

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

Филологический факультет



УТВЕРЖДАЮ:

Декан ФилФ

И.В. Тубалова

«15»

03

2022 г.

Рабочая программа дисциплины

**Корпусная лингвистика**

по направлению подготовки

**45.03.01 Филология**

Направленность (профиль) подготовки  
«Отечественная филология (русский язык и литература)»

Форма обучения

**Очная**

Квалификация

**Бакалавр**

Год приема

**2021**

Код дисциплины в учебном плане: Б1.В.ДВ.12.01

СОГЛАСОВАНО:

Руководитель ОПОП

В.С. Киселев

Председатель УМК

Ю.А. Тихомирова

Томск – 2022

## **1. Цель и планируемые результаты освоения дисциплины (модуля)**

Целью освоения дисциплины является формирование следующих компетенций:

ПК-1 способность выполнять отдельные задания в рамках решения исследовательских задач в области филологии под руководством более квалифицированного работника.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИПК-1.1 Проводит поиск, сбор и обработку информации, необходимую для решения исследовательских задач.

ИПК-1.2 Знает и использует методы и способы решения исследовательских задач по тематике проводимых исследований.

## **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, педагогических.

Таким образом, курс позволяет реализовать деятельностный подход к изучению современного русского языка на основе использования электронных ресурсов. Содержание дисциплины способствует формированию у обучающихся как профессиональных (использование методов корпусной лингвистики для решения исследовательских и педагогических задач), так и универсальных (поиск и критический анализ информации, применение информационно-коммуникационных технологий) компетенций.

## **3. Место дисциплины (модуля) в структуре образовательной программы**

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, предлагается обучающимся на выбор (модуль Б1.В.ДВ.12).

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

Семестр 2, зачет.

## **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:

- Б1.О.18.04 Морфология;
- Б1.О.20 Общее языкознание;
- Б1.В.03 Методика преподавания русского языка;
- Б1.В.04 Методика преподавания литературы;
- Б1.В.ДВ.07.01 Методы лингвистических исследований.

## **6. Язык реализации**

Русский.

## **7. Объем дисциплины (модуля)**

Общая трудоемкость дисциплины составляет 2 з.е., 72 часа, из которых:

- лекции: 8 ч.;
  - практические занятия: 24 ч.;
  - контактная работа во время теоретического обучения: 1,85 ч.
- Объем самостоятельной работы студента определен учебным планом.

## **8. Содержание дисциплины (модуля), структурированное по темам**

**Тема 1.** Корпусная лингвистика как часть цифровой гуманитаристики, направление научно-практической деятельности. Электронные ресурсы: библиотека текстов, словарь, база данных, корпус. История корпусной лингвистики.

**Тема 2.** Корпус как инструмент филологических исследований. Типология корпусов.

**Тема 3.** Понятие аннотации (разметки). Типы аннотирования. Экстралингвистическая аннотация.

**Тема 4.** Репрезентативность и сбалансированность как базовые характеристики корпуса. Объем корпуса. Сравнение разных по объёму корпусов, понятие *ipm*.

**Тема 5.** Морфологическая разметка. Морфологический стандарт НКРЯ.

**Тема 6.** Проблемы морфологического аннотирования русского языка. Понятие грамматической омонимии. Способы снятия омонимии.

**Тема 7.** Семантическая разметка. Принципы семантической разметки в НКРЯ.

**Тема 8.** Синтаксическая разметка. Корпус СинТагРус.

**Тема 9.** Учебные корпуса. Специфика материала. Виды разметки учебного корпуса. Многоязычные (параллельные) корпуса. Способы применения корпусов в преподавательской деятельности.

**Тема 10.** Мультиязычные корпуса. Специфика материала. Виды разметки мультиязычного корпуса. Диалектные корпуса.

**Тема 11.** Лингвистические исследования на базе корпусов. Исследование художественного текста. Corpus-based и corpus-driven подходы. Количественные исследования на базе корпусов.

**Тема 12.** Ресурсы для самостоятельного создания корпуса.

## **9. Текущий контроль по дисциплине**

Текущий контроль по дисциплине проводится путем контроля посещаемости, выполнения домашних заданий, опроса на занятиях, тестов по лекционному материалу, проведения и оценки практических работ и фиксируется в форме контрольной точки не менее одного раза в семестр.

### Примерные задания для проведения текущего контроля.

*Тестовые задания множественного выбора:*

1. Обязательными свойствами хорошего лингвистического корпуса являются:

- а) представительность;
- б) присутствие разметки;
- в) снятая грамматическая омонимия;
- г) баланс между разными типами текстов.

2. Разметка Обучающего корпуса отличается от аннотации других частей НКРЯ тем что:

- а) она содержит разметку ошибок;
- б) она проведена в соответствии со школьной грамматикой;
- б) в ней присутствует синтаксическая разметка;

г) она произведена вручную.

*Тестовые задания открытого типа:*

1. Интерпретация корпусных данных возможна, если исследователю известен \_\_\_\_\_ корпуса, по которому велся поиск.
2. Нормализация значений обычно предполагает вычисление отношения количества вхождений к объему корпуса, умноженного на \_\_\_\_\_.

*Примеры домашних заданий.*

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

*Примеры заданий для работы в аудитории.*

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

## **10. Порядок проведения и критерии оценивания промежуточной аттестации**

**Зачет в восьмом семестре** проходит в устной форме. От его сдачи освобождаются студенты, показавшие стабильно высокие результаты на практических занятиях, не имеющие пропусков по неуважительным причинам. Билет содержит теоретический вопрос и практическое задание.

*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 *irm* 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'.

### **11. Учебно-методическое обеспечение**

- а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=14411>
- б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.
- в) План семинарских / практических занятий по дисциплине.
- г) Методические указания по организации самостоятельной работы студентов.

### **12. Перечень учебной литературы и ресурсов сети Интернет**

- а) основная литература:
  - Захаров В.П., Богданова С.Ю. [Корпусная лингвистика: Учебник для студентов направления «Лингвистика»](#). 2-е изд., перераб. и дополн. – СПб.: СПбГУ. РИО. Филологический факультет, 2013. – 148 с.
  - Копотев М.В. [О некоторых следствиях корпусной лингвистики для общей теории языка](#) // Филологический класс. 2021. Т. 26. № 2. С. 90–102.
  - 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.
  - 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.
- 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. № 4. С. 25–54.

– 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.

– Копотев М. В. Введение в корпусную лингвистику. Учебное пособие для студентов филологических и лингвистических специальностей университетов / М. В. Копотев. – Прага, 2014. – 195 с.

в) ресурсы сети Интернет:

– Русская корпусная грамматика <http://rusgram.ru/index>

– Русский мультимедийный дискурс <https://multidiscourse.ru/main/>

– EAGLES. Preliminary recommendations on Corpus Typology. EAG--TCWG--СТУР/P. Version of May, 1996 // <http://www.ilc.cnr.it/EAGLES/corpus/typology.html>

– Левинзон А. И. Введение в корпусную лингвистику [https://openedu.ru/course/hse/CORPUS/?session=fall\\_2020](https://openedu.ru/course/hse/CORPUS/?session=fall_2020)

### **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);

– 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](https://losl.tsu.ru/losl_search)

#### **14. 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.

#### **15. Информация о разработчиках**

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