

Министерство образования Республики Беларусь
Белорусский государственный университет
Филологический факультет

УТВЕРЖДАЮ
Проректор по учебной работе

Рег. № _____

Программа вступительного испытания
по специальности 1-21 80 07 "Прикладная и математическая лингвистика"
для поступающих в магистратуру

Минск
2008

Авторы: Головня Анастасия Ивановна, доцент, кандидат филологических наук
Лаврененко Анжелла Валерьевна, доцент, кандидат филологических наук

Рецензент: Рычкова Людмила Васильевна, зав.кафедрой общего языкознания
ГрГУ, доцент, кандидат филологических наук

Одобрена на заседании Ученого совета (методической комиссии)
филологического факультета
(протокол № _____ от " ____ " _____ 2007 г.)

Ответственный за редакцию: Головня Анастасия Ивановна

КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

Компьютерная лингвистика в ряду других лингвистических дисциплин, ее цели и задачи.

Теоретические и практические основания компьютерной лингвистики: искусственный интеллект, теория программирования, когнитивистика.

Математизация знаний о деятельности мозга.

Естественный язык как модель действительности и источник компьютерного моделирования.

Соотношение знания и понимания. Уровни понимания текста. Язык как многоуровневая иерархо-неиерархическая система.

ИНЖЕНЕРИЯ ЗНАНИЙ

Инженерия знаний как поиск знаний в информмассивах, разработка средств экономной упаковки знаний, создание систем, позволяющих пользоваться машинными знаниями. Проблема полноты/ неполноты формальных и реальных систем. Теория представления знаний о языке и мире.

Виды знаний: декларативные и процедурные. Алгоритмы работы со знаниями.

Фреймы как способы представления понятия и компьютерная модель значения.

Семантические сети, их достоинства и недостатки.

Языковая картина мира, отраженная в ассоциативном словаре.

Сценарий, план, схема, сеть.

Моделирование ситуации (сюжета) в диалоговых системах на базе вопросных операторов.

Словарь и грамматика как способы представления знаний.

Текст как способ представления знаний.

Паралингвистические способы представления знаний.

Гипертекст как теория и компьютерная технология представления знаний.

КВАНТИТАТИВНАЯ ЛИНГВИСТИКА

Основные статистические категории: выборка и совокупность, типы переменных. Применение статистики в лингвистических исследованиях.

Системная и прагматическая выборки лингвистических данных. Частоты: индивидуальная, частота класса, абсолютная, относительная, условная и позиционная.

Статистико-комбинаторные, дистрибутивно-статистические и дешифровочные методы в грамматике.

Частотные словари, словари-индексы, машинные фонды лексики.

ПРИКЛАДНАЯ ЛЕКСИКОГРАФИЯ

Задачи прикладной лексикографии. Теоретическое и методическое обеспечение разработок машинных словарей.

Существующие машинные словари.

Словарь словоформ, его достоинства и недостатки, принципы разработки. Моделирование полной парадигмы слова. Способы подачи омонимичных словоформ.

Словарь типа «машинная основа + машинная флексия», его достоинства и недостатки, принципы разработки. Омонимия машинных основ и машинных флексий.

Системное тегирование (кодирование) морфологической информации на уровне словаря. Требования к тегам: полнота, минимальность, непересекаемость (однозначность), избыточность. Каноническая (начальная) форма слова.

Возможные пути распознавания машиной слов, отсутствующих в словаре.

Моделирование парадигмы незнакомого слова на основе симметрии финалей словоформ. Возможности вариантного моделирования парадигмы.

Концепция матричного гнездового (словообразовательного) словаря как один из способов снятия неполноты словарей.

Семантические машинные словари.

Автоматизация лексикографических работ.

Словарь в системах естественного языкового анализа и синтеза текстов, информационно-поисковых системах, системах машинного перевода.

МАШИННЫЕ ГРАММАТИКИ

Методы выявления и обработки синтаксической структуры словосочетаний и предложений в системах МП и автоматического анализа ЕЯ-текстов. Локальный синтаксический анализ. Выделение именных групп в структуре предложения.

Принципы контекстно-свободной грамматики (Context Free Grammar).

Универсальные алгоритмы многовариантного синтаксического анализа по непосредственно составляющим.

Понятие и структура «признака» в обобщенной грамматике составляющих (Generalized Phrase Structure Grammar). Представление структуры предложения в рамках лексико-функциональной грамматики (Lexical Functional Grammar).

Возможности учета семантических аспектов естественного языка при синтаксическом анализе текста (предложения). Синтаксическая семантика и ее описание с помощью формальных методов. Концепция «глубинных» падежей Ч.Филлмора и В.В.Бондарко. Трансформационная грамматика. Модели управления глаголов и алгоритмы выделения глагольных групп в структуре предложения в грамматике зависимостей, непосредственно составляющих и контекстно-свободной грамматике.

Проблема синтеза связного осмысленного текста на заданную тему в системах ИИ.

ОМОНИМИЯ КАК СИСТЕМА

Концептуальное поле омонимии.

Межпарадигматическая омонимия.

Формальная система омонимии. Общая классификация омонимов в письменных и устных текстах.

Комбинированная грамматическая омонимия..

Обобщение внутрисистемной языковой омонимии (внутриклассовой и межклассовой омонимии). Омофакторы.

Межъязыковая омонимия.

Снятие омонимии в речи и на письме. Помехи, которые создает омонимия при автоматическом переводе, и возможные пути их устранения.

*

КОРПУСНАЯ ЛИНГВИСТИКА

Исходные понятия корпусной лингвистики: проблемная область, корпус данных, корпус текстов, параллельный многоязычный корпус текстов.

Динамические и статические корпуса текстов. Требования к корпусу текстов: репрезентативность, экономичность, полнота. Аннотированные корпуса текстов, автоматизация их создания и коррекции. Программное обеспечение корпусной лингвистики. Опыт разработки корпусов текстов в Беларуси и за рубежом.

ФИЛОСОФИЯ ЯЗЫКА, ЯЗЫК КАК СИСТЕМА

Мозг как универсальное отражательное устройство. Опережающее отражение действительности. Функциональный подход.

Неданность языка как объекта. Деятельностный подход.

Методы анализа языка. Системный подход. ОТСУ как общая методология.

Симметрия как сохранение, совпадение признаков; асимметрия, диссимметрия в языке на разных уровнях.

Рефлексивность, симметричность и транзитивность на разных языковых уровнях как исходные системные характеристики. Теоретико-множественный характер лингвистического универсума на разных уровнях членения.

Изоморфизм и полиморфизм языковых объектов.

Словообразование, словоизменение и текстопорождение как направленные и ненаправленные изменения в системе.

Конвергенция и дивергенция на уровне словообразования и словоизменения.

Отраженный характер лингвистических взаимодействий.

Экстралингвистические факторы.

Понятие лингвистического универсума как предельно достижимого множества однородных текстов. Неполнота лингвистического универсума.

Синхрония/диахрония/панхрония в симметрично-асимметричном представлении. Квазисимметрия (омонимия) и ее системный характер в разных предметных областях.

Метафора как система. Формализация метафоры.

Отношения единства первичных элементов как системообразующий фактор.

Закон композиции. Вывод модусов, моделей и модификаций на уровне

синтаксиса.

Тождественные, количественные, качественные и относительные преобразования. Групповой характер преобразований.

Порядок слов в ядерной и расширенных композициях. Языковая изомерия, ее уровни. Понятие левизны/ правизны / лево - правизны. Системный изоморфизм. Семантика возможных миров. Неединственность критериев смысла. Логика и язык, их соотношения.

Симметрия / асимметрия / диссимметрия и информация. Вероятностные модели с переменными. Асимметризация систем и приращение информации.

Закон композиции как упаковка информации о мире.

Текст как суперинформационная творческая структура. Белковые цепочки как тексты без пробелов.

ОТСУ как база для создания общей теории языка.

Математизация языковых представлений – основа доказательности гипотез.

Лингвистические банки данных и Машинный фонд русского языка.

Системные структуры представления знаний о языке.

ТЕОРИЯ И ПРАКТИКА МАШИННОГО ПЕРЕВОДА

Перевод как вид языковой деятельности. Виды переводов.

Машинный (автоматизированный и автоматический) перевод как одно из направлений искусственного интеллекта.

Три подхода к моделированию процесса перевода в системах МП: прямой (подстрочный) перевод; перевод через язык-посредник; перевод как трехэтапный процесс, включающий анализ, межъязыковые операции и синтез.

Лингвистические проблемы машинного перевода.

Основные типы «человеко-машинных» систем МП.

Способы организации машинных словарей в зависимости от стратегии перевода. Полисемия и способы выбора лексического эквивалента при переводе.

Омонимия при автоматическом переводе. Способы обнаружения морфологических омонимов в тексте. Синтаксический анализ как способ снятия грамматической омонимии при автоматическом переводе.

Перевод имен собственных, морфологические и орфографические трудности при передаче имен собственных.

Способы отражения при переводе грамматических категорий, отсутствующих в языке перевода (определенности/ неопределенности, несовпадающих падежных форм имен и временных форм глагола, специфических для ряда языков синтаксических конструкций).

Подстрочный перевод, редактирование и оценка качества.

КОМПЬЮТЕРНЫЕ ЭКСПЕРТНЫЕ СИСТЕМЫ

Экспертизация знаний. Верификация истинности знаний. Критерии надежности

(количественные, качественные, структурные).

Матричное представление знаний. Матрицы как порождающие и классификационные системы. Прямая и треугольная декартовы матрицы. Теоретическая и системная полнота матричных представлений.

Экспертизация языковых знаний.

Экспертизация лексики. Построение классификационных систем-матриц открытого типа для словообразования и словоизменения. Многозначность как симметричная многослойная упаковка формы при асимметрии значений.

Антонимия как система. Синонимия как система.

Экспертизация синтаксиса.

Экспертизация школьных учебников на базе машинных словарей.

Дифференциальный словарь.

АВТОМАТИЧЕСКИЙ АНАЛИЗ ЕСТЕСТВЕННОЯЗЫКОВОГО ТЕКСТА

Автоматическая обработка текста как проблема компьютерной лингвистики.

Автоматизированные системы обработки текста и их архитектура.

Лингвистический процессор как сложная автоматизированная система обработки текста.

Анализ, синтез, диалоговое взаимодействие, логическая обработка информации.

Структура лингвистического банка данных: машинные словари, базы данных и знаний. Проблема ограничения естественного языка для действующей системы обработки ЕЯ-текстов. Проблема универсальности лингвистических средств обработки текста.

Основные виды понимания текста и методы формализации понимания.

Анализаторы текста как модели различных видов понимания.

Процедуры экспликационного и информационного анализа текста.

Морфологический, синтаксический и семантический этапы экспликационного анализа текста.

Автоматизация синтеза текста.

Диалоговые системы. Принципы организации информационного диалога.

Автоматическое аннотирование и индексирование научно-технической документации. Автоматическое реферирование.

Обучающие диалоговые системы. Автоматическое извлечение знаний из текстов для экспертных систем. Системы автоматизированного редактирования.

Терминологические банки данных. Системы машинных фондов естественных языков. Создание АРМ лингвиста.

Литература

Автоматизация подготовки словарей. М., 1988.

Арапов М. В. Квантитативная лингвистика. М., 1988.

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

Баранов А.Н. Фреймы и сценарии. Категории искусственного интеллекта в

лингвистической семантике. М., 1987.

Баранов А.Н., Кобозева И.М. Метаязыковые средства описания семантики предложения: опыт типологии // Лингвистическое обеспечение информационных систем. М., 1987.

Баранов А.Н., Сергеев В.М. Искусственный интеллект и традиционные методы анализа текста // Когнитивные исследования за рубежом. М., 1990.

Баранов А.Н. Автоматизация лингвистических исследований: корпус текстов как лингвистическая проблема // Русистика сегодня. 1998. № 1 – 2, с. 179 – 191.

Головня А.И. Омонимия как системная категория языка: Диссертация ... канд. Фил. Наук:10.02.19. // Бгу. Мн., 1996.

Гомон Д.Н. Системные проблемы лексической омонимии: Диссертация ... канд. фил. наук:10.02.19. // Бгу. Мн., 2004

Карпов В.А. Язык как система. Мн., 1992.

Минский М. Структура для представления знаний // Психология машинного зрения. – М., 1978.

Михайлов М.Н. Компьютерное обеспечение корпуса текстов (взгляд пользователя) // Русистика сегодня. 1998. № 1 – 2, с. 192 – 201.

Новое в зарубежной лингвистике. Вып. XXIV. Компьютерная лингвистика. М., 1989.

Представление и использование знаний. М., 1989.

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

Сухотин В. В. Исследование грамматики числовыми методами. М., 1990.

Шенк Р. Обработка концептуальной информации. М., 1980.

Анохин П.К. Избранные труды. Философские аспекты теории функциональных систем. М., 1978.

Карпов В.А. Язык как система. Мн., 1992; М., 2003.

Урманцев Ю.А. Общая теория систем: состояние, приложения и перспективы // Система. Симметрия. Гармония. М., 1988.

Апресян Ю. Д., Богуславский И. М., Иомдин Л. Л. и др. Лингвистический процессор для сложных информационных систем. М., 1992

Марчук Ю.Н. Проблемы машинного перевода. М., 1983.

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