

ІКТ у викладанні філологічних дисциплін

УДК 81'3-047.37-027.44

Маргарита Надутенко, Максим Надутенко, Олена Семенов

Застосування цифрового методу у викладанні філологічних дисциплін (на прикладі віртуальної лексикографічної лабораторії)

У статті схарактеризовано віртуальну лексикографічну лабораторію «Мультимедійний словник з інфомедійної грамотності» як цифровий інструмент у викладанні філологічних дисциплін. На прикладі віртуальної лексикографічної лабораторії презентовано застосування цифрового методу для викладання філологічних дисциплін, який використано для оброблення та візуалізації цифрової лінгвістичної інформації. Серед методів виділено лексикографічний метод дослідження, розкрито його значення, доцільність застосування, особливості впровадження та подальшого використання у загальноукраїнській та європейській співпраці.

Мета запропонованої розвідки – презентація віртуальної лексикографічної лабораторії «Мультимедійний словник з інфомедійної грамотності» у викладанні філологічних дисциплін як одного із основних методів дослідження цифрової лінгвістичної інформації, яку використано для обробки цифрових лінгвістичних ресурсів та для створення відповідних програмних продуктів у межах загальноукраїнської та міжнародної співпраці. Наукова новизна дослідження полягає в практичній пропозиції нових методів навчання у вищій школі.

Результати проведеної роботи дали можливість встановити основні методи цифрових лінгвістичних досліджень, які доцільно використовувати для дослідження мовного матеріалу. У групі презентованих методів виділено основні: статистичний, корпусний, лексикографічний метод дослідження, в основі якого теорія лексикографічних систем, та метод штучного інтелекту: машинне навчання, глибоке навчання, нейромережі.

Ключові слова: цифрові методи лінгвістичних досліджень, корпусні технології, мультимедійний словник, лексикографічний метод дослідження, теорія лексикографічних систем, штучний інтелект, машинне навчання, нейромережі.

Вступ

У діджиталізованому суспільстві XXI століття знання стають одним із найважливіших чинників культурного, економічного, технологічного та гуманітарного розвитку кожної держави. Знаннєва інформація може зберігатися, передаватися й оброблятися в різних формах, у тому числі й у символній (знаковій) формі. Одна й та сама інформація може мати також різні форми вираження, що здійснюється засобами певних знакових систем. Такі знакові системи конструюються з базових елементів («алфавітів»), що «споряджені» актуальними правилами для виконання операцій. Для забезпечення інформаційного процесу необхідні щонайменше три елементи: джерело інформації, її одержувач (реципієнт) та пристрій зв'язку, що забезпечує доправлення інформації від джерела до реципієнта. Таким чином, інформація завжди є інформацією про певний об'єкт, який і є її джерелом.

Основну частину нових знань отримуємо безпосередньо з текстів, що створені природною мовою (мовами) і виконують роль носіїв лінгвістичної інформації. Динамічний розвиток інформаційно-комунікаційних технологій обумовлює трансформацію природної мови у мову символів, за допомогою яких кодифікуються та зберігаються великі масиви мовної інформації. Наприклад, цифровий пристрій зберігає й обробляє інформацію за допомогою двійкової системи.

На рисунку 1 представлено приклади інформації у цифровому форматі, наведено приклади трансформації різноманітного контенту (наскельні малюнки, мова символів) у цифровий формат.

Одна з сутнісних ознак людини є **мова**.

За допомогою мови ми не тільки обмінюємось інформацією, а й визначаємо та формуємо своє буття.

Мова значною мірою відображає **спосіб мислення**, вона є однією з ключових ознак людського інтелекту.

Людина, в процесі пізнання дійсності, створює **штучні мови** оскільки спілкується за допомогою знаків, символів, цифр, кодів.

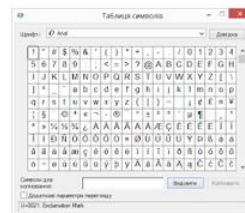
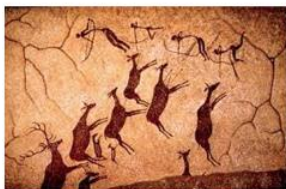


Рис. 1. Приклади інформації у цифровому форматі

Висока швидкість надходження нових даних і зростання обсягів накопиченої інформації стимулюють потребу у високоефективних інформаційних цифрових технологіях й інтелектуальних системах, пошуку нових способів автоматичної обробки інформації, її систематизації, відповідної класифікації та презентації на вимогу користувача.

Зокрема, в основу цифрових технологій проведення лінгвістичних досліджень покладено перехід до цифрової інформації, трансформацію різноманітного контенту у відповідний цифровий ресурс. Рисунок 2 відображає процес переходу мовної інформації у цифровий контент. На початковому етапі виділено технологічний процес обробки мовних одиниць (малюнки, смайли, тексти на різноманітних носіях, графіка тощо), що характеризується вибором необхідних засобів цифровізації та створенням цифрового контенту з можливістю подальшого статистично-математичного моделювання й аналізу лінгвістичного контенту.

ЦИФРОВІ ТЕХНОЛОГІЇ У ЛІНГВІСТИЧНИХ ДОСЛІДЖЕННЯХ

в основі цифрових технологій є перехід до цифрової інформації
– трансформація мовних одиниць у цифровий ресурс

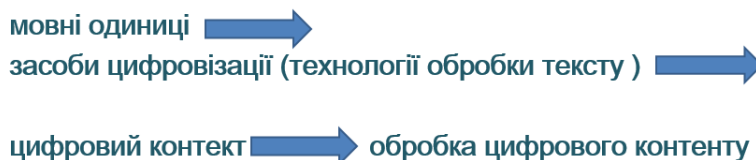


Рис. 2. Цифрові технології у лінгвістичних дослідженнях

У сучасних лінгвістичних дослідженнях (Лінгвістично-інформаційні студії, 2018; Mintser, O., Babintseva, L., Zaliskyi, V., Nadutenko, M., Kharchenko, N., & Ladychuk, O., 2021; Nadutenko, M., Prykhodniuk, V., Shyrokov, V., & Stryzhak, O., 2022) перевагу надають цифровому (автоматичному) методу обробки лінгвістичної інформації. Означений метод забезпечує можливість опрацювання великих масивів баз даних, здатність працювати з будь-яким контентом (паперові та цифрові носії інформації тощо), а також можливість подальшого цифрового опрацювання: кількісно-статистичні підрахунки, моделювання, надання доступу широкому колу користувачів; можливість трансформації в інші середовища, і,

таким чином, забезпечує відкритість української науки на європейському та світовому філологічному просторі.

До основних трендів мережевого світу, що визначають розвиток цифрових технологій у лінгвістичних дослідженнях, фахівці на основі аналізу тенденцій відносять такі, як віртуалізація мовного середовища, інтелектуалізація, робота з Великими даними.

Відповідно до основних трендів виникає необхідність пошуку нових ефективних методів дослідження лінгвістичної інформації, створення програмних засобів-інструментів та їх подальше ефективне застосування для викладання філологічних дисциплін у вищій школі.

Для забезпечення доступності програмних продуктів студентам – майбутнім учителям-філологам, а також широкому колу зацікавлених користувачів важливою є співпраця дослідників, які створюють лінгвістичні програмні інструменти, здійснюють цифровізацію контенту, автоматичну обробку, систематизацію, класифікацію великих масивів даних, з викладачами філологічних дисциплін закладів вищої освіти. У 2018–2022 роках така науково-дослідна співпраця відповідно до угоди виконується між Українським мовно-інформаційним фондом Національної академії наук України (далі – УМІФ НАН України) та кафедрою української мови і літератури Сумського державного педагогічного університету імені А. С. Макаренка (далі – СумДПУ).

Упродовж кількох останніх років дослідники УМІФ НАН України створили Український національний лінгвістичний корпус (Український національний лінгвістичний корпус), Інтегровану лексикографічну систему «Словникі України online» (Інтегрована лексикографічна система «Словникі України online»), банк океанографічних даних на основі онтологічних інтерактивних документів (Stryzhak, O., Prykhodniuk, V., Popova, M., Nadutenko, Naiko, S., & Shepko, R., 2021), онтологічно керовані лексикографічні системи (Nadutenko, M., Prykhodniuk, V. Shyrov, V., & Stryzhak, O., 2022).

У співпраці з фахівцями закладів вищої освіти, як показує аналіз, науковці досліджують парадигмальні засади лінгвістики першої половини ХХІ століття, яка вже «кваліфікується як еволюційно-інформаційно-феноменологічна» (Широков, В. & Лучик, А., 2021), розробляють онтологічно керовані лексикографічні

системи (Nadutenko, M., Prykhodniuk, V., Shyrokov, V., & Stryzhak, O., 2022).

У співпраці УМІФ НАН України та освітньо-дослідницького центру «МЕДІА&ВЧИТЕЛЬСЬКИЙ кампус» кафедри української мови і літератури СумДПУ імені А. С. Макаренка окреслено технічні особливості віртуальної лабораторії «Мультимедійний словник з інфомедійної грамотності» (Семенов, О. М., & Надутенко, М. В., 2022), здійснено огляд створення та використання платформи «Трансдисциплінарний кластер «МЕДІА&КАПСУЛИ» як засобу підвищення рівня медіакультури (Semenog, O., Nadutenko, M., & Nadutenko, M., 2021; Широков, В., Загнітко, А., Надутенко, М., Надутенко, М. & Семенов, О. та ін., 2022).

Проведений аналіз засвідчує відсутність системних досліджень щодо застосування цифрового методу у викладанні, що й актуалізувало потребу схарактеризувати застосування цифрового методу у викладанні філологічних дисциплін.

Мета статті – схарактеризувати застосування цифрового методу у викладанні філологічних дисциплін, завдання – окреслити методи дослідження цифрової лінгвістичної інформації, зокрема методи статистичної лінгвістики, корпусних технологій, штучного інтелекту, лексикографічного методу, з'ясуємо специфіку використання віртуальної лексикографічної лабораторії «Мультимедійний словник з інфомедійної грамотності» у викладанні філологічних дисциплін.

Методи й методики

У дослідженні враховуємо тезу про те, що метод формує підходи до аналізу мовних і мовленнєвих явищ. Для вирішення поставлених завдань використано термінологічний і поняттєвий аналіз з метою уточнення сутнісних характеристик методів дослідження цифрової лінгвістичної інформації

У межах власне проведеного дослідження представлено характеристики таких методів, як описовий, статистичний (математичний), метод корпусних технологій, лексикографічний, метод штучного інтелекту (машинне навчання, «глибоке» навчання, нейромережі).

Виклад основного матеріалу

Відповідно до завдань окреслимо методи дослідження цифрової лінгвістичної інформації. Лінгвометодологію дослідники характеризують як самостійну галузь теоретичного знання, що

проектується в загальну методологію – учення про способи пізнання й осмислення дійсності й формування внутрішнього рефлексивного досвіду людини, а також про дослідницькі настанови, принципи та процедури аналізу об'єктів мовного середовища.

УМІФ НАН України розроблено теоретичні та науково-технічні засади методів дослідження цифрової лінгвістичної інформації.

Серед основних методів дослідження цифрової лінгвістичної інформації, які слугують основою сучасних лінгвістичних досліджень, виділяємо статистичний метод обробки інформації (метод статистичної лінгвістики), метод корпусних технологій, лексикографічний, штучного інтелекту. Класифікація методів дослідження формується відповідно до мети їх використання та має враховувати специфіку джерельної бази дослідження. Запропоновані методи дослідження не суперечать класичним методам дослідження та можуть бути використані у комплексі.

Статистичний метод, зокрема, розглядає мову як системно-структурне утворення, що має окремі підсистеми – рівні, представлені відповідними одиницями: фонемами, морфемами, лексемами, синтагмами. Таким чином, мова характеризується не лише якісними, а й кількісними показниками. Кількісна методика стала більш ефективною з появою відповідного програмного забезпечення відкритого доступу.

На рис. 3 представлено приклад застосування статистичного методу в лінгвістичних дослідженнях, зокрема, для виявлення особливостей функціональних стилів мови та особливостей стилю окремих авторів.

• **Застосування статистичного методу:**

- для виявлення особливостей функціональних стилів мови та особливостей стилю окремих авторів;
- для виявлення хронологічних особливостей мовних одиниць;
- лінгводидактика (добір слів для вивчення іноземної мови);
- фонетичний устрій мови та частота появи звуків, букв, фонем;
- лексичний склад мови та **укладання частотних словників**;
- авторський та функціональний стилі, спрямовані на встановлення авторства творів;
- одиниці різних рівнів мови та встановлення їх кількісних характеристик;
- дешифрування історичних писемностей тощо



Рис. 3. Приклад застосування статистичного методу у лінгвістичних дослідженнях

Методи статистичної лінгвістики використовують для лінгвістичного моніторингу функціонування мови в конкретному типі дискурсу (політичному, науковому, засобів масової інформації тощо), для контент-аналізу (виявлення стану суспільної свідомості). Предметом моніторингу можуть бути типи мовних помилок, сфера іншомовних запозичень, нові слова та значення, нові метафори, тематичний розподіл лексики (наприклад, лексика часових та просторових відношень, лексика на позначення почуттів та емоцій та ін.), особливості використання в текстах певних граматичних форм чи синтаксичних конструкцій.

Приклад виявлення та порівняння синонімічних рядів у лексемах *благодать* (розм.), *благо* (заст.) та сучасного багатозначного відповідника *добро* в Інтегрованій лексикографічній системі «Словникі України online» подано на рис. 4.



Рис. 4. Приклад виявлення та порівняння синонімічних рядів у лексемах: *благодать* (розм.), *благо* (заст.), *добро* (Інтегрована лексикографічна система «Словникі України online»)

Відповідно до аспектів дослідження сучасна лінгвістика пропонує цілий ряд методик та прийомів аналізу одиниць лексичної системи. Метод корпусних технологій вважаємо за доцільне використовувати як необхідну складову для створення мінікорпусів відкритого типу. Нейромережі розглядаємо як підмножину штучного інтелекту, що зосереджена переважно на проєктуванні систем, які дозволяють навчатися та робити прогнози на основі певного досвіду.

Штучний інтелект розглядаємо як форму індивідуалізації систем, якій властивий мовний статус. Ідеться про імітацію чи

реплікацію когнітивних рис людини машинами – комп'ютерними програмами.

Загальне поняття «штучний інтелект» охоплює кілька напрямів: експертні системи, системи для характеристичного аналізу і робототехніка; здатність інженерної системи обробляти, застосовувати та вдосконалювати набуті знання. Це система, що має певні ознаки інтелекту, тобто здатна розпізнавати та розуміти; знаходити спосіб досягнення результату та приймати рішення.

Наприклад, для визначення семантичної подібності між реченнями та великими контекстами, необхідної для індексування, нами використовується метод семантичного хешування, що не потребує значних обчислювальних потужностей на попередню обробку текстів, як того вимагають методи на основі нейронних мереж.

Для семантичного хешування застосовується метод розбиття тексту на N-грами слів. Далі для побудованих наборів N-грам розраховується набір із 1000 хеш-значень на основі комбінації таких некриптографічних хеш функцій, як CityHash, spookyHash та xxHash, що мають високу колізійну якість та якість псевдовипадковості.

Для визначення семантичної подібності речень використовується набір алгоритмів типу Word2Vec. Водночас для побудови семантичного вектору контексту може використовуватись набір кластерів, які відображають тематику тексту. Тематики визначаються автоматично щоразу при індексації на основі кластеризації. Якщо слово є реєстровим, у якості початкових значень може використовуватися семантична ремарка зі «Словника української мови» у 20 томах (URL: <https://services.ulif.org.ua/expl/>).

Машинне навчання здійснюється без викладача на основі контекстуального представлення Українського національного лінгвістичного корпусу: кожне слово поєднується контекстним зв'язком як із наступним, так і з попереднім словом. Відповідно нечіткий пошук за реченнями нами реалізовано на основі трьох методів: вибору оптимально близьких за наборами хешів речень, речень, ранжованих за найменшою косинусною відстанню між їхніми семантичними векторами розмірністю 3000 та речень, отриманих за допомогою векторного представлення на основі алгоритму BERT.

Практичне застосування зазначених вище методів дослідження засвідчує високу ефективність запроваджених засобів у формі

віртуальних лексикографічних лабораторій та згрупованих лінгвістичних платформ, зокрема мультимедійних словникових продуктів, які активно впроваджуються у загальноукраїнський та європейський філологічний простір.

Актуальним є застосування методів і технологій аналізу Великих даних та інтегрованих платформ, розроблених УМІФ НАН України (див. рис.5): «Словникі України online», «Система лінгвістичної взаємодії «ВЛЛ», «Український лінгвістичний корпус» та ін.).

**Інтегровані платформи,
розроблені Українським мовно-інформаційним фондом НАН України:**

<https://svc2.ulif.org.ua/>

ПРИКЛАДИ РОЗРОБОК

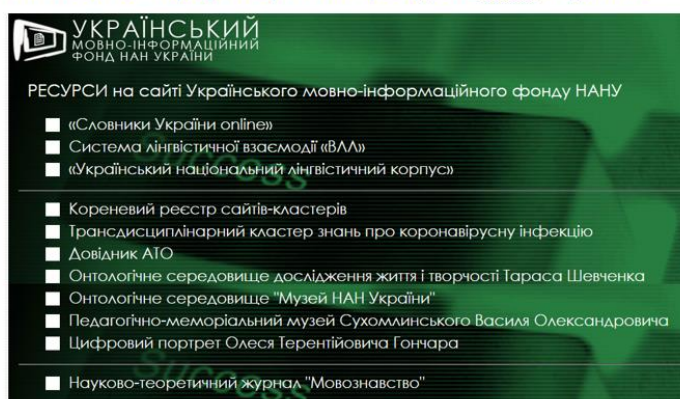


Рис. 5. Інтегровані платформи, розроблені УМІФ НАН України

Для сучасних досліджень лексичної системи мови доцільно застосовувати лексикографічний метод. Метод спрямований передусім на послідовне виділення та вибіркове вивчення окремих елементів та їх відношень у мовній системі.

Наразі триває активний пошук нових прийомів, які використовують лексикографічний метод дослідження в поєднанні з автоматичними лексикографічними системами. Наведемо приклади лексикографічних методів: класифікація лексики з метою створення мультимедійних словників нових типів; відкритий вільний асоціативний експеримент із використанням автоматичних лексикографічних систем; дистрибутивний аналіз семантем для визначення відмінності в семемах окремої лексеми у хронологічному зрізі; аналіз зіставлень та порівнянь на межі декількох мов, що здійснюється з метою дослідження встановлення частотних характеристик відповідних лексем (див. табл. 1 порівняння

частотності вживання лексем в українському та болгарському національних лінгвістичних корпусах).

Таблиця 1

Порівняння частотності вживання лексем
в Українському національному лінгвістичному корпусі (УНЛК)
та Болгарському національному корпусі (BNC)

Українська мова Український національний лінгвістичний корпус (УНЛК) https://svc.ulif.org.ua/UNLC/virt_unlc_4.5/	Болгарська мова Български национален корпус/ Bulgarian National Corpus (BNC) http://search.dcl.bas.bg/
Україна 669441	Украина 12
український народ 12968	украински народ 1
українець 56432	украински 551
відродження 11506	възраждане 1598
громадянин 67104	гражданин 19466
держава 157059	държава 46866
Радянський Союз 6857	Съветския Съюз 2536
Болгарія 5768	България 37652
болгарський народ 80	български народ 1482
Європа 61464	Европа 20651
влада 147282	власт 42531
національні інтереси 2480	национални интереси 366

Серед програмних розробок Українського мовно-інформаційного фонду НАН України, які застосовуються у викладанні філологічних дисциплін у закладах вищої освіти, відзначимо такі:

- 1) відеословник як новий метод презентації онімного матеріалу;
- 2) віртуальні лабораторії з ономастики як інструмент активної взаємодії фахівців у віддаленому режимі;
- 3) бази даних для обробки онімної інформації та масштабного аналізу;
- 4) лінгвістичні корпуси з ономастики з використанням розробленої лінгвістичної розмітки;
- 5) онотологічні екскурсії 3D панорамою віртуального ономастичного музею як новий метод презентації онімного матеріалу;

б) система «ПОЛІЕДР» як мережевий інструмент для забезпечення взаємодії з інформаційними ресурсами. URL: <http://ekmair.ukma.edu.ua/handle/123456789/20620>

Схарактеризуємо особливості застосування цифрового методу у викладанні філологічних дисциплін на прикладі віртуальної лексикографічної лабораторії «Мультимедійний словник з інфомедійної грамотності», яка створена для освітніх потреб партнерами – УМІФ НАН України та освітньо-дослідницьким центром «МЕДІА&ВЧИТЕЛЬський кампус» кафедри української мови і літератури СумДПУ імені А. С. Макаренка (Широков, В., Загнітко, А., Надутенко, М., Надутенко, М. & Семенов, О. та ін., 2022).

Цифровий продукт апробований у навчально-дослідницькій діяльності майбутніх учителів-філологів Сумського державного педагогічного університету імені А. С. Макаренка, зокрема у студентському науковому гуртку з академічної культури. Особливість віртуальної лексикографічної лабораторії полягає в тому, що це мультимедійне середовище відкритого типу, призначене для взаємодії учасників освітнього процесу з цифровими медіаресурсами.

Програмний продукт комплексно представляє інструментарій, методологію, методи та важливу для медіалінгвістики аспектологію наукових підходів. Словник побудований за класичним алфавітним принципом (див. рис. 6).

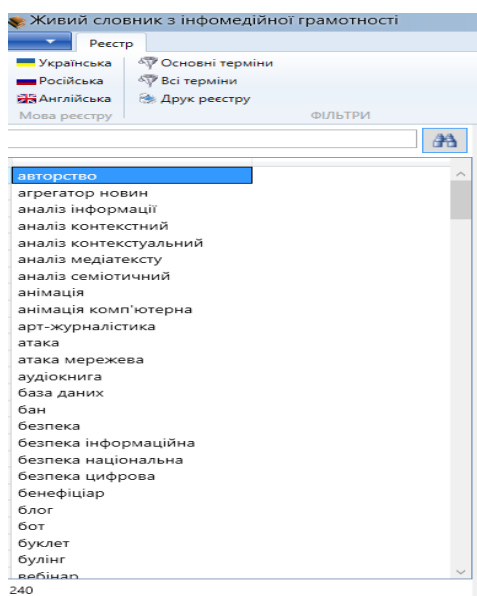


Рис. 6. Будова словника «Мультимедійний словник з інфомедійної грамотності»

Терміни й наукові поняття віртуальної лексикографічної лабораторії становлять основну проблематику медіалінгвістики, її методи, типи й види аналізованих текстів і структурно-функційних мовних одиниць цих текстів.

Словникові статті містять дотичні до власне лінгвістичних поняття, без яких неможливо проаналізувати специфіку мови у масмедіа, що є синкретичною сферою сучасної гуманітаристики (насамперед журналістські, деякі політологічні та ін.).

Зокрема, лексема «маніпулювання масовою свідомістю» (рис. 7) схарактеризована як психологічне програмування думок і прагнень людей, їхніх настроїв і психічного стану з метою забезпечення такої поведінки, яка необхідна групі небагатьох власників засобів масової інформації, що здійснюють такий вплив, переслідуючи свої особисті корисливі цілі.

Реєстр лексем ґрунтується на основі концепції інфомедійної грамотності, опрацьованої Радою міжнародних наукових досліджень та обмінів IREX. Автори концепції вважають, що компонентами інфомедійної грамотності є медіаграмотність; критичне мислення; соціальна толерантність; стійкість до впливів і маніпулювань; перевірка фактів; інформаційна грамотність; цифрова безпека; візуальна грамотність; інноваційність, розвиток креативності.

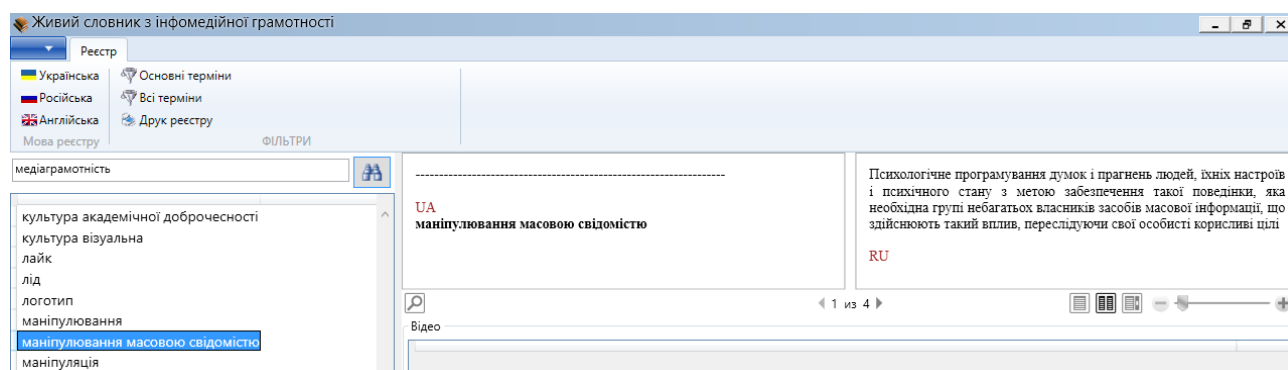


Рис. 7. Лексема «Маніпулювання масовою свідомістю» у «Мультимедійному словнику з інфомедійної грамотності»

Мультимедійний словник з інфомедійної грамотності увиразнює міждисциплінарний підхід у навчанні шляхом уведення в реєстр основних термінів із навчальних дисциплін «Медіалінгвістика», «Медіаграмотність в освітньому процесі», «Культура української мови і стилістика» і т. ін. Наприклад, лексема «Вебкартографія» презентується як географічний термін – сукупність технологій,

пов'язаних зі створенням різноманітних електронних карт, їх розміщенням та опрацюванням у вебпросторі (рис. 8).

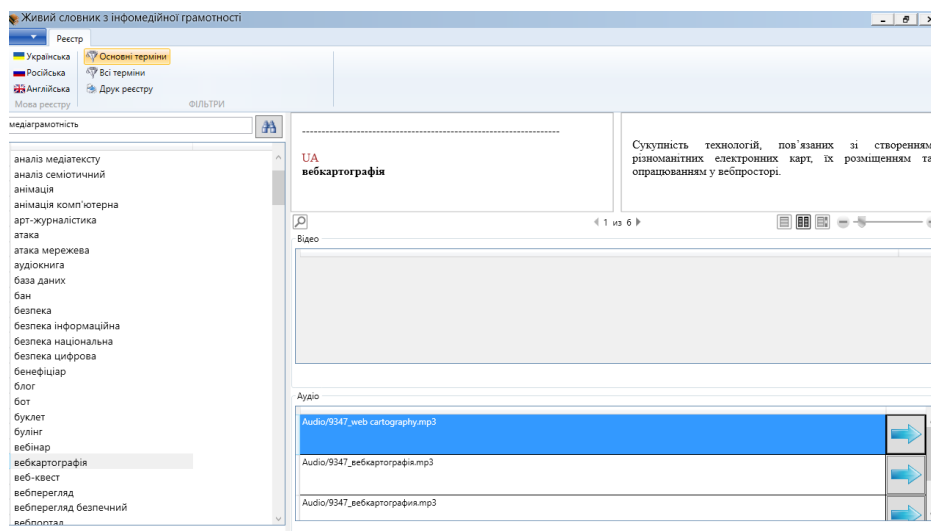


Рис. 8. Лексема «Вебкартографія» у «Мультимедійному словнику з інфомедійної грамотності»

Мультимедійний словник є унікальним цифровим продуктом віртуальної взаємодії між укладачами-учасниками проєкту (науковцями, викладачами різних дисциплін, журналістами, вчителями, студентами) з різних населених пунктів України та зарубіжжя (Semenog, O. M, Nadutenko, M. V, & Nadutenko, M. V., 2021).

Особливо посилено стали використовувати словник в умовах війни з метою розвитку життєво важливих навичок студентської молоді інфомедійної та цифрової грамотності: умінь раціонально споживати медіаконтент, критично сприймати інформацію, відрізнити факти від суджень, ідентифікувати емоційний вплив медіа, виявляти маніпулятивний контент, фейки, володіти прийомами комунікативної взаємодії.

Теми навчальної дисципліни «Культура української мови і стилістика» збагатили мультимедійною візуалізацією медіатермінів *булінг, відеоблог, гаджет, медіабезпека, медіаграмотність, медіависловлювання, онлайн-шахрайство, цифрова безпека, цифровий слід, фактчекінг, фейк, троль, чат-бот та ін.* Запропоновано їхні визначення, переклад і тлумачення англійською мовою, а також відеоролик відповідно характеристики.

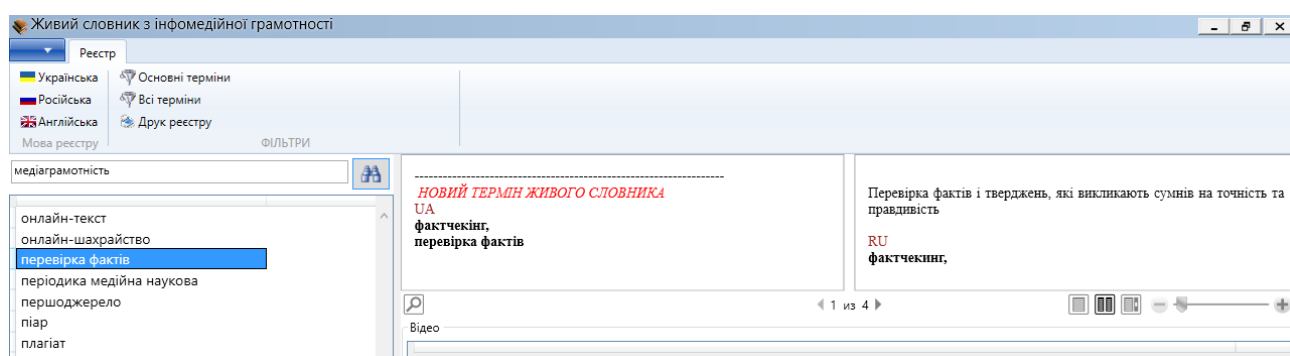


Рис. 9. Лексема «фактчекінг» у «Мультимедійному словнику з інфомедійної грамотності»

Використання віртуальної лабораторії надає освітньому процесу більшої інтерактивності: виникає можливість вирішувати специфічні завдання за допомогою різних інтерактивних методів (мозковий штурм, проєктна робота, ситуаційні завдання тощо), підвищує рівень медіакультури, критичного мислення в умовах війни. З цією метою у співпраці студентів спеціальності 014 Середня освіта. Українська мова і література СумДПУ імені А. С. Макаренка, спеціальності 061 Журналістика Машинобудівного фахового коледжу Сумського державного університету, Києво-Могилянської академії, науковцями УМІФ НАН України було створено презентаційні відеоролики (<https://www.youtube.com/watch?v=HN0bsmnM6gE>). Зокрема, на рис. 10 наведено тлумачення слова *гаджет* у презентаційних роликах.

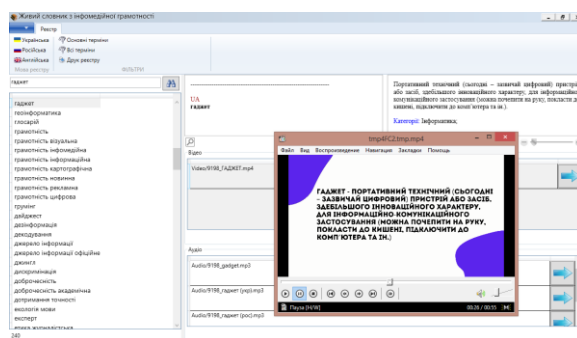


Рис. 10. Характеристика терміна «гаджет» у «Мультимедійному словнику з інфомедійної грамотності»

Автори відео Аріна Столбцова, Мирослав Деревянко, Наталія Пономаренко пропонують короткий вступ – сценку-діалог між студентами як інтерактивний формат подання інформації. На заняттях аналізуємо, що ризиком для такої інформації є динамічний характер мови. Якщо донедавна слово «гаджет» не використовувалось активно в повсякденній мові, то наразі через свою

поширеність воно вже не є неологізмом, і здивування героїв відео щодо цього слова може видаватися дивним. Отже, потребує подальших спільних досліджень щодо використання цифрового методу у філологічних дисциплінах.

Мультимедійний словник з інфомедійної грамотності увійшов складником проєкту з мовно-методичних дисциплін «Навчання української мови у Новій українській школі»: студенти створювали короткі відеоролики, візуалізації з мовної теми, укладали вправи для робочого шкільного зошита «Українська мова. 5 клас, зокрема, для рубрики «Мовленнєвий практикум. Говори правильно!».

Сталість проєкту використання віртуальної лексикографічної лабораторії у викладанні філологічних дисциплін полягає в розширенні аудиторії, долучення до співпраці викладачів-філологів та студентів Національної академії статистики обліку та аудиту і використання віртуальної лабораторії як тренажера та створення нових проєктів під час проходження навчально-виробничої практики: «Захищені месенджери», «Фінансова безпека та шахрайство» (див. рис. 11).



Рис. 11. Приклади робіт студентів Національної академії статистики обліку та аудиту для занять з лінгвістичних дисциплін

Методичні матеріали для занять, науково-дослідної роботи з філології укладаємо з урахуванням освітніх трендів 4.0: цифровізація освітнього середовища, дослідницьке, проблемно-орієнтоване і

проектне навчання; дотримуємось таких принципів, як принцип особистісного розвитку, інтерактивності, візуалізації, міждисциплінарної інтеграції, м'якої інтеграції інфомедійної грамотності, естетичної вартісності медіатекстів, розвитку критичного мислення.

Наукова новизна

Результати проведеної роботи дали можливість встановити основні методи цифрових лінгвістичних досліджень, які доцільно використовувати для дослідження мовного матеріалу у вищій школі. У групі презентованих методів виділено основні: статистичний, корпусний, лексикографічний метод дослідження, в основі якого теорія лексикографічних систем, та метод штучного інтелекту: машинне навчання, глибоке навчання, нейромережі.

Висновки

Отже, з урахуванням досліджень, які виконують дослідники УМІФ НАН України та освітньо-дослідницького центру «МЕДІА&ВЧИТЕЛЬСЬКИЙ кампус» кафедри української мови і літератури СумДПУ імені А. С. Макаренка, схарактеризовано застосування цифрового методу у викладанні філологічних дисциплін.

Методи лінгвістичних досліджень визначено як учення про способи пізнання й осмислення дійсності й формування внутрішнього рефлексивного досвіду людини, а також про дослідницькі настанови, принципи та процедури аналізу об'єктів мовного середовища. Серед основних методів дослідження цифрової лінгвістичної інформації, які є основою сучасних лінгвістичних досліджень, виділено статистичний метод обробки інформації (математичний), метод корпусних технологій, лексикографічний, штучного інтелекту.

Фахівцями УМІФ НАН України розроблено теоретичні та науково-технічні засади методів дослідження цифрової лінгвістичної інформації.

Статистичний метод дослідження розглядає мову як системно-структурне утворення. Методи статистичної лінгвістики використовують для лінгвістичного моніторингу функціонування мови в конкретному типі дискурсу (політичному, науковому, засобів масової інформації тощо), для контент-аналізу (виявлення стану суспільної свідомості).

Метод корпусних технологій вважаємо за доцільне використовувати на основі вже наявних програмних продуктів як необхідну складову для створення мінікорпусів відкритого типу. Нейромережі розглядаємо як підмножину штучного інтелекту, що зосереджена переважно на проектуванні систем, які дозволяють навчатися та робити прогнози на основі певного досвіду.

Штучний інтелект розглядаємо як форму індивідуалізації систем, якій властивий мовний статус.

Лексикографічний метод вважаємо за доцільне застосовувати для сучасних досліджень лексичної системи мови. Метод спрямований передусім на послідовне виділення та вибіркоче вивчення окремих елементів та їх відношень у мовній системі.

На прикладі віртуальної лексикографічної лабораторії «Мультимедійний словник з інфомедійної грамотності» презентовано застосування цифрового методу для викладання філологічних дисциплін, використаного для оброблення та візуалізації цифрової лінгвістичної інформації. Як результат занять із мовно-методичних дисциплін, студенти поглибили знання про критичне мислення, медійну/ візуальну грамотність, культуру читання медіатексту, розвинули навички протидії фейкам, мові ворожнечі, уміння міжособистісної комунікації та ефективної взаємодії з іншими людьми.

Розроблене програмне забезпечення може стати основою для подальшого впровадження зазначених методів, які стануть важливою складовою розвитку лексикографічних технологій у галузі мовознавства та забезпечить ефективну фахову взаємодію у загальноукраїнському та європейському філологічному просторі.

Література

Інтегрована лексикографічна система «Словники України online». URL: <https://lcorp.ulif.org.ua/dictua/> (дата звернення: 10.11.2022).

Лінгвістично-інформаційні студії (2018). Праці Укр. мов.-інформ. фонду НАН України: в 5 т. Т. 1. Наукова парадигма та основні мовно-інформаційні структури. Київ: Український мовно-інформаційний фонд НАН України.

Семенов, О. & Надутенко, М. (2022). Платформа «Медіа&капсули» як засіб підвищення рівня медіакультури. *Академічні студії. Серія: Гуманітарні науки*, (2), 39–50. URL: <https://doi.org/10.52726/as.humanities/2022.2.6>.

Український національний лінгвістичний корпус (УНЛК). URL: https://svc.ulif.org.ua/UNLC/virt_unlc_4.5/ (дата звернення: 10.11.2022).

- Широков, В., Загнітко, А., Надутенко, М., Надутенко, М. & Семенов, О. та ін. (2022). Віртуальна лексикографічна лабораторія «Мультимедійний словник з інфомедійної грамотності». *Український мовно-інформаційний фонд НАН України*. URL: <https://lcorp.ulif.org.ua/InfoMediaVLL/> (дата звернення: 15.09.2022).
- Широков, В. & Лучик, А. (2021) Парадигмальні засади лінгвістики першої половини XXI століття. *Мовознавство*, 5, 3–163.
- Български национален корпус. Bulgarian National Corpus (BNC). URL: <http://search.dcl.bas.bg/> (дата звернення: 10.11.2022).
- Mintser, O., Babintseva, L., Zaliskyi, V., Nadutenko, M., Kharchenko, N. & Ladychuk, O. (2021). Theoretical approaches to the creation of systemic biomedicine (based on the materials of the report on SRW «System-Biological And System-Medical Regularities Of Development And Course Of Ischemic Heart Disease». *Medical Informatics and Engineering*, 4, 16–72. URL: <https://doi.org/10.11603/mie.1996-1960.2020.4.11889>
- Nadutenko, M., Prykhodniuk, V., Shyrov, V. & Stryzhak, O. (2022). Ontology-Driven Lexicographic Systems. *Advances in Information and Communication*. In: *FICC 2022. Lecture Notes in Networks and Systems*. Cham: Springer, 204–215. URL: https://doi.org/10.1007/978-3-030-98012-2_16
- Semenog, O., Nadutenko, M. & Nadutenko, M. (2021). *VIRTUAL LABORATORY «MULTIMEDIA DICTIONARY OF INFOMEDIA LITERACY»*. *International scientific and practical conference «Philological sciences, intercultural communication and translation studies: an experience and challenges»*. Vol. 1. Czestochowa: «Baltija Publishing». URL: <http://baltijapublishing.lv/omp/index.php/bp/catalog/view/128/3621/7733-1>. (дата звернення: 10.11.2022).
- Stryzhak, O., Prykhodniuk, V., Popova, M., Nadutenko, M., Haiko, S. & Chepkov, R. (2021). Development of an Oceanographic Databank Based on Ontological Interactive Documents. In: *Arai, K. (eds) Intelligent Computing. Lecture Notes in Networks and Systems*, vol 284. Springer, Cham. URL: https://doi.org/10.1007/978-3-030-80126-7_8

References

- Intehrovana leksykohrafichna systema «Slovnkyý Ukrainy online» [Integrated lexicographic system "Dictionaries of Ukraine online"]. URL: <https://lcorp.ulif.org.ua/dictua/> (data zvernennia: 10.11.2022).
- Linhvistychno-informatsiini studii (2018) [Лінгвістично-інформаційні студії.] Pratsi Ukr. mov.-inform. fondu NAN Ukrainy: v 5 t. T. 1. Naukova paradyhma ta osnovni movno-informatsiini struktury. Kyiv: Ukrainskyi movno-informatsiinyi fond NAN Ukrainy.
- Semenoh, O. & Nadutenko, M. (2022). Platforma «Media&kapsuly» yak zasib pidvyshchennia rivnia mediakultury [The "Media&capsules" platform as a means of

- raising the level of media culture]. *Akademichni studii. Serii: Humanitarni nauky*, (2), 39–50. URL: <https://doi.org/10.52726/as.humanities/2022.2.6>.
- Ukrainskyi natsionalnyi linhvistychnyi korpus [Ukrainian national linguistic corpus] (UNLK). URL: https://svc.ulif.org.ua/UNLC/virt_unlc_4.5/ (data zvernennia: 10.11.2022).
- Shyrov, V., Zahnitko, A., Nadutenko, M., Nadutenko, M. & Semenoh, O. ta in. (2022). Virtualna leksykohrafichna laboratoriiia «Multymediinyi slovnyk z infomediinoi hramot-nosti» [Virtual lexicographic laboratory "Multimedia dictionary on infomedia literacy"]. *Ukrainskyi movno-informatsiinyi fond NAN Ukrainy*. URL: <https://lcorp.ulif.org.ua/InfoMediaVLL/> (data zvernennia: 15.09.2022).
- Shyrov, V. & Luchyk, A. (2021) Paradyhmalni zasady linhvistyky pershoi polovyny XXI stolittia [Paradigmatic principles of linguistics of the first half of the 21st century]. *Movoznavstvo*, 5, 3–163.
- Vъlharsky natsyonalen korpus [Bulgarian National Corpus] (BNC). URL: <http://search.dcl.bas.bg/> (data zvernennia: 10.11.2022).
- Mintser, O., Babintseva, L., Zaliskyi, V., Nadutenko, M., Kharchenko, N. & Ladychuk, O. (2021). Theoretical approaches to the creation of systemic biomedicine (based on the materials of the report on SRW «System-Biological And System-Medical Regularities Of Development And Course Of Ischemic Heart Disease». *Medical Informatics and Engineering*. vol 4. pp. 16-72. URL: <https://doi.org/10.11603/mie.1996-1960.2020.4.11889>.
- Nadutenko, M., Prykhodniuk, V., Shyrov, V., & Stryzhak, O. (2022) Ontology-Driven Lexicographic Systems. *Advances in Information and Communication. FICC 2022. Lecture Notes in Networks and Systems*. Cham: Springer, 204–215. URL: https://doi.org/10.1007/978-3-030-98012-2_16
- Semenog, O., Nadutenko, M., & Nadutenko, M. (2021) *VIRTUAL LABORATORY «MULTIMEDIA DICTIONARY OF INFOMEDIA LITERACY»*. *International scientific and practical conference «Philological sciences, intercultural communication and translation studies: an experience and challenges»*: Vol. 1. Czestochowa: «Baltija Publishing». URL: <http://baltijapublishing.lv/omp/index.php/bp/catalog/view/128/3621/7733-1>. (дата звернення: 10.11.2022).
- Stryzhak, O., & Prykhodniuk, V., & Popova, M., & Nadutenko, & Haiko, S., & Chepkov, R. (2021). Development of an Oceanographic Databank Based on Ontological Interactive Documents. *Lecture Notes in Networks and Systems*. Cham : Springer. 97–114 . URL: https://doi.org/10.1007/978-3-030-80126-7_8

Margarita Nadutenko, Maksym Nadutenko, Olena Semenog. Application of the Digital Method in Teaching Philological Disciplines (on the Example of a Virtual Lexicographic Laboratory). The article describes the virtual lexicographic laboratory "Multimedia Dictionary of Infomedia Literacy" as a digital tool for teaching philological disciplines. The article presents the application of digital methods for teaching philological disciplines, using the example of a virtual

lexicographic laboratory. This virtual laboratory is used for processing and visualization of digital linguistic information. Among the digital methods for teaching philological disciplines, we especially mentioned the lexicographic method of research, its meaning, feasibility of application, features of implementation and further use in all-Ukrainian and European cooperation.

The purpose of the proposed work to present the virtual lexicographic laboratory "Multimedia Dictionary of Infomedia Literacy" in the teaching of philological disciplines as one of the main methods for researching digital linguistic information. The laboratory is used to process digital linguistic resources and to create relevant software products within the framework of all-Ukrainian and international cooperation. The scientific novelty of the research lies in the practical proposal of new methods of teaching in higher education.

The results of the work made it possible to establish the main methods of digital linguistic research, which are appropriate to use for the study of language material. In the group of presented methods, the main ones are highlighted: statistical, corpus, lexicographic research method based on the theory of lexicographic systems, and the methods of artificial intelligence: machine learning, deep learning and neural networks.

Key words: digital methods of linguistic research, corpus technologies, multimedia dictionary, theory of lexicographic systems, artificial intelligence, machine learning, neural networks.

Надутенко Маргарита Володимирівна – кандидат філологічних наук, старший науковий співробітник відділу лінгвістики українського мовно-інформаційного фонду Національної академії наук України, керівник Загальноукраїнського центру словникарства, <http://orcid.org/0000-0003-3215-0970>; margo.nadutenko@gmail.com

Надутенко Максим Вікторович – кандидат технічних наук, завідувач відділу інформатики Українського мовно-інформаційного фонду Національної академії наук України

Семенов Олена Миколаївна – доктор педагогічних наук, професор, завідувач кафедри української мови Сумського державного педагогічного університету імені А.С.Макаренка, <http://orcid.org/0000-0002-8697-8602>; olenasemenog@gmail.com