

# ОСНОВЫ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ

*Автор программы проф. Ю.Н.Марчук*

Данный курс (32 часа) рассчитан на повышение осведомленности специализирующихся в компьютерной лингвистике, которая к настоящему времени сложилась как самостоятельная прикладная лингвистическая дисциплина. Специалист по деловой прозе обязательно должен иметь дело с современными компьютерными программами обработки естественно-языковых текстов, от простых программ типа корректоров орфографии до сложных систем автоматической обработки текстов, например, систем машинного перевода и автоматического информационного поиска, экспертных систем и других прикладных программ, которые, с одной стороны, моделируют искусственный интеллект в аспекте работы с обширными базами данных, а, с другой, требуют обязательных знаний современных аспектов компьютерной лингвистики, поскольку все эти системы базируются на естественном языке и на алгоритмах обработки языковых сведений.

Предполагается, что студент знает основы языкознания и начала деловой прозы, состав и структуру делового документа, а также один из иностранных языков. Начальные разделы курса посвящены общей роли информатизации современного общества, составу современных текстов с точки зрения лексики, преобладанию информатических текстов в массовой коммуникации. Вводятся основные понятия информатики и само представление об информатике как науке о наиболее общих закономерностях построения и преобразования современной модели мира.

Лингвистические основы компьютерной лингвистики строятся по принципу укрупнения рассматриваемых лингвистических единиц – части слова (автоматический морфологический анализ и синтез), слово (лексический уровень языка в автоматической обработке, лингвистический знак, денотативные аспекты слова), словосочетание (основной носитель информации в компьютерной лингвистике), предложение (способы изображения, представления и автоматического обнаружения синтаксической структуры), текст (способы кодирования смыслового содержания текста и автоматической его обработки).

Каждая единица названных уровней имеет свои особенности в информатике как с точки зрения формы, так и в содержании. В последнем особое место занимают способы инвентаризации экстралингвистических знаний и работы с ними в рамках компьютерной лингвистики.

Особо рассматриваются основные проблемы практической компьютерной лингвистики – машинный перевод, информационный поиск, экспертные системы, ведение деловой документации. Предполагается, что студент имеет представление об основных приемах работы с компьютером, поэтому собственно программирование в круг данного курса не входит. Однако для освоения содержания курса не требуется каких-либо глубоких знаний по программированию и навыков работы с компьютером.

Определенное внимание в курсе уделяется использованию формальных грамматик в работе с лингвистическими данными разных уровней. При этом используются новые и новейшие достижения отечественной и мировой науки, оглашенные на семинарах и международных конференциях по проблеме. В качестве методической и теоретической основы данного курса используются фундаментальные положения прикладной филологии, разработанные на кафедре общего и сравнительно-исторического языкознания филологического факультета МГУ, а также смежных наук (культуроведения, теоретического и общего языкознания, терминоведения, математической лингвистики, количественной и алгоритмической лингвистики, машинного перевода и искусственного интеллекта и др.).

В программе, которая приводится ниже, даются основные темы курса с указанием предполагаемого числа часов, и основная литература, используемая в данном разделе. Предусматривается также проведение спецсеминаров по теме и по отдельным разделам – практических занятий в компьютерном классе.

## ПРОГРАММА

**Информатизация современного общества.** Роль информатики в социальной сфере, в науке, технике, деловом общении. Совершенствование массовой и индивидуальной коммуникации.

**Понятие об алгоритме.** Лингвистический алгоритм. Нечеткие множества. Лингвистический знак и математический знак.

**Ввод языковой информации в компьютер.** Распознавание устной речи. Автоматический морфологический анализ. Словарь основ и словарь словоформ. Графематический уровень текста и работа с ним.

**Лексический уровень языка.** Машинные словари и терминологические банки данных. Тезаурусы и их роль в информационном поиске и организации информации, в обучении и систематизации знаний. Автоматический контекстологический словарь.

**Автоматический синтаксический анализ.** Предложение и словосочетание. Автоматический семантический анализ и синтез. Проблемы генерации предложений и текста.

**Современный машинный перевод.** Информационный поиск. Системы распознавания устной речи. Экспертные системы. Моделирование человеческого интеллекта.

**Перспективы развития компьютерной лингвистики.** Ближайшие и отдаленные задачи. Связь развития компьютерной лингвистики с прогрессом в компьютерных науках.

## **ЛИТЕРАТУРА** *(К КАЖДОМУ РАЗДЕЛУ)*

1. Гуревич П.С. Культурология. – М.: Знание, 1996, 286.
2. Златоустова Л.В., Королев Э.И., Марчук Ю.Н., Поликарпов А.А. Актуальные аспекты развития прикладной лингвистики // Вестник МГУ, серия 2, филология, 5/1989, с. 55–64.
3. Рождественский Ю.В. Введение в культуроведение. – М.: ЧеРо, 1996, 285 с.
4. Рождественский Ю.В. Введение в общую филологию. – М.: Высшая школа, 1979, 222 с.
5. Шемакин Ю.И. Начала компьютерной лингвистики. – М.: МГОУ, 1992, 112 с.
6. Зубов А.В., Зубова И.И. Основы лингвистической информатики. – Минск: МГПИИЯ, 1992, 137 с.
7. Пиотровский Р.Г. Инженерная лингвистика и теория языка. – Л.: Наука, 1979, 111 с.
8. Белоногов Г.Г., Новоселов А.П. Автоматизация процессов накопления, поиска и обобщения информации. – М.: Наука, 1979, 255 с.
9. Нелюбин Л.Л. Компьютерная лингвистика и машинный перевод. – М.: ВЦП, 1991, 151 с.
10. Денисов П.Н. Лексика русского языка и принципы ее описания. – М.: Русский язык, 1993, 245 с.
11. Марчук Ю.Н. Вычислительная лексикография. – М.: ВЦП, 1976, 175 с.
12. Труды Международного семинара “Диалог 96” по компьютерной лингвистике и ее приложениям / Под ред. Нариньяни А.С. – М., 1996, 305 с.
13. Актуальные проблемы российского языкознания. 1992–1996. – М., ИНИОН РАН, 1997, 204 стр.
14. Диалог 97. Труды Международного семинара по компьютерной лингвистике и ее приложениям. – М., 1997, 315 с.
15. Кривоносов А.Т. Язык, логика, мышление. – М. – Нью-Йорк, 1996, 682 с.
16. Карпов В.А. Язык как система. – Минск: Вышэйшая школа, 1992, 301 с.
17. Марчук Ю.Н. Проблемы машинного перевода. – М.: Наука, 1983, 232 с.
18. Его же: Методы моделирования перевода. – М.: Наука, 1985, 202 с.
19. Потапова Р.К. Речь: коммуникация, информация, кибернетика. – М.: Радио и связь, 1997, 527 с.
20. Диалог 97. Труды Международного семинара по компьютерной лингвистике и ее приложениям. – М., 1997, 315 с.