

2.3. Размеченный	140
2.4. Прагматически ориентированный	141
3. Какие бывают корпуса	141
3.1. Параллельные корпуса	142
3.2. Корпусы устной речи	143
4. Разметка корпусов	143
4.1. Средства разметки	143
4.2. Лингвистическая разметка	144
5. Интернет как корпус	146
6. Сервис корпусного менеджера	148
7. Как сделать корпус самому	150
8. Корпусы как инструмент будущего	151
<i>Литература</i>	152
<i>Электронные ресурсы</i>	154
Часть 2. Направления	156
Глава 1. Машинный перевод	156
1. Три подхода к машинному переводу	156
2. Перевод на основе правил	158
2.1. Три способа перевода с помощью правил	158
2.2. Трансферный подход	159
2.3. Пример словарей и грамматик компании PROMT	160
3. Статистический машинный перевод	162
3.1. Главная формула перевода	162
3.2. Модель языка и цепи Маркова	163
3.3. Оценка максимального правдоподобия	166
3.4. Методы сглаживания	168
3.5. Модель перевода	170
4. Гибридный перевод	182
5. Методы оценки качества перевода	183
6. Некоторые современные системы машинного перевода	184
<i>Литература</i>	188
<i>Электронные ресурсы</i>	188

Глава 2. Информационный поиск	190
1. Где ищем?	190
2. Что ищем?	191
3. Как ищем?	192
3.1. Индекс	192
3.2. В идеальном мире	194
3.3. Тем временем в реальности	194
4. Что такое хорошо?	196
4.1. Релевантность, полнота, точность	196
4.2. Фильтрация и ранжирование	197
4.3. Факторы ранжирования	198
4.4. Оценки релевантности	199
4.5. Не все слова одинаково полезны	199
5. А где же лингвистика?	203
5.1. Стандартные запчасти	203
5.2. Расширения	204
5.3. Расстояния	206
5.4. Еще немного поисковой лингвистики	207
<i>Литература</i>	<i>209</i>
<i>Электронные ресурсы</i>	<i>210</i>
Глава 3. Извлечение информации	211
1. Какую информацию извлекаем?	211
2. Распознавание сущностей	214
2.1. Какие сущности извлекаем	214
2.2. Зависимость категории от контекста	215
2.3. Неоднозначность идентификации	216
2.4. Концептуальные сложности	216
2.5. Разрешение анафоры и кореферентности	217
2.6. Установление референта	218
2.7. Автоматические подходы	218
2.8. Использование экстратекстуальных сигналов	220
3. Извлечение отношений	222
3.1. Какие отношения извлекаем	222
3.2. Обучение моделей на размеченных текстах	223
3.3. Полуавтоматическое создание размеченного корпуса	224
3.4. Временное измерение	225

4. Извлечение событий	226
5. Для тех, кто хочет попробовать сам	230
<i>Литература</i>	231
<i>Электронные ресурсы</i>	232
Глава 4. Диалоги и чат-боты	233
1. Компьютер притворяется человеком	233
2. Особенности диалога на естественном языке	234
3. Архитектура диалоговых систем	235
3.1. Модуль понимания естественного языка	236
3.2. Диалоговый менеджер	236
3.3. Модуль генерации естественного языка	237
4. Как работают чат-боты	238
4.1. Имитация беседы	238
4.2. Язык AIML и другие подходы	239
5. Обучение диалоговых систем на реальных диалогах	241
6. Углубление диалога	242
<i>Литература</i>	243
<i>Электронные ресурсы</i>	244
Глава 5. Анализ тональности	245
1. Компьютер отслеживает чувства	245
2. С чего начинается оценка?	246
3. Как измерить тональность текста	248
3.1. Подход с использованием правил и словарей	248
3.2. Подход с использованием машинного обучения	253
4. Как это выглядит на практике	254
5. Оценка качества работы алгоритмов	255
<i>Литература</i>	256
<i>Электронные ресурсы</i>	258
Глава 6. Компьютерная текстология	259
1. Что такое текстология	259
2. Этапы текстологического исследования рукописной традиции	260

2.3. Размеченный	140
2.4. Прагматически ориентированный	141
3. Какие бывают корпуса	141
3.1. Параллельные корпуса	142
3.2. Корпусы устной речи	143
4. Разметка корпусов	143
4.1. Средства разметки	143
4.2. Лингвистическая разметка	144
5. Интернет как корпус	146
6. Сервис корпусного менеджера	148
7. Как сделать корпус самому	150
8. Корпусы как инструмент будущего	151
<i>Литература</i>	152
<i>Электронные ресурсы</i>	154
Часть 2. Направления	156
Глава 1. Машинный перевод	156
1. Три подхода к машинному переводу	156
2. Перевод на основе правил	158
2.1. Три способа перевода с помощью правил	158
2.2. Трансферный подход	159
2.3. Пример словарей и грамматик компании PROMT	160
3. Статистический машинный перевод	162
3.1. Главная формула перевода	162
3.2. Модель языка и цепи Маркова	163
3.3. Оценка максимального правдоподобия	166
3.4. Методы сглаживания	168
3.5. Модель перевода	170
4. Гибридный перевод	182
5. Методы оценки качества перевода	183
6. Некоторые современные системы машинного перевода	184
<i>Литература</i>	188
<i>Электронные ресурсы</i>	188

Глава 2. Информационный поиск	190
1. Где ищем?	190
2. Что ищем?	191
3. Как ищем?	192
3.1. Индекс	192
3.2. В идеальном мире	194
3.3. Тем временем в реальности	194
4. Что такое хорошо?	196
4.1. Релевантность, полнота, точность	196
4.2. Фильтрация и ранжирование	197
4.3. Факторы ранжирования	198
4.4. Оценки релевантности	199
4.5. Не все слова одинаково полезны	199
5. А где же лингвистика?	203
5.1. Стандартные запчасти	203
5.2. Расширения	204
5.3. Расстояния	206
5.4. Еще немного поисковой лингвистики	207
<i>Литература</i>	<i>209</i>
<i>Электронные ресурсы</i>	<i>210</i>
Глава 3. Извлечение информации	211
1. Какую информацию извлекаем?	211
2. Распознавание сущностей	214
2.1. Какие сущности извлекаем	214
2.2. Зависимость категории от контекста	215
2.3. Неоднозначность идентификации	216
2.4. Концептуальные сложности	216
2.5. Разрешение анафоры и кореферентности	217
2.6. Установление референта	218
2.7. Автоматические подходы	218
2.8. Использование экстратекстуальных сигналов	220
3. Извлечение отношений	222
3.1. Какие отношения извлекаем	222
3.2. Обучение моделей на размеченных текстах	223
3.3. Полуавтоматическое создание размеченного корпуса	224
3.4. Временное измерение	225

4. Извлечение событий	226
5. Для тех, кто хочет попробовать сам	230
<i>Литература</i>	231
<i>Электронные ресурсы</i>	232
Глава 4. Диалоги и чат-боты	233
1. Компьютер притворяется человеком	233
2. Особенности диалога на естественном языке	234
3. Архитектура диалоговых систем	235
3.1. Модуль понимания естественного языка	236
3.2. Диалоговый менеджер	236
3.3. Модуль генерации естественного языка	237
4. Как работают чат-боты	238
4.1. Имитация беседы	238
4.2. Язык AIML и другие подходы	239
5. Обучение диалоговых систем на реальных диалогах	241
6. Углубление диалога	242
<i>Литература</i>	243
<i>Электронные ресурсы</i>	244
Глава 5. Анализ тональности	245
1. Компьютер отслеживает чувства	245
2. С чего начинается оценка?	246
3. Как измерить тональность текста	248
3.1. Подход с использованием правил и словарей	248
3.2. Подход с использованием машинного обучения	253
4. Как это выглядит на практике	254
5. Оценка качества работы алгоритмов	255
<i>Литература</i>	256
<i>Электронные ресурсы</i>	258
Глава 6. Компьютерная текстология	259
1. Что такое текстология	259
2. Этапы текстологического исследования рукописной традиции	260

3. Компьютер в работе текстолога	265
3.1. Автоматическое сравнение рукописей	265
3.2. Компьютерная классификация рукописей	266
<i>Литература</i>	271
<i>Электронные ресурсы</i>	272

Глава 7. Квантитативная лингвистика:

что можно сосчитать в языке и речи?	273
1. Буквы и звуки: как определить, на каком языке написан текст? — Дешифровка	273
2. Морфемы: как оценить сложность языка? — Типология	275
3. Части речи: можно ли определить, о чем текст? — Стилеметрия	277
4. Сто слов: как определить возраст языков? — Глоттохронология	279
5. Слова, слова, слова: сколько слов мы знаем и сколько нам нужно знать? — Частотные словари	283
6. Порядок, строй, парадигма: насколько стройна грамматика? — Квантитативная морфология	289
<i>Литература</i>	291
<i>Электронные ресурсы</i>	292

Глава 8. Речевое воздействие и манипулирование..... **294**

1. Что такое речевое воздействие?	294
2. Кто и зачем изучает речевое воздействие?	296
3. Разновидности речевого воздействия	297
4. Языковое манипулирование: приемы и ресурсы	301
<i>Литература</i>	307
<i>Электронные ресурсы</i>	308

Указатель терминов..... **309**