

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:**

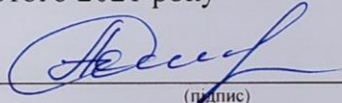
Дроботенко Валентина Юріївна, кандидат філологічних наук, доцент.  
Горохова Тетяна Олександрівна, кандидат педагогічних наук, старший викладач кафедри української мови.

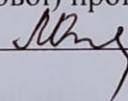
**Викладач:**

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

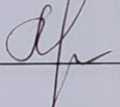
Робочу програму розглянуто і затверджено на засіданні кафедри української мови

Протокол № 8 від 3 лютого 2021 року

Завідувач кафедри  (С.О. Караман)  
(підпис) (ініціали, прізвище)

Гарант освітньої (професійної/наукової) програми  
(керівник проектної групи)  (М. О. Вінтонів)  
(ініціали, прізвище)

Робочу програму перевірено  
08.02.2021 р.

Заступник директора/декана  ( А. І. Золін )  
Пролонговано:

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
<b>Лінгвістика комп'ютерного аналізу і обробки інформації</b>	
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-конкорданси 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
<b>Змістовий модуль 1.</b>							
<b>Корпусна лінгвістика як галузь прикладного мовознавства</b>							
Тема 1. Корпусна лінгвістика як галузь прикладного мовознавства.	4	2		2			
Тема 2. Загальна характеристика лінгвістичних корпусів. Машинні фонди, бази даних, корпус.	4	2		2			
Тема 3. Розвиток лінгвістичних корпусів у світі.	11	2		2		7	
Тема 4. Види корпусів : усні і писемні, одномовні і багатомовні. Анотовані і неанотовані корпуси.	9			2		7	
Тема 5. Розмітка. Види розміток. Лінгвістична анотація і метадані.	10	2	4	2			2
Тема 6. Методи вилучення інформації з корпусу.	2			2			
Тема 7. Створення формалізованих описів для розроблення корпусу	9			2		7	
Тема 8. Огляд корпусних систем української мови	6	2		4			
Тема 9. Актуальні проблеми корпусної лінгвістики української мови (слов'янських мов)	12			2		8	2
Всього за 1 модуль	67	10	4	20		29	4

<b>Змістовий модуль 2.</b>							
<b>Технологія створення корпусів.</b>							
<b>Лінгвістичний та лінгводидактичний потенціал корпусних систем</b>							
Тема 10. Етапи створення корпусу. Формат даних і стандартизація даних корпусу	4			4			

Тема 11. Комп'ютерні програми для аналізу даних корпусу. Візуалізація корпусних даних	15		4	2		9	
Тема 12. Сфери застосування лінгвістичних корпусів	2			2			
Тема 13. Лінгводидактичний потенціал корпусних систем	15			4		9	2
Тема 14. Корпусні дослідження в сучасній лінгвістиці	13			4		9	
Тема 6. Захист проєктів.	4			2			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.

**Тема 1.** Вступ. Корпусна лінгвістика як галузь прикладного мовознавства

**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:** 1–4.

**Додаткова:** 1, 2, 5, 6.

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

**Основні питання теми:**

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

**Рекомендована література:**

**Основна:** 1–4.

**Додаткова:** 3, 4, 5, 6, 20.

**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 ХХ ст. до пост 2000-х.

**Рекомендована література:**

**Основна:** 1–4.

**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:** 1–4.

**Додаткова:** 1, 2, 3, 4, 5, 6.

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

**Основні питання теми:**

- Метарозмітка корпусів текстів: стандарти і реалізація
- Типи корпусної розмітки.
  - \* Структурне маркування
  - \* Лінгвістична анотація.
- Вимоги до розмітки.

**Рекомендована література:**

**Основна:** 1–4.

**Додаткова:** 1, 2, 6.

**Тема 6.** Методи вилучення інформації з корпусу.

**Основні питання теми:**

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

**Рекомендована література:**

**Основна:** 1–4.

**Додаткова:** 1, 3, 4, 5.

**Тема 7.** Створення формалізованих описів для розроблення корпусу. 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 література:**

**Основна:** 1–4.

**Додаткова:** 1, 2, 3, 6, 17, 18.



## **Тема 8.** 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:** 1–4.

**Додаткова:** 1, 2, 3, 4, 5, 6, 10, 13, 15, 17.

## **Тема 9.** 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:** 1–4.

**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 10.** 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).

Міжнародні стандарти в галузі корпусної лінгвістики.

- Стандарти кодування корпусів.
- Стандарти оцінки корпусів.

Огляд міжнародних стандартів корпусної лінгвістики.

Стандарти TEI. Структура тексту TEI.

### **Рекомендована література:**

**Основна:** 5, 7, 8, 9

**Додаткова:** 1, 2, 3, 4, 5, 6.

## **Тема 11.** Комп'ютерні програми для аналізу даних корпусу. Візуалізація корпусних даних

### **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 (Авторматичний словник Мультигран, Google Перекладач).

Словники і тезауруси.

Візуалізація корпусних даних. Створення графіків.

***Рекомендована література:***

***Основна:*** 5, 7, 8, 9

***Додаткова:*** 1, 2, 3, 4, 5, 6, 9, 11.

**Тема 12.** 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 13.** Лінгводидактичний потенціал корпусних систем

***Основні питання теми:***

Корпусний підхід у навчанні української мови в закладах загальної середньої освіти.

Корпусний підхід у навчанні української мови в закладах вищої освіти.

Корпуси для навчання української мови як іноземної.

***Рекомендована література:***

***Основна:*** 5, 7, 8, 9

***Додаткова:*** 1, 2, 3, 4, 5, 6.

**Тема 14.** Корпусні дослідження в сучасній лінгвістиці.

***Основні питання теми:***

Корпусні дослідження в галузі лексикології.

Корпусні дослідження в галузі граматики.

Корпусні дослідження семантики.

Корпусні дослідження прагматики і дискурсивний аналіз.

Корпусні дослідження в галузі соціолінгвістики.

Корпусні дослідження в галузі стилістики та лінгвістичного аналізу текстів.

Корпусні дослідження в галузі діалектології.

Корпусні системи і лексикографія.

***Рекомендована література:***

***Основна:*** 5, 7, 8, 9

*Додаткова: 1, 2, 3, 4, 5, 6.*

**Тема 15.** Створення свого корпусу. Планування. Збір і оцифрування даних. Кодування тексту. Захист проєктів.

**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, 15,17.

## **6. Контроль навчальних досягнень**

### **6.1. Система оцінювання навчальних досягнень студентів**

Вид діяльності студента	Максимальна кількість	Модуль 1		Модуль 2	
		кількість одиниць	максимальна	кількість одиниць	максимальна
Відвідування лекцій	1	5	5		
Відвідування семінарських занять	1	10	10	9	9
Відвідування практичних занять	1	2	2	2	2
Робота на семінарському занятті	10	10	100	9	90
Робота на практичному занятті	10	2	20	2	20
Лабораторна робота (зокрема допуск, виконання, захист)	-	-	-	-	-
Виконання завдань для самостійної роботи	5	4	20	3	15
Виконання модульної роботи	25	2	50	2	50
Виконання ІНДЗ	-	-	-	-	-
Разом		-	207	-	186
Максимальна кількість балів:			393		
Розрахунок коефіцієнта:			3,93		

### **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.

Головашук 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](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 5. 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 6. Лінгвістичні дослідження на базі корпусу: укладання словників.

Коссаk O., Коруд O., 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 7. 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 балів** – 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 необгрунтоване 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, обгрунтовано, із наведенням прикладів і застосуванням методу аналізу та творчого підходу. Студент самостійно працює оцінює комунікативні явища, професійні ситуації; робить висновки; вміє формулювати й обгрунтовувати 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, обгрунтовано, 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 необгрунтовано, 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 необгрунтоване 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.
23. Корпусні дослідження в галузі лексикології.
24. Корпусні дослідження в галузі граматики.
25. Корпусні дослідження семантики.
26. Корпусні дослідження прагматики і дискурсивний аналіз.
27. Корпусні дослідження в галузі соціолінгвістики.
28. Корпусні дослідження в галузі стилістики та лінгвістичного аналізу текстів.
29. Корпусні дослідження в галузі діалектології.
30. Корпусні системи і лексикографія.
31. Корпусний підхід у лінгводидактиці.

**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 I.							ЗМІСТОВИЙ МОДУЛЬ II								
Назва модуля	<b>Корпусна лінгвістика як галузь прикладного мовознавства.</b>							<b>Технологія створення корпусів. Лінгвістичний та лінгводидактичний потенціал корпусних систем</b>								
Кількість балів за модуль	207							186								
Загальна кількість балів	393/3,93 = 100															
Лекції	2	2	2		2			2								
Дати																
Теми семінарських занять	Корпусна лінгвістика як галузь прикладного мовознавства (1 б+ 10 б)	Загальна характеристика лінгвістичних корпусів. Машинні фонди, бази даних, корпус. (1 б+ 10 б)	Розвиток лінгвістичних корпусів у світі. (1 б+ 10 б).	Види корпусів : усні і писемні, одномовні і багатомовні. Анотовані і неанотовані корпуси	Розмітка. Види розміток. Лінгвістична анотація і метадані. (2 б+ 20 б)	Методи вилучення інформації з корпусу. (1 б+ 10 б).	Створення формалізованих описів для розроблення корпусів. (1 б+ 10 б)	Огляд корпусних систем української мови (2 б+ 20 б)	Актуальні проблеми корпусної лінгвістики української мови (1 б+10 б).	Етапи створення корпусу. Формат даних і стандартизація даних корпусу	Формат даних і стандартизація даних	Комп'ютерні програми для аналізу даних корпусу (1+10 балів)	Сфери застосування лінгвістичних корпусів (1 + 10 балів)	Лінгводидактичний потенціал корпусних систем (2+20 балів)	Корпусні дослідження в сучасній лінгвістиці (2+20 балів)	Захист проєктів. (1+10 балів)
Теми практичних занять	Розмітка. Види розміток. Лінгвістична анотація і метадані (2 б+ 20 б)							Комп'ютерні програми для аналізу корпусу. Візуалізація корпусних даних (2 б+ 20 б)								
Самостійна робота			5 балів	5 балів			5 балів		5 балів			5 балів		5 балів		
Види поточного контролю	Модульна контрольна робота 1 (25 балів)				Модульна контрольна робота 2 (25 балів)				Модульна контрольна робота 3 (25 балів)				Модульна контрольна робота 4 (25 балів)			
Підсумковий контроль	Залік															



### VIII. Рекомендована література

Основна:

1. Волошин В.Г. Комп'ютерна лінгвістика: Навчальний посібник. – Суми: ВТД "Університетська книга", 2004. – 382 с.
2. Жуковська В.В. Вступ до корпусної лінгвістики. Навчальний посібник. — Житомир: Вид-во ЖДУ ім. І. Франка, 2013. — 142 с.
3. Карпіловська Є.А. Вступ до комп'ютерної лінгвістики. – Донецьк: ТОВ "Юго-Восток, Лтд", 2006. – 184 с.
4. Широков В. Лінгвістично-інформаційні студії. У 5 томах. Том 4. Корпусна та когнітивна лінгвістика / Український мовно-інформаційний фонд НАН України, 2018, 246 с.

2) додаткова:

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. Гвишиани Н.Б. Практикум по корпусной лингвистике: Учеб. пособие по английскому языку. – М.: Издательство Высшая школа, 2008.
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.
3. Карпіловська Є.А. Машинні версії традиційних словників як основа для укладання комп'ютерних словників і тезаурусів // Мовознавство. – 1996. – № 4-5.
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](http://skowal.narod.ru/research/corpora2006/Koval_Corpora.2006.htm)
4. Лосев А.Ф. Введение в общую теорию языковых моделей. – Изд. 2-ое, М.: Едиториал УРСС, 2004.– 296 с.
5. Рыков В.В. Курс лекций по корпусной лингвистике. [Электронный ресурс]. – Режим доступа: <http://rykov-cl.narod.ru/c.html>
6. Фисун Р. С. Составные неопределенные местоимения в украинском языке. Науковий часопис Національного педагогічного університету імені М. П. Драгоманова. Серія 10 : Проблеми граматики і лексикології української мови : збірник наукових праць / Відп. редактор М. Я. Плющ. Київ : НПУ імені М. П. Драгоманова, 2019. Вип. 14. С. 117-126
7. Шаров С.А. Параметры описания текстов корпуса. [Электронный ресурс]. – Режим доступа: <http://bokrcorpora.narod.ru/header.html>.
8. Шведова М. Генеральний регіонально анотований корпус української мови (ГРАК) як інструмент дослідження лексико-граматичної варіативності. Людина. Комп'ютер. Комунікація: Збірник наукових праць / за ред. проф. О. П. Левченко. Видавництво Львівської політехніки, 2019. С. 145-148
9. Шведова М. Динаміка синтаксичних конструкцій на позначення шляху при дієсловах руху: корпусне дослідження. Українська мова, 2018. №3. С. 67-79
10. Шведова М. Становлення присвійного займенника третьої особи множини в українській літературній мові. Мовознавство. 2018, №4. С. 40-53

11. 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.
12. 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.
13. Anisimov A., Marchenko O., Taranukha V., Vozniuk T. Semantic and Syntactic Model of Natural Language Based on Non-negative Matrix and Tensor Factorization. // PolTal-2014, Lecture Notes in Artificial Intelligence. – 8686. – Springer. – PP. 177-184.
14. 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>
15. Bookmarks for corpus-based linguistics. [Электронный ресурс]. – Режим доступа: URL: <http://www.devoted.to/corpora>
16. 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.
17. Corpus Linguistics Beyond the Word: Corpus Research from Phrase to Discourse / Ed. by E. Fitzpartick. – Amsterdam, New York: Rodopi, 2007.
18. Corpus Linguistics: An International Handbook / Ed. by A. Lüdeling, M. Kytö. – Vols. 1-2. – Berlin et al: Walter de Gruyter, 2008.
19. Developing linguistic corpora: a guide to good practice / Ed. by M. Wynne. [Электронный ресурс]. – Режим доступа: URL: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/index.htm>
20. Kennedy G. An Introduction to Corpus Linguistics. – London: Longman, 1998.
21. Leech G. Corpus Annotation Schemes // Literary and Linguistics Computing. 2003. Vol. 8. N. 4. P.275–281
22. Levchenko O., Tyshchenko O. and Dilai M. Modeling of the Associative Verbal Network and its Cognitive Interpretation (the Conceptual Domain ЗРАДА ‘BETRAYAL’). 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT), 2020, pp. 304-307
23. McEnery T., Wilson A. Corpus Linguistics. – Edinburgh University Press, 2001. [Электронный ресурс]. – Режим доступа: URL: [http://books.google.com/books?id=nwmgdvN\\_akAC](http://books.google.com/books?id=nwmgdvN_akAC)
24. 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>
25. Shvedova M., R. von Waldenfels. Regional Annotation within GRAC, a Large Reference Corpus of Ukrainian: Issues and Challenges. CEUR Workshop Proceedings. Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2021). Volume I: Main Conference. Kharkiv, Ukraine, April 22-23, 2021. P. 32-45
26. Shvedova M.. The General Regionally Annotated Corpus of Ukrainian (GRAC, uacorpora.org): Architecture and Functionality. CEUR Workshop Proceedings. Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020). Volume I: Main Conference. Lviv, Ukraine, April 23-24, 2020. P. 489-506
27. Sinclair J.M. Corpus, Concordance, Collocation. – Oxford: Oxford University Press, 1991.
28. Starko V.. Implementing Semantic Annotation in a Ukrainian Corpus. CEUR Workshop Proceedings. Proceedings of the 5th International Conference on Computational

- Linguistics and Intelligent Systems (COLINS 2021). Volume I: Main Conference. Kharkiv, Ukraine, April 22-23, 2021. P. 435-447
29. Starko V. Semantic Annotation for Ukrainian: Categorization Scheme, Principles, and Tools. CEUR Workshop Proceedings. Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020). Volume I: Main Conference. Lviv, Ukraine, April 23-24, 2020

## **ІХ. Інформаційні ресурси**

1. ЕНК дисципліни «Лінгвістика комп'ютерного аналізу і обробки інформації»: <https://elearning.kubg.edu.ua/course/view.php?id=22184&section=2>
2. Lang-uk. Корпуси українських текстів: <https://lang.org.ua/uk/corpora/>
3. UA-GEC: корпус текстів з розміченими граматичними помилками: <https://ua-gec-dataset.grammarly.ai/>
4. Браунський український корпус: <https://github.com/brown-uk/corpus>
5. Великий електронний словник української мови (ВЕСУМ): <https://r2u.org.ua/vesum/>
6. Генеральний регіонально анотований корпус української мови: <http://uacorpora.org/>
7. Інструмент аналізу української мови на основі словника ВЕСУМ та двигуна LanguageTool. Має підтримку токенізації, лематизації, частиномовного аналізу та базового зняття омонімії: [https://github.com/brown-uk/nlp\\_uk](https://github.com/brown-uk/nlp_uk)
8. Корпус текстів української мови: <http://www.mova.info/corpus.aspx?11=209>
9. Корпус української мови бібліотеки «Чтиво»: <http://korpus.org.ua/>
10. Корпуси східнослов'янських мов: <http://uacorpora.org/Kyiv/ua/other-ukrainian-corpora>
11. Лабораторія Української. Веб-корпус із синтаксичною розміткою: <https://mova.institute/>
12. Паралельні українсько-російський і російсько-український корпуси у складі Національного корпусу російської мови: <https://ruscorporu.ru/new/search-para-uk.html>
13. Стенфордська бібліотека для опрацювання мови: <https://stanfordnlp.github.io/stanza/>
14. Український веб-корпус Лейпцизького університету: [https://corpora.uni-leipzig.de/en?corpusId=ukr-ua\\_web\\_2019&fbclid=IwAR16-gXBSiSsSOMTDTr8vpDThRdYj4-NSZAmP6YIKHMvVueNUZ79sN2c4VQ](https://corpora.uni-leipzig.de/en?corpusId=ukr-ua_web_2019&fbclid=IwAR16-gXBSiSsSOMTDTr8vpDThRdYj4-NSZAmP6YIKHMvVueNUZ79sN2c4VQ)
15. Центр культури української мови: <https://centrkum.blogspot.com/>