимое в скобках при лексическом маркере, обозначает его абсолютную частоту в ремарках (f), второе число — мера S.

## Распределение слов по микрожанрам

А — авторская речь; Р — ремарки, сопровождающие прямую речь;  
Г — прямая речь героев (в т.ч.: м — монологи, рс — рассказы);  
Пс — письма; В — внутренняя речь героев; Л — литературные  
жанры не от лица автора

|                   | ВСЕГО | А    | Р  | Г    | м   | рс  | Пс  | В   | Л  |
|-------------------|-------|------|----|------|-----|-----|-----|-----|----|
| Всего слов        | 1835  | 856  | 88 | 752  | 140 | 91  | 49  | 58  | 26 |
| знаменитый        | 53    | 42   |    | 9    | 4   | 2   |     | 1   | 1  |
| знамо             | 10    | 4    |    | 6    |     |     |     |     |    |
| знамо дело        | 4     |      |    | 4    |     |     |     |     |    |
| знамши            | 1     |      |    | 1    | 1   |     |     |     |    |
| знамя             | 16    | 3    |    | 5    | 1   |     |     | 1   | 6  |
| знание            | 67    | 50   |    | 16   | 3   | 3   | 1   |     |    |
| знатнейший        | 3     | 2    |    | 1    | 1   |     |     |     |    |
| знатно            | 1     |      |    | 1    |     |     |     |     |    |
| знатность         | 2     | 1    |    | 1    | 1   |     |     |     |    |
| знатный           | 66    | 31   |    | 29   | 7   | 5   | 4   | 1   | 1  |
| знаток            | 22    | 13   | 2  | 7    |     | 1   |     |     |    |
| знать             | 6972  | 1739 | 45 | 4641 | 700 | 350 | 251 | 199 | 79 |
| Бог знает         | 208   | 98   | 10 | 69   | 4   | 6   | 12  | 11  | 8  |
| Господь знает     | 8     | 1    |    | 4    |     |     | 3   |     |    |
| дать знать        | 3     | 1    |    | 2    |     |     |     |     |    |
| почем знать       | 7     | 1    |    | 5    | 2   |     |     | 1   |    |
| честь знать       | 2     | 1    |    | 1    |     | 1   |     |     |    |
| знаете ли (-есть) | 58    |      |    | 49   | 8   | 3   | 9   |     |    |
| знай наших        | 1     | 1    |    |      |     |     |     |     |    |
| знать (вв.сл.)    | 51    | 2    |    | 42   | 5   | 12  | 4   | 3   |    |
| знать (сущ.)      | 5     |      |    | 3    | 1   | 2   | 2   |     |    |
| значение          | 90    | 52   | 1  | 36   | 13  | 2   |     |     | 1  |
| значит            | 401   | 68   | 3  | 288  | 51  | 40  | 2   | 37  | 3  |
| значительно       | 44    | 17   | 16 | 11   | 2   | 2   |     |     |    |
| значительный      | 121   | 86   | 8  | 25   | 6   | 2   | 2   |     |    |
| значить           | 206   | 58   |    | 124  | 16  | 8   | 9   | 12  | 3  |
| золовка           | 23    | 18   |    | 4    |     |     | 1   |     |    |
| золото            | 51    | 30   | 1  | 17   | 6   | 2   | 1   | 1   | 1  |
| золотой           | 157   | 85   | 1  | 63   | 12  | 14  | 4   | 1   | 3  |
| зонтик            | 30    | 19   | 1  | 10   | 3   |     |     |     |    |

**Распределение слов по микрожанрам.  
Относительные частоты на 100 тыс. словоупотреблений**

А — авторская речь; Р — ремарки, сопровождающие прямую речь;  
Г — прямая речь героев (в т.ч.: м — монологи, рс — рассказы);  
Пс — письма; В — внутренняя речь героев; Л — литературные  
жанры не от лица автора

|             | Средняя | А    | Р    | Г    | м    | рс   | Пс   | В    | Л    |
|-------------|---------|------|------|------|------|------|------|------|------|
|             | 1835    | 856  | 88   | 752  | 140  | 91   | 49   | 58   | 26   |
| и           | 4730    | 5428 | 4024 | 4062 | 4723 | 5003 | 4367 | 4783 | 5123 |
| ибо         | 35      | 27   | 5    | 41   | 95   | 38   | 71   | 26   | 144  |
| играть      | 21      | 22   | 3    | 22   | 13   | 22   | 24   | 7    | 12   |
| идеал       | 5       | 5    |      | 7    | 21   | 2    |      | 2    |      |
| идея        | 35      | 31   | 17   | 41   | 57   | 14   | 26   | 49   | 16   |
| идиот       | 4       | 2    |      | 5    | 6    | 6    |      | 14   | 29   |
| идти        | 78      | 73   | 17   | 89   | 62   | 80   | 55   | 129  | 70   |
| из          | 347     | 454  | 298  | 248  | 298  | 274  | 243  | 239  | 431  |
| из-за       | 22      | 21   | 20   | 23   | 30   | 18   | 15   | 29   | 12   |
| изверг      | 4       | 1    |      | 8    | 10   | 9    |      | 4    |      |
| известие    | 13      | 23   | 6    | 6    | 6    | 6    | 4    | 7    | 4    |
| извинить    | 14      | 1    | 1    | 32   | 9    | 10   | 2    | 7    |      |
| изволить    | 17      | 2    | 2    | 36   | 15   | 25   | 15   | 8    |      |
| изволь (те) | 5       |      |      | 11   | 1    | 5    |      | 4    |      |
| измениться  | 6       | 9    | 7    | 3    | 4    | 3    | 2    | 4    |      |
| измученный  | 7       | 6    | 1    | 1    | 1    |      |      |      | 4    |
| изредка     | 5       | 10   | 1    | 1    | 2    | 1    | 2    | 4    |      |
| изумление   | 11      | 18   | 46   | 2    | 1    |      | 4    |      | 12   |
| или         | 161     | 162  | 25   | 173  | 185  | 121  | 122  | 225  | 193  |
| именно      | 92      | 88   | 12   | 108  | 148  | 80   | 58   | 117  | 58   |
| иметь       | 86      | 80   | 16   | 101  | 113  | 52   | 91   | 45   | 119  |
| имя         | 18      | 16   | 1    | 21   | 16   | 15   | 41   | 7    | 74   |
| иначе       | 19      | 17   | 1    | 25   | 36   | 9    | 24   | 22   | 16   |
| иногда      | 59      | 85   | 29   | 35   | 38   | 55   | 53   | 29   | 82   |
| иной        | 39      | 51   | 9    | 30   | 44   | 28   | 53   | 35   | 21   |
| искать      | 23      | 32   | 7    | 15   | 16   | 23   | 17   | 11   | 24   |
| искренний   | 7       | 8    | 9    | 7    | 8    | 3    | 17   | 2    | 4    |
| искренно    | 12      | 12   | 8    | 13   | 22   | 8    |      | 16   |      |

Довольно обычна идентификация персонажей при помощи имен нарицательных. Чаще всего мы наблюдаем здесь любимый Достоевским титул князь, обычны также воинские звания *генерал, майор, капитан, штабс-капитан, поручик*. Указание на должность, общественное положение также обычные способы идентификации – *крупер, мещанин, студентка, председатель, старец*. Наконец, типично указание на пол и возраст персонажей: *парень, старичок, юноша и т. п.* Конечно, и эти обозначения совместимы с именами собственными: *Николай Сергеевич и старик, Анна Андреевна и старушка, Пульхерия Александровна и бедная женщина, Иван Федорович и генерал, Лизавета Прокофьевна и генеральша, Кирилов и инженер, Макар Иванович и старик, Федор Павлович и старик, Николай Парфенович и следователь, Ипполит Кириллович и прокурор*. Особого упоминания требует прием предварительного называния персонажа именем нарицательным, прежде чем будет введено его имя: *толстяк, молодой человек, боксер, чиновник, черномазый*.

В какой-то мере в ремарках отражено и пространство, в котором находится говорящий, окружающие его предметы обстановки: *комната (116 – 4), передняя (13 – 3), кабинет (9 – 2), порог (14 – 3), дверь (84 – 3), стол (95 – 7), стул (85 – 11), кресло (52 – 9), диван (54 – 7)*.

Речевое взаимодействие персонажей, как правило, предполагает и визуальный контакт: *смотреть (385 – 19), глядеть (357 – 14), повернуться, обернуться и т. п. (136 – 16)*.

Центральный элемент ремарки – *verbum dicendi*, вводящий прямую речь, или *verbum cogitandi*, вводящий внутреннюю речь. В русском языке это, главным образом, глагол совершенного вида, ср.

|   |                             |
|---|-----------------------------|
| <i>сказать</i> (1127 – 53)                                  | <i>говорить</i> (430 – 6)   |
| <i>проговорить</i> (889 – 116), <i>заговорить</i> (84 – 13) |                             |
| <i>заметить</i> (393 – 33)                                  | <i>замечать</i> (40 – 5)    |
| <i>воскликнуть</i> (299 – 68)                               | <i>воскликать</i> (99 – 33) |
| <i>вскричать</i> (320 – 60), <i>произнести</i> (216 – 48)   |                             |
| <i>промолвить</i> (50 – 25), <i>провозгласить</i> (25 – 12) |                             |
| <i>заявить</i> (22 – 5)                                     |                             |
| <i>подумать</i> (331 – 40)                                  | <i>думать</i> (171 – 4)     |
| (также про себя 169 – 34)                                   |                             |

Некоторые из этих глаголов включают реплики героев в диалогическую цепь:

|                               |                              |
|-------------------------------|------------------------------|
| <i>спросить</i> (705 – 80)    | <i>спрашивать</i> (53 – 5)   |
| <i>осведомиться</i> (37 – 17) |                              |
| <i>ответить</i> (313 – 57)    | <i>отвечать</i> (574 – 68)   |
| <i>подхватить</i> (221 – 56)  | <i>подхватывать</i> (11 – 9) |
| <i>отозваться</i> (68 – 28)   |                              |
| <i>возразить</i> (31 – 17)    | <i>возражать</i> (11 – 4)    |
| <i>подтвердить</i> (58 – 21)  | <i>подтверждать</i> (11 – 5) |
| <i>перебить</i> (172 – 48)    | <i>перебивать</i> (17 – 9)   |
| <i>прервать</i> (99 – 29)     | <i>прерывать</i> (26 – 11)   |

Ту же роль играют и некоторые другие глаголы, описывающие весьма характерную для Достоевского ситуацию «коллективного говорения»: *вернуть (25 – 15), вставить (12 – 8), ввязаться (22 – 13), включить (6 – 5), подсказать (11 – 7)*.

Вплетенность реплик в структуру дискурса манифестируется такими словами, как *вдруг (1364 – 71), внезапно (41 – 7), тотчас (136 – 4), тотчас же (105 – 5), неожиданно (49 – 13), опять (505 – 32), снова (74 – 15), в свою очередь (16 – 5), брякнуть (17 – 10)*.

К внутренней структуре монологов отсылают нас некоторые фазовые глаголы – *начать (409 – 35), приступить (7 – 4), продолжать (539 – 70), прибавить (483 – 71), прибавлять (25 – 10), повторить (108 – 32), повторять*

(98 - 21), *заклЮчить* (122 - 27). Сходную функцию выполняют и некоторые наречия: *всё* (111 - 6), *всё ещё* (47 - 7), *поминутно* (20 - 3), *наконец* (274 - 14).

Частица *было* (198 - 23) указывает на колебания говорящего в своей речи или в своих действиях, ту же функцию выполняет глагол *осесться* (15 + 9).

Перечисленные выше глаголы говорения (более 6000 словоупотреблений) обозначают передачу информации без каких бы то ни было дополнительных семантических нюансов. Наряду с ними в ремарках Достоевского широко представлены самые разнообразные лексические средства, обозначающие различные особенности говорения. В этом фонетическом пространстве можно выделить несколько измерений; в выражении соответствующих смыслов большую роль играют не только глаголы, но также существительные, прилагательные и наречия. Синтаксические различия между этими частями речи оказываются несущественными в семантике ремарок: всегда есть возможности разного грамматического выражения, причем роль грамматической «подпорки» часто принимает на себя слово *голос* (557 - 73) [также *голосок* (35 - 20)], гипероним рассматриваемой семантической области.

Первое измерение связано с различиями по громкости:

|   |   |
|---|---|
| (вс) <i>крикнуть, за/про/кричать</i> (1161 - 102) |   |
| <i>громко/ий/че</i> (112 - 25)                    | <i>вполголоса</i> (46 - 24)               |
| <i>вслух</i> (20 - 4)                             | <i>тихо/ий</i> (159 - 24)                 |
| <i>за/про/визжать, взвизгнуть</i> (49 - 16)       | <i>за/про/шептать, шепнуть</i> (330 - 65) |
| <i>за/вз/про/реветь</i> (46 - 25)                 | <i>шепот</i> (94 - 30)                    |
| (за) <i>вопить</i> (120 - 9)                      | <i>полушепотом</i> (22 - 15)              |
| <i>возвысить</i> (16 - 11)                        | <i>понизить</i> (12 - 11)                 |

Второе измерение ассоциировано с темпом речи (или мысли) персонажа:

|  |                                       |
|--|---------------------------------------|
| (за) <i>торопиться/ливо/ый</i> (98 - 20) | <i>протянуть, -тягивать</i> (90 - 18) |
| <i>быстро/ый</i> (130 - 19)              | <i>медленно</i> (29 - 8)              |
| <i>скороговоркой</i> (40 - 22)           | <i>сыпать</i> (6 - 3)                 |
| <i>вполыхах</i> (9 - 6)                  | <i>поспешить/но</i> (58 - 10)         |
| (про) <i>мелькнуть</i> (59 - 13)         | <i>пронестись</i> (20 - 9)            |

Третье измерение основано на контрасте разборчивости/неразборчивости речи:

|  |                                    |
|--|------------------------------------|
|  | <i>за/про/бормотать</i> (372 - 75) |
| <i>раздельно</i> (16 - 14)               | <i>проворчать</i> (39 - 23)        |
| <i>отчеканить</i> (14 - 12)              | <i>промямлить</i> (23 - 16)        |
| <i>с расстановкой</i> (9 - 12)           | <i>промычать</i> (16 - 14)         |
| (также <i>размеренно, скандировать</i> ) | <i>прошамкать</i> (6 - 8)          |

Итак, если в первых двух измерениях ремарки Достоевского заметно сдвинуты в сторону громкости и быстроты, то в третьем измерении наблюдается яркая асимметрия - торжествует невнятность.

Эти три оси, конечно, не покрывают семантику всех глаголов речи, многие из которых привносят свои индивидуальные добавки смысла, ср. *за/про/лепетать* (98 - 36), *за/про/скрежетать* (33 - 17), *защебетать* (14 - 14), *за/про/звенеть, звонко/ий* (27 - 11), *пропищать* (16 - 15), *прошипеть* (14 - 13), *сюсюкать* (5 - 6), *затараторить, протрещать, хриплый, шепелявить, загалдеть* и др.

Рассмотренное семантическое поле незаметно переходит в другое, где внешние свойства речи все теснее переплетаются либо с общим накалом эмоций, либо с конкретным эмоциональным содержанием и где гиперонимом выступает уже не *голос*, а *тон* (61 - 12), также *нотка* (7 - 4). Неконтролируемая эмоция видна в характерном для Достоевского глаголе *задышаться* (65 - 19), также в глаголах *захлебываться* (8 - 6), *заикаться* (11 - 9), *запинаться* (10 - 6), *прерываться* (14 - 7). Намеренная демонстрация эмоций говорящего видна в глаголе *отрезать* (81 - 30), в словах *резко/ий* (86 - 23).



Видное место в ремарках отводится внешнему (неречевому) выражению психологического состояния говорящего. И здесь мы сталкиваемся с неожиданной асимметрией. Плач, конечно, представлен: слезы (74 - 5), рыдать (23 - 5), залиться (20 - 7), простонать (17 - 14), всхлипывать (7 - 5); но смех встречается намного чаще: за/рас/смеяться, смех (314 - 23), усмехаться/нуться, усмешка (166 - 35), за/хохотать, хохот (57 - 16), улыбаться, улыбнуться, улыбка (333 - 36), хихикать/нуть (18 - 9), осклабиться (20 - 13), скалить (5 - 4). В целом, однако, наблюдаемые физиологические реакции скорее ассоциируются с отрицательными эмоциями. Характерны обозначения: дрожи - за/дрожать, вздрогнуть и т. п. (329 - 30), затрястись (9 - 3); кровотока - бледный, (по)бледнеть (86 - 10), (по)краснеть, за/раскраснеться (118 - 20), вспыхнуть (37 - 11), побагроветь (5 - 3),

Граница между контролируемой и неконтролируемой эмоциями довольно зыбка. Переход в область мимики и жестов одновременно означает переход в область семиотических явлений, Специфические гиперонимы здесь редки - есть жест (19 - 4), есть мина (8 - 4), но лишь в слабой степени характерны взгляд (86 - 4), взор (6 - 2), существительные зрения - скорее примета собственно авторской речи.

Лексические маркеры области взглядов, мимики и жестов охватывают многочисленные существительные, обозначающие части тела и лица, и самые разнообразные глаголы:

|                   | существительные  | глаголы  |
|-------------------|------------------|--|
| глаза (298 - 16)  |                  | скосить, по/накоситься (также искоса) (18 - 8)   |
| глазки (16 - 4)   |                  | подмигивать/нуть (15 - 7)                        |
|                   |                  | уставиться (16 - 8) потупить(ся) (28 - 10)       |
|                   |                  | переводить (13 - 5) опустить глаза (14 - 10)     |
|                   |                  | прищури(ва)ться (15 - 9) выпучить (6 - 2)        |
|                   |                  | вытаращить (5 - 4), (за) сверкать/нуть (38 - 16) |
| брови (15 - 5)    | лоб (19 - 3)     |  |
| рот (23 - 3)      | губ(к)а (48 - 7) |  |
| зубы (32 - 8)     | лицо (168 - 6)   | исказиться, искаженный (10 - 6)                  |
|                   |                  | (ис/по/с)кривиться (также криво) (51 - 21)       |
| голова (227 - 15) |                  | поднять, поднимать, подымать (91 - 12)           |
|                   |                  | качать, покач(ив)ать (30 - 12)                   |
|                   |                  | кивать/нуть (41 - 12) вскинуть (13 - 7)          |
|                   |                  | наклонить/ять (8 - 4) склонить/ять (9 - 3)       |
| плечо (65 - 12)   |                  | пожимать (11 - 7)                                |
| рука (477 - 25)   |                  | сжимать (48 - 16) указ(ыв)ать (104 - 20)         |
|                   |                  | стиснуть (8 - 3) ударить/ять (34 - 6)            |
|                   |                  | (в)сплеснуть (44 - 20) хлопнуть (12 - 10)        |
|                   |                  | ломать руки (15 - 8) стукнуть/чать (33 - 8)      |
|                   |                  | (за)махать, махнуть (80 - 20) дернуть (10 - 4)   |
|                   |                  | сложить (15 - 4) потирать (14 - 9)               |
| кулак (34 - 10)   | ладонь (21 - 8)  |  |
| палец (50 - 5)    | колени (25 - 3)  |  |
|                   |                  | (за)топать, топнуть (26 - 12)                    |

Наконец, в ремарках в изобилии представлены реальные позы, действия и движения, которые нельзя отнести к жестам. Рука отходит на второй план [ср. однако, хватать(ся), схватить/ывать (109 - 15), ткнуть (5 - 3), толкать (7 - 4)]. Движения и изменения позы распространяются на человека в целом, ср. нагнуться (18 - 5) и наклоняться (6 - 6). Очень характерны размашистые, судорожные движения, ср. бросаться/иться (74 - 5), рвануться (5 - 3), стремительно (15 - 7), судорожно/ый (14 - 4).

Лексические маркеры ремарок ясно указывают на ту ситуацию сидения, в которой чаще всего и ведутся диалоги персонажей Достоевского, и на завершение этой ситуации:

|                      |                      |                       |
|----------------------|----------------------|-----------------------|
| развалиться (5 - 3)  | усесться (40 - 8)    | подсаживаться (3 - 4) |
| вскинуться (28 - 19) | выпрямиться (13 - 7) | привстать (21 - 7)    |