

Федеральное государственное автономное образовательное учреждение высшего образования

**«Российский университет дружбы народов»
(РУДН)**

Рекомендовано МСЧН/МО

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Квантитативная лингвистика и ИКТ (на английском языке)

**Рекомендуется для направления подготовки/специальности:
45.04.02 Лингвистика**

Направленность программы (профиль):

Translator and Interpreter for Public Services and Institutions

1. Цель освоения дисциплины

Корпусная лингвистика - изучение языка, находящего отражение в корпусах или текстах. Данный метод представляет собой систематический подход к получению набора абстрактных правил, по которым один естественный язык соотносится с другим. Изначально корпуса составлялись вручную, но в настоящее время они создаются с помощью программного обеспечения.

Современные лингвистические исследования все чаще прибегают к количественному анализу: ученые подсчитывают или измеряют различные аспекты языка и речи и используют эти данные, чтобы ответить на вопросы о том, как работает язык. Данный курс поможет студентам описать, визуализировать и провести базовый статистический анализ лингвистических данных с помощью статистического программного обеспечения с открытым исходным кодом.

Результаты освоения дисциплины

Курс способствует развитию ряда компетенций, в том числе (согласно образовательным стандартам РФ):

В области универсальных компетенций студент:

- Способен управлять проектом на всех этапах его жизненного цикла. (УК-2)
- Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели (УК-3)
- Способен применять современные коммуникативные технологии на государственном языке Российской Федерации и иностранном(ых) языке(ах) для академического и профессионального взаимодействия. (УК-4)
- Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач;
- проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных. (УК-7)

В области общепрофессиональных компетенций студент:

- Способен учитывать в практической деятельности специфику иноязычной научной картины мира и научного дискурса в русском и изучаемом иностранном языках; (ОПК-2)

По завершении курса студент:

владеет распространенным программным обеспечением и применяет его надлежащим образом

может критически оценивать количественный лингвистический анализ

принимает правильные решения в области методологии при планировании будущих исследовательских проектов

владеет навыками управления данными, аналитической работой и результатами анализа

владеет навыками обобщения, визуализации и анализа данных с помощью различных методов

2. Общая трудоемкость дисциплины

Общая трудоемкость дисциплины составляет 2 зачетные единицы.

Виды учебной работы	Всего, ак.ч.	Учебные модули			
		1	2	3	4

Контактная работа, ак.ч.	18	18			
В том числе:					
Лекции					
Семинары	18	18			
Лабораторные работы					
Самостоятельная работа обучающихся, ак.ч.	54	54			
Контроль (зачет)					
Общая трудоемкость: 72 ак.ч., 2 зач. ед.	72	72			
	2	2			

3. Содержание дисциплины

3.1 Разделы.

№	Разделы	Темы
	Модуль 1	Квантитативная лингвистика
1.	Введение в курс и методологию научной работы	Квантитативные методы. Основы методологии научной работы: правильная постановка вопросов.
2.	Описательная статистика.	Обобщение данных с помощью чисел: меры центральной тенденции, дисперсии, стандартной погрешности и интервалов. Дискретные и непрерывные распределения.
3.	Корреляционный анализ и регрессия	Виды корреляции. Изучение взаимосвязи между двумя переменными измерения
4.	Программное обеспечение для квантитативной лингвистики	SPSS, Voyant Tools, инструменты для анализа цифрового контента
	Модуль 2	Корпусная лингвистика
5.	Создание корпуса	Ручное создание одноязычного корпуса DIV. Автоматическое создание двуязычного сопоставительного корпуса DIV.
6.	Кодирование и аннотирование корпусов	Исследования перевода на основе корпусов. Структурирование информации. Языковые и текстовые данные. Корпусная лингвистика для междисциплинарных исследований: https://dh2018.adho.org/corpus-linguistics-for-multidisciplinary-research-coptic-scriptorium-as-case-study/ (Coptic Scriptorium As Case Study)
7	Создание документа XML TEI	Добавление простого заголовка. Добавление лингвистической аннотации. Индексация корпуса.
8	Инструменты корпусной лингвистики и корпусного анализа	Компиляция текстов. Аннотирование. Управление корпусом. Список слов. Конкорданс.
9	Создание многоязычных корпусов	Сопоставление корпусов. Параллельные корпуса. Выравнивание. Контрольный список по созданию корпусов Корпусная лингвистика – YouTube:

		https://www.youtube.com/watch?v=YH6RcHzNDmQ Исследование параллельных корпусов юридических текстов
10	Многоязычные корпусы юридических текстов	Корпус юридических текстов ООН. Презентация проектов.

4 Разделы дисциплины и практические занятия

№	Раздел	Семи нары	Самостоя тельное обучение	Общая трудоемк ость, часов
1.	Введение в курс и методологию научной работы	2	4	6
2.	Описательная статистика.	2	4	6
3.	Корреляционный анализ и регрессия	2	4	6
4.	Программное обеспечение для квантитативной лингвистики	2	6	8
5.	Создание корпуса. Проект по созданию корпуса.	2	4	6
6.	Кодирование и аннотирование корпусов	2	4	6
7.	Одноязычные и многоязычные корпусы.	2	3	5
8.	Конкорданс.	2	4	6
9.	Задания и проекты. Презентации.	2	6	7

4. Семинары

№ п/п	Раздел	Темы и вопросы для обсуждения	Общая трудое мкость, часов
1.	Введение в курс и методологию научной работы	Квантитативные методы. Основы методологии научной работы: правильная постановка вопросов.	2
2.	Описательная статистика.	Обобщение данных с помощью чисел: меры центральной тенденции, дисперсии, стандартной погрешности и интервалов. Дискретные и непрерывные распределения.	2
3	Корреляционный анализ и регрессия	Виды корреляции. Изучение взаимосвязи между двумя переменными измерения	2
4	Программное обеспечение для квантитативной лингвистики	SPSS, Voyant Tools, инструменты для анализа цифрового контента	2
5	Создание корпуса	Создание одноязычного и двуязычного сопоставительного корпуса.	2
6	Кодирование и аннотирование корпусов	Корпусная лингвистика и перевод. Практические задания.	2
7	Задания	Разметка текста. Поиск по корпусу.	2
8	Корпусные	Конкорданс, сочетания, кластеры.	2

	инструменты и анализ.		
9	Создание многоязычного корпуса.	Учебные примеры. Параллельные корпуса.	2

5. Мастер-классы: Квантитативная лингвистика, корпус и конкорданс

п/п	№	Рзадел	Вид деятельности	Трудоемкость
	1.	Задание 1	Рассмотрение учебной ситуации	3
	2.	Задание 2	Рассмотрение учебной ситуации	3

6. Материально-техническое обеспечение.

Учебная аудитория, оснащенная комплектом специализированной мебели; магнитно-маркерная доска; технические средства: портативный мультимедийный проектор, ноутбук, проекционный экран, стабильное беспроводное подключение к сети Интернет (10 Мбит/с).

7. Информационное обеспечение дисциплины:

- a. Corpus Linguistics: The Basics – YouTube
- b. <https://www.youtube.com/watch?v=32RjJ-1A-8Q>
- c. What is CORPUS LINGUISTICS? What does CORPUS LINGUISTICS ...
- d. <https://www.youtube.com/watch?v=d3InPxxwZESU>
- e. Increasing the impact of corpus linguistics in disciplinary education ...
- f. <https://www.youtube.com/watch?v=iP9iJ9vHQoM>
- g. A Glossary of Corpus Linguistics - Edinburgh University Press
- h. <https://edinburghuniversitypress.com/book-a-glossary-of-cor>.
- i. A Glossary of Corpus Linguistics
- j. https://books.google.ru/books/about/A_Glossary_of_Corpus_Linguistics.html?id=bQFNAGS9_NkC&redir_esc=y

8. Учебно-методическое и информационное обеспечение дисциплины

Основная литература:

1. Gries, S.(2009). Statistics for Linguistics with R. De Gruyter Mouton.
2. Johnson, Keith. (2008). Quantitative Methods in Linguistics. Blackwell.
3. Tognini-Bonelli, T. 2001. Corpus Linguistics at Work. Amsterdam: John Benjamins.

Дополнительная литература:

2. Baayen, R.H. 2008. Analyzing Linguistic Data: A Practical Introduction to Statistics Using R. Cambridge: Cambridge University Press.
3. Oakes, M. 1998. Statistics for Corpus Linguistics. Edinburgh: Edinburgh University Press.
4. McEnery, T., Xiao, R.Z. and Tono, Y. 2006. Corpus-based Language Studies: An Advanced Resource Book. London: Routledge.
5. Garside, R., Leech, G. and McEnery, T. (eds) 1997. Corpus Annotation: Linguistic Information from Computer Text Corpora. Harlow: Longman.
6. Semino, E. and Short, M. 2004. Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing. London: Routledge.

9. Методические рекомендации по организации изучения дисциплины.

Занятия посвящены введению в специальную терминологию и понятия, связанные с квантитативной лингвистикой.

Занятия проводятся два часа в неделю, включая лекции и семинары.

В периоды самостоятельной работы студентам предлагается рассмотреть и расширить теоретический материал, а также выполнить некоторые задания и упражнения. Для этого в начале курса лектор предоставляет студентам подборку литературы и возможных тем и упражнений, которые необходимо выполнить.

Студенты должны выполнить проект по квантитативной лингвистике. Методические указания даются на первой неделе занятий.

На лекциях вводятся основные понятия, теории и методы работы.

Студенты читают заданные тексты и/или другие материалы и посещают семинары, подготовившись к обсуждению.

Задания, упражнения и перевод различных типов текстов для разных аудиторий обсуждается и проверяется в небольших группах.

Студенты обязаны выполнять задания, данные на дом, и посещать занятия, подготовившись к участию в них.

Проект по квантитативной лингвистике выполняется в небольших группах (3-4 студента) и проверяется на трех семинарах.

10. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Оценка включает в себя различные системы оценивания знаний с целью достижения объективности. Периодически выполняемая работа и участие в занятиях влияют на итоговую оценку. Студенты должны будут активно участвовать в занятиях, а также во всех других онлайн- и групповых мероприятиях. Посещаемость на уровне 80% является обязательной. Оценка включает в себя:

1. Посещение занятий
2. Участие в занятиях
3. Сданные задания (не менее 2 за каждый раздел до окончания срока)
4. Групповой проект, который должен быть сдан до 16 недели.
5. Два письменных теста, которые состоят из двух частей (7 и 16 неделя):
6. Вопросы по теоретическому содержанию курса.
7. Комментарии к конкретной ситуации, анализу текста и/или переводу.

Раздел	Посещение лекций	Посещение практических занятий	Количество сданных работ	Текущий контроль и итоговая аттестация	Разделы курса и баллы
	2	4	4		10
	2	4	4		10
	2	4	4		10
	2	4	4		10
Создание корпуса. Проект по созданию		4	4		8

корпуса.					
Кодирование корпуса	2	4	4		10
Инструменты для кодирования и анализа корпуса	2		4		6
Одноязычные и многоязычные корпуса.	2	4	4		10
Конкорданс	2	4	4		10
Всего: 100 баллов	16	32	36	16	

Таблица соответствия баллов и оценок

Баллы БРС	Традиционные оценки РФ	Оценки ECTS
95-100	5	A
86-94	B	
69-85	4	C
61-68	3	D
51-60	E	
31-50	2	FX
0-30	F	
51 - 100	Зачет	Passed

Материалы для оценки уровня освоения учебного материала дисциплины (оценочные материалы), включающие в себя вопросы по темам дисциплины типовые контрольные задания и задания для семинарских и практических занятий, примеры тестовых заданий, темы докладов / рефератов / курсовых работ разработаны в полном объеме и доступны для обучающихся на странице дисциплины в ТУИС РУДН.

Разработчики:

доценты кафедры иностранных языков Ф.Н. Новиков
Н.Н. Удина, к.ф.н.

Руководитель программы

Заведующая кафедрой
иностраннх языков Юридического института
д.ф.н., профессор А.А. Атабекова

Заведующая кафедрой

иностраннх языков Юридического института
д.ф.н., профессор А.А. Атабекова

**FEDERAL STATE AUTONOMOUS EDUCATIONAL INSTITUTION OF HIGHER
EDUCATION**

**PEOPLES' FRIENDSHIP UNIVERSITY OF RUSSIA
(RUDN UNIVERSITY)**

COURSE SYLLABUS

Subject: *Quantitative Linguistics and ICT*

Higher Education Field *45.04.02 Linguistics*

Specialization *Translator and Interpreter for Public Services and Institutions*

COURSE SYLLABUS

1. Course Goals

Corpus linguistics is the study of language as expressed in samples (corpora) or "real world" text. This method represents a systematic approach to obtaining a set of abstract rules by which a natural language is governed or else relates to another language. Originally done by hand, corpora are now largely derived by an automated process.

Modern linguistic research increasingly involves quantitative analysis: we count or measure various aspects of linguistic behavior, and want to use this data to answer questions about the way language works. This course will guide students through the description, visualization, and basic statistical analysis of both discrete and continuous linguistic data using open-source statistical software environment.

Competence requirements on the course completion:

The course is expected to contribute to the development of a number of competences including the following ones as required by the Russian Higher Education Standard for Training in Linguistics, so that

in the field of Universal competences the student:

- is able to manage the project at all stages of its lifecycle (UC-2)
- is able to create and lead a team, designing a team strategy aimed at achieving the goals (UC-3)
- is able to apply modern communication techniques using the official language of the Russian Federation and foreign language(s) in academic and professional interaction. (UC-4)
- is able to find the necessary sources of information and data, comprehend, analyze, memorize information while using digital devices and algorithms; use various types of information from different sources in an effective way in problem-solving; assess the information, its authenticity, infer and deduct based on the input data and information. (UC-7)

in the field of linguist's core professional competences the student:

- is able to take into consideration the specifics of foreign scientific worldview and academic discourse in Russian and foreign languages they are currently learning while engaging in their professional activities (GPC-2)

On completion of the course the student

knows the conceptual underpinnings of some common statistical tests, and apply them appropriately

can critically evaluate quantitative analyses in a range of linguistic subfields

can make appropriate methodological choices when planning future research projects

masters skills in the management of data, analytical work, and analysis results

masters skills to summarize, visualize, and otherwise explore data using a variety of methods

2. The course total workload

The course total workload is equal to 2 credits.

Academic studies activities	Total hours	Semesters			
Sessions	18	18			
Including					
Lectures					
Seminars	18	18			
Workshops					
Independent learning	54	54			
Course paper					

Mid-term assessment					
Total workload 72 academic hours, 2 credits	72	72			
	2	2			

3. Course Contents

4.1 Course Units

No	Units	Contents
	Module 1	Quantitative Linguistics
1.	Introduction to the course & Research design	Quantitative methods. Fundamentals of research design: asking answerable questions.
2.	Descriptive statistics.	Summarizing data with numbers: measures of central tendency, dispersion, standard error, and confidence intervals. Discrete and continuous distributions.
3.	Correlation Analysis and regression	Types of correlation. Exploring the relationship between two measurement variables
4.	Software for quantitative linguistics	SPSS, voyant tools, tools for digital content analysis
	Module 2	Corpus Linguistics
5.	Corpus building	Manual creation of a DIV monolingual corpus. Automatic creation of a DIV bilingual comparable corpus.
6.	Corpus encoding and annotation	Corpus- based translation studies. Structural information. Text-linguistic information. Tasks Corpus Linguistics for Multidisciplinary Research: Coptic Scriptorium As Case Study https://dh2018.adho.org/corpus-linguistics-for-multidisciplinary-research-coptic-scriptorium-as-case-study/
7	Creating an XML TEI document	Adding a simple header. Adding linguistic annotation. Indexing the corpus.
8	Corpus tools and corpus analysis	Text acquisition. Annotation. Corpus management. Wordlist. Concordances.
9	Creating multilingual corpora	Comparable corpora. Parallel corpora. Alignment. Corpus creation checklist Corpus linguistics – YouTube . https://www.youtube.com/watch?v=YH6RcHzNDmQ Exploring parallel corpus of legal texts
10	Multilingual corpora of legal texts	UN Corpus of legal texts. Projects presentations

5.1 Course modules and teaching activities

No	Course unit	Lectures	Seminars	Independent Learning	Total workload, hours
1.	Introduction to the course & Research design		2	4	6
2.	Descriptive statistics.		2	4	6

3.	Correlation Analysis and regression		2	4	6
4.	Software for quantitative linguistics		2	6	8
5.	Corpus building. Corpus building project		2	4	6
6.	How to encode a corpus		2	4	6
7.	Corpus tools and analysis		2	3	5
8.	Monolingual corpora. Multilingual corpora.		2	4	6
9.	Concordance		2	6	7
10.	Tasks and projects. Presentations		2	4	6

4. Small group training

№ п/п	Course unit	Topics and issues for discussion	Total workload, hours
1.	Introduction to the course & Research design	Quantitative methods. Fundamentals of research design: asking answerable questions.	2
2.	Descriptive statistics.	Summarizing data with numbers: measures of central tendency, dispersion, standard error, and confidence intervals. Discrete and continuous distributions.	2
3	Correlation Analysis and regression	Types of correlation. Exploring the relationship between two measurement variables	2
4	Software for quantitative linguistics	SPSS, voyant tools, tools for digital content analysis	2
5	Corpus building	Creation of a monolingual and bilingual comparable corpus.	2
6	Corpus encoding	Corpus- based translation studies. Tasks	2
7	Tasks	Marking- up text structure. Searching the corpus.	2
8	Corpus tools and corpus analysis	Concordances, collocations, clusters.	2
9	Creating multilingual corpora	Case study. Parallel corpora	2

5. Workshops: Quantitative Linguistics, Corpora and Concordance

п/п	Course units	Activities	Workload
.	Task 1	Case Study	3
.	Task 2	Case Study	3

6. Software and technical support requirements

A classroom is equipped with TV, VCR, and transparency projector. CD players and DVD players are available upon request.

7. Recommendations to study the course

Lessons will be dedicated to introduce specific terminology and concepts related to Quantitative Linguistics.

The class will meet two hours per week, including lectures and seminars.

In the independent-study periods, students are asked to review and expand the theoretical explanations as well as to complete some activities and exercises. To this end, the lecturer provides students with a selection of readings and possible topics and activities to be completed at the beginning of the course.

The students are expected to complete the project on Quantitative Linguistics. Instructions shall be given the first week of class.

Main concepts, theories and conventions will be explained at lectures.

Students are required to read the texts assigned and/or other materials assigned and come to seminars ready for discussion.

Activities, exercises and the translation of different types of texts for varied readerships will be discussed and checked in class in smaller groups.

Students are required to completed the tasks given at home and come to class ready to participate

A project on Quantitative Linguistics will be complete in small groups (3-4 students), and reviewed in 3 workshops.

9. Assessment and Grading

Assessment will include different evaluation systems in order to reach an overall continuous and formative assessment. Continuous work and participation will determine the final mark. Students will have to participate actively and effectively in classes as well as in all other online and group activities. 80% attendance is compulsory.

Assessment includes:

1. Class attendance -
2. Class participation
3. Activities handed (minimum 2 per unit before deadline)
4. Project in group, to be submitted before week 16.
5. Two written tests that will consist of two parts (week 7 and week 16):
 - questions about the theoretical contents.
 - comments on a specific situation, case/ analysis of a text and /or translation in

class.

Grading

Course Unit	Lecture attendance	Class participation	Activiti es handed	Mid-term and Final Tests	Co-urse Units and points
	2	4	4		10
	2	4	4		10
	2	4	4		10
	2	4	4		10
. Corpus building. Corpus building project		4	4		8
How to encode a corpus	2	4	4		10

Corpus tools and analysis	2		4		6
Monolingual corpora. Multilingual corpora.	2	4	4		10
Concordance	2	4	4		10
Total: 100 points	16	32	36	16	

Assessment

Points	Marks in accordance with Russian system of grading	ECTS
95 – 100		A
86 – 94	5	B
69 – 85	4	C
61 – 68		D
51 – 60	3	E
31 – 50		FX
0 – 30	2	F
51 – 100	Passed	Passed

10. Recommended Sources

Main readings:

Gries, S. (2009). Statistics for Linguistics with R. De Gruyter Mouton.

Johnson, Keith. (2008). Quantitative Methods in Linguistics. Blackwell.

Tognini-Bonelli, T. 2001. Corpus Linguistics at Work. Amsterdam: John Benjamins.

Other readings

Baayen, R.H. 2008. Analyzing Linguistic Data: A Practical Introduction to Statistics Using R. Cambridge: Cambridge University Press.

Oakes, M. 1998. Statistics for Corpus Linguistics. Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R.Z. and Tono, Y. 2006. Corpus-based Language Studies: An Advanced Resource Book. London: Routledge.

Garside, R., Leech, G. and McEnery, T. (eds) 1997. Corpus Annotation: Linguistic Information from Computer Text Corpora. Harlow: Longman.

Semino, E. and Short, M. 2004. Corpus Stylistics: Speech, Writing and Thought Presentation in a Corpus of English Writing. London: Routledge.

Web-sites

Corpus Linguistics: The Basics – YouTube

<https://www.youtube.com/watch?v=32RjJ-1A-8Q>

What is CORPUS LINGUISTICS? What does CORPUS LINGUISTICS ...

<https://www.youtube.com/watch?v=d3InPxwZESU>

Increasing the impact of corpus linguistics in disciplinary education ...

<https://www.youtube.com/watch?v=iP9iJ9vHQoM>

A Glossary of Corpus Linguistics - Edinburgh University Press

<https://edinburghuniversitypress.com/book-a-glossary-of-cor>.

A Glossary of Corpus Linguistics

https://books.google.ru/books/about/A_Glossary_of_Corpus_Linguistics.html?id=bQFNnGS9_NkC&redir_esc=y

Course developer

Associate professors of Department of Foreign Languages
Institute of Law
Novikov F.N.

N. N. Udina, in Linguistics

Head of Foreign Languages Department

A.A. Atabekova, Full Professor, Dr.habil in Comparative Linguistics