

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЛІНГВІСТИЧНИЙ УНІВЕРСИТЕТ

Ваховська О.В.

ОСНОВИ КОМП'ЮТЕРНОЇ ЛІНГВІСТИКИ

Навчально-методичний посібник

Київ
Видавничий центр КНЛУ
2023

УДК 81'322 : 004 (075.8)
В 22

Друкується за рішенням вченої ради
Київського національного лінгвістичного університету
(протокол № 13 від 27 лютого 2023 р.)

Рецензенти: **Андрущенко І.О.** - кандидат філологічних наук, завідувач кафедри фонетики і практики англійської мови Київського національного лінгвістичного університету;
Жихарєва О.О. - доктор філологічних наук, доцент, доцент кафедри іноземних мов Національної академії образотворчого мистецтва і архітектури;
Присяжнюк Л.Ф. - кандидат філологічних наук, доцент, доцент кафедри іноземних мов Інституту міжнародних відносин Київського національного університету імені Тараса Шевченка.

Ваховська, О.В.

В 22 Basics of computational linguistics / Основи комп'ютерної лінгвістики.
Навчально-методичний посібник / Ваховська О.В. - К.: Видавничий центр КНЛУ, 2023. - 112 с.

ISBN 978-966-638-361-0

© Ваховська О.В., 2023
© Видавничий центр КНЛУ, 2023

ЗМІСТ

Передмова	4
Розділ 1	
Комп'ютерна лінгвістика в інформаційному суспільстві	7
<i>Проектні роботи</i>	29
Розділ 2	
Обробка природної мови: синтаксис і парсинг.....	34
<i>Проектні роботи</i>	38
Розділ 3	
Обробка природної мови: семантика і підходи до її формалізації.....	54
<i>Проектні роботи</i>	57
Розділ 4	
Комп'ютерна лінгвістика і штучний інтелект.....	85
<i>Проектні роботи</i>	86
Завдання для самостійної роботи	90
Післямова	101
Література	103
Нотатки	110

ПЕРЕДМОВА

У навчально-методичному посібнику розглянуто основні поняття комп'ютерної лінгвістики - сучасної галузі знання, метою якої є вивчення природної мови із точки зору комп'ютерного обчислення й розробка артефактів, здатних обробляти і продукувати природну мову в алгоритмічний спосіб, - та оглянуто науково-технічне тло, у яке ці поняття включаються.

Посібник зосереджено, зокрема, на сутності, термінологічному апараті, цілях і методах комп'ютерної лінгвістики, її витоках, сучасному стані і міждисциплінарних зв'язках (Розділ 1), на провідних підходах до автоматичного синтаксичного (Розділ 2) і семантичного (Розділ 3) аналізу природної мови й деяких можливостях їхнього застосування, і на ключових аспектах взаємодії комп'ютерної лінгвістики і штучного інтелекту (Розділ 4), із акцентом на визначальних для цієї взаємодії рушіях. Кожен розділ посібника передбачає поглиблене опрацювання матеріалу, що його попередньо представлено на лекції, із залученням різноманітних завдань і проєктної роботи, задуманих у SMART-форматі як інструменти, за допомогою вибору і поєднання яких можливо сконструювати - згідно з метою і обраним викладачем підходом, а також відповідно до інтересів студентів - навчальний процес як різної форми організації, так і різного рівня складності. Посібник оздоблений завданнями для самостійної роботи студентів над проблемами, що їх окреслено у розділах, і списком літератури й електронних ресурсів.

Перевагою посібника є його вкоріненість в український ґрунт, що її запропоновано студентам у вигляді, по-перше, виконаних нами перекладів фрагментів автентичної англійської фахової літератури на українську мову і, по-друге, дібраної нами інформації насамперед про українські наукові дослідження, інжинірингові розробки і технологічні компанії, які постали криголами у галузі комп'ютерної лінгвістики в Україні, а часом і в світі, і сьогодні продовжують цю галузь розвивати. Студентам пропонується ознайомитися із дотичними роботами працівників провідних українських університетів, зокрема Київського національного лінгвістичного університету, та авторитетних науково-дослідних установ України. Принадою посібника, на нашу думку, є й те, що вміщені у нього завдання і проєктні роботи спрямовані на розвиток у студентів не лише жорстких і цифрових, але й м'яких навичок, як-от планування, самоорганізація, відповідальність (зокрема, за своє навчання), проєктний і тайм-менеджмент, міжособистісна комунікація, емпатія і командна робота, адаптивність до змін, креативність, мистецтво публічних виступів тощо.

Вихідним для навчання вважаємо принцип *Non scholae, sed vitae discimus*, сформульований Сенекою. У курсі комп'ютерної лінгвістики, що його супроводжує посібник, прагнемо втілити цей принцип через орієнтацію на прикладний характер знань, аби зв'язок між отримуваною студентами інформацією і практичною реальністю, в яку студенти зрештою мають - і воліють - інтегруватися, був максимально унаочнений. Зокрема, у викладанні курсу принагідно посилаємося на реалії сучасного українського і міжнародного ринків праці, знайомимо студентів із релевантними кейсами, скеровуємо студентів на усвідомлення ними свого покликання і на той дієвий пошук можливостей професійної самореалізації, який вони можуть здійснити вже сьогодні. У такий спосіб переносимо увагу з академічних результатів як таких на розвиток насамперед індивідуальних умінь і навичок студентів.

Курс, хоча і задуманий як українськомовний, заохочує студентів знати іноземні мови, зокрема англійську, саме яка наразі домінує на ринку ІТ-технологій та у міжнародному корпоративному ландшафті. Ба більше, курс надихає студентів на вивчення мов програмування, знання яких поступово стає таким, що буде прийматися роботодавцями як належне: студенти мають підготуватися до перспектив і викликів (не такого вже і далекого) майбутнього відносно своєї конкурентоздібності і працевлаштування у ньому. Так, справжнім здобутком курсу вважатимемо розширення кругозору, розвиток загального та емоційного інтелекту студентів, їхнього системного, критичного й аналітичного мислення, здатності до навчання, і, головне, мотивацію студентів до саморозвитку, освіти і самоосвіти протягом життя.

Звертаючись до студентів, дозволимо собі із винятковою вдячністю процитувати нині вже покійного професора Петера Боша, у якого свого часу мали нагоду вчитися комп'ютерній лінгвістиці і чийм неодмінним вітальним словом студентам було таке: "This class is part of an academic study, i.e., we make you an offer to learn a few things, and we will support you in doing so. The decision about how you learn and how much you learn is yours."¹ Вітаємо студентів Університету на нашому академічному курсі! Академічну культуру визначаємо як насамперед комунікацію: усну і письмову, - й залучення Вас до цієї комунікації вважаємо одним із пріоритетних завдань курсу.

Ця комунікація спирається на відповідний рівень освіченості, що його Вам пропонує підвищити курс комп'ютерної лінгвістики, і передбачає практику академічних свобод і академічної доброчесності. Так, велику увагу у курсі приділено проєктним роботам, коли у дискусії висловлюються думки, формулюються та обґрунтовуються точки зору. В якості проєктних матеріалів у

¹ "Цей курс є складовою академічного навчання, що означає, що ми пропонуємо Вам навчитися деяким речам і що ми підтримаємо Вас у цьому навчанні. Як Ви вчитиметесь і чому Ви зрештою навчитеся, вирішувати Вам." (Тут і далі переклад наш.)

курс введено наукові статті, робота із якими покликана прилучити Вас до поціновування конвенцій академічного письма і слугувати базою для Ваших власних досліджень. Запрошуємо Вас перейняти погляд на академічну культуру як на спілкування - із підкресленою спільністю дій у досягненні спільної мети. Курс сформувався у своєму нинішньому вигляді саме у спілкуванні і взаємодії із Вашими попередниками - випускниками Університету і студентами старших років навчання. За традицією, курс залишається відкритим для Ваших думок і пропозицій, саме як вітаються цікаві Вам теми й питання і цінується зворотний зв'язок.

Озираючись на той шлях, що цим посібником пройдено від задуму до друку, щиросердо дякуємо за натхнення, конструктив і підтримку шанованим рецензентам й, особно, кафедрі англійської філології і філософії мови, науково-методичній раді і вченій раді Київського національного лінгвістичного університету, в якому здійлася наша праця.

О.В. Ваховська

РОЗДІЛ 1

Комп'ютерна лінгвістика в інформаційному суспільстві

- ▷ Комп'ютерна лінгвістика як наукова й інжинірингова дисципліна
 - Об'єкт і предмет комп'ютерної лінгвістики
 - Цілі і методи комп'ютерної лінгвістики
 - Витоки комп'ютерної лінгвістики
 - Міждисциплінарні зв'язки комп'ютерної лінгвістики
- ▷ Основні поняття комп'ютерної лінгвістики
 - Поняттєвий фундамент комп'ютерної лінгвістики
 - Інформація, інформаційна модель і знання
 - Лінгвістичний алгоритм і лінгвістичний процесор
- ▷ Комп'ютерна лінгвістика в Україні
 - Провідні наукові установи й основні здобутки
 - Витоки, сучасність і перспективи комп'ютерної лінгвістики в Україні
 - Комп'ютерний фонд української мови

*Людське знання виражено у мові.
Отже, комп'ютерна лінгвістика дуже важлива.*

Марк Стідман

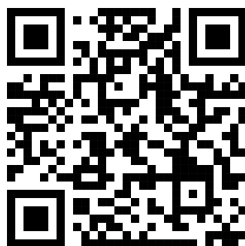
Комп'ютерна лінгвістика визначається як наукова й інжинірингова дисципліна, завданням якої постає розуміння письмової та усної мови із точки зору її (мови) обчислення, а також розробка артефактів, що здатні ефективно обробляти і продукувати мову, у контексті як великих даних, так і окремого діалогу. Оскільки мова є дзеркалом мислення, розуміння того, як мова обчислюється, дає і розуміння того, як влаштовані людська думка і інтелект. А також, оскільки мова є природним і універсальним засобом спілкування людей, компетентні у лінгвістичному відношенні комп'ютери сприятимуть кращій взаємодії людини й техніки і різноманітного програмного забезпечення, та нададуть людині надійний доступ до безмежних текстових, й не лише, ресурсів Інтернету (Schubert, 2020).

Сутність комп'ютерної лінгвістики розкривається у тих характеристиках цієї дисципліни, що їх протягом курсу розглядається на лекціях і обговорюється під час виконання різних типів завдань на семінарах.

▷ Комп'ютерна лінгвістика як наукова й інжинірингова дисципліна



Завдання 1. Уважно прочитайте розділ “Комп’ютерна лінгвістика - новий етап розвитку теоретичної та прикладної лінгвістики” у підручнику (Карпіловська, 2006, с. 7-20).²



Створіть ментальну мапу³ інформації, що її подано у розділі. Сформулюйте для себе *a take-home message*⁴ із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.



Завдання 2. Уважно прочитайте визначення комп'ютерної лінгвістики.⁵

Комп'ютерна лінгвістика - це ...

... вивчення мови у термінах алгоритмів (Dunlop, & Fetzer, 1993).

² Завдання базового рівня складності.

³ Студентам пропонується ознайомитися із принципами і підходами до створення ментальних мап, а також із відповідними (онлайн-)інструментами.

Ми вдячні професору Світлані Анатоліївні Жаботинській за натхнення залучити практику - і інструменти - створення ментальних мап у викладання. Переконані, що ментальні мапи володіють надзвичайно потужним евристичним потенціалом. Із спілкування з нашими випускниками, знаємо, що багатьом із них саме набуті в Університеті досвід і вміння створювати ментальні мапи відкрили перші двері у світ ІТ-професій, саме як і у світ професійного викладання іноземних мов, де і досі повсюдно ними використовуються. Створення ментальних мап, зрештою, навчає і привчає структурувати інформацію в ефективний спосіб, що в інформаційну епоху набуває особливої ваги.

⁴ *A take-home message* (англ.) розуміємо як “основний посил, основне повідомлення, основна інформація, що їх людина виносить для себе із будь-якого досвіду, зокрема із навчання; висновок” і на семінарах використовуємо як фасилітаційний інструмент.

⁵ Завдання середнього рівня складності.

... наукове вивчення мови із обчислювальної точки зору (Gallego, & Martin, 2020).

... застосування комп'ютерної науки при аналізі і задля розуміння письмової і усної мови (Gillis, 2022).

... застосування комп'ютерної науки до аналізу, синтезу і розуміння письмової і усної мови (Hunston, 2022).

... застосування лінгвістичних теорій і обчислювальних технік до проблем обробки природної мови (Bolshakov, & Gelbuch, 2004).

... наука про мову, де особлива увага приділяється тим обмеженням, які накладаються на складність обробки природної мови особливостями архітектури людського мислення. Як і більшість інших наук, комп'ютерна лінгвістика має інжинірингові застосування (Hausser, 2014).

... вивчення комп'ютерних систем із метою зрозуміти і генерувати природну мову (Grishman, 1986).

Чому, на Вашу думку, комп'ютерна лінгвістика має ці різні визначення? Чи є вони доцільними?

Чи можна сказати, що серед термінів із Завдання 5 (див. нижче) деякі більше, ніж інші, відповідають окремим визначенням із цього завдання?

Знайдіть у цих визначеннях інваріант vs. варіанти. За потреби, доповніть список визначеннями комп'ютерної лінгвістики, що їх подано у вподобаній Вами фаховій літературі.

Сформулюйте власне - на Ваш погляд, необхідне і достатнє - визначення комп'ютерної лінгвістики.



Завдання 3. Уважно прочитайте уривок із першоджерела:⁶

⁶ Завдання базового рівня складності.

Теоретичні цілі комп'ютерної лінгвістики полягають у створенні таких підходів до граматики і семантики природної мови, які уможливили б обчислення цієї мови із застосуванням **методів синтаксичного і семантичного аналізу**, що є здійсненними практично; виявлення і пояснення механізмів обробки природної мови у мозку і мисленні людини, а також формулювання принципів навчання мови, які охопили б її структурні і дистрибуційні (статистичні) властивості; створення обчислювальних моделей обробки і навчання мови: оскільки ці процеси здійснюються у мозку і мисленні людини, створювані комп'ютерними лінгвістами моделі мають бути нейрофізіологічно і когнітивно правдоподібними.

Практичні цілі комп'ютерної лінгвістики є дуже різноманітними. Деякими із основних є ефективний текстовий пошук за заданою тематикою; ефективний машинний переклад; відповіді на запитання, які включають як прості запитання про певні факти, так і запитання, що вимагають умовиводів і дескриптивних або дискурсивних відповідей (можливо, відповідей із обґрунтуваннями); реферування текстів; аналіз теми, тональності або інших психологічних характеристик текстів, а також усного мовлення; створення діалогових помічників і діалогових систем задля вирішення окремих завдань (здійснення покупок, надання технічного супроводу, організація подорожей, створення розкладу, надання медичної консультації тощо); і, в підсумку, створення комп'ютерних систем, чий механізм засвоєння мови, володіння мовою у діалозі й набуття знань із тексту дорівнювали б тим, які наявні у людей (Schubert, 2020).

Схарактеризуйте своїми словами основні теоретичні і практичні цілі комп'ютерної лінгвістики.

Наведіть приклади реалізації практичних цілей комп'ютерної лінгвістики із свого повсякденного життя. Підсумуйте, наскільки Ви як користувач задоволені досвідом своєї взаємодії із окремими інструментами реалізації цих цілей, що їх пропонує ринок сучасних технологій. У чому Ви як фахівець вбачали б можливість вдосконалення цих інструментів?



Завдання 4. Схарактеризуйте коротко підходи до обробки і представлення інформації, зокрема, тієї, що об'єктивована у мові.⁷

⁷ Завдання високого рівня складності.

Аналогові підходи (Analogical approaches)
Конекціоністські підходи (Connectionist approaches)
(Формальні) Логічні підходи ((Formal) Logical approaches)
Модульні підходи (Modular approaches)
Корпусні, статистичні підходи (Corpus-based, statistical approaches)
Підходи на основі нейронних мереж (Neural-network approaches)
Підсимвольні підходи (Subsymbolic approaches)
Процедурні підходи (Procedural approaches)
Символьні підходи (Symbolic approaches)

Підходи подані в алфавітному порядку. Згрупуйте їх за провідними принципами. Намалюйте стрілу часу для домінування цих підходів у когнітивній науці взагалі й у комп'ютерній лінгвістиці зокрема.

Якщо вважаєте доцільним, на основі фахової літератури доповніть список. Схарактеризуйте коротко підходи, що Ви їх додали.



Завдання 5. У сучасній англійській фаховій літературі на позначення комп'ютерної лінгвістики як дисципліни використовуються такі терміни:⁸

Computational Linguistics
Natural Language Processing
Human Language Technology
Engineering Linguistics
Language Engineering

На Вашу думку, чи є ці терміни тотожними у плані їхнього змісту? Які аспекти комп'ютерної лінгвістики висвітлює кожен із них?

Знайдіть еквіваленти цих термінів в українській мові. Ви задоволені наявним перекладом? Якщо так, обґрунтуйте. Якщо ні, запропонуйте власний переклад.

⁸ Завдання високого рівня складності.



Завдання 6. На основі авторитетних джерел, підготуйте коротку інформацію про такі системи:⁹

ENIAC - Mauchly, Eckert et al., 1946

МЕЛІМ - Лебедев, 1951-1953

Sad Sam - Lindsay, 1963

ELIZA - Weizenbaum, 1966

Sir - Raphael, 1968

Student - Bobrow, 1968

CYRUS - Kolodner, 1970

Lunar - Woods et al., 1972

PARRY - Colby, 1972

SHRDLU - Winograd, 1972

HACKER - Sussman, 1973

BORIS - Dyer, 1980

Відстежте еволюцію підходів до створення, використовуваних у цих системах.
Визначте спільні й відмінні риси цих систем.



Завдання 7. Вкажіть внесок цих особистостей у становлення і розвиток когнітивної науки, зокрема комп'ютерної науки і комп'ютерної лінгвістики. Якщо вважаєте доцільним, доповніть список:¹⁰

Л. Баум (Leonard Baum)

Ч. Бебідж (Charles Babbage)

Е. Бут (Andrew Booth)

Н. Вінер (Norbert Wiener)

Т. Віноград (Terry Winograd)

С. Возняк (Stephen Wozniak)

С. Джобс (Steven Jobs)

Р. Квіліан (Ross Quillian)

А. Лавлейс (Ada Lovelace)

⁹ Завдання базового рівня складності.

¹⁰ Завдання базового рівня складності.

Сергій Олексійович Лебедєв
М. Мінський (Marvin Minsky)
А. Тьюринг (Alan Turing)
У. Уївер (Warren Weaver)
Р. Хартлі (Ralph Hartley)
Н. Чомські (Noam Chomsky)
К. Шеннон (Claude Shannon)

Прізвища у списку розташовані за алфавітом. Намалуйте стрілу часу, яка б унаочнювала послідовність розробок і праць цих особистостей в еволюції науки і техніки.



Завдання 8. На основі фахової літератури, дайте визначення:¹¹

Інформатика
Когнітивна лінгвістика
Когнітивна наука
Комп'ютерологія (Комп'ютерна наука)
Корпусна лінгвістика
Квантитативна лінгвістика
Лінгвостатистика
Логіка
Математика
Математична лінгвістика
Нейролінгвістика
Обчислювальна лінгвістика
Прикладна лінгвістика
Психолінгвістика
Структурна лінгвістика
Штучний інтелект

Терміни у списку розташовані за алфавітом. Розташуйте їх у вигляді семантичної мережі, подайте у вигляді поняттєвої схеми або ментальної мапи тощо за міждисциплінарними зв'язками.

¹¹ Завдання середнього рівня складності.

Зважаючи на предмет курсу, чому у списку наведені саме ці терміни, а не інші? Як позначені ними дисципліни взаємодіють із комп'ютерною лінгвістикою? Якщо вважаєте доцільним, доповніть список.

Знайдіть відповідники цим термінам в англійській мові. Чи представився Вам цей пошук простим і нехитрим?



Завдання 9. На Вашу думку, комп'ютерна лінгвістика є більше лінгвістикою чи комп'ютерологією? Чому?¹²

Визначте місце комп'ютерної лінгвістики у структурі наук (про мову).



Завдання 10. Протягом навчання в Університеті, Ви вивчаєте “Вступ до мовознавства.” На Вашу думку, чому зараз ми із Вами вивчаємо основи комп'ютерної лінгвістики, а не основи комп'ютерного мовознавства?¹³

Як, на Ваш погляд, співвідносяться, або можуть співвідноситися, терміни “мовознавство” і “лінгвістика”? Зважте на англійські відповідники термінів.

¹² Завдання базового рівня складності.

¹³ Завдання базового рівня складності.

▷ Основні поняття комп'ютерної лінгвістики

До основних понять комп'ютерної лінгвістики, що із ними пропонується ознайомитися студентам, належать:¹⁴

Алгоритм (Algorithm) - абсолютно надійна послідовність дій або процедур, яку можна виконати кінцевою кількістю кроків, аби вирішити проблему. Алгоритми інколи також називаються ефективними процедурами.

Гомункул (Homunculus) - буквально, “маленька людина,” уявлення про гомункулуса закладено у теоріях мислення, які посилаються на одні когнітивні процеси у своїх поясненнях інших когнітивних процесів. Наприклад, інколи стверджується, що зорове сприйняття передбачає утворення у мисленні образів. Якщо це твердження правильне, то слід було б передбачити, що на ці образи як на внутрішні картини має дивитися певний внутрішній споглядач - гомункулус, але тоді слід було б також передбачити, що і у гомункулуса має бути у мисленні свій гомункулус, який і дивиться на ці картини, і так до безкінечності. Див. також (Кришталь, 2022).

Глибинна структура (Deep structure) - рівень аналізу в *трансформаційній граматиці* (transformational grammar), протиставлений *поверхневій структурі* (surface structure), розрізнення яких було запропоноване Н. Чомскі. Відношення між двома рівнями виявляється у трансформаційній граматиці, за якою поверхнева структура є похідною від глибинної структури. Як вказує термін “глибинна структура,” цей рівень аналізу не відповідає поверхневій структурі конкретного речення безпосередньо, але дозволяє дійти до певних узагальнень стосовно прихованих подібностей у глибинній структурі низки речень, які поверхнево відрізняються одне від одного.

Грамматика структури синтаксичних груп (Phrase structure grammar) - грамматика, яка передбачає групування синтаксичних компонентів за ієрархічним принципом. Так, наприклад, речення може бути розкладене на *іменникову групу* (noun phrase) і *дієслівну групу* (verb phrase), які, за принципами *ієрархії мов за Н. Чомскі* (Chomsky hierarchy), будуть розкладені у свою чергу на менші синтаксичні компоненти.

¹⁴ Якщо не зазначено інакше, терміни, що ними позначаються основні поняття комп'ютерної лінгвістики, у цьому підрозділі визначено услід за (Dunlop, & Fetzer, 1993).

Дерево (Tree) - ієрархічна структура, яка складається із одного **вузла** (a node) (термінологічно позначеного як **кореневий вузол** (the root node)), до якого сходять численні **гілки** (branches), що зв'язують цей вузол із іншими вузлами цього дерева. <...> Використовується для представлення граматичної структури речення; для представлення можливих вирішень однієї проблеми шляхом розбиття цієї проблеми на підпроблеми тощо.

Дієслівна група (Verb phrase, VP) - одиниця граматичного аналізу, яка у найпростішому вигляді складається лише із дієслова або із дієслова і додатка (представленого як **іменникова група** (noun phrase)). Прикладами дієслівних груп є “eat” і “played the banjo.” Один із рівнів аналізу у **граматиці структури синтаксичних груп** (phrase structure grammar).

Експертна система (Expert system) - система штучного інтелекту, яка демонструє рівень володіння специфічним знанням на рівні людини-експерта. Експертна система складається із трьох основних компонентів: **бази знань** (knowledge base), зібраної зі знань людей-експертів певної галузі шляхом процесу, що зветься інженерія знань; набір **правил для прийняття рішень** (decision rules), які визначають, що саме системі слід зробити в певних умовах; і **механізму логічного висновку** (inference engine), який застосовує по відношенню до бази знань відповідні умовам правила.

Ієрархія мов за Н. Чомскі (Chomsky hierarchy) - упорядкування типів граматик, яке відображає спектр формальних мов, що їх може генерувати кожна така граматика. У цьому упорядкуванні, мови, що їх генерують слабші граматики, є підмножинами тих мов, що їх генерують сильніші граматики. В ієрархії виокремлюються чотири рівня, позначені як 0,1,2 і 3, де рівень 0 репрезентує найсильнішу граматику, а рівень 3 - найслабшу. Кожен рівень в ієрархії точно відповідає окремому типу **автомату** (automaton).

Іменникова група (Noun phrase, NP) - одиниця граматичного аналізу, яка складається із займенника або детермінатива (необов'язково) плюс із одного або двох прикметників (необов'язково) плюс іменника. Прикладами іменникових груп є “it,” “she,” “frog,” “a slimy frog” і “the beautiful old guitar.” Один із рівнів аналізу у **граматиці структури синтаксичних груп** (phrase structure grammar).

Інтелект (Intelligence) - інтелект традиційно визначається як здатність істот, наділених мисленням, вчитися. Коли ця здатність оцінюється відносно хронологічного віку людини за використання стандартизованих тестів, то результатом такої оцінки постає певне число як **коефіцієнт інтелекту** (intelligence quotient). Виникнення терміну “штучний інтелект” свідчить про те, що машини, хоча і не наділені мисленням, теж можуть описуватися у термінах інтелекту, тому що машини успішно і передбачувано стабільно виконують певні завдання - особливо такі завдання, вирішення яких зазвичай вимагає залучення людей.

Інтерлінгва (Interlingua) - система символів для представлення значень речень природної мови; використовується у деяких системах обробки природної мови. У межах цього підходу, речення із значно схожими значеннями представляються однаково, навіть якщо їхні поверхневі форми значно відрізняються одна від одної. Наприклад, речення “A threw the ball to B,” “The ball was thrown by A to B” і “B caught the ball that A threw” виражають одну й ту саму головну ідею, що *м'яч був переданий від А до В у просторі*. Використання такого виразу, поданого курсивом, у якості спільної інтерлінгви для цих трьох речень полегшує завдання їхнього схемного представлення. Так, використання єдиної інтерлінгви для представлення значень речень будь-якої природної мови спрощує завдання машинного перекладу, оскільки замість правил, за якими одна мова має бути співвіднесена з іншою, машинний перекладач, базований на інтерлінгві, вимагає лише правил перекладу цих мов на інтерлінгву. Тоді будь-які дві природні мови можуть бути перекладені одна на одну за використання інтерлінгви як їхнього єдиного проміжного представлення. Штучна мова Есперанто була запропонована у якості інтерлінгви для міжнародної комунікації.

Китайська Кімната (Chinese Room) - приклад, наведений Джоном Серлем як ілюстрація його твердження, що тест Тьюринга є неадекватним критерієм для визначення наявності справжнього мислення. У Китайській Кімнаті знаходиться людина, яка не знає китайської, але у якої наявний повний і витончений набір правил поєднання символів, завдяки чому ці символи, хоча і залишаються без значень, притаманних їм у китайській мові, можна комбінувати у граматично правильні для китайської мови послідовності. Людина отримує зовні символи, які вона у Кімнаті комбінує за правилами й потім повертає назовні. Людина цього не знає, але отримувані символи вона комбінує у граматично правильні відповіді на питання до розповіді,

що її людина також не знає, - цю розповідь знають люди назовні: вони знають китайську і на основі відповідей людини із середини приходять до хибного висновку, що ця людина вільно володіє китайською. Уподібнюючи те, що відбувається у Китайській Кімнаті, до того, що відбувається у комп'ютері із його програмним забезпеченням, Серл стверджує, що оперування символами за правилами, але без знання значень цих символів, не є інтенціональним, а, отже, і не дає підстав для суджень про наявність чи відсутність справжнього мислення.

Комбінаторний вибух (Combinatorial explosion) - різке зростання кількості можливих варіантів аналізу синтаксичної структури речення або його окремих частин при комп'ютерній обробці цього речення.

Лексична субституція (Lexical substitution) - заміна кінцевих вузлів синтаксичного дерева словами мови.

Лематизація (Lemmatization) (від лат. *lemma* "словникова форма") - процедура комп'ютерної і корпусної лінгвістики, що відновлює словникову форму за її словоформою.

Лінгвістичний процесор (Natural language processor) - складна система автоматичної обробки природної мови, здатна до аналізу і синтезу текстів, що забезпечує діалог у системах штучного інтелекту.

Машина Тьюрінга (Turing machine) - абстрактний прилад, який складається із безкінечної "стрічки," розділеної на комірки, кожна із яких може містити лише або 0, або 1. Сканер працює над стрічкою, зчитуючи вміст однієї комірки за один раз і виконуючи набір серійних операцій, кожна з яких сумарно залежить від вмісту цієї конкретної комірки і від стану машини у цей конкретний момент її оперування. Програма машини визначає, що саме машина має зробити у кожен момент її оперування, - наприклад, "Якщо зчитано символ 0 і в цей час машина знаходиться у стані 6, тоді 0 слід замінити на 1 і переключитися у стан 4."

Машинна мова (Machine language) - мова, на яку безпосередньо реагує комп'ютер, на відміну від **мов вищого рівня** (higher-level languages) (Pascal, BASIC, Fortran тощо), які повсюдно використовуються програмістами. Представлені двійковими числами, версії програм вищого

рівня, написані машинною мовою, створюються за допомогою *компілятора* (a compiler) або *інтерпретатора* (an interpreter).

Метамова (Metalanguage) - “засіб спілкування людей, який виник в еволюції людства як засіб передачі почуттів. У мистецтві стільки мистецтва, скільки у ньому метамови. Проте терміни не ідентичні, існує потреба в обох. Метамова - емоційне тло життя, а мистецтво - засіб, щоб тло прикрасити і зробити спільним” (Кришталь, 2020).

Мова думки (Language of thought) - гіпотетична вроджена схема кодування, яка пропонує достатньо можливостей для вираження будь-якої відмінності об'єктів світу, що її взагалі може описати будь-яка природна мова. У сучасній когнітивній науці, ця ідея належить Джеррі Фодору, який стверджує, що така вроджена мова (інколи її називають *менталіз* (mentalese)) є логічно необхідною задля вивчення природної мови, адже дитина, яка опиняється у своєму першому мовному середовищі і починає вчити свою першу природну мову, має почати формулювати припущення про те, до яких об'єктів відносяться мовні вирази, але задля формулювання таких припущень вже потрібна якась попередньо наявна мова. Аби уникнути безкінечної регресії, потрібно постулювати наявність вродженої мови думки.

Мова програмування (Programming language) - набір конструктив *формальної мови* (formal language), які уможливають запис *алгоритмів* (algorithms) у вигляді *програм* (programs), тобто із використанням словника і структури, на які безпосередньо може реагувати комп'ютер. Інколи розрізняються *мови програмування високого і низького рівнів* (high- vs. low-level programming languages). Мови програмування високого рівня, як правило, мають великий спектр конструктив, які можуть у достатній мірі відобразити більшість категорій людської думки і зв'язків між цими категоріями. Такі мови достатньо легко запам'ятовуються і використовуються. Тим не менш, мови програмування високого рівня не повністю відображають ті процеси, що відбуваються у комп'ютері, і тому мають перекладатися *компілятором* (compiler) або *інтерпретатором* (interpreter) в *об'єктний код* (object code), який може використовувати комп'ютер. На відміну від мов високого рівня, мови програмування низького рівня достатньо точно відображають принципи роботи комп'ютера, але їхня невідповідність категоріям і процесам людської думки робить їх значно менш інтуїтивними.

Модель (Model) - спрощений об'єкт (система, процес тощо), що зберігає лише важливі властивості справжнього об'єкта (системи, процесу тощо), який моделюється і саме для вивчення якого створюється модель: дослідження моделі є засобом отримати інформацію про модельований об'єкт.

Комп'ютерні лінгвісти прагнуть створити **обчислювальні моделі** (computational models) різноманітних мовних явищ. Побудова цих моделей залучає правила (такі моделі визначаються як knowledge-based, rule-based або hand-crafted) та/або мовні дані (такі моделі є data-driven, statistical або empirical). Особливості обох типів моделей, услід за (Gallego, & Martin, 2020), можна узагальнити у таблиці:

<i>Моделі, базовані на правилах</i>	<i>Моделі, базовані на даних</i>
Кодування знання про мову, експліцитне	Кодування знання у мові, імпліцитне
Залучення набору створених або відібраних вручну граматичних правил	Залучення статистичних методів або методів машинного навчання
Легкі тестування і відладка	Дещо ускладнені тестування і відладка
Потребують значної участі людини	Потребують не такої значної участі людини
Часто базуються на обмеженому огляді даних, акцент на прототипних прикладах	Повністю керовані даними і вимагають великих даних
Часто не покривають у достатній мірі модельовану царину	Ступінь покриття модельованої царини прямо пропорційна обсягу даних
Часто втрачають ефективність при роботі із зашумленими даними	Легше адаптуються до зашумлених даних

Навчання (Learning) - (а) Будь-яка зміна у поведінці суб'єкта, що її викликає попередній досвід. (б) Самоадаптація суб'єкта до покращеної продуктивності у відповідь на зміну середовища і обставин. Це

покращення зазвичай пояснюється через набуття нових знань і здатність використовувати ці знання.

Неоднозначність (Ambiguity) - якщо слово, фраза або речення мають більше, ніж одну можливу інтерпретацію, то про них говорять, що вони є неоднозначними. Неоднозначність може виникати із різних джерел. **Синтаксична неоднозначність** (Syntactic ambiguity) виникає тоді, коли одна й та сама лінгвістична структура припускає кілька граматичних інтерпретацій - наприклад, різні інтерпретації фрази “old men and women” (Жінки теж були старими?) і речення “I greeted the woman with the flowers” (Квіти були у неї, чи у мене?) <...> У випадках **семантичної неоднозначності** (semantic ambiguity) слово, фраза або речення можуть інтерпретуватися як такі, що мають різні значення, і вимагають контексту для усунення такої неоднозначності - наприклад, іменник “an organ” на позначення частини тіла і музичного інструменту¹⁵ або речення “Everybody isn't here,” яке можна інтерпретувати як те, що не всі тут, або як те, що всі не тут.

Об'єктний код (Object code) - програма *машинною мовою* (machine language), створена компілятором із програми, написаної мовою вищого рівня, наприклад, мовою Pascal. Програма об'єктного рівня виконує інструкції, які були визначені в програмі вищого рівня, із якої вона була створена, але інструкції при цьому перекладаються на машинну мову, саме безпосередньо на яку може відреагувати комп'ютер.

Обробка інформації (Information processing) - введення, виведення, зберігання, пошук і оперування символами, здійснювані за допомогою алгоритмів.

Обробка природної мови (Natural language processing) - аналіз генерування і розуміння речення із урахуванням процесів його обчислення. Кінцева мета цього підходу у когнітивній науці - побудувати комп'ютерну модель, яка відображала б усі ті процеси обробки мови, що використовуються людьми, при тому, що наразі повне розуміння цих процесів у науці відсутнє.

Обчислення (Computation) - систематична трансформація вхідних даних у вихідні дані, при цьому зміст цих даних зберігається.

¹⁵ Див. також (Vakhovska, 2022), де, зокрема, здійснено аналіз мережі полісемії іменника *an eye*, який у сучасній англійській мові має 44 лексико-семантичних варіанти.

Онтологія (Ontology) - в інжинірингу знань (knowledge engineering), онтологія визначається як артефакт, що у межах певної царини знань специфікує об'єкти (або поняття про ці об'єкти) і зв'язки між ними, а також властивості цих об'єктів (або понять про них) (Peters, & Shrobe, 2003).

Парсер, Лінгвістичний аналізатор (Linguistic parser) - комп'ютерна програма, що аналізує поверхневу синтаксичну структуру речення із метою виявити глибинну синтаксичну структуру цього речення.

Поверхнева структура (Surface structure) - рівень аналізу у *трансформаційній граматиці* (transformational grammar), протиставлений *глибинній структурі* (deep structure) (розрізнення було запроваджене Н. Чомскі). На рівні поверхневої структури, спостерігається пряма відповідність між реченням і його граматичним представленням. Відношення між поверхневою і глибинною структурами речення виявляється трансформаційною граматикою.

Прийменникова група (Prepositional phrase, PP) - мовна структура, яка складається із прийменника і іменникової групи. Одна із одиниць аналізу у *граматиці структури синтаксичних груп* (phrase structure grammar).

Природна мова (Natural language) - будь-яка конвенційна система символів і правил їхнього поєднання, яка виникла природним шляхом, на відміну від *штучних мов* (artificial languages), що виникають внаслідок задуму і свідомої роботи людей, які ці мови створюють. Символи у цій системі можуть бути письмовими, усними або знаковими (у мовах, які використовуються для спілкування з глухими). <...> Природні мови зазвичай слугують широкому спектру комунікативних потреб і тяжіють до багатства і різноманіття вираження, але не до його ефективності та економності.

Семантична мережа (Semantic network) - структура представлення знань (knowledge representation), яка складається із *вузлів* (nodes), які представляють слова або поняття, і *зв'язків* (links), які представляють відношення між словами або поняттями.

Сильний штучний інтелект (Strong AI) - твердження про те, що, у відповідних умовах (наприклад, при запуску вірної програми), комп'ютер і

дійсно отримав би розум; тобто, комп'ютер мав би здатність до мислення у неметафоричному смислі.

Символьний штучний інтелект (Symbolic AI) - підхід до штучного інтелекту, який наголошує важливість правил для оперування символами різного типу. <...> Відображає думку тих когнітологів, які вважають, що людський інтелект виникає внаслідок оперування символами, яке здійснює людський мозок. Протиставляється **конекціонізму** (connectionism).

Синтаксичний аналіз, або **парсинг** (Parsing) - процес аналізу граматичної структури речення із метою її визначення. Вирізняють **синтаксичний аналіз знизу догори** (bottom-up parsing) і **синтаксичний аналіз згори донизу** (top-down parsing). Метою обох підходів є аналіз структури речення відносно набору граматичних правил, але, як свідчать самі терміни, один підхід є протилежним іншому. Наприклад, речення "Pat ate some strawberries" має таку граматику (символ "→" має значення "складається з"):

Sentence → *Noun Phrase* + *Verb Phrase*

Noun Phrase → *Noun*, or *Article* + *Noun*

Verb Phrase → *Verb* + *Noun Phrase*

Відповідний словник включає наступні статті, із відповідними категоріями слів:

Pat: *Noun*

strawberries: *Noun*

ate: *Verb*

some: *Article*

Словник показує, що речення "Pat ate some strawberries" складається із *Noun* + *Verb* + *Article* + *Noun*. Аналіз знизу догори починає із цих одиниць і намагається знайти структури вищого рівня (в кінцевому підсумку *Sentence*), що їх утворюють ці одиниці. Цього можна досягти, якщо зчитувати граматичні правила справа наліво. Так, ця процедура визначає *Article* + *Noun* як складники *Noun Phrase*; отже, початкове *Noun* + *Verb* + *Article* + *Noun* можна переписати у вигляді структури вищого рівня *Noun* + *Verb* + *Noun Phrase*. Послідовне (справа наліво) використання цих саме правил зрештою дає структуру найвищого рівня *Sentence*. Отже, речення "Pat ate some strawberries" є прийнятним згідно із цією граматиною.

Для порівняння, аналіз згори донизу починається із структури найвищого рівня *Sentence* і намагається розкласти цю структуру на складники так, аби отримати вихідне речення. Починаючи із *Sentence* і

використовуючи правила зараз вже зліва направо, отримуємо *Noun Phrase* + *Verb Phrase* і так далі, аж поки не буде отримано речення “Pat ate some strawberries,” яке є граматично прийнятним згідно із цими правилами. Утім, при цьому можна отримати і інші речення, наприклад “Strawberries ate some strawberries,” які, хоча і будуть абсурдними, але все ж таки відповідатимуть заданій граматиці. Граматична правильність не дорівнює правильності змістовній.

Обидва підходи - знизу догори і згори донизу - мають використовуватися залежно від мети аналізу і від того, яке саме речення аналізується. Інколи ці два підходи комбінуються між собою.

Слабкий штучний інтелект (Weak AI) - твердження про те, що комп'ютери є корисними і дієвими інструментами для вивчення мислення, наприклад, шляхом формулювання і перевірки гіпотез про механізми обробки мови або роботи пам'яті. Утім, на відміну від **сильного штучного інтелекту (strong AI)**, слабкий штучний інтелект не припускає, що комп'ютери буквально є (або могли б бути) мислячими суб'єктами.

Теорія інформації (Information theory) - заснована Клодом Шенноном, теорія інформації є математичним підходом до вивчення комунікації. Цей підхід наголошує кількість інформації в одиницях (біт - базова одиниця вимірювання інформації), а не її зміст, і уможливорює цифрове представлення цих одиниць. Наприклад, якщо збирається інформація про автомобілі, то категорії старі vs. нові автомобілі, автомобілі вітчизняного vs. іноземного виробництва, великі vs. невеликі автомобілі можуть бути представлені двійковим числом: старі = 0 і нові = 1, вітчизняні = 0 і іноземні = 1, великі = 0 і невеликі = 1, і тоді інформація про будь-який автомобіль може бути представлена трьома бітами, а, наприклад, старий вітчизняний невеликий автомобіль отримав би код 001.

Тест Тьюрінга (Turing test) - мисленнєвий експеримент; операційний тест, запропонований А. Тьюрінгом з метою надати відповідь на питання “Чи може машина думати?” Тест включає людину, яка ставить різноманітні питання на будь-яку тему чоловікові і жінці, які знаходяться в ізольованому від цієї людини приміщенні; комунікація здійснюється через систему передачі повідомлень. Спочатку, завданням цієї людини є визначити, хто із двох відповідачів є чоловіком, а хто - жінкою. Потім, чоловіка замінює комп'ютер. Якщо людина, яка ставить запитання, продовжує вірити, що обидва відповідачі - люди, то робота комп'ютера

точно відповідає роботі людини, і тоді слід визнати, що комп'ютер здатен думати. Відповідність тесту Тьюринга меті визначити наявність мислення була поставлена під сумнів Джоном Серлем із його відомим антиприкладом під назвою *Китайська Кімната* (Chinese Room).

Трансформаційна граматики (Transformational grammar) - запропонована і розвинена Н. Чомскі граматики, яка залучає два рівні репрезентації: глибинну структуру і поверхневу структуру. Сутність цієї гіпотези полягає у тому, що продукування речень природною мовою є результатом деривацій від глибинних структур, які лежать в основі мови і які мовець не усвідомлює.

Універсальна граматики (Universal grammar) - запропонована Н. Чомскі, універсальна граматики є набором обмежень, яким має відповідати граматики будь-якої природної мови. На думку Н. Чомскі, надзвичайно мала кількість мовних даних, які отримує дитина під час засвоєння мови, має бути пояснена існуванням *вродженої мовної бази* (innate linguistic framework). <...> Із цієї точки зору, усі природні мови (як існуючі, так і можливі) мають у своїй основі певні спільні риси.

Формальна мова (Formal language) - специфікація набору *правильно оформлених формул* (well-formed formulae), шляхом визначення (1) набору символів і (2) набору правил, які застосовуються для допустимого, тобто граматично правильного, комбінування цих символів.

Штучна мова (Artificial language) - будь-яка система запису і правила її використання, яка була спеціально створена із особливою метою, що не є метою звичайної комунікації між членами мовної спільноти. До прикладів належать системи формальних мов у теоретичній математиці (як неінтерпретовані системи), мови програмування у комп'ютерній науці (як інтерпретовані системи) і Есперанто (людська мова загального призначення, покликана полегшити міжнародну комунікацію). На деякі окремі випадки (як, наприклад, азбука Морзе) можна дивитися як на граничні між природними і штучними мовами. Протиставляється *природній мові* (natural language).

Штучний інтелект (Artificial intelligence) - розділ комп'ютерної науки, який зосереджується на тому, наскільки ментальні можливості людських істот можуть бути відтвореними за допомогою машин. Дослідження зі

штучного інтелекту є формуючими для міждисциплінарної царини когнітивної науки у цілому.

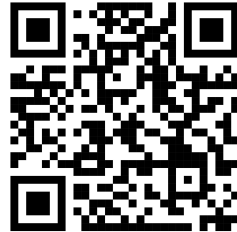


Завдання 11. Подані у цьому підрозділі терміни розташовані за алфавітом. Розташуйте їх у вигляді (фрагментів) семантичної мережі, (фрагментів) поняттєвої схеми або (фрагментів) ментальної мапи тощо за поняттєвими зв'язками.¹⁶

¹⁶ Завдання високого рівня складності. Можливе виконання цього завдання у вигляді підсумкової проєктної роботи для структурації всієї вивченої інформації.



Завдання 12. Уважно прочитайте розділ “Інформаційні моделі лінгвістичних об’єктів” у підручнику (Карпіловська, 2006, с. 20-34).¹⁷



Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.



Завдання 13. Уважно прочитайте розділ “Лінгвістичний алгоритм та лінгвістичний процесор” у підручнику (Карпіловська, 2006, с. 50-54).¹⁸



Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.

¹⁷ Завдання базового рівня складності.

¹⁸ Завдання базового рівня складності.

▷ Комп'ютерна лінгвістика в Україні



Завдання 14. Уважно прочитайте статтю “Комп'ютерна лінгвістика” в Енциклопедії Сучасної України (Карпіловська, 2014).¹⁹



Створіть ментальну мапу інформації, що її подано у статті. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.



Завдання 15. Уважно прочитайте розділ “Комп'ютерний фонд української мови в Інституті мовознавства ім. О.О. Потебні НАН України” у підручнику (Карпіловська, 2006, с. 96-106).²⁰



Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.

¹⁹ Завдання базового рівня складності.

²⁰ Завдання базового рівня складності.

Проектні роботи до Розділу 1



Проектна робота 1. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] У “Короткому англо-українському словнику з комп’ютерної лінгвістики” (Коломієць, 2022), знайдіть приклади, які ілюструють основні положення статті. Підготуйте і проведіть мультимедійну презентацію²¹ на основі статті і знайдених Вами прикладів.²²

Завадська, В. (2013). Коли “вікно” не є вікном, або ще раз про сучасну українську ІТ-термінологію. *Українське мовознавство*, 43, 20-27.



***Анотація.** Розглядаються особливості перекладу термінів українською мовою, зокрема в ІТ-сфері, в галузі комп’ютерної термінології. Окреслено проблеми, пов’язані з мовою-посередницею та явищем калькування термінів, вказано причини появи термінів, що є нехарактерними для української літературної мови. Подано деякі рекомендації практичного характеру.*

²¹ Студентам пропонується ознайомитися із принципами і підходами до створення презентацій, а також із різними типами презентацій (мультимедійна, постер, “презентація у ліфті” (англ. *an elevator pitch*) тощо) і відповідними (онлайн-)інструментами. Студенти заохочуються до набуття досвіду й навичок публічних виступів.

Ми вдячні професору Каю-Уве Кюнбергеру за натхнення залучити практику підготовки і проведення різних типів презентацій у викладання. Переконані, що презентації, окрім того, що володіють надзвичайно потужним евристичним потенціалом, тренують водночас цілий набір м’яких навичок.

У нашому курсі, ця практика розтлумачується студентам як завдання підготувати і провести презентацію на тему, що її представлено у відповідній науковій статті: студентам слід виокремити у статті саме той матеріал, який є для них важливим і цікавим; студенти можуть додавати матеріал до презентації, якщо, на їхню думку, цей матеріал сприятиме кращому розумінню й сприйняттю презентації аудиторією; студентам під час презентації слід критично обговорити окремі положення і приклади із статті, додавши своє бачення проблеми та/або власні приклади. Важливо: по завершенню презентації, у студентів є можливість задати питання аудиторії - цей етап покликаний залучити у дискусію усіх учасників семінару (при цьому, не очікується, що як самі доповідачі, так і їхня аудиторія матимуть чіткі відповіді на поставлені питання, адже мета цього етапу - генерувати цивілізовану дискусію як обмін думок). Тривалість презентації - орієнтовно 45 хвилин, плюс час на дискусію.

²² Уміщені у навчально-методичний посібник проектної роботи намічені як такі, що мають високий рівень складності.

Ключові слова: переклад, термін, ІТ-термінологія, ІТ-переклад, стандартизація термінів.

Коломієць, В.О. (2022). *Короткий англо-український словник з комп'ютерної лінгвістики*. Київ: Видавничий центр КНЛУ.

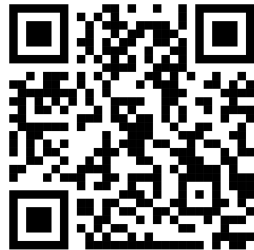


Анотація. Словник містить переклад близько 1000 англomовних термінів із комп'ютерної лінгвістики, а також слів і словосполучень, які не є термінами, але вживаються в науковій літературі. Укладений на основі корпусу анотацій наукових статей, опублікованих з 2000 по 2017 рік у журналі “*Computational linguistics*”, який є офіційним виданням Міжнародної асоціації комп'ютерної лінгвістики. Для студентів, аспірантів, науковців і спеціалістів з автоматичного опрацювання інформації природною мовою.



Проектна робота 2. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Дослідіть сучасний - на момент вивчення Вами курсу - стан української комп'ютерної лінгвістики. Проаналізуйте еволюцію комп'ютерної лінгвістики в Україні. Підготуйте і проведіть мультимедійну презентацію на основі статті і Вашого аналізу. Висловіть і обґрунтуйте власну думку про стан, потреби і перспективи розвитку комп'ютерної лінгвістики в Україні.

Балог, В. (2005). Сучасний стан української комп'ютерної лінгвістики. *Лексикографічний бюлетень*, 11, 28-35.



Анотація. У статті розглянуто стан української комп'ютерної лінгвістики на сучасному етапі розвитку мовознавчої науки. Оглядом подано інформацію про наукові установи, де працюють у галузі комп'ютерної лінгвістики, й основні здобутки науковців у цій галузі, а також коротко окреслено перспективи її розвитку.

Ключові слова: мовознавство, минуле і сучасність, перспективи, українська комп'ютерна лінгвістика.



Проектна робота 3. Шановні студенти, ця проектна робота відкрита для Ваших ідей відповідно до тематики Розділу 1.

Будьте винахідливі й розвивайте власне почуття напрямку.

РОЗДІЛ 2

Обробка природної мови: синтаксис і парсинг

- ▷Лінгвістичне моделювання синтаксису
 - Синтаксична неоднозначність
- ▷Автоматичний синтаксичний аналіз, або парсинг
 - Сутність, особливості і основні сфери застосування
 - Автоматичний синтаксичний аналіз за безпосередніми складниками
 - Автоматичний синтаксичний аналіз за зв'язком залежностей
 - Автоматичні синтаксичні аналізатори, або парсери
- ▷Комп'ютерний синтаксис української мови
 - Парсинг речень української, англійської та німецької мов
- ▷Корпуси текстів
 - Корпус української мови
 - Генеральний регіонально анотований корпус української мови
 - Інші корпуси української мови та слов'янських мов
 - Провідні корпуси англійської мови
 - Провідні корпуси німецької мови

Мова структурується на різних своїх рівнях, починаючи, у випадку із усним мовленням, із певних закономірностей, наявних в акустичному сигналі, що можуть бути визначені як *фони* (відмінні один від одного і, як такі, впізнавані звуки, які змінюють один одного й із яких вибудовується мова). Групи фонів, які є специфічними для конкретної мови (і які, якщо й дещо змінюються, не перешкоджають впізнаванню слів слухачем), визначаються як *фонем* цієї мови. Фонем, у свою чергу, є складовими *морфем* (мінімальних значущих частин слів), які є вже складовими слів. (У випадку письмового мовлення, на відміну від усного, говорять про символи, графем, склади і слова.) Слова об'єднуються у *словосполучення*, як-от іменникові, дієслівні, прикметникові і прийменникові, які виступають структурними компонентами *речень*, здатних виражати завершені думки. На ще вищому рівні, виокремлюють різні типи структури *дискурсу*, яка, втім, є дещо вільнішою, порівняно із структурами нижчих рівнів.

Були створені відповідні техніки автоматичного аналізу мови на всіх її структурних рівнях <...>. [Т]ехніки, створені у 1980-ті та 1990-ті роки з метою розпізнавання мовлення мали великий вплив на переорієнтацію досліджень із обробки природної мови на нові - корпусні і статистичні - підходи (Schubert, 2020).



Завдання 16. Уважно прочитайте розділ “Автоматичний синтаксичний аналіз тексту” у підручнику (Карпіловська, 2006, с. 139-144).²³



Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.



Завдання 17. Ознайомтеся із сервісом автоматичного синтаксичного аналізу “Parse a sentence.” Проаналізуйте кілька речень англійською мовою. Визначте основні, на Ваш погляд, переваги й недоліки цього сервісу.



Знайдіть інші подібні сервіси і порівняйте із “Parse a sentence.” Не обмежуйте свій пошук англійською мовою, а й поекспериментуйте із сервісами для української, німецької та інших мов. Узагальніть особливості їхнього функціоналу.²⁴

²³ Завдання базового рівня складності.

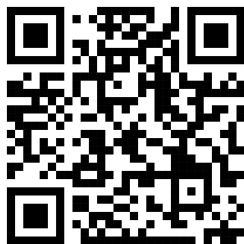
²⁴ Завдання середнього рівня складності.



Завдання 18. Наведіть приклади 5 українських та англійських або німецьких речень, де наявна синтаксична неоднозначність. Чи може проблема синтаксичної неоднозначності цих речень бути вирішена сучасними сервісами автоматичного синтаксичного аналізу? Якщо так, то яким чином? Якщо ні, то чому?²⁵



Завдання 19. Уважно прочитайте розділ “Корпусна лінгвістика: предмет дослідження і завдання” у підручнику (Карпіловська, 2006, с. 74-96).²⁶



Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.



Завдання 20. Дослідіть призначення і функціонал Корпусу української мови.²⁷



²⁵ Завдання середнього рівня складності.

²⁶ Завдання базового рівня складності.

²⁷ Завдання високого рівня складності.

Дослідіть призначення і функціонал Генерального регіонально анотованого корпусу української мови.²⁸



Знайдіть і дослідіть інші корпуси української мови та слов'янських мов, а також провідні корпуси англійської та німецької мов. Поекспериментуйте із цими корпусами. Як Ви можете використати їх у навчанні, наукових дослідженнях й (майбутній) професійній діяльності?²⁹

Структуруйте результати своєї розвідки і представте їх на семінарі.

²⁸ Див. також (Shvedova, 2020).

Закликаємо студентів відстежувати Генеральний регіональний анотований корпус української мови (ГРАК) у соціальних мережах, зокрема у Фейсбучі, де ГРАК є доволі активним і часто запрошує українських студентів долучитися до своїх проєктів.

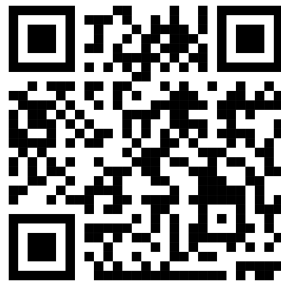
²⁹ Див. також (Шведова, 2022).

Проектні роботи до Розділу 2



Проектна робота 4. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Гирин, О.В. (2020). Основні проблеми систем обробки природної мови. *Філологічні студії. Збірник наукових праць*, 14, 41-45.



Анотація. Стаття присвячена обробці природної мови, а саме автоматичній синтаксичній обробці англійських речень. Висвітлено проблеми, спричинені цим процесом, що пов'язані з графічною, семантичною та синтаксичною неоднозначністю. Визначено шляхи вирішення цих проблем, зумовлених застосуванням автоматичного синтаксичного аналізу, та яким чином такі методи аналізу можуть бути корисні для розробки його нових алгоритмів. Дослідження зосереджене на питаннях, які унеможливають основу обробки природної мови - парсинг. Це процес аналізу речень за їх структурою, змістом і значенням, метою якого є визначення граматичної структури речення, розподіл речень на складові компоненти і визначення зв'язків між ними.

Ключові слова: синтаксичний аналіз, обробка природної мови, статистичне машинне навчання, неоднозначність.



Проектна робота 5. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Шипнівська, О.О. (2013). Визначення синтаксично неоднозначних конструкцій в україномовному тексті. *Мовознавство*, 4, 67-76.



Анотація. Стаття присвячена розглядові теоретичних та практичних засад вивчення в українській мові синтаксичної омонімії як одного з типів мовленнєвої неоднозначності. На підставі аналізу найвідоміших наукових розвідок з традиційної та комп'ютерної лінгвістики виділено типи синтаксично неоднозначних конструкцій і відзначено закономірність їх функціонування в українському тексті.

Ключові слова: синтаксична омонімія, синтаксична конструкція, автоматичний синтаксичний аналіз, синтаксичний зв'язок, словосполучення, синтагма.



Проектна робота 6. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Шипнівська, О. (2013). Визначення типів синтаксичної неоднозначності у знаннеорієнтованій системі машинного перекладу. *Українське мовознавство*, 43, 104-113.



Анотація. Розглянуто теоретико-практичні засади визначення синтаксично неоднозначних конструкцій в автоматичному синтаксичному аналізі. На матеріалі текстів військової тематики проаналізовано основні типи синтаксичної омонімії.

Ключові слова: автоматичний синтаксичний аналіз, знаннеорієнтована система машинного перекладу, синтаксична неоднозначність, типи синтаксичних омонімів, синтаксична структура.



Проектна робота 7. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Лангенбах, М. (2015). Автоматичний синтаксичний аналіз речення за принципами граматики залежностей. *Науковий вісник Східноєвропейського національного університету імені Лесі Українки*, 3, 249-254.



Анотація. Стаття висвітлює питання організації автоматичного синтаксичного аналізу речень. Актуальність обраної теми визначається її належністю до класу завдань комп'ютерного моделювання мови - однієї з ключових проблем у сфері інформаційних технологій. Уперше наведено алгоритм автоматичного моделювання структури речення в термінах граматики залежностей, побудований на основі інформації про його безпосередні складники. Описано хід виконання основних завдань: формалізацію правил установлення зв'язків у реченні та їх автоматизацію; з'ясування основних переваг та недоліків обраної формальної моделі. Схарактеризовано структуру й основні етапи роботи аналізатора. Теоретичне значення дослідження полягає в накопиченні бази синтаксичних моделей речень. Практичне застосування отриманих результатів передбачається в системах комп'ютерного опрацювання мови, у навчально-методичних розробках тощо.

Ключові слова: граMATика залежностей, автоматичний синтаксичний аналіз, автоматичне опрацювання тексту, лінгвістичний алгоритм, дерево залежностей, універсальна граMATика.



Проектна робота 8. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Лозинська, О.В., Давидов, М.В., & Пасічник, В.В. (2016). Трансформація дерев граматики складових у дерева граматики залежностей для граматичного розбору українських речень. *Штучний інтелект*, 3, 22-31.



Анотація. У статті розглянуто метод трансформації дерев граматики складових у дерева граматики залежностей, який використовується для перекладу речень української словесної мови у речення анотованої української жестової мови. Зроблено граматичний розбір корпусу речень «Українська словесна мова» та побудовано дерева граматичного розбору цих речень (дерева граматики складових). Описано кроки алгоритму трансформації дерев граматики складових у дерева граматики залежностей для речень української словесної мови.

Ключові слова: граматичний розбір, дерево синтаксичного розбору, трансформація дерев, граматика складових, граматика залежностей, машинний переклад.



Проектна робота 9. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Дарчук, Н.П. (2013). Автоматичний синтаксичний аналіз текстів корпусу української мови. *Українське мовознавство*, 43, 11-20.



Анотація. Розглядається автоматичне представлення синтаксичної структури речення на рівні словосполучення: автоматичне виокремлення словосполучення, приписування йому типу синтаксичного зв'язку (підрядного, сурядного, предикативного).

Ключові слова: автоматичний синтаксичний аналіз, словосполучення, синтаксичний зв'язок, підрядний зв'язок, сурядний зв'язок, предикативний зв'язок, ядровий підрядний зв'язок, ад'юнктивний підрядний зв'язок.



Проектна робота 10. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Дарчук, Н.П. (2020). АГАТ-синтаксис як компонент комп'ютерної граматики української мови. *Українська мова. Науково-теоретичний журнал*, 4(76), 89-103. <https://doi.org/10.15407/ukrmova2020.04.089>



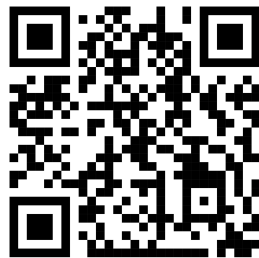
Анотація. Статтю присвячено описові синтаксичного модуля комп'ютерної граматики української мови АГАТ. Простежено, як унаслідок роботи правил АГАТ-граматики лінійна морфологічна структура речення перетворюється на двомірну деревоподібну синтаксичну структуру із приписуванням типу синтаксичного зв'язку і типу синтаксичного відношення. Це дало змогу автоматично побудувати алфавітно-частотні словники сполучуваності всіх частин мови.

Ключові слова: автоматичний синтаксичний аналіз, дерево залежностей, синтаксичні зв'язки, синтаксичні відношення.



Проектна робота 11. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Гирин, О.В. (2017). Автоматичний синтаксичний аналіз англійського речення: застосування та перспективи. *Вісник Житомирського державного університету імені Івана Франка*, 1(85), 26-30.



Анотація. Стаття присвячена автоматичному синтаксичному аналізу англійської мови; проведено огляд основних сфер використання автоматичного синтаксичного аналізу та визначено завдання, які достатньою мірою не вирішені. У статті обґрунтовано, як автоматичний синтаксичний аналіз може покращити вирішення того чи іншого завдання в межах проаналізованих сфер використання. У результаті проведеного аналізу було визначено коло лінгвістичних питань, що сприятиме розробці більш досконалої моделі автоматичного синтаксичного аналізу.

Ключові слова: парсинг, обробка природних мов, стилімерія, формальна логіка.



Проектна робота 12. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Марчук, К.І. (2017). Автоматичний синтаксичний аналіз англійського тексту: застосування та перспективи. *Сучасні філологічні дослідження та навчання іноземної мови в контексті міжкультурної комунікації*, 10, 328-333.



Анотація. Сучасний світ лінгвістики потребує постійного розвитку, тому опрацювання текстової інформації незмінно залишається провідним завданням лінгвістики. Наші знання про дійсність втілюються у певній вербалізованій формі. Дати можливість комп'ютеру "розуміти" текст означає, що комп'ютер матиме властивість виокремлювати з нього необхідну інформацію. Таке "розуміння" тексту полягає у вмінні аналізувати його на різних рівнях. Метою статті є пояснення проблематики обробки природної мови за допомогою цифрових та комп'ютерних ресурсів. Автоматичний аналіз тексту є перспективним напрямом, адже він є основою для розуміння будь-якої мови. Синтаксичний аналіз має на меті виявлення зв'язків між членами речення, встановлення смислового значення, членування речення. Автоматичний синтаксичний аналіз тексту передбачає використання комп'ютерного синтаксису. Він спрямований на виявлення в тексті синтаксичних структур та їхнє формалізоване представлення.

Ключові слова: автоматичний аналіз тексту, автоматичний синтаксичний аналіз, комп'ютерний синтаксис, парсинг.



Проектна робота 13. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Парнус, К.І., Свісюк, О.В. (2022). Алгоритм аналізу англomовного тексту за допомогою парсерів. *SWorldJournal*, 14(01), 108-113. <https://doi.org/10.30888/2663-5712.2022-14-01-020>



***Анотація.** Потреба в науковому дослідженні алгоритму обробки англomовного тексту за допомогою парсерів актуальна для вивчення граматичних структур студентами в англomовних текстах сучасного соціуму. Метою дослідження є обґрунтування методології концептуального підходу до парсерування англomовних текстів. Необхідність вивчення сполучуваності лексичних одиниць зумовлена нерозробленістю широкого кола як теоретичних, так і прикладних проблем. Теоретичні аспекти, які потребують вивчення, - це, зокрема, граматична і лексична валентність слів, типова сполучуваність, синонімія словосполучень різних структурних типів, лексична і граматична валентність як критерій синонімічності, закони комбінаторики словосполучень різних типів і розрядів, лексична валентність як критерій розмежування вільних і фразеологічних словосполучень, взаємодія стійкості й ідіоматичності тощо. До прикладних проблем можна віднести автоматизацію лінгвістичних досліджень, автоматичне визначення меж словосполучень, установлення критеріїв членування фрази на синтагми, автоматичний синтаксичний аналіз речення, автоматичне реферування й анотування тексту на основі*

сполучувальнісних критеріїв тощо. Об'єктами досліджень є: власне природні мови; використання комп'ютерних програм при аналізі англомовних текстових матеріалів. Предметом цього дослідження є сучасні англомовні тексти. Об'єктом дослідження є граматичні структури сучасних англомовних текстів. Мета дослідження - обґрунтування методології концептуального підходу до парсерування текстів. Методи дослідження. Пропоноване дослідження проведене з використанням загальнонаукових методів аналізу, синтезу і узагальнення, а також бібліографічно-описового методу з метою залучення та систематизації англомовних текстових матеріалів.

Ключові слова: *парсинг, парсер, граббер, синтаксичний аналізатор.*



Проектна робота 14. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Толубко, В.В., Шипнівська, О.О., & Ляшенко, А.В. (2010). Задачі автоматичної обробки синтаксичної структури в знання-орієнтованій системі машинного перекладу. *Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка*, 27, 136-141.



Анотація. У статті розглядаються задачі розробки процедур автоматичного синтаксичного аналізу й синтезу природно-мовних речень відповідно до положень знання-орієнтованої системи машинного перекладу. Розглядаються існуючі формалізовані моделі синтаксису. Пропонується фреймова структура синтаксису як така, що забезпечує адекватний синтаксичний аналіз і синтез.

Ключові слова: автоматичний синтаксичний аналіз, формальна граматики, фрейм, словник синтаксичних правил, словник інтерпретацій, просте речення, складне речення.



Проектна робота 15. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Міщенко, А. (2013). Створення паралельного банку дерев для німецької та української мов. *Українське мовознавство*, 43, 51-63.



Анотація. *Описано створення паралельного банку дерев для української та німецької мов. Було здійснено ручне тегування й лематизацію токенів для української мови, створено набір тегів для анотування українських речень на рівні синтаксичної структури; проведено вирівнювання й визначення повних або часткових відповідників на рівні як термінальних, так і нетермінальних символів. Для анотування банку дерев українською мовою було застосовано формат TIGERXML, адаптований до потреб формального опису граматичної структури української мови.*

Ключові слова: *банки дерев, анотація, тегування частин мови, вирівнювання, переклад.*



Проектна робота 16. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті. Обміркуйте можливості застосування корпусів у викладанні англійської і німецької мов.

Шведова М.О. (2022). Застосування корпусу у викладанні української мови як іноземної. *Слов'янські мови, культури, переклад і викладання: сучасний стан і перспективи*, 227-232.



Abstract. *The article describes the applications of General Regionally Annotated Corpus of Ukrainian (GRAC: uacorporus.org) in teaching Ukrainian as a foreign language. The importance of using authentic language material in teaching Ukrainian as a foreign language is emphasized. The following issues are discussed: frequency criteria in building the lexical minimum lists, searching for collocations, idiomatic variants and words unattested in dictionaries, using the corpus for specifying grammatical features, analyzing frequencies of synonyms and their historical change.*

Keywords: *Ukrainian as a foreign language, corpus, vocabulary, collocations, frequency.*³⁰

³⁰ Стаття українською мовою; анотацію і ключові слова англійською мовою подано авторкою статті.



Проектна робота 17. Шановні студенти, ця проектна робота відкрита для Ваших ідей відповідно до тематики Розділу 2.

Будьте винахідливі й розвивайте власне почуття напрямку.

РОЗДІЛ 3

Обробка природної мови: семантика і підходи до її формалізації

- ▷Лінгвістичне моделювання семантики
 - Семантична неоднозначність
 - Формалізація семантики із введенням граматики
 - Відмінкова граматика Ч. Філлмора
 - Породжувальна граматика Н. Чомскі
- ▷Бази даних і бази знань
 - Внутрішня і зовнішня формалізація даних
 - Декларативні і процедурні знання
- ▷Представлення знань
 - Логічні моделі представлення знань
 - Статичні моделі представлення знань
 - Семантичні мережі
 - Онтології і мова OWL³¹
 - Динамічні моделі представлення знань
 - Організація знань в експертних системах
- ▷Словники: традиційні, комп'ютерні, ідеографічні (тезауруси)
- ▷Автоматичний семантичний аналіз тексту
 - Автоматичне індексування тексту і інформаційний пошук
 - Сентимент-аналіз тексту
- ▷Автоматичний семантичний аналіз і моделювання людського інтелекту

Мова слугує передачі значення. Отже, аналіз синтаксичної структури є лише половиною шляху до механізації цієї головної функції мови, й усі можливі переваги і досягнення окремих підходів до синтаксису залежать перш за все від їхньої можливості підтримати семантичний аналіз і генерувати мову із тих значень, що їх потрібно передати.

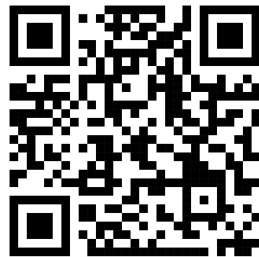
Це не означає, що синтаксичний аналіз у собі не має жодної цінності - він може бути корисним у таких прикладних аспектах, як перевірка граматичної правильності або статистичний машинний переклад. Але задля амбіційнішої мети виведення і вираження значення мови необхідно є насамперед теорія *семантичної репрезентації*, рівно як і її зв'язок із поверхневою структурою, і її взаємодія із репрезентацією фонових знань і їхнім використанням. <...> [У цьому розділі] ми розумітимемо семантичну інтерпретацію як процес виведення семантичних репрезентацій із потоку слів, і вважатимемо, що цьому

³¹ англ. *Web Ontology Language*.

процесу передує, або здійснюється із ним одночасно, процес парсингу (Schubert, 2020).



Завдання 21. Уважно прочитайте розділ “Бази даних і бази знань (=інтелектуальні бази даних)” у підручнику (Карпіловська, 2006, с. 34-50).³²



Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.



Завдання 22. Уважно прочитайте розділ “Лінгвістичні проблеми створення баз знань” у підручнику (Карпіловська, 2006, с. 114-126).³³



Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3

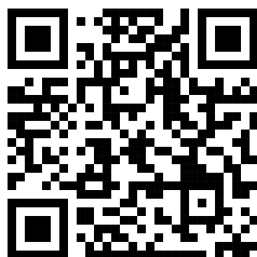
³² Завдання базового рівня складності.

³³ Завдання базового рівня складності.

запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.



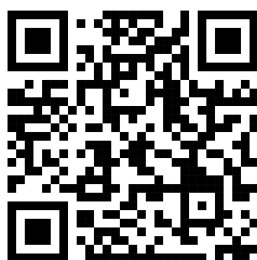
Завдання 23. Уважно прочитайте розділ “Комп’ютерна лексикографія: її предмет та завдання” у підручнику (Карпіловська, 2006, с. 54-74).³⁴



Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.



Завдання 24. Уважно прочитайте розділ “Автоматичний логіко-семантичний аналіз тексту” у підручнику (Карпіловська, 2006, с. 144-153).³⁵



Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.

³⁴ Завдання базового рівня складності.

³⁵ Завдання базового рівня складності.

Проектні роботи до Розділу 3



Проектна робота 18. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Путіліна, О. (2012). Універсальна та відмінкова граматики: корелятивно-ієрархічні відношення. *Лінгвістичні студії. Збірник наукових праць*, 24, 15-20.



Анотація. *Стаття продовжує цикл публікацій, присвячених проблемам відмінкової граматики в загальнолінгвістичній парадигмі, встановленню витоків відмінкової граматики, а також простеженню причин, передумов і специфіки її розвитку і формування як самостійного лінгвістичного напрямку та корелятивності / ієрархічності з універсальною граматиною.*

Ключові слова: *відмінкова граMATика, універсалізм, структуралізм, універсалія, універсальна граMATика.*



Проектна робота 19. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Ситар, Г. (2015). Конструкційна граматики як теоретичне підґрунтя дослідження фразеологізованих речень. *Типологія та функції мовних одиниць. Науковий журнал на пошану член-кореспондента НАН України Івана Романовича Вихованця*, 2(4), 192-205.



Анотація. Статтю присвячено аналізу основних положень конструкційної граматики як напряму теоретичної лінгвістики, у межах якого основною мовною одиницею є конструкція як некомпозиційний мовний знак. Центральним положенням конструкційної граматики є заперечення потреби протиставлення основних і периферійних граматичних одиниць. Визначено ознаки конструкцій, розмежовано конструкції та конструкти, виділено синтаксичні, семантичні, прагматичні та просодичні аспекти аналізу конструкцій, наголошено на розмежуванні внутрішніх і зовнішніх ознак конструкцій, що стане надійним теоретичним підґрунтям для аналізу фразеологізованих речень української мови. Розглянуто принципи побудови бокс-діаграми як основного формального методу конструкційної граматики, який покликаний втілити не тільки форму конструкції, але й відобразити особливості семантики, прагматики та просодики, що загалом відбиває теоретичне положення конструкційної граматики про взаємовплив мовних рівнів. Позичено колексемний аналіз, розмежувальний колексемний аналіз і коваріацію лексем як корпусно зорієнтовані статистичні методи дослідження конструкцій.

Ключові слова: бокс-діаграма, конструкція, конструкційна граматика, синтаксичний фразеологізм, фразеологізоване речення.



Проектна робота 20. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Шомко, А.В. (2017). Генеративна граматики як галузь лінгвістичної науки. Проблеми та перспективи її розвитку. *Сучасні філологічні дослідження та навчання іноземної мови в контексті міжкультурної комунікації*, 10, 355-359.



Анотація. У статті проаналізовано етапи становлення та розвитку генеративної граматики Н. Хомські, окреслено її основні проблеми, завдання та здобутки з можливістю практичного застосування у сучасному світі.

Ключові слова: генеративізм, синтаксична структура, генеративна граматики, породжувальний механізм, стандартна теорія, мінімалістська програма.



Проектна робота 21. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Калько, Р.М. (2017). Універсалізм Н. Хомського у межах когнітивної парадигми світової лінгвістики. *Наукові записки Національного університету «Острозька академія»: Серія «Філологія», 1(64), 143-146.*



Анотація. У статті висвітлено проблему універсалізму Н. Хомського у межах когнітивної парадигми світової лінгвістики; узагальнено різні наукові підходи до вивчення проблеми; окреслено основні риси універсальної теорії Н. Хомського про вроджені здібності до мови, що має за підґрунтя проблему співвідношення мови й мислення; обґрунтовано, що ідеї трансформаційно-генеративної граматики Н. Хомського про мову як породжувальний механізм, про ментальні репрезентації та їхнє моделювання знайшли своє відображення в когнітивній лінгвістиці.

Ключові слова: універсалізм, когнітивна лінгвістика, вроджені здібності, генеративна граматика, породжувальний механізм, ментальні репрезентації.



Проектна робота 22. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Комарницька, О. (2018). Методи автоматизованого семантичного аналізу природномовної інформації. *Філологічний дискурс*, 7. 92-100.



***Анотація.** У статті розглянуто особливості автоматизованого семантичного аналізу тексту, досліджено проблеми створення автоматизованих лінгвістично-програмних засобів, придатних для застосування в системах екстракції семантики з тексту. Автором систематизовано та здійснено порівняльний аналіз результатів наукових досліджень у галузі розробки моделей і методів семантичного аналізу природномовної інформації. Визначено два основні підходи в напрямку комп'ютерної обробки природномовних текстів: лінгвоаналітичний і статистичний. Автором аргументовано, що найбільш перспективними та ефективними з них є, відповідно, експліцитні методи семантичного аналізу текстової інформації (алгоритми онтологічного семантичного аналізу) та методи латентно-семантичного аналізу. Окреслено можливі шляхи удосконалення існуючих комп'ютерних засобів діагностування релевантності природномовної інформації; обґрунтовано, що найпопулярнішими методами обробки природномовної інформації з метою екстракції та репрезентації семантики мають бути системи, що ґрунтуються на ефективному поєднанні лінгвістичних технологій аналізу (графематичного, морфологічного, синтаксичного, семантичного), зокрема із застосуванням онтологій, та методу латентно-семантичного аналізу. Доведено, що інтеграція технологій експліцитного семантичного*

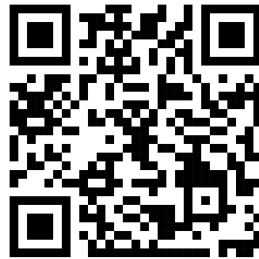
аналізу, латентно-семантичного аналізу, методів теорії нечіткої логіки, штучного інтелекту та ін. є перспективним шляхом розв'язання проблеми автоматизованого семантичного аналізу природномовної інформації.

Ключові слова: *автоматизований семантичний аналіз, природномовна інформація, лінгвістичний аналіз, латентно-семантичний аналіз, метод, онтологія.*



Проектна робота 23. [Усі студенти групи] Уважно прочитайте запропонований розділ монографії. Створіть ментальну мапу розділу. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі розділу. Висловіть і обґрунтуйте власну думку щодо основних положень розділу.

Вавіленкова, А.І. (2016). Концепція автоматизованого лінгвістичного аналізу тексту. *Теоретичні основи аналізу електронних текстів: монографія*. К.: ТОВ «СІК ГРУП УКРАЇНА.» С. 175-181.



Анотація. У дослідженні запропоновано підхід до вирішення основної проблеми автоматичної обробки текстової інформації: відсутності формального апарату, який дозволив би аналізувати текстову інформацію за єдиним принципом, систематизувати процес змістовного пошуку та уникнути неоднозначності при аналізі текстових документів. Зокрема, підхід заснований на формуванні змістовних моделей текстових документів на базі використання математичного апарату логіки предикатів.

Ключові слова: автоматична обробка текстової інформації, автоматичний семантичний аналіз, змістовна модель тексту, логіко-лінгвістична модель речення.



Проектна робота 24. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Яровий, А., Кудрявцев, Д., & Крилик, Л. (2020). Удосконалення методу семантичного аналізу тексту. *Інтернет-Освіта-Наука-2020: XII Міжнародна науково-практична конференція*. Вінниця: Вінницький національний технічний університет. С. 2-4.



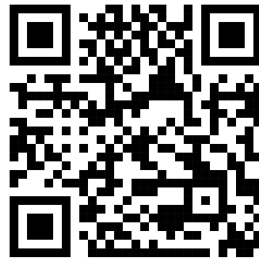
Анотація. Під час проведеного дослідження було розглянуто сучасні методи аналізу текстової інформації. Визначено фактори, що негативно впливають на семантичний аналіз тексту. Запропоновано способи покращення семантичного аналізу тексту. Реалізовано семантичний аналіз тексту на прикладі чат-бота із застосуванням запропонованих рішень та порівняно результат їх роботи.

Ключові слова: автоматичний аналіз тексту, автоматичний семантичний аналіз, інтелектуальний аналіз даних, чат-бот.



Проектна робота 25. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Дарчук, Н.П. (2018). До питання про створення автоматичного словника словосполучень української мови. *Лінгвістичні студії*, 36, 148-158. <https://doi.org/10.31558/1815-3070.2018.36.23>



Анотація. У статті розглянуто засади створення автоматичного словника словосполучень української мови, створюваного на кафедрі української мови та в лабораторії комп'ютерної лінгвістики Інституту філології Київського національного університету імені Тараса Шевченка. Подано типологію синтаксичних зв'язків та семантичних відношень, на яких ґрунтовано створювану систему автоматичного виокремлення словосполучень, принципи семантичного розмічення словника. Описано експеримент з вилучення з корпусу публіцистичних текстів словосполучень з орудним відмінком, прокоментовано його результати.

Ключові слова: автоматичний семантичний аналіз, автоматичний синтаксичний аналіз, автоматичний словник словосполучень, комп'ютерна граматики, орудний відмінок, семантичне відношення, синтаксичний зв'язок.



Проектна робота 26. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Романюк, А., Романишин, М. (2013). Тональний словник української мови на основі сентимент-анотованого корпусу. *Українське мовознавство*, 43, 63-75.



Анотація. Докладно розглянуто всі етапи створення сентимент-анотованого корпусу українськомовних відгуків і тонального словника на його основі.

Ключові слова: емоційно-смісловий аналіз, тональний словник, сентимент-анотований корпус, засоби для анотування текстів.



Проектна робота 27. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Дарчук, Н., Зубань, О., Лангенбах, М., & Ходаківська, Я. (2016). АГАТ-семантика: семантичне розмічування корпусу української мови. *Українське мовознавство*, 1(46), 92-102.



Анотація. У статті розглянуто лінгвістичні засади семантичного розмічування Корпусу української мови як четвертого етапу представлення інформації про одиниці Корпусу. В основу розмічування покладено таксономічну класифікацію Національного корпусу російської мови³⁶, але доповнену та видозмінену. Створено програмне забезпечення для роботи в он-лайн режимі. Матеріалом слугував частотний словник публіцистичного стилю обсягом 40 тис. лексем, укладений на вибірці 16 млн. слів українськомовного тексту.

Ключові слова: корпус текстів, семантичне розмічування, таксономічна класифікація, таксон.

³⁶ Див. також (Приймак, 2021).



Проектна робота 28. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Зубань, О. (2018). Корпус української мови - комп'ютерна експертна система лінгвістичного аналізу українськомовного тексту. *ТЕКА Komisji Polsko-Ukraińskich Związków Kulturowych*, 6(13), 191-206. <https://doi.org/10.31743/teka.5723>



Анотація. У статті представлено структуру та засади автоматичного укладання експертної системи лінгвістичного аналізу «Корпус української мови». Методика формалізованого опису мовних одиниць тексту, запропонована у створенні Корпусу, забезпечує проведення автоматичного морфологічного, морфемного, синтаксичного, семантичного аналізів українськомовного тексту, а також автоматичне укладання різноманітних електронних частотних словників за текстовими вибірками.

Ключові слова: Корпус української мови, Електронний частотний словник, база даних, автоматичний лінгвістичний аналіз.



Проектна робота 29. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Вавіленкова, А. (2017). Структура системи порівняльного аналізу електронних текстових документів за змістом. *Вісник Національного університету «Львівська політехніка.» Комп'ютерні науки та інформаційні технології*, 864(1), 277-285.



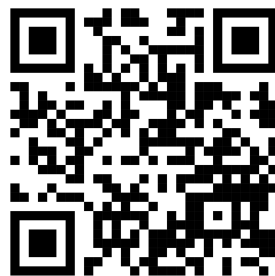
Анотація. Проаналізовано схему функціонування сервісів для визначення унікальності електронних текстових документів, розглянуто їхні основні характеристики під час перевірки на оригінальність наукової статті. Наведено структуру системи порівняльного аналізу електронних текстових документів за змістом, описано принцип функціонування кожної з її основних компонент.

Ключові слова: природна мова, порівняльний аналіз, пошук, логіко-лінгвістична модель, база знань.



Проектна робота 30. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Бараннік, В.В., Белікова, Т.В., Капко, М.О., & Гуржій, І.А. (2019). Комплексний метод автоматичного фоносемантичного аналізу текстової інформації на основі оцінки вагомих семантичних одиниць в умовах інформаційного протиборства. *Кибербезпека: освіта, наука, техніка*, 3(3), 53-62. <https://doi.org/10.28925/2663-4023.2019.3.5362>



Анотація. У статті запропоновано метод аналізу текстових повідомлень для виявлення їх семантичної спрямованості та сугестивного впливу на підсвідомість людини, в умовах інформаційного протиборства. Метод заснований на оцінці фонетичної складової слова, за допомогою фоносемантичного аналізу. А також структурної, ймовірнісної оцінки та оцінки згідно місця в реченні, за допомогою ймовірнісного семантичного аналізу. Описано метод автоматичного фоносемантичного аналізу, запропонована схема алгоритму та концепцію методу автоматичного семантичного аналізу на основі аналізу структури слова, з схемою алгоритмом. Запропонований комплексний метод фоносемантичного аналізу на основі оцінки вагомих семантичних одиниць, який вбачає об'єднання двох вище описаних методів та продемонстровано приклад оцінки випадкового, завідома семантично-негативного речення. Для демонстрації оцінок проведено порівняння результатів роботи запропонованого методу, з результатами роботи загальної фоносемантичної оцінки речення. На основі результатів сформовані висновки.

Ключові слова: *фоносемантичний аналіз, семантичний аналіз на основі структури слова, інформаційне протиборство, сугестивний негативний вплив.*



Проектна робота 31. [Усі студенти групи] Уважно прочитайте запропонований розділ магістерської роботи. Створіть ментальну мапу розділу. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі розділу. Висловіть і обґрунтуйте власну думку щодо основних положень розділу.

Ляхоцький, О.С. (2016). Огляд технологій Semantic Web. *Програмний модуль для побудови семантичних порталів: Магістерська робота: 8.05010302 - Інженерія програмного забезпечення*. Тернопіль: Тернопільський національний економічний університет. С. 10-33.



Анотація. Метою наукового дослідження є побудова веб-семантичного порталу для того, щоб зробити інформацію, передану в Web, більш формалізованою й зручною для машинного сприйняття, для її ідентифікації та класифікації. Для досягнення мети було поставлено такі задачі: здійснено аналіз існуючих підходів до розробки порталів знань на основі технологій Semantic Web; спроектовано оптимальну модель структури зберігання й доступу до інформаційних ресурсів; здійснено дослідження порталів знань і семантичних порталів; розроблено тематичний рубрикатор для СКВІР; здійснено програмну реалізацію автоматичного рубрикування на основі семантичного аналізу вмісту інформаційних ресурсів; здійснено тестування розробленого програмного продукту. Об'єктом дослідження є семантичні портали. Предметом дослідження є web-орієнтований програмний комплекс для побудови семантичних порталів. Методи досліджень базуються на використанні технологій Ruby on Rails, Microsoft SharePoint, бібліотеки та методів інтелектуального аналізу даних Data Mining. Наукова новизна одержаних результатів: Дістав подальший розвиток семантичний аналіз вмісту інформаційних ресурсів для реалізації автоматичного

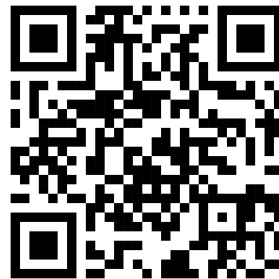
рубрикування. Вперше для структуризації та класифікації інформаційних ресурсів використовується підхід Topic Maps.

Ключові слова: програмний засіб, семантичний портал, рубрикатор, метадані, веб-ресурс.



Проектна робота 32. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.³⁷

Ваховська, О.В. (2021). Метафори емоцій: на шляху до формалізації, що базується на мовних даних. [Metaphors of emotions: towards a data-driven formalization.] *Вісник Київського національного лінгвістичного університету. Серія Філологія*, 24(1), 33-48. doi: 10.32589/2311-0821.24%20(1).2021.236030



***Анотація.** У статті представлено результати практичної розвідки, виконаної на перетині теорії штучного інтелекту та когнітивної лінгвістики - дисциплін у складі когнітивної науки, чия методологічна комплементарність, хоча й проголошується, досі має недостатнє унаочнення в практичних дослідженнях. Ця робота присвячена питанню створення формальної моделі метафори. Обчислення метафор протягом тривалого часу залишається однією з невирішених проблем у царині штучного інтелекту й низки суміжних дисциплін. Перешкодою до подолання зазначеної проблеми є семантична невизначеність метафор, що її неможливо вирішити за допомогою формул та алгоритмів. Мета статті полягає в демонстрації того, як за використання методологічного інструментарію когнітивної лінгвістики, а саме доробку семантики лінгвальних мереж³⁸, метафоричне значення може бути репрезентоване в такий спосіб, який постане придатним для обчислення в системах штучного інтелекту. Матеріалом дослідження*

³⁷ Цій проектній роботі передуює лекція українською мовою на основі обговорюваної статті.

³⁸ Семантику лінгвальних мереж розроблено професором С.А. Жаботинською.

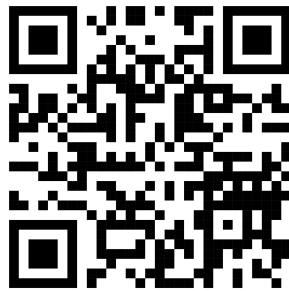
служували метафоричні вирази, які маніфестують у сучасному англомовному дискурсі концепти емоцій страху, смутку, щастя та безтурботності. Емоції як квалія є особистим досвідом, що його можливо словесно описати переважно за допомогою метафор. Це дає багатий матеріал та підкреслює практичну значущість дослідження, адже емоції, як і метафора, є складними для математичного моделювання сутностями. За точку відліку в статті взято алгебраїчну теорію метафори, керовану евристичним пошуком. Саме ця теорія спрямувала дослідження на послідовне розв'язання двох практичних завдань. По-перше, у статті у вигляді інтенціональних дефініцій визначено та впорядковано за ступенем проміантності парадигматичні властивості концептів FEAR, SADNESS, HAPPINESS та RELAXATION/SERENITY. По-друге, ці властивості для кожного концепту організовано у вигляді онтології - моделі репрезентації знань, що її можна обчислити математично та описати за допомогою спеціалізованих мов програмування. Онтології експлікують метафоризовані властивості концептів емоцій у їхньому взаємозв'язку з урахуванням ступеня проміантності.

Ключові слова: емоції, інтенціональне визначення, когнітивна лінгвістика, онтологія, схема, теорія штучного інтелекту, формалізація метафор.



Проектна робота 33. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Жаботинська, С.А. (2021). Типи онтологій для наративних концептів (на прикладі медійного наративного концепту COVID-19). *Виклики та парадокси соціальної взаємодії в постмодерному світі: лінгвістичні та психологічні аспекти: II Міжнародна науково-практична конференція*. Луцьк: Волинський національний університет імені Лесі Українки. С. 17-21.



Анотація. *Темою доповіді є способи структурування наративного концепту як інформації, маніфестованої тематично гомогенними наративними текстами. Алгоритм структурування уможливорює систематизацію такої інформації у вигляді концептуальних онтологій різних типів.*

Ключові слова: *наративний концепт, концептуальна онтологія, структурування, типи онтологій.*



Проектна робота 34. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Жаботинська, С.А. (2019). Семантика лінгвальних мереж у навчальному комбінаторному тезаурусі. *Studia Philologica. Філологічні студії*, 13, 17-27. doi: 10.28925/2311-2425.2019.13.3



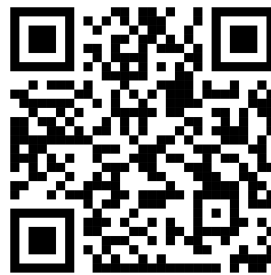
Анотація. У статті показано застосування положень лінгвокогнітивної концепції, названої “семантика лінгвальних мереж” (СЛМ), у розробленні методичної системи Лінгвакон, де у навчанні англійської мови вживаються концептуальні схеми і концептуальні онтології. У роботі надається короткий опис положень СЛМ у їхній проєкції на систему Лінгвакон і розглядається процедура укладання комбінаторного словника-тезаурусу як центрального елемента цієї системи. У якості ілюстрації пропонується комбінаторний словник-тезаурус за темою SCHOOL. Його вживання в навчальному процесі демонструється із залученням вправ, спрямованих на засвоєння граматики та розвиток мовленнєвих навичок за темою.

Ключові слова: прикладна когнітивна лінгвістика, семантика лінгвальних мереж, методика викладання англійської мови, система Лінгвакон, комбінаторний словник-тезаурус, концептуальна онтологія.



Проектна робота 35. [Усі студенти групи] Уважно прочитайте запропонований розділ магістерської роботи. Створіть ментальну мапу розділу. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі розділу. Висловіте і обґрунтуйте власну думку щодо основних положень розділу.

Войтенко, Я.С. (2020). Огляд методів семантичного аналізу текстів. Використання методів семантичного аналізу текстів в мережі Інтернет. *Дослідження методів семантичного аналізу текстів для інтелектуалізації Web-сайтів: Магістерська робота: 8.05010104 - Системи штучного інтелекту.* Харків: Харківський національний університет радіоелектроніки. С. 10-30.



Анотація. *Об'єкт дослідження - методи семантичного аналізу даних. Предмет дослідження - методи семантичного аналізу природномовних текстів. Мета роботи - дослідження методів семантичного аналізу текстів для інтелектуалізації web-сайтів. Згідно з метою роботи необхідно виконати такі завдання, як аналіз існуючих наукових джерел за темою дослідження, огляд існуючих систем семантичного аналізу, аналіз алгоритмів семантичного аналізу, дослідження проблем в сучасних методах інтелектуалізації web-сайтів, дослідження найперспективніших шляхів використання семантичного аналізу для інтелектуалізації сайтів, висновки та рекомендації з приводу опрацьованого матеріалу. Методи дослідження - методи семантичного аналізу природномовних текстів, латентно-семантичний аналіз, методи лексичного аналізу природномовних текстів, методи семантико-синтаксичного аналізу природномовних текстів.*

Ключові слова: семантичний метод, штучний інтелект, латентно-семантичний аналіз, контент-аналіз, нейронні мережі, інтернет, машинний переклад, web-сайт.



Проектна робота 36. [Усі студенти групи] Уважно прочитайте запропоновану статтю. Створіть ментальну мапу статті. Підготуйтеся до обговорення. [Учасники проектної роботи] Підготуйте і проведіть мультимедійну презентацію на основі статті. Висловіть і обґрунтуйте власну думку щодо основних положень статті.

Титенко, С.В. (2006). Проблема подання знань на основі природної мови у освітніх системах штучного інтелекту. *Лабораторія SET - Семантичні технології в електронному навчанні. Штучний інтелект в освіті.* Режим доступу: http://www.setlab.net/?view=Philosophy_Knowledge



Анотація. У статті розглядається проблема подання знань, зобов'язана своїм виникненням процесам, що відбуваються у розвитку комп'ютерної техніки і дослідженнях із штучного інтелекту, а саме появи і досить широкому поширенню систем, які називають системами, заснованими на знаннях. Це насамперед інтелектуальні інформаційно-пошукові і експертні системи. Досліджуються освітні процеси і застосування технологій штучного інтелекту для створення комп'ютерних систем навчання. Оглядаються підходи до моделювання процесів пов'язаних із навчанням, роботи вчителів і суб'єктів навчання, пояснюються вивчення і навчання як складові людської системи обробки інформації. Увага у статті зосереджується на поданні й передачі знань, об'єктивних засобами природної мови, як основних завдань освіти. Розглянуто сутність знань, мови і тексту з філософської і загальнонаукової точок зору із метою синтезувати технології штучного інтелекту у поданні знань і природну силу людської мови у завданні створення автоматизованих освітніх систем.

Ключові слова: знання, природна мова, освіта, штучний інтелект, інтелектуальні освітні системи, інформаційно-пошукові системи, експертні системи.



Проектна робота 37. Шановні студенти, ця проектна робота відкрита для Ваших ідей відповідно до тематики Розділу 3.

Будьте винахідливі й розвивайте власне почуття напрямку.

РОЗДІЛ 4

Комп'ютерна лінгвістика і штучний інтелект

- ▷ Інтелект людини як інструмент пізнання дійсності
 - Визначення інтелекту людини
 - Складники інтелектуальної діяльності людини
 - Типи інтелекту
- ▷ Штучний інтелект як модель інтелекту людини
 - Визначення штучного інтелекту
 - Машина та тест Тьюрінга
 - Лінгвістичне забезпечення в системах штучного інтелекту
- ▷ Прикладні аспекти штучного інтелекту
- ▷ Штучний інтелект в Україні: дослідження й інжинірингові розробки
- ▷ Штучний інтелект у цивілізаційному вимірі

Штучний інтелект визначається як галузь, метою якої є побудова штучних тварин (або, принаймні, штучних істот, які - у відповідних контекстах - *схожі на тварин*) і, для багатьох, штучних людей (або, принаймні, штучних істот, які - у відповідних контекстах - *схожі на людей*) (Bringsjord, & Naveen, 2022).



Завдання 25. Уважно прочитайте розділ “Природний інтелект (=інтелект людини) і штучний інтелект (=інтелект комп'ютера) як його модель” у підручнику (Карпіловська, 2006, с. 106-114).³⁹



Створіть ментальну мапу інформації, що її подано у розділі. Сформулюйте для себе *a take-home message* із прочитаного. Поставте щонайменше 3 запитання до прочитаного. Підготуйтеся до обговорення прочитаного на семінарі.

³⁹ Завдання базового рівня складності.

Проектні роботи до Розділу 4



Проектна робота 38. [Усі студенти групи є учасниками цієї проектної роботи]

Оберіть одну із запропонованих тем так, аби між усіма студентами групи та обраними темами встановилося взаємно однозначне відношення. На основі авторитетних джерел, дослідіть обрану Вами тему; дайте темі ємну і привабливу назву. Створіть постерну доповідь для обраної Вами теми. Підготуйтеся до її обговорення; будьте готові висловити і обґрунтувати власну думку.⁴⁰

1. Історія штучного інтелекту
2. Підходи до визначення штучного інтелекту
3. Загальний і спеціальний штучний інтелект
4. Природний і штучний емоційний інтелект
5. Підходи до створення систем штучного інтелекту
6. Лінгвістичне забезпечення в системах штучного інтелекту
7. Штучний інтелект в Україні: наукові й інжинірингові розробки, провідні установи, компанії та стартапи
8. Прикладні аспекти штучного інтелекту: інтелектуальні гаджети
9. Прикладні аспекти штучного інтелекту: віртуальна реальність
10. Прикладні аспекти штучного інтелекту: Інтернет-пошук
11. Прикладні аспекти штучного інтелекту: розпізнавання образів
12. Прикладні аспекти штучного інтелекту: автономні автомобілі
13. Прикладні аспекти штучного інтелекту: комп'ютерні ігри

⁴⁰ Семінар(и) намічено як постер-конференцію, де доповіді, що із ними попередньо мають ознайомитися усі учасники, обговорюватимуться у форматі круглого столу.

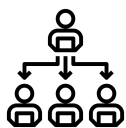
14. Прикладні аспекти штучного інтелекту: відкриті математичні задачі
15. Прикладні аспекти штучного інтелекту: моделювання природної мови
16. Штучний інтелект: історії комерційного успіху
17. Штучний інтелект: філософські передумови, аспекти і виклики
18. Етичний штучний інтелект і моральні роботи
19. Сильний і слабкий штучний інтелект
20. Сингулярність
21. Штучний інтелект: можливості та ризики для *Homo sapiens*
22. Штучний інтелект і технологічне безробіття
23. Суперкомп'ютер Watson від IBM
24. Штучний інтелект AlphaGo від Google DeepMind
25. Штучний інтелект Libratus aka Pluribus від Facebook AI Research
26. Майбутнє штучного інтелекту
27. *Запропонуйте власну - доладну і варту - тему*



Проектна робота 39. Шановні студенти, ця проектна робота відкрита для Ваших ідей відповідно до тематики Розділу 4.

Будьте винахідливі й розвивайте власне почуття напрямку.

ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ

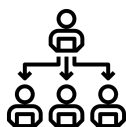


Завдання 1. В *Українській Вікіпедії* запит “Комп’ютерна лінгвістика” перенаправляється на сторінку “Математична лінгвістика.” На Вашу думку, чи допустимий такий стан речей?⁴¹



Якщо так, обґрунтуйте цю допустимість.

Якщо ні, створіть статтю “*Комп’ютерна лінгвістика,*” орієнтовану на Українську Вікіпедію та академічну доброчесність. Представте статтю на семінарі; по підсумках обговорення, створіть в Українській Вікіпедії⁴² сторінку “Комп’ютерна лінгвістика.”



Завдання 2. Онлайн-платформа *Grammarly*: дослідіть витoki, профіль компанії, напрямки діяльності, продукти і сервіси тощо. Будьте винахідливі у своєму дослідженні й здійсніть ґрунтовний, дійсно інтелектуальний веб-пошук, включивши у нього афілійовані ресурси, нещодавні і майбутні

⁴¹ Завдання цього розділу намічені як такі, що можуть бути виконані також у форматі проєктної роботи, або бути заплановані для написання підсумкових есе, курсових проєктів (англ. *a term paper*) тощо.

⁴² Студентам пропонується ознайомитися із поняттям і принципами відкритої науки (англ. *Open Science*) і, зокрема, подивитися на Українську Вікіпедію як на просвітницьку громадську ініціативу поширення наукового знання в Інтернет-доду, із розумінням того, що це знання має бути якісним.

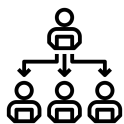
У нашому курсі, студентів запрошено до створення та редагування Вікіпедія-статей на відповідні теми: ця практика, окрім того, що надасть безпрецедентний досвід і дозволить студентам відкрито опублікувати результати своєї роботи, також тренуватиме широкий спектр цифрових і м’яких навичок студентів й сприятиме покращенню якості статей у Вікіпедії. Закликаємо студентів опікуватися своїми статтями у Вікіпедії і після завершення курсу й долучатися до спілкування за інтересами, яке відіграє значну роль як у науковій роботі, так і в професійному житті.

Ми вдячні доценту Тімо Торстену Шмідту, *who is an active Wikipedian*[®], за натхнення залучити роботу студентів із Вікіпедією у викладання.

події, соціальні мережі (зокрема, Фейсбук) й останні активності компанії у них, контакти, особисті сайти фахівців тощо.

Особливу увагу зверніть на **можливості професійного розвитку** (зокрема, ті, що Grammarly традиційно пропонує **українським студентам** - наприклад, воркшопи і літні школи), кар'єри, вимоги до кандидатів, історії успіху, особливості корпоративної культури, сучасний стан і тенденції розвитку ринку праці тощо. Чи відгукуються вони Вам особисто? Чи відгукнулися б Ви їм? Чи можуть слугувати певним дороговказом (англ. *a road map*)?

Структуруйте результати своєї розвідки і представте їх на семінарі. Проведіть для колег розгорнуту онлайн-екскурсію.



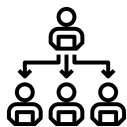
Завдання 3. Дослідіть призначення і функціонал ресурсів, об'єднаних у **“The Stanford Natural Language Processing Group.”**



Також, дослідіть афілійовані до ресурсу групи, нещодавні і майбутні події, соціальні мережі й останні активності у них, контакти, особисті сайти й персональні сторінки учасників тощо - будьте винахідливі у своєму дослідженні й здійсніть ґрунтовний, дійсно інтелектуальний веб-пошук.

Особливу увагу зверніть на **кар'єрні можливості** (зокрема, ті, що пропонуються (міжнародним) студентам), вимоги до кандидатів, особливості корпоративної культури, сучасний стан і тенденції розвитку ринку праці тощо. Чи відгукуються вони Вам особисто? Чи відгукнулися б Ви їм? Чи можуть слугувати певним дороговказом - і з чого цю дорогу розпочати?

Структуруйте результати своєї розвідки і представте їх на семінарі. Проведіть для колег розгорнуту онлайн-екскурсію.



Завдання 4. На основі авторитетних джерел, дослідіть тему **Мови програмування**. Опрацюйте такі аспекти цієї теми, але не обмежуйтеся ними: мова програмування - визначення; класифікація мов програмування; синтаксис і семантика мов програмування; діалекти мов програмування; мова програмування vs. комп'ютерна мова; особливості застосування; затребувані по галузях й поширені у світі мови програмування.

Поміркуйте над співвідношенням мов штучного інтелекту й штучних мов.

Дослідіть призначення і функціонал ресурсів, об'єднаних у **“Github.”**



Дослідіть призначення і функціонал ресурсів, об'єднаних у **“Ontohub.”**

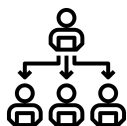


Також, дослідіть афілійовані до ресурсів продукти, послуги, проекти і групи, нещодавні і майбутні події, соціальні мережі й останні активності у них, контакти, особисті сайти й персональні сторінки учасників тощо - будьте винахідливі у своєму дослідженні й здійсніть ґрунтовний, дійсно інтелектуальний веб-пошук.

Особливу увагу зверніть на **кар'єрні можливості** (зокрема, ті, що пропонуються (міжнародним) студентам), вимоги до кандидатів, особливості корпоративної культури, сучасний стан і тенденції розвитку ринку праці тощо. Чи відгукуються вони Вам особисто? Чи відгукнулися б Ви їм? Чи можуть слугувати певним дороговказом?

Знайдіть досяжні Вам пропозиції **вивчення мов програмування**: в українському та міжнародному освітньому (онлайн-)просторі. Зважте на ці пропозиції, порівняйте їх. Чи відгукуються вони Вам?

Структуруйте результати своєї розвідки і представте їх на семінарі. Проведіть для колег розгорнуту онлайн-екскурсію.



Завдання 5. Дослідіть ресурси *IBM Training* і *Microsoft Learn*.

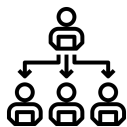


Юні колеги, слова тут зайві. “Як нерозумно випрошувати те, чого можеш сам досягти!” (Г. Сковорода).

Із точки зору (майбутніх) освітян, розгляньте можливості і цифрові рішення, що їх об'єднано у **Microsoft | Освіта**. Принагідно, не забудьте згадати про них у своєму професійному житті.



Структуруйте результати своєї розвідки і представте їх на семінарі.



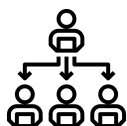
Завдання 6. Зважте на перспективи *викладання англійської мови ІТ-спеціалістам*. Наш курс може стати в неабиякій пригоді.

Дослідіть досяжний Вам сегмент сучасного українського та міжнародного ринків праці English for IT, знайдіть і зважте на можливості такого працевлаштування. Чи відгукуються вони Вам? Які вимоги висуваються до кандидатів? Що треба зробити, аби цим вимогам відповідати?

Обміркуйте відомі Вам підходи до викладання мови, використовувані методи(ки) тощо й оберіть ті, які б Ви як педагог залучили у роботу саме із ІТ-спеціалістами, враховуючи особливості їхнього психологічного типу.

Протягом вивчення дисциплін педагогічного циклу в Університеті, не проминіть можливість розширити свій набір дієвих інструментів викладача.

Структуруйте результати своєї розвідки і представте їх на семінарі.



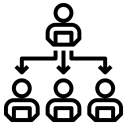
Завдання 7. На основі авторитетних джерел, дослідіть тему *Машинне навчання*.⁴³ Визначте, у чому саме полягає і як здійснюється це навчання. Хто/Що є суб'єктом, а хто/що - об'єктом цього навчання? Що є предметом цього навчання? Що таке глибинне навчання⁴⁴ і якою є його роль у машинному навчанні?

Запрошуємо Вас подивитися широко на освітянство. У чому машинне навчання і Ваша майбутня професія перетинаються?

Структуруйте результати своєї розвідки і представте їх на семінарі.

⁴³ англ. *Machine Learning*.

⁴⁴ англ. *Deep Learning*.



Завдання 8. На основі авторитетних джерел, дослідіть тему **Взаємодія людини і комп'ютера**.⁴⁵ Опрацюйте такі аспекти цієї теми, але не обмежуйтеся ними: взаємодія людини і комп'ютера як інтерфейс; принципи і методології комп'ютерного дизайну; соціальні обчислення й інженерія⁴⁶; емоції і комп'ютер; нейрокомп'ютерні інтерфейси⁴⁷; напрямки досліджень і прикладні аспекти.

Значне місце у Human-computer interaction, а особливо у Cognitive human-computer interaction, посідають психологія і педагогіка. На Вашу думку, чому? Якими саме є **психолого-педагогічні аспекти** цієї взаємодії? Чи відгукується Вам перспектива цю взаємодію медіювати?

На **(нейро)лінгвістичні аспекти** цієї взаємодії Вас можуть найсучасніше скерувати:

Moens, J. (September 8, 2022). An AI can decode speech from brain activity with surprising accuracy. *ScienceNews. Neuroscience*.



Oxley, T. (June 2, 2022). A brain implant that turns your thoughts into text. *TED2022. TED: Ideas Worth Spreading*.

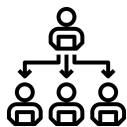


⁴⁵ англ. *Human-Computer Interaction*.

⁴⁶ англ. *Social Computing*.

⁴⁷ англ. *Brain-Computer Interfaces*.

Структуруйте результати своєї розвідки і представте їх на семінарі.



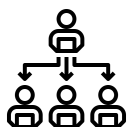
Завдання 9. На основі авторитетних джерел, дослідіть тему **Інформація. Теорія інформації**. Опрацюйте такі аспекти цієї теми, але не обмежуйтеся ними: інформація: визначення і властивості, види і функції; теорія інформації⁴⁸: визначення, витоки, ключові поняття, завдання, міждисциплінарні зв'язки.

Чим (інтелектуальний) інформаційний пошук відрізняється від пошуку інформації? Що таке інформаційний шум?

Осібнo, дайте визначення інформатиці⁴⁹, квантовій інформатиці⁵⁰ й нейроінформатиці⁵¹ через зв'язок цих дисциплін із завданням (квантового) обчислення⁵².

Схарактеризуйте відношення між інформацією і **знанням**. Чому говорять, що “знання - сила,” а “той, хто володіє інформацією, володіє світом”? Як співвідносяться знання і **вміння**?

Структуруйте результати своєї розвідки і представте їх на семінарі.



Завдання 10. На основі авторитетних джерел, дослідіть тему **Експертні системи**. Опрацюйте такі аспекти цієї теми, але не обмежуйтеся ними: експертна система - визначення; витоки й еволюція експертних систем; сучасні підходи до створення; архітектури програмного забезпечення в експертних системах; застосування і функції експертних систем; сильні й слабкі сторони експертних систем.

⁴⁸ англ. *Information Theory*.

⁴⁹ англ. *Information Science*.

⁵⁰ англ. *Quantum Information Science*.

⁵¹ англ. *Neuroinformatics*.

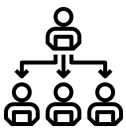
⁵² англ. *(Quantum) Computing*.

Дослідіть і узагальніть світовий досвід створення і використання *експертних систем в освіті*. Яку експертну систему в (українській) освіті Ви б вважали за потрібне, й запропонували б, створити? Як саме Ваші фахові компетентності можуть бути залучені у таке створення? Які жорсткі та м'які, а також цифрові, навички Вам знадобляться?

На Вашу думку, наскільки ефективними є тести на коефіцієнт інтелекту?⁵³ Дослідіть історію, проблематику і підходи до їхнього створення. Чи співвідносяться тести зовнішнього незалежного оцінювання (ЗНО) із тестами на коефіцієнт інтелекту? Проаналізуйте свій досвід складання ЗНО.

Експерт - хто це? Якими якостями має володіти людина, щоб бути визнаною експертом? Поміркуйте над критеріями такого визнання. Людина-експерт vs. експертна система: чи зможуть експертні системи свого часу повністю замінити людей-експертів? Чому?

Структуруйте результати своєї розвідки і представте їх на семінарі.



Завдання 11. Тема *Природна мова: у межах і поза межами розуміння*: подивіться і законспекуйте лекцію Н. Чомскі “Language and the Limits of Understanding” від 30 вересня 2021 року, присвячену Міжнародному дню перекладу.

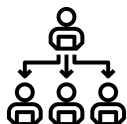


Особливу увагу зверніть на фрагмент лекції з 24 по 30 хвилини. Схарактеризуйте креативний, або генеративний, аспект використання мови, у його витлумаченні Н. Чомскі. Як відбувається генерування думок? Як одна думка обирається із багатьох у процесі мовлення? Чим

⁵³ англ. IQ tests.

генерування думки відрізняється від продукування мовлення? Що саме знаходиться поза межами людського розуміння?

Структуруйте результати своєї розвідки і представте їх на семінарі.



Завдання 12. Тема *Мозок і свідомість: українська нейронаука у медіа-контексті*: на основі відкритих джерел, дослідіть наукову (і творчу) діяльність академіка О.О. Кришталя, прочитайте і подивіться кілька визначних інтерв'ю Олега Олександровича українським медіа-виданням.

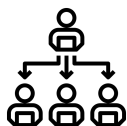
Що Ви дізналися про свій мозок? Про свої мислення і свідомість? Про мову? Хто такий Гомункулус, і якою є мова думки на відміну від мови слів?

Як влаштований і як працює мозок людини? Якою є його робота? Як людині дбайливо й ефективно використовувати свій мозок? Що шкодить мозкові?

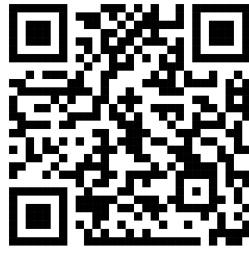
Як роздуми, читання і письмо *стимулюють мозок*, сприяючи появі і розвиткові нових нейронів? Чому “багатство людини має вимірюватися не грошима, а кількістю нейронних зв'язків” (О.О. Кришталь)?

Поясніть, будь ласка, як термін “штучна нейронна мережа” виконує свої функції наукової метафори?

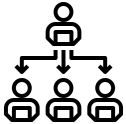
Структуруйте результати своєї розвідки і представте їх на семінарі.



Завдання 13. Тема *Штучний інтелект: (не)фантастична наукова фантастика*: Оберіть і подивіться один або кілька фільмів із запропонованих на ресурсі “Best Artificial Intelligence Movies.” Створіть короткий синопсис, підсумуйте й узагальніть ідею, вбудувавши її у ширший філософський, науковий, науково-популярний або науково-фантастичний контекст. Додайте власні міркування.



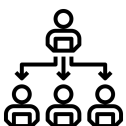
Структуруйте результати своєї розвідки і представте їх на семінарі.



Завдання 14. Знайдіть і поспілкуйтеся (наприклад, у соціальних і професійних мережах, на відповідних форумах тощо) із **випускниками** Вашої спеціальності (як варіант, суміжних або цікавих Вам особисто спеціальностей) про те, як вони влаштовувалися і як зрештою влаштувалися у професії.

Узвіть їх поділитися досвідом працевлаштування, професійного (само)розвитку, побудови кар'єри тощо. Як цьому сприяли здобуті в Університеті знання, вміння і навички? Якою була роль власного досвіду, завзяття, мотивації і відповідальних рішень? *Додайте цікаві Вам питання.*

Структуруйте результати своєї розвідки і представте їх на семінарі.

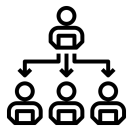


Завдання 15. Чергування на семінарах (англ. *Seminar duty*): протокольний супровід семінарів (англ. *minutes*), рекап (англ. *a recap*) на початку і підведення підсумків (англ. *a take-home message*) у кінці кожного семінару, формулювання питань, організаційна комунікація з учасниками семінару.

Можливість додати тему для обговорення, питання, напрямок, ідею, релевантну літературу й електронні ресурси, які б Ви рекомендували колегам, тощо.

Чергування розраховане як таке, що триває протягом усього семестру вивчення курсу й сприяє розвитку Ваших лідерських якостей і навичок управління проектами.⁵⁴

Наприкінці семестру, звітайте про свій досвід.



Завдання 16. Шановні студенти, це завдання відкрите для Ваших ідей відповідно до тематики курсу.

Будьте винахідливі й розвивайте власне почуття напрямку.

⁵⁴ Зацікавлених студентів закликаємо ознайомитися на дозвіллі із методологіями управління проектами: Agile, Kanban і Scrum (зверніться також до описів вакансії (англ. *a job description*) Scrum-master на українському та міжнародному ринках праці - якими є вимоги?); а також, ознайомитися із (цифровими) інструментами управління часом і управління знанням, використовуваними для командної роботи у бізнес-середовищі, наприклад, Trello, Jira, Slack, Sharepoint, Notion тощо.

ПІСЛЯМОВА

Цей посібник є спробою схопити академічний дух тих найкращих практик вищої освіти, що ми їх засвоїли протягом свого навчання, стажувань і викладання у можних університетах України, Німеччини і Польщі та використовуємо, зокрема, у викладанні курсу комп'ютерної лінгвістики, що його покликаний супроводжувати цей посібник.

Курс базується на нашому досвіді вивчення компютерної лінгвістики в Інституті когнітивної науки Оснабрюцького університету і надихнутий перебуванням саме у цьому академічному середовищі. Шанобливо вдячні Вчителю і нашим викладачам!

До наших студентів звертаємося із тим, що розуміємо і поділяємо Вашу охоту до якостей європейської освіти і прагнемо Вам усіляко у набутті цих якостей сприяти. Через це прагнення, дозволимо собі наголосити на тій активній, відповідальній ролі, яку студенти європейських університетів радо беруть на себе у навчанні й яку, на нашу думку, українські студенти поки що не зовсім налаштовані помічати: оспівуючи академічні свободи, варто було б оспівати - чи не у першу чергу - й академічні добродетності. Однією із таких чеснот вважаємо відповідальність, що її доросла людина несе за своє навчання, і мотивацію до такого навчання: навіть *-ся* у слові “вчитися” означає, що вчать себе. Закликаємо студентів на дозвіллі поцікавитися етимологією слова “студент” і звернутися до витоків й перекладів “*Gaudeamus igitur*” - потверджуємо, юні колеги, ту неабияку мудрість, що там міститься.

Курс комп'ютерної лінгвістики запрошує Вас ширше подивитися на обрану спеціальність, на її сутність і можливості застосування здобутих в Університеті знань і вмінь, й зважити на перспективи самореалізації, які ця спеціальність Вам надає - у вимірі фахових і загальних компетентностей. Вибудова цих компетентностей - у фундаментальній університетській освіті - передуює вбудовуванню свого професійного талану у життя, у той реальний контекст, що оточуватиме Вас на момент працевлаштування й у подальшій трудовій діяльності. Протягом курсу, закликатимемо Вас усвідомлювати особливості цього контексту: дивитися на себе як на кандидата очима HR, складати CV, набувати досвід із work.ua і dou.ua, заглядати на michaelpage.com, проходити job interviews, шукати professional networking у LinkedIn, долучатися до освітніх і профорієнтаційних подій, навідуватися до coursera.org і udemy.com, визначатися з пріоритетами розвитку себе, пам'ятати “отже, веселімося” і зрештою впевнитися, що на момент випуску з Університету Ви встигли

створити для себе достатньо можливостей набуття практичного досвіду за спеціальністю, аби не лише все знати, але й дещо вміти...

“З усіх утрат втрата часу найтяжча” (Г. Сковорода). Юні колеги, бажаємо Вам пожити час навчання із щонайбільшою користю!

О.В. Ваховська

ЛІТЕРАТУРА

ОСНОВНА

- Балог, В. (2005). Сучасний стан української комп'ютерної лінгвістики. *Лексикографічний бюлетень*, 11, 28-35.
- Бараннік, В.В., Белікова, Т.В., Капко, М.О., & Гуржій, І.А. (2019). Комплексний метод автоматичного фоносемантичного аналізу текстової інформації на основі оцінки вагомих семантичних одиниць в умовах інформаційного протиборства. *Кібербезпека: освіта, наука, техніка*, 3(3), 53-62. <https://doi.org/10.28925/2663-4023.2019.3.5362>
- Вавіленкова, А.І. (2016). Концепція автоматизованого лінгвістичного аналізу тексту. *Теоретичні основи аналізу електронних текстів: монографія*. К.: ТОВ «СІК ГРУП УКРАЇНА.» С. 175-181.
- Вавіленкова, А.І. (2017). Структура системи порівняльного аналізу електронних текстових документів за змістом. *Вісник Національного університету «Львівська політехніка.» Комп'ютерні науки та інформаційні технології*, 864(1), 277-285.
- Войтенко, Я.С. (2020). Огляд методів семантичного аналізу текстів. Використання методів семантичного аналізу текстів в мережі Інтернет. *Дослідження методів семантичного аналізу текстів для інтелектуалізації Web-сайтів: Магістерська робота: 8.05010104 - Системи штучного інтелекту*. Харків: Харківський національний університет радіоелектроніки. С. 10-30.
- Волошин, В.Г. (2004). *Комп'ютерна лінгвістика: Навчальний посібник*. Суми: ВТД «Українська книга.»
- Гирич, О.В. (2017). Автоматичний синтаксичний аналіз англійського речення: застосування та перспективи. *Вісник Житомирського державного університету імені Івана Франка*, 1(85), 26-30.
- Гирич, О.В. (2020). Основні проблеми систем обробки природної мови. *Філологічні студії. Збірник наукових праць*, 14, 41-45.
- Дарчук, Н.П. (2008). *Комп'ютерна лінгвістика (автоматичне опрацювання тексту): Підручник*. Київ: Видавничо-поліграфічний центр «Київський університет.»
- Дарчук, Н.П. (2013). Автоматичний синтаксичний аналіз текстів корпусу української мови. *Українське мовознавство*, 43, 11-20.
- Дарчук, Н.П. (2018). До питання про створення автоматичного словника словосполучень української мови. *Лінгвістичні студії*, 36, 148-158. <https://doi.org/10.31558/1815-3070.2018.36.23>

- Дарчук, Н.П. (2020). АГАТ-синтаксис як компонент комп'ютерної граматики української мови. *Українська мова. Науково-теоретичний журнал*, 4(76), 89-103. <https://doi.org/10.15407/ukrmova2020.04.089>
- Дарчук, Н., Зубань, О., Лангенбах, М., & Ходаківська, Я. (2016). АГАТ-семантика: семантичне розмічування корпусу української мови. *Українське мовознавство*, 1(46), 92-102.
- Жаботинська, С.А. (2019). Семантика лінгвальних мереж у навчальному комбінаторному тезаурусі. *Studia Philologica. Філологічні студії*, 13, 17-27. doi: 10.28925/2311-2425.2019.13.3
- Жаботинська, С.А. (2021). Типи онтологій для наративних концептів (на прикладі медійного наративного концепту COVID-19). *Виклики та парадокси соціальної взаємодії в постмодерному світі: лінгвістичні та психологічні аспекти: II Міжнародна науково-практична конференція*. Луцьк: Волинський національний університет імені Лесі Українки. С. 17-21.
- Завадська, В. (2013). Коли “вікно” не є вікном, або ще раз про сучасну українську ІТ-термінологію. *Українське мовознавство*, 43, 20-27.
- Зубань, О. (2018). Корпус української мови - комп'ютерна експертна система лінгвістичного аналізу українськомовного тексту. *ТЕКА Komisji Polsko-Ukraińskich Związków Kulturowych*, 6(13), 191-206. <https://doi.org/10.31743/teka.5723>
- Калько, Р.М. (2017). Універсалізм Н. Хомського у межах когнітивної парадигми світової лінгвістики. *Наукові записки Національного університету «Острозька академія»: Серія «Філологія»*, 1(64), 143-146.
- Карпіловська, Є.А. (2006). *Вступ до прикладної лінгвістики: Комп'ютерна лінгвістика: Підручник*. Донецьк: ТОВ «Юго-Восток, Лтд.»
- Карпіловська, Є.А. (2014). Комп'ютерна лінгвістика. *Енциклопедія Сучасної України : енциклопедія [електронна версія] ; ред.: І.М. Дзюба, А.І. Жуковський, М.Г. Железняк та ін.; НАН України, НТШ. Київ : Інститут енциклопедичних досліджень НАН України. Т. 14. URL: <https://esu.com.ua/article-4396> (дата перегляду: 28.11.2022)*
- Коломієць, В.О. (2022). *Короткий англо-український словник з комп'ютерної лінгвістики*. Київ: Видавничий центр КНЛУ.
- Комарницька, О. (2018). Методи автоматизованого семантичного аналізу природномовної інформації. *Філологічний дискурс*, 7. 92-100.
- Кришталь, О.О. (2020). *До співу пташок: приватна подорож до себе*. Київ: Паливода А.В.
- Кришталь, О.О. (2022). *Гомункулус*. (Готується до публікації українською мовою).

- Лангенбах, М. (2015). Автоматичний синтаксичний аналіз речення за принципами граматики залежностей. *Науковий вісник Східноєвропейського національного університету імені Лесі Українки*, 3, 249-254.
- Лозинська, О.В., Давидов, М.В., & Пасічник, В.В. (2016). Трансформація дерев граматики складових у дерева граматики залежностей для граматичного розбору українських речень. *Штучний інтелект*, 3, 22-31.
- Ляхоцький, О.С. (2016). Огляд технологій Semantic Web. *Програмний модуль для побудови семантичних порталів: Магістерська робота: 8.05010302 - Інженерія програмного забезпечення*. Тернопіль: Тернопільський національний економічний університет. С. 10-33.
- Марчук, К.І. (2017). Автоматичний синтаксичний аналіз англomовного тексту: застосування та перспективи. *Сучасні філологічні дослідження та навчання іноземної мови в контексті міжкультурної комунікації*, 10, 328-333.
- Міщенко, А. (2013). Створення паралельного банку дерев для німецької та української мов. *Українське мовознавство*, 43, 51-63.
- Парнус, К.І., Свісюк, О.В. (2022). Алгоритм аналізу англomовного тексту за допомогою парсерів. *SWorldJournal*, 14(01), 108-113. <https://doi.org/10.30888/2663-5712.2022-14-01-020>
- Партико, З.В. (2008). *Прикладна і комп'ютерна лінгвістика: Вступ до спеціальності: Навчальний посібник*. Львів: Афіма.
- Пещак, М.М. (1999). *Нариси з комп'ютерної лінгвістики*. Ужгород: Закарпаття.
- Приймак Д.М. (2021). *Принципи укладання двомовних термінологічних словників: перекладацький аспект (на матеріалі англійської, французької, італійської та української мов)*: Кваліфікаційна наукова праця на правах рукопису. Дисертація на здобуття наукового ступеня доктора філософії в галузі філології за спеціальністю 10.02.16 - перекладознавство (035 - Філологія). Київ: Київський національний лінгвістичний університет.
- Путіліна, О. (2012). Універсальна та відмінкова граматики: корелятивно-ієрархічні відношення. *Лінгвістичні студії. Збірник наукових праць*, 24, 15-20.
- Романюк, А., Романишин, М. (2013). Тональний словник української мови на основі сентимент-анотованого корпусу. *Українське мовознавство*, 43, 63-75.
- Ситар, Г. (2015). Конструкційна граMATика як теоретичне підґрунтя дослідження фразеологізованих речень. *Типологія та функції мовних одиниць. Науковий журнал на пошану член-кореспондента НАН України Івана Романовича Вихованця*, 2(4), 192-205.

- Титенко, С.В. (2006). Проблема подання знань на основі природної мови у освітніх системах штучного інтелекту. *Лабораторія CET - Семантичні технології в електронному навчанні. Штучний інтелект в освіті*. Режим доступу: http://www.setlab.net/?view=Philosophy_Knowledge
- Толубко, В.В., Шипнівська, О.О., & Ляшенко, А.В. (2010). Задачі автоматичної обробки синтаксичної структури в знання-орієнтованій системі машинного перекладу. *Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка*, 27, 136-141.
- Шведова М.О. (2022). Застосування корпусу у викладанні української мови як іноземної. *Слов'янські мови, культури, переклад і викладання: сучасний стан і перспективи*, 227-232.
- Шипнівська, О.О. (2013). Визначення синтаксично неоднозначних конструкцій в україномовному тексті. *Мовознавство*, 4, 67-76.
- Шипнівська, О.О. (2013). Визначення типів синтаксичної неоднозначності у знаннеорієнтованій системі машинного перекладу. *Українське мовознавство*, 43, 104-113.
- Шомко, А.В. (2017). Генеративна граматики як галузь лінгвістичної науки. Проблеми та перспективи її розвитку. *Сучасні філологічні дослідження та навчання іноземної мови в контексті міжкультурної комунікації*, 10, 355-359.
- Яровий, А., Кудрявцев, Д., & Крилик, Л. (2020). Удосконалення методу семантичного аналізу тексту. *Інтернет-Освіта-Наука-2020: XII Міжнародна науково-практична конференція*. Вінниця: Вінницький національний технічний університет. С. 2-4.
- Bolshakov, I., & Gelbuch, A. (2004). *Computational Linguistics: Models, Resources, Applications*. Electronic edition IPN-UNAM-FCE.
- Bringsjord, S., & Naveen, S.G. (2022). Artificial Intelligence. *The Stanford Encyclopedia of Philosophy*. E.N. Zalta & U. Nodelman (Eds.). Режим доступу: <https://plato.stanford.edu/archives/fall2022/entries/artificial-intelligence/>
- Clark, A., Fox, C., & Lappin, S. (2010). *The Handbook of Computational Linguistics and Natural Language Processing*. Blackwell Publishing Ltd.
- Dunlop, C., & Fetzer, J. (1993). *Glossary of Cognitive Science*. Saint Paul, MN: Paragon House.
- Gallego, A.J., & Martin R. (2020). *Language, Syntax, and the Natural Sciences*. Cambridge University Press.
- Gillis, A.S. (2022). *What is Computational Linguistics?* Режим доступу: <https://www.techtarget.com/searchenterpriseai/definition/computational-linguistics-CL>

- Grishman, R. (1986). *Computational Linguistics: An Introduction*. Cambridge University Press.
- Hammond, M. (2020). *Python for Linguists*. Cambridge University Press.
- Hausser, R. (2014). *Foundations of Computational Linguistics: Human-Computer Communication in Natural Language*. Heidelberg: Springer.
- Hunston, S. (2022). *Corpora in Applied Linguistics*. Cambridge University Press.
- McCulloch, G. (2020). *Because Internet: Understanding how Language is Changing*. Vintage Publishing.
- Moens, J. (September 8, 2022). An AI can decode speech from brain activity with surprising accuracy. *ScienceNews. Neuroscience*.
- Oxley, T. (June 2, 2022). A brain implant that turns your thoughts into text. *TED2022. TED: Ideas Worth Spreading*.
- Peters, S., & Shrobe, H.E. (2003). Using semantic networks for knowledge representation in an intelligent environment. *IEEE International Conference on Pervasive Computing and Communications (PerCom'03), 1*, 323-337. doi: 10.1109/PERCOM.2003.1192756
- Poole, G. (2011). *Syntactic Theory*. Bloomsbury Publishing PLC.
- Powell, M. (2019). *Grammar Geek*. Octopus Publishing Group.
- Powell, M. (2019). *Word Nerd*. Octopus Publishing Group.
- Schubert, L. (2020). Computational Linguistics. *The Stanford Encyclopedia of Philosophy*, E.N. Zalta (Ed.). Режим доступу: <https://plato.stanford.edu/archives/spr2020/entries/computational-linguistics/>
- Shvedova, M.O. (2020). The General Regionally Annotated Corpus of Ukrainian (GRAC, uacorporus.org): Architecture and Functionality. *The 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2020)*. Lviv. P. 489-506.
- Vakhovska, O.V. (2021). Metaphors of Emotions: Towards a Data-driven Formalization. *Вісник Київського національного лінгвістичного університету. Серія Філологія*, 24(1), 33-48. doi: 10.32589/2311-0821.24%20(1).2021.236030
- Vakhovska, O.V. (2022). Must the Pot of Seven Holes be a Translation Theorist's Riddle? *KELM Knowledge, Education, Law, Management*, 4(48), 117-125. DOI <https://doi.org/10.51647/kelm.2022.4.19>

ДОДАТКОВА

- Карпенко, Ю.О. (2006). *Вступ до мовознавства*. Київ: Академія.
- Кочерган, М.П. (2006). *Вступ до мовознавства: Підручник*. Київ: Академія.

- Селіванова, О.О. (2008). *Сучасна лінгвістика: напрями та проблеми: Підручник*. Полтава: Довкілля-К.
- Акмаджян, А., Deemers, R.A., Farmer, A.K., & Harnish, R.M. (1995). *Linguistics. An Introduction to Language and Communication*. MIT Press.
- Fromkin, V. (2000). *Linguistics. An Introduction to Linguistic Theory*. Blackwell. Oxford.
- O'Grady, W., Archibald, J., Aronoff, M., & Rees-Miller, J. (2006). *Contemporary Linguistics. An Introduction. 5th edition*. Bedford/St.Martin's, Boston & New York.
- Pinker, S. (1994). *The Language Instinct*. London: Penguin.
- Radford, A., Atkinson, M., Britain, D., Clahsen, H., & Spencer, A. (1999). *Linguistics. An Introduction*. Cambridge UP.
- (Доповніть список тією літературою, яку Ви мали нагоду додатково опрацювати протягом вивчення курсу.)

ЕЛЕКТРОННІ РЕСУРСИ

ОСНОВНІ

[Вікіпедія](#)

[Головна | Освіта Microsoft](#)

[ГРАК](#)

[Корпус української мови](#)

[AlphaGo](#)

[Best Artificial Intelligence Movies: 20+ Titles To Watch Now](#)

[Coursera](#)

[DOU](#)

[Facebook, Carnegie Mellon build first AI that beats pros in 6-player poker](#)

[Grammarly](#)

[GitHub](#)

[IBM Training](#)

[IBM Watson](#)

[Language and the Limits of Understanding - International Translation Day 2021](#)

[LinkedIn](#)

[Michael Page](#)

[Ontohub.org](#)

[Parse a sentence](#)

[Stanford NLP Group](#)

[Training | Microsoft Learn](#)

[Udemy](#)
[Work.ua](#)

ДОДАТКОВІ

(Доповніть список тими електронними ресурсами, із якими Ви мали нагоду додатково ознайомитися протягом вивчення курсу.)

Підписано до друку 28.03.2023 р. Формат 60x84 1/16
Спосіб друку офсетний. Умовн. друк. арк. 7,67
Умовн. фарбо-відб. 7,78 Обл.-вид. арк. 7,78
Тираж 100. Зам. № 23 - 11

Видавничий центр КНЛУ
Свідоцтво: серія ДК 1596 від 08.12.2003 р.

Віддруковано "Видавництво Ліра-К"
03115, Київ, вул. Василя Стуса 22/1
Свідоцтво про внесення до державного реєстру
Серія ДК № 3981.