

А.В. Гребенщикова

**ОСНОВЫ
КВАНТИТАТИВНОЙ
ЛИНГВИСТИКИ
И НОВЫХ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ**

Учебное пособие

3-е издание, стереотипное

*Рекомендовано Учебно-методическим объединением
по образованию в области лингвистики
Министерства образования и науки
Российской Федерации в качестве учебного пособия
для студентов, обучающихся по направлению подготовки
магистров «Лингвистика»*

Москва
Издательство «ФЛИНТА»
2018

УДК 81'32:81'33(075.8)
ББК 81.1-5-923
Г79

Р е ц е н з е н т:

д-р филол. наук, профессор, академик Международной академии информатизации, директор Института прикладной и математической лингвистики *Р.К. Потапова*

Гребенщикова А.В

Г79 Основы квантитативной лингвистики и новых информационных технологий : учеб. пособие / А.В. Гребенщикова. – 3-е изд., стер. – М. : ФЛИНТА, 2018. – 152 с.

ISBN 978-5-9765-2137-7

Учебное пособие содержит теоретический и практический материал, задания для самостоятельной работы, глоссарий по дисциплине «Квантитативная лингвистика и новые информационные технологии». В теоретическую часть включен материал, связанный с особенностями использования квантитативных методов в лингвистических исследованиях. Практическая часть содержит задания, направленные на обобщение знаний, отработку умений и навыков, необходимых для проведения лингвостатистических исследований на базе новых информационных технологий.

Для студентов лингвистических факультетов высших образовательных учреждений, а также аспирантов и соискателей, работающих над проблемами лингвистики, преподавателей-филологов и лингводидактов.

УДК 81'32:81'33(075.8)
ББК 81.1-5-923

ISBN 978-5-9765-2137-7

© Издательство «ФЛИНТА», 2014
© Гребенщикова А.В., 2014

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	6
1. КВАНТИТАТИВНАЯ ЛИНГВИСТИКА КАК НАПРАВЛЕНИЕ ОБЩЕГО ЯЗЫКОЗНАНИЯ	9
1.1. Понятие «квантитативная лингвистика» и взаимосвязь дисциплины с другими науками	9
1.2. История взаимодействия математики и лингвистики. Математическая (комбинаторная и квантитативная) лингвистика	11
1.3. Характеристика квантитативных методов в лингвистике и их роль	15
1.4. Основные области применения структурно-вероятностной модели языка	20
Вопросы для самоконтроля	22
Основные теоретические аспекты для обсуждения	22
Список рекомендуемой литературы	23
Практические задания	23
Задания для самостоятельной работы	24
2. ПОНЯТИЕ И СУЩНОСТЬ ЛИНГВОСТАТИСТИЧЕСКОГО АНАЛИЗА. КЛЮЧЕВЫЕ ПОНЯТИЯ КВАНТИТАТИВНОЙ ЛИНГВИСТИКИ	25
2.1. Условия успешного осуществления лингвостатистического анализа	25
2.2. Понятие статистического закона и вероятности	27
2.3. Понятие цели и единицы лингвостатистического анализа	30
2.4. Методика сбора информации для лингвостатистического анализа	31
2.5. Минимально-необходимые статистические инструменты: частота, генеральная и выборочная совокупности	32
Вопросы для самоконтроля	35
Основные теоретические аспекты для обсуждения	35
Список рекомендуемой литературы	36
Практические задания	36
Задания для самостоятельной работы	39

3. КЛЮЧЕВЫЕ ПОНЯТИЯ КВАНТИТАТИВНОЙ ЛИНГВИСТИКИ	40
3.1. Проблема репрезентативности лингвистической выборки	41
3.2. Выборочная частота, средняя частота и отклонение от средней частоты	42
3.3. Проблема определения рационального объема выборки	47
Вопросы для контроля	49
Основные теоретические аспекты для обсуждения	50
Список рекомендуемой литературы	50
Практические задания	51
Задания для самостоятельной работы	51
4. КВАНТИТАТИВНЫЕ ИССЛЕДОВАНИЯ ЛЕКСИКИ. ЗАКОН ЦИПФА — МАНДЕЛЬБРОТА	54
4.1. Значение количественных и статистических методов в исследовании лексики	54
4.2. Частота как характеристика употребительности слова в тексте. Частотные словари, модель «ранг-частота»	56
4.3. Закон Ципфа. Уточнение закона Ципфа: закон Ципфа — Мандельброта	58
4.4. Другие закономерности в функционировании лексики	63
Вопросы для контроля	65
Основные теоретические аспекты для обсуждения	65
Список рекомендуемой литературы	66
Практические задания	66
Задания для самостоятельной работы	67
5. КОРПУСНАЯ ЛИНГВИСТИКА. НАЦИОНАЛЬНЫЕ КОРПУСЫ ТЕКСТОВ	72
5.1. Корпусная лингвистика как наука. Понятие «корпус текста»	72
5.2. Классификация корпусов	74
5.3. Особенности применения корпусов текстов	78
5.4. Национальные корпусы текстов	80
Вопросы для контроля	84
Основные теоретические аспекты для обсуждения	84
Список рекомендуемой литературы	85
Практические задания	85
Задания для самостоятельной работы	90

6. НАПРАВЛЕНИЯ ЛИНГВИСТИКИ, ИСПОЛЬЗУЮЩИЕ СТАТИСТИЧЕСКИЙ АНАЛИЗ ТЕКСТА. АТРИБУЦИЯ	94
6.1. Понятие стилеметрии и атрибуции. Область их применения	95
6.2. История развития автороведения и стилеметрии	97
6.3. Основные аспекты процесса атрибуции текста	101
6.4. Обзор программного обеспечения для идентификации авторства текстов	102
Вопросы для контроля	104
Основные теоретические аспекты для обсуждения	104
Список рекомендуемой литературы	105
Практические задания	105
7. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАБОТКЕ ТЕКСТОВ. АВТОМАТИЧЕСКОЕ РЕФЕРИРОВАНИЕ И АННОТИРОВАНИЕ ТЕКСТОВ	110
7.1. Обработка текстов на естественных языках и искусственный интеллект	110
7.2. Реферат и аннотация текста. Общие понятия	114
7.3. Автоматическое реферирование и аннотирование	116
7.4. Системы автоматического реферирования и аннотирования текстов	121
Вопросы для контроля	122
Основные теоретические аспекты для обсуждения	122
Список рекомендуемой литературы	123
Практические задания	123
Задания для самостоятельной работы	129
ЗАКЛЮЧЕНИЕ	139
ГЛОССАРИЙ	140
БИБЛИОГРАФИЯ	147