

ОСНОВИ КВАНТИТАТИВНОЇ ЛІНГВІСТИКИ: ПІДХОДИ, ІНСТРУМЕНТИ, КЕЙСИ

Силабус курсу, осінній семестр 2025-2026 н.р.
Загальна інформація про курс

Викладачі:



Соломія Бук, д. філол. н., професор кафедри загального мовознавства (Львівський національний університет імені Івана Франка, Україна); асоційований науковий співробітник, Інституту славістичних та кавказознавчих досліджень (Єнський університет ім. Фрідріха Шиллера, Німеччина):

- досвід викладання у вишах понад 20 років
- експертка Національного фонду досліджень України, Міжнародних науково-технічних конференцій Інформатика та інформаційні технології / IEEE International Conference on Computer Science and Information Technologies (CSIT–2021, 2023, 2024); членкиня Журі визначення "Найкращої книги 22 Форуму видавців 2015"
- понад 100 наукових публікацій в Україні та за кордоном, 7 словників, відгуки на автореферати дисертацій; рецензії курсових, магістерських і бакалаврських робіт; її монографію "Велика проза Івана Франка: електронний корпус, частотні словники та інші міждисциплінарні контексти" висунуто на здобуття Міжнародної премії І. Франка
- доповіді на понад 100 симпозіумах, конференціях та семінарах (Австрія, Болгарія, Великобританія, Канада, Китай, Молдова, Німеччина, Польща, Словаччина, Чехія, Україна, Франція, Швеція)
- учасниця понад 10 науково-дослідних проєктів і стажувань в Австрії, Великобританії, Німеччині, Польщі, Чехії, Україні
- членкиня редколегії міжнародних та всеукраїнських фахових наукових журналів "Limbaј și context = Speech and Context", "Іноземна філологія", "Вісник Львівського університету. Сесія Філологічна", "Теорія і практика викладання української мови як іноземної"
- членкиня 4 міжнародних наукових організацій; Експертної ради Міжнародної премії імені Івана Франка (2025).

Сфера наукових інтересів: прикладна лінгвістика (зокрема, корпусна, квантитативна, комп'ютерна, а також лінгводидактика і лексикографія), лінгвістична семантика та прагматика, філософія науки, психологія щастя.

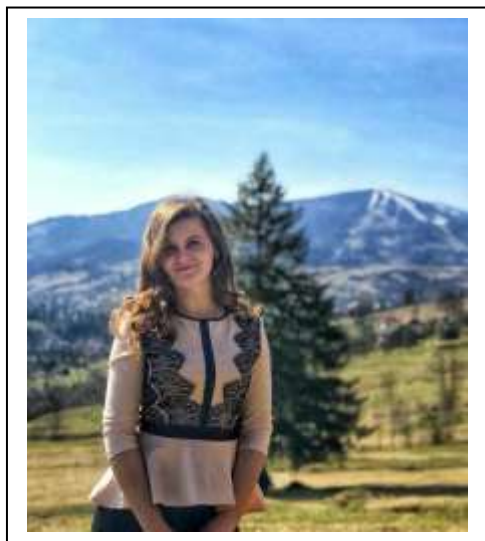
Профіль в ORCID: [0000-0001-8026-3289](https://orcid.org/0000-0001-8026-3289)

Профіль у Google Scholar: scholar.google.com.ua/citations?user=dGagnL8AAAAJ

Профіль у Scopus: <https://www.scopus.com/authid/detail.uri?authorId=26631678100>

Профіль в Academia.edu: <https://lnulviv.academia.edu/SolomijaBuk>

Христина Пронь, старший викладач (Львівська національна академія ім. М. Лисенка)



Старший викладач Львівської національної музичної академії ім. Миколи Лисенка. Дослідниця, фахівець з прикладної та комунікативної лінгвістики. Займається проблематикою дослідження мовленнєвого жанру «докір» в українському мовознавстві, аналізом корпусів текстів та їх маркуванням. З 2018 викладає українську мову для іноземців в Ukrainian Language Academy.

Коротка анотація курсу:

Мета курсу – сформувати знання щодо застосування основних квантитативних (кількісних) методів у філології: мовознавстві та літературознавстві, стилometriї, корпусній, квантитативній та історичній лінгвістиці, лінгводидактиці та оволодінні іноземною мовою тощо; навчити студентів застосовувати статистичні та математичні інструменти для дослідження мовних явищ, розвинути навички обробки великих текстових даних.

Завдання курсу – (1) осмислити карколомний кількісний поворот у лінгвістиці: протягом останніх 20 років у світовій лінгвістиці частка кількісно-статистичних методів неухильно зростає, а частка якісного підходу різко зменшується (Kortmann 2021);

(2) дати студентам уявлення про те, як покращити якість своїх досліджень, застосовуючи основні кількісні інструменти, і таким чином зробити молоде покоління більш конкурентоспроможним в міжнародній галузі досліджень;

(3) проаналізувати основні статистичні інструменти текстових корпусів, програм Sketch Engine, Stylo, графіків CQL, Word, Excel тощо та сформувати навички їх використання; навчитися візуалізувати текстові дані;

(4) з'ясувати роль квантитативного опису мови і тексту, а також визначити місце цієї галузі серед цифрових гуманітарних наук;

(5) показати значення основоположних кількісних понять (частота, відносна частота та її порівняння; ключові слова, вимірювання колокацій, багатовимірний аналіз тощо) та їхнє значення в гуманітарних галузях, згаданих вище.

Практичні заняття передбачають використання корпусних та інших цифрових технологій, написання ефективних промптів для ШІ (зокрема, ChatGPT), зіставлення мовних одиниць та їх кількісних показників, формат академічної дискусії, командну роботу, створення та аналіз частотного словника, автоматичне визначення авторства, лінгвістичну експертизу конкретних текстів,

дослідження закономірностей вживання слів залежно від жанру, автора, часу написання, які разом із індивідуальним проектом і підсумковим есе визначають успішність роботи студентів за семестр.

Пререквізити. Базові теоретичні знання з мовознавства (вивчили "Вступ до мовознавства", "Вступ до прикладної лінгвістики") бажані, але не обов'язкові для успішного проходження курсу. Можлива участь студентів-першокурсників. Запрошуємо також студентів з дотичних до мовознавства спеціальностей: літературознавства, лінгводидактики, фольклористики, діалектології, культурології, гендерних студій тощо.

Формат курсу: синхронно-асинхронне навчання онлайн з чат-підтримкою викладача.

- Синхронне навчання відбувається на платформі Zoom — це лекції і практичні заняття, а також тестування.
- Асинхронне навчання — це доступні онлайн протягом курсу відеолекції, завдання до практичних занять, наукова література і чат з підтримкою викладачів в одному з каналів (Facebook messenger, Viber або ін.).
- Індивідуальний проект, який демонструє застосування знань прикладної лінгвістики у власній курсовій/дипломній/магістерській/іншій дослідницькій роботі студента/ки. Студентам-першокурсникам, які ще не пишуть курсової, запропонуємо посильну тему дослідження.

Результати навчання

У результаті вивчення навчальної дисципліни здобувач освіти має бути здатним продемонструвати такі **результати навчання**.

Знати:

- найважливіші поняття і категорійний апарат квантитативної лінгвістики;
- лексикографічні та електронні джерела знаходження кількісної інформації про мовні одиниці;
- основні статистичні інструменти текстових корпусів, програм Sketch Engine, Stylo, графіків CQL, Word, Excel тощо;
- ключові інструменти візуалізації статистичної інформації про текстові дані;
- частотні словники української мови та їх визначальні характеристики;
- основні підстави застосування квантитативних методів у мовознавстві;
- розвиток та становлення квантитативних методів у мовознавстві;
- поняття статистичної структури тексту, типологію та проблеми укладання частотних словників, методика їх порівняння;
- лінгвостатистичні характеристики, закони, залежності;
- кількісні характеристики одиниць різних рівнів мови та мовлення.

Уміти:

- знаходити необхідну кількісну інформацію про мовні одиниці в лексикографічних та електронних джерелах;

- застосовувати квантитативні методи для аналізу тексту у своїй дослідницькій роботі та для розв'язання різних типів лінгвістичних завдань;
- укладати і зіставляти частотні словники, правильно інтерпретувати їх інформацію;
- виявляти і правильно інтерпретувати квантитативні дані про мовну одиницю, отримані з лексикографічних та електронних джерел, основні їх залежності і закони;
- користуватися основними статистичними інструментами текстових корпусів, програм Sketch Engine, Stylo, графіків CQL, Word, Excel тощо;
- візуалізувати статистичну інформацію про текстові дані за допомогою різних ресурсів;
- оперувати найважливішими поняттями квантитативної лінгвістики;
- аналізувати мовлення (текст) на предмет кількісних характеристик;
- послуговуватися корпусами текстів для виявлення зазначених характеристик на різних рівнях;
- критично оцінювати кількісні дані текстів, отримані з частотних словників, корпусів текстів та комп'ютерних програм.

Мова викладання: українська

Викладачі:

Соломія Бук, д. філол. н., проф. кафедри загального мовознавства Львівського національного університету ім. І. Франка (лекції);
Христина Пронь, старший викладач ЛНМА ім. М. Лисенка (практичні).

Тривалість курсу: 8 тижнів

Обсяг навчального навантаження:

Обсяг курсу: 100 год. (4 кредити ECTS).

Лекції: 16 ак. год.

Практичні заняття: 16 ак. год.

Самостійне навчання: 68 год.

Схема курсу:

Теми лекційних занять:

Тема 1. Кількісний поворот у лінгвістиці. Квантитативна лінгвістика як наука та її основні завдання. Нові тренди галузі.

Тема 2. Що і для чого рахувати у мові? – Основні поняття та категорійний апарат квантитативної лінгвістики: вибірка, частота, розподіл.

Тема 3. Що показують частотні словники, як ними користуватися і укладати самим? – Частотні словники, їх типологія та проблеми укладання. Статистична структура тексту. Електронні інструменти укладання частотних списків.

Тема 4. Який текст багатший? – Методика порівняння частотних словників. Лінгвостатистичні характеристики, закони, залежності.

Тема 5. Як визначити автора тексту? – Стилеметрія. Атрибуція тексту, її види (авторська, часова тощо). Визначення авторства тексту. Методи й інструменти атрибуції тексту. Програма аналізу українських текстів TextAttributor

Тема 6. «Відбиток пальця» І. Франка у текстах – Лінгвостатистична модель ідіолекту: приклад І. Франка.

Тема 7. Скільки слів потрібно, щоб вивчити мову? – Частотність як основний критерій компресії словника. Словники-мінімуми. Лексична основа української мови.

Тема 8. Як поміряти мову? – Лінгвометрія. Кількісні параметри мови та мовлення. Лінгвостатистичні закони (переваги, Ципфа, Ципфа-Мандельброта). Лінгвометричні методи у виявленні варіативності української мови ХХ ст. на прикладі корпусу текстів.

Теми практичних занять:

1. Квантитативна лінгвістика як наука та її основні завдання. Кількісний поворот у лінгвістиці. Статистика корпусних ресурсів української мови. Визначення орієнтовної теми індивідуального проєкту.

2. Основні поняття та категорійний апарат статистичної лінгвістики: вибірка, частота, розподіл. Звідки брати статистику мовних одиниць і про що вона свідчить. Аналіз вибірок (підкорпусів) електронних ресурсів української мови.

3. Частотні словники, їх типологія та проблеми укладання. Статистична структура тексту. Електронні інструменти укладання частотних списків. Практичне укладання частотного словника обраного тексту. Командна робота – співвідношення стилів у репрезентативному частотному словнику мови

4. Методика порівняння частотних словників. Лінгвостатистичні характеристики і закони. Визначення найважливіших характеристик свого укладеного власноруч частотного словника.

5. Стилеметрія. Атрибуція тексту, її види (авторська, часова тощо). Визначення авторства тексту. Методи й інструменти атрибуції тексту.

6. Лінгвостатистична модель ідіолекту. Приклад І. Франка. Кількісно-візуальний розподіл прозових творів письменника; Потенціал моделі на прикладі творчості інших письменників. Ключові інструменти візуалізації статистичної інформації про текстові дані. Представлення індивідуальних проєктів

7. Частотність – критерій компресії словника. Електронні ресурси візуалізації кількісних даних сполучуваності мовних одиниць. Представлення індивідуальних проєктів

8. Лінгвометрія. Кількісні параметри мови та мовлення. Представлення індивідуальних проєктів.