

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ВОЛИНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ЛЕСІ УКРАЇНКИ
Факультет іноземної філології
Кафедра прикладної лінгвістики**

СИЛАБУС

нормативного освітнього компонента

КОМП'ЮТЕРНА ЛІНГВІСТИКА

підготовки здобувачів освіти першого (бакалаврського) рівня

галузі знань 03 Гуманітарні науки

спеціальності 035 Філологія

спеціалізації 035.10 Прикладна лінгвістика

освітньо-професійної програми Прикладна лінгвістика. Переклад і комп'ютерна лінгвістика

Силабус освітнього компонента КОМП'ЮТЕРНА ЛІНГВІСТИКА
підготовки здобувачів освіти першого (бакалаврського) рівня галузі знань 03
Гуманітарні науки спеціальності 035 Філологія спеціалізації 035.10
Прикладна лінгвістика освітньо-професійної програми ПРИКЛАДНА
ЛІНГВІСТИКА. ПЕРЕКЛАД І КОМП'ЮТЕРНА ЛІНГВІСТИКА
денної/заочної форм навчання.

Розробник: кандидат філологічних наук, доцент кафедри прикладної
лінгвістики Старко В. Ф.

Погоджено

Гарант освітньо-професійної програми:  (Калиновська І. М.)

**Силабус освітнього компонента затверджено на засіданні кафедри
прикладної лінгвістики, протокол № 1 від 30.08.2024 р.**

**Силабус освітнього компонента перезатверджено на засіданні кафедри
прикладної лінгвістики**

протокол № 1 від 29.08. 2025 р.

В.о. завідувача кафедри:



І.М. Калиновська

I. Опис освітнього компонента

Таблиця 1.1 (Денна форма)

Найменування показників	Галузь знань, спеціальність, освітня програма, освітній рівень	Характеристика освітнього компонента
Денна форма здобуття освіти	03 Гуманітарні науки 035 Філологія Прикладна лінгвістика. Переклад і комп'ютерна лінгвістика. Бакалавр	Нормативний
Кількість годин/кредитів 210 год. / 7 кредитів		Рік навчання: 3-й
		Семестри: 5, 6-й
		Лекції: 44 год.
		Практичні (семінарські): 66 год. Консультації: 12 год.
ІНДЗ: немає		Самостійна робота 88 год.
		Форма контролю: екзамен
Мова навчання		англійська

Таблиця 1.2 (Заочна форма)

Найменування показників	Галузь знань, спеціальність, освітня програма, освітній рівень	Характеристика освітнього компонента
Заочна форма здобуття освіти	03 Гуманітарні науки 035 Філологія Прикладна лінгвістика. Переклад і комп'ютерна лінгвістика. Бакалавр	Нормативний
Кількість годин/кредитів 210 год. / 7 кредитів		Рік навчання: 3-й
		Семестри: 5, 6-й
		Лекції: 14 год.
		Практичні (семінарські): 22 год. Консультації: 26 год.
ІНДЗ: немає		Самостійна робота: 148 год.
	Форма контролю: екзамен	
Мова навчання		англійська

II. Інформація про викладачів

Прізвище, ім'я та по батькові	Старко Василь Феодосійович
Науковий ступінь	кандидат філологічних наук
Вчене звання	доцент
Посада	доцент кафедри прикладної лінгвістики
Контактна інформація	м.т.: +380977519850
	@vnu.edu.ua
Дні занять	http://194.44.187.20/cgi-bin/timetable.cgi
	Заняття за розкладом. Консультації в день проведення лекцій/ семінарських занять /перед іспитом (за попередньою домовленістю)

III. Опис освітнього компонента

1. Анотація. ОК Комп'ютерна лінгвістика призначений для здобувачів освіти третього року денної та заочної форм навчання ОПП Прикладна лінгвістика. Переклад і комп'ютерна лінгвістика першого (бакалаврського) рівня вищої освіти факультету іноземної філології. Вивчення ОК передбачає ознайомлення здобувачів освіти з основними поняттями, методами та сучасними тенденціями комп'ютерної лінгвістики.

2. Пререквізити й постреквізити. Пререквізити ОК Комп'ютерна лінгвістика: ОК Вступ до фаху, ОК Англійська мова, ОК Математичне моделювання, ОК Математична логіка, ОК Прикладна лінгвістика, ОК Теоретичний курс англійської мови, ОК Інформаційно-комунікаційні технології, ОК Машинний та автоматизований переклад. До постреквізитів належать: ОК Курсова робота з ІТ та комп'ютерної лінгвістики, ОК Теоретична та прикладна лінгвістика, ОК Мовленнєві технології і мовні моделі, ОК Штучний інтелект та прикладні інформаційні технології.

3. Мета і завдання. Метою викладання ОК Комп'ютерна лінгвістика є формування у здобувачів освіти теоретичних знань про базові поняття, методи і технології комп'ютерної лінгвістики та практичних навичок використання корпусних технологій і комп'ютерних інструментів для аналізу мовних даних.

Основними завданнями курсу є: 1) сформувати у здобувачів освіти систему теоретичних знань про базові поняття і термінологію комп'ютерної лінгвістики; 2) навчити використовувати методи і технології формального опису та дослідження мовних явищ; 3) ознайомити з актуальними напрямками розвитку комп'ютерної лінгвістики; 4) розвинути практичні навички роботи з корпусами та інструментами обробки мови; 5) сформувати практичні навички роботи з інструментами корпусного і комп'ютерного аналізу.

Методи навчання: викладання курсу ґрунтується на поєднанні традиційних та інноваційних методів і технологій навчання. Традиційні методи охоплюють пояснювально-ілюстративний метод, метод проблемного викладу, частково-пошуковий та дослідницький методи. Вивчення ОК також здійснюється з залученням методу кейсів (case study), ділової гри та командо-орієнтованого навчання, сукупність яких спрямована на пошук оптимальних рішень під час обговорення питань та розв'язання завдань, що потребують як індивідуальних, так і групових алгоритмів дій. В основі дистанційних методів навчання лежать сучасні веб-технології і мультимедійні засоби, які впроваджуються у всіх можливих формах онлайн роботи.

4. Результати навчання (компетентності). По завершенню вивчення ОК Комп'ютерна лінгвістика здобувачі освіти володітимуть такими компетентностями:

Інтегральна компетентність

Здатність розв'язувати складні спеціалізовані задачі та практичні проблеми в галузі філології (лінгвістики, літературознавства, фольклористики, перекладу) в процесі професійної діяльності або навчання, що передбачає застосування теорій та методів філологічної науки і характеризується комплексністю та невизначеністю умов.

Загальними компетентностями:

ЗК 2. Здатність зберігати та примножувати моральні, культурні, наукові цінності й досягнення суспільства на основі розуміння історії та закономірностей розвитку предметної області, її місця у загальній системі знань про природу і суспільство та у розвитку суспільства, техніки і технологій, використовувати різні види та форми рухової активності для активного відпочинку та ведення здорового способу життя.

ЗК 3. Здатність спілкуватися державною мовою як усно, так і письмово.

ЗК 4. Здатність бути критичним і самокритичним.

ЗК 5. Здатність учитися й оволодівати сучасними знаннями.

ЗК 6. Здатність до пошуку, опрацювання та аналізу інформації з різних джерел.

ЗК 7. Уміння виявляти, ставити та вирішувати проблеми.

ЗК 8. Здатність працювати в команді та автономно.

- ЗК 9.** Здатність спілкуватися іноземною мовою.
ЗК 10. Здатність до абстрактного мислення, аналізу та синтезу.
ЗК 11. Здатність застосовувати знання у практичних ситуаціях.
ЗК 12. Навички використання інформаційних і комунікаційних технологій.
ЗК 13. Здатність проведення досліджень на належному рівні.

Фаховими компетентностями:

- ФК 1.** Усвідомлення структури філологічної науки та її теоретичних основ.
ФК 2. Здатність використовувати в професійній діяльності знання про мову як особливу знакову систему, її природу, функції, рівні.
ФК 3. Здатність використовувати в професійній діяльності знання з теорії та історії мов, що вивчаються.
ФК 6. Здатність вільно, гнучко й ефективно використовувати мови, що вивчаються, в усній та письмовій формі, у різних жанрово-стильових різновидах і регістрах спілкування (офіційному, неофіційному, нейтральному), для розв'язання комунікативних завдань у різних сферах життя.
ФК 7. Здатність до збирання й аналізу, систематизації та інтерпретації мовних літературних, фольклорних фактів, інтерпретації та перекладу тексту (залежно від обраної спеціалізації).
ФК 8. Здатність вільно оперувати спеціальною термінологією для розв'язання професійних завдань.
ФК 9. Усвідомлення засад і технологій створення текстів різних жанрів і стилів державною та іноземними мовами.
ФК 13. Здатність застосовувати лексико-семантичні, граматичні й стилістичні трансформації під час письмового та усного перекладу текстів різних жанрів і функціональних стилів для забезпечення синтаксичної, семантичної та прагматичної еквівалентності вихідного та цільового текстів.
ФК 15. Здатність використовувати сучасні інформаційні системи та технології під час виконання функціональних завдань та обов'язків, знати основи безпечної роботи в інформаційних системах, методи створення баз даних та вебресурсів.
ФК 17. Здатність використовувати базові знання розділів математики та логіки у завданнях комп'ютерної лінгвістики та розробці програмного забезпечення.

ОК формує такі програмні результати навчання:

- ПРН 1.** Вільно спілкуватися з професійних питань із фахівцями та нефахівцями державною та іноземними мовами усно й письмово, використовувати їх для організації ефективної міжкультурної комунікації.
ПРН 2. Ефективно працювати з інформацією: добирати необхідну інформацію з різних джерел, зокрема з фахової літератури та електронних баз, критично аналізувати й інтерпретувати її, впорядковувати, класифікувати й систематизувати.
ПРН 3. Організовувати процес свого навчання й самоосвіти.
ПРН 6. Використовувати інформаційні й комунікаційні технології для вирішення складних спеціалізованих задач і проблем професійної діяльності.
ПРН 7. Розуміти основні проблеми філології та підходи до їх розв'язання із застосуванням доцільних методів та інноваційних підходів.
ПРН 8. Знати й розуміти систему мови, загальні властивості літератури як мистецтва слова, історію мови і літератури, що вивчаються, і вміти застосовувати ці знання у професійній діяльності.
ПРН 11. Знати принципи, технології і прийоми створення усних і письмових текстів різних жанрів і стилів державною та іноземними мовами.
ПРН 12. Аналізувати мовні одиниці, визначати їхню взаємодію та характеризувати мовні явища і процеси, що їх зумовлюють.

ПРН 14. Використовувати мови, що вивчаються, в усній та письмовій формах, у різних жанрово-стильових різновидах і регістрах спілкування (офіційному, неофіційному, нейтральному), для розв'язання комунікативних завдань у побутовій, суспільній, навчальній, професійній, науковій сферах життя.

ПРН 15. Здійснювати лінгвістичний, літературознавчий та спеціальний філологічний аналіз текстів різних стилів і жанрів.

ПРН 16. Знати й розуміти основні поняття, теорії та концепції обраної філологічної спеціалізації, уміти застосовувати їх у професійній діяльності.

ПРН 17. Збирати, аналізувати, систематизувати й інтерпретувати факти мови й мовлення та використовувати їх для розв'язання складних задач і проблем у спеціалізованих сферах професійної діяльності та/або навчання.

ПРН 19. Мати навички участі в наукових та/або прикладних дослідженнях у галузі філології.

5. Структура освітнього компонента

Таблиця 2.1. (Денна форма)

Назви змістових модулів і тем	Усього	Лек.	Практ.	Сам. роб.	Конс.	Методи контролю / Бали
5-й семестр						
Змістовий модуль 1. Foundations of Computational Linguistics						
Тема 1. Introduction to Computational Linguistics: Goals, Scope, and Applications	8	2	2	4	–	ДС+ РМГ / 2
Тема 2. Historical Development and Key Milestones in Computational Linguistics.	8	2	2	4	–	ДС+ РМГ / 4
Тема 3. Core Terminology and Fundamental Concepts in Computational Linguistics	12	2	4	4	2	ДС+ РМГ / 4
Разом за модулем 1	28	6	8	12	2	10
Змістовий модуль 2. Computational Lexicography and Natural Language Processing						
Тема 4. Digital Lexicography: Building and Analyzing Electronic Dictionaries	13	2	6	4	1	ДС+ РМГ / 10
Тема 5. Introduction to Natural Language Processing (NLP): Basic Techniques and Challenges	10	2	2	6	–	ДС+ РМГ / 5
Тема 6. Approaches in NLP: Rule-Based, Statistical, and Neural Models	11	2	4	4	1	ДС+ РМГ / 5
Разом за модулем 2	34	6	12	14	2	20
Змістовий модуль 3. Computational Linguistics, Machine Translation and AI						
Тема 7. Machine Translation: History and Techniques	8	2	2	4	–	ДС+ РМГ / 5
Тема 8. Evaluating Machine Translation Quality	9	2	2	4	1	ДС+ РМГ / 10
Тема 9. Applications of AI in Computational Linguistics	11	2	4	4	1	ДС+ РМГ / 5
Разом за модулем 3	28	6	8	12	2	20

Змістовний модуль 4. Computational Linguistics and Corpus Linguistics						
Тема 10. Corpus Linguistics: Fundamental Concepts and Principles	8	2	2	4	–	ДС+ РМГ / 5
Тема 11. Corpus-Based Linguistic Studies	9	2	2	4	1	ДС+ РМГ / 5
Тема 12. Corpus Linguistics in Computational Linguistics: Building and Analyzing Corpora	13	2	6	4	1	ДС+ РМГ / 10
Разом за модулем 4	30	6	10	12	2	40
Екзамен						30
Всього годин/Балів	120	24	38	50	8	100

6-й семестр						
Змістовий модуль 1. Theoretical and Practical Background of Corpus Linguistics						
Тема 1. The Evolution of Corpus Linguistics	8	2	2	4	–	ДС+ РМГ / 4
Тема 2. Keys terms of Corpus Linguistics	10	2	2	4	2	ДС+ РМГ / 4
Тема 3. Methods and Procedures Used in Corpus Linguistics.	10	2	4	4	–	КР+ РМГ / 4
Тема 4. Types of Corpora: their Usage and Applications	10	2	4	4	–	ДС+ РМГ / 4
Тема 5. Application of Corpora in Applied Linguistics and Different Areas of Linguistic Research	8	2	2	4	–	ДС+Т / 4
Разом за модулем 1	46	10	14	20	2	20
Змістовий модуль 2. Building and Designing Corpora						
Тема 6. Building a spoken corpus: the basics	10	2	4	4	–	ДС+РМГ / 4
Тема 7. Building a written corpus: the basics	6	2	2	2	–	ДС+РМГ / 4
Тема 8. Building specialized corpora	8	2	2	4	–	ДС+Т / 4
Тема 9. The production of dictionaries and grammars	10	2	4	4	–	ДС+РМГ / 4
Тема 10. Development of Corpus Linguistics in Ukraine	10	2	2	4	2	ДС+ РМГ / 4
Разом за модулем 2	44	10	14	18	2	20
Екзамен						30
Всього годин / Балів	90	20	28	38	4	100

*Форма контролю: ДС – дискусія, Т – тести, РМГ – робота в малих групах, МКР / КР – модульна контрольна робота / контрольна робота, Т – тести, К – колоквиум, Р – рецензування, ПР – презентація.

Таблиця 2.2 (Заочна форма)

Назви змістових модулів і тем	Усього	Лек.	Практ.	Сам. роб.	Конс.	Методи контролю / Бали
5-й семестр						
Змістовий модуль 1. Defining Computational Linguistics. Computational Linguistics in Information Society						
Тема 1. CL as scientific discipline: key concepts. Computational Linguistics and Other Sciences.	14	2	2	8	2	ДС / 5
Тема 2. Development of theories, approaches and methods of Computational Linguistics.	9	–	–	8	1	–
Тема 3. Modern Computational Linguistics. Computational Linguistics: Applications in Real Life.	11	–	2	8	1	ПР / 5
Разом за модулем 1	34	2	4	24	4	10
Змістовий модуль 2. Methods of Natural Language Processing						
Тема 4. NLP: Syntax and Parsing	13	2	2	8	1	ДС / 2
Тема 5. NLP: Semantics and its approaches to formalization	11	–	2	8	1	ДС / 2
Тема 6. Discourse Analysis and information structuring.	13	2	2	8	1	ПР / 6
Разом за модулем 2	37	4	6	24	3	10
Змістовий модуль 3. Computational Linguistics and Corpus Linguistics						
Тема 7. Corpus Linguistics and its basic concepts. Principles of Corpus Linguistics. Corpora for Computational Linguistics.	12	2	–	8	2	–
Тема 8. Computational Corpus Analysis. Corpus annotation. Part-of-speech tagging.	12	–	2	8	2	ПР / 10
Разом за модулем 3	24	2	2	16	4	30
ІНДЗ						10
Види підсумкових робіт						Бал
Модульна контрольна робота						30
Усне опитування						30
Всього годин/Балів	90	8	12	64	11	100

6-й семестр						
Змістовий модуль 1. Theoretical and Practical Background of Corpus Linguistics						
Тема 1. Key terms of corpus linguistics	6	1	1	4	–	ДС+ РМГ / 4
Тема 2. Methods and Procedures Used in Corpus Linguistics.	10	1	1	6	2	ДС+ РМГ / 4
Тема 3. Corpus-based vs corpus-driven approaches.	10	1	1	6	2	КР+ РМГ / 4
Тема 4. Types of Corpora: their Usage and Applications.	8	1	1	6	–	ДС+ РМГ / 4
Тема 5. Application of Corpora in Applied Linguistics and Different Areas of Linguistic Research.	8	1	1	6	–	ДС+Т / 4
Разом за модулем 1	42	5	5	28	4	20

Змістовий модуль 2. Building and Designing Corpora						
Тема 6. Building a spoken corpus: the basics.	10	1	1	6	2	ДС+РМ / 4
Тема 7. Building a written corpus: the basics.	8	1	1	6	–	ДС+РМГ / 4
Тема 8. Building specialized corpora.	8	1	1	6	–	ДС+Т / 4
Тема 9. The production of dictionaries and grammars.	12	1	1	8	2	ДС+РМГ / 4
Тема 10. Development of corpus linguistics in Ukraine.	10	1	1	6	2	ДС+ РМГ / 4
Разом за модулем 2	48	5	5	32	6	20
Види підсумкових робіт						Бал
Модульна контрольна робота						30
Усне опитування						30
Всього годин / Балів	90	10	10	60	10	100

*Форма контролю: ДС – дискусія, Т – тести, РМГ – робота в малих групах, МКР / КР – модульна контрольна робота / контрольна робота, Т – тести, К – колоквіум, Р – рецензування, ПР – презентація.

Зміст та види практичних завдань

Introduction to Computational Linguistics: Goals, Scope, and Applications

Introduction to the goals, scope, and main directions of computational linguistics; overview of modern applications of computational linguistics in various fields; working with computer systems for processing information of different types.

Instruments, Platforms, Technologies: Basic text processing tools, text processing libraries

Historical Development and Key Milestones in Computational Linguistics

Studying the historical development and key milestones in the field of computational linguistics; analysis of the development of corpus linguistics in Ukraine; introduction to the evolution of natural language processing methods.

Instruments, Platforms, Technologies: Historical corpora, archival linguistic resources, systems for analyzing the development of language technologies

Core Terminology and Fundamental Concepts in Computational Linguistics

Mastering basic terminology and fundamental concepts of computational linguistics; studying key terms of corpus linguistics; introduction to the conceptual framework of the field.

Instruments, Platforms, Technologies: Terminological databases, glossaries, electronic reference systems

Text Checking Tools. Checking and correcting text grammar and style; preparing materials for automated tests using corpus data; extraction of lexical and grammatical patterns; building word lists and frequency dictionaries; tasks for stylistic analysis of texts; adaptation of educational materials using corpus-based approaches; creation of exercises based on authentic language usage; development of error detection and correction systems; using corpus data for language pedagogy; analyzing learner errors through corpus linguistics methods.

Instruments, Platforms, Technologies: LanguageTool (languagetool.org), Grammarly, online grammar and style checking instruments, language teaching tools

Digital Lexicography: Building and Analyzing Electronic Dictionaries

Creating and analyzing electronic dictionaries; working with lexicographic databases; structuring dictionary entries; developing search and filtering methods for dictionary data; updating VESUM (vesum.nlp.net.ua); working on the development of dictionary portals r2u.org.ua and e2u.org.ua; applying computer algorithms for compiling electronic

dictionaries; defining types of hypertext systems and features of their functioning in electronic dictionaries; creating dictionaries based on corpus data.

Instruments, Platforms, Technologies: VESUM (vesum.nlp.net.ua), online dictionary portals r2u.org.ua and e2u.org.ua, dictionary data management systems, HTML, XML, semantic knowledge bases (OWL, ontologies), hypertext TEI (Text Encoding Initiative), electronic dictionaries, databases

Working with Linguistic Portals. Extracting and analyzing data from web sources; working with markup languages and data formats; building systems for text classification and categorization; creating and managing structured databases; web scraping and data mining; processing and normalizing text from diverse sources; implementing search and retrieval systems; developing web-based linguistic resources; working with metadata; analyzing web content for linguistic research.

Instruments, Platforms, Technologies: Mova.info, Google Analytics

Search and Classification Systems for Electronic Texts. Extracting and analyzing data from web sources; working with markup languages and data formats; building systems for text classification and categorization; creating and managing structured databases; web scraping and data mining; processing and normalizing text from diverse sources; implementing search and retrieval systems.

Instruments, Platforms, Technologies: Elasticsearch, Apache Solr, quality assessment tools, statistical analysis software, annotation platforms

Introduction to Natural Language Processing (NLP): Basic Techniques and Challenges

Applying basic natural language processing techniques; introduction to main NLP challenges; text tokenization; morphological analysis; performing tasks based on applied methods of language description and processing for computer systems; data processing (analysis, search, filtering, reproduction, etc.) using computer technology.

Instruments, Platforms, Technologies: NLP libraries, TMKX, tokenizers, language corpus processing systems, morphological analysis tools

Approaches in NLP: Rule-Based, Statistical, and Neural Models

Studying rule-based approaches to language processing; applying statistical methods in NLP; introduction to neural network models for natural language processing; comparative analysis of different approaches; practical work with various models.

Instruments, Platforms, Technologies: Statistical analysis libraries, machine learning frameworks, LanguageTool (languagetool.org/uk), Grammarly, spelling and grammar modules, NLP tools

Machine Translation Systems and CAT-tools. Translation memory systems; creating and managing glossaries and terminology databases; tasks for alignment and segmentation of parallel texts; translation quality assessment; working with TMX and TBX formats; localization of software and multimedia content; using machine translation engines within CAT tools; extraction and management of multilingual terminology; creating specialized translation memories for different domains; developing bilingual lexical resources.

Instruments, Platforms, Technologies: SDL Trados, memQ, SmartCAT, MemoQ, Translation Memory systems

Automatic Text Summarization and Indexing. Automatic text generation and refactoring; building tools for creating document outlines; obtaining structured information from text; indexing and classification of documents; keyword extraction; automatic summarization of texts; generating abstracts and annotations; text clustering and categorization; topic modeling and semantic analysis; building recommendation systems based on text content; entity recognition and relationship extraction; implementing information retrieval systems; creating searchable text databases; developing text mining applications; sentiment analysis and opinion extraction.

Instruments, Platforms, Technologies: Python, NLTK, spaCy, Gensim, TextRank, advanced summarization algorithms

Applications of AI in Computational Linguistics. Exploring AI-powered language technologies; implementing chatbots and virtual assistants; working with language models; analyzing sentiment and emotion; applications in information extraction and text generation.

Instruments, Platforms, Technologies: HTML, XML, JSON processing, SQL databases, TEI (Text Encoding Initiative), AI libraries, neural network frameworks

Corpus Linguistics: Fundamental Concepts and Principles. Corpus Linguistics in Ukraine

Studying fundamental concepts and principles of corpus linguistics; introduction to key terms of corpus linguistics; understanding methods and procedures used in corpus linguistics; working with electronic corpora.

Instruments, Platforms, Technologies: ГПАК (uacorporus.org), corpus managers, text annotation and markup tools, Sketch Engine

Corpus-Based Linguistic Studies

Conducting corpus-oriented linguistic studies; frequency analysis, collocations and concordances; processing ГПАК corpus data for lexicographic purposes; application of corpora in applied linguistics and various areas of linguistic research.

Instruments, Platforms, Technologies: ГПАК (uacorporus.org), corpus managers, statistical analysis tools, Sketch Engine, Pandas library, TF-IDF

Corpus Linguistics in Computational Linguistics: Building and Analyzing Corpora

Collection and preparation of text data for text corpora; development and standardization of text markup; compiling word lists based on corpora; creating thematically-oriented text collections; part-of-speech analysis; search and analysis of collocations; construction and statistical analysis of corpus data; export of knowledge about grammaticalized lexemes; extraction of lexical semantics phenomena (synonyms, antonyms, etc.); building a spoken corpus - basics; building a written corpus - basics; creating specialized corpora; creating grammars based on corpus data; development of corpus linguistics in Ukraine.

Instruments, Platforms, Technologies: ГПАК (uacorporus.org), corpus managers, text annotation and markup tools, Pandas library, Sketch Engine, TF-IDF, markup tools

6. Завдання для самостійного опрацювання

Таблиця 3

Тема	Кількість годин	
	денна форма	заочна форма
5-й семестр		
Тема 1. Computational and Theoretical Linguistics	4	8
Тема 2. Computational Linguistics as engineering	4	8
Тема 3. CL in Ukraine	5	8
Тема 4. The notion of computer metaphor	4	8
Тема 5. Differences between the work of the human mind and computer	5	8
Тема 6. Constraints in the context-free grammar	4	8
Тема 7. The Earley algorithm. Comparing top-down and bottom-up parsing	4	8
Тема 8. Deep learning for Automated Sentiment Analysis of social media.	4	8
Тема 9. AI and Sentiment Analysis	5	8
Тема 10. Theory-driven and Corpus-driven Computational Linguistics and the Use of Corpora	5	6
Тема 11. Computational Linguistics and AI	6	10
Разом	50	88

6-й семестр		
Тема 1. The use of corpora in language teaching	3	6
Тема 2. The use of corpora in language teaching	4	6
Тема 3. The use of corpora in teacher education	3	4
Тема 4. The use of corpus linguistics in the study of health communication	4	6
Тема 5. The use of corpora in language testing	4	8
Тема 6. The use of use of corpus linguistics in sociolinguistics.	4	6
Тема 7. The use of corpus linguistics in the study of media discourse	4	6
Тема 8. The use of corpus linguistics in forensic linguistics	4	6
Тема 9. The use of corpus linguistics in the study of political discourse	4	6
Тема 10. The use of corpus linguistics in translation studies	4	6
Разом	38	60

IV. Політика оцінювання

Оцінювання здобувачів освіти регулюється [Положенням про поточне та підсумкове оцінювання знань здобувачів вищої освіти ВНУ імені Лесі Українки](#).

Політика викладача щодо здобувача освіти. При вивченні ОК Корпусна лінгвістика здобувач освіти опановує теми, зазначені у силабусі. Для оцінювання знань студента використовують два види контролю: поточний та підсумковий. Поточний контроль включає тестування, опитування, роботу в малих групах, ведення дискусії. Такі форми контролю дозволяють перевірити рівень підготовки студентів до кожного заняття, вміння оперувати вивченим матеріалом, логічно і послідовно виконувати завдання. Питання (завдання для самостійного опрацювання) перевіряють вміння працювати самостійно, виокремлювати головні та другорядні аспекти, ставити запитання іншим учасникам освітнього процесу. Максимальна кількість балів, яку може набрати студент за змістові модулі як у першому, так і другому семестрі, становить 70.

Політика щодо визнання результатів навчання, отриманих у формальній, неформальній та/або інформальній освіті. Згідно з [ПОЛОЖЕННЯМ про визнання результатів навчання, отриманих у формальній, неформальній та/або інформальній освіті у ВНУ імені Лесі Українки](#) та рішенням науково-методичної комісії факультету іноземної філології від 02.02.2022 р., протокол № 7, здобувачам освіти можуть бути зараховані результати навчання, отримані у формальній, неформальній та/або інформальній освіті.

Політика щодо академічної доброчесності. Політика щодо академічної доброчесності унормована в [Кодексі академічної доброчесності ВНУ імені Лесі Українки](#), згідно з яким учасники освітнього процесу мають усвідомлювати значущість норм академічної доброчесності, дотримуватися навчальної етики, толерантно ставитися до всіх учасників процесу навчання, дотримуватися розкладу навчального процесу, зокрема заборонено списування під час тесту (у т. ч. із використанням мобільних телефонів). Здобувач освіти необхідно скласти всі завдання самостійно, без допомоги сторонніх осіб; надавати для оцінювання лише результати власної роботи; не публікувати відповідей до тесту та письмових завдань; зазначати джерела інформації та посилатися на роботу інших авторів у разі використання результатів їхньої роботи; не вдаватися до кроків, що можуть нечесно погіршити чи покращити результати інших здобувачів освіти.

Політика щодо дедлайнів та перескладання. У разі відсутності з об'єктивних причин (наприклад, хвороба, навчання за програмою подвійного диплома, з використанням елементів дуальної форми здобуття освіти, міжнародна академічна мобільність) на лекційному чи семінарському занятті здобувач освіти самостійно проходить пропущену тему/теми й може звернутися за консультацією до викладача. Матеріал пропущеного семінарського заняття здобувач освіти за погодженням з викладачем може відпрацювати в онлайн/офлайн-форматі (відповідно до затвердженого графіку консультацій кафедри прикладної лінгвістики), але не

пізніше прикінцевого заняття.

Здобувачі освіти мають можливість порушити будь-яке питання, яке стосується процедури проведення чи оцінювання завдань та очікувати, що воно буде розглянуто згідно із наперед визначеними процедурами у ЗВО (див. [Положення про порядок і процедури вирішення конфліктних ситуацій у ВНУ імені Лесі Українки](#), пункт 5 «ВРЕГУЛЮВАННЯ КОНФЛІКТІВ У НАВЧАЛЬНОМУ ПРОЦЕСІ»).

Політика щодо визнання результатів навчання, отриманих у формальній, неформальній та/або інформальній освіті. Згідно з [Положенням про визнання результатів навчання, отриманих у формальній, неформальній та/або інформальній освіті у ВНУ імені Лесі Українки](#) та рішенням науково-методичної комісії факультету іноземної філології від 02.02.2022 р., протокол № 7, здобувачам освіти можуть бути зараховані результати навчання, отримані у формальній, неформальній та/або інформальній освіті.

V. Підсумковий контроль

Формою підсумкового контролю ОК Комп'ютерна лінгвістика є екзамен. Для оцінювання знань ЗО використовують два види контролю: поточний та підсумковий. Максимальна кількість балів, яку ЗО може отримати за поточний контроль, становить 70 балів. Протягом семестру ЗО повинен набрати не менше 35 балів (як допуск до складання іспиту). У випадку, якщо ЗО отримав менше, ніж 35 балів, він не може бути допущеним до екзамену.

Підсумкову семестрову оцінку з ОК Комп'ютерна лінгвістика буде виставлено за результатами складання екзамену (30 балів), який включає виконання практичного завдання (15 балів) та усне опитування (15 балів, 3 питання по 5 балів). Розподіл балів поточного й підсумкового контролю наведено в таблицях 4.1. та 4.2. У випадку незадовільної підсумкової оцінки (менше 60 балів) здобувач освіти перескладає екзамен з освітнього компоненту у період, визначений розкладом залікової сесії. Перескладання екзамену допускається не більше, як два рази: один раз – викладачеві, другий – комісії, яку створює декан факультету.

Таблиця 4.1. (5-й семестр, денна та заочна форми)

Поточний контроль (має 70 балів)				Підсумковий контроль (має 30 балів)		Загальна кількість балів
Змістовий модуль 1	Змістовий модуль 2	Змістовий модуль 3	Змістовий модуль 4 / ІНДЗ*	Практичне завдання	Усне опитування	100
10	20	20	20	15	15	

* для заочної форми навчання

Таблиця 4.2. (6-й семестр, денна та заочна форми)

Поточний контроль (має 70 балів)		Підсумковий контроль (має 30 балів)		Загальна кількість балів
Змістовий модуль 1	Змістовий модуль 2	Практичне завдання	Усне опитування	100
30	40	15	15	

Питання до екзамену (5-й семестр)

1. Computational linguistics and its basic concepts.
2. Definitions of Computational Linguistics.
3. The history and development of Computational Linguistics.

4. The state of Computational Linguistics in Ukraine.
5. Formal foundations and current methods of Computational Linguistics.
6. The relationship between Computational Linguistics and Theoretical Linguistics.
7. The distinction between Computational Linguistics and Applied Linguistics.
8. Computational linguistics as engineering.
9. Common applications of Computational Linguistics.
10. How Computational Linguistics benefits language processing.
11. Differences between the work of the human mind and a computer.
12. Tackling problems in Computational Linguistics.
13. Overcoming challenges in language processing.
14. The role of algorithms in addressing Computational Linguistics problems.
15. Illustrating Computational Linguistics through examples.
16. Theories and approaches in Computational Linguistics.
17. The future of Computational Linguistics and language processing.
18. The notion of the computer metaphor.
19. Men and machines: The computational metaphor.
20. The Computational Theory of Mind.
21. The connection between the Computational Metaphor and Cognitive Psychology.
22. The connection between the Computational Metaphor and Artificial Intelligence.
23. Syntax analysis and its role in Computational Linguistics.
24. The importance of AI-driven syntax tree generation in Computational Linguistics.
25. Syntax analysis software.
26. Levels of natural language: Phonology, Morphology, Syntax, Semantics, Discourse, Pragmatics.
27. Natural language as a formal language.
28. Sentence structure in English.
29. Syntax analysis: Definitions and key concepts.
30. The role of tokens in formal approaches to syntax analysis.
31. Formation of parse trees and types of parsers.
32. Types of syntax analysis models.
33. Applications of syntax analysis in natural language processing (NLP).
34. Challenges in syntax analysis.
35. Phrase-structure grammar and its components.
36. Context-free grammars in syntax analysis.
37. Constraints in context-free grammar.
38. Parsing algorithms for context-free grammars.
39. Comparing top-down and bottom-up parsing.
40. The Earley algorithm in parsing.
41. Chomsky hierarchy and its relevance to the theory of computation.
42. Discourse analysis in Computational Linguistics.
43. Discourse analysis and information structuring.
44. Discourse analysis: Text grammar and organization.
45. Discourse analysis: Organizing world knowledge.
46. Grouping facts by topic in discourse analysis.
47. The role of frames in discourse analysis.
48. The role of scripts in discourse analysis.
49. The role of plans in discourse analysis.
50. Analyzing narratives using scripts and plans.

Питання до екзамену (6-й семестр)

1. The origins of corpus linguistics.
2. The evolution of corpus linguistics.
3. The future outlook for corpus linguistics.
4. The theoretical and practical applications of corpus linguistics.

5. The different types of corpora in corpus linguistics.
6. The key terms in corpus linguistics.
7. The advantages and disadvantages of using corpora.
8. The necessity of using corpora in certain situations.
9. The process of searching and presenting concordance lines.
10. The peculiarities of concordance lines.
11. The issues in accessing and interpreting concordance lines.
12. The use of frequency and key-word lists in corpus analysis.
13. The measurement of collocation in corpus linguistics.
14. The application of collocational information in language analysis.
15. The categories and annotation in corpus linguistics.
16. The roles of tagging and parsing in corpus analysis.
17. The process of corpus annotation.
18. The steps involved in building a corpus.
19. The challenges in building a multimodal corpus.
20. The key aspects of building a spoken corpus.
21. The process of building a written corpus.
22. The creation of small specialised corpora.
23. The methods used to build a corpus representing a language variety.
24. The construction of a specialised audio-visual corpus.
25. The differences between parallel and comparable corpora.
26. The concept of phraseology in linguistic studies.
27. The relationship between lexis and grammar in corpus linguistics.
28. The manifestation of language variation in corpora.
29. The key parameters of variation in language.
30. The measurement of co-occurrence of variation in corpora.
31. The role of variation in language description.
32. The use of corpora in academic language studies.
33. The role of corpora in language testing.
34. The methods of corpus analysis.
35. The use of corpora in the production of dictionaries and grammars.
36. The application of corpora in translation studies.
37. The contribution of corpora to literary studies and stylistics.
38. The use of corpora in forensic linguistics.
39. The concept of data-driven learning in corpus linguistics.
40. The design of materials based on corpus data.
41. The concept of reciprocal learning and the use of parallel concordances.
42. The integration of corpora in language teaching.
43. The concept and distinctive features of multimodal corpora.
44. The application of corpora in sociolinguistics.
45. The use of corpora in the study of media discourse.
46. The relationship between corpus linguistics and computational linguistics.
47. The distinctive features of multimodal corpora.
48. The different types of multimodal corpora.
49. The application of multimodal corpora in linguistic studies.
50. The practical applications of corpora in applied linguistics.

VI. Шкала оцінювання

Оцінка в балах	Лінгвістична оцінка
90 – 100	Відмінно
82 – 89	Дуже добре
75 – 81	Добре
67 – 74	Задовільно
60 – 66	Достатньо
1 – 59	Незадовільно

VIII. Рекомендована література

1. Біскуб І., Крестьянполь Л. Вебформи як інструмент автоматизованого збору інформації : монографія. Луцьк : СНУ імені Лесі Українки, 2020. 163 с.
2. Біскуб І. П., Розвод Е. В. Основи прикладної лінгвістики : методичні рекомендації. Луцьк : Вежа-Друк, 2020. 28 с.
3. Карпіловська Є.А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика: Підручник. Донецьк: ТОВ «Юго-Восток, Лтд», 2006. 188 с.
4. An Introduction to Language and Linguistics / Ed. by Ralph Fasold and Jeff Connor-Linton. Cambridge: CUP. 2006. 540 p.
5. Berns M. Concise Encyclopedia of Applied Linguistics. Oxford: Elsevier. 2010. 571 p.
6. Biskub I. Modeling natural language communication in chat robots. *Computer science and information technologies* : proceedings of the International Conference on Computer Science and Information Technologies, 28-30 September. 2006. Lviv : Institute of Computer Science and Information Technologies, 2006. P. 12–18.
7. Biskub I. Applied and Computational Linguistics : підручник для студентів спеціальності «Прикладна лінгвістика». Луцьк : Вид-во Волин. держ. ун-ту «Вежа», 2007. 303 с.
8. Davies A. An Introduction to Applied Linguistics. Edinburgh: EUP. 2007. 214 p.
9. Grabe W. Applied Linguistics: An Emerging Discipline for the Twenty-first Century. Oxford Handbook of Applied Linguistics. Oxford University Press, 2002. P. 3–12.
10. Hunston S. Corpora in Applied Linguistics. Cambridge University Press, 2022. 300 p.
11. Jurafsky Daniel, James H. Martin. Speech and language processing: An introduction to natural language processing, speech recognition, and computational linguistics. 2d ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2009. 1024p.
12. Kamath Uday, Liu John, Whitaker James. Deep Learning for NLP and Speech Recognition. Springer Publishing Company, 2019. 621p.
13. Mitkov R. The Oxford Handbook of Computational Linguistics. 2nd edn. Oxford University Press, 2022. 1392p.
14. The Concise Oxford Dictionary of Linguistics. P. H. Matthews. Oxford University Press, 2007. Oxford Reference Online. Oxford University Press. UBL: <http://www.oxfordreference.com/views/ENTRY.html?subview=Main&entry=t36.e210>
15. The Routledge Handbook of Applied Linguistics. Edited by Malcolm Coulthard and Alison Johnson. London, New York: Routledge. 2011. P. 792.
16. The Routledge handbook of research methods in applied linguistics. Ed. by Heath Rose, Jim McKinley. London, New York. 2020. P. 292.
17. Schmitt, R. & Celce-Murcia M. An Overview of Applied Linguistics. In R. Schmitt (Ed.), An Introduction to Applied Linguistics. London: Arnold. 2002. P. 184.