

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

Методические указания к изучению дисциплины

Подпись руководителя ИОНЦ

Дата

**Екатеринбург
2007**

Утверждено учебно-методической комиссией филологического факультета

Методические указания состоят из следующих частей:

1. Введение

2. Программа учебного курса «Основы прикладной лингвистики», включающая содержание лекционного курса и материалы для самостоятельной работы студента по основным темам, сформулированные в тезисном виде.

3. Списки рекомендуемой учебной и научной литературы (основной и дополнительной).

4. Вопросы и задания для самоконтроля

5. Перечень вопросов к зачету.

ВВЕДЕНИЕ

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.

Курс решает следующие задачи: 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.

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.

Студенты должны получить общую ориентацию в направлениях и методах прикладной лингвистики, представление о современных возможностях автоматической обработки текстов, умение работать с некоторыми программами. Желательно научиться формулировать техническое задание для прикладного проекта, для чего необходима языковая и чисто логическая компетенция, связанная с необходимостью формализации языковых данных.

Данный курс должен постоянно меняться в соответствии с быстрым развитием отраслей прикладной лингвистики, появлением новых программ автоматической обработки речи, лингвистических корпусов и т.п.

ПРОГРАММА УЧЕБНОГО КУРСА «ОСНОВЫ ПРИКЛАДНОЙ ЛИНГВИСТИКИ»

Содержание лекционного курса и материалы для самостоятельной работы студента

I. Введение в прикладную лингвистику (ПЛ)

Содержание лекций

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

Содержание лекций

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.

Материалы для работы студента

Анализ устной речи

Компьютерный интерфейс: *ввод, преобразование, вывод данных.*

Проблемы анализа и синтеза речи, обработка информации:

понять — среагировать — преобразовать и выдать.

Спектрограммы (сонограммы) речи как единственный источник звуковой информации.

Классификация система распознавания речи:

- по назначению:

командные системы / системы диктовки текста.

- по потребительским качествам:

*диктороориентированные / дикторонезависимые,
распознающие отдельные слова / распознающие слитную речь.*

- по функционированию:

простейшие детекторы / экспертные системы / модели принятия решения.

Требования к системам:

- распознавание слитной речи,
- устойчивость к изменениям голоса, физическому состоянию говорящего, стилю говорения, помехам и фоновому шуму, характеристикам компьютеров, характеристикам телефонной сети, характеристикам микрофонов и т. п.,
- отсутствие ограничений на задания и словарь.

Синтез речи-1. Моделирование речевого аппарата:

артикуляторный синтез.

Синтез речи-2. Моделирование звукового сигнала:

а) формантный синтез по правилам,

б) компилятивный синтез

Формантный синтез:

генерация сигнала источником (синтезатором) — фильтрация исходного сигнала — моделирование голосового тракта.

Компилятивный синтез:

«склейка» речи из имеющихся единиц (см. пример с сайта кафедры фонетики СПбГУ).

Анализ письменной речи

Распознавание письменного текста.

Проблемы интерпретации символов:

- *разные начертания букв, «шум», неровность строк,*
- *опечатки (в том числе правдоподобные).*

Важные параметры рисунка:

- *яркость,*
- *контрастность,*
- *разрешение.*

Методы распознавания текста:

- *шаблонные,*
- *признаковые,*
- *структурные,*
- *структурно-пятенные.*

Сложные графические системы (вербальные и невербальные). Распознавание символов. Формы, иконические элементы, нотная запись и т.п.

2. Машинная дериватология и морфология

Содержание лекций

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.

Материалы для работы студента

Словообразование (дериватология)

Проблема идиоматичности словообразовательной структуры:

Подснежник = предмет (-ник), находящийся ниже (под-) [снега].

Снег – минимальная непроеизводная единица.

Проблема неуниверсальности СО-моделей:

тигр + -иц- = тигрица

тигр + -енок = тигренок

Аналогично:

лев – левица – левенок (?)

бык – бычица – бычонок (?)

Словообразовательные структуры проще давать списком, гнездами, тем более что сегодня можно позволить себе затратность в плане компьютерных ресурсов.

С другой стороны, описание аффиксов может быть полезно в процедуре морфологического и семантического анализа слов, отсутствующих в базе данных:

** компьютерность (мышления) ← компьютерн-ый + -ость.*

Применимость словообразовательного анализа: *в практике обучения РКИ, в системах машинного перевода, как дополнение разных видов АОТ.*

Ограниченность использования из-за перечисленных проблем.

Морфологический анализ

Соотносимые понятия: *лемматизация, стемминг, жарг. «морфология»).*

Лемма – исходная словоформа. Stem (англ.) – основа слова.

Необходимость морфоанализа:

- *обобщение* грамматических форм (*ректор – ректора – ректору* и т.п.) в различных программах, связанных с АОТ,
- *разведение* форм, относящихся к разным парадигмам (*disambiguation*, снятие неоднозначностей): *начала – начать, начала – начало.*

Морфологическая классификация языков:

- *изолирующие,*
- *агглютинирующие,*
- *инкорпорирующие,*
- *флективные.*

Грамматический словарь А. А. Зализняка. Примеры словарных статей:

1/бол<еть нсв нп 1a (_о живом существе_)

2/бол<еть нсв нп 5b (_о частях тела_)

б>олеутол<ение с 7a

б>олеутол<яющий п 4а

болив<ар м 1а (_ шляпа _)

бол<ивар м 1а (_ денежная единица _)

*болив<иец мо 5*а*

...

Особенности именной парадигмы падежей и условные обозначения:

- *nom* — именительный падеж (*голова, сын, степь, сани, который*),
- *gen* — родительный падеж (*головы, сына, степи, саней, которого*),
- *dat* — дательный падеж (*голове, сыну, степи, саням, которому*),
- *acc* — винительный падеж (*голову, сына, степь, сани, который / которого*),
- *ins* — творительный падеж (*головой, сыном, степью, санями, которыми*),
- *loc* — предложный падеж (*[о] голове, сыне, степи, санях, котором*),
- *gen2* — второй родительный падеж (*чашка чаю*),
- *acc2* — второй винительный падеж (*постричься в монахи; по два человека*),
- *loc2* — второй предложный падеж (*в лесу, на оси*),
- *voc* — звательная форма (*Господи, Серёж, ребят*),
- *adnum* — счётная форма (*два часа́, три шара́*).

Особенности глагольной парадигмы:

- инфинитив,
- личные формы,
- причастия,
- деепричастия.

Итого около 150 глагольных форм + возвратные (≈ 80).

Другие проблемы морфоанализа:

[Пример чисто флективного класса: *стол – стола – столы*]

- словоизменение с чередованием: *потолок – потолка – потолки* (?)

● словоизменение с супплетивизмом: *человек – человека – человеки (?)*,
меня – меняю –мень (?).

По словарю Зализняка:

● 8 типов именного словоизменения (с подтипами),
● 16 типов глагольного словоизменения (с подтипами),
● наличие «изолированной» лексики, не подчиняющейся общим моделям (глаголы типа *БЫТЬ*) и прочие проблемы.

● кроме того, «многословные» слова, фразеологизмы:

в течение, в виде, без устали, на ощупь, до упаду,
крутиться (крутится, кручусь, крутилась...) как белка в колесе,
положить (положишь, положил...) зубы на полку,
черт с... (тобой, ним, этим, работой...)

Пример работы морфоанализатора (Mystem):

Все{все=S, ед, сред, неод=(им | вин) | все=S, мн, од=им | весь=A=(
им, ед, сред | им, мн | вин, ед, сред | вин, мн, неод) | все=PART=}смешало
сь{смешиваться=V=прош, ед, изъяв, сред, сов}

в{в=PR=}

доме{дом=S, муж, неод=пр, ед}

Облонских{облонский?=A=(род, мн | вин, мн, од | пр, мн)}

Проблема грамматической омонимии:

стекла – стечь, стекла – стекло,

*при – при (предлог), при – переть (императив), при – пря (напр. род. п. ед.
ч.) = устар. ссора, спор, состязание; то же, что распря.*

Виды морфоанализа:

● со словарем основ,
● со словарем словоформ,
● методом логического умножения,

- без словаря (с помощью таблиц).

Чистый стеммер малопродуктивен для русского языка:

* *кровать* – *кроватью* – *кроватьеишь* – по типу *читать* (?),

* *кровать* – *кровлю* – *кровишь* – по типу *спать* (?),

* *кровать* – *крую* – *круеишь* – по типу *снова* (?).

Лучше всего использовать комбинированные модели.

Формально-графический принцип анализа (основа + флексия):

мыш-Ь, мыш-И, мыш-И, мыш-Ь, мыш-ЬЮ, мыш-И.

Требования к современному алгоритму:

- объем словаря (покрытие 98-99%),
- словоизменяющий алгоритм,
- эффективная гипотетическая обработка новых слов,
- обучаемость системы,
- технические характеристики.

Снятие неоднозначностей: постморфология.

- Анализ ближайшего контекста:

Л.П.Быков, профессор, зав. кафедрой русской литературы XX в.

- Анализ повторных употреблений:

Быков – Быкова – Быковым и проч.

Собственно синтаксический анализ дополняет морфологический: функция в предложении, особенности согласования, управления и т.п.

Общие проблемы морфоанализа:

- объем словаря (больше или меньше?),
- развитие системы (пополнение и чистка базы).

3. Автоматический синтаксический анализ

Содержание лекций

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.

Материалы для работы студента

Интернациональные эквиваленты: *parser, parsing engine*.

Русские жаргонизмы: *па́рсить* и даже *отпа́рсить* (предложение, текст).

- Проблема синтаксической омонимии:

Делегация из Петербурга прибыла в Москву.

Преступница представлялась им красивой девушкой...

Дожди вызвали аварии на дорогах.

- Проблема порядка слов

Мой отец – хороший специалист.

Тире при инверсии может отсутствовать:

Хороший специалист мой отец.

Ср. также:

Белеет парус одинокий

В тумане моря голубом.

«Нормальный» порядок слов: * *Одинокий парус белеет в голубом тумане моря.*

- Дистантное расположение и эллипсис элементов структуры:

Специалист мой отец хороший.

Прикладная лингвистика, как мне кажется, интересный предмет.

Шапочка знала, что, если дернуть за веревочку, дверь откроется.

Мне два до Москвы.

Я за свечку, свечка – в печку!

1-й этап анализа. Сегментация предложения.

Построение дерева синтагм и дерева сегментов.

Дерево («граф»).

«Синтагма» в теоретической и прикладной лингвистике.

«Сегмент» = *клауза* (фрагмент предложения) в теоретической и прикладной лингвистике.

Выделение сегментов

Вчера выпал снег, заваливший всю область.

Вчера выпал снег, который завалил всю область.

2-й этап. Способы представления синтаксических структур:

- *грамматика зависимостей,*
- *грамматика составляющих.*

Классификация парсеров:

- по способу движения по тексту: *непрерывные / циклические.*
- по выявляемым отрезкам: *интегральные / локальные.*

Языковые средства, применяемые при парсинге:

словоизменение, учет морфологических признаков, морфоанализ.

Классический пример Л.В.Щербы:

Глок-ая куздр-а штеко будланул-а бокр-а и кудряч-ит бокрѐнк-а.

Связь морфологических и синтаксических явлений:

Выбрал президента – выбрал «Президент» (сыр, шоколад, отель...).

Другие языковые средства, применяемые при парсинге:

- языковые универсалии (члены предложения и и т.п.),
- служебные слова,
- пунктуация,
- порядок слов,
- интонация (?).

Интонация как логико-синтаксическое средство

Пример А. Е. Кибрика:

В этой гимназии учился впоследствии всемирно известный киноартист.

Полезные синтаксические ограничения:

- «проективность» предложения,
- зависимость любой синтаксической структуры от одной вершины,

- один субъект и один предикат в структуре сегмента, кроме случаев сочинительных отношений между субъектами и предикатами.

Общие проблемы парсинга:

- сложность языковой системы,
- свободный порядок слов,
- неуниверсальность синтаксических моделей в плане анализа разных языков,
- коррекция результатов человеком («ручная» работа).

4. Автоматический семантический анализ

Содержание лекций

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.

Материалы для работы студента

Текст и семантика. Семантический анализ.

Генеральная цель – представить структуру предложения и текста в виде однозначной формальной схемы.

Возможный последующий семантический синтез:

- перефразирование,
- автоматический перевод,
- выделение ключевых слов («терминов»),
- построение парадигматических сетей (в первую очередь синонимов),
- реферирование, в том числе синтез нового текста, связанного с содержанием исходного.

Множественность проблем:

- идиоматичность семантики $\Rightarrow$ неформализованность данных,
- пересекаемость семантических парадигм:
храпеть – спать (физиология),
храпеть – издавать звуки (звучание);
- асимметричность языкового знака — многозначность («омонимия») и синонимия,

- наличие составных наименований,
- контекстная обусловленность, влияние речевой ситуации,
- изменяемость действительности.

Идеографическая лексикография как перспективное направление в плане последующего семантического анализа.

Пример структуры идеографического словаря:

[Большой толковый словарь русских глаголов: Идеографическое описание. Синонимы. Антонимы. Английские эквиваленты. Сер. «Фундаментальные словари». Под ред. проф. Л. Г. Бабенко. М., 2007. 576 с.]

I. Действие и деятельность

1.1. Движение

1.2. Перемещение объекта

1.3. Помещение

1.4. Физическое воздействие

1.5. Созидательная деятельность

... ..

II. Бытие, состояние, качество

2.1. Бытие

2.2. Качественное состояние

III. Отношение

3.1. Взаимоотношение

3.2. Владение

3.3. Межличностные отношения

3.4. Социальные отношения

Как применять идеографическую систему?

Лексикологические установки:

- связь синтагматики с парадигматикой,
- контекст и «семная поддержка»,
- валентность слова и контекстные факторы.

Пример для контекстологического анализа многозначного слова.

Назначить

1. Заранее наметить, определить (срок, совершение какого-л. действия и т.п.)

3. *кого кем.* Поставить, направить на какую-л. должность, работу.

5. *Мед.* Предписать (лечение, прием лекарства и т.п.)

Теоретические проблемы:

- границы словосочетания,
- границы предложения.

Ответы возможны с теоретической и прикладной точек зрения.

Модель «СМЫСЛ $\Leftrightarrow$ ТЕКСТ» (И. Мельчук, А. Жолковский, Ю. Апресян)

— многоуровневая формализованная модель преобразования смысла в текст и обратно.

«Семантический язык». Опора на семантические примитивы.

Слово как структурированная комбинация элементарных заданных смыслов:

стол = *стол* + *маленький*

учительница = *учитель* + *женщина* = *человек* + *учить* + *женщина* = ?)

Семантическая структура предложения. Компоненты пропозиции:

- субъект
- предикат
- актанты

Семантические характеристики: система «Диалинг» (www.aot.ru).

Семантическое отношение

$R(A, B)$, где R – название отношения,

- A – зависимое понятие
- B – управляющее понятие

Принцип: *A является R для B.*

Семантические отношения: система «Диалинг» (aot.ru)

Пример анализа фразы (Диалинг):

Мы разговаривали так, как будто расстались вчера, как будто знали друг друга много лет (М.Булгаков, «Мастер и Маргарита»).

Формализация структуры текста:

- микроструктура,
- макроструктура,
- суперструктура.

Единство темы как важное свойство текста:

- регулярная повторяемость ключевых слов,
- тождество референции (соотношение с одним объектом номинации),
- наличие ситуативных связей.

Семантическая сеть как интегральное представление смысла текста – множество понятий (слов, словосочетаний), связанных по смыслу и выстроенных иерархически.

Пример: программа TextAnalyst (www.analyst.ru).

Выводы. Связь семантического анализа с проблемами искусственного интеллекта. Перспективы использования семантического анализа.

Машинный перевод как наиболее сложное направление в ИИ.

Междисциплинарность МП.

Привлекательность МП:

- АОТ на всех уровнях языка,
- отработка лингвистических теорий,
- лексико-синтаксическая основа,
- «интеллектуальность».

→ МП называют центральным направлением искусственного интеллекта.

Лингвистические проблемы:

- разрешение неоднозначностей,

- текстовый уровень их появления (ситуативная обусловленность текста, экстралингвистические знания),
- возможность построения альтернативных структур текста,
- несовершенство формальных методов анализа текста, необходимость их уточнения,
- обучаемость системы МП.

III. Место прикладной лингвистики среди направлений, задействующих информационные технологии

Содержание лекций

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

Содержание лекций

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

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА (ОСНОВНАЯ)

1. Баранов А. Н. Введение в прикладную лингвистику. М., 2001.
2. Баранов А. Н. Автоматизация лингвистических исследований: корпус текстов как лингвистическая проблема // Русистика сегодня. 1998. № 1–2.
3. Богданов В. В. Статистические концепции языка и речи // Статистика речи и автоматический анализ текста. М., 1973. С. 9–19.

4. Герд А. С. и др. Информационно-поисковый тезаурус как объект лексикографии // Структурная и прикладная лингвистика. Вып. 1. Л., 1978.
5. Герд А.С. Предмет и основные направления прикладной лингвистики // Прикладное языкознание. СПб., 1996.
6. Городецкий Б. Ю. Актуальные проблемы прикладной лингвистики // Новое в зарубежной лингвистике. М., 1983.
7. Городецкий Б. Ю. К созданию Машинного фонда русского языка (определение, применения, актуальные проблемы) // Машинный фонд русского языка: идеи и суждения. М., 1986.
8. Звегинцев В. А. Теоретическая и прикладная лингвистика. М., 1968.
9. Использование ЭВМ в лингвистических исследованиях. Киев, 1989.
10. Караулов Ю.Н. Лингвистическое конструирование и тезаурус литературного языка. М., 1981.
11. Кибрик А. Е. Очерки по общим и прикладным вопросам языкознания. М., 1992.
12. Лингвистический энциклопедический словарь. М., 1990. Ст.: "Автоматическая обработка текстов", "Количественные методы" и др..
13. Марчук Ю. Н. Основы компьютерной лингвистики. М., 2000.
14. Машинный перевод и прикладная лингвистика. М., 1986.
15. Машинный фонд русского языка: идеи и суждения. М., 1986.
16. Мельников Г.П. Системология и языковые аспекты кибернетики. М., 1978.
17. 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.)
18. Моделирование языковой деятельности в интеллектуальных системах. М., 1987.
19. Пиотровский Р. Г. Инженерная лингвистика и теория языка. Л., 1979.
20. Прикладное языкознание. СПб., 1996.

21. 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.)
22. 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.
23. Эндрю А. Искусственный интеллект. М., 1985.
24. Интернет-сайты, связанные с направлениями прикладной лингвистики: dialog-21.ru, aot.ru, ruscorpora.ru и др.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА (ДОПОЛНИТЕЛЬНАЯ)

1. Андреев Д. Н. Статистико-комбинаторные методы в теоретическом и прикладном языковедении. Л., 1967.
2. Андрющенко В. М. Концепция и архитектура машинного фонда русского языка. М., 1989.
3. Апресян Ю. Д. Избранные труды. М., 1995.
4. Белоногов, Г.Г., Кузнецов, Б.А. Языковые средства автоматизированных информационных систем. М., 1983.
5. Блюменау, Д.И. Проблемы свёртывания научной информации. Л., 1982.
6. Богданов В. В. Статистические концепции языка и речи // Статистика речи и автоматический анализ текста. М., 1973. С. 9–19.
7. Грязнухина Т.А. и др. Использование ЭВМ в лингвистических исследованиях. Киев, 1990.
8. Златоустова Л.В. и др. Актуальные проблемы прикладной лингвистики // Вестник МГУ. Сер. 9. Филология. 1989, № 5.
9. Искусственный интеллект. В 3-х кн. М., 1990. — Кн. 1. Системы общения и экспертные системы. Кн. 2. Модели и методы.

10. Мальковский, М.Г. Диалог с системой искусственного интеллекта. М., 1986.
11. Мельников Г.П. Системология и языковые аспекты кибернетики. М., 1978.
12. Минский М. Фреймы для представления знаний // Психология машинного зрения. М., 1978.
13. Моделирование языковой деятельности в интеллектуальных системах. М., 1987.
14. Морфологический анализ научного текста на ЭВМ. Киев, 1989.
15. Москович В.А. Информационные языки. М., 1971.
16. Пospelов Д.А. Логико-лингвистические модели в системах управления. М., 1981.
17. Прикладная лингвистика и автоматический анализ текста. Тарту, 1988.
18. Прикладная лингвистика и информационная технология. М., 1987.
19. Прикладная лингвистика. М., 1976.
20. Прикладные аспекты лингвистики. М., 1989.
21. Прикладные и экспериментальные лингвистические процессоры. Новосибирск, 1982.
22. Промышленные системы машинного перевода. Вып. 20. М., 1991.
23. 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.)
24. Субботин М.М. Гипертекст. Новая форма письменной коммуникации. // Итоги науки и техники. Сер. Информатика. Т. 18. М., 1994.
25. Субботин М.М. Новая информационная технология: создание и обработка гипертекстов // НТИ. Сер. 2. 1988. № 5.
26. Трибис Л. И. Об одной модели распознавания лексических значений неоднозначных слов. Статистика речи и автоматический анализ текста. М., 1973. С. 131–180.

27. 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.

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

ВОПРОСЫ И ЗАДАНИЯ ДЛЯ САМОКОНТРОЛЯ

I. Введение:

1) общее ознакомление с источниками, посвященными автоматическому анализу устной и письменной речи.

II. Автоматизация фонетического анализа речи:

2) выбор материала по автоматическому фонетическому анализу (в том числе из Интернет),

3) ознакомление с программами автоматического анализа и синтеза устной речи,

4) обобщение существующих проблем автоматического фонетического анализа.

III. Автоматизация графематического анализа речи:

5) выбор материала по автоматическому графематическому анализу (в том числе из Интернет),

6) сопоставление различных программа автоматического распознавания символов,

7) работа с программой ABBYY FineReader,

8) обобщение существующих проблем автоматического графематического анализа.

IV. Автоматизация словообразовательного анализа речи:

- 9) выбор материала по автоматическому словообразовательному анализу (в том числе из Интернет),
- 10) ознакомление со словообразовательными модулями программ АОТ,
- 11) обобщение существующих проблем автоматического словообразовательного анализа.

V. Автоматизация морфологического анализа речи:

- 12) Выбор материала по автоматическому морфологическому анализу (в том числе из Интернет),
- 13) работа с программой Mystem,
- 14) работа с программой «Диалинг»,
- 15) ознакомление с возможностями постморфологической обработки текста,
- 16) обобщение достоинств и недостатков программ русского морфоанализа.

VI. Автоматизация синтаксического анализа речи:

- 17) выбор материала по автоматическому синтаксическому анализу (в том числе из Интернет),
- 18) работа со свободно распространяемыми программами парсинга,
- 19) обобщение существующих проблем автоматического синтаксического анализа

VII. Автоматизация семантического анализа речи:

- 20) выбор материала по автоматическому семантическому анализу (в том числе из Интернет),
- 21) реферирование источников по идеографическому представлению данных, понятиям «семантическое отношение» и «семантическая сеть»,
- 22) ознакомление с идеографическими словарями,

23) оценка существующих программ семантического анализа и элементов автоматизированной обработки семантики в различных приложениях,

24) обобщение существующих проблем автоматического семантического анализа.

ПЕРЕЧЕНЬ ВОПРОСОВ К ЗАЧЕТУ

1. Прикладная лингвистика и другие науки. Предпосылки возникновения прикладной лингвистики. Основные направления ПЛ.
2. Области приложения прикладной лингвистики: общая характеристика.
3. Фонетический и графический уровень автоматической обработки текста. Основные проблемы анализа и синтеза устной речи.
4. Синтез устной речи: проблемы и решения. Методы синтеза речи.
5. Методы распознавания письменного текста. Программы оптического распознавания символов.
6. Возможности автоматической обработки печатного текста при верстке в современных текстовых процессорах и издательских системах.
7. Общие задачи автоматического словообразовательного и морфологического анализа.
8. Языковые ограничения использования автоматического дериватологического анализа.
9. Автоматический морфоанализ. Основные принципы и виды морфоанализа.
10. Требования к современным алгоритмам морфоанализа. Существующие системы русского морфоанализа, их достоинства и недостатки.
11. Возможности "постморфологической" обработки текста для установления грамматических признаков слов.
12. Автоматический синтаксический анализ. Общие принципы и проблемы.
13. Сегментация предложения при синтаксическом анализе.

14. Способы изображения и выявления синтаксических структур при автоматическом анализе.
15. Языковые средства автоматического синтаксического моделирования. Языковые ограничения, применяемые в моделях автоматического синтаксического анализа (закон проективности предложения и др.)
16. Автоматический семантический анализ. Общие принципы и проблемы.
17. Категоризация лексики и семантические характеристики слов, используемые при автоматическом семантическом анализе.
18. Понятие "семантическое отношение". Возможности представления семантических отношений.
19. Работа с тематической структурой текста. Понятие "семантическая сеть".
20. Искусственный интеллект как направление, смежное с прикладной лингвистикой. Элементы искусственного интеллекта в системах АОТ.
21. Гипертекст как лингвистическая проблема.
22. Перспективы развития прикладной лингвистики. Методы прикладной и теоретической лингвистики: различие и взаимодействие.