

Київський університет імені Бориса Грінченка
Інститут філології
Кафедра української мови

“ЗАТВЕРДЖУЮ”

Проректор з науково-методичної
та навчальної роботи



О.Б.Жильцов
2021 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Лінгвістика комп'ютерного аналізу і обробки інформації: Корпусна лінгвістика

для студентів

Галузі знань	03 Гуманітарні науки
Спеціальності	035 Філологія
Спеціалізації	035.01 Українська мова та література
Освітньої програми	035.01.01 Українська мова та література
Вибіркова спеціалізація	Прикладна лінгвістика
Освітнього рівня	першого (бакалаврського)



Київ – 2021

Розробники:

Дроботенко Валентина Юріївна, кандидат філологічних наук, доцент.

Горохова Тетяна Олександрівна, кандидат педагогічних наук, старший викладач кафедри української мови.

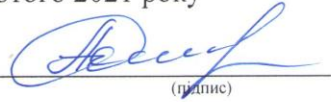
Викладач:

Горохова Тетяна Олександрівна, кандидат педагогічних наук, старший викладач кафедри української мови.

Робочу програму розглянуто і затверджено на засіданні
кафедри української мови

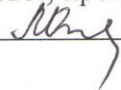
Протокол № 8 від 3 лютого 2021 року

Завідувач кафедри



(С.О. Караман)
(ініціали, прізвище)

Гарант освітньої (професійної/наукової) програми
(керівник проектної групи)

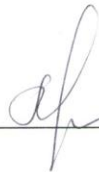


(М. О. Вінтонів)
(ініціали, прізвище)

Робочу програму перевірено

08.02.2020 р.

Заступник директора/декана
Пролонговано:



(А.І. Золотарь)

на 20__/20__ н.р. _____. «__» 20__ р., протокол № ____
(підпис) (ПІБ)

на 20__/20__ н.р. _____. «__» 20__ р., протокол № ____
(підпис) (ПІБ)

на 20__/20__ н.р. _____. «__» 20__ р., протокол № ____
(підпис) (ПІБ)

на 20__/20__ н.р. _____. «__» 20__ р., протокол № ____
(підпис) (ПІБ)

1. Опис навчальної дисципліни

Найменування показників	Характеристика дисципліни за формами навчання
	денна
Лінгвістика комп'ютерного аналізу і обробки інформації	
Вид дисципліни	<i>вибіркова</i>
Мова викладання, навчання та оцінювання	<i>українська</i>
Загальний обсяг кредитів / годин	<i>20 /600</i>
Курс	<i>2-4</i>
Семестр	<i>3-7</i>
Кількість змістових модулів з розподілом:	<i>5</i>
Обсяг кредитів	<i>20</i>
Обсяг годин, у тому числі:	<i>600</i>
Аудиторні	<i>280</i>
Модульний контроль	<i>40</i>
Самостійна робота	<i>280</i>
Форма семестрового контролю	<i>залік (3,4, 6,7)</i>

Змістовий модуль 1. Корпусна лінгвістика.

Найменування показників	Характеристика дисципліни за формами навчання
	денна
Вид дисципліни	<i>вибіркова</i>
Мова викладання, навчання та оцінювання	<i>українська</i>
Загальний обсяг кредитів / годин	<i>4/ 120</i>
Курс	<i>2</i>
Семестр	<i>4</i>
Кількість змістових модулів з розподілом:	<i>2</i>
Обсяг кредитів	<i>4</i>
Обсяг годин, у тому числі:	<i>120</i>
Аудиторні	<i>56</i>
Модульний контроль	<i>8</i>
Самостійна робота	<i>56</i>
Форма семестрового контролю	<i>залік</i>

ПОЯСНЮВАЛЬНА ЗАПИСКА

II. Мета та завдання навчальної дисципліни

Мета навчальної дисципліни «Корпусна лінгвістика»: ознайомити студентів з базовими поняттями корпусної лінгвістики, охарактеризувати її мету та завдання, описати методи дослідження в цій галузі, представити основи корпусних технологій та показати теоретичне й практичне значення корпусів для проведення наукових досліджень в галузі лінгвістики.

Завдання навчальної дисципліни:

- виховання в студентів почуття національної свідомості, що значною мірою сприяє підвищенню мотивації вивчення української мови;
- формування комунікативної, мовної, мовленнєвої, предметної, прагматичної, соціокультурної компетентностей студентів на основі свідомого опанування мовної і мовленнєвої теорії;
- корекція уявлення студентів про можливе застосування корпусів у лінгвістичних дослідженнях, зокрема в українському мовознавстві.

Основними **завданнями** вивчення дисципліни «Корпусна лінгвістика» є:

- 1) ознайомити студентів із предметами, завданнями, методами, понятійним (термінологічним) апаратом корпусної лінгвістики;
- 2) ознайомити студентів з типологією корпусів, а також з наявними мовними національними та дослідницькими корпусами української, англійської та інших мов і паралельними (двомовними) корпусами;
- 3) сформувати знання про комп'ютерний інструментарій корпусної лінгвістики (корпусні менеджери, програми-конкордансери тощо);
- 4) надати студентам первинне уявлення про можливі застосування корпусів у лінгвістичних дослідженнях.

Під час занять, індивідуальної навчально-дослідницької та самостійної роботи студенти, майбутні бакалаври з прикладної лінгвістики, **набувають таких програмних компетентностей:**

• **ЗК3:** Інформаційна компетенція. Здатність до самостійного пошуку та оброблення інформації з різних джерел для розгляду конкретних питань; до ефективного використання інформаційних технологій в соціальній і професійній діяльності. Сервіс-орієнтування.

ЗК7: Самоосвітня компетенція. Здатність до самостійної пізнавальної діяльності, самоорганізації та саморозвитку. Спрямованість на розкриття особистісного потенціалу та самореалізацію. Прагнення до особистісно-професійного лідерства та успіху.

ФК2 Лінгвістична (глибокі знання та розуміння в галузі лінгвістики). Здатність аналізувати лінгвістичні явища з погляду фундаментальних філологічних принципів і знань, а також на основі відповідних загальнонаукових методів. Знання природи і сутності мови, походження й закономірностей розвитку мов, особливостей структури мови, класифікації мов світу, шляхів і методів вивчення мовного матеріалу. Знання системи сучасної української літературної мови, еволюції її граматичної будови

ФК5 Прикладна лінгвістично-орієнтована Здатність застосовувати нелінгвістичні знання в практичній філологічній діяльності. Уміти застосовувати знання ІКТехнологій для виконання лінгвістичних завдань. Володіти базовими знаннями з юриспруденції, вміти застосовувати їх під час аналізу мовної політики та мовної ситуації в державі, проведення експертиз юридично значимих текстів.

ФК6 Комунікативна професійно-орієнтована Здатність застосовувати в професійній діяльності нормативні засоби мови в усному та писемному мовленні з урахуванням змістового наповнення, соціально-демографічних особливостей співрозмовника, специфіки ситуації спілкування та контексту. Здатність вільно й

ефективно використовувати мову(и), що вивчається(ються), для розв'язання комунікативних завдань.

Організація навчальної діяльності в центрі культури української мови

Реформування системи вищої освіти на сучасному етапі зумовлює концептуальні зміни в напрямках, цілях, завданнях, змісті освіти, філологічної зокрема, які зорієнтовують передовсім на розвиток особистості, формування творчої ініціативи, мобільності, підвищення конкурентоздатності майбутнього фахівця.

Задля ефективної реалізації концепції нової освітньої стратегії університету програмою передбачено застосування можливостей Центру культури української мови.

Для цього впроваджено систематичне застосування інноваційних тренінгів, дослідницьких завдань, що передбачають карткування укладання мультимедійних словників, опрацювання спеціальної літератури тощо.

Крім того, студентів залучено до роботи в проблемних групах, до участі в наукових конференціях; мовно-інформаційних конкурсах тощо.

У Центрі заплановано презентації наукових видань з проблем культури української мови, зустрічі з науковцями.

3. Результати навчання за дисципліною

Опанувавши дисципліну «Корпусна лінгвістика», студенти повинні мати такі *програмні результати навчання*:

ПРН-3-6: Володіння методами і прийомами здійснення усної та писемної комунікації; методами і прийомами створення різних типів і видів текстів (мовних творів); методами і прийомами аналізу й інтерпретації різних типів і видів текстів (мовних творів); методами і прийомами обробки та трансформації різних типів і видів текстів (творів).

ПРН-3-7: Знання основних бібліографічних показників і пошукових систем, володіння прийомами бібліографічного опису різних джерел; сучасними методами і прийомами обробки інформації в галузі професійної діяльності.

ПРН-3-9: Здатність застосовувати нелінгвістичні знання у філологічній діяльності. Знання мовного законодавства, уміння застосовувати його у філологічній та педагогічній діяльності, аналізувати конфліктні та юридично значимі тексти. Оволодіння навичками застосовувати ІКТ- технології, лінгвістичні програмні продукти для створення баз даних, автоматизованих лексикографічних систем, систем машинного перекладу і користування ними.

ПРН-У-1: Вільне володіння базовим професійним категоріальнопоняттєвим та дослідницьким апаратом сучасної філологічної науки. Уміння пояснювати професійно з історичних позицій явища і факти української мови та літератури

ПРН-У-7: Ефективно застосовувати ІКТ-технології у науково-дослідній і науково-педагогічній діяльності.

ПРН-У-8: Уміння працювати самостійно, працювати з метою підвищення фахової компетентності, самоосвіти; уміння працювати в умовах обмеженого часу з акцентом на професійну сумлінність та унеможливлення плагіату.

ПРН-С-1: Фахово формувати й аргументувати власні теоретичні і практичні судження, професійні позиції

ПРН-С-3: Здатність формувати судження про значення і наслідки своєї професійної діяльності з урахуванням соціальних, професійних і етичних позицій.

III. Структура навчальної дисципліни

Тематичне планування навчальної дисципліни

Назви змістових модулів і тем	Кількість годин						
	денна форма						
	Усього	у тому числі					МК
		л	пр	сем.	семестр. контроль	с.р.	
1	2	3	4		6	7	8
Змістовий модуль 1. Корпусна лінгвістика як галузь прикладного мовознавства. Корпусні студії: історична перспектива та сучасний стан. Корпус текстів природної мови							
Тема 1. Корпусна лінгвістика як галузь прикладного мовознавства. Вступ. Завдання і основні напрями корпусної лінгвістики. Співвідношення корпусної лінгвістики і комп'ютерної лінгвістики	6	2		2		2	
Тема 2. Предмет дослідження. Загальна характеристика лінгвістичних корпусів. Машинні фонди, бази даних, корпус.	7	2		2		3	
Тема 3. Розвиток лінгвістичних корпусів у світі.	5			2		3	
Тема 4. Види корпусів : усні і писемні, одномовні і багатомовні. Анотовані і неанотовані корпуси.	7	2		2		3	
Тема 5. Розмітка. Види розміток. Лінгвістична анотація і метадані.	9		2	2		3	2
Тема 6. Методи вилучення інформації з корпусу.	7		2	2		3	
Тема 7. Створення формалізованих описів для розроблення корпусу. Словниковий і безсловниковий спосіб.	5			2		3	
Тема 8. Анотування корпусу. Збереження, публікація і оновлення корпусів.	5			2		3	
Тема 9. Огляд корпусних систем по слов'янських мовах	7	2		2		3	

Тема 10. Актуальні проблеми корпусної лінгвістики української мови (слов'янських мов)	9	2		2		3	2
Всього за 1 модуль	67	10	4	20		29	4

Змістовий модуль 2. Технологія створення корпусів та комп'ютерні програми для роботи з корпусом. Можливості використання корпусів у лінгвістичних дослідженнях							
Тема 1. Етапи створення корпусу	7		2	2		3	
Тема 2. Формат даних і стандартизація даних корпусу	5			2		3	
Тема 3. Комп'ютерні програми для аналізу даних корпусу	5			2		3	
Тема 4. Сфери застосування лінгвістичних корпусів	5			2		3	
Тема 5. Лінгвістичні дослідження на базі корпусу: вивчення лексики.	7			2		3	2
Тема 6. Лінгвістичні дослідження на базі корпусу: укладання словників.	7		2	2		3	
Тема 7. Лінгвістичні дослідження на базі корпусу: вивчення граматики.	5			2		3	
Тема 8. Створення свого корпусу. Планування. Збір і оцифрування даних. Кодування тексту.	5			2		3	
Тема 9. Захист проєктів.	7			2		3	2
Всього за 2 модуль	53		4	18		27	4
Всього	120	10	8	38		56	8

5. Програма навчальної дисципліни

Змістовий модуль 1.

Корпусна лінгвістика як галузь прикладного мовознавства.

Корпусні студії: історична перспектива та сучасний стан.

Корпус текстів природної мови

Тема 1. Вступ. Завдання і основні напрями корпусної лінгвістики. Співвідношення корпусної лінгвістики і комп'ютерної лінгвістики

Основні питання теми:

Сутність, предмет і завдання корпусної лінгвістики
Корпусна лінгвістика в системі мовознавчих наук
Відмінності між корпусною та традиційною лінгвістикою; корпусною лінгвістикою і комп'ютерною лінгвістикою.

Рекомендована література:

Основна: 1–5.

Додаткова: 1, 2, 5, 6.

Тема 2 Предмет дослідження. Загальна характеристика лінгвістичних корпусів. Машинні фонди, бази даних, корпус.

Основні питання теми:

Стратегії створення тексторієнтованих баз даних
Повнотекстові бази (=корпуси текстів)
Електронні картотеки (=ілюстративні бази даних, бази цитат)
Продукти опрацювання тексторієнтованих баз даних: комп'ютерні
словопокажчики, конкорданси та частотні словники

Рекомендована література:

Основна: 1–5.

Додаткова: 3, 4, 5, 6.

Тема 3 Розвиток лінгвістичних корпусів у світі.

Основні питання теми:

Історія становлення корпусної лінгвістики: від паперових конкордансів і картотек до перших електронних корпусів.
Корпусна лінгвістика з 60-х років XX ст. до пост 2000-х.

Рекомендована література:

Основна: 1–5.

Додаткова: 1, 2, 3, 6.

Тема 4 Види корпусів : усні і писемні, одномовні і багатомовні. Анотовані і неанотовані корпуси.

Основні питання теми:

Поняття лінгвістичного корпусу
Усні – писемні корпуси
Статичні – динамічні корпуси
Одномовні – багатомовні корпуси
Анотовані – неанотовані корпуси

Рекомендована література:

Основна: 1–5.

Додаткова: 1, 2, 3, 4, 5, 6.

Тема 5 Розмітка. Види розміток. Лінгвістична анотація і метадані.

Основні питання теми:

Метарозвідка корпусів текстів: стандарти і реалізація
Типи корпусної розмітки.

- * Структурне мараування
 - * Лінгвістична анаація.
- Вимоги до розмітки.

Рекомендована ліаература:

Основа: 1–5.

Додаткова: 1, 2, 6.

Тема 6. Методи вилучення інформації з корпуса.

Основні питання теми:

- Спеціальні комп'ютерні програми для вилучення інформації.
- Типи вилученої інформації.
- Списки слів і конкорданси.

Рекомендована ліаература:

Основа: 1–5.

Додаткова: 1, 3, 4, 5.

Тема 7 Створення формалізованих описів для розроблення корпуса. Словниковий і безсловниковий спосіб.

Основні питання теми:

- Мовний корпус як комп'ютерна база даних.
- Процеси токенізації, лематизації, стемінгу і парсингу: специфіка, цілі, результати.
- Корпусні менеджери.

Рекомендована ліаература:

Основа: 1–5.

Додаткова: 1, 2, 3, 6.

Тема 8 Анотування корпуса. Збереження, публікація і оновлення корпусів.

Основні питання теми:

- Одиниця збереження корпусу даних.
- Формат подачі інформації.
- Онлайн корпусу.
- Офлайн корпусу.
- Програми анотування корпусів.
- Проблема оновлення корпусів.

Рекомендована ліаература:

Основа: 1–5.

Додаткова: 1, 2, 3, 4, 5, 6.

Тема 9 Огляд корпусних систем української мови

Основні питання теми:

- Історія створення і розроблення корпусних систем української мови.
- Генеральний регіонально анотований корпус української мови.
- Корпус текстів української мови на лінгвістичному порталі mova.info.
- Загальномовний (або національний) неанотований та несистематизований корпус української мови.
- Корпус українських інтернет-текстів у Лейпцизькому університеті.
- Корпус інтернет-текстів на платформі «Лабораторія української»

Рекомендована література:

Основна: 1–5.

Додаткова: 1, 2, 3, 4, 5, 6.

Тема 10 Актуальні проблеми корпусної лінгвістики української мови (слов'янських мов)

Основні питання теми:

Дослідження в галузі граматики і лексики слов'янських мов (української мови).

Проблеми створення корпусу усного мовлення слов'янських мов.

Відсутність корпусних розроблень з деяких мов слов'янської групи.

Створення нових типів розмітки для наявних корпусів.

Проблема глибини розмітки.

Проблема омонімії у розроблених корпусних системах.

Рекомендована література:

Основна: 1–5.

Додаткова: 1, 2, 3, 4, 5, 6.

Змістовий модуль 2.

Технологія створення корпусів та комп'ютерні програми для роботи з корпусом.

Можливості використання корпусів у лінгвістичних дослідженнях

Тема 1. Етапи створення корпусу

Основні питання теми:

Визначення переліку джерел.

Оцифрування текстів (перетворення на комп'ютерну форму).

Попереднє оброблення текстів (філологічне корегування, підготовка екстралінгвістичного опису тексту).

Рекомендована література:

Основна: 5, 7, 8, 9

Додаткова: 1, 2, 3, 4, 5, 6.

Тема 2 Формат даних і стандартизація даних корпусу

Основні питання теми:

Міжнародні стандарти в галузі корпусної лінгвістики.

- Стандарти кодування корпусів.

- Стандарти оцінки корпусів.

Огляд міжнародних стандартів корпусної лінгвістики.

Стандарти TEI. Структура тексту TEI.

Рекомендації щодо створення мовних корпусів.

Рекомендована література:

Основна: 5, 7, 8, 9

Додаткова: 1, 2, 3, 4, 5, 6.

Тема 3 Комп'ютерні програми для аналізу даних корпусу

Основні питання теми:

Програма AntConc як одна із найзручніших безкоштовних комп'ютерних програм для роботи

з корпусом.

Програми аналізу і лінгвістичного оброблення текстів (TextAnalyst, AOT та ін.)

Програми перетворення текстів (TextTransformer, XReplacer)

Психолінгвістичні програми (ПСИ-Офіс, ВААЛ-міні)

Генератори текстів (RWC)

Системи оброблення природної мови і машинного перекладу (Автоматичний словник Мультитран, Google Перекладач)

Словники і тезауруси

Рекомендована література:

Основна: 5, 7, 8, 9

Додаткова: 1, 2, 3, 4, 5, 6.

.

Тема 4 Сфери застосування лінгвістичних корпусів

Основні питання теми:

Корпуси з інформацією про динаміку процесів зміни лексичного складу мови.

Корпуси щодо функціонування граматичних та синтаксичних одиниць.

Корпуси для аналізу вживання мовних одиниць у різних стилях, жанрах, у різних авторів тощо.

Корпуси для навчання іноземної мови.

Корпуси для аналізу дискурсу, у літературознавстві, перекладознавстві, судовій лінгвістиці.

Рекомендована література:

Основна: 5, 7, 8, 9

Додаткова: 1, 2, 3, 4, 5, 6.

Тема 5 Лінгвістичні дослідження на базі корпусу: вивчення лексики.

Основні питання теми:

Вивчення семантики на базі лінгвістичних корпусів: PropBank, WordNet, FrameNet.

Collins для підготовки словників світових лексикографічних видавництв.

Clause patterns in Modern British English: A corpus-based (quantitative) study як результат корпусного аналізу структури складного речення.

Лінгвістичні корпуси для прагматичних досліджень (corpus of London teenage language, COLT).

Рекомендована література:

Основна: 5, 7, 8, 9

Додаткова: 1, 2, 3, 4, 5, 6.

Тема 6 Лінгвістичні дослідження на базі корпусу: укладання словників.

Основні питання теми:

Лексикографія як лінгвістична дисципліна. Її термінологічний апарат. Розвиток лексикографії як науки. Принципи сучасної лексикографії.

Підходи до класифікації лінгвістичних словників англійської мови. Систематичний опис типів словників.

Лексикографічні дослідження, засновані на корпусах.

Рекомендована література:

Основна: 5, 6, 7, 8, 9

Додаткова: 1, 2, 3, 4, 5, 6.

Тема 7 Лінгвістичні дослідження на базі корпусу: вивчення граматики.

Основні питання теми:

Корпусні дані як база для написання граматики (An Empirical Grammar of the English Verb: Modal Verbs Дієтера Міндта (1995) та Longman Grammar of Spoken and Written English (Д. Бібер, С. Конрад).

Роль національних корпусів української мови у вивченні граматики (роль корпусів у дослідженні частин мови, синтаксису)

Рекомендована література:

Основна: 5, 7, 8, 9

Додаткова: 1, 2, 3, 4, 5, 6.

Тема 8 Створення свого корпусу. Планування. Збір і оцифрування даних. Кодування тексту.

Основні питання теми:

Принципи створення корпусних систем.

Відомі кодування символів. Переваги того чи іншого виду кодування.

Збір і оцифрування для свого корпусу.

Рекомендована література:

Основна: 5, 7, 8, 9

Додаткова: 1, 2, 3, 4, 5, 6.

Тема 9. Захист проєктів.

Основні питання теми:

Підготувати опис переваги ручного і автоматичного анотування корпусу.

Підготувати опис наявних корпусних менеджерів.

Підготувати опис переваг онлайн і офлайн корпусу.

Рекомендована література:

Основна: 5, 7, 8, 9

Додаткова: 1, 2, 4, 5, 6.

6. Контроль навчальних досягнень

6.1. Система оцінювання навчальних досягнень студентів

Вид діяльності студента	Максимальна кількість	Модуль 1		Модуль 2	
		кількість одиниць	максималь на	кількість одиниць	максималь на
Відвідування лекцій	1	5	5		
Відвідування семінарських занять	1	20	20	18	18

Відвідування практичних занять	1	4	4	4	4
Робота на семінарському занятті	10	20	200	18	180
Робота на практичному занятті	10	4	40	4	40
Лабораторна робота (зокрема допуск, виконання, захист)	-	-	-	-	-
Виконання завдань для самостійної роботи	5	4	20	4	20
Виконання модульної роботи	25	2	50	2	50
Виконання ІНДЗ	-	-	-	-	-
Разом		-	339	-	312
Максимальна кількість балів:	651				
Розрахунок коефіцієнта:	6,5				

6.2. Завдання для самостійної роботи та критерії її оцінювання

Опрацювання наукової літератури і письмові відповіді на запропоновані питання (5 балів за кожну опрацьовану тему)

Модуль 1.

Тема 1. Корпусна лінгвістика як галузь прикладного мовознавства.

Леміш Н. Є. Корпусна метарозмітка спеціальних текстів з лінгвоантропогенезу // Науковий часопис НПУ ім. М.П.Драгоманова. - Серія 9. Сучасні тенденції розвитку мов. – Вип. 15. – 2017. – С.139 - 150

Ковбасюк Л. А. Корпусна лінгвістика та германістика : теоретичні засади та перспективи // Філологічні науки. – 2017. – Книга 1. – С.9 – 13.

Тема 2. Розвиток лінгвістичних корпусів у світі.

Головащук С.Л. Історичні передумови розвитку корпусної лінгвістики. - https://www.academia.edu/38226753/%D0%86%D0%A1%D0%A2%D0%9E%D0%A0%D0%98%D0%A7%D0%9D%D0%86_%D0%9F%D0%95%D0%A0%D0%95%D0%94%D0%A3%D0%9C%D0%9E%D0%92%D0%98_%D0%A0%D0%9E%D0%97%D0%92%D0%98%D0%A2%D0%9A%D0%A3_%D0%9A%D0%9E%D0%A0%D0%9F%D0%A3%D0%A1%D0%9D%D0%9E%D0%87_%D0%9B%D0%86%D0%9D%D0%93%D0%92%D0%86%D0%A1%D0%A2%D0%98%D0%9A%D0%98

Жуковська В.В. Вступ до корпусної лінгвістики: навчальний посібник. // В. В. Жуковська – Житомир : Вид-во ЖДУ ім. І. Франка, 2013. – 140 с.

Тема 3. Методи вилучення інформації з корпусу.

Кочерган М.П. Загальне мовознавство. – К.: Академія, 1999. – С. 224-240, 246-255 (Структурний метод. Застосування математичних методів у мовознавстві)

Шведова М.О. Корпусні методи дослідження регіональних відмінностей у межах однієї мови (на матеріалах регіональних корпусів української та російської мов) // Вісник Харківського національного університету імені В. Н. Каразіна. - Серія : Філологія. - Вип. 77. – 2017 – С. 33-36

Тема 4. Огляд корпусних систем по слов'янських мовах

Карпов В.А. Язык как система. – М.: Едиториал УРСС, 2003. – С. 256-263 Або:
Щипицина Л.Ю. Информационные технологии в лингвистике : учеб. пособие / Л.Ю. Щипицина. — М. : ФЛИНТА : Наука, 2013. — С.57 - 65.
Демська-Кульчицька О. Національний корпус української мови: концептуальний аспект / О. Демська-Кульчицька // Лексикографічний бюлетень: Зб. наук. пр. — К.: Ін-т української мови НАН України, 2006. — Вип. 13. — С. 5-9 Або:
Балог В., Демська-Кульчицька О. Національний корпус української мови: підкорпус художньої літератури // Лексикографічний бюлетень. -
<http://dspace.nbuv.gov.ua/bitstream/handle/123456789/72811/02-Balog.pdf?sequence=1>

Модуль 2.

Тема 1. Сфери застосування лінгвістичних корпусів

Мейзерська І. В. Корпусний підхід у сучасній лінгвістиці : перспективи і можливості застосування// Науковий вісник кафедри Юнеско КНЛУ Серія Філологія - Педагогіка - Психологія. Випуск 28 - 2014. – С.53 – 58.

Тема 2. Лінгвістичні дослідження на базі корпусу: укладання словників.

Коссак О., Коруд О., Хвищун Л. Система комп'ютерної підтримки словникових баз даних СЛОВО // Проблеми українізації комп'ютерів. – К., 1993. – С. 73-76.
Клименко Н.Ф., Русанівський В.М. Від універсальної бази лінгвістичних знань до комп'ютерного укладання словників // Мовознавство. – 1995. – № 4-5. – С. 3-10.

Тема 3. Комп'ютерна лексикографія, предмет, завдання, напрями

Автоматичний словник // Українська мова: Енциклопедія. – К.: Українська енциклопедія, 2000, 1-е вид або 2004, 2-е вид. – С. 10.
Ардан Р.В., Бацевич Ф.С., Партико З.В. Комп'ютерний словник-мінімум української мови // Мовознавство. – 1996. – № 4-5.

Тема 4. Створення власного корпусу.

Провести збір і оцифрування для свого корпусу.

- Які принципи створення корпусних систем?
- Які існують кодування символів?
- Описати переваги того чи того типу кодування.
- Які способи та інструментарій створення відомих Вам корпусних систем?
Зробіть повідомлення по кожному із корпусів (історія і сучасний стан)

Критерії оцінювання виконання студентами завдань для самостійної роботи

Самостійна робота студентів з навчальної дисципліни «Корпусна лінгвістика» оцінюється за такими параметрами:

5 балів – завдання для самостійної роботи виконані в повному обсязі, лаконічно, змістовно, обґрунтовано, із наведенням прикладів і застосуванням методу аналізу та творчого підходу. Студент самостійно працює з науковою літературою; уміє формулювати й обґрунтовувати власну позицію; володіє правилами написання та підготовки виступів, доповідей, бесід; креативно підходить до їх підготовки; має комунікативні вміння і навички; прагне до самовдосконалення і саморозвитку.

4 бали – завдання для самостійної роботи виконані в повному обсязі, лаконічно, змістовно, обґрунтовано, але з певними помилками щодо викладу матеріалу та без наведення прикладів до окремих фактів і явищ. Студент працює із запропонованою науковою літературою; робить висновки; володіє правилами написання та підготовки виступів, доповідей, бесід і втілює їх у практику. Однак у процесі відтворення самостійно дібраного матеріалу простежується брак власних суджень, прояву креативності й ініціативності.

3 бали – завдання для самостійної роботи виконані в повному обсязі, але з порушенням логіки й послідовності викладу матеріалу та без ілюстрування прикладами. У роботі допускаються стилістичні помилки, неточне вживання термінів, довільне витлумачення фактів. Студент працює із запропонованою науковою літературою, але аналізує її примітивно, без висвітлення власної позиції; володіє правилами написання та підготовки виступів, доповідей, бесід, але втілює їх у практику допускаючи грубі помилки та уникаючи використання фахової термінології.

2 бали – завдання для самостійної роботи виконані не в повному обсязі, частково самостійно і за певним зразком. Студент володіє матеріалом на початковому рівні, викладає його хаотично і необґрунтовано, без дотримання мовних норм; має фрагментарні навички роботи з науковими джерелами і не вміє робити висновки; володіє правилами написання та підготовки виступів, доповідей, бесід, але не втілює їх у практику; комунікативні вміння і навички мають низький рівень розвитку.

1 бал – виконано менше половини завдань для самостійної роботи. Матеріал фрагментарний, не систематизований, не ілюстрований прикладами, порушена структура викладу. Його основу складає необґрунтоване копіювання чужих думок, фраз і позицій. Студент не вміє працювати з науковими джерелами і робити власні висновки; не володіє правилами написання та підготовки виступів, доповідей, бесід; відсутній творчий компонент даного виду діяльності; комунікативні вміння і навички мають дуже низький рівень розвитку.

0 балів – більшу частину самостійних завдань не виконано. Студент лише частково володіє навчальним матеріалом, не вміє чітко, лаконічно й послідовно висвітлювати його; не працює з науковими і навчальними джерелами і не володіє правилами написання та підготовки виступів, доповідей, бесід; майже відсутні творчі та комунікативні уміння і навички.

6.3. Форми проведення модульного контролю та критерії оцінювання

Модульний контроль проводиться в письмовій формі

Критерії оцінювання виконання студентами модульних контрольних робіт

21-25 балів – завдання модульної контрольної роботи виконані в повному обсязі, лаконічно, змістовно, обґрунтовано, із наведенням прикладів і застосуванням методу аналізу та творчого підходу. Студент самостійно працює оцінює комунікативні явища, професійні ситуації; робить висновки; вміє формулювати й обґрунтовувати власну позицію; володіє правилами написання та підготовки виступів, доповідей, бесід; креативно підходить до їх підготовки; має комунікативні вміння і навички.

15-20 балів – завдання модульної контрольної роботи виконані в повному обсязі, лаконічно, змістовно, обґрунтовано, але з певними помилками щодо викладу матеріалу та без наведення прикладів до окремих фактів і явищ. Студент робить висновки; володіє правилами написання та підготовки виступів, доповідей, бесід і втілює їх у практику. Однак у процесі відтворення самостійно дібраного матеріалу простежується брак власних суджень, прояву креативності й ініціативності.

9-14 балів – завдання модульної контрольної роботи виконані в повному обсязі, але з порушенням логіки й послідовності викладу матеріалу та без ілюстрування прикладами. У роботі допускаються стилістичні помилки, неточне вживання термінів, довільне витлумачення фактів. Студент не висвітлює власної позиції; володіє правилами написання та підготовки виступів, доповідей, бесід, але втілює їх у практику допускаючи грубі помилки та уникаючи використання фахової термінології.

3-8 балів – завдання модульної контрольної роботи виконані не в повному обсязі, частково самостійно і за певним зразком. Студент володіє матеріалом на початковому рівні, викладає його хаотично і необґрунтовано, без дотримання мовних норм; має фрагментарні навички роботи з науковими джерелами і не вміє робити висновки; володіє правилами написання та підготовки виступів, доповідей, бесід, але не втілює їх у практику; комунікативні вміння і навички мають низький рівень розвитку.

0-2 бали – виконано менше половини завдань для самостійної роботи. Матеріал фрагментарний, не систематизований, не ілюстрований прикладами, порушена структура викладу. Його основу складає необґрунтоване копіювання чужих думок, фраз і позицій. Студент не вміє робити власні висновки; не володіє правилами написання та підготовки виступів, доповідей, бесід; відсутній творчий компонент даного виду діяльності; комунікативні вміння і навички мають дуже низький рівень розвитку.

6.4. Форми проведення семестрового контролю та критерії оцінювання

Семестровий контроль із навчальної дисципліни «Корпусна лінгвістика» здійснюється у формі заліку.

6.5. Орієнтовний перелік питань для семестрового контролю

1. Корпус. База даних. Машинний фонд.
2. Історія лінгвістичних корпусів: від картотеки до корпусу.
3. Класифікація (типологія) корпусів.
4. Корпусна лінгвістика: сучасний стан.
5. Основні поняття корпусної лінгвістики.
6. Цілі і основні напрямки корпусної лінгвістики.
7. Історія створення лінгвістичних корпусів.
8. Основні характеристики корпусів.
9. Корпусні менеджери.
10. Класифікація корпусів за різними підставами.
11. Особливі типи корпусів.
12. Поняття розмітки. Розмітка. Засоби розмітки корпусів.
13. Стандартизація в корпусній лінгвістиці.
14. Огляд наявних корпусів різних типів.
15. Зарубіжні національні корпуси.
16. Спеціальні корпуси.

17. Корпусні дослідження.
18. Користувачі корпусів.
19. Формалізований опис мови.
20. Словниковий і безсловниковий спосіб.
21. Лексикографічні дослідження, що ґрунтуються на корпусах.
22. Історія корпусів української мови.

VII. Навчально-методична картка дисципліни «Корпусна лінгвістика»

Разом 120 годин: лекції – 10 годин, практичні заняття – 8 годин, семінарські заняття - 38 годин, самостійна робота – 56 годин, модульний контроль – 8 годин.

Тиждень	3 семестр																		
Модулі	ЗМІСТОВИЙ МОДУЛЬ І.										ЗМІСТОВИЙ МОДУЛЬ ІІ								
Назва модуля	Корпусна лінгвістика як галузь прикладного мовознавства. Корпусні студії: історична перспектива та сучасний стан. Корпус текстів природної мови										Технологія створення корпусів та комп'ютерні програми для роботи з корпусом. Можливості використання корпусів у лінгвістичних дослідженнях								
Кількість балів за модуль	339										213								
Загальна кількість балів	651/6,5																		
Лекції	2	2		2					2	2									
Дати																			
Теми практичних/ семінарських занять	Вступ. Співвідношення корпусної лінгвістики і комп'ютерної лінгвістики	Загальна характеристика лінгвістичних корпусів. Машинні фонди, бази даних, корпус. (2+ 20	Розвиток лінгвістичних корпусів у світі. (1 6+ 10 б).	Види корпусів : усні і писемні, одномовні і багатомовні. Анотовані і неанотовані корпуси.	Розмітка. Види розміток. Лінгвістична анотація і метадані. (2	Методи вилучення інформації з корпусу.	Створення формалізованих описів для розроблення корпусу. (1 6+ 10 б)	Анотування корпусу. Збереження, публікація і оновлення корпусів. (1 6+ 10 б)	Огляд корпусних систем по слов'янських мовах (1 6+ 10 б)	Актуальні проблеми корпусної лінгвістики української мови (1 6+10 б).	Етапи створення корпусу (2+20 балів)	Формат даних і стандартизація даних корпусу (1 6+ 10 б).	Комп'ютерні програми для аналізу даних корпусу (1+10 балів)	Сфери застосування лінгвістичних корпусів (1 6+ 10 б)	Лінгвістичні дослідження на базі корпусу: вивчення лексики (1+10 балів)	Лінгвістичні дослідження на базі корпусу: укладання словників. (2+20 балів)	Лінгвістичні дослідження на базі корпусу: вивчення граматики (1 6+ 10 б)	Створення свого корпусу. Планування. Збір і оцифрування даних. (1 6+ 10 б)	Захист проєктів. (1+10 балів)
Самостійна робота		5 балів	5 балів	5 балів	5 балів						5 балів	5 балів		5 балів		5 балів			
Види поточного контролю	Модульна контрольна робота (25 балів)					Модульна контрольна робота (25 балів)					Модульна контрольна робота (25 балів)			Модульна контрольна робота (25 балів)					
Підсумковий контроль	Залік																		

VIII. Рекомендована література

1. Бобкова Т. В. До визначення корпусної лінгвістики в сучасному мовознавстві. // Наукові записки Національного університету "Острозька академія. Серія "Філологічна". Випуск 45. – Острог, 2014. – С. 3–6.
 2. Волошин В.Г. Комп'ютерна лінгвістика: Навчальний посібник. – Суми: ВТД "Університетська книга", 2004. – 382 с.
 3. Гвишиани Н.Б. Практикум по корпусной лингвистике: Учеб. пособие по английскому языку. – М.: Издательство Высшая школа, 2008.
 4. Жуковська В.В. Вступ до корпусної лінгвістики. Навчальний посібник. — Житомир: Вид-во ЖДУ ім. І. Франка, 2013. — 142 с.
 5. Карпіловська Є.А. Вступ до комп'ютерної лінгвістики. – Донецьк: ТОВ "Юго-Восток, Лтд", 2006. – 184 с.
 6. Карпіловська Є.А. Машинні версії традиційних словників як основа для укладання комп'ютерних словників і тезаурусів // Мовознавство. – 1996. – № 4-5.
 7. Лосев А.Ф. Введение в общую теорию языковых моделей. – Изд. 2-ое, М.: Едиториал УРСС, 2004.– 296 с.
 8. Рыков В.В. Курс лекций по корпусной лингвистике. [Электронный ресурс]. – Режим доступа: <http://rykov-cl.narod.ru/c.html>
- 2) додаткова:
1. Доклады научной конференции «Корпусная лингвистика и лингвистические базы данных» / Под ред. А.С. Герда. - СПб., 2002.
 2. Захаров В.П. Корпусная лингвистика. Учебно-метод. пособие. – СПб.: Издательство СПбГУ, 2006.
 3. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике. – М.: Академия, 2004. – 206 с.
 4. Караулов Ю.Н. Лингвистическое конструирование й тезаурус литературного языка. – М., 1981. – 307 с.
 5. Коваль С.А. Роль корпуса в создании реалистичных моделей словоизменительной морфологии [Электронный ресурс]. – Режим доступа: http://skowal.narod.ru/research/corpora2006/Koval_Corpora.2006.htm
 6. Плунгян В.А. Почему современная лингвистика должна быть лингвистикой корпусов? (Публичная лекция, прочитанная 01.10.2009). [Электронный ресурс]. – Режим доступа: <http://www.polit.ru/lectures/2009/10/23/corpus.html>.
 7. Шаров С.А. Параметры описания текстов корпуса. [Электронный ресурс]. – Режим доступа: <http://bokrcorpora.narod.ru/header.html>.
 8. Шаров С.А. Формат выходного представления корпуса текстов. [Электронный ресурс]. – Режим доступа: <http://bokrcorpora.narod.ru/format.html>.
 9. Широков В.А., Рабулець О.Г., Шевченко І.В., Костишин О.М., Якименко К.М. Технологічні основи сучасної тлумачної лексикографії // Мовознавство. – 2002. – № 6. – С. 49-86.
 10. Aarts Jan. Комментарий к статье «A New Corpus of English» (Sidney Greenbaum) // Directions in Corpus Linguistics. Berlin, 1992.
 16. Baker P., Hardie A., McEnery T. A Glossary of Corpus Linguistics. – Edinburgh: Edinburgh University Press, 2006.
 11. Biber D., Conrad S., Reppen R. Corpus Linguistics: Investigating Language Structure and Use. – Cambridge University Press, 1998. [Электронный ресурс]. – Режим доступа: URL: <http://books.google.com/books?id=2h5F7TXa6psC>

12. Bookmarks for corpus-based linguistics. [Електронний ресурс]. – Режим доступа: URL: <http://www.devoted.to/corpora>
13. Corpora in cognitive linguistics : corpus-based approaches to syntax and lexis / Ed. by S.Th. Gries, A. Stefanowitsch. – Berlin et al: Mouton de Gruyter, 2006.
14. Corpus Linguistics Beyond the Word: Corpus Research from Phrase to Discourse / Ed. by E. Fitzpatrick. – Amsterdam, New York: Rodopi, 2007.
15. Corpus Linguistics: An International Handbook / Ed. by A. Lüdeling, M. Kytö. – Vols. 1-2. – Berlin et al: Walter de Gruyter, 2008.
16. Developing linguistic corpora: a guide to good practice / Ed. by M. Wynne. [Електронний ресурс]. – Режим доступа: URL: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/index.htm>
17. Kennedy G. An Introduction to Corpus Linguistics. – London: Longman, 1998.
18. Leech G. Corpus Annotation Schemes // Literary and Linguistics Computing. 2003. Vol. 8. N. 4. P.275–281
19. McEnery T., Wilson A. Corpus Linguistics. – Edinburgh University Press, 2001. [Електронний ресурс]. – Режим доступа: URL: http://books.google.com/books?id=nwmgdvN_akAC
20. Sharoff S. Towards Basic Categories for Describing Properties of Texts in a Corpus. In Proc. of Language Resources and Evaluation Conference (LREC04). May, 2004, Lisbon, Portugal //
- <http://www.comp.leeds.ac.uk/ssharoff/texts/lrec-04.pdf>
21. Sinclair J.M. Corpus, Concordance, Collocation. – Oxford: Oxford University Press, 1991.

IX. Інформаційні ресурси

1. Британський Національний корпус: <http://www.natcorp.ox.ac.uk/>
2. Корпус сучасної американської англійської мови: <http://www.americancorpus.org/>
3. Програма фундаментальних досліджень Президиума РАН «Корпусная лингвистика» <http://www.corplingran.ru/links.html>
4. XML Corpus Encoding Standard Document <http://www.xces.org/>
5. Professor Tony McEnery (Lancaster University) delivers the annual John Sinclair Lecture at the University of Birmingham's English department on Thursday 9 May 2013. "Primed for Violence? A corpus analysis of jihadist discourse" <https://www.youtube.com/watch?v=dj5SXGgT11U>
6. Виктор Бочаров «Корпусная лингвистика» - лекция: <https://vimeo.com/27140749>
7. <http://linguistlist.org/sp/GetWRLListings.cfm?WRAbbrev=Software#wr97> - програмний каталог.
8. <http://linguistlist.org/sp/get-software.cfm?submissionid=12195716> – PRESEMT програма (безкоштовна) для перекладачів.
9. <http://www.fon.hum.uva.nl/praat/> - програма-аналізатор фонетичної складової мовлення..
10. <http://www.speech.kth.se/snack/> - аналізатор звукових волн.
11. <http://wordnet.princeton.edu/> - лексична база знань англійської мови.
12. <http://www.lx.ugent.be/toolbox/> - зовнішній каталог програм для лінгвістичних досліджень.
13. http://www.bmanuel.org/clar2_tt.html - каталог посилань з корпусної лінгвістики.
14. <http://www.langsoft.ch/> - програмне забезпечення для обробки й розбору природних мов, перевірка орфографії, машинний переклад, тезауруси, атрибуції текста для англійської, німецької, французької, італійської мов.
15. <http://similis.org/linguaetmachina.www/index.php> - similis повнофункціональний інструмент машинного перекладу для менеджерів проектів і перекладачів. Similis аналізує попередні переклади, генерує пам'ять перекладів(TM) і застосовує його до всіх нових проектів.