

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

Филологический факультет

УТВЕРЖДЕНО:
Декан
И. В.Тубалова

Рабочая программа дисциплины

Методы корпусной лингвистики

по направлению подготовки

45.04.01 Филология

Направленность (профиль) подготовки:

Академическая филология: современные исследовательские технологии

Форма обучения

Очная

Квалификация

Магистр

Год приема

2025

СОГЛАСОВАНО:
Руководитель ОП
Н.А. Мишанкина

Председатель УМК
Ю.А. Тихомирова

Томск – 2025

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-2 Способен использовать в профессиональной деятельности, в том числе педагогической, знания современной научной парадигмы в области филологии и динамики ее развития, системы методологических принципов и методических приемов филологического исследования.

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

ПК-1 Выполнение отдельных заданий в рамках решения исследовательских задач в сфере филологии под руководством более квалифицированного работника.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-2.1 Демонстрирует знание современной научной парадигмы в области филологии и динамики ее развития, системы методологических принципов и методических приемов филологического исследования.

ИОПК-2.2 Осуществляет для решения профессиональных задач выбор методологических принципов и методических приемов филологического исследования, способен аргументировать свой выбор.

ИОПК-2.3 Способен к самостоятельному поиску информации в области новых, в том числе междисциплинарных, отраслей филологии, их методологических принципов и методических приемов для расширения сферы своей научной деятельности.

ИОПК-3.1 Демонстрирует углубленные знания в избранной конкретной области филологии и владение характерным для нее спектром методов и приемов филологической работы с различными типами текстов.

ИОПК-3.2 Осуществляет для решения профессиональных задач выбор методологических принципов и методических приемов исследования в избранной конкретной области филологии, способен аргументировать свой выбор.

ИОПК-3.3 Способен к самостоятельному поиску информации о новом эмпирическом материале, концепциях и теориях в избранной конкретной области филологии, новациях в ее методологических принципах и методических приемах для расширения сферы своей научной деятельности.

ИПК-1.1 Владеет методами и способами решения научных задач по тематике проводимого исследования, под руководством более квалифицированного работника намечает путь решения исследовательской задачи, методологию и методику исследования.

2. Задачи освоения дисциплины

- Освоить основные понятия корпусной лингвистики;
- Овладеть 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 Блоку 1 «Дисциплины (модули)».

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

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

Третий семестр, зачет с оценкой

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются компетенции, сформированные в ходе освоения образовательных программ предшествующего уровня образования.

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».

6. Язык реализации

Русский

7. Объем дисциплины

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

-лекции: 4 ч.

-практические занятия: 8 ч.

в том числе практическая подготовка: 8 ч.

Объем самостоятельной работы студента определен учебным планом.

8. 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. Мультимодальные корпуса. Виды разметки мультимодального корпуса.

Тема 10. Лингвистические исследования на базе корпусов.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, опроса на занятиях (ИОПК-2.1, ИОПК-3.1), выполнения самостоятельных работ (ИОПК-2.2, ИОПК-2.3, ИОПК-3.2), выполнения домашних заданий (ИОПК-3.3, ИПК-1.1) и фиксируется в форме контрольной точки не менее одного раза в семестр.

Оценочные материалы текущего контроля размещены на сайте ТГУ в разделе «Информация об образовательной программе» - <https://www.tsu.ru/sveden/education/eduop/>.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Зачет с оценкой в третьем семестре проводится в письменной форме по билетам. Экзаменационный билет состоит из двух частей. Продолжительность зачета с оценкой 1 час.

Оценочные материалы для проведения промежуточной аттестации размещены на сайте ТГУ в разделе «Информация об образовательной программе» - <https://www.tsu.ru/sveden/education/eduop/>.

11. 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 семинарских / практических занятий по дисциплине.

г) Методические указания по проведению лабораторных работ.

д) Методические указания по организации самостоятельной работы студентов.

12. 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-obshchey-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.- Петерб. 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: cfri.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/corpus.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

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 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 (*при наличии*):

- 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

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.

Аудитории для проведения занятий лекционного и семинарского типа индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации в смешенном формате («Актру»).

15. Информация о разработчиках

Земичева Светлана Сергеевна, кандидат филологических наук, старший научный сотрудник лаборатории общей и сибирской лексикографии ТГУ.

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

Шиляев Константин Сергеевич, кандидат филологических наук, доцент, доцент кафедры общей, компьютерной и когнитивной лингвистики ТГУ.