

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

Государственное образовательное учреждение высшего профессионального образования «Уральский государственный университет им. А.М. Горького»

ИОНЦ «Русский язык»

филологический факультет

кафедра современного русского языка

Основы прикладной лингвистики

Программа дисциплины

Подпись руководителя ИОНЦ

Дата

**Екатеринбург
2007**

I. Введение

1. Цель дисциплины — сформировать у студентов системное представление о практической приложимости и перспективах развития прикладной лингвистики. Принципы автоматической обработки текста на разных уровнях предполагают высокую степень формализации языка, что существенно влияет на общетеоретические представления. Многие лингвистические исследования сегодня невозможны без привлечения масштабного анализа большого количества данных. Современный специалист должен понимать применимость своего теоретического и практического опыта в любых сферах деятельности, в которых востребованы лингвистические знания.

2. Задачи дисциплины

Ознакомить студентов с направлениями и методами прикладной лингвистики. Определить базисные лингвистические знания, необходимые для разработки программ автоматической обработки текста и привести примеры существующих программ. Обсудить теоретические проблемы, возникающие при формализации языковых данных. Дать представление о сегодняшнем состоянии и дальнейшем развитии прикладной лингвистики.

3. Место дисциплины в системе высшего профессионального образования (какие дисциплины используются в качестве основы для данной и для каких используется данная дисциплина). Данная дисциплина, с одной стороны, базируется на общелингвистической информации, которую студенты получают в рамках курсов «Современный русский язык», «Введение в языкознание» и др. С другой стороны, синтетический характер курса предполагает привлечение информации естественно-научных дисциплин: информатики, математики, статистики.

4. Требования к уровню освоения содержания курса (приобретаемые компетенции, знания, умения, навыки).

Общая ориентация в направлениях и методах прикладной лингвистики. Представление о современных возможностях автоматической обработки текстов

и умение работать с некоторыми программами . Умение грамотно сформулировать техническое задание для прикладного проекта, для чего необходима языковая и чисто логическая компетенция, связанная с необходимостью формализации языковых данных.

5. Методическая новизна курса (новые методики, формы работы, авторские приемы в преподавании курса).

Изменчивость курса в соответствии с развитием отраслей прикладной лингвистики, использование демонстраций на лекциях. Новые по характеру задания для самостоятельной работы (формализация словарных данных, лингвистические решения для снятия неоднозначностей и т.п.).

II. Содержание курса

1. Разделы курса, темы, их краткое содержание

I. Введение в прикладную лингвистику (ПЛ).

1. Общие сведения о ПЛ. Задачи и методы, области приложения ПЛ.

2. Широкое и узкое понимание интересов ПЛ. Традиционные задачи, связанные с разработкой систем письма, лексикографией, орфоэпией, развитием терминосистем, транскрипцией и транслитерацией. Современные актуальные прикладные задачи: лингвистическая экспертиза, атрибуция устного и письменного текста и др.

Узкое понимание ПЛ как направления, связанного с автоматической обработкой текста (АОТ), необходимой для распознавание устной речи, машинного перевода, автоматической классификации документов, аннотирования, кодирования текстов и т.п.

II. Автоматический анализ и обработка речи.

1. Фонетический и графический уровень. Развитие компьютерного интерфейса. Основные проблемы анализа и синтеза речи. Классификация систем распознавания речи и примеры работающих систем. Синтез устной речи: проблемы и решения. Методы синтеза речи.

Распознавание письменного текста. Современные программы распознавания символов. Чисто технические и собственно лингвистические аспекты интерпретации.

Автоматическая обработка печатного текста в современных текстовых процессорах и издательских системах (АОТ в узком смысле). Поиск и преобразование символов и форматирования в процессоре Word. Возможности макропрограммирования.

2. Машинная дериватология и морфология. Вводные замечания.

Использование дериватологического анализа. Возможности формализации словообразовательного анализа и синтеза. Словообразовательные аспекты в морфологических и семантических модулях.

Виды морфоанализа. Требования к современному алгоритму морфоанализа. Существующие системы русского морфоанализа. Постморфологические модули. Общие проблемы морфоанализа.

3. Автоматический синтаксический анализ. Вводные замечания

Сегментация предложения. Способы изображения и выявления синтаксических структур. Языковые средства, применяемые при синтаксическом моделировании. Синтаксические (языковые) ограничения, применяемые в прикладных моделях анализа. Общие проблемы парсинга.

4. Автоматический семантический анализ. Вводные замечания. Общие принципы анализа. Работа на уровне словосочетания и фразы. Общая категоризация лексики и семантические характеристики слов. Семантическое отношение и возможности его представления. Работа с тематической структурой текста. «Семантическая сеть». Общие проблемы семантического анализа и перспективы его использования.

III. Место прикладной лингвистики среди направлений, задействующих информационные технологии.

1. Прикладная лингвистика и искусственный интеллект (ИИ).
Элементы ИИ в современных системах.

2. ПЛ и гипертекстовые технологии. Понятие "гипертекст" в информатике и филологии. Языки разметки гипертекста и программы-браузеры (общий обзор). Перспективы развития лингвистических основ гипертекстовых технологий.

IV. Заключение. Перспективы развития прикладной лингвистики. Взаимовлияние теоретической и прикладной лингвистики. «Философия» ПЛ.

2. Темы лабораторных, семинарских занятий и коллоквиумов (если предусмотрены)

3. Перечень примерных контрольных вопросов и заданий для самостоятельной работы

Проблемы автоматизации:

- 1) фонетического анализа речи;
- 2) графематического анализа речи;
- 3) словообразовательного анализа речи;
- 4) морфологического анализа речи;
- 5) синтаксического анализа речи;
- 6) семантического анализа речи.

III. Распределение часов курса по темам и видам работ

№ п/п	Наименование разделов и тем	ВСЕГ О (часов)	Аудиторные занятия (час)		Самостоятель ная работа
			в том числе		
			Лекции	Практичес кие (семинары , лаборатор ные работы)	
1.	Введение в прикладную лингвистику. Широкое и узкое понимание задач ПЛ.	2	2		
2.	Автоматический анализ и обработка речи. Фонетический и графический уровень.	6	6		+
3.	Машинная дериватология и морфология.	8	8		+
4.	Автоматический синтаксический анализ.	4	4		+
5.	Автоматический семантический анализ.	6	6		+
6.	Прикладная лингвистика и искусственный интеллект.	4	4		
7.	Прикладная лингвистика и гипертекстовые технологии.	2	2		
8.	Перспективы развития прикладной	2	2		

	лингвистики. Взаимовлияние теоретической и прикладной лингвистики.				
	ИТОГО:	34	34		

IV. Форма итогового контроля

Теоретико-практический зачет, реферат, самостоятельная работа.

V. Учебно-методическое обеспечение курса

1. Рекомендуемая литература (основная)

Баранов А. Н. Введение в прикладную лингвистику. М., 2001.

Баранов А. Н. Автоматизация лингвистических исследований: корпус текстов как лингвистическая проблема // Русистика сегодня. 1998. № 1–2.

Богданов В. В. Статистические концепции языка и речи // Статистика речи и автоматический анализ текста. М., 1973. С. 9–19.

Герд А. С. и др. Информационно-поисковый тезаурус как объект лексикографии // Структурная и прикладная лингвистика. Вып. 1. Л., 1978.

Герд А.С. Предмет и основные направления прикладной лингвистики // Прикладное языкознание. СПб., 1996.

Городецкий Б. Ю. Актуальные проблемы прикладной лингвистики // Новое в зарубежной лингвистике. М., 1983.

Городецкий Б. Ю. К созданию Машинного фонда русского языка (определение, применения, актуальные проблемы) // Машинный фонд русского языка: идеи и суждения. М., 1986.

Звегинцев В. А. Теоретическая и прикладная лингвистика. М., 1968.

Использование ЭВМ в лингвистических исследованиях. Киев, 1989.

Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. М., 1981.

Кибрик А. Е. Очерки по общим и прикладным вопросам языкознания. М., 1992.

Лингвистический энциклопедический словарь. М., 1990. Ст.: "Автоматическая обработка текстов", "Количественные методы" и др..

Марчук Ю. Н. Основы компьютерной лингвистики. М., 2000.

Машинный перевод и прикладная лингвистика. М., 1986.

Машинный фонд русского языка: идеи и суждения. М., 1986.

Мельников Г.П. Системология и языковые аспекты кибернетики. М., 1978.

Мельчук И. А. Опыт теории лингвистических моделей «Смысл — Текст»: семантика, синтаксис. М., 1999. (Или: Мельчук И. А. Опыт теории лингвистических моделей «Смысл — Текст». М., 1974.)

Моделирование языковой деятельности в интеллектуальных системах. М., 1987.

Пиотровский Р. Г. Инженерная лингвистика и теория языка. Л., 1979.

Прикладное языкознание. СПб., 1996.

Рождественский, Ю.В., Марчук, Ю.Н., Волков, А.А. Введение в прикладную филологию. М., 1998. (Рождественский Ю. В. и др. Введение в прикладную филологию. М., 1987.)

Турыгина Л. А. Моделирование языковых структур средствами вычислительной техники. М., 1988. Введение. Глава I. Общие сведения о статистических подходах к исследованию языковых структур. М., 1988.

Эндрю А. Искусственный интеллект. М., 1985.

Интернет-сайты, связанные с направлениями прикладной лингвистики: dialog-21.ru, aot.ru, ruscorpora.ru и др.

2. Рекомендуемая литература (дополнительная)

Андреев Д. Н. Статистико-комбинаторные методы в теоретическом и прикладном языковедении. Л., 1967.

Андрющенко В. М. Концепция и архитектура машинного фонда русского языка. М., 1989.

- Апресян Ю. Д. Избранные труды. М., 1995.
- Белоногов, Г.Г., Кузнецов, Б.А. Языковые средства автоматизированных информационных систем. М., 1983.
- Блюменау, Д.И. Проблемы свёртывания научной информации. Л., 1982.
- Богданов В. В. Статистические концепции языка и речи // Статистика речи и автоматический анализ текста. М., 1973. С. 9–19.
- Грязнухина Т.А. и др. Использование ЭВМ в лингвистических исследованиях. Киев, 1990.
- Златоустова Л.В. и др. Актуальные проблемы прикладной лингвистики // Вестник МГУ. Сер. 9. Филология. 1989, № 5.
- Искусственный интеллект. В 3-х кн. М., 1990. — Кн. 1. Системы общения и экспертные системы. Кн. 2. Модели и методы.
- Мальковский, М.Г. Диалог с системой искусственного интеллекта. М., 1986.
- Мельников Г.П. Системология и языковые аспекты кибернетики. М., 1978.
- Минский М. Фреймы для представления знаний // Психология машинного зрения. М., 1978.
- Моделирование языковой деятельности в интеллектуальных системах. М., 1987.
- Морфологический анализ научного текста на ЭВМ. Киев, 1989.
- Москович В.А. Информационные языки. М., 1971.
- Поспелов Д.А. Логико-лингвистические модели в системах управления. М., 1981.
- Прикладная лингвистика и автоматический анализ текста. Тарту, 1988.
- Прикладная лингвистика и информационная технология. М., 1987.
- Прикладная лингвистика. М., 1976.
- Прикладные аспекты лингвистики. М., 1989.

Прикладные и экспериментальные лингвистические процессоры. Новосибирск, 1982.

Промышленные системы машинного перевода. Вып. 20. М., 1991.

Рождественский, Ю.В., Марчук, Ю.Н., Волков, А.А. Введение в прикладную филологию. М., 1998. (Рождественский Ю. В. и др. Введение в прикладную филологию. М., 1987.)

Субботин М.М. Гипертекст. Новая форма письменной коммуникации. // Итоги науки и техники. Сер. Информатика. Т. 18. М., 1994.

Субботин М.М. Новая информационная технология: создание и обработка гипертекстов // НТИ. Сер. 2. 1988. № 5.

Трибис Л. И. Об одной модели распознавания лексических значений неоднозначных слов. Статистика речи и автоматический анализ текста. М., 1973. С. 131–180.

Турыгина Л. А. Моделирование языковых структур средствами вычислительной техники. М., 1988. Введение. Глава I. Общие сведения о статистических подходах к исследованию языковых структур. М., 1988.

Шубик С. А. Статистические методы в лингвистике // Статистика речи и автоматический анализ текста. М., 1980. С. 52–63.

3. Перечень обучающих, контролирующих компьютерных программ, кино- и телефильмов, мультимедиа и т.п.

Демонстрационные версии программ автоматической обработки текстов, презентации лекций.

VI. Ресурсное обеспечение (если требуется)

1. Лаборатории (в том числе, вузовско-академические), музеи, гербарии, биостанция, ботанический сад, астрономическая обсерватория и т.д.

Лаборатория компьютерной лексикографии УрГУ.

2. Базы данных

3. Компьютерные классы