

Минобрнауки России

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«Оренбургский государственный университет»

Кафедра теории и практики перевода

РАБОЧАЯ ПРОГРАММА

ДИСЦИПЛИНЫ

«Б1.Д.Б.30 Компьютерная лексикография»

Уровень высшего образования

БАКАЛАВРИАТ

Направление подготовки

45.03.04 Интеллектуальные системы в гуманитарной сфере

(код и наименование направления подготовки)

Языковые модели и искусственный интеллект

(наименование направленности (профиля) образовательной программы)

Квалификация

Бакалавр

Форма обучения

Очная

Год набора 2024

Рабочая программа дисциплины «Б1.Д.Б.30 Компьютерная лексикография» рассмотрена и утверждена на заседании кафедры

Кафедра теории и практики перевода

наименование кафедры

протокол № 6 от "19" 02 2024г.

Заведующий кафедрой

Кафедра теории и практики перевода

наименование кафедры

подпись

Е.Д. Андреева

расшифровка подписи

Исполнители:

доцент

должность

подпись

Е.Д. Андреева

расшифровка подписи

должность

подпись

расшифровка подписи

СОГЛАСОВАНО:

Председатель методической комиссии по направлению подготовки

45.03.04 Интеллектуальные системы в гуманитарной сфере

код наименование

личная подпись

расшифровка подписи

Е.Д. Андреева

Заведующий отделом формирования фонда и научной обработки документов

личная подпись

расшифровка подписи

Уполномоченный по качеству факультета

личная подпись

Т.В. Сапух

расшифровка подписи

№ регистрации _____

© Андреева Е.Д., 2024
© ОГУ, 2024

1 Цели и задачи освоения дисциплины

Цель (цели) освоения дисциплины:

формирование базовой системы компетенций в области современной лексикографической практики с использованием цифровых технологий и знаний основных программных продуктов для выполнения лексикографических исследований и разработок.

Задачи:

- приобрести и (или) расширить знания основ современной теоретической и прикладной, в частности компьютерной (цифровой), лексикографии; принципов представления значения в словаре; основных типов словарей;
- сформировать умение различать значение и употребление; различные типы значений (денотативное и коннотативное, системное и актуальное и др.);
- сформировать умение осуществлять лексикографический анализ с учетом прагматических факторов значения;
- овладеть основными методами лексикографических исследований и разработок;
- сформировать умение использовать понятийный аппарат и основные концепции лексикографии в смежных областях лингвистики и гуманитарного знания.

2 Место дисциплины в структуре образовательной программы

Дисциплина относится к базовой части блока Д «Дисциплины (модули)»

Пререквизиты дисциплины: *Б1.Д.Б.24 Компьютерная лингвистика, Б1.Д.Б.28 Корпусная лингвистика, Б1.Д.В.12 Семантика*

Постреквизиты дисциплины: *Б1.Д.В.5 Большие языковые модели*

3 Требования к результатам обучения по дисциплине

Процесс изучения дисциплины направлен на формирование следующих результатов обучения

Код и наименование формируемых компетенций	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
ОПК-1 Способен применять в профессиональной деятельности методы математического анализа, логики и моделирования, теоретического и экспериментального исследования в информатике, лингвистике и гуманитарных науках	ОПК-1-В-1 Способен использовать основы математического анализа, логики и математического моделирования ОПК-1-В-2 Способен использовать математические методы для построения моделей в информатике, лингвистике и других гуманитарных дисциплинах ОПК-1-В-3 Владеет методами теоретического и экспериментального исследования в информатике, лингвистике и других гуманитарных дисциплинах	<u>Знать:</u> базовые понятия лингвистики и смежных гуманитарных наук; основы математических наук, применимые для разработки цифровых продуктов; терминологический аппарат общей и компьютерной лексикографии, ее базовые концепции; основные методологические подходы к ключевым проблемам компьютерной лексикографии; основные методы и приемы анализа и синтеза лексикографических данных. <u>Уметь:</u> пользоваться различного рода электронными словарями; соотносить междисциплинарные теоретические знания с решением прикладной лексикографической задачи; проводить формализацию

Код и наименование формируемых компетенций	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
		лексикографического материала; применять цифровые технологии для анализа и синтеза лексикографических данных. <u>Владеть:</u> базисным лингвистическим аппаратом, необходимым для работы с компьютерными программами по обработке лексикона; методикой отбора лексикографических данных и их репрезентации в словаре; математическими методами обработки лексикографического материала; навыками критической оценки эффективности различных методов и технологий современной лексикографической практики.

4 Структура и содержание дисциплины

4.1 Структура дисциплины

Общая трудоемкость дисциплины составляет 3 зачетные единицы (108 академических часов).

Вид работы	Трудоемкость, академических часов	
	6 семестр	всего
Общая трудоёмкость	108	108
Контактная работа:	52,25	52,25
Лекции (Л)	18	18
Практические занятия (ПЗ)	34	34
Промежуточная аттестация (зачет, экзамен)	0,25	0,25
Самостоятельная работа: - выполнение индивидуального творческого задания (ИТЗ); - самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий; - изучение разделов курса в системе электронного обучения; - подготовка к практическим занятиям)	55,75	55,75
Вид итогового контроля (зачет, экзамен, дифференцированный зачет)	диф. зач.	

Разделы дисциплины, изучаемые в 6 семестре

№ раздела	Наименование разделов	Количество часов				
		всего	аудиторная работа			внеауд. работа
			Л	ПЗ	ЛР	
1	Лексикография как раздел лингвистики	12	2	4		6
2	Понятие и строение словаря	16	2	4		10
3	Системная лексикография	26	6	10		10
4	Компьютерная лексикография	54	8	16		30

№ раздела	Наименование разделов	Количество часов				
		всего	аудиторная работа			внеауд. работа
			Л	ПЗ	ЛР	
	Итого:	108	18	34		56
	Всего:	108	18	34		56

4.2 Содержание разделов дисциплины

1 Лексикография как раздел лингвистики

Место лексикографии в системе лингвистического знания. Круг проблем теоретической лексикографии. Содержание и задачи прикладной лексикографии. Соотношение лексикографии и лексикологии. Теоретическая база лексикографии как самостоятельной лингвистической дисциплины. История лексикографии. Древнейшие глоссарии и вокабулярии; тематические подборки слов как основа будущего развития идеографии. Словарная разработка языковых идиомов (в основном сакральных языков), значительно отличающихся от разговорного языка народа. Создание переводных словарей. Развитие литературных языков и словарей национальных литературных языков (словарная деятельность первых европейских академий). Русская лексикография: словари-ономастиконы, словари символов, славяно-русские словари, разговорники, азбуковники, активное развитие словарного дела в XVIII в., «Словарь Академии Российской», словарные поиски и труды А.Х. Востокова, И.И. Срезневского, Я.К. Грота, А.А. Шахматова, В.И. Даля. Становление научного языкознания и появление новых типов словарей (орфографических, фразеологических, синонимических, терминологических и др.). Современные тенденции развития лексикографии.

2 Понятие и строение словаря

Словарь в жизни языка и общества. Словарь как памятник культуры. Проблема типологии словарей. Типология Л.В. Щербы и ее развитие в типологиях Ю.Н. Караулова, П.Н. Денисова, В.Г. Гака и других исследователей. Словари энциклопедические и лингвистические. Активные и пассивные словари. Нормативные и описательные словари. Словари идеографические и семантические. Бумажные и электронные версии. Общие и частные толковые словари. Онтологии. Словари, описывающие различные пласты лексики. Словари, отображающие внутрилингвистические группировки слов: словари синонимов, антонимов, паронимов, омонимов, фразеологические. Словари, отражающие динамику лексики: словари неологизмов, устаревших и редких слов, исторические. Словари лексики, социально и территориально ограниченной. Словари, отражающие происхождение слов: этимологические, иноязычных слов. Ономастические словари: идеографические, ассоциативные. «Функционально-языковые» словари: сочетаемости, грамматические, словообразовательные, морфемные, орфографические, орфоэпические, трудностей, частотные. «Функционально-текстовые» словари: словари языка писателей, эпитетов, метафор, рифм. «Культурологические» словари: ономастические, лингвострановедческие, лингвокультурологические. Комплексные словари. Переводные словари. Двуязычные и многоязычные словари. Новые типы словарей. Терминологические электронные тезаурусы (структура и использование) Словари-тезаурусы. Современные терминологические тезаурусы, принципы организации, назначение, прикладное использование. Словари активного типа.

Современные словари и их информативные возможности. Принципы отбора информации в словарях разного типа. Лексикографические параметры и их словарная реализация. Макро- и микроструктура словаря. Структура и состав словарной статьи. Словарные входы. Словник и заголовочное слово, состав правой части различных типов словарей; важнейшие словарные зоны (толкование, перевод, значение, примечания, конструкции, сочетаемость, иллюстрации). Требования к словарю. Вспомогательные тексты, инструкции по использованию, указатели, объем словаря. Этапы построения словаря. Проблема источников словаря. Основные источники языкового материала для словарей.

3 Системная лексикография

Основные принципы системной лексикографии. Реконструкция языковой картины мира. Системный характер лексики. Системообразующий смысл. Метаязык описания. Асимметрия

языкового знака: моносемия, полисемия, омонимия. Изменение значения слов: расширение – сужение – смещение – перенос. Регулярная и окказиональная многозначность. Значение – употребление (Ю.Д. Апресян). Смысл (sens) – значение (acception) – употребление (emploi) (Р. Мартен). Правила сочетаемостных ограничений. Лексика как система. Семантические классы в языке и контексте. Таксономии и доха. Взаимодействие лексических значений. Семантическое согласование. Семантическая модификация (конденсация, нейтрализация, транспозиция (сдвиг) семантических признаков). Взаимодействие грамматических значений. Взаимодействие лексических значений с грамматическими значениями. Семантико-прагматическое представление лексемы (ингерентные и афферентные признаки). Прагматическая информация. Интегральная характеристика лексемы по просодическим, морфологическим, сочетаемостным, словообразовательным, синтаксическим, стилистическим свойствам. Семантические связи лексемы с другими лексемами: (фразеологизмы, синонимы, антонимы, конверсивы, дериваты). Лексикографический тип.

4. Компьютерная лексикография

История развития компьютерной лексикографии. Этапы формирования компьютерной лексикографии. Эксперимент Р. Бузы. Автоматизация и оптимизация работы лексикографа по созданию традиционных (бумажных) словарей. «Машиночитаемые словари». Структурирование, кодифицирование, обработка информации на естественных языках. Автоматическая обработка текста. Извлечение информации. Типы информации для компьютерных словарей. Алгоритмы сортировки и создания выборки. Банки эмпирических данных и банки рабочих лексикографических результатов. Терминологические банки данных. Технология электронных корпусов и корпусных словарей. Методологические и лингвистические аспекты корпусной лексикографии. Разрешение лексической многозначности: WSD. Автоматические конкордансы. Электронные ассоциативные словари. Электронные словари-онтологии. Компьютерная поддержка создания и функционирования онтологий. Словарные системы типа WordNet (назначение, структура, использование). Описания лексики. Исследование частей речи. Структура машинной словарной статьи и ее отличия от структуры словарной статьи традиционного словаря. Зоны словарной статьи в компьютерной лексикографии. Лексический вход (вокабула, лемма). Зона грамматической информации. Зона стилистических помет. Зона значения. Зона фразеологизмов. Зона этимологии. Зона примера и источника примера. Структура машинной словарной статьи. Зона морфологических сведений. Кодировки для обозначения частей речи и представления морфологической информации. Зона семантических и зона лексических сведений машинной словарной статьи. Работа над электронными терминологическими словарями и банками данных перспективных направлений компьютерной лексикографии и прикладной лингвистики. Принципы компьютерной терминографии. Унификации терминосистем и терминологические стандарты. Терминологические базы (банки) данных (ТБД). Переводческие (многоязычные) и информационно-нормативные (одноязычные) ТБД. Свойства терминов, сопоставимые с отдельными зонами словарной статьи. Перспективы современной компьютерной лексикографии.

4.3 Практические занятия (семинары)

№ занятия	№ раздела	Тема	Кол-во часов
1	1	Теоретическая и прикладная лексикография	2
2	1	История лексикографии	2
3	2	Типология словарей	2
4	2	Структура словаря и словарной статьи	2
5	3	Системный характер лексики	2
6	3	Семантические классы и лексическое значение	2
7	3	Лексическое и грамматическое значение	2
8	3	Прагматическая информация	2
9	3	Связи лексем. Лексикографический тип	2
10	4	История развития компьютерной лексикографии. Типы информации для компьютерных словарей	2
11	4	Электронные словари. Структура машинной словарной статьи.	2

№ занятия	№ раздела	Тема	Кол-во часов
12	4	Компьютерная терминология. Машинные переводные словари.	2
13	4	Проблемы компьютерной формализации лексики.	2
14	4	Разрешение омонимии и многозначности.	2
15	4	Современные технологии в компьютерной лексикографии	2
16	4	Перспективы компьютерной лексикографии	2
17	4	Представление проекта	2
		Итого:	34

5 Учебно-методическое обеспечение дисциплины

5.1 Основная литература

1. Иванищева, О.Н. Лексикографирование культуры : учебное пособие / О.Н. Иванищева. – Москва ; Берлин : Директ-Медиа, 2015. – 171 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=428738>. – ISBN 978-5-4475-5685-3. – Текст : электронный.

2. Кишина, Е.В. Русская лексикография : учебное пособие / Е.В. Кишина. – Кемерово : Кемеровский государственный университет, 2012. – 140 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=232403>. – ISBN 978-5-8353-1286-3. – Текст : электронный.

5.2 Дополнительная литература

1. Антопольский, А.Б. Лингвистические информационные ресурсы / А.Б. Антопольский. – Москва : Институт научной информации по общественным наукам (ИНИОН) РАН, 2022. – 466 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=695077>. – ISBN 978-5-248-01030-1. – Текст : электронный.

2. Апресян, Ю.Д. Избранные труды : сборник научных трудов / Ю.Д. Апресян. – Москва : Школа «Языки русской культуры», 1995. – Том 2. Интегральное описание языка и системная лексикография. – 768 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=210920>. – ISBN 5-88766-045-7. – Текст : электронный.

3. Апресян, Ю.Д. Исследование по семантике и лексикографии / Ю.Д. Апресян. – Москва : Языки славянской культуры (ЯСК), 2009. – Том I. Парадигматика. – 569 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=73322>. – ISBN 978-5-9551-0304-4. – Текст : электронный.

4. Богачева, Г.Ф. Лексическое значение как объект словарного толкования / Г.Ф. Богачева. – Москва : ФЛИНТА, 2013. – 208 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=375628>. – ISBN 978-5-9765-1452-2. – Текст : электронный.

5. Болотнова, Н.С. Современный русский язык : лексикология. Фразеология. Лексикография. Контрольно-тренировочные задания / Н.С. Болотнова, А.В. Болотнов. – Москва : ФЛИНТА, 2019. – 224 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=83070>. – ISBN 978-5-9765-0739-5. – Текст : электронный.

6. Влавацкая, М.В. Теоретические проблемы комбинаторной лингвистики. Лексикология. Лексикография / М.В. Влавацкая. – Новосибирск : Новосибирский государственный технический университет, 2011. – 506 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=135592>. – ISBN 978-5-7782-1824-6. – Текст : электронный.

7. Гончарова, Т.В. Лексикология. Лексикография. Фразеология : контрольные задания и справочные материалы / Т.В. Гончарова, Е.А. Попова. – Липецк : Липецкий государственный педагогический университет им. П.П. Семенова-Тян-Шанского, 2017. – 111 с. – Режим доступа: по

подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=576652>. – ISBN 978-5-88526-805-9. – Текст : электронный.

8. Евсюкова, Т.В. Лексикология / Т.В. Евсюкова, Е.А. Чередникова. – Ростов-на-Дону : Издательско-полиграфический комплекс РГЭУ (РИНХ), 2021. – 112 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=685531>. – ISBN 978-5-7972-2902-5. – Текст : электронный.

9. Рубашкин, В.Ш. Онтологическая семантика. Знания. Онтологии. Онтологически ориентированные методы информационного анализа текстов / В.Ш. Рубашкин. – Москва : Физматлит, 2012. – 347 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=457687>. – ISBN 978-5-9221-1439-4. – Текст : электронный.

10. Шестакова, Л.Л. Русская авторская лексикография : теория, история, современность / Л.Л. Шестакова. – Москва : Языки славянской культуры (ЯСК), 2011. – 464 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=211593>. – ISBN 978-5-9551-0456-0. – Текст : электронный.

5.3 Периодические издания

Не предусмотрены

5.4 Интернет-ресурсы

<https://ruscorpora.ru/> – Национальный корпус русского языка
<http://philologos.narod.ru/index.html> – портал лингвистических ресурсов
<https://wordnet.princeton.edu/> – WordNet, лексическая база данных английского языка
<https://www.english-corpora.org/bnc/> – Британский национальный корпус
<https://wordnet.ru/> – Русский WordNet
<https://dic.academic.ru/> – словари на Академике
<https://www.slovari.ru/start2.aspx?s=0&p=5638> – собрание словарей русского языка
<http://glossword.biz/> – свободное ПО для создания и публикации словарей
<https://tcdh.uni-trier.de/en/thema/digital-edition-and-lexicography> – Центр цифровой гуманитаристики, раздел компьютерной лексикографии
<https://www.digitallexicography.com/> – проект по созданию словарей для малоресурсных языков
<https://www.bbaw.de/en/bbaw-digital/zentrum-fuer-digitale-lexikographie-der-deutschen-sprache-zdl> – Center for Digital Lexicography for the German Language
<https://journals.openedition.org/jtei/> – Journal of the Text Encoding Initiative
https://elex.link/elex2021/wp-content/uploads/eLex_2021-proceedings.pdf – Electronic lexicography in the 21st century (eLex 2021): Post-editing lexicography/ Proceedings of the eLex 2021 conference
<https://link.springer.com/journal/40607/volumes-and-issues> – журнал Lexicography на сайте Springer
<https://journall.org/index.php/main> – Journal of Applied Linguistics and Lexicography
«Лингвистические подходы к дискурсу» [Электронный ресурс]: онлайн-курс на платформе <https://openedu.ru> – «Открытое образование» / Разработчик курса: Национальный исследовательский университет «Высшая школа экономики», режим доступа: <https://openedu.ru/course/hse/DISCOURSE/>

5.5 Программное обеспечение, профессиональные базы данных и информационные справочные системы

1. Операционная система РЕД ОС
2. Пакет офисных приложений LibreOffice
3. Программная система для организации видео-конференц-связи МТС Линк
4. ГАРАНТ Платформа F1 [Электронный ресурс]: справочно-правовая система. / Разработчик ООО НПП «ГАРАНТ-Сервис», 119992, Москва, Воробьевы горы, МГУ, [1990–2023]. – Режим доступа в сети ОГУ для установки системы: \\fileserver1\GarantClient\garant.exe

5. КонсультантПлюс [Электронный ресурс]: электронное периодическое издание справочная правовая система. / Разработчик ЗАО «Консультант Плюс», [1992–2023]. – Режим доступа к системе в сети ОГУ для установки системы: \\fileserver1\!CONSULT\cons.exe

6. <http://edu.garant.ru/garant/study/> – Интернет-версия ГАРАНТ-Образование, Система ГАРАНТ для студентов, аспирантов и преподавателей

7. Мультиплатформенная программа для проведения корпусных лингвистических исследований и управления данными AntConc. Доступна бесплатно после принятия условий лицензионного соглашения. Разработчик: Laurence Anthony (Center for English Language Education (CELESE), Faculty of Science and Engineering Waseda University). Режим доступа: <http://www.laurenceanthony.net/software.html>.

6 Материально-техническое обеспечение дисциплины

Учебные аудитории для проведения занятий семинарского типа, для проведения групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации.

Аудитории оснащены комплектами ученической мебели, техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой, подключенной к сети «Интернет», и обеспечены доступом в электронную информационно-образовательную среду ОГУ.