

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение высшего
образования
"САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ"

Кафедра № 63

УТВЕРЖДАЮ

Ответственный за образовательную
программу

к.ф.н., доц.

(должность, уч. степень, звание)

М.А. Чиханова

(инициалы, фамилия)

(подпись)

«27» июня 2024 г

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Корпусная лингвистика»

(Наименование дисциплины)

Код направления подготовки/ специальности	45.03.02
Наименование направления подготовки/ специальности	Лингвистика
Наименование направленности	Теоретическая и прикладная лингвистика
Форма обучения	заочная
Год приема	2024

Санкт-Петербург– 2024

Лист согласования рабочей программы дисциплины

Программу составил (а)

доц., к. ф. н.

(должность, уч. степень, звание)



27.05.2024

(подпись, дата)

Е.Ю. Дубинина

(инициалы, фамилия)

Программа одобрена на заседании кафедры № 63

«27» мая 2024 г, протокол № 10

Заведующий кафедрой № 63

к. ф. н., доц.

(уч. степень, звание)



27.05.2024

(подпись, дата)

М.А. Чиханова

(инициалы, фамилия)

Заместитель декана факультета №6 по методической работе

проф., д. и. н., доц.

(должность, уч. степень, звание)



27.05.2024

(подпись, дата)

Л.Ю. Гусман

(инициалы, фамилия)

Аннотация

Дисциплина «Корпусная лингвистика» входит в образовательную программу высшего образования – программу бакалавриата по направлению подготовки/специальности 45.03.02 «Лингвистика» направленности «Теоретическая и прикладная лингвистика». Дисциплина реализуется кафедрой «№63».

Дисциплина нацелена на формирование у выпускника следующих компетенций:

УК-2 «Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений»

ПК-3 «Владение методами проведения лингвистических экспертиз»

ПК-5 «Способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-лингвистических технологий и с учетом основных требований информационной безопасности»

ПК-7 «Владение основными математико-статистическими методами обработки лингвистической информации с учетом элементов программирования и автоматической обработки лингвистических корпусов»

Содержание дисциплины охватывает круг вопросов, связанных с анализом языка на базе корпусных данных. Корпусная лингвистика как направление прикладной и компьютерной лингвистики. Основные понятия корпусной лингвистики. История создания лингвистических корпусов. Корпусная лингвистика в России. Корпусная русистика.

Типология корпусов. Фундаментальные корпусы - Британский Национальный корпус, Национальный корпус русского языка. Параллельные корпусы. Учебные (Learner) корпусы. Обзор иных существующих корпусов различных типов. Зарубежные национальные корпусы. Корпусы русского языка. Специальные корпусы.

Создание корпусов. Проблемы репрезентативности и выборки текстов.

Разметка. Средства создания и разметки корпусов. Типы разметки. Металингвистическая разметка. Автоматический морфологический и синтаксический анализ в корпусах на примере национальных корпусов. Стандартизация в корпусной лингвистике.

Использование корпусов – корпусный подход.

Корпусные исследования на примерах. Лексические исследования, базирующиеся на корпусах. Грамматические исследования. Семантические исследования, базирующиеся на корпусах. Использование корпусов в социологии, политологии, методике иностранных языков, переводе и др.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточная аттестация в форме зачета.

Общая трудоемкость освоения дисциплины составляет 2 зачетных единицы, 72 часа.

Язык обучения по дисциплине «русский»

1. Перечень планируемых результатов обучения по дисциплине

1.1. Цели преподавания дисциплины

выработка у студентов навыков и умений работы с различными источниками информации в одно- и многоязычной среде;

формирование системы знаний о современных методах компьютерного лингвистического анализа;

развитие у студентов умений организовать свою исследовательскую, лингводидактическую и методическую работу с использованием современных российских и зарубежных корпусных ресурсов;

формирование системы знаний о существующих лингвистических корпусах, ориентированных на выполнение различных операций с текстом.

1.2. Дисциплина входит в состав части, формируемой участниками образовательных отношений, образовательной программы высшего образования (далее – ОП ВО).

1.3 Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения ОП ВО.

В результате изучения дисциплины обучающийся должен обладать следующими компетенциями или их частями. Компетенции и индикаторы их достижения приведены в таблице 1.

Таблица 1 – Перечень компетенций и индикаторов их достижения

Категория (группа) компетенции	Код и наименование компетенции	Код и наименование индикатора достижения компетенции
Универсальные компетенции	УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.У.1 уметь проводить анализ поставленной цели и формулировать задачи, которые необходимо решить для ее достижения
Профессиональные компетенции	ПК-3 Владение методами проведения лингвистических экспертиз	ПК-3.З.1 знать методы лингвистических экспертиз, методы фоносемантического и ассоциативного анализа лингвистических объектов ПК-3.В.1 владеть методами проведения лингвистических экспертиз, навыками анализа макроструктуры и микроструктуры текстов в практике решения экспертных задач
Профессиональные компетенции	ПК-5 Способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с	ПК-5.З.1 знать основные принципы организации информационно-поисковых систем с применением информационно-лингвистических технологий ПК-5.В.1 владеть технологиями построения тезаурусов и онтологий

	применением информационно-лингвистических технологий и с учетом основных требований информационной безопасности	
Профессиональные компетенции	ПК-7 Владение основными математико-статистическими методами обработки лингвистической информации с учетом элементов программирования и автоматической обработки лингвистических корпусов	ПК-7.3.1 знать основные математико-статистические методы обработки лингвистической информации ПК-7.У.1 уметь с учетом элементов программирования и автоматической обработки лингвистических корпусов обрабатывать лингвистическую информацию ПК-7.В.1 владеть навыками программирования и навыками автоматической обработки корпусов

2. Место дисциплины в структуре ОП

Дисциплина базируется на знаниях, ранее приобретенных студентами при изучении следующих дисциплин:

- Базы данных
- Информационный поиск и извлечение информации
- Технологии обработки текстов и звучащей речи
- Формальные модели в лингвистике

Знания, полученные при изучении материала данной дисциплины, имеют как самостоятельное значение, так и используются при написании ВКРБ.

3. Объем и трудоемкость дисциплины

Данные об общем объеме дисциплины, трудоемкости отдельных видов учебной работы по дисциплине (и распределение этой трудоемкости по семестрам) представлены в таблице 2.

Таблица 2 – Объем и трудоемкость дисциплины

Вид учебной работы	Всего	Трудоемкость по семестрам
		№10
1	2	3
Общая трудоемкость дисциплины, ЗЕ/ (час)	2/ 72	2/ 72
Из них часов практической подготовки	9	9
Аудиторные занятия, всего час.	20	20
в том числе:		
лекции (Л), (час)	8	8
практические/семинарские занятия (ПЗ), (час)	12	12
лабораторные работы (ЛР), (час)		
курсовой проект (работа) (КП, КР), (час)		

экзамен, (час)		
Самостоятельная работа , всего (час)	52	52
Вид промежуточной аттестации: зачет, дифф. зачет, экзамен (Зачет, Дифф. зач, Экз.**)	Зачет	Зачет

Примечание: ** кандидатский экзамен

4. Содержание дисциплины

4.1. Распределение трудоемкости дисциплины по разделам и видам занятий.

Разделы, темы дисциплины и их трудоемкость приведены в таблице 3.

Таблица 3 – Разделы, темы дисциплины, их трудоемкость

Разделы, темы дисциплины	Лекции (час)	ПЗ (СЗ) (час)	ЛР (час)	КП (час)	СРС (час)
Семестр 10					
Раздел 1. Тема 1.1. Тема 1.2. Тема 1.3. Тема 1.4. Тема 1.5.	1	2			10
Раздел 2. Тема 2.1. Тема 2.2. Тема 2.3. Тема 2.4.	1	2			10
Раздел 3. Тема 3.1. Тема 3.2. Тема 3.3. Тема 3.4.	2	2			10
Раздел 4. Тема 4.1. Тема 4.2. Тема 4.3. Тема 4.4.	2	2			10
Раздел 5. Тема 5.1. Тема 5.2. Тема 5.3. Тема 5.4.	2	4			12
Итого в семестре:	8	12			52
Итого:	8	12	0	0	52

Практическая подготовка заключается в непосредственном выполнении обучающимися определенных трудовых функций, связанных с будущей профессиональной деятельностью.

4.2. Содержание разделов и тем лекционных занятий.

Содержание разделов и тем лекционных занятий приведено в таблице 4.

Таблица 4 – Содержание разделов и тем лекционного цикла

Номер раздела	Название и содержание разделов и тем лекционных занятий
1	<p>Базовые понятия и термины корпусной лингвистики</p> <p>Тема 1.1. Объект и предмет корпусной лингвистики.</p> <p>Тема 1.2. Содержание понятия компьютерный корпус как базового понятия корпусной лингвистики.</p> <p>Тема 1.3. Компьютерный корпус как объект лингвистического аннотирования.</p> <p>Тема 1.4. Корпусный менеджер как система с программными средствами, применяемая для поиска и статистической обработки текстов.</p> <p>Тема 1.5. Глобальные информационно-поисковые системы и примеры их использования в лингвистических исследованиях</p>
2	<p>Компьютерный лингвистический корпус как база текстовых данных</p> <p>Тема 2.1. Компьютерный корпус как объемная, представленная в электронном виде и размеченная база текстовых данных</p> <p>Тема 2.2. Репрезентативность объема корпусного материала для решения лингвистических задач</p> <p>Тема 2.3. Понятия мегакорпус и подкорпус</p> <p>Тема 2.4. Сопоставления объектных и предметных областей традиционной и корпусной лингвистики</p>
3	<p>Технологический процесс создания лингвистического корпуса</p> <p>Тема 3.1. Общая характеристика лингвистических информационно-программных объектов</p> <p>Тема 3.2. Технологические этапы создания корпуса</p> <p>Тема 3.3. Использование при создании корпусов программ автоматического аннотирования</p> <p>Тема 3.4 Проблемные зоны создания лингвистического корпуса</p>
4	<p>Корпусная типология</p> <p>Тема 4.1. Краткая история создания зарубежных и отечественных корпусных собраний</p> <p>Тема 4.2. Корпусы первого и второго поколения</p> <p>Тема 4.3. Типы корпусов</p> <p>Тема 4.4. Примеры корпусных проектов каждого типа</p>
5	<p>Использование корпусов</p> <p>Тема 5.1. Корпус как поисковая система</p> <p>Тема 5.2. Языки запросов</p> <p>Тема 5.3. Корпусные исследования</p> <p>Тема 5.4. Корпусная лингвистика и проблемы перевода</p>

4.3. Практические (семинарские) занятия

Темы практических занятий и их трудоемкость приведены в таблице 5.

Таблица 5 – Практические занятия и их трудоемкость

№ п/п	Темы практических занятий	Формы практических занятий	Трудоемкость, (час)	Из них практической подготовки,	№ раздела дисцип

				(час)	лины
Семестр 10					
1	Базовые понятия и термины корпусной лингвистики	Решение ситуационных задач	2	2	1
2	Компьютерный лингвистический корпус как база текстовых данных	Решение ситуационных задач	2	2	2
3	Технологический процесс создания лингвистического корпуса	Решение ситуационных задач	2	2	3
4	Корпусная типология	Решение ситуационных задач	2	2	4
5	Использование корпусов		4	4	5
Всего			12		

4.4. Лабораторные занятия

Темы лабораторных занятий и их трудоемкость приведены в таблице 6.

Таблица 6 – Лабораторные занятия и их трудоемкость

№ п/п	Наименование лабораторных работ	Трудоемкость, (час)	Из них практической подготовки, (час)	№ раздела дисциплины
Учебным планом не предусмотрено				
Всего				

4.5. Курсовое проектирование/ выполнение курсовой работы

Цель курсовой работы:

Часов практической подготовки:

Примерные темы заданий на курсовую работу приведены в разделе 10 РПД.

4.6. Самостоятельная работа обучающихся

Виды самостоятельной работы и ее трудоемкость приведены в таблице 7.

Таблица 7 – Виды самостоятельной работы и ее трудоемкость

Вид самостоятельной работы	Всего, час	Семестр 10, час
1	2	3
Изучение теоретического материала дисциплины (ТО)	10	10
Курсовое проектирование (КП, КР)		
Расчетно-графические задания (РГЗ)		

Выполнение реферата (Р)		
Подготовка к текущему контролю успеваемости (ТКУ)		
Домашнее задание (ДЗ)	10	10
Контрольные работы заочников (КРЗ)	20	20
Подготовка к промежуточной аттестации (ПА)	12	12
Всего:	52	52

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
Учебно-методические материалы для самостоятельной работы обучающихся указаны в п.п. 7–11.

6. Перечень печатных и электронных учебных изданий
Перечень печатных и электронных учебных изданий приведен в таблице 8.
Таблица 8– Перечень печатных и электронных учебных изданий

Шифр/ URL адрес	Библиографическая ссылка	Количество экземпляров в библиотеке (кроме электронных экземпляров)
	Дубинина Е.Ю. Лингвистические корпуса текстов. СПб. ГУАП, 2024. 76 с.	50
	Дубинина Е.Ю. Корпусная лингвистика: учебно-метод. пособие. СПб. ГУАП, 2020. 90с.	50
	Барсебян, А. А., Куприянов М. С., Степаненко В.В., Холод А.А., Методы и модели анализа данных: OLAP и Data Mining. БХВ-Петербург, 2004. - 336 с.	
	<u>Себастьян</u> Рашка, А. В. <u>Логунов</u> Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения, ДМК Пресс, 2017. https://znanium.com/catalog/document?id=341047	
	В. В. Вьюгин Математические основы машинного обучения и прогнозирования, МЦНМО, 2014, https://znanium.com/catalog/document?id=328779	
ЭБС "ИНФРА-М" http://www.znanium.com	1. J. Han and M. Kamber, Data Mining: Concepts and Techniques, Second Edition, Morgan Kaufman, 2006. (Глава 1)	
ЭБС «Лань» http://www.elan.com	2. Tom Mitchell, Machine Learning, McGraw Hill, 1997. (Глава 1)	

	<p>3. Ian H. Witten and Eibe Frank. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco, 2 editions, 2005. (Главы 1, 2, 3)</p> <p>4. Биркгоф Г. Теория решеток. - М.: Наука, 1984.</p> <p>5. Szathmary, L. Symbolic Data Mining Methods with the Coron Platform. PhD thesis, University Henri Poincaré -- Nancy 1, France, Nov 2006. (глава The Coron User Toolkit)</p> <p>6. Janez Demsar, Blaz Zupan, Gregor Leban, Tomaz Curk: Orange: From Experimental Machine Learning to Interactive Data Mining. J.-F. Boulicaut et al. (Eds.): PKDD 2004, LNAI 3202, pp. 537–539, Springer, 2004</p> <p>7. Николенко, С.И., Тулупьев, А.Л., Самообучающиеся системы. – М.: МЦНМО, 2009. – 288 с. (Глава 6)</p> <p>8. L. Zhukov, Spectral Clustering of Large Advertiser Datasets. Technical report, Overture R&D, 2003.</p> <p>9. Автоматическое порождение гипотез в интеллектуальных системах/ Сост. Е.С. Панкратова, В.К. Финн; Под общ. ред. В.К.Финна. – М.: Книжный дом «ЛИБРОКОМ», 2009. – 528 с. (Главы 1,2)</p>	
--	--	--

7. Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»

Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины приведен в таблице 9.

Таблица 9 – Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»

URL адрес	Наименование
http://www.dblang.ru/Default.aspx	База данных «Языки мира»
http://www.philol.msu.ru/~lex/corpus/	Компьютерный корпус текстов русских газет конца XX-ого века
http://narusco.ru/	Корпус русского литературного языка
http://www.artint.ru/projects/frqlist.asp	Частотный словарь С.И. Шарова
http://www.ling.helsinki.fi/projects/hanco/	Корпус русского языка ХАНКО
http://www.petrso.ru	Весь Достоевский
http://dostoevskij.karelia.ru/	Словарь-конкорданс публицистики Ф.М. Достоевского
http://www.rvb.ru	Русская виртуальная библиотека

http://feb-web.ru/feb/	Словарь-корпус языка А.С. Грибоедова
http://www.sfb441.uni-tuebingen.de/b1/en/korpora.html	Уппсальский корпус русских текстов
http://sable.ox.ac.uk/ota/	Oxford Text Archive
http://info.ox.ac.uk/bnk	British National Corpus
http://titania.cobuild.collins.co.uk/boe_info.html/	Bank of English
http://www.americancorpus.org/	Corpus of Contemporary American English
http://www.warwick.ac.uk/	British Academic Spoken English
http://www.uclouvain.be/en-cecl-icle.html	International Corpus of Learner English
http://www.ivacs.mic.ul.ie/corpora/limerick-corpus-irish-english-l-cie/	Limerick Corpus of Irish-English
https://www ldc.upenn.edu/collaborations/current-projects	Santa Barbara Corpus of Spoken American English
http://www.americanrhetoric.com/speechbank.htm	American Rhetoric Online Speech Bank

8. Перечень информационных технологий

8.1. Перечень программного обеспечения, используемого при осуществлении образовательного процесса по дисциплине.

Перечень используемого программного обеспечения представлен в таблице 10.

Таблица 10– Перечень программного обеспечения

№ п/п	Наименование
1	MS Office 2013: Номер лицензии: 62042637 Родительская программа: 92010750ZZE1506 Договор на приобретение: 402-7
2	Офис: Microsoft® Office Professional Plus 2007 Russian Тип лицензии: Academic Номер лицензии 44260430

8.2. Перечень информационно-справочных систем, используемых при осуществлении образовательного процесса по дисциплине

Перечень используемых информационно-справочных систем представлен в таблице 11.

Таблица 11– Перечень информационно-справочных систем

№ п/п	Наименование
1.	Электронно-библиотечная система Лань URL: https://e.lanbook.com/
2.	Электронно-библиотечная система Znanium URL: https://znanium.com/
3.	Информационно-правовой портал ГАРАНТ.РУ URL: http://www.garant.ru/
4.	Правовая поддержка КОНСУЛЬТАНТПЛЮС URL: http://www.consultant.ru

9. Материально-техническая база

Состав материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине, представлен в таблице 12.

Таблица 12 – Состав материально-технической базы

№ п/п	Наименование составной части материально-технической базы	Номер аудитории (при необходимости)
1	Аудитория для практических занятий	Аудитория укомплектована

	Компьютерный класс	специализированной мебелью, техническими средствами обучения, служащими для представления учебной информации большой аудитории Ауд. 34-10
2	Аудитории общего пользования (для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации)	Аудитория укомплектована специализированной мебелью, техническими средствами обучения, служащими для предоставления учебной информации большой аудитории
3	Библиотека, Интернет-класс ГУАП (для самостоятельной работы)	Помещения укомплектованы специализированной мебелью, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечено доступом в электронную информационно-образовательную среду ГУАП

10. Оценочные средства для проведения промежуточной аттестации

10.1. Состав оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине приведен в таблице 13.

Таблица 13 – Состав оценочных средств для проведения промежуточной аттестации

Вид промежуточной аттестации	Перечень оценочных средств
Зачет	Список вопросов; Задачи.

10.2. В качестве критериев оценки уровня сформированности (освоения) компетенций обучающимися применяется 5-балльная шкала оценки сформированности компетенций, которая приведена в таблице 14. В течение семестра может использоваться 100-балльная шкала модульно-рейтинговой системы Университета, правила использования которой, установлены соответствующим локальным нормативным актом ГУАП.

Таблица 14 – Критерии оценки уровня сформированности компетенций

Оценка компетенции	Характеристика сформированных компетенций
5-балльная шкала	
«отлично» «зачтено»	<ul style="list-style-type: none"> – обучающийся глубоко и всесторонне усвоил программный материал; – уверенно, логично, последовательно и грамотно его излагает; – опираясь на знания основной и дополнительной литературы, тесно привязывает усвоенные научные положения с практической деятельностью направления; – умело обосновывает и аргументирует выдвигаемые им идеи; – делает выводы и обобщения; – свободно владеет системой специализированных понятий.
«хорошо» «зачтено»	<ul style="list-style-type: none"> – обучающийся твердо усвоил программный материал, грамотно и по существу излагает его, опираясь на знания основной литературы; – не допускает существенных неточностей; – увязывает усвоенные знания с практической деятельностью направления; – аргументирует научные положения;

Оценка компетенции	Характеристика сформированных компетенций
5-балльная шкала	
	<ul style="list-style-type: none"> – делает выводы и обобщения; – владеет системой специализированных понятий.
«удовлетворительно» «зачтено»	<ul style="list-style-type: none"> – обучающийся усвоил только основной программный материал, по существу излагает его, опираясь на знания только основной литературы; – допускает несущественные ошибки и неточности; – испытывает затруднения в практическом применении знаний направления; – слабо аргументирует научные положения; – затрудняется в формулировании выводов и обобщений; – частично владеет системой специализированных понятий.
«неудовлетворительно» «не зачтено»	<ul style="list-style-type: none"> – обучающийся не усвоил значительной части программного материала; – допускает существенные ошибки и неточности при рассмотрении проблем в конкретном направлении; – испытывает трудности в практическом применении знаний; – не может аргументировать научные положения; – не формулирует выводов и обобщений.

10.3. Типовые контрольные задания или иные материалы.

Вопросы (задачи) для экзамена представлены в таблице 15.

Таблица 15 – Вопросы (задачи) для экзамена

№ п/п	Перечень вопросов (задач) для экзамена	Код индикатора
	Учебным планом не предусмотрено	

Вопросы (задачи) для зачета / дифф. зачета представлены в таблице 16.

Таблица 16 – Вопросы (задачи) для зачета / дифф. зачета

№ п/п	Перечень вопросов (задач) для зачета / дифф. зачета	Код индикатора
1	Возможные средства обработки корпусных данных	ПК-3.3.1
2	Понятие корпусной разметки	ПК-3.В.1
3	Технологические этапы создания корпуса	ПК-5.3.1
4	Возможные средства обработки корпусных данных	ПК-5.В.1
5	Лингвистические исследования, базирующиеся на корпусах	ПК-7.3.1
6	Прикладное применение корпусных материалов	УК-2.У.1
7	Общая характеристика конкордансов	ПК-7.В.1

Перечень тем для курсового проектирования/выполнения курсовой работы представлены в таблице 17.

Таблица 17 – Перечень тем для курсового проектирования/выполнения курсовой работы

№ п/п	Примерный перечень тем для курсового проектирования/выполнения курсовой работы
	Учебным планом не предусмотрено

Вопросы для проведения промежуточной аттестации в виде тестирования представлены в таблице 18.

Таблица 18 – Примерный перечень вопросов для тестов

№ п/п.	Перечень вопросов (задач) для зачета	Код индикатора
1.	<p>Выберите правильный вариант ответа: Текстовые материалы, входящие в корпус:</p> <ol style="list-style-type: none"> 1. предназначены для чтения; 2. не содержат разговорную речь; 3. относятся к одному жанру; 4. содержат разметку; 5. охватывают небольшой промежуток времени. 	УК-2.У.1
2.	<p>Запишите одним словом как называется комбинация двух или более слов, имеющих устойчивую совместную встречаемость</p>	УК-2.У.1
3.	<p>Запишите одним словом как называется присвоение каждому элементу текста особого «тега».</p>	УК-2.У.1
4.	<p>Запишите словосочетанием как называется словник, в котором каждому слову приписывается информация о частоте его встречаемости в тексте</p>	УК-2.У.1
5.	<p>Назовите объект исследования корпусной лингвистики</p>	УК-2.У.1
6.	<p>Выберите правильные варианты ответа: выберите, какие задачи решает лингвистическая экспертиза</p> <ol style="list-style-type: none"> 1. выявить пониженную или повышенную экспрессию выражений, употребляемых в тексте 2. установить онтологическое понимание времени 3. проанализировать признаки текста с целью установления авторства 4. исследовать динамику развития знаний о конфликтах <p>Ответы: 1, 3</p>	ПК-3.3.1
7.	<p>Выберите правильный вариант ответа:</p> <p>Комплексная экспертиза проводится:</p> <ol style="list-style-type: none"> а) судьей, следователем, прокурором; б) экспертами одной специальности; в) несколькими экспертами разных специальностей; г) несколькими комиссиями. 	ПК-3.3.1
8.	<p>Выберите правильный вариант ответа:</p> <p>Выберите программу, которая используется для выравнивания параллельных текстов</p> <ol style="list-style-type: none"> a. ABBYY Aligner b. Online OCR c. WIX d. Mendeley 	ПК-3.3.1
9.	<p>Напишите, какие три вида исследований содержит</p>	ПК-3.В.1

	лингвистическая экспертиза	
	Ответ: терминологические, смысловые, автороведческие	
10.	Выберите правильный вариант ответа: корпусы, содержащие материалы, которые можно использовать при проведении исследований в области американского варианта английского языка: 1. Lob Corpus; 2. COCA; 3. Brown Corpus; 4. Bank of English.	ПК-5.3.1
11.	Выберите правильный вариант ответа: Сведения, которые не являются частью текста, но содержат информацию о нем: а. Метаданные б. Коллокации с. Стоп-слова d. N-граммы	ПК-5.3.1
12.	Запишите как называется компьютерная программа, позволяющая производить синтаксическую разметку	ПК-5.В.1
13.	Установите, какой тип разметки включает следующие элементы: название текста, автор текста, год и место издания текста	ПК-5.В.1
14.	Вставьте пропущенное словосочетание _____ системы – это программные системы для хранения, поиска и выдачи интересующей пользователя информации.	ПК-5.В.1
15.	Выберите правильные варианты ответа: Выберите виды разметки, существующие в Национальном корпусе русского языка: 1. морфологическая разметка; 2. синтаксическая разметка; 3. семантическая разметка; 4. анафорическая разметка; 5. метаразметка.	ПК-7.3.1
16.	Выберите правильные варианты ответа: Метаразметка в корпусе содержит информацию: 1. о семантической структуре контекста; 2. о грамматических признаках словоформы; 3. об авторе и названии документа; 4. о времени создания документа.	ПК-7.3.1
17.	Выберите правильные варианты ответа: Корпусы текстов, объем которых составляет 1 млн. слов – это: 1. Брауновский корпус;	ПК-7.3.1

	<p>2. Ланкастер-Осло-Берген; 3. Bank of English; 4. COCA.</p>	
18.	<p>Выберите правильный вариант ответа:</p> <p>Выберите, какая программа Вы используется для парсинга:</p> <p>a. SQL b. AntConc c. ParseHub d. Topic Modelling Tool</p>	ПК-7.3.1
19.	<p>Выберите правильный вариант ответа:</p> <p>Автоматический анализ структуры текстовых данных, который позволяет осуществлять процедуру сбора синтаксических конструкций (предложений) на естественном языке из Интернет-источников – это:</p> <p>a. Парсинг b. Коллокация c. Онтология d. Тезаурус</p>	ПК-7.3.1
20.	<p>Выберите правильный вариант ответа:</p> <p>Символ * в поисковом запросе заменяет</p> <p>а) один символ; б) одну морфему; в) любое количество символов в словоформе; г) два символа.</p>	ПК-7.3.1
21.	<p>Сопоставьте критерии классификации корпусов с их описаниями:</p> <p>1. тип языковых данных А. размеченные и неразмеченные корпуса 2. язык представления текстов Б. полнотекстовые, фрагментнотекстовые корпуса 3. представление языкового материала В. одноязычные, двуязычные, многоязычные корпуса 4. способ обработки материала Г. устные, письменные, смешанные корпуса</p>	ПК-7.У.1
22.	<p>Соотнесите виды разметки с их описанием:</p> <p>1. Морфологическая разметка А. обозначение семантических категорий, к которым приписывают слово 2. Синтаксическая разметка Б. приписывание лексическим единицам корпуса грамматических характеристик 3. Семантическая разметка В. приписывание элементам текста специальных меток, обозначающих ударение и интонацию 4. Просодическая разметка Г. описание синтаксических связей между лексическими единицами</p>	ПК-7.У.1

23.	<p>Задание закрытого типа на установление последовательности Расположите в правильном порядке этапы создания корпуса: А. Разметка текстов и корректировка результатов Б. Обеспечение многоаспектного поиска В. Отбор текстовых источников Г. Представление текстов в электронном виде</p>	ПК-7.У.1
24.	<p>Вставьте пропущенное слово: _____ поиска – отношение между количеством выданных релевантных текстов к общему количеству выданных системой текстов.</p>	ПК-7.У.1
25.	<p>Выберите правильные варианты ответа: выберите лингвистические корпуса, сформированные из текстов, написанных на русском языке: 1. Тюбингенский корпус; 2. Брауновский корпус; 3. Упсальский корпус; 4. Ланкастер-Осло-Берген.</p>	ПК-7.У.1
26.	<p>Выберите правильные варианты ответа: выберите корпуса, которые относятся к специализированным корпусам: 1. Корпус Childes 2. Британский национальный корпус 3. Корпус английских текстов петербургских школьников 4. Национальный корпус русского языка</p>	ПК-7.У.1
27.	<p>Сопоставьте процедуры обработки естественного языка с их названиями:</p> <p>1. векторизация А. процесс выявления базовой формы лексической единицы 2. токенизация Б. синтаксический анализ 3. парсинг В. разделение символов естественного языка на словоформы 4. лемматизация Г. преобразование текстового документа в цифровой вектор</p>	ПК-7.У.1
28.	<p>Соотнесите названия корпусов с их описаниями: 1. Мультимедийный корпус А. Корпус, содержащий тексты с размеченными слогами 2. Акцентологический корпус Б. Корпус, содержащий кинотранскрипты фильмов 3. Основной корпус В. Корпус, в котором сохранена лексическая специфика текстов 4. Корпус диалектных текстов Г. Корпус, содержащий тексты, представляющие русский язык</p>	ПК-7.У.1
29.	<p>Соотнесите название корпуса с его описанием 1. Параллельный корпус А. корпус, содержащий тексты, относящиеся к большому промежутку времени</p>	ПК-7.У.1

	2. Устный корпус Б. корпус, содержащий структурированную совокупность речевых фрагментов 3. Динамический корпус В. корпус, содержащий тексты, относящиеся к различным предметным областям 4. Терминологический корпус Г. корпус, в котором представлены тексты-оригиналы и соответствующие им тексты-переводы	
30.	Запишите как называются слова, которые используются в определенном корпусе (или тексте) чаще, чем они обычно употребляются в языке, называются ключевыми словами	ПК-7.В.1
31.	Запишите как называется представленный в алфавитном порядке список слов (или словоформ) с указанием всех вхождений конкретного слова в заданный массив текстов.	ПК-7.В.1
32.	Запишите как называются слова, которые используются в определенном корпусе (или тексте) чаще, чем они обычно употребляются в языке, называются ключевыми словами	ПК-7.В.1

Примечание: Система оценивания тестовых заданий

1 тип) Задание комбинированного типа с выбором одного верного ответа из четырех предложенных и обоснованием выбора считается верным, если правильно указана цифра и приведены конкретные аргументы, используемые при выборе ответа. Полное совпадение с верным ответом оценивается 1 баллом, неверный ответ или его отсутствие – 0 баллов.

2 тип) Задание комбинированного типа с выбором нескольких вариантов ответа из предложенных и развернутым обоснованием выбора считается верным, если правильно указаны цифры и приведены конкретные аргументы, используемые при выборе ответов. Полное совпадение с верным ответом оценивается 1 баллом, если допущены ошибки или ответ отсутствует – 0 баллов.

3 тип) Задание закрытого типа на установление соответствия считается верным, если установлены все соответствия (позиции из одного столбца верно сопоставлены с позициями другого столбца). Полное совпадение с верным ответом оценивается 1 баллом, неверный ответ или его отсутствие – 0 баллов

4 тип) Задание закрытого типа на установление последовательности считается верным, если правильно указана вся последовательность цифр. Полное совпадение с верным ответом оценивается 1 баллом, если допущены ошибки или ответ отсутствует – 0 баллов.

5 тип) Задание открытого типа с развернутым ответом считается верным, если ответ совпадает с эталонным по содержанию и полноте. Правильный ответ за задание оценивается в 3 балла, если допущена одна ошибка \ неточность \ ответ правильный, но не полный - 1 балл, если допущено более 1 ошибки \ ответ неправильный \ ответ отсутствует – 0 баллов.

Перечень тем контрольных работ по дисциплине обучающихся заочной формы обучения, представлены в таблице 19.

Таблица 19 – Перечень контрольных работ

№ п/п	Перечень контрольных работ
	Не предусмотрено

10.4. Методические материалы, определяющие процедуры оценивания индикаторов, характеризующих этапы формирования компетенций, содержатся в локальных нормативных актах ГУАП, регламентирующих порядок и процедуру

проведения текущего контроля успеваемости и промежуточной аттестации обучающихся ГУАП.

11. Методические указания для обучающихся по освоению дисциплины

11.1. Методические указания для обучающихся по освоению лекционного материала

Основное назначение лекционного материала – логически стройное, системное, глубокое и ясное изложение учебного материала. Назначение современной лекции в рамках дисциплины не в том, чтобы получить всю информацию по теме, а в освоении фундаментальных проблем дисциплины, методов научного познания, новейших достижений научной мысли. В учебном процессе лекция выполняет методологическую, организационную и информационную функции. Лекция раскрывает понятийный аппарат конкретной области знания, её проблемы, дает цельное представление о дисциплине, показывает взаимосвязь с другими дисциплинами.

Планируемые результаты при освоении обучающимися лекционного материала:

- получение современных, целостных, взаимосвязанных знаний, уровень которых определяется целевой установкой к каждой конкретной теме;
- получение опыта творческой работы совместно с преподавателем;
- развитие профессионально-деловых качеств, любви к предмету и самостоятельного творческого мышления.
- появление необходимого интереса, необходимого для самостоятельной работы;
- получение знаний о современном уровне развития науки и техники и о прогнозе их развития на ближайшие годы;
- научиться методически обрабатывать материал (выделять главные мысли и положения, приходить к конкретным выводам, повторять их в различных формулировках);
- получение точного понимания всех необходимых терминов и понятий.

Лекционный материал может сопровождаться демонстрацией слайдов и использованием раздаточного материала при проведении коротких дискуссий об особенностях применения отдельных тематик по дисциплине.

Структура предоставления лекционного материала:

- Постановка проблемы;
- Обзор предметной области;
- Обзор литературы;
- Разработка проекта решения;
- Реализация и анализ результатов;
- Дальнейшие направления исследований.

Корпусная лингвистика [Текст]: учебно-методическое пособие / С.-Петербург. гос. ун-т аэрокосм. приборостроения; сост. Е. Ю. Дубинина. - СПб.: Изд-во ГУАП, 2020. - 90 с. Гастелло (50).

11.2. Методические указания для обучающихся по прохождению самостоятельной работы

В ходе выполнения самостоятельной работы, обучающийся выполняет работу по заданию и при методическом руководстве преподавателя, но без его непосредственного участия.

Для обучающихся по заочной форме обучения, самостоятельная работа может включать в себя контрольную работу.

В процессе выполнения самостоятельной работы, у обучающегося формируется целесообразное планирование рабочего времени, которое позволяет им развивать умения и навыки в усвоении и систематизации приобретаемых знаний, обеспечивает высокий

уровень успеваемости в период обучения, помогает получить навыки повышения профессионального уровня.

Методическими материалами, направляющими самостоятельную работу обучающихся являются:

- учебно-методический материал по дисциплине.

Записи имеют первостепенное значение для самостоятельной работы студентов. Они помогают понять построение изучаемого материала, выделить основные положения, проследить их логику и тем самым проникнуть в творческую лабораторию автора.

Ведение записей способствует превращению чтения в активный процесс, мобилизует, наряду со зрительной, и моторную память. Следует помнить: у студента, систематически ведущего записи, создается свой индивидуальный фонд подсобных материалов для быстрого повторения прочитанного, для мобилизации накопленных знаний. Особенно важны и полезны записи тогда, когда в них находят отражение мысли, возникшие при самостоятельной работе.

Корпусная лингвистика [Текст]: учебно-методическое пособие / С.-Петерб. гос. ун-т аэрокосм. приборостроения; сост. Е. Ю. Дубинина. - СПб.: Изд-во ГУАП, 2020. - 90 с. Гастелло (50).

11.3. Методические указания для обучающихся по прохождению текущего контроля успеваемости.

Текущий контроль успеваемости предусматривает контроль качества знаний обучающихся, осуществляемого в течение семестра с целью оценивания хода освоения дисциплины.

Главное условие успешности в освоении учебной дисциплины - систематические занятия. Для полного понимания изучаемого материала следует задавать вопросы непосредственно на практических занятиях, чтобы не оставлять пробелов в изучении. За дополнительными разъяснениями и рекомендациями студент может обращаться к преподавателю во время консультаций. Систематическая работа в семестре, активное участие в семинарах (доклады, сообщения, самостоятельно подготовленные презентации по интересующим темам по теории и практике перевода) и выполнение практических заданий обеспечит высокую оценку при прохождении промежуточной аттестации.

11.4. Методические указания для обучающихся по прохождению промежуточной аттестации.

Промежуточная аттестация обучающихся предусматривает оценивание промежуточных и окончательных результатов обучения по дисциплине. Она включает в себя:

- зачет – это форма оценки знаний, полученных обучающимся в ходе изучения учебной дисциплины в целом или промежуточная (по окончании семестра) оценка знаний обучающимся по отдельным разделам дисциплины с аттестационной оценкой «зачтено» или «не зачтено».

При подготовке к аттестации следует не только разобраться в материале, но попробовать, не подглядывая в конспекты или учебники, изложить письменно наиболее существенные понятия, утверждения, точки зрения по каждому разделу программы, составить план-конспекты ответов на вопросы.

На зачете в процессе подготовки к ответу прежде, чем приступить к подробному изложению ответа на вопрос, следует составить (письменно или устно) план предстоящего ответа, обязательно привести примеры, указать авторов, на которых ссылаются при ответе.

Промежуточная аттестация проводится не только в традиционном формате «вопрос-ответ», но и в форме дискуссии, в процессе которой определяется умение студента быстро мыслить, формулировать свой ответ при линейном развитии речи, владение устной и письменной версией официально-деловой нормы современного русского языка.

Лист внесения изменений в рабочую программу дисциплины

Дата внесения изменений и дополнений. Подпись внесшего изменения	Содержание изменений и дополнений	Дата и № протокола заседания кафедры	Подпись зав. кафедрой