

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

САЕ «Институт человек цифровой эпохи»

УТВЕРЖДАЮ:
Руководитель ОПОП:

 3.И. Резанова

« 31 » августа 20 22 г.

Рабочая программа дисциплины

**Формальные грамматические модели естественного языка и
их прикладное применение**
по направлению подготовки

45.03.03 Фундаментальная и прикладная лингвистика

Направленность (профиль) подготовки :
Компьютерная и когнитивная лингвистика


Форма обучения
Очная

Квалификация
Магистр

Год приема
2022

Код дисциплины в учебном плане: Б1.В.ДВ.1.1.6

СОГЛАСОВАНО:
Руководитель ОПОП

 3. И. Резанова

Председатель УМК

 Ю.А. Тихомирова

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

УК-6 – способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки;

ОПК-1 – способен решать профессиональные задачи, применяя основные понятия, категории и положения лингвистических теорий и актуальные концепции в области лингвистики;

ОПК-2 – способен анализировать, сопоставлять и критически оценивать различные лингвистические направления, теории и гипотезы при решении задач профессиональной деятельности;

ОПК-3 – способен выбирать оптимальные подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий;

ПК-1 – способен проводить самостоятельные исследования и получать новые научные результаты в области междисциплинарных лингвистических исследований;

ПК-3 – способность разрабатывать системы автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистические компоненты электронных ресурсов и интеллектуальных электронных систем (лингвистические корпуса, словари, онтологии, базы данных);

ПК-4 – способность разрабатывать проекты прикладной направленности в области когнитивной и компьютерной лингвистики с применением современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИУК-6.2 Реализует и корректирует стратегию личностного и профессионального развития с учетом конъюнктуры и перспектив развития рынка;

ИОПК-1.3 Оценивает и производит коррекцию в использовании понятий, категорий и положений лингвистических теорий при решении профессиональных задач;

ИОПК-2.3 Совершает выбор лингвистического направления, теории на основе их самостоятельного поиска и анализа, сопоставления, критической оценки при решении задач профессиональной деятельности;

ИОПК-3.2 Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий;

ИПК-1.2 Способен формулировать исследовательскую проблему на основе изучения актуальных направлений междисциплинарных лингвистических исследований и разрабатывать программу исследований;

ИПК-1.3 Последовательно реализует исследовательскую программу, получает новые научные результаты;

ИПК-3.3 Разрабатывает лингвистические компоненты интеллектуальных информационных систем (онтологии, базы данных);

ИПК-4.3 Обеспечивает выполнение проекта в области когнитивной и компьютерной лингвистики с применением современных технических средств и информационных технологий, в том числе в области искусственного интеллекта, в соответствии с установленными целями, сроками и затратами.

2. Задачи освоения дисциплины

– Освоить основные формальные грамматические модели естественного языка, принципы их применения к речевому материалу

– Научиться анализировать формальные грамматические модели естественного языка и результаты их применения к речевому материалу

– Овладеть навыками лингвистического моделирования вербального материала в соответствии с требованиями формальных грамматических моделей естественного языка

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, предлагается обучающимся на выбор. Дисциплина входит в модуль *Компьютерная лингвистика*.

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Третий семестр, экзамен

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам «Лингвистика в контексте современного гуманитарного и естественнонаучного знания», «Профессиональный иностранный язык», «Основные направления лингвистического обеспечения новых информационных технологий», «Когнитивная лингвистика: "передний край"», «Введение в анализ естественного языка (NLP)».

6. Язык реализации

Русский

7. Объем дисциплины

Общая трудоемкость дисциплины составляет 5 з.е., 180 часов, из которых:

-лекции: 10 ч.

-практические занятия: 44 ч.

в том числе практическая подготовка: 0 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Введение в проблематику

Понятие модели в лингвистике. Виды моделей.

Тема 2. Грамматика зависимостей

Основные принципы устройства модели. Проективность. Принцип синтаксической несовместимости. Грамматика зависимостей в модели «Смысл-Текст». Общие принципы организации. Поверхностно-синтаксические структуры. Глубинно-синтаксические структуры. Семантический компонент модели.

Тема 3. Грамматика составляющих

Грамматика составляющих: основные принципы устройства модели. Порождающая грамматика: трансформационные правила. Нулевые единицы грамматической структуры. Актантная деривация. Теория принципов и параметров. X-штрих-теория

Тема 4. Типы формальных грамматик

Типы формальных грамматик по иерархии Хомского: Неограниченные грамматики, Контекстно-зависимые грамматики, Контекстно-свободные грамматики, Регулярные грамматики. Контекстно-зависимые грамматики в системах автоматизированной обработки неструктурированного текста в прикладных задачах Natural Language Processing (NLP). Регулярные выражения как шаблоны для текстового поиска в корпусной лингвистике и NLP.

Тема 5. Проблемы синтаксической синонимии и синтаксической омонимии.

Типология и анализ синтаксической синонимии. Типология и анализ синтаксической омонимии. Анализ синтаксических аспектов фразеологии и афористики. Шаблоны автоматического извлечения и автогенерация синонимических и омонимических синтаксических структур методами контекстно-свободных грамматик.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, проведения контрольных работ, тестов, выполнения домашних заданий, выполнение и представление обзора литературы по заданной теме, устный опрос, письменный опрос, проверяющие знание теоретической базы, и фиксируется в форме контрольной точки не менее одного раза в семестр.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Зачет с оценкой в третьем семестре представлен в форме онлайн-теста, который включает задания разного типа и проверяет ИУК-6.2; ИОПК-1.3; ИОПК-2.3; ИОПК-3.2; ИПК-1.2; ИПК-1.3; ИПК-3.3; ИПК-4.3.

Примеры заданий

Контрольные задания, проверяющие навыки лингвистического моделирования вербального материала в соответствии с требованиями формальных грамматических моделей естественного языка:

Построение деревьев зависимостей для предложений.

Анализ критериев синтаксической связанности в словосочетаниях. Построение моделей разметки дерева подчинения

Контрольные задания, проверяющие умения анализировать формальные грамматические модели естественного языка и результаты их применения к речевому материалу:

Анализ проективности предложений в структуре зависимостей.

Анализ предикативных вершин и грамматической омонимии частей речи в сегментации предложения.

Контрольные задания, проверяющие знания основных формальных грамматических моделей естественного языка, принципов их применения к речевому материалу, необходимых для разработки и совершенствования системы автоматизации и информационной поддержки лингвистических исследований:

Автоматизация составления конкордансов в программе Томита-парсер для извлечения и изучения контекстов семантических сущностей в прикладных задачах NLP.

Конкатенация фактологических цепочек и фреймовых структур в программе Томита-парсер.

Анализ синтаксической синонимии на примере извлечения каузирующих связей и генерированию вопросно-ответных синтаксических конструкций с помощью программы Томита-парсер.

Анализ синтаксической омонимии с помощью формальных шаблонов языка поисковых запросов в НКРЯ

Автогенерация фраз по лексико-синтаксическому шаблону с помощью программы Томита-парсер на примере анализа афористики.

Максимальное количество баллов за тест – 100 баллов.

Критерии оценивания

	Требования	
	Ответ на теоретические вопросы	Выполнение практического задания
Отлично	Обучающийся показал знание заданного аспекта устройства формальной грамматической	Задание выполнено полностью, без ошибок. Представлены

	концепции естественного языка, умение его анализировать и оценивать.	возможные вариативные способы решения.
Хорошо	Обучающийся показал знание заданного аспекта устройства формальной грамматической концепции естественного языка, но не в полной мере проявил умение его анализировать и оценивать.	Задание выполнено не полностью (от 75%). Не представлены возможные вариативные способы решения.
Удовлетворительно	Обучающийся показал неполное знание заданного аспекта устройства формальной грамматической концепции естественного языка, не проявил умение его анализировать и оценивать.	Задание выполнено не полностью (от 50%). Не представлены возможные вариативные способы решения.
Неудовлетворительно	Обучающийся не показал знание заданного аспекта устройства формальной грамматической концепции естественного языка.	Задание выполнено в объеме менее чем 50%.

11. Учебно-методическое обеспечение

- а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=14682>
- б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

1. Кибрик А.А., Кобозева И.М., Секерина И.А. (ред.) Современная американская лингвистика: фундаментальные направления. М.: Едиториал УРСС, 2002.
2. Тестелец Я.Г. Введение в общий синтаксис. М.: Издательство РГГУ, 2001. 798 с.
3. Митренина О.В, Романова Е.Е., Слюсарь Н.А. Введение в генеративную грамматику. М.: Либроком, 2012. 376 с.
4. Мельчук И.А. Опыт теории лингвистических моделей «Смысл \leftrightarrow Текст». М.: Школа «Языки русской культуры», 1999. 346 с.

б) дополнительная литература:

1. Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л., Санников В.З., Теоретические проблемы русского синтаксиса: Взаимодействие грамматики и словаря. М.: Языки славянских культур, 2010. 408 с.
2. Гладкий А. В. Формальные грамматики и языки. М.: Наука, 1973. 368 с.
3. Гладкий, А.В. Синтаксические структуры естественного языка в автоматизированных системах общения. М.: Наука, 1985. - 144 с.
4. Гросс М., Лантен А. Теория формальных грамматик. М.: Мир, 1971. 296 с.
5. Золотова, Г.А. Очерк функционального синтаксиса русского языка. М.: Наука, 1973. - 351 с.; Изд. 3. URSS. 2009. - 352 с.
6. Золотова, Г.А. Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса. URSS. 2006. - 440 с.

7. Иванов, Е. Е. Лингвистика афоризма: учебно-методическое пособие. Могилев: МГУ имени А. А. Кулешова, 2016. - 156 с.
8. Касевич В.Б. Элементы общей лингвистики. М.: Наука, 1977. 177 с.
9. Лайонз Д. Введение в теоретическую лингвистику. М.: Прогресс, 1978. 540 с.
10. Медведева Т.Н. Формальные модели в лингвистике: Учебное пособие. Саратов: Научная книга, 2010. 56 с.
11. Падучева Е.В. О способах представления синтаксической структуры предложения // Вопросы языкознания. 1964. № 2. С. 99-113.
12. Сухотин, В. П. Синтаксическая синонимика в современном русском языке. Москва, 1960. - 160 с.
13. Хомский Н. Синтаксические структуры // Новое в лингвистике. М., 1962. Вып. II. С. 412-527.
14. Хомский Н. Аспекты теории синтаксиса. Пер. с англ. В.А. Звегинцева. М., Изд-во Моск. ун-та, 1972. 129 с.
15. Хомский Н., Миллер Дж. Введение в формальный анализ естественных языков // Кибернетический сборник / Под ред. А.А.Ляпунова и О.Б.Лупанова.—М.: Мир, 1965.

в) ресурсы сети Интернет:

1. Elibrary.ru: научная электронная библиотека [Электронный ресурс]. URL: http://elibrary.ru/project_risc.asp (01.09.2016).
2. ЕТАР-3 МТ SYSTEM // Laboratory of Computational Linguistics. Institute For Information Transmission Problems (Kharkevich Institute), RAS [Электронный ресурс]. URL: <http://www.cl.iitp.ru/etap3>
3. Scopus: database [Электронный ресурс]. URL: <https://www.scopus.com/>
4. Web of Science: database [Электронный ресурс]. URL: <http://login.webofknowledge.com/>
5. Компьютерная лингвистика и интеллектуальные технологии (материалы ежегодного сборника) [Электронный ресурс]. URL: <http://www.dialog-21.ru/digest/>
6. Национальный корпус русского языка: Синтаксический корпус [Электронный ресурс]. URL: <http://www.ruscorpora.ru/search-syntax.html>
7. Томита-парсер. URL: <https://yandex.ru/dev/tomita/>

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

– Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office OneNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook); Томита-Парсер (<https://github.com/yandex/tomita-parser>)

– публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

б) информационные справочные системы:

– Электронный каталог Научной библиотеки ТГУ – <http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>

– Электронная библиотека (репозиторий) ТГУ – <http://vital.lib.tsu.ru/vital/access/manager/Index>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

15. Информация о разработчиках

Тубалова Инна Витальевна, д-р филол. наук, доцент, декан филологического факультета ТГУ;

Бочаров Алексей Владимирович, канд. ист. наук, доцент, доцент кафедры общей, компьютерной и когнитивной лингвистики ТГУ.