

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Наумова Наталия Александровна
Должность: Ректор
Дата подписания: 21.07.2025 17:13:13
Уникальный программный ключ:
6b5279da4e034bffa79172803da5b7b5559fc69e2

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования

«ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ»

(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ)

Лингвистический факультет

Кафедра переводоведения и когнитивной лингвистики

Согласовано

деканом лингвистического факультета

«20» июля 2025 г.

/Вековищева С.Н./

Рабочая программа дисциплины

Корпусная лингвистика

Специальность

45.05.01 Перевод и переводоведение

Специализация:

Перевод и переводоведение (английский язык+китайский язык)

Квалификация

Лингвист-переводчик

Форма обучения

Очная

Согласовано учебно-методической комиссией
лингвистического факультета

Протокол «20» июля 2025 г. № 4

Председатель УМКом /Горбачева О.А./

Рекомендовано кафедрой переводоведения
и когнитивной лингвистики

Протокол от «20» июля 2025 г. № 5

Зав. кафедрой /Жирова И.Г./

Москва

2025

Автор-составитель:

Улиткин И.А., доцент, кандидат филологических наук

Рабочая программа дисциплины «Корпусная лингвистика» составлена в соответствии с требованиями Федерального Государственного образовательного стандарта высшего образования по специальности 45.05.01 Лингвистика, утвержденного приказом МИНОБРНАУКИ России от 12.08.2020 г. № 989.

Дисциплина относится к обязательной части, Блока 1 «Дисциплины (модули)», и является обязательной для изучения.

Год начала подготовки (по учебному плану) 2025

СОДЕРЖАНИЕ

1. Планируемые результаты обучения.....	4
2. Место дисциплины в структуре образовательной программы.....	4
3. Объем и содержание дисциплины.....	5
4. Учебно-методическое обеспечение самостоятельной работы обучающихся	7
5. Фонд оценочных средств для проведения текущей и промежуточной аттестации по дисциплине.....	8
6. Учебно-методическое и ресурсное обеспечение дисциплины.....	16
7. Методические указания по освоению дисциплины.....	17
8. Информационные технологии для осуществления образовательного процесса по дисциплине	17
9. Материально-техническое обеспечение дисциплины.....	17

1. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

1.1. Цель и задачи дисциплины

Целями освоения дисциплины «Корпусная лингвистика» являются ознакомление обучающихся с основными понятиями и возможностями корпусной лингвистики, освоение на практике технологически несложных методов корпусной лингвистики, формирование практических навыков работы с наиболее известными мировыми лингвистическими ресурсами, использование лингвистических корпусов для решения конкретных переводческих и исследовательских задач, а также приобретение опыта работы с корпусными средствами для построения своего собственного переводческого и исследовательского корпусов.

Задачи дисциплины:

1. представить возникновение и развитие корпусной лингвистики как науки;
2. сформировать представление о предмете корпусной лингвистики, о категориях ее понятийного аппарата;
3. ознакомить студентов с существующими корпусами русского языка, их возможностями;
4. выработать методику анализа параметров корпуса.

1.2. Планируемые результаты дисциплины

В результате освоения данной дисциплины у обучающихся формируется следующая компетенция:

ОПК-1. Способен применять знания иностранных языков и знания о закономерностях функционирования языков перевода, а также использовать систему лингвистических знаний при осуществлении профессиональной деятельности

ОПК-2. Способен применять систему знаний о видах, приемах, стратегиях, технологиях и закономерностях перевода, а также требованиях, предъявляемых к переводу

СПК-1. Способен использовать понятийный аппарат теоретической и прикладной лингвистики, переводоведения, теории межкультурной коммуникации для решения профессиональных задач

СПК-5. Способен применять систему лингвистических знаний об основных фонетических, лексических, грамматических, словообразовательных явлениях о закономерностях функционирования языков мира

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина относится к обязательной части, Блока 1 «Дисциплины (модули)», и является обязательной для изучения.

Учебная дисциплина является частью читаемых кафедрой переводоведения и когнитивной лингвистики дисциплин в области формирования профессиональных компетенций переводчика и преподавателя перевода.

Данная дисциплина ориентирована на формирование гуманистического мировоззрения студентов, расширения их общекультурной, языковой и практической профессиональной подготовки, а также совершенствования навыков работы с корпусами текстов для решения профессиональных переводческих задач.

В методическом плане дисциплина является теоретическим курсом, получаемым студентами на третьем курсе и погружающих обучающихся в специфику профессиональной деятельности наравне с учебными дисциплинами *«Практический курс первого иностранного языка (английский язык)»* и *«Теория перевода»*, *«Информационные технологии в профессиональной деятельности»*. Изучая дисциплину «Корпусные

исследования в переводе», студенты приобретают знания, необходимые для таких дисциплин, как «Синхронный перевод», «Письменный перевод», «Устный последовательный перевод», «Технический перевод», «Судебный перевод», «Основы научного и литературного редактирования».

Требования к «входным» знаниям, умениям и готовностям обучающегося, необходимые при освоении данной дисциплины и приобретенные в результате освоения предшествующих дисциплин. Обучающийся, освоивший дисциплину (модуль) должен:

знать:

- ключевые понятия, актуальные проблемы, перспективные направления развития корпусной лингвистики;
- основные типы корпусов, наиболее известные и широко применяемые иноязычные и русскоязычные корпусы;
- способы использования корпусов для различных целей;
- технологию создания корпусов, основные форматы и способы организации корпусов;

уметь:

- определять тип и способы применения незнакомых корпусов;
- осваивать новые компьютерные инструменты;
- интерпретировать данные, полученные из корпусов;
- сравнивать и оценивать функциональность корпусов и инструментов по результатам работы с ними;

владеть навыками

- работы с несколькими существующими иноязычными и русскоязычными корпусами и корпусными менеджерами;
- формулирования сложных лингвистических запросов с использованием специальных обозначений (тэгов) и регулярных выражений, разметки текстов и чтения разметки;
- проектирования собственного корпуса;
- создания небольших специализированных корпусов.

3. ОБЪЁМ И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ.

3.1. Объём дисциплины

Показатель объёма дисциплины	Форма обучения
	Очная
Объём дисциплины в зачетных единицах	2
Объём дисциплины в часах	180
Контактная работа	90,4
Практические занятия	90
Из них в форме практической подготовки	50
Контактные часы на промежуточную аттестацию	0.2
Зачет	0.2
Самостоятельная работа	74
Контроль	15,6

Форма промежуточной аттестации – зачет в 7 семестре, зачет с оценкой в 8 семестре

3.2. Содержание дисциплины

Наименование разделов (тем) дисциплины с кратким содержанием	Количество часов	
	Практические занятия	
	Общее количество	Из них в форме практической подготовки
Тема 1. Введение в корпусную лингвистику Определение корпуса, его свойства (репрезентативность, аннотирование). Типология корпусов (общие, параллельные, сопоставимые). Роль корпусных данных в переводе: выявление узуса, коллокаций и переводческих соответствий.	18	
Тема 2. Практика по работе в наиболее известных корпусах. Формирование навыков поиска в электронных корпусах (английских и китайских), создание конкордансов и частотных списков. Особенности лингвистической разметки для китайского языка.	18	10
Тема 3. Качественные и количественные данные в корпусе Различие между качественным анализом контекста и количественными метриками (частотность, статистическая значимость). Интерпретация числовых показателей для выявления переводческих закономерностей.	18	10
Тема 4. Принципы и способы составления корпусов Этапы проектирования: целеполагание, отбор источников, балансировка по жанрам. Форматы хранения и проблема разметки (сегментация слов, тегирование). Оценка репрезентативности корпуса.	18	15
Тема 5. Корпусные методы в применении к переводу Использование параллельных и сопоставимых корпусов для подбора эквивалентов, выявления переводческих универсалий и оценки качества перевода. Применение корпусных данных при переводе с английского на китайский и обратно.	18	15
Итого:	90	50

Практическая подготовка

Тема	Задания на практическую подготовку	Часы
Тема 2. Практика по работе в наиболее известных	Составить частотный список 20 наиболее употребительных слов в выбранном жанре	10

корпусах.	(например, новости) для каждого языка с помощью инструментов корпуса.	
Тема 3. Качественные и количественные данные в корпусе	Используя встроенные статистические инструменты корпуса (MI-score, Log-likelihood), определить, какая из двух коллокаций является более устойчивой и значимой для перевода.	10
Тема 4. Принципы и способы составления корпусов	Оценить репрезентативность предложенного учебного корпуса (набора текстов) по критерию жанрового и временного баланса, указать возможные перекосы.	15
Тема 5. Корпусные методы в применении к переводу	С помощью сопоставимого корпуса (оригинальные тексты на целевом языке vs. переводные тексты) выявить один случай переводческой универсалии (эксплицитация, упрощение, нормализация).	15
		50

4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Темы для самостоятельного обучения	Изучаемые вопросы	Кол-во часов	Формы самостоятельной работы	Методическое обеспечение	Форма отчётности
Введение в корпусную лингвистику	1. Теоретическая значимость корпусных данных. 2. Практическое использование корпусных данных.	15	проработка конспектов лекций, изучение теоретического материала, подготовка практического задания	Основная и дополнительная литература, интернет-источники	практическое задание
Практика по работе в наиболее известных корпусах.	1. Использование интерфейса и языка запросов, использование экспортированных данных. British National Corpus, НКРЯ, Intercorp, синтаксические корпуса. 2. Наиболее известные корпуса русского языка	15	проработка конспектов лекций, изучение теоретического материала, подготовка практического задания подготовка проекта	Основная и дополнительная литература, интернет-источники	практическое задание, проект
Качественные и количественные	1. Получение корпусных примеров: как и зачем. 2. Корпусное изучение вариативности.		проработка конспектов лекций, изучение	Основная и дополнительная литература,	практическое задание

данные в корпусе	3. Проблема взаимодействия факторов. 4. Основы статистики.	15	теоретического материала, подготовка практического задания	интернет-источники	
Принципы и способы составления корпусов	1. Принципы составления корпусов (репрезентативность, сбалансированность, металингвистическая разметка, типы корпусов). 2. Корпусные средства. 3. Лингвистическое аннотирование	15	проработка конспектов лекций, изучение теоретического материала, подготовка практического задания	Основная и дополнительная литература, интернет-источники	практическое задание
Корпусные методы в применении и к переводу	1. Корпусная фонетика и фонология. 2. Корпусная лексикология и лексикография. 3. Корпусный морфосинтаксис. 4. Использование корпуса для литературоведения	14	проработка конспектов лекций, изучение теоретического материала, подготовка практического задания	Основная и дополнительная литература, интернет-источники	практическое задание
Итого		74			

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

5.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы:

<i>Код компетенции</i> <i>Наименование компетенции</i>	<i>Этапы формирования компетенции</i>
ОПК-1. Способен применять знания иностранных языков и знания о закономерностях функционирования языков перевода, а также использовать систему лингвистических знаний при осуществлении профессиональной деятельности	1. Работа на учебных занятиях 2. Самостоятельная работа.
ОПК-2. Способен применять систему знаний о видах, приемах, стратегиях, технологиях и закономерностях перевода, а также требованиях, предъявляемых к переводу	1. Работа на учебных занятиях 2. Самостоятельная работа.
СПК-1. Способен использовать понятийный аппарат теоретической и прикладной лингвистики, переводоведения, теории межкультурной коммуникации для решения профессиональных задач	1. Работа на учебных занятиях 2. Самостоятельная работа.

СПК-5. Способен применять систему лингвистических знаний об основных фонетических, лексических, грамматических, словообразовательных явлениях о закономерностях функционирования языков мира	1. Работа на учебных занятиях 2. Самостоятельная работа.
--	---

5.2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

<i>Оцениваемые компетенции</i>	<i>Уровень сформированности</i>	<i>Этап формирования</i>	<i>Описание показателей</i>	<i>Критерии оценивания</i>	<i>Шкала оценивания</i>
ОПК-1	Пороговый	1. Работа на учебных занятиях 2. Самостоятельная работа.	Знать: основы осуществления письменного и устного перевода с использованием корпусов в соответствии с коммуникативной ситуацией Уметь: осуществлять письменный и устный перевод текстов различной тематической направленности на иностранном (иностраннных) языках с использованием корпусного подхода в соответствии с коммуникативной ситуацией	Устный опрос, практическое задание, тест	Шкала оценивания устного опроса Шкала оценивания практического задания Шкала оценивания теста
	Продвинутый	1. Работа на учебных занятиях 2. Самостоятельная работа.	Знать: основы применения стратегий и тактик осуществления письменного и устного перевода с использованием корпусов в соответствии с коммуникативной ситуацией Уметь:	Устный опрос, проект задание по практической подготовке	Шкала оценивания устного опроса Шкала оценивания проекта Шкала оценивания практической подготовки

			осуществлять письменный и устный перевод текстов различной тематической направленности на иностранном (иностраннных) языках с использованием корпусного подхода в соответствии с коммуникативной ситуацией Владеть: навыками осуществления письменного и устного перевода текстов различной тематической направленности на иностранном (иностраннных) языках с использованием корпусного подхода в соответствии с коммуникативной ситуацией		
ОПК-2	Пороговые	1. Работа на учебных занятиях 2. Самостоятельная работа.	Знать: основные современные методы научного исследования, в том числе и в смежных областях, в самостоятельных исследованиях с учетом наработок корпусной лингвистики Уметь: использовать на базовом уровне основные современные методы научного исследования, в том числе и в	Устный опрос, практическое задание, тест	Шкала оценивания устного опроса Шкала оценивания практического задания Шкала оценивания теста

			смежных областях, в самостоятельных исследованиях с учетом наработок корпусной лингвистики		
	Продвину- тый	1. Работа на учебных занятиях 2. Самостоятельная работа.	Знать: возможности использования основных современных методов научного исследования, в том числе и в смежных областях, в самостоятельных исследованиях с позиций лингвопрагматики Уметь: использовать на продвинутом уровне основных современных методов научного исследования, в том числе и в смежных областях, в самостоятельных исследованиях с учетом наработок корпусной лингвистики Владеть: техниками использования основных современных методов научного исследования, в том числе и в смежных областях, в самостоятельных исследованиях с учетом наработок корпусной лингвистики	Устный опрос, проект задание по практической подготовке	Шкала оценивания устного опроса Шкала оценивания проекта Шкала оценивания практической подготовки
СПК-1	Порогов- ый	1. Работа на учебных занятиях 2.	Знать: типы электронных словарей и сферы их использования;	Устный опрос, практическое задание, тест	Шкала оценивания устного опроса

		Самостоятельная работа.	преимущества и способы наиболее эффективного использования электронных словарей и различных прикладных обучающих и тестирующих программ, интернет ресурсов для лингвистов Уметь: работать с электронными ресурсами для решения лингвистических задач; оценивать программное обеспечение (обучающие и тестирующие программы, электронные словари, машинные переводчики) и перспективы его использования с учетом решаемых профессиональных задач		Шкала оценивания практического задания Шкала оценивания теста
	Продвину- тый	1. Работа на учебных занятиях 2. Самостоятельная работа.	Знать: типы электронных словарей и сферы их использования; преимущества и способы наиболее эффективного использования электронных словарей и различных прикладных обучающих и тестирующих программ, интернет ресурсов для лингвистов Уметь: работать с электронными	Устный опрос, проект задание по практической подготовке	Шкала оценивания устного опроса Шкала оценивания проекта Шкала оценивания практической подготовки

			ресурсами для решения лингвистических задач; оценивать программное обеспечение (обучающие и тестирующие программы, электронные словари, машинные переводчики) и перспективы его использования с учетом решаемых профессиональных задач Владеть: навыками извлечения необходимой информации из электронных словарей и электронных ресурсов; навыками создания и периодического обновления тематических глоссариев при помощи электронных ресурсов для решения лингвистических задач		
СПК-5	Пороговый	1. Работа на учебных занятиях 2. Самостоятельная работа.	Знать: способы технической и психологической подготовки к различным видам перевода при помощи гипертекстовой системы, каталогов ресурсов, поисковых машин. Уметь: искать, отбирать и обрабатывать информацию в	Устный опрос, практическое задание, тест	Шкала оценивания устного опроса Шкала оценивания практического задания Шкала оценивания теста

			справочной литературе и компьютерных сетях для выполнения перевода; создавать переводы текстов различной жанрово-стилистической принадлежности при помощи поиска информации в справочной литературе и компьютерных сетях		
	Продвинутый	1. Работа на учебных занятиях 2. Самостоятельная работа.	Знать: способы технической и психологической подготовки к различным видам перевода при помощи гипертекстовой системы, каталогов ресурсов, поисковых машин. Уметь: искать, отбирать и обрабатывать информацию в справочной литературе и компьютерных сетях для выполнения перевода; создавать переводы текстов различной жанрово-стилистической принадлежности при помощи поиска информации в справочной литературе и компьютерных сетях Владеть: навыками корректного и	Устный опрос, проект задание по практической подготовке	Шкала оценивания устного опроса Шкала оценивания проекта Шкала оценивания практической подготовки

			эффективного использования информационно-справочных ресурсов при выполнении переводов; навыками оценивать результаты поиска информации и собственного предпереводческого о анализа и корректировать их для достижения наибольшей		
--	--	--	--	--	--

Шкала оценивания устного опроса

Критерии оценивания	Баллы
Студент раскрывает вопрос полностью, без пробелов. Приводит точные определения корпусных понятий, перечисляет типы корпусов, объясняет разницу между качественными и количественными данными.	18–20
Студент раскрывает вопрос в целом верно, но допускает небольшие пропуски (например, не называет все типы корпусов или не упоминает одну из статистических метрик).	14–17
Раскрыто не менее 50% материала. Студент даёт лишь базовые определения, не углубляясь в детали (например, знает, что такое конкорданс, но не объясняет, как его строить)	10–13
менее 50% вопроса раскрыто, либо ответ не по теме.	0–9

Шкала оценивания проекта

Максимальный балл за выполнение проекта – 10.

– цель и задачи чёткие, актуальные для перевода с англ./кит. – цель и задачи в целом верны, но неточности – цель и задачи поверхностны или неполны – цель и задачи отсутствуют	6
– источники, объём, разметка обоснованы полно и логично – описание есть, но без деталей по сбалансированности – описание скудное, ошибки в выборе источников – описание отсутствует или не соответствует теме	2
– правильно применены конкордансы, частоты, статистика для обоих языков – методы применены в основном верно, но с 1–2 ошибками	2

– методы применены, но с грубыми ошибками или только для одного языка – методы не применены или полностью ошибочны	
– анализ отсутствует или ошибочен – рекомендации отсутствуют – хаотичное оформление, невозможно читать – защита отсутствует или не по теме	0

Шкала оценивания теста

Критерии оценивания	Баллы
Теоретические вопросы	40 – все ответы верные, развёрнутые определения точны. 30–39 – 1–2 ошибки или неполные формулировки. 20–29 – 3–5 ошибок, определения приблизительные. <20 – более 5 ошибок или грубые искажения.
Терминология	24 – все термины (корпус, конкорданс) названы и определены верно. 18–23 – 1–2 термина перепутаны или нет перевода на иностранный язык. 12–17 – 3–4 термина с ошибками. <12 – более 4 ошибок.
Анализы корпусных данных	20 – правильно интерпретирован конкорданс, рассчитана частота, верно истолкована статистическая мера (MI/LL). 15–19 – допущена 1 неточность в интерпретации. 10–14 – 2 ошибки, анализ поверхностный. <10 – непонимание результатов.
Применение к переводу	16 – для каждой задачи выбран корректный переводческий эквивалент на основе корпусных данных, аргументировано качество перевода. 12–15 – 1 задание выполнено неверно или без ссылки на корпус. 8–11 – 2–3 задания с ошибками. <8 – более 3 ошибок или задания не выполнены.

Шкала оценивания выполнения практического задания

Критерии оценивания	Баллы
содержание вопроса раскрыто полностью: информация изложена логично с акцентом на ключевых особенностях литературного явления, предлагаются выводы/рекомендации по его оценке; терминология курса используется адекватно, нет подмены понятий и терминов; для практического анализа в качестве иллюстративных примеров используются адекватные примеры.	5 баллов
содержание вопроса раскрыто полностью: информация изложена логично с акцентом на ключевых особенностях описываемого явления, отсутствуют объективные выводы/рекомендации по его оценке; есть неточности в использовании терминологии курса; для практического анализа в качестве иллюстративных примеров используются адекватные примеры.	4 балла
содержание вопроса раскрыто: информация изложена логично, однако не выделены ключевые положения, выводы/рекомендации носят субъективный характер; терминология курса используется в целом точно, представлены иллюстративные примеры адекватные рассматриваемым явлениям.	3 балла
содержание вопроса раскрыто: информация изложена схематично, не выделены ключевые положения, выводы/рекомендации носят субъективный характер; терминология курса используется в целом точно, но не представлены иллюстративные примеры адекватные рассматриваемым явлениям	2 балла
содержание вопроса раскрыто неполностью: изложение материала не структурировано, отсутствуют выводы/рекомендации; представленные примеры недостаточно иллюстративны; не приводятся иллюстративные примеры адекватные рассматриваемым явлениям	1 балл
ответ неверный или нет ответа на вопрос задания/ не соответствует теме/ не представляет авторского вклада (отсутствуют иллюстративные примеры/выводы/рекомендации).	0 баллов

Шкала оценивания практической подготовки

Полнота выполнения задания	– задание выполнено полностью, все пункты соблюдены 5 – выполнено не менее 75% задания 4 – выполнено 50–74% задания 2–3 – выполнено менее 50% 1 – задание не выполнено 0
Корректность использования корпусных инструментов	– все инструменты применены верно, результаты воспроизводимы 5 – инструменты применены в основном верно, но 1–2 ошибки 4 – ошибки в применении 3 и более инструментов 2–3

	– инструменты применены неверно или не применены 0–1
Анализ и интерпретация результатов (качество выводов)	– выводы точны, опираются на количественные и качественные данные, связаны с переводом (английский/китайский) 5 – выводы верны, но неполны или слабо связаны с переводом 4 – выводы поверхностны, содержат ошибки 2–3 – анализ отсутствует или полностью ошибочен 0–1
Оформление и воспроизводимость	– оформление чёткое, все шаги документированы, можно повторить 5 – оформление в целом понятно, но есть пропуски 4 – оформление небрежное, часть данных отсутствует 2–3 – оформление отсутствует 0–1
Итого: 18–20 отлично 14–17 хорошо 10–13 удовлетворительно 0–9 неудовлетворительно	

5.3. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Примерные практические задания

Работа с известными корпусами

Задание 1.1. В корпусе COCA (английский) или CCL (китайский) найти по 10 конкордансов для заданного слова. Определить 3 типичных коллоката (с опорой на MI-score или частоту).

Задание 1.2. Составить частотный список 15 наиболее употребительных существительных в подкорпусе «новости» для английского и в подкорпусе «газетные тексты» для китайского. Сравнить результаты.

Задание 1.3. В параллельном корпусе (например, MultiPar или Sino-European) найти 5 переводческих соответствий для английской идиомы и для китайского чэньюй (成语). Оценить вариативность перевода.

Задание 1.4. С помощью функции KWIC проанализировать окружение многозначного английского глагола (например, get) и китайского иероглифа (например, 打). Определить основные значения по контексту.

Качественный и количественный анализ

Задание 2.1. Для двух английских синонимов (например, begin – start) и двух близких по значению китайских слов (например, 开始 – 启动) рассчитать частоту в корпусе. Сделать вывод о частотном предпочтении.

Задание 2.2. Используя встроенную статистику корпуса, сравнить силу коллокации (MI-score) для двух словосочетаний. Определить, какая коллокация более устойчива.

Задание 2.3. Выбрать 5 контекстов из параллельного корпуса, где исходное слово переведено по-разному. Качественно объяснить причины различий, затем подтвердить или опровергнуть свои гипотезы количественными данными.

Составление и разметка корпусов

Задание 3.1. Собрать 10 небольших текстов на английском и 10 на китайском по одной тематике (например, инструкции к бытовой технике). Привести к формату plain text, очистить от разметки.

Задание 3.2. Для собранных китайских текстов выполнить сегментацию слов с помощью Jieba или аналогичного инструмента. Для английских текстов выполнить POS-разметку в Treetagger. Оценить качество автоматической разметки (найти 3 ошибки).

Задание 3.3. На основе отобранных текстов составить небольшой параллельный корпус (5–7 предложений, выровненных вручную или с помощью инструмента выравнивания). Сохранить в формате TMX или plain text с разметкой.

Корпусные методы в переводе

Задание 4.1. На материале параллельного корпуса (английский–китайский) найти не менее 5 способов перевода английского пассива на китайский. Сделать вывод о доминирующей стратегии (использование маркеров 被, 让, 由 или активной конструкции).

Задание 4.2. С помощью сопоставимого корпуса (оригинальные китайские тексты vs. переводные с английского) выявить один случай переводческой универсалии (эксплицитация, упрощение, нормализация). Привести примеры.

Задание 4.3. Оценить качество предложенного перевода (3–5 предложений). Проверить естественность коллокаций в корпусе целевого языка. Найти не менее 3 сомнительных случаев и предложить корпусно-обоснованные замены.

Задание 4.4. Создать мини-параллельный корпус из 10 предложений (оригинал + ваш собственный перевод). С помощью конкордансера (AntConc) найти в нём повторяющиеся переводческие приёмы. Представить отчёт с выводами для своей будущей переводческой практики.

Задания по практической подготовке

1. Составить частотный список 20 наиболее употребительных слов в выбранном жанре (например, новости) для каждого языка с помощью инструментов корпуса.
2. Используя встроенные статистические инструменты корпуса (MI-score, Log-likelihood), определить, какая из двух коллокаций является более устойчивой и значимой для перевода.
3. Оценить репрезентативность предложенного учебного корпуса (набора текстов) по критерию жанрового и временного баланса, указать возможные перекосы.
4. С помощью сопоставимого корпуса (оригинальные тексты на целевом языке vs. переводные тексты) выявить один случай переводческой универсалии (эксплицитация, упрощение, нормализация)

Примерные вопросы для обсуждения на практических занятиях

1. Понятие корпусной лингвистики. Роль и прикладное назначение корпусных исследований на современном этапе развития языкознания. Отечественные и зарубежные исследователи, ведущие разработки в русле корпусной лингвистики. Значение корпусной лингвистики для современного языкознания. Специфика корпусных баз данных как лексикографического источника. Сбалансированные и несбалансированные корпуса.

2. Прикладные задачи переводоведения и лингводидактики через призму корпусных исследований. Развитие и ключевые современные направления корпусных исследований. Перспективы разработок в области корпусных баз данных. Преимущества и недостатки корпусных баз данных.

3. Первое и второе поколение корпусов. Типы корпусов: устные письменные, одноязычные и многоязычные, аннотированные и неаннотированные. Параллельные корпусы. Существующие корпусы: Brown Corpus, British National Corpus, Lancaster Corpus, Uppsala Corpus, COCA, COHA, Learners' Corpus, Scottish Corpus, НКРЯ. Проекты Linguae и Reverso Context.

4. Корпусоподобные интерфейсы. Принципы внутренней организации корпусов. Интерфейс корпусной базы данных. Разметка. Виды разметок: частеречная разметка, грамматический парсинг, просодическая аннотация, анафорическая аннотация, семантическая разметка и под. Лемма, лемматизация. Корпусный менеджер. Конкордансер. Программы для работы с корпусными базами данных.

5. Алгоритмы решения типовых задач в области теории перевода и методики обучения иностранным языкам. Коллострукции. Коллострукционный анализ А. Штефановича. Взаимная информация (MI). Частотность. Поиск по регистру. Методика диахронического изучения языковых трендов. Инструменты межкорпусного сопоставления.

6. Раздел корпусной базы данных KWIC, метод лексико-семантических полей, поиск синонимов и антонимов в корпусных базах данных. Параметры ранжирования результатов в корпусных базах данных.

Примерные тестовые задания

Раздел 1. Теоретические вопросы (множественный выбор)

1. Что такое корпус текстов в определении корпусной лингвистики?

- А) Любая коллекция текстов на жёстком диске
- Б) Структурированное, репрезентативное, машиночитаемое собрание текстов
- В) База данных грамматических правил
- Г) Библиотека художественной литературы

2. Какой тип корпуса наиболее полезен для поиска переводческих эквивалентов?

- А) Одноязычный частотный
- Б) Параллельный (с выравненными текстами перевода)
- В) Диахронический
- Г) Акустический

3. Какой параметр характеризует силу связи между словами в коллокации?

- А) TF-IDF
- Б) MI-score (взаимная информация)
- В) F1-score
- Г) Перплексия

4. Что такое конкорданс?

- А) Список всех слов корпуса с частотой
- Б) Список контекстных употреблений искомого слова с окружением
- В) Алгоритм автоматического перевода
- Г) Статистический критерий для сравнения корпусов

Раздел 2. Верно / Неверно

1. Репрезентативность корпуса означает, что его результаты можно обобщить на весь язык или жанр.

(Верно / Неверно)

2. POS-тегирование китайского языка не требует предварительной сегментации слов, так как иероглифы уже являются отдельными токенами.

(Верно / Неверно)

3. Параллельный корпус содержит оригинальные тексты на одном языке и их переводы на другой, выровненные на уровне предложений или абзацев.

(Верно / Неверно)

4. Значение MI-score, близкое к 0, указывает на сильную коллокационную связь между словами.

(Верно / Неверно)

Раздел 3. Сопоставление терминов и определений

Соедините термин с его определением (укажите соответствия в виде буква–цифра):

Термин	Определение
А. Лемма	1. Автоматическая грамматическая разметка слова (часть речи, число, время)
В. Токен	2. Словарная форма слова, объединяющая словоформы (например, go для goes, went, gone)
С. Таггинг	3. Количество словоупотреблений в корпусе (каждое вхождение слова)
Д. Коллокация	4. Устойчивое сочетание слов, встречающееся чаще случайного

Ответ: А–2, В–3, С–1, D–4.

Раздел 4. Короткий ответ (терминология на русском, английском и китайском)

1. Напишите русский, английский и китайский эквивалент термина:

«параллельный корпус»

(Ответ: параллельный корпус – parallel corpus – 平行语料库)

2. Какая статистическая мера используется в корпусной лингвистике для определения значимости коллокации, кроме MI-score? Назовите одну.

(Ответ: Log-likelihood, T-score, Dice coefficient – любой один)

3. Как называется файл обмена переводческими памятьями (формат XML для параллельных корпусов)?

(Ответ: TMX)

Раздел 5. Анализ корпусных данных (интерпретация конкорданса)

Задание. Дан фрагмент конкорданса из китайского корпуса ВСС для иероглифа 打:

他打了一辆车回家。

小孩在打游戏。

打毛衣需要耐心。

她打了人，很后悔。
电话打不通。

Вопросы:

Сколько различных значений иероглифа 打 представлено в примерах? Перечислите их.

Какое значение встречается чаще всего?

Предложите английский перевод для каждого китайского предложения.

(Пример правильного ответа: четыре значения: «взять (такси)», «играть (в игру)», «вязать», «ударить»; чаще – «играть»; переводы соответственно.)

Раздел 6. Применение к переводу (выбор эквивалента на основе корпусных данных)

1. В параллельном корпусе английский–китайский для английского словосочетания heavy rain чаще всего встречается перевод:

- А) 大雨
- Б) 强降雨
- В) 暴雨
- Г) 重雨

(Правильный ответ: А – 大雨)

2. Анализ частотности в корпусе СОСА показывает, что для перевода русского «однако» в академическом тексте наиболее уместным английским эквивалентом является:

- А) however
- Б) but
- В) nevertheless
- Г) though

(Ответ зависит от задания, но предполагается however как самый частотный вариант в формальном стиле. В тесте должен быть чёткий ключ, например, А.)

Раздел 7. Открытое задание на расчёт / интерпретацию

1. В корпусе объёмом 10 млн слов слово happy встретилось 500 раз, а слово joyful – 50 раз. Во сколько раз happy частотнее? Какое слово более употребительно в письменной речи?

(Ответ: в 10 раз, happy более употребительно)

2. Для двух коллокаций MI-score составил: strong coffee – 8.5, powerful coffee – 1.2. Какая коллокация является естественной? Почему?

(Ответ: strong coffee, так как MI-score значительно выше нуля, указывая на сильную неслучайную связь.)

В рамках курса для текущего контроля используется формат проектных заданий.

Примерная тематика проектов:

1. Сравните пару синонимов *кинуть* и *бросить*: проверьте влияние на их распределение не менее чем двух факторов. Приведите количественные данные и запросы, по которым они были получены, примените статистические критерии для проверки значимости различий.
2. Постройте частотный словарь словоформ для заданного текста.
3. Выполните лемматизацию и автоматическую частеречную разметку предложенного текста.
4. Создайте собственный корпус по заданной тематике.

Примерные вопросы к зачету / зачету с оценкой

1. Предмет корпусной лингвистики
2. Полевая лингвистика и корпусная лингвистика
3. Связь корпусной лингвистики с этнолингвистикой
4. История лингвистических корпусов: от картотеки к корпусу.
5. Классификация (типология) корпусов.
6. Корпусная лингвистика: современное состояние.
7. Корпусная лингвистика в России.
8. Обзор существующих корпусов различных типов.
9. Корпус как поисковая система.
10. Корпусоподобные интерфейсы между лингвистом и поисковыми системами Интернета.
11. Лингвистические исследования, базирующиеся на корпусах.
12. Проблемы репрезентативности корпусов.
13. Проблемы хронологии в общезыковых корпусах.
14. Отбор текстов для корпусов.
15. Графематический анализ.
16. Понятие разметки.
17. Типы разметки.
18. Морфологическая разметка.
19. Синтаксические корпуса (treebanks).
20. Семантическая разметка.
21. Технология создания корпусов. Стадии работы.
22. Понятие корпусоида.
23. Автоматическая морфоразметка.
24. Автоматический синтаксический анализ (parsing).

25. Языковые средства представления размеченных текстов (языки SGML, XML).
26. Международные стандарты (TEI, EAGLES, CDIF, XCES).
27. Способы использования корпусов в лингвистических исследованиях.
28. Исследование способов использования корпусов в лексикографии.
29. Изучение средств обработки корпусных данных, представленных на языке XML.
30. Обзор существующих корпусов различных типов.
31. Классификация (типология) корпусов по различным основаниям.
32. Типы корпусов по задачам. Типы корпусов по формальным признакам.

5.4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

В рамках освоения дисциплины предусмотрены следующие формы текущего контроля: устный опрос, тестирование, выполнение практических заданий, проект, практическая подготовка

Общее количество баллов по дисциплине – 100 баллов.

Максимальное количество баллов, которое может набрать обучающийся в течение семестра за текущий контроль, равняется 80/70 баллам.

Промежуточная аттестация проводится устно по вопросам в форме зачета и зачета с оценкой. На зачете и зачете с оценкой обучающийся должен дать продемонстрировать знание теоретических аспектов изучаемого курса. Максимальное число баллов, которые выставляются обучающемуся по итогам зачета - 20 баллов. Максимальное число баллов, которые выставляются обучающемуся по итогам зачета с оценкой - 30 баллов.

На зачете и зачете с оценкой дается оценка по следующим критериям:

- понимание и степень усвоения теоретического материала курса в соответствии с учебной программой;
- владение понятийным аппаратом и терминологией данной дисциплины;
- умение связать теорию с практическим применением и владение практическими навыками корпусного анализа и использования корпусного подхода и корпусной лингвистики в объеме программы.

Обязательным условием допуска студентов к зачету и экзамену является выполнение всех предусмотренных учебной программой видов текущего контроля. На зачете осуществляется комплексная проверка компетенций, знаний, навыков, умений студентов.

Требования к зачету - один теоретический вопрос на проверку владения материалом курса корпусные исследования в переводе

Шкала оценивания зачета

Критерии оценивания	Баллы
- 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	20-15

направления; – умело обосновывает и аргументирует выдвигаемые им идеи; – делает выводы и обобщения; – свободно владеет системой специализированных понятий.	
– 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; – делает выводы и обобщения; – владеет системой специализированных понятий	14-10
- 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.	9-5
– 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.	4-0

Итоговая шкала оценивания дисциплины

Итоговая оценка складывается из оценки за выполнения всех предусмотренных в программе дисциплины форм отчетности в рамках текущего контроля, а также оценки на промежуточной аттестации.

Баллы, полученные по текущему контролю и промежуточной аттестации	Оценка в традиционной системе
81-100	зачтено
61-80	зачтено
41-60	зачтено
0-40	не зачтено

Требования к зачету с оценкой – один теоретический вопрос на проверку владения материалом курса корпусные исследования в переводе

Шкала оценивания зачета с оценкой

Критерии оценивания	Баллы
Активность на занятиях (устные ответы)	5
Практические работы (4 работы по 5 баллов)	20
Тест (итоговый)	5
Итого	30

Итоговая шкала оценивания дисциплины

Итоговая оценка складывается из оценки за выполнения всех предусмотренных в

программе дисциплины форм отчетности в рамках текущего контроля, а также оценки на промежуточной аттестации.

Баллы, полученные по текущему контролю и промежуточной аттестации	Оценка в традиционной системе
81-100	отлично
61-80	хорошо
41-60	удовлетворительно
0-40	не удовлетворительно

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Основная литература

1. Захаров, В. П. Корпусная лингвистика : учебник / В. П. Захаров, С. Ю. Богданова. - 3-е изд., перераб. - Санкт-Петербург : СПбГУ, 2020. - 234 с. - ISBN 978-5-288-05997-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1244746> (дата обращения: 09.06.2024). – Режим доступа: по подписке.
2. Моделирование в корпусной лингвистике: специализированные корпуса русского языка : монография / В. П. Захаров, И. В. Азаров, О. А. Митрофанова [и др.]. - СПб : Изд-во С.-Петерб. ун-та, 2019. - 208 с. - ISBN 978-5-288-05902-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1080953> (дата обращения: 09.06.2024). – Режим доступа: по подписке.
3. Черняховская, Л. А. Перевод и смысловая структура : монография / Л. А. Черняховская. - 2-е изд, доп. - Москва : Флинта, 2021. - 280 с. - ISBN 978-5-9765-4637-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1852018> (дата обращения: 09.06.2024). – Режим доступа: по подписке.

6.2. Дополнительная литература

1. Грудева, Е.В. Корпусная лингвистика : учеб. пособие / Е.В. Грудева. - 3-е изд., стер. - Москва : ФЛИНТА, 2017. - 165 с. - ISBN 978-5-9765-1497-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1032488> (дата обращения: 09.06.2024). – Режим доступа: по подписке.
2. Гацура, Н. И. Перевод и лингвистический анализ текста : учебно-методическое пособие / Н.И. Гацура. - Омск : Издательство Омского государственного университета, 2018. - 148 с. - ISBN 978-5-7779-2263-2. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2144433> (дата обращения: 09.06.2024). – Режим доступа: по подписке.

6.3. Ресурсы информационно-телекоммуникационной сети «Интернет»

1. <http://ruscorpora.ru/instruction-main.pdf> (основной подкорпус);
2. <http://ruscorpora.ru/instruction-syntax.html> (синтаксический подкорпус);
3. <http://ruscorpora.ru/instruction-accent.pdf> (акцентологический подкорпус);
4. <http://ruscorpora.ru/instruction-murco.html> (мультимедийный подкорпус);
5. Национальный корпус русского языка (<http://www.ruscorpora.ru>)
6. Корпусы на базе Corpus Technologies (<http://web-corpora.net>)
7. Корпусы Марка Дэвиса: английские, испанский и португальский (<http://corpus.byu.edu>)
8. Корпусы, доступные через интерфейс университета Лидс (<http://corpus.leeds.ac.uk>)

9. Британский национальный корпус <http://www.natcorp.ox.ac.uk/>
10. Корпусы, доступные через интерфейс Kontext (https://kontext.korpus.cz/first_form), в частности:
11. Параллельный корпус Intercorp (<http://www.korpus.cz/intercorp>)
12. Корпусы Araneum
13. Параллельный корпус ParaSol (<http://parasol.unibe.ch>, требуется регистрация)
14. Генеральный интернет-корпус русского языка (<http://int.webcorpora.ru/drake>, требуется регистрация)
15. «Рассказы о сновидениях и другие корпуса звучащей речи» (<http://spokencorpora.ru/>)
16. LinguistList, программное обеспечение: <http://linguistlist.org/sp/software-index.cfm>
17. SIL International, программное обеспечение: <http://software.sil.org/products/>
18. Sketch Engine: <https://the.sketchengine.co.uk/> (требуется регистрация)
- Электронно-библиотечная система Лань – <https://e.lanbook.com/>
- Электронно-библиотечная система Юрайт – <https://urait.ru/>
- Notepad++ <https://notepad-plus-plus.org/>
- Ms office (word, excel) или openoffice
- Google chrome <https://www.google.ru/chrome/>
- Яндекс.браузер <https://browser.yandex.ru/>
- Морфологический анализатор mystem <https://tech.yandex.ru/mystem/>
- Корпусный менеджер antconc <http://www.laurenceanthony.net/software.html>
- Google книги n-грамм анализ: <https://books.google.com/ngrams>
- Elan – <http://tla.mpi.nl/tools/tla-tools/elan/>
- Flex: language explorer – <http://fieldworks.sil.org/flex/>

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

1. Методические рекомендации по организации самостоятельной работы студентов

8. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ

Лицензионное программное обеспечение:

Microsoft Windows

Microsoft Office

Kaspersky Endpoint Security

Информационные справочные системы:

Система «Консультант Плюс»

Профессиональные базы данных:

fgosvo.ru – Портал Федеральных государственных образовательных стандартов высшего образования

pravo.gov.ru - Официальный интернет-портал правовой информации

www.edu.ru – Федеральный портал Российское образование

Свободно распространяемое программное обеспечение, в том числе отечественного производства

ОМС Плеер (для воспроизведения Электронных Учебных Модулей)

7-zip

Google Chrome

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Материально-техническое обеспечение дисциплины включает в себя:

- учебная аудитория для проведения учебных занятий, оснащенная оборудованием и техническими средствами обучения (комплект учебной мебели, ПК, аудиосистема, экран, проектор подвесной);

- помещения для самостоятельной работы обучающихся, оснащенные компьютерной техникой, подключенной к сети Интернет, обеспечено доступом к электронной информационно-образовательной среде Государственного университета просвещения. Персональные компьютеры с подключением к сети Интернет и обеспечением доступа к электронным библиотекам и в электронную информационно-образовательную среду Университета. Доска. Программное обеспечение: Лицензионное программное обеспечение: Зарубежное: Microsoft Windows, Microsoft Office Отечественное: Kaspersky Endpoint Security Свободно распространяемое программное обеспечение: Зарубежное: Google Chrome, 7-zip Отечественное: ОМС Плеер (для воспроизведения Электронных Учебных Модулей) Информационные справочные системы: система «КонсультантПлюс» Профессиональные базы данных: fgosvo.ru pravo.gov.ru www.edu.ru;

- помещение для самостоятельной работы для инвалидов и лиц с ограниченными возможностями здоровья, оснащенное компьютерной техникой, подключенной к сети Интернет, обеспечено доступом к электронно-образовательной среде Университета. Комплект специализированной учебной мебели, беспроводный ресивер; специализированный программно-технический комплекс для обучающихся с ограниченными возможностями здоровья; гарнитура компактная; портативная информационная индукционная система; беспроводной компьютерный джойстик в комплекте с двумя выносными кнопками; беспроводная выносная компьютерная кнопка большая; портативное устройство для чтения печатных материалов; беспроводная клавиатура с большими кнопками и разделяющей клавиши накладкой; увеличитель для работы с удаленными объектами;

- **в том числе, материально-техническое обеспечение практической подготовки по дисциплине:**

- учебную аудиторию для проведения учебных занятий, оснащенную оборудованием и техническими средствами обучения: комплект учебной мебели, персональные компьютеры с подключением к сети Интернет и обеспечением доступа к электронным библиотекам и в электронную информационно-образовательную среду Государственного университета просвещения.