

РОССИЙСКАЯ АКАДЕМИЯ НАУК
ИНСТИТУТ РУССКОГО ЯЗЫКА им. В. В. ВИНОГРАДОВА

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

ДИСТРИБУТИВНО-
СТАТИСТИЧЕСКИЙ АНАЛИЗ
ЯЗЫКА РУССКОЙ ПРОЗЫ
1850—1870-х гг.

Том 2



Издательский Дом ЯСК
Москва 2016

УДК 811.161.1
ББК 81.2 Рус
Ш 17



Издание подготовлено при финансовой поддержке
Российского гуманитарного научного фонда (РГНФ)
проект № 16-04-16022

Рецензенты:

д. ф. н. А. Ф. Журавлев, д. ф. н. С. А. Крылов

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

Ш 17 Дистрибутивно-статистический анализ языка русской прозы 1850—1870-х гг. Т. 2. М.:
Издательский Дом ЯСК, 2016. — 848 с. — (Studia philologica.)

ISBN 978-5-9907947-7-1

В соответствии с двояким характером исследования в целом настоящий том включает две части. В части 4 дистрибутивно-статистический анализ (ДСА) проводится на материале 15 миллионов слов, непосредственно примыкающих друг к другу в текстах русской прозы 1850—1870-х гг. Тот же формальный метод прилагается к английскому корпусу (4 миллиона слов романов Диккенса). И в том, и в другом случае ДСА приводит к открытию дистрибутивных классов и некоторых грамматических конструкций. Часть 5 представляет собой частотный словарь русского корпуса, показывающий распределение слов по основным жанрам, по микрожанрам и по текстам 25 писателей.

В существенно расширенном виде том 2 представлен на компакт-диске.

УДК 811.161.1
ББК 81.2 Рус

Электронная версия данного издания является собственностью издательства,
и ее распространение без согласия издательства запрещается.

ISBN 978-5-9907947-7-1

© А. Я. Шайкевич, В. М. Андрущенко, Н. А. Ребецкая, 2016
© Издательский Дом ЯСК, 2016

СОДЕРЖАНИЕ

Часть 4. Дистрибутивно-статистический анализ в минимальном интервале	5
4.1. Работа ДСА в минимальном интервале на материале корпуса русской прозы	7
4.1.0. Систематизация морфологических данных	7
4.1.0.1. Адъективные парадигмы	8
4.1.0.2. Субстантивные парадигмы	15
4.1.0.3. Глагольные парадигмы	24
4.1.1. Минимальный интервал в русском корпусе	29
4.1.1.1. Частые слова в левой позиции и суффиксы правых слов	32
4.1.1.2. Частые слова в правой позиции и суффиксы левых слов	39
4.1.1.3. Поиск конструкций в зоне слов средней частоты	45
4.1.1.4. Поиск аномалий в конструкциях	50
4.2. Минимальный интервал в английском корпусе	53
4.2.1. Дистрибутивные классы частых слов	54
4.2.2. Комбинаторика дистрибутивных классов	62
4.2.3. Морфология и дистрибутивные классы	69
Часть 5. Дифференциация лексики русской прозы 1850—1870-х годов	75
5.1. Меры лексического сходства частотных словарей	77
5.2. Распределение лексики по основным жанрам прозы 1850—1870-х гг.	80
5.3. Распределение лексики по микрожанрам прозы 1850—1870-х гг.	108
5.4. Распределение лексики по текстам восьми авторов	136
5.5. Расширение круга сравниваемых авторов	168
5.6. Лексические маркеры авторов	171
5.7. Интерпретация авторских лексических маркеров	206
5.8. Авторские маркеры и дистрибутивно-статистическая морфология	216
5.9. Межавторские изоглоссы	231
5.10. Проблема взаимодействия факторов	236
Таблица 5-А. Распределение лексики по основным жанрам	242
Таблица 5-В. Распределение лексики по микрожанрам	327
Таблица 5-С. Распределение лексики по текстам важнейших авторов	441
Таблица 5-Д. Распределение лексики по текстам восьми авторов, родившихся в 1812—1829 гг.	625
Таблица 5-Е. Распределение лексики по текстам девяти авторов, родившихся в 1829—1840 гг.	715
Таблица 5-Ф. Основные параметры распределения лексики по 25 авторам	825
Таблица 5-Г. Ранговый словарь 1500 самых частых лемм	831
Перечень больших таблиц в электронной версии тома 2	839
Литература	843