

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

Филологический факультет



УТВЕРЖДАЮ:

Декан филологического факультета
И.В. Тубалова

« 30 » августа 2023 г.

Рабочая программа дисциплины

Введение в корпусную лингвистику

по направлению подготовки

45.04.01 Филология

Направленность (профиль) подготовки:

Юридическая лингвистика

Форма обучения

Очная

Квалификация

Лингвист-эксперт

Год приема

2023

Код дисциплины в учебном плане: Б1.В.ДВ.01.02

СОГЛАСОВАНО:

Руководитель ОПОП

Шевчик А.В. Шевчик

Председатель УМК

Тихомирова Ю.А. Тихомирова

Томск – 2023

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

БК-2 Способен использовать научные методы для решения профессиональных задач.

ОПК-3 Способен владеть широким спектром методов и приемов филологической работы с различными типами текстов.

Результатами освоения дисциплины являются следующие результаты обучения:

РОБК-2.1 Знает основные методы научных исследований.

РОБК-2.2 Умеет выстраивать систематическую и логическую цепочку анализа и принимаемых решений в контексте задачи профессиональной деятельности.

РООПК-3.1 Знает методы и приемы профессиональной обработки, филологического анализа и оценки текстов различного типа.

РООПК-3.2 Умеет осуществлять выбор методов и приемов профессиональной обработки, анализа и оценки текстов различного типа и применять их для решения профессиональных задач конкретной области филологии.

2. Задачи освоения дисциплины

- Освоить основные понятия корпусной лингвистики;
- Владеть навыком использования специализированных электронных лингвистических ресурсов – корпусов;
- Научиться применять корпусные ресурсы для решения практических задач профессиональной деятельности – исследовательских и педагогических.

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений (Блок 1), предлагается обучающимся на выбор.

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

Третий семестр, зачет с оценкой

Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: «Методы филологических исследований»; «Проектная и исследовательская деятельность в филологии»; «Лингвистические и когнитивные методы исследования семантики».

5. Язык реализации

Русский

6. Объем дисциплины

Общая трудоемкость дисциплины составляет 3 з.е., 108 часов, из которых:

-лекции: 4 ч.

-практические занятия: 8 ч.

-практическая подготовка: 8 ч.

Объем самостоятельной работы студента определен учебным планом.

7. Содержание дисциплины, структурированное по темам

Тема 1. Корпусная лингвистика как научная и практическая дисциплина. База данных и корпус. История корпусной лингвистики.

Тема 2. Корпус как инструмент лингвистических исследований. Типология корпусов.

Тема 3. Понятие разметки. Типы разметки

- Тема 4. Репрезентативность, сбалансированность и объем корпуса. Понятие *ipm*.
- Тема 5. Морфологическая разметка. Морфологический стандарт НКРЯ. Понятие грамматической омонимии. Способы снятия омонимии.
- Тема 6. Синтаксическая разметка. Корпус СинТагРус.
- Тема 7. Семантическая разметка. Семантическая разметка в НКРЯ.
- Тема 8. Учебные корпуса. Многоязычные (параллельные) корпуса. Диалектные корпуса. Способы применения корпусов в преподавательской деятельности.
- Тема 9. Мультиязычные корпуса. Виды разметки мультиязычного корпуса. Тема 10. Лингвистические исследования на базе корпусов.

8. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, опроса на занятиях (РООПК-3.1), выполнения самостоятельных работ (РОБК-2.1, РОБК-2.2, РООПК-3.1, РООПК-3.2), выполнения домашних заданий (РОБК-2.1, РОБК-2.2, РООПК-3.1, РООПК-3.2) и фиксируется в форме контрольной точки не менее одного раза в семестр.

Примерные задания для проведения контрольного опроса.

Тестовые задания множественного выбора:

- Обязательными свойствами хорошего лингвистического корпуса являются:
 - представительность;
 - присутствие разметки;
 - снятая грамматическая омонимия;
 - баланс между разными типами текстов.
- Разметка Обучающего корпуса отличается от аннотации других частей НКРЯ тем что:
 - она содержит разметку ошибок;
 - она проведена в соответствии со школьной грамматикой; б) в ней присутствует синтаксическая разметка;
 - она произведена вручную. Тестовые задания открытого типа:
 - Интерпретация корпусных данных возможна, если исследователю известен _____ корпуса, по которому велся поиск.
 - Нормализация значений обычно предполагает вычисление отношения количества вхождений к объему корпуса, умноженного на _____.

Примеры заданий для самостоятельных работ.

- Зайдите на сайт Национального корпуса русского языка, и выясните, каков объем представленных в нём корпусов в настоящий момент.
- На странице выбора метаразметки НКРЯ (ruscorpora.ru/mycorpora-main.html) соберите подкорпус текстов Вашего любимого автора или писателя, поэта, чьё творчество Вы изучаете.
- Изучите систему семантических помет в НКРЯ. Проведите поиск по одной из выбранных семантических групп. Зафиксируйте найденные ошибки. Чем они обусловлены?

Примеры домашних заданий.

- Опишите, какие перспективы открывает перед вашим собственным исследованием «корпусная идеология».
- Изучите системы экстралингвистической разметки двух корпусов (по выбору).
- Соотнесите целевую направленность корпуса и систему его метаразметки.
- Охарактеризуйте один из известных вам корпусов по признакам сбалансированности

и репрезентативности.

9. Порядок проведения и критерии оценивания промежуточной аттестации

Зачет с оценкой в первом семестре проводится в письменной форме по билетам. Билет содержит теоретический вопрос и практическое задание. Продолжительность зачета 1,5 часа.

Зачёт с оценкой «отлично» выставляется за полный, логично выстроенный ответ. Обучающийся свободно ориентируется в проблематике курса, знает терминологию и основные положения дисциплины, отвечает на дополнительные вопросы, обобщающие содержание курса.

Зачёт с оценкой «хорошо» выставляется за правильные и достаточно полные ответы. Обучающийся свободно ориентируется в большинстве вопросов курса, но испытывает небольшие затруднения при ответе на отдельные вопросы (1 или 2), имеются незначительные погрешности в использовании терминологии.

Зачёт с оценкой «удовлетворительно» выставляется при недостаточно полном ответе, при наличии заметных пробелов в знаниях обучающегося, затруднениях при ответе на более чем 2 вопроса экзаменатора.

Незачёт выставляется в случае отсутствия знаний по курсу, неспособности ответить на вопросы экзаменатора.

Вопрос и задание оцениваются по шкале 3-5 баллов, текущий контроль также учитывается в промежуточной аттестации, оценка выставляется по шкале 3-5 баллов. Интегральная оценка высчитывается по формуле «оценка за текущий контроль $\times 0,4$ + оценка за промежуточную аттестацию $\times 0,6$ » и округляется до целого числа в пользу обучающегося.

Примерный перечень теоретических вопросов по курсу.

1. Что такое лингвистический корпус? В чём его отличие от других типов лингвистических ресурсов?
2. Какие корпусные ресурсы Вам известны? (Не менее 5). Охарактеризуйте один из них по 5 типологическим признакам.
3. По каким признакам классифицируются корпуса? Какие типологии корпусов существуют?
4. Какие параметры включаются в экстралингвистическую аннотацию? Какова их функция?
5. Охарактеризуйте понятие «репрезентативность» корпуса. Для чего нужна ipm и как рассчитывается эта характеристика?
6. В чём основные отличия морфологического стандарта НКРЯ от классической теории частей речи?
7. Что такое грамматическая омонимия? Какие способы снятия омонимии Вам известны?
8. Как организована синтаксическая разметка НКРЯ?
9. Как организована семантическая разметка НКРЯ? В чём её преимущества и ограничения?
10. В чём специфика учебных корпусов? Для каких целей они используются?
11. В чём специфика параллельных корпусов? Для каких целей они используются?
12. В чём специфика мультимодальных корпусов? Для каких целей они используются?
13. В чём специфика диалектных корпусов? Для каких целей они используются?
14. В чём отличие corpus-based и corpus-driven подходов к исследованию языка на базе корпуса?

15. Как корпусная лингвистика повлияла на изменение представлений о языке?

Примеры практических заданий.

1. Пользуясь любым корпусом русского языка, расположите падежи имён существительных в соответствии с частотой их употребления в корпусе.
2. Подсчитайте, сколько раз в НКРЯ используется имя существительное *быль*.
3. Соберите с помощью НКРЯ подкорпус примеров с прилагательными-обозначениями цвета в текстах В.П. Астафьева.
4. Сравните относительную частотность использования слова *тятя* в основном и диалектном подкорпусах НКРЯ.
5. Пользуясь НКРЯ, определите приблизительный период возникновения у слова *передовой* значения 'прогрессивный'.

Первая часть представляет собой вопрос, проверяющий РОБК-2.1, РООПК-3.1. Ответ на вопрос первой части дается в развернутой форме.

Вторая часть содержит практическое задание, проверяющее РОБК-2.2, РООПК-3.1, РООПК-3.2.

10. Учебно-методическое обеспечение

- а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=3652>.
- б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.
- в) План практических занятий по дисциплине: <https://moodle.tsu.ru/mod/resource/view.php?id=780232>.
- г) Самостоятельная работа студентов включает чтение текстов из списка литературы, освоение материалов массового открытого онлайн-курса и подготовку к лекциям по технологии перевернутого класса. Подробные методические указания представлены в курсе «Moodle» - <https://moodle.tsu.ru/course/view.php?id=3652>.

11. Перечень учебной литературы и ресурсов сети Интернет

- а) основная литература:
 - Захаров В. П. Корпусная лингвистика : [учебник для студентов направления "Лингвистика"] / В. П. Захаров, С. Ю. Богданова ; Санкт-Петербургский гос. ун-т. - 2-е изд., перераб. и доп.. - Санкт-Петербург : СПбГУ РИО, 2013. - 144, [3] с.: ил., табл.. URL: <http://sun.tsu.ru/limit/2018/000517760/000517760.pdf>
 - Копотев М. В. Введение в корпусную лингвистику : учебное пособие для студентов филологических и лингвистических специальностей университетов / М. Копотев. - Прага : Animedia Company, 2014. - 1 онлайн-ресурс (194, [1] с.): ил.. URL: <http://sun.tsu.ru/limit/2016/000553839/000553839.pdf>
 - Копотев М.В. О некоторых следствиях корпусной лингвистики для общей теории языка // Филологический класс. 2021. Т. 26. № 2. С. 90–102. <https://cyberleninka.ru/article/n/o-nekotoryh-sledstviyah-korpusnoy-lingvistiki-dlya-obschey-teorii-yazyka?ysclid=lnvf7ndq9z497873389>
 - Плунгян В.А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении. 2008. №. 16 (2). С. 7–20. https://ruscorpora.ru/new/tool_ideology.pdf
 -
- б) дополнительная литература:

- Два века в двадцати словах / М. К. Данова, М. А. Даниэль, Н. Р. Добрушина [и др.]; отв. ред. Н.Р. Добрушина, М.А. Даниэль. – М.: Национальный исследовательский университет "Высшая школа экономики", 2016. 456 с.
- Корпус как язык: от масштабируемости к дифференциальной полноте / В. И. Беликов [и др.] // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной Международной конференции «Диалог». – М., 2013. – Вып. 12 (19). – С. 84–95.
- Малышева А. В., Ронько Р. В. Русский объектный генитив при отрицании по данным диалектных корпусов и устного корпуса НКРЯ. // Вопросы языкознания, 2020. № С. 25–54.
- Пиперски А.Ч. Русский язык и корпусное разнообразие // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной международной конференции. Российский государственный гуманитарный университет. 2020. С. 615–627.
- Рахилина Е.В., Кустова Г.И., Ляшевская О.Н., Резникова Т.И., Шеманаева О.Ю. Задачи и принципы семантической разметки лексики в НКРЯ // Национальный корпус русского языка: 2006—2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 215–239.
- Рахилина Е.В., Бычкова П.А., Жукова С.Ю. Речевые акты как лингвистическая категория: дискурсивные формулы // Вопросы языкознания. 2021. № 2. С. 7–27.
- Резникова Т.И. Славянская корпусная лингвистика: современное состояние ресурсов // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 402–461.
- Савчук С.О. Метатекстовая разметка в Национальном корпусе русского языка: базовые принципы и основные функции // Национальный корпус русского языка: 2003- 2005. Результаты и перспективы. — М., 2005. С. 62–88.
- Шерстинова Т.Ю. О подготовке к веб-публикации корпуса повседневной русской речи «Один речевой день»: анонимизация текстов и выборочное кодирование лексики // Труды международной конференции «Корпусная лингвистика-2019». – СПб.: Изд-во С.-Петербург. ун-та, 2019. – С. 366–372.
- Шмурак Р.И. К уточнению понятия упрека с помощью корпусных инструментов // Известия РАН. Серия литературы и языка. 2020. № 3. Т. 79. С. 24–48.
- Biber D. Representativeness in corpus design // Literary and linguistic computing. 1993. Vol. 8. № 4. P. 243–257.
- Biber D., Conrad S., Reppen R. Corpus linguistics: Investigating language structure and use. Cambridge University Press, 1998.
- McEnery T., Hardie A. Corpus linguistics: method, theory and practice. Cambridge University Press, 2011.

в) ресурсы сети Интернет:

- Весь Толстой в один клик: readingtolstoy.ru
- Корпус М. В. Ломоносова: lomonosov.pro
- Левинзон А. И. Введение в корпусную лингвистику. https://openedu.ru/course/hse/CORPUS/?session=fall_2020
- Машинный фонд русского языка: cfl.ruslang.ru
- Мультимедийный русский корпус (МУРКО): www.ruscorpora.ru/instruction-murco.html
- Русская корпусная грамматика <http://rusgram.ru/index>
- Русский мультимедийный дискурс <https://multidiscourse.ru/main/>
- «Слово о полку Игореве»: Параллельный корпус переводов: nevmenandr.net/slovo
- Упсальский и Тюбингенский корпуса: www.lingexp.uni-tuebingen.de/sfb441/b1/en/korpora.html

- Corpus of Bilingual Russian child speech (BiRCh):
people.brandeis.edu/~smalamud/MILa/birch_ru.html
- EAGLES. Preliminary recommendations on Corpus Typology. EAG--TCWG-CTYP/P. Version of May, 1996 // <http://www.ilc.cnr.it/EAGLES/corpus/typ/corpus/typ.html>
- Google Books Ngram Viewer: books.google.com/ngrams
- Lancaster-Oslo-Bergen Corpus: purl.ox.ac.uk/ota/0167
- MLibrary Digital Collections: quod.lib.umich.edu/lib/collist
- ParaSol: A Parallel Corpus of Slavic and other languages: parasolcorpus.org

12. 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 On-eNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);
- 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>
- ЭБС Лань – <http://e.lanbook.com/>
- ЭБС Консультант студента – <http://www.studentlibrary.ru/>
- Образовательная платформа Юрайт – <https://urait.ru/>
- ЭБС ZNANIUM.com – <https://znanium.com/>
- ЭБС IPRbooks – <http://www.iprbookshop.ru/>

в) профессиональные базы данных:

- НКРЯ – Национальный корпус русского языка: <https://ruscorpora.ru/new/index.html>
- Учебный корпус русского языка <http://web-corpora.net/RussianLearnerCorpus/search/>
- MICASE – Корпус устного академического английского: <https://quod.lib.umich.edu/m/micase/>
- Russian Error-Annotated Learner English Corpus: <https://realec.org/>
- ГИКРЯ – генеральный интернет-корпус русского языка <http://www.webcorpora.ru/>
- Рассказы о сновидениях и другие корпуса звучащей речи: <http://spokencorpora.ru/>
- Корпус «Манускрипт» <http://mns.udsu.ru/>
- Томский диалектный корпус: https://losl.tsu.ru/losl_search

13. Материально-техническое обеспечение

Лекционная аудитория, оборудованная компьютером с мультимедийным проектором. Компьютерный класс, оснащенный компьютерной техникой и доступом к сети Интернет. Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

14. Информация о разработчиках

Шевчик Анна Валерьевна, кандидат филологических наук, доцент кафедры русского языка ТГУ.