

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЦЕНТРАЛЬНОУКРАЇНСЬКИЙ ДЕРЖАВНИЙ ПЕДАГОГІЧНИЙ
УНІВЕРСИТЕТ ІМЕНІ ВОЛОДИМИРА ВИННИЧЕНКА

ФАКУЛЬТЕТ ІНОЗЕМНИХ МОВ

**Кафедра перекладу, прикладної та
загальної лінгвістики**

КУРС ЛЕКЦІЙ
«ВСТУП ДО ПРИКЛАДНОЇ ЛІНГВІСТИКИ»
для студентів факультету іноземних мов
спеціальності 035.10 Прикладна лінгвістика (англійська, німецька
мови/німецька, англійська мови)

Кропивницький – 2020

Курс лекцій «Вступ до прикладної лінгвістики» для студентів факультету іноземних мов спеціальності 035.10 Прикладна лінгвістика (англійська, німецька мови/німецька, англійська мови) / Укл. Тарнавська М.М. – Кропивницький: РВВ ЦДПУ ім. В.Винниченка, 2020 – 45 с.

Рецензенти: Кочубей Вікторія Юріївна, кандидат філологічних наук, доцент кафедри англійської мови та методики її викладання;

Пянковська Ірина Володимирівна, кандидат філологічних наук, доцент кафедри німецької мови та методики її викладання.

Матеріали містять тексти лекцій з курсу «Вступ для прикладної лінгвістики». У курсі детально розглянуті базові поняття прикладної лінгвістики, надані дефініції, що складають понятійний апарат цієї галузі мовознавства, окреслені основні напрямки сучасної прикладної науки, перспективи її розвитку. Кожна лекція побудована за планом, а також супроводжується питаннями для контролю засвоєння лекційного матеріалу.

Затверджено методичною радою
Центральноукраїнського державного
педагогічного університету ім. В.Винниченка
Протокол №4 від 18 червня 2020 року

ПОЯСНЮВАЛЬНА ЗАПИСКА

«Вступ до прикладної лінгвістики» є базовою дисципліною для студентів спеціальності «Прикладна лінгвістика», оскільки знайомить їх з основними категоріями цієї галузі мовознавства та напрямками її розвитку, як з точки зору сучасності, так і в історичному контексті. Курс дає студентам можливість дізнатися про потенційні сфери діяльності спеціаліста з прикладної лінгвістики, а також сформуванню чітких та об'єктивних уявлень про базові задачі, що стоять перед цією наукою та перспективи їх розв'язання. «Вступ до прикладної лінгвістики» є частиною блоку загальнотеоретичних дисциплін, які вивчаються студентами спеціальності «Прикладна лінгвістика», тому матеріал курсу тісно пов'язаний з понятійним апаратом інших лінгвістичних дисциплін, таких як «Вступ до мовознавства» та «Вступ до перекладознавства».

Курс лекцій «Вступ до прикладної лінгвістики» містить основні поняття прикладної лінгвістики, їх дефініції та тлумачення у даній конкретній галузі. Понятійний апарат засвоюється студентами поступово, поняття ускладнюються та пов'язуються між собою для формування у майбутнього фахівця понятійного апарату, яким він може послуговуватися, здобуваючи практичні знання у сфері прикладної лінгвістики. У курсі подані базові історичні дані про те чи інше явище, але акцент зроблено на сучасних здобутках галузі та проблемах, що є актуальними наразі, а також на перспективах розв'язання цих проблем.

Кожна лекція побудована за планом, що розкриває один з аспектів прикладної лінгвістики. Наприкінці кожної лекції вміщено питання для перевірки рівня засвоєння матеріалу.

Курс лекцій «Вступ до прикладної лінгвістики» може бути використаний студентами спеціальності для кращого засвоєння тем, що вивчаються у курсі, як посібник для виконання самостійної роботи та практичних завдань курсу, а також для повторення та узагальнення знань з дисципліни під час підготовки до тематичного контролю.

ЛЕКЦІЯ 1

ТЕОРЕТИЧНА І ПРИКЛАДНА ЛІНГВІСТИКА. ЗАДАЧІ І НАПРЯМКИ ПРИКЛАДНОЇ ЛІНГВІСТИКИ

План

- I. *Лінгвістика як наука. Система лінгвістичних дисциплін і напрямків.*
- II. *Теоретична і прикладна (практична) лінгвістика. Задачі і напрямки прикладної лінгвістики.*
- III. *Зв'язок лінгвістики з іншими науками – природничими і гуманітарними. Зв'язок лінгвістики з інформатикою.*

I. Лінгвістика як наука. Лінгвістика (мовознавство) - це наука про природну людську мову взагалі і про всі мови світу як індивідуальних її представників.

Як будь-яка наука, мовознавство виникло у зв'язку з практичними потребами людей. Вважається, що „паростки” лінгвістики з'явилися з появою писемності - 2-3 тисячі років до н.е. на Давньому Сході. Першим практичним лінгвістичним результатом були примітивні словники (Месопотамія), а теоретичним – граматики санскриту (Індія, 5 ст. до н.е.). У Давній Греції мова розглядалася з теоретичної точки зору у тісному зв'язку з логічним мисленням (Аристотель). Саме ж мовознавство оформилося як наука лише в 17 -18 століттях (Гумбольдт), що було обумовлено його величезною складністю і браком знань про мову – об'єкту цієї науки. У ХІХ-ХХ ст. спостерігається сплеск наукової активності у сфері мовознавства. **Фердинанд де Соссюр** розробляє структурну лінгвістику. **Бодуен де Куртене** вивчає фонетику і фонологію. Термін «Прикладна лінгвістика» з'явився лише у 20-х роках ХХ століття з появою більш досконалих технічних засобів.

Поступово лінгвістика розвилася у складну і розгалужену систему дисциплін як теоретичного, так і прикладного характеру. Наведемо деякі, найважливіші з них.

Система основних лінгвістичних дисциплін:

1. Загальне мовознавство – вивчає властивості мови взагалі, загальні елементи мови.
2. Описове мовознавство – вивчає сучасні мови, їхню структуру в даний час.
3. Діалектологія – вивчає місцеві територіальні різновиди однієї мови.
4. Порівняльно-історичне мовознавство (компаративістика).
5. Типологія (універсалізм) – порівняльне вивчення структурних і функціональних властивостей мов з погляду деяких обраних спільних рис.
6. Ареальна лінгвістика – досліджує поширення мовних явищ, виявляє ареали мов.
7. Польова лінгвістика – спостерігає за мовою чи діалектом у безпосередньому середовищі носія мови.
8. Дешифрування – досліджує тексти на незнайомому коді (мові) для одержання інформації.
9. Інтернаціональна лінгвістика – вивчає міжнародні мови як засіб спілкування між людьми.
10. Паралінгвістика – вивчає немовні засоби в мові.
11. Етнолінгвістика – вивчає мову у її відношенні до культури народу.
12. Психолінгвістика – пов'язана з вивченням мови, з точки зору мовоутворення.
13. Соціолінгвістика – вивчає роль мови в суспільстві, вплив суспільства на мову.

14. Прикладна лінгвістика – до її складу входять математична лінгвістика (структурна), комп'ютерна лінгвістика (машинний переклад, мови програмування, лінгвістичні основи інформатики, інформаційний пошук), лінгводидактика (навчання мові) та ін.

II. Теоретична та прикладна лінгвістика. Одним з найважливіших напрямків розвитку теоретичної лінгвістики (або «Теорії мовознавства») є прикладна лінгвістика, тобто реалізація традиційних лінгвістичних задач на практиці. Прикладна лінгвістика також займається розробкою методів рішення практичних задач, пов'язаних з використанням мови.

Прикладні сфери мовознавства здавна відрізнялися широкою різноманітністю. Найбільш прадавні з них – письмо (графіка), методика навчання рідній й нерідній мовам, лексикографія. Надалі з'явилися переклад, дешифрування, орфографія, транслітерація, розробка термінології. Один із традиційних напрямків прикладної лінгвістики – участь у мовній політиці держави. Мовна політика включає:

- 1) розробку алфавіту, принципів орфографії, ліквідацію неосвіченості;
- 2) вибір і встановлення державної мови;
- 3) визначення положення інших мов стосовно мови державної;
- 4) розробку, уніфікацію й стандартизацію національної термінології;
- 5) питання найменування й перейменування (міст, вулиць, площ).

Одночасно з розвитком і вдосконалюванням класичних прикладних областей мовознавства у другій половині ХХ ст. намітилася й визначилася **низка нових напрямків прикладної лінгвістики**. Ці напрямки являють собою логічне продовження історичного розвитку прикладної лінгвістики на новому етапі, обумовленому сучасною науково-технічною революцією, що характеризується зміцненням взаємозв'язків суспільних, природних і технічних наук.

У цілому прикладні аспекти лінгвістичного забезпечення різноманітних сфер людської діяльності зводяться насамперед до однієї загальної проблеми – проблеми обробки інформації, що функціонує в суспільстві. Це й текстова інформація в її письмовому вигляді, і усне мовлення як найбільш звичний спосіб комунікації. Особлива роль мовознавства у вирішенні практичних проблем і потреб суспільства визначається самою сутністю природної людської мови, що є унікальним засобом зберігання й передачі інформації.

Актуальні практичні проблеми й потреби суспільства – це впровадження у виробництво верстатів і приладів, керованих голосом оператора, систем автоматизованого пошуку й обробки інформації на різних мовах, удосконалювання організації зв'язку (телефонного, радіозв'язку та ін.), використання фонетичних даних при лікуванні порушень мовлення й слуху.

Названі проблеми містять у собі ряд нових напрямків прикладної лінгвістики, пов'язаних з автоматичною обробкою тексту (мовлення), із широким використанням комп'ютерів, таких, як лінгвістичне забезпечення інформаційних систем різних типів; машинний переклад; комп'ютеризація навчання; розробка систем, що розуміють природню мову (лінгвістичні завдання в системах штучного інтелекту); розробка систем використання інформації, що міститься у звуковому мовному сигналі.

У зв'язку з розробкою лінгвістичного забезпечення інформаційних систем різних типів в останні роки особливу актуальність набула проблема уніфікації та стандартизації термінології. Це пов'язано з різким зростанням кількості наукових і технічних термінів, що обслуговують різні сфери знання. Природно, що при цьому не

залишається осторонь і лексикографія, для якої питання, зв'язані зі стандартизацією термінології, є найвищою мірою актуальними. Поступово із традиційної лексикографії все чіткіше виділяється спеціальна сфера лексикографічної діяльності, пов'язана із широким застосуванням комп'ютерів, – автоматична лексикографія.

Під час розв'язання низки практичних завдань потрібна ефективна взаємодія людини з комп'ютером у процесі діалогу природньою мовою. Такі завдання розробки систем «питання-відповідь», керування роботом, забезпечення діалогу в процесі прийняття рішень у системі керування. Основна проблема тут полягає у тому, щоб забезпечити розуміння письмового тексту або усного мовлення (якщо використовується усне введення) автоматичним обладнанням.

Прикладна лінгвістика виникла в результаті гострої потреби знати, що таке мова у дії, у комунікації, оскільки теоретична лінгвістика вивчала переважно мову у її стані, у системі. При цьому вона довго цуралася свого споріднення з науково-технічною й діловою прозою. Не було описів, моделей, орієнтованих на власне прикладні завдання. Нарешті, в 70-х роках ХХ ст. дослідники дійшли висновку, що вирішення багатьох прикладних проблем не може бути чисто лінгвістичним, а лежить на зовсім інших шляхах, на шляхах моделювання поведінки й мислення людини, семантики, синтезу формальних і семантичних засобів мови. Так з'явилася одна з найважливіших міжгалузевих фундаментальних проблем прикладної спрямованості – проблема моделювання знань.

Моделювання знань останнім часом виявилось сферою інтересів різних наук: логіки, лінгвістики, математики, психології, кібернетики. Причина такого загального інтересу до цієї проблеми одна – створення автоматичних систем штучного інтелекту. У той же час знання нам дані у мові й через мову. Знання існують як у текстах спонтанного мовлення (монолог, діалог, репліки в різних ситуаціях), так і в письмових текстах (пам'ятки прадавньої писемності, художня література, наукова й технічна література). Через мову ми передаємо свої знання від покоління до покоління. Мова, таким чином, виступає й. як форма збереження знань, і як засіб і спосіб їх передачі. У науці й техніці, очевидно, взагалі немає знань поза текстами, а моделювання семантики науково-технічних текстів – це моделювання системи знань цієї галузі. Так, поступово проблема моделювання знань зблизилася із проблемою моделювання змісту тексту. І тут центральним є питання про створення семантичних представлень тексту.

У той же час питання про те, як вирішувати те або інше конкретне прикладне завдання у сфері технічної і ділової комунікації, визначається саме тим, наскільки детально описані тексти цих документів, їх структура, граматики, лексика, семантика, чи є термінологічний словник, чи виявлена статистична структура тексту, чи побудовані семантичні представлення для текстів цього типу.

Отже, визначимо **коло основних задач прикладної лінгвістики**:

1. Переклад з/на іноземну мову.
2. Навчання іноземній мові (методики навчання), лінгводидактика.
3. Комунікація за допомогою технічних засобів.
4. Створення штучних мов.
5. Інформаційний пошук.
6. Анотування і реферування текстів.
7. Укладання словників (практична лексикографія).
8. Упорядкування, стандартизація й уніфікація науково-технічної термінології.

9. Організація бібліографічної інформації.
10. Ефективний типографський набір.
11. Транскрибування усного мовлення.

У наш час упровадження нових інформаційних технологій в усі сфери людського спілкування прикладна лінгвістика розвивається у напрямку **автоматизації основних задач**, а саме:

1. Машинний переклад + машинні словники.
2. Комп'ютерна лінгводидактика.
3. Комп'ютерна лінгвістика (усі додатки лінгвістики в комп'ютерних середовищах).
4. Математична лінгвістика (розробка формальних моделей мов).
5. Автоматична обробка природних мов:
 - а) розпізнавання і синтез мови,
 - б) автоматизація інформаційних робіт,
 - в) автоматичні системи інформаційного пошуку,
 - г) квантитативна лінгвістика (частотний аналіз текстів).

III. Зв'язок лінгвістики з іншими науками. Лінгвістика – складна багатопланова наука, що дотична до багатьох наук, і навіть з'явилися лінгвістичні напрямки на стиці тих чи інших наук, що активно розвиваються за рахунок нових теоретичних і практичних додатків.

Багато теоретичних понять і методів дослідження й аналізу, що уперше з'явилися серед гуманітарних наук у структурній, математичній і прикладній лінгвістиці, сьогодні усе більш активно проникають у літературознавство, бібліотекознавство, історію, етнографію, психологію, діловодство. У різних гуманітарних науках спостерігається багато спільного й у спрямованості до автоматизації досліджень.

Автоматизація наукових досліджень у гуманітарних науках повинна йти насамперед по лінії створення більших банків зберігання документальних і фактографічних даних. Такий тип автоматизації визначається насамперед характером основних інформаційних потреб фахівців з гуманітарних наук. Якого роду довідки найчастіше потрібні науковцеві-гуманітарієві? По перше, це запити бібліотечно-бібліографічного типу: знайти й видати всю вітчизняну й закордонну літературу (книги, статті, тези конференцій, хронікальні замітки) з окремого питання. Наприклад, археологові потрібна література про стародавні прикраси певного виду, етнографові – про типи господарських будівель у Північній Європі, лінгвістові – про сполучуваність морфем.

По друге, в археології, історії, етнографії, антропології, мовознавстві в ході дослідження найчастіше буває потрібно зробити добір як опублікованих, так і неопублікованих джерел, що містять матеріали того або іншого типу. Робота ця, як правило, у край трудомістка. Археологові потрібно, наприклад, зібрати всі конкретні описи предметів текстильної кераміки, етнографові – відшукати всі джерела, що містять опис знарядь рибальства в Східній Європі, філологів – усі текстові джерела, що відбивають, наприклад, конфесійно-оповідальний тип українських текстів XVI в, істориків – усі роботи про торгівлю українських монастирів в XVI-XVII ст. Добір джерел у фундаментальних гуманітарних дослідженнях не знає ані хронологічних, ані географічних, ані мовних меж. Більше того, нерідко в край цінна, власне фактографічна інформація міститься саме у старих джерелах.

По третє, перед дослідником-гуманітарієм стоїть завдання відбору із джерел «своїх» фактів з іншого матеріалу. Так, наприклад, археологові з величезної кількості виробів з кераміки, зібраних в ХІХ-ХХ вв., потрібно відібрати тільки об'єкти «свого» розгляду, наприклад, кераміку саме з відбитками тканини, етнографові – відомості про дерев'яний посуд, філологів – слова із суфіксом - *ість* або дієслова минулого часу. В археології, етнографії, антропології, частково у лінгвістиці є безліч даних, конкретних фактів, отриманих «у полі», в експедиціях, що звичайно зберігаються в неопублікованому вигляді – в архівах, фондах музеїв, інститутів, у ручних описах і картотеках. Неопубліковані матеріали такого роду, розкидані до того ж по різних краях і містам, – безцінне джерело будь-якого наукового дослідження в області історико-філологічних наук у цілому. Визначити тип, характер і генезис того або іншого явища в гуманітарних науках можна тільки на широкому порівняльно-історичному й типологічному тлі.

Особливе значення набувають окреслені вище проблеми в міждисциплінарних, комплексних дослідженнях. Так, наприклад, сучасна історична географія оперує одночасно фактами з фізичної й економічної географії, археології, історії, етнографії, демографії. Фахівцеві з теорії етногенезу й етнічної історії потрібні відомості з археології, антропології, мовознавства, джерелознавства, палеозоології, палеоботаніки, музикознавства, фольклористики. Розв'язання проблем етнічної історії й історичної географії Східної Європи вимагає звертання не тільки до фактів археології, до ареалу й генезису окремих культур, але й до зіставлення прадавніх текстів, аналізу старих карт, до даних мовознавства (топоніміки). А як відомо, саме на стиці різних наук і з'являються найцікавіші розв'язки й гіпотези.

Нарешті, перед дослідником в області гуманітарних наук постає завдання систематизації зібраного матеріалу. При цьому для більшості гуманітарних досліджень, як правило, буває потрібна класифікація даних по різних параметрах. Наприклад, в археології: типи поховань можуть бути охарактеризовані за їхньою структурою, по хронології, по ареалу; аналогічно в етнографії: види будівель – за матеріалом, плануванням, призначенням, за ареалом; у мовознавстві: словоформи – за внутрішньою структурою, за їхнім відношенням до системи мови, за їхньою реалізацією у мовленні, у текстах різних жанрів тощо. Усе це припускає кількаразову перекласифікацію того самого матеріалу по різних параметрах.

Усе сказане визначає й специфічні первинні форми представлень фактів у гуманітарних науках. У кожній з них велике значення має створення капітальних картотек, каталогів, реєстрів і індексів фактів, словників, що містять опис даних і їх детальну характеристику. Створення таких картотек багато в чому являє собою самостійне дослідницьке завдання. Створення таких картотек, словників аж ніяк не механічна справа, тому що кожний об'єкт забезпечується описом усіх його змістовних параметрів, релевантних для фахівців. При цьому самі описи будуються на різних принципах – за алфавітом, тематично, хронологічно, за ареалом. Розробка багатьох гуманітарних проблем, особливо мовознавства, етнографії, антропології, історії, пов'язана із залученням карт як джерел. От чому проблема введення в комп'ютер і виводу даних карт представляється надзвичайно актуальною не тільки для географії, геології або біології, але й для гуманітарних досліджень. Більше того, саме обробка картографічних даних і побудова нових карт являє собою один з найбільш трудомістких етапів такої роботи.

Отже, ядром сучасного прикладного мовознавства є структурна й математична лінгвістика у всьому різноманітті їх методів і досягнень. Основний предмет структурної й математичної лінгвістики – розробка й удосконалювання структурних і формальних методів аналізу й опису мови. У цілому, у всій своїй сукупності цілей, напрямків, методів і підходів, прикладна лінгвістика – це насамперед комплексна дисципліна, що постійно сполучається з філософією, психологією, фізіологією, математикою, логікою, соціологією, інформатикою.

Наведемо кілька наукових напрямків, з якими стикається прикладна лінгвістика і нових напрямків, що виникають на межі дисциплін:

1. Філологія (лінгвістика тексту).
2. Соціологія (соціолінгвістика).
3. Логіка (моделювання суджень, штучний інтелект).
4. Математика (математична лінгвістика).
5. Психологія (психолінгвістика).
6. Біологія (фонетика).
7. Фізика (акустика звуків мови).
8. Історія, археологія, етнографія (етнолінгвістика).
9. Математична статистика (лінгвостатистика).
10. Інформатика (лінгвістичні основи інформатики, штучний інтелект).
11. Семіотика (знакова теорія мови).

Питання для самоперевірки:

- I. Наведіть приклади основних галузей мовознавства, які є актуальними на сучасному етапі розвитку науки. Поясніть чому.*
- II. Які задачі прикладної лінгвістики є найбільш перспективними у сфері машинного перекладу? Поясніть свій вибір.*
- III. Оберіть кілька (2-3) галузей науки, суміжних з прикладною лінгвістикою, і доведіть їх взаємозв'язок, наводячи приклади.*

ЛЕКЦІЯ 2

СЕМІОТИКА. ІНФОРМАТИКА. КОМП'ЮТЕРНА ЛІНГВІСТИКА.

План

- I. Семіотика. Знак і теорія знаків. Мова як мовна система.*
- II. Інформація. Інформатика.*
- III. Комп'ютерна лінгвістика.*

I. Семіотика. Знак і теорія знаків. Мова як мовна система. Семіотика (сема – «значення») – це наука про загальну теорію знака, що досліджує будь-які знакові системи як засоби позначення і передачі значення чи інформації. Семіотика – молода наукова дисципліна. Вона сформувалася лише у ХХ ст., хоча знакові підходи до

вивчення певних явищ і процесів можна зустріти вже у творах античних та середньовічних вчених. Засновниками цієї науки є швейцарський лінгвіст **Фердинанд де Соссюр** (1857-1913) та американські філософи **Чарльз Пірс** (1839-1914) і **Чарльз Морріс** (1901-1978). **Фердинанд де Соссюр** спробував розглянути природні мови як знакові системи у рамках нової наукової дисципліни, яку назвав «семіологією». Чарльз Пірс вніс до наукового співтовариства власне термін «семіотика». Чарльз Морріс у 1938 р. видав працю «Основи теорії знаків», що стала першим систематичним викладом семіотики як науки.

Семіотика на сучасному етапі її розвитку – це міждисциплінарна наука. Вона не має чітко окресленого предмета дослідження. Будь-яке явище або процес, що розглядаються з точки зору їх знакового втілення, можуть стати предметом аналізу цієї наукової дисципліни.

Лінгвістику цікавлять загальні положення семіотики про знаки, розпізнавальні ознаки знаків, способи класифікації знаків, комбінація їх у систему для конкретних задач.

Серед численних визначень знака виберемо найбільш зрозуміле для первісного пояснення. «Знак – це деяке Б, навмисно поставлене кимсь замість деякого А з метою інформувати когось про це А». Наприклад, значок © („копірайт”) ставлять на книзі, щоб повідомити про захист прав того, хто написав і видав цю книгу.

Знак матеріальний і має спрямоване значення. У знаку виділяють 2 основних аспекти (чи «плани»):

1. План вираження (матеріальний, як виглядає).
2. План змісту (що позначає, зміст).

Знак – це член визначеної знакової системи. Існує кілька основних типів знакових систем:

- 1) натуральні чи природні;
- 2) іконічні;
- 3) системи художніх образів;
- 4) системи мовних знаків;
- 5) системи письмових знаків;
- 6) формалізовані чи кодові знакові системи.

Наприклад, усім відомі такі знакові системи як дорожні, телеграфні коди, сигнальні (прапорці), позначення на кресленнях, шифри, комп'ютерні піктограми і т.д.

Знак представляє не тільки об'єкт (предмет, явище, процес, дію, подію), але й окремі властивості об'єктів і відношення між ними.

У семіотиці всі знаки поділяють на *мовні* та *позамовні*.

Позамовні знаки, в свою чергу, поділяються на

- знаки-копії (іконічні знаки),
- знаки-ознаки (знаки-прикмети, знаки-індекси),
- знаки-символи.

Знаки-копії (іконічні знаки) – це знаки, значення яких повністю визначається тим предметом, якому вони відповідають. Знаки-копії об'єднуються за принципом схожості, подібності предмета та його позначення. Прикладами іконічних знаків можуть бути фотографії, картини, відбитки пальців, відображення у дзеркалі, копії документів тощо.

Знаки-ознаки (знаки-прикмети, знаки-індекси) – це знаки, значення яких повністю визначається тим контекстом, у якому вони виявляються і позначають

відношення між об'єктами, а також між об'єктом і його властивостями. Прикладом таких знаків можуть бути займенники, деякі прислівники (тут, зараз, завтра), положення флюгера, сліди на снігу. Дим – знак вогню, розбитий автомобіль – знак транспортної аварії на дорозі, крадіжка автомобіля – знак злочину, безпідставне звільнення людини "х" з роботи – знак правопорушення тощо.

Знаки-символи – це знаки, що фізично не зв'язані з об'єктами, які вони позначають, їхнє значення встановлюється переважно за умовною згодою. У зв'язку з цим вони набувають статусу умовного позначення і всезагального правила. Знаки-символи – це щось конкретне (предмет, образ), те, що має свій особистий зміст і, разом з тим, відображає дещо загальне, абстрактне (поняття, ідею, гіпотезу, концепцію). «Чистий» знак відрізняється від знака-символа тим, що перший просто позначає об'єкт, проте сам не має особистого змісту, не несе якоїсь додаткової смислової інформації, що її можна інтерпретувати, тоді як знак-символ узагальнено, абстрактно відображає предмети та їхні властивості, а також явище через свій особистий зміст, і встановлює певні відношення схожості, подібності між різними предметами та явищами. Наприклад, знаки-символи української культури – Т.Шевченко, Л.Українка, червона калина, вишита сорочка, писанка, танець "Гопак", бандура та ін; знаки-символи Української держави – жовто-блакитний Державний прапор, Державний герб, Державний гімн, опис яких подається у ст. 20 Конституції України.

У мистецтві той чи інший образ реальних живих істот або вигаданих об'єктів часто постає як знак-символ абстрактних властивостей – розуму, таланту, сміливості, мужності, хитрощів, скупості і т.д. Наприклад, Шерлок Холмс – символ славетного слідчого (детектива), носія високого рівня логічного мислення. Знаками-символами можуть стати реальні індивіди, які в силу своїх соціально-психологічних особливостей в певних соціальних умовах є носіями загальних і абстрактних ідей. Наприклад, в історії людства давньогрецький філософ Сократ сприймається як символ мудрості.

Знаки-символи полісемічні, тобто, мають багато значень у різних культурах і їх можна по різному інтерпретувати у тому чи іншому контексті. Знаки-символи використовують в усіх видах практичної і теоретичної, матеріальної і духовної діяльності людей – мистецтві, науці, філософії, релігії і т.д.

У правовій діяльності також широко використовують знаки-символи. Наприклад, знаки-символи судової влади в Україні: в залі засідань вміщені зображення Державного герба України та Державний прапор України. До символів судової влади відносяться також суддівські посвідчення, нагрудні знаки і мантия, зразки яких затверджуються Верховною Радою України. Тепер логічно перейти до того, чому ж, коли дають визначення мови, говорять, що «Мова – це система знаків». Тут саме і простежується чіткий зв'язок лінгвістики і семіотики. Візьмемо будь-як слово, наприклад, «будинок». Це слово і є «деяке А», навмисно поставлене замість натурального будинку («деякого Б»), щоб сказати нам про той чи інший будинок. Причому, слово «будинок» має план вираження (воно або вимовлено, або написано) і план змісту (позначає будинок).

Знаковими одиницями мови вважають морфеми і лексеми, тому що вони несуть інформацію і значення на відміну від одиниць нижчого рівня мови – звуків і букв (так називані знаки 1-го роду).

Основні соціальні функції мовних знаків: позначення предметів, вираження людського духу, думок, почуттів, настрою, бажань, потреб людини; пізнавальна (людина пізнає світ за допомогою мислення, а мислення реалізується через мову); інформаційна, тобто за допомогою мовних знаків передаються відомості, знання від людини до людини, від покоління до покоління; комунікативна, тобто функція спілкування; культурологічна (мовні знаки, знакові системи є засобом засвоєння

національної та загальнолюдської культури окремою людиною або соціальною групою, засобом передачі культурних традицій, особистого та колективного досвіду, навичок, умінь).

У мові виділяють **3 семіотичних аспекти знака**:

1. Синтактика (відносини між знаками).
2. Семантика (відносини між знаком і значимим).
3. Прагматика (відносини між знаками і тими, хто їх використовує).

Синтаксичний вимір – це вимір, який фіксує відношення між знаками в структурі знакового процесу.

Семантичний вимір – це вимір, який фіксує відношення між знаком та його значенням в структурі знакового процесу.

Прагматичний вимір – це вимір, який фіксує відношення між знаком та його інтерпретатором в структурі знакового процесу.

Найчастіше в знаковому процесі присутні всі три виміри. Проте можливі й такі ситуації, у яких деякі виміри зникають. Так, знак може не мати зв'язків з іншими знаками (відсутність синтаксичного виміру); знак може мати зв'язки з іншими знаками, але при цьому не мати значення (відсутність семантичного виміру); і, нарешті, у знака може бути відсутнім інтерпретатор (відсутність прагматичного виміру).

Відповідно до трьох вимірів семіозису виділяють **три основні розділи семіотики як науки**:

1. синтактику;
2. семантику;
3. прагматику.

Синтаксис вивчає систему відношень між знаками всередині певної мови, виявляє різні зв'язки між словами і реченнями, що утворюють певну мовну систему.

Прагматика вивчає відношення знакових систем до тих, хто їх використовує, умови використання знакових систем, мовні знаки як засіб встановлення взаєморозуміння між людьми.

Семантика (гр. – «той, що позначає») – це особлива теорія (складова частина семіотики), яка у наш час аналізує природну і штучну мову у двох напрямках – відношення мовних виразів (слів, імен) до предметів, які вони позначають, і зміст мовних виразів.

Відповідно, в семантиці розрізняють, теорію смислу і теорію реферації (позначення).

Основні семантичні категорії: висловлювання, ім'я, смисл, значення, реферація. Дамо їм загальну характеристику. Висловлювання — це граматично правильно побудоване речення, яке має певний смисл. Складовою частиною висловлювань є імена. ***Ім'я*** – це слово або словосполучення, яке позначає будь-який предмет (об'єкт) або властивість предмета.

Імена бувають

1. одиничні або власні (Сократ, Шевченко, Юстиніан),
2. загальні (людина, юрист, право),
3. конкретні (книга, автомобіль, злочинець),
4. абстрактні (доброта, дієздатність, законність),
5. нульові (Кентавр, Зевс, Перун).

Слово "термін" (гр. – «межа, кінець, границя») має два значення:

а) у широкому значенні "термін" – це слово чи словосполучення, які позначають реальний або абстрактний предмет. У цьому випадку слово "термін" збігається за смислом із "ім'ям", тобто "термін" означає те ж, що й "ім'я". У вузькому значенні

„термін” – це слово або словосполучення, що введені в науку для позначення предметів, явищ, процесів, які вивчає наука, або для побудови наукової теорії відповідно з певними правилами введення наукових термінів. Наукові терміни специфічні за своїм змістом. Сукупність термінів, якими оперують вчені в галузі певної науки, називають науковою термінологією (математична термінологія, юридична термінологія).

б) у логіко-семантичному значенні "терміни" – це слова і словосполучення, які вводять до складу висловлювань і які поділяються на описові (deskриптивні) і логічні.

До deskриптивних термінів належать:

- а) вирази, які називають або позначають емпіричний та абстрактний об'єкт. Вони мають назву "терм". Терм – це одиничні (власні) імена і загальні імена;
- б) предикатний вираз (предикатор) – слово чи словосполучення, яке позначає властивості і якості предметів і відношення між предметами;
- в) предметно-функціональні вирази, або предметні функтори-вирази, які позначають предметні функції і операції, до яких належать спеціальні математичні і логічні знаки, а також слова, як обсяг, вага, ріст, колір. До логічних термінів належать логічні зв'язки, логічні оператори, описові вирази, про які мова піде у відповідних розділах.

Слово "смысл" вживають у таких значеннях:

- 1) У повсякденній мові як синонім слова "значення".
- 2) Як зміст мовного виразу.
- 3) Як думка, що виражена в словах і словосполученнях.
- 4) Як мета (призначення) певної події, вчинку (смысл життя, смысл навчання у вузі, тобто для чого людина живе, для чого вона вчиться у вузі).

Мовні вирази можуть мати прямий та переносний, широкий та вузький смысл, а залежно від контексту – філософський і релігійний, науковий і буденний, юридичний і політичний смысл.

Усі імена, в тому числі нульові, мають певний смысл, проте можуть не мати конкретного предметного значення.

Предметне значення імені (мовного виразу, мовних знаків) – це предмет (об'єкт), який позначається даним іменем. Для позначення окремих предметів або сукупності предметів, тобто класу, до яких застосовується мовний знак (ім'я), введено термін "денотат". Денотат – предмет або клас предметів, який позначається особистим або загальним іменем у певній мовній (семантичній) системі, або, інакше, предметне значення імені (слова), або об'єкт позначення. Наприклад, слово "Київ" є особистим ім'ям, яке позначає столицю України, а саме місто Київ є денотатом, тобто реально існуючим предметом, який позначають особистим іменем "Київ".

Німецький логік і математик Г.Фреге (1848-1925) розробив концепцію смыслу і значення, згідно з якою предмет, який носить дане ім'я, – це предметне значення даного імені, а зміст імені – це його смысл, тобто за Фреге, смысл – це спосіб, яким мовний знак (ім'я) позначає предмет.

Предметне значення мовного виразу може бути істинним або хибним, а смысл мовного виразу, тобто його зміст, може бути зрозумілим або незрозумілим для тієї чи іншої людини. Зрозуміти смысл мовного виразу означає його засвоєння. У процесі засвоєння мовних виразів потрібно спочатку правильно зрозуміти смысл мовних виразів (що саме виражає те або інше ім'я, термін, в певному контексті, яку інформацію воно містить у собі), а потім адекватно встановлювати їх предметне значення, тобто співвідносити з реально існуючими або з абстрактними об'єктами.

У сучасній логіко-лінгвістичній літературі співвідношення між іменем і його смыслом та предметним значенням зображають за допомогою наглядної моделі (схеми), яка дістала назву "семантичного трикутника".

Слово "референція" (лат. – «називати, позначати») означає те ж саме що денотат, тобто стосується процесу встановлення відношення між предметом і його іменем. Для референції (вказівки, назви, позначення) окремого предмета з метою виділення його з певного класу використовують такі способи:

а) називають особисте (власне) ім'я предмета;

б) дають теоретичний опис предмета, як носія певних специфічних властивостей та ознак, без вказівки особистого імені. Такий спосіб називають **дескрипцією** (лат. – «опис»). Автором теорії дескрипції є філософ, логік, математик Б. Рассел (1872-1970). Наведемо приклади дескрипцій, які вказують на предмет: "автор Кобзаря", "засновник науки логіки", "римський юрист", "Кримінальний кодекс України". Усі дескрипції поділяються на визначені, тобто такі, що вказують на конкретний предмет ("перший президент України", "автор першої кодифікації права"), і невизначені, які вказують на будь-який елемент з певного класу ("якась книга", "український юрист", "працівник органів внутрішніх справ"). Для дескрипції предметів використовується висловлювання з операторами (знаками) дескрипції: "існує "x" ...", "такий "x" .., що ..."; "той "x", якому притаманна властивість P" та ін. Символічне позначення оператора (знака) дескрипції – "IxP/x/", де "I" - позначає дескрипцію, "x" - предметна змінна (предмет думки), "P" - властивість (предикат), яка приписується предмету думки.

Повернемося до визначення семіотики. Показавши, що мова – це деяка система знаків, і говорячи, що мова – це засіб комунікації (основне визначення у прикладній лінгвістиці), чи засіб передачі значення (визначення семіотики), ми знаходимо іще один зв'язок лінгвістики і семіотики. Крім того, звідси випливає, що мова є засобом передачі інформації, а тим самим ми встановлюємо зв'язок з такою перспективною наукою, як інформатика. Отже, розглянемо цей взаємозв'язок більш детально.

II. Інформація. Інформатика. Спочатку дамо кілька визначень поняття інформації.

«Інформація – зведення, дані, що несуть значення для конкретної людини». Під інформацією розуміються не самі предмети і процеси, а їхні значення і характеристики (їхнє відображення у вигляді чисел, текстів, формул, таблиць, графіки), тобто говорять про текстову, графічну інформацію. Цікаве і таке визначення, що «інформація є засіб усунення невизначеності у знаннях». Проблемами інформації займається «Теорія інформації», обґрунтована Р. Шенноном у 50-х рр ХХ ст. Основним питанням теорії інформації є процес передачі інформації по каналах зв'язку від відправника до одержувача. Це питання стало найбільш актуальним у наш час з розвитком різноманітних засобів зв'язку, у т.ч. комп'ютерних і телекомунікаційних.

Інформація є фундаментальним науковим поняттям. Воно є первинним і не має чіткого наукового визначення. Особливість цього поняття це його універсальність – воно використовується практично у всіх сферах людської діяльності: природничих науках, філософії, біології, економіці. Разом з тим фундаментальної природничо-наукової дисципліни, яка б займалась природою інформації, не існує. Фізика, хімія, біологія вивчають властивості інформації, а не її природу. Конкретний смисл поняття "інформація" залежить від того, в якому контексті воно використовується.

В останні роки інформатика почала формуватись як природничо-наукова дисципліна, однак вона не ще вийшла за межі прикладної технічної науки, тому також не ввела до цих пір чіткого поняття інформації. У кращому випадку її розглядають як змістовну частину даних, яка інтерпретується людиною. Однак наукове визначення інформації дається досить просто, якщо припустити, що інформація – це динамічний

об'єкт, не існуючий в природі сам по собі, а той, що утворюється в результаті взаємодії даних та методів. Він існує стільки, скільки триває ця взаємодія, а весь інший час він перебуває у вигляді даних. Отже, інформація – це продукт взаємодії даних та методів, який розглядається в контексті цієї взаємодії. Наприклад, книга – це сховище даних. Вона призначена для одержання інформації візуально, шляхом читання. Якщо спробувати різні книжки навпомацки, також можна одержати інформацію, хоча, напевно, не ту, що передбачалася авторами. Даний приклад пояснює сутність контексту взаємодії даних та методів. Контекстний метод опрацювання даних – це той метод, який призначений для обробки даних певного типу. Цей метод відомий як тому, хто створює дані, так і тому, хто їх використовує. Для графічних даних контекстним є метод спостереження. Для текстових даних контекстним є метод, оснований на візуальних образах та знаннях мови тексту. Для даних, що подаються радіохвилями, контекстними є методи перетворення даних за допомогою радіо або телевізійного приймача.

Є свої особливості і для комп'ютерної інформації. Для даних, що подаються в числовій формі і зберігаються у вигляді сигналів, зареєстрованих на магнітних (або інших) носіях, контекстними є апаратні та програмні методи обчислювальної техніки.

З визначення інформації випливає важлива властивість інформації – **динамічність**. Тобто, інформація існує досить нетривалий термін часу – рівно стільки, скільки триває взаємодія даних та методів під час її створення, використання та перетворення. Як тільки ця взаємодія завершується, ми знову маємо справу з даними, але поданими вже в іншій формі. У цьому випадку ми маємо справу з **інформаційним процесом**. Такий процес являє собою цикл утворення інформації з даних і збереження її у вигляді нових даних. Інформація існує досить нетривалий термін часу, але інформаційний процес триває стільки, скільки існують носії інформації. Наприклад, вивчаючи старовинні надписи, вчені продовжують інформаційний процес, розпочатий багато років тому.

Особливість інформаційного процесу в обчислювальній техніці полягає у тому, що певні його етапи відбуваються автоматично, без участі людини. У ході цих етапів дані, подані у вигляді зареєстрованих сигналів, взаємодіють як з апаратними методами (комп'ютерами та іншим обладнанням), так і з програмними методами (комп'ютерними програмами). При цьому важливою особливістю комп'ютерних програм є їх двоїстість. З одного боку, вони ведуть себе як дані, а з іншого – як методи опрацювання даних.

Дані – діалектична складова інформації. Вони являють собою зареєстровані сигнали. При цьому фізичний метод реєстрації може бути довільним: механічним, електричним, магнітним, оптичним, тощо. Відповідно до методу реєстрації дані можуть зберігатись і транспортуватись на носіях різних видів. Найбільш розповсюдженим видом носія, хоча і не найбільш економічним та надійним, є папір. В обчислювальній техніці у ролі носіїв інформації виступають різноманітні магнітні диски, стрічки, оптичні диски, тощо.

В ході виконання інформаційного процесу дані перетворюються з одного виду в інший за допомогою **методів**. Опрацювання даних містить в собі багато операцій, серед яких можна виділити наступні:

1. Збір даних – накопичення даних з метою забезпечення їх повноти для прийняття рішень.
2. Формалізація даних – зведення даних, одержаних з різних джерел, до однакової форми.

3. Фільтрація даних – відкидання “зайвих” даних, які не потрібні для прийняття рішення.
4. Сортування даних – впорядкування даних за певною ознакою.
5. Групування даних – об’єднання даних за певною ознакою з метою їх більш зручного використання.
6. Архівація даних – організація збереження даних в зручній та легкодоступній формі, як правило, у більш економному форматі.
7. Захист даних – комплекс заходів, спрямованих на запобігання втрати, модифікації або відтворення даних.
8. Транспортування даних – прийом та передача даних між віддаленими учасниками інформаційного процесу.
9. Перетворення даних – перевід даних з однієї форми або однієї структури до іншої.

Повний перелік операцій з даними набагато більший. Тому можна зробити висновок: опрацювання інформації має високу трудомісткість і тому її потрібно автоматизувати. Для автоматизації опрацювання даних різних типів їх потрібно подати в єдиній формі. Для цього використовують кодування даних. Звичайна людська мова є прикладом кодування даних. Азбука Морзе, нотна грамота, малюнки є також прикладами систем кодування. Своя система існує і в обчислювальній техніці. Вона заснована на використанні двійкової системи числення. Для подання інформації використовується дві цифри: 0 та 1. Ці цифри називаються двійковими цифрами, або бітами. Таке подання є досить зручним. По перше, одним бітом може бути виражене одно з двох понять: «так» (1) або «ні» (0). Один біт легко подати у вигляді сигналів різної фізичної природи: намагнічено (1) – не намагнічено (0); є струм (1) – немає струму (0); високий рівень напруги (1) – низький рівень напруги (0). Двома бітами можна подати вже чотири стани, трьома – 8. Всього n бітами можливо подати 2^n різних станів. Збільшуючи кількість розрядів на один, ми збільшуємо кількість можливих станів удвічі. Біт є досить дрібною, а отже, і незручною одиницею виміру даних. Тому біти об’єднуються у групи по 8 біт – байти. Два байти називаються машинним словом. Байти об’єднуються в кілобайти ($1\text{Кб} = 1024 \gg 1000$ байт), мегабайти ($1\text{Мб} = 1024 \gg 1000\text{Кб} \gg 1000\ 000$ байт), гігабайти, терабайти і т. д. Враховуючи сучасні можливості виготовлення носіїв інформації і об’єми даних, які потрібно зберігати, найбільш вживаними одиницями виміру даних є мегабайти та гігабайти. Для кодування числової інформації використовуються групи з декількох байтів. При цьому як правило, старший байт відповідає за знак числа: 0 – число додатне, 1 – число від’ємне. Так, для цілих чисел використовується два байти, для дійсних чисел – 10 байт.

Для кодування текстової інформації кожному символу ставиться у відповідність деяке ціле додатне число – його код. Так, одним байтом можна подати 256 різних символів. Цього достатньо для запису всіх символів англійського, російського (українського) алфавітів, цифр та спеціальних символів типу знаків арифметичних операцій, дужок, розділових знаків, тощо. Кодування символів визначається стандартами, яких на даний час існує досить багато. Це призвело до певних суперечностей – текст, поданий в одній кодовій таблиці, неправильно відтворювався в іншій. Для того, щоб уникнути невідповідностей у даний час використовується універсальна таблиця кодування UNICODE, яка основана на 16-бітному поданні символів. Цього достатньо для розміщення в одній кодовій таблиці символів більшості мов нашої планети. Графічне зображення на екрані складається з дрібних точок, які

називаються пікселями. Вони утворюють характерний візерунок, який називається растром. Для кодування графічної інформації використовується принцип декомпозиції. Його сутність полягає у тому, що будь-який колір подається у вигляді комбінації трьох основних кольорів: червоного, зеленого та блакитного. Якщо для кодування інтенсивностей кожної складової використовувати 8 біт (256 різних значень), то на кодування однієї точки екрану потрібно 24 розряди. При цьому така система забезпечує кодування близько 16,5 млн. кольорів. Такий режим подання графічного зображення називається повнокольоровим (True Color). Всі дані зберігаються на зовнішньому носіїві. Для швидкого доступу до даних потрібно, щоб вони мали впорядковану структуру. У ролі такої структури використовується об'єкт змінної довжини, який називається файлом. Отже, файл – це послідовність довільної кількості байтів, записаних на зовнішньому носіїві даних, яка має ім'я. Тип даних визначає тип файлу. Збереження файлів організовується в ієрархічній структурі, яка називається файловою системою. Для зручності роботи файли групуються в каталоги, які, у свою чергу, можуть містити інші каталоги. Вони являють собою іменовані сукупності файлів. На найвищому рівні ієрархії розташований кореневий каталог. Унікальність імені файлу полягає у тому, що повне ім'я файлу містить власне його ім'я разом з маршрутом доступу до файлу. Маршрут завжди починається з логічного імені носія.

Інформація піддається:

1. сприйняттю (розпізнавання образів, текстів і т.д.);
2. передачі (по каналах зв'язку) у вигляді сигналів (аналогових/цифрових);
3. обробці (перетворення в зручну форму);
4. представленню (для демонстрації людині).

У середині ХХ століття в зв'язку зі зростаючими обсягами інформації і з появою нових інформаційних технологій і середовищ склалася і стала активно розвиватися така наука, як інформатика.

Під **інформатикою** розуміють науку про закономірності запису, збереження, переробки, передачі і використання інформації за допомогою технічних засобів. Інформатика – це технічна наука, що систематизує прийоми створення збереження, накопичення, опрацювання та передачі даних засобами обчислювальної техніки, а також принципи функціонування цих засобів і методи управління ними. Інформатика досить близька до технології, тому її предмет часто називають інформаційними технологіями.

Предмет інформатики складають такі поняття:

1. Апаратне забезпечення засобів обчислювальної техніки.
2. Програмне забезпечення засобів обчислювальної техніки.
3. Засоби забезпечення апаратної та програмної складових.
4. Засоби взаємодії людини з апаратними та програмними складовими.

Як видно, в інформатиці велика увага надається взаємодії. Для цього використовується спеціальне поняття – **інтерфейс**. Відповідно до наведених завдань розрізняють апаратний, програмний, програмно-апаратний інтерфейси та інтерфейс користувача.

Основним завданням інформатики є систематизація прийомів та методів роботи з апаратними та програмними засобами обчислювальної техніки. Мета систематизації полягає у виділенні, впровадженні і розвиткові передових, найбільш ефективних технологій, в автоматизації опрацювання даних, а також методичному забезпеченні нових досліджень. Інформатика – практична наука. Її досягнення повинні

підтверджуватись практикою і прийматись в тих випадках, коли вони відповідають критерію підвищення ефективності. Основними завданнями для практичних застосувань є:

- архітектура обчислювальних систем;
- інтерфейси обчислювальних систем;
- програмування;
- перетворення даних;
- захист інформації від несанкціонованого доступу;
- автоматизація опрацювання даних.

При цьому головним є не як виконати ту чи іншу операцію, а як виконати її ефективно.

Основні поняття в інформатиці – **код, система умовних знаків чи символів; алфавіт - набір знаків коду, текст - послідовність знаків даного повідомлення, імовірність – частота появи якого-небудь коду в повідомленні.**

В інформатиці існує такий термін, як «**кодування інформації**» – це представлення повідомлення у формі, зручній для передачі по даному каналу. Більшість кодів для кодування інформації засновані на тих чи інших системах числення. Будь-якому дискретному повідомленню, букві повідомлення можна приписати який-небудь порядковий номер. Передача і збереження інформації при цьому – це робота з числами. Числа виражаються в якій-небудь системі числення. При порівнянні систем числення ефективною виявилася двійкова (1; 0). Логічні елементи повинні мати 2 стійкі стани (включений/виключений). Таким чином, з технічної точки зору, інформація – це сигнал (1,0) чи елементарна подія «так» – «ні», виміряна у двійкових одиницях – бітах (міра Хартлі).

III. Комп'ютерна лінгвістика. Сучасна людина живе у величезному інформаційному середовищі – mass media (телебачення, радіо, газети і т. д.), WWW (“World Wide Web” – всесвітня комп'ютерна мережа). Обсяги інформації зростають на порядок щорічно. Сучасні інформаційні технології включають зростаюче число автоматизованих інформаційних систем, засобів масової комунікації, систем інформаційного пошуку, систем машинного перекладу. Спілкування з комп'ютером розвивається у бік використання зручної для людини мови. Комп'ютер – це складна технічна система, у якій усе повинно бути жорстко формалізовано і запрограмовано, але природна мова (ПМ) – занадто складна структура, тому робота з інформацією на ПМ у сучасних комп'ютерних середовищах обмежується браком знань про мову. Проблеми мовної комунікації «людина – комп'ютер – людина» і моделювання мови лежать в області досліджень такої науки як **комп'ютерна лінгвістика**, яка утворилася на межі інформатики та лінгвістики.

Математична лінгвістика є галуззю науки про штучний інтелект. Усе почалося в Сполучених Штатах Америки, в 1950-х роках. З винаходом транзистора й появою нового покоління комп'ютерів, а також перших мов програмування, почалися експерименти з машинним перекладом. В 1960-х роках подібні дослідження проводилися й у СРСР (наприклад, стаття про переклад з російської на вірменську в зб. «Проблеми кібернетики» за 1964 рік). Однак якість машинного перекладу і дотепер поступається якості перекладу, зробленого людиною. Були створені перші системи штучного інтелекту, такі як SHDSL. Вони, щоправда, застаріли, але користуються популярністю в студентів і наукових співробітників Академій Наук, що займаються

комп'ютерною лінгвістикою. Як особливий науковий напрямок комп'ютерна лінгвістика оформилася у 1960-ті роки. Термін «**комп'ютерна лінгвістика**» є калькою від англійського слова **computational linguistics**. Оскільки прикметник computational може перекладатися і як «обчислювальний», у літературі зустрічається також термін «**обчислювальна лінгвістика**», однак у вітчизняній науці він здобуває більш вузьке значення, що наближається до поняття «квантитативної лінгвістики». **Комп'ютерна лінгвістика**, як напрямок прикладної лінгвістики, вивчає лінгвістичні основи інформатики і всі аспекти зв'язку мови і мислення, моделювання мови і мислення у комп'ютерному середовищі за допомогою комп'ютерних програм.

Інтереси комп'ютерної лінгвістики поширюються на:

1. Створення спільних, загальних мов для представлення інформації.
2. Удосконалення мов доступу до інформації.
3. Створення природного мовного інтерфейсу для спілкування людини з машиною (одна з основних проблем штучного інтелекту (artificial intelligence)).
4. Створення і моделювання інформаційних систем.

Питання для самоперевірки:

- I. Наведіть приклади різних типів знаків у сучасному житті, поясніть, чому вони утворюють систему.*
- II. Перерахуйте основні відмінності мовних знаків від немовних.*
- III. Дайте визначення інформації та інформатики. Поясніть відмінності та спільні риси поданих понять.*
- IV. У чому полягає складність визначення предмету інформатики?*

ЛЕКЦІЯ 3

ПРИРОДНІ І ШТУЧНІ МОВИ. МОВИ ПРОГРАМУВАННЯ. МЕТАМОВА.

План

- I. Природні і штучні мови.*
- II. Мови програмування як штучні мови.*
- III. Формальні методи опису штучних мов. Формальна граматики. Математична лінгвістика.*
- IV. Поняття метамови. БНФ-нотації. Синтаксичні діаграми.*

I. Природні і штучні мови. У попередній лекції ми говорили про лінгвістичні основи інформатики і про те, що у зв'язку з неможливістю повної формалізації природних мов (ПМ) для комунікації у комп'ютерних середовищах необхідна розробка спеціальних штучних мов (ШМ), що входять до кола інтересів лінгвістики.

Інформація стає зрозумілою, якщо вона виражена мовою, якою говорять ті, кому призначена інформація. У процесі розвитку людського суспільства люди створили велику кількість мов. Наведемо приклади різних мов:

- розмовні мови (у наш час у світі їх нараховують більш 2000);

- мови міміки й жестів;
- мови креслень, малюнків, схем;
- мови науки (математики, хімії, біології і т.д.);
- мови мистецтва (живопису, музики, скульптури, архітектури і т.д.);
- спеціальні мови (абетка Брайля для сліпих, абетка Морзе, Есперанто, морський семафор і т.д.);
- алгоритмічні мови (блок-схеми, мови програмування).

Мова – це знакова система, використовувана для цілей комунікації та пізнання. Основою більшості мов є алфавіт – набір символів, з яких можна створювати слова й фрази даної мови.

Мова характеризується:

- набором використовуваних знаків;
- правилами утворення із цих знаків таких мовних конструкцій, як “слова”, “фрази” і “тексти” (у широкому тлумаченні цих понять);
- набором синтаксичних, семантичних і прагматичних правил використання цих мовних конструкцій.

Усі мови можна розділити на **природні й штучні**.

Природними називаються “звичайні”, “розмовні” мови, які складаються стихійно й протягом довгого часу. Історія кожної такої мови невіддільна від історії народу, що володіє нею. Природна мова, призначена, насамперед, для повсякденного спілкування, має цілий ряд **характерних рис**:

- майже всі слова мають не одне, а кілька значень;
- часто зустрічаються слова з неточним і неясним змістом;
- значення окремих слів і виразів залежать не тільки від них самих, але й від їхнього оточення (контексту);
- поширеними є синоніми (різне звучання - однаковий зміст) і омоніми (однакове звучання - різний зміст);
- ті самі предмети можуть мати кілька назв;
- є слова, що не позначають жодних предметів;
- багато угод щодо вживання слів не формулюються явно, а тільки передбачаються і для кожного правила є виключення.

Основними функціями природної мови є:

- комунікативна (функція спілкування);
- когнітивна (пізнавальна функція);
- емоційна (функція формування особистості);
- директивна (функція впливу).

Штучні мови створюються людьми для спеціальних цілей або для певних груп людей: мова математики, морський семафор, мова програмування. Характерною рисою штучних мов є однозначність одиниць їх словника, правил утворення виразів і правил додання їм значень.

Штучні мови поділяються на:

- 1). ***Неспеціалізовані мови загального призначення*** (наприклад, есперанто, волапюк);
- 2). ***Спеціалізовані мови різного призначення*** (наприклад, символічні мови науки (математика у XVI ст., логіка у XVIII ст.).

До другої групи входять також мови людино-машинного (комп'ютерного) спілкування (мови програмування, мови операційних систем, мови інформаційних систем тощо.).

Будь-яка мова – і природна, і штучна – має набір певних правил. Вони можуть бути явно й строго сформульованими (формалізованими), а можуть допускати різні варіанти їх використання.

Формалізована (формальна) мова – мова, що характеризується точними правилами побудови виразів і їх розуміння. Вона будується відповідно до чітких правил, забезпечуючи несуперечливе, точне й компактне відображення властивостей і відносин досліджуваної предметної області (об'єктів, що моделюються). На відміну від природних мов, формальним мовам властиві чітко сформульовані правила семантичної інтерпретації та синтаксичного перетворення використовуваних знаків, а також те, що зміст і значення знаків не змінюється залежно від певних прагматичних обставин (наприклад, від контексту).

Більшість формальних мов (створених конструкцій) будується за наступною схемою: спочатку обирається алфавіт або сукупність вихідних символів, з яких будуватимуться всі вирази мови; потім описується синтаксис мови, тобто правила побудови осмислених виразів. Буквами в алфавіті формальної мови можуть бути і букви алфавітів природних мов, і дужки, і спеціальні знаки і т. д. З букв за певними правилами можна будувати слова та вирази. Осмислені вирази у формальній мові виходять тільки, якщо дотримані прийняті у мові правила їх утворення. Для кожної формальної мови сукупність цих правил повинна бути чітко визначена й модифікація кожного з них найчастіше призводить до появи нового різновиду (діалекту) цієї мови.

Формальні мови широко застосовуються у науці та техніці. У процесі наукового дослідження й практичної діяльності формальні мови звичайно використовуються у тісному взаємозв'язку із природною мовою, оскільки остання має набагато більші виражальні можливості. У той же час формальна мова є засобом більш точного представлення знань, ніж природна мова, а отже, засобом більш точного й об'єктивного обміну інформацією між людьми.

Формальні мови часто конструюються на базі мови математики. Епохою бурхливого розвитку різних формальних мов можна вважати ХХ століття. З погляду інформатики, серед формальних мов найбільш значну роль відіграють формальна мова логіки і мови програмування.

II. Мови програмування як штучні мови. Мови програмування (МП) – це клас штучних мов, призначених для обробки інформації за допомогою комп'ютера. Мов програмування та їх діалектів (різновидів) налічується кілька тисяч. Класифікувати їх можна по-різному. Деякі автори розбивають усе різноманіття мов програмування на процедурні й декларативні. У процедурних мовах перетворення даних задається за допомогою опису послідовності дій над ними. У декларативних мовах перетворення даних задається за допомогою опису відносин між самими даними. Згідно з іншою класифікацією, мови програмування можна розділити на процедурні, функціональні, логічні, об'єктно-орієнтовані. Однак будь-яка класифікація дещо умовна, оскільки, як правило, більшість мов програмування містять у собі можливості мов різних типів. Особливе місце серед мов програмування займають мови, що забезпечують роботу систем керування базами даних (СКБД). Часто в них виділяють дві підсистеми: мова опису даних і мова маніпулювання.

Історично мови програмування виникли у 40-х роках ХХ ст. і якісно удосконалювалися у бік спрощення свого опису, методів програмування і додатків для обробки інформації людиною. Можна виділити наступні якісні рівні розвитку МП (тобто те, як їх класифікують у програмуванні):

1. Низький рівень – Ассемблери (40-50-і роки ХХ ст.) (робота з машинними кодами).
2. Середній рівень – Фортран, Алгол (60-і роки ХХ ст.).
3. Високий рівень – Паскаль, С++, Prolog (з 80-х років ХХ ст.).

Якщо поглянути на мову програмування з лінгвістичної точки зору, можна побачити, що МП так само, як і будь-яка мова, має рівневу структуру. Перший нижчий рівень – символний; його елементи – букви, спецсимволи (за аналогією з ПМ – графічний рівень). Другий рівень – це рівень імен, зарезервованих слів, висловлювань (в ПМ – це лексичний рівень). Третій рівень – операторний (командний), синтаксичні конструкції мають наказовий характер (в ПМ – аналог синтаксичного рівня), і останній – рівень програми завжди є синтаксично і семантично закінченою послідовністю розпоряджень-команд. Програма – це текст, записаний за заданими правилами штучної мови, тобто МП.

III. Формальні методи опису штучних мов. Формальна граматики. Математична лінгвістика. Загальною ознакою опису спеціалізованих штучних мов є формальний метод їхнього опису і визначення шляхом завдання алфавіту, словника і системи правил утворення і перетворення висловлювань – це граматики. Формальний метод служить для породження „правильних текстів” („правильних” – значить „записаних за визначеними правилами”). Наприклад, для мов програмування задаються визначені форми мовних елементів (алфавіт, слова), правила побудови текстів, точно описуються семантика і синтаксис для однозначного розуміння програм комп’ютером. Узагалі, при написанні „правильних текстів” (програм на МП чи будь-яких інших текстів на ШМ) широко використовуються формальні граматики.

Математична лінгвістика – математична дисципліна, що розробляє формальний апарат для опису будови природних і деяких штучних мов. Виникла в 50-х роках 20 століття у зв’язку з назрілою в мовознавстві потребою уточнення його основних понять. У математичній лінгвістиці використовуються переважно ідеї та методи алгебри, теорії алгоритмів та теорії автоматів. Математична лінгвістика, яка не є лінгвістичною дисципліною, незважаючи на її назву, розбудовується у тісній взаємодії з нею. Математичною лінгвістикою називають іноді лінгвістичні дослідження, у яких застосовується будь-який математичний апарат.

Математичний опис мови починається з **Ф. де Соссюра**, що представляв мову як механізм, функціонування якого проявляється у мовній діяльності його носіїв; її результатом є «правильні тексти» – послідовності мовних одиниць, що підкоряються певним закономірностям, багато з яких допускають математичний опис. Вивчення способів математичного опису правильних текстів (у першу чергу речень) становить зміст одного з розділів математичної лінгвістики – теорії способів опису синтаксичної структури. Для опису будови (синтаксичної структури) речення можна або виділити в ньому «складові» – групи слів, що функціонують як цілісні синтаксичні одиниці, або вказати для кожного слова ті слова, які від нього безпосередньо залежать (якщо такі є). Так, у реченні «Коні їдять овес» при описі за 1-м способом складовими будуть: все речення, кожне окреме слово й словосполучення «їдять овес».

Інший розділ математичної лінгвістики – теорія формальних граматики, що виникла головним чином завдяки роботам **Н. Хомського**. Вона вивчає способи опису закономірностей, які характеризують уже не окремий текст, а всю сукупність правильних текстів тієї або іншої мови. Ці закономірності описуються шляхом побудови «формальної граматики» – абстрактного «механізму», що дозволяє за допомогою однакової процедури одержувати правильні тексти даної мови разом з описами їх структури. **Формальна граматика** – це абстрактний апарат, що дозволяє за допомогою однакової процедури одержувати правильні тексти даної мови.

Питаннями формальних граматики і їхньої теорії займається математична лінгвістика (виникла в 50-і роки ХХ ст.). Вона є одним з напрямків прикладної лінгвістики, що межує з математикою. У 1959 році американський учений-лінгвіст Н. Хомський запропонував першу класифікацію формальних мов за типами правил формальної граматики. Існує безліч формальних граматики, як наприклад:

1. Граматика, що породжує, Хомського.
2. Контекстно-вільна граматика.
3. Граматика безпосередніх складових.
4. Трансформаційна граматика (граматика дерев).
5. Граматика Монтегю.

Коротко опишемо граматику, що породжує. Формальна граматика, що породжує – це система $\Gamma = \langle VT, VNT, S, R \rangle$, де Γ – граматика;

VT – безліч термінальних (кінцевих) символів мови;

VNT – безліч нетермінальних символів (з яких можна виводити наступні дані), замикаються у кутові дужки $\langle \dots \rangle$;

S – початковий символ нетермінальної множинності;

R – система правил виводу типу $X \rightarrow Y$ (де X, Y – ланцюжки символів з VT, VNT).

Безліч ланцюжків, виведених у Γ з її початкового символу S , і є мова, породжувана цією граматиною Γ (тобто висновок ланцюжків завжди починається з нетермінала S).

Математична лінгвістика вивчає також аналітичні моделі мови, у яких на основі даних про мовлення, що вважаються відомими (наприклад, безлічі правильних речень), проводяться формальні побудови, що дають деякі відомості про структуру мови. Застосування методів математичної лінгвістики до конкретних мов відноситься до області лінгвістики.

IV. Поняття метамови. БНФ-нотації. Синтаксичні діаграми. Для того щоб описати факти мови та установити істотні зв'язки між ними, необхідно не тільки розробити ефективні методи дослідження, але й знайти раціональні способи вичленування елементів, що підлягають виченню, ефективно представляти отримані результати, класифікувати й узагальнити їх. Оскільки будь-яке знання повинне бути виражене словами або іншими знаками, усі ці завдання здійсненні тільки за умови, якщо процеси пізнання й інтелектуальної комунікації забезпечені адекватно розвиненою метамовою.

Метамовою називають мову, призначену й використовувану для опису іншої мови, тобто мову, **об'єктом якої є зміст і форма іншої мови**. Цією іншою мовою може бути окрема національна мова (у нашому випадку англійська) або навіть будь-яка інша семіотична система. Грецький префікс «мета» – означає «інший, після» й уживається в науковому стилі досить часто (порівн.: метатеорія – теорія, що вивчає властивості іншої

теорії, метаматематика – теорія математичних доказів). У метамову лінгвістики входять термінологія, загальнонаукова лексика, властива науковій літературі взагалі в будь-якій галузі знання, буквена символіка й графіка. Графічна метамова включає таблиці, схеми, графіки, графи, матриці й т. д. Графічні засоби мають допоміжний характер і повинні сполучатися з вербальними.

Науковий рівень будь-якого дослідження суттєво залежить від застосування метамови, її строгості та точності. Метамова впливає не тільки на ясність концептуального апарата й зрозумілість його для читача, але й на самого автора: ускладнюючи мову, автор сам собі заважає мислити. Усяка наука є процес перебоязки інформації й обміну нею, тому, якщо колеги не розуміють термінології й взагалі метамови, якою користується автор, обмін інформацією стає або взагалі неможливим, або відчутно ускладнюється, а у результаті й внесок автора в науку знецінюється.

Грамматика Хомського є прикладом метамови. На практиці застосовується ще одна метамова, що не настільки математична. Це **Бекус-Наурови форми (БНФ-форми чи БНФ-нотації)**, що, як і формальна граматика, служать для задання правил одержання правильних текстів.

У програмуванні найпоширенішою метамовою є БНФ, тому що синтаксис більшості мов програмування описується за її допомогою. Основне призначення БНФ полягає у **представленні в стислому й компактному вигляді формальних правил написання основних конструкцій мови.** Будь-яке представлення мови в такий спосіб розглядають як ланцюжок основних символів і за допомогою БНФ можна побудувати будь-яку правильну конструкцію. Один із класів об'єктів, який використовує БНФ – це основні символи мови, іншим класом об'єктів є імена конструкцій. Ці імена називають ідентифікаторами, які є металінгвістичними змінними. Кожна формальна БНФ має дві основні частини, з'єднані через стандартну зв'язку. У правій частині вказується один з декількох варіантів побудов конструкції у лівій частині. Причому металінгвістичні змінні є класом нетермінальних символів і поміщуються в <>.

Аналогічна метамова, що має графічне, наочне представлення – це синтаксичні діаграми. **Синтаксична діаграма – це схема, що пояснює правило побудови певного поняття, виразу або текстів.** Синтаксична діаграма – це графічне правило визначення конструкції мови за допомогою спеціальних позначень (елементів). У будь-якій мові є споконвічні базові поняття, які перебувають не в роз'ясненні, а в перерахуванні. В українській мові, наприклад, це букви кирилиці – вони просто є і не несуть особливого змістового навантаження. У Паскалі таку роль відіграють символи, що становлять алфавіт мови, і службові слова. Ознакою такого об'єкта, що не потребує роз'яснень в синтаксичній діаграмі є ті, що замкнені в овал (або коло). Усі інші об'єкти на синтаксичній діаграмі замкнені у прямокутники. Кожний такий об'єкт потребує точного однозначного визначення того, що він означає; відповідне визначення дається окремо для кожного об'єкта синтаксичної діаграми. Кожна діаграма має вхідну й вихідну стрілки, що означають початок і кінець читання синтаксичної конструкції, що й відбивають процес її аналізу. Із кожного елемента виходить одна або кілька стрілок, що вказують на ті елементи, які можуть впливати безпосередньо з даного елемента. Усяке роздвоєння передається словом «або» і означає можливість рухатися по будь-якій гілці.

Обидві ці форми знайшли широке застосування при описі мов програмування в інформатиці. Для прикладної лінгвістики побудова БНФ і синтаксичних діаграм цікава

як спосіб зрозуміти структуру будь-якої мови, побачити можливості моделювання штучних мов.

Питання для самоперевірки:

- I. Перерахуйте основні відмінності між штучними та природними мовами.*
- II. Наведіть відповідності між рівнями природної мови та мовами програмування.*
- III. Дайте визначення математичної лінгвістики, формальної мови, формальної граматики.*
- IV. У чому полягає сутність БНФ-нотацій? Наведіть приклади можливого застосування БНФ-нотацій у прикладній лінгвістиці.*

ЛЕКЦІЯ 4

ПРОБЛЕМА МОДЕЛЮВАННЯ ЗНАНЬ У ПРИКЛАДНІЙ ЛІНГВІСТИЦІ

План

- I. Моделювання у комп'ютерній лінгвістиці.*
- II. Моделі знань у штучному інтелекті.*
- III. Мови представлення знань як варіант штучних мов.*

I. Моделювання у комп'ютерній лінгвістиці. У різних науках доводиться мати справу з різними моделями (зразками) тих чи інших об'єктів дослідження цих наук. Як ми вже підкресливали, у лінгвістиці об'єктом дослідження є мова, отже, у даній лекції ми будемо розглядати саме моделювання мовних здібностей людини на основі моделювання знань, що є однією з центральних задач штучного інтелекту та комп'ютерної лінгвістики. Існує більше 50 визначень терміну «модель». Поняття «лінгвістичної моделі» виникло у математичній лінгвістиці, тому що формалізований запис, що дозволяє одержувати правильні тексти, і є моделлю певної мови. Будь-яка формальна модель припускає наявність однозначно заданих об'єктів, що зв'язуються взаємовідносинами і правилами поводження з ними.

Якщо узагальнити бачення моделі в лінгвістиці, то це:

- а) опис (завдання форми) тексту (формальним методом);
- б) спосіб запису певних даних чи тексту (символьний, графічний);
- в) метамова, у термінах якої фіксуються вихідні дані для побудови правильних текстів.

Таким чином, ***модель у лінгвістиці*** – це «синонім» формалізованої структури з фіксованою метамовою (див. лекцію 3), що служить зразком для дослідження породжуваної мови, аналізу її характеристик і функцій. Метод моделювання мови та мовних процесів широко використовується лінгвістами, оскільки дає можливість реалізувати теоретичні знання на практиці. Прикладна лінгвістика прагне будувати моделі, що відображають конкретні процеси мовленнєво-мисленнєвої діяльності людини у тій чи іншій ситуації. У лінгвістиці особливо важливими є моделі мовної діяльності, тісно пов'язані з моделями мислення (моделі мислення досліджуються

логікою), тому що логічні і граматико-семантичні форми – це форми одного й того ж самого об'єкта.

II. Моделі знань у штучному інтелекті. Моделювання мовного і людського мислення – це задачі **штучного інтелекту** як одного з провідних наукових напрямків інформатики, що займається **створенням машин, здатних розуміти, моделювати й аналізувати тексти, зберігати та переробляти інформацію, приймати рішення.**

Метою штучного інтелекту є побудова комп'ютерних систем, які можуть продемонструвати рівень інтелекту, подібний до людського розуму. Основною проблемою досліджень в області штучного інтелекту є побудова машиною моделі, яка б здійснювала складні перетворення інформації, що виконуються людським мозком. Ідея практичного застосування досліджень в області штучного інтелекту полягає у використанні накопиченого людського досвіду, готових знань, методів і навичок розв'язання завдань. Перші класичні системи штучного інтелекту містили в собі:

- розв'язок завдання в цілому;
- проходження тестів на інтелект;
- доказ теорем і пошук цілей;
- дослідження в області робототехніки;
- зорове розпізнавання малюнків;
- усвідомлення абстрактних понять;
- навченість;
- ігри;
- розуміння природньої мови.

Дослідження штучного інтелекту проходять по двом основним напрямкам:

1. Програмно-прагматичний напрямок – створюються програми розв'язання завдань (ігрові, логічні, пошуку, класифікації), розв'язок яких уважався прерогативою людини. «Мисляче обладнання», незалежно від його побудови, на задані вхідні впливи реагує як людський мозок.

2. Біонічний напрямок – штучне відтворення тих структур і процесів, які характерні для живого людського мозку і які лежать в основі процесу розв'язання завдань людиною. «Мисляче обладнання» повинне відтворювати структуру людського мозку. У цьому напрямку сформувалася нова наука – нейроінформатика.

У наш час класичним напрямком є програмно-прагматичний, за якого використовуються наступні моделі алгоритмів:

- модель лабіринтового пошуку;
- евристичне програмування;
- використання методів математичної логіки.

Евристичне програмування являє собою методи розв'язання завдань прийняття, що опираються на досвід подібний розв'язків. Методи евристичного програмування застосовуються при розв'язанні завдань великого обсягу в умовах обмеженого запасу часу. Евристичне програмування реалізується:

- використанням інтуїтивного досвіду;
- завданням експертного варіанта;
- заміною одного завдання іншим;
- звуженням області дослідження.

При складанні евристичної програми використовується досвід фахівців у конкретній області, який реалізується у вигляді правил, емпіричних залежностей, обчислювальних алгоритмів. Керування процесом полегшується, якщо фахівець

пропонує експертний (опорний) варіант розв'язання цього завдання. Можна перевірити критерій зміни ефективності при варіюванні окремих параметрів. У випадку заміни одного завдання на інше, модель не буде чітко реагувати на сутність розглянутої ситуації, але для знаходження рішення можна використовувати алгоритм обраного завдання. Пошук оптимального варіанта може спроститися, якщо ввести додаткові обмежуючі умови.

Так, наприклад, перед прийняттям багатьох управлінських розв'язків необхідно розрахувати вартість продажів, витрат, оцінити ряд аналітичних коефіцієнтів (частка валового прибутку, частка чистого прибутку, частка експлуатаційних витрат та ін.) й оцінити ситуацію за оперативним даними. Для розв'язання таких завдань менеджери використовують системи штучного інтелекту – програмні системи, що імітують мислення людини. **Системи штучного інтелекту визначають також як складну програму, що маніпулює знаннями з метою одержання задовільного результату й ефективного рішення у вузькій предметній області.** Системи у цьому випадку виконують роль експертів-консультантів, тому що побудовані на знаннях компетентних експертів і мають їх компетентність (штучно відтворюють компетентність експертів).

Головними прикладними технологіями штучного інтелекту є: експертні системи, інтелектуальні системи підтримки рішень, обробка природньої мови, розуміння мовлення, нечітка логіка, робототехніка й сенсорні системи, комп'ютерний зір і розпізнавання образів. Області застосування систем штучного інтелекту охоплюють медичну діагностику, інтерпретацію геологічних даних, наукові дослідження в хімії й біології, військову справу, виробництво, фінанси й інші сфери економіки.

Знання – це інформація, яка необхідна програмі, для того, щоб вона стала «інтелектуальною». У системах штучного інтелекту знання структуровані й організовані таким чином, що вони відділені від знань інших предметних областей і від загальних знань. Виділені знання про предметну область називають **базою знань**, а загальні знання, що використовуються для знаходження рішень, називають **механізмом висновку**. Сукупність фактів, за допомогою яких оцінюються знання, називається **базою фактів**. Наприклад, бухгалтерських баланс, додатки до нього й фінансовий звіт утворюють базу фактів, а правила осмислення фінансового стану підприємства утворюють базу знань.

Із безлічі показників у базі фактів аналітики використовують лише певні показники, тобто у процесі використання фактів застосовується **метод спрощення**. У системах штучного інтелекту це найважливіше початкове положення називається **процедурою спрощення фактів**.

В узагальненій концептуальній структурі систем штучного інтелекту виділяють **три головні елементи: база фактів, база знань і механізм висновку**. У деяких випадках поняття бази знань включає не тільки структуровані знання, але й факти, для оцінки яких використовуються знання. Наприклад, електронна таблиця з обчисленими значеннями економічних показників – це база фактів, з яких можна зробити висновок про стан об'єкта керування лише тому, хто знає правила економічного аналізу. У базі фактів зберігаються дані, що мають відношення до аналізованої проблеми, що й дають можливість застосувати до них наявний набір правил із бази знань. У базу фактів менеджер переносить дані шляхом запиту із бази даних, у якій зберігаються первинні дані. Крім того, менеджер може сформулювати частину даних самостійно.

Традиційні комп'ютерні програми ґрунтуються на алгоритмі, який чітко визначає послідовну процедуру для рішення проблеми. Інтелектуальне програмне забезпечення ґрунтується на символічній презентації та маніпуляції. При використанні символів створюється база знань. Для розв'язання завдання при маніпулюванні символами під час

генерації порад або рекомендацій використовуються різні процеси. Базовою технологією є пошук і зіставлення зразків. Одержавши первісну інформацію, програма переглядає базу знань, здійснюючи пошук специфічних умов або зразків, які відповідали б установленому критерію розв'язання завдання. Для здійснення процесу пошуку використовуються алгоритми.

Функціональна структура систем штучного інтелекту складається із трьох комплексів обчислювальних засобів:

- перший комплекс – сукупність виконавчих засобів (обчислювальних, логічних, пошукових) програми-виконавця;
- другий комплекс – сукупність засобів інтелектуального інтерфейсу, що мають гнучку структуру, яка забезпечує можливість адаптації у широкому спектрі інтересів кінцевих користувачів;
- третій комплекс засобів – це база знань, яка організує взаємодію перших двох комплексів обчислювальними засобами та містить систему знань про проблемне середовище.

Виконавча система поєднує всю сукупність засобів, що забезпечують виконання програми. Взаємодія інтелектуальної системи із прикладним програмним забезпеченням здійснюється при виконанні спеціальних обчислень, тому що часто виникає необхідність використовувати стандартні операції з обробки даних в якості підзадач. Зв'язок з розподіленою базою даних інтегрованої системи керування використовується для одержання даних і знань, розосереджених на різних рівнях ієрархії системи керування.

Інтелектуальний інтерфейс забезпечує для кінцевого користувача використання комп'ютера для розв'язання завдань: зв'язок на обмеженій природній мові, мовне введення й візуальне представлення (графіку, технічний зір). Для деяких виробничих процесів використовуються засоби для автоматичного одержання даних та їх обробки, а також зворотні зв'язки з керування.

У середині 70-х років сформувався новий підхід до розв'язання інтелектуальних завдань представлення та використання знань, коли на зміну пошуку універсального алгоритму прийшла ідея моделювання знань фахівців-експертів. Перші експертні системи з'явилися в Сполучених Штатах Америки. Експертні системи, які базуються на системному акумулюванні, узагальненні, аналізі та оцінці знань висококваліфікованих фахівців – експертів, є однією з форм реалізації автоматизованих інформаційних технологій керування економічною діяльністю організації.

Експертна система – це прикладна діалогова система штучного інтелекту, здатна одержувати, накопичувати, коректувати знання з певної предметної області, виводити нові знання, знаходити на основі цих знань рішення практичних завдань, що близькі за якістю до рішень експертів, і за запитом користувача пояснювати хід розв'язання у зрозумілій для нього формі. Експертна система являє собою сукупність методів та засобів організації, накопичення та застосування знань для розв'язання складних завдань у певній предметній області.

Експертна система досягає високої ефективності за рахунок перебору великої кількості альтернативних розв'язань під час вибору рішення, спираючись на високоякісний досвід групи фахівців, аналізує вплив великого обсягу нових факторів, оцінюючи їх при побудові стратегій, додаючи можливості прогнозу. Такі системи розробляються з перспективою навчання та здатні обґрунтувати логіку вибору рішення, тобто мають властивості адаптивності та аргументування. Експертні системи проектують для об'єднання та копіювання дій окремих фахівців у конкретних областях,

таких як медична діагностика, пошук несправностей у складних системах, інтерпретація зображень, дозвіл на надання кредитів і т. д. Вони широко використовуються у функціональних сферах керування:

- бухгалтерський облік і керування фінансами;
- стратегія (планування проекту, аналіз його результатів та ін.);
- виробництво (процеси моніторингу та контролювання якості продукції, аналіз несправностей, планування розміщення устаткування, розробка продуктів, включаючи комп'ютерні системи й ін.);
- маркетинг (відповідь на запити покупців, експлуатаційне обслуговування обладнань і устаткування, діагностика устаткування, вибір моделей довгострокового прогнозування збуту, визначення прийнятних знижок для покупця та ін.);
- керування людськими ресурсами (навчання в окремих областях, визначення кваліфікації та ін.);
- підтримка прийняття, вибору, проведення порівняльного аналізу варіантів управлінських розв'язків.

Експертні системи дозволяють вирішувати складні неординарні завдання у таких областях як: промисловість, будівництво, сільське господарство, військова справа, інженерна справа, космічна техніка, керування процесами, економіка, медицина, електроніка, метеорологія, юриспруденція, геологія, інформатика, математика, фізика, хімія. Основою експертної системи є база знань, що накопичується в процесі побудови системи й структурована з метою формалізації процесу прийняття рішень. База знань складається із сукупності моделей, правил і фактів, що породжують аналіз і висновки для знаходження розв'язків складних завдань у певній предметній області. База знань обумовлює компетентність експертної системи, втілює у собі знання фахівців установи, відділу, досвід групи фахівців і являє собою звід кваліфікованих стратегій, що оновлюються, методів, розв'язків і т. д.

У базу знань входить підсистема набуття знань, яка дає можливість користувачам застосовувати систему для одержання знань індивідуальних експертів у зручній формі. Зміст бази знань використовується для одержання ефективних управлінських рішень.

Структура бази знань включає: а) **експерта** – фахівця, що вміє знаходити ефективні розв'язки в конкретній предметній області; б) **блок набуття знань**, який відбиває нагромадження й модифікацію знань та даних; в) **блок логічних висновків**, що здійснює зіставлення правил із фактами, що й породжує ланцюжок висновків; г) **блок пояснень**, який використовує знання для пояснення того, яким чином система прийшла до даного рішення, і відбиває послідовність кроків користувача у технології використання бази знань.

Області застосування баз знань і систем на їхній основі постійно розширюються. Створюється цілий спектр баз знань – від невеликих за обсягом для портативних систем до потужних, призначених для професіоналів, що експлуатують складні, технічно оснащені АРМ. Величезні масиви баз даних зберігаються у централізованих єдиних сховищах, доступ до яких здійснюється через мережі користувачами різних систем.

Створення та використання експертних систем є одним з концептуальних етапів розвитку інформаційних технологій. В основі інтелектуального вирішення проблем у певній предметній області лежить принцип відтворення знань досвідчених фахівців – експертів. Виходячи із власного досвіду, експерт аналізує ситуацію й розпізнає найбільш корисну інформацію, оптимізує прийняття рішень. Експертні системи повинні демонструвати компетентність, тобто досягати у конкретній предметній області того ж

рівня, що й експерти. Методи знаходження рішень проблем розробляються на основі міркувань, що виходять із фундаментальних принципів у випадку неконкретних даних або неповних наборів правил.

Стандартні компоненти експертної системи:

- користувач – фахівець, для якого призначена система;
- інтерфейс користувача – комплекс програм, що реалізує діалог користувача з експертною системою у зручному для нього режимі;
- база знань – ядро експертної системи;
- «розв'язувач» – програма (система), що моделює хід міркувань експерта на підставі знань, завдяки вбудованій у неї загальній стратегії знаходження розв'язання, наприклад, шляхом логічного висновку;
- підсистема пояснень – підсистема, яка пояснює користувачеві хід міркувань;
- інтелектуальний редактор бази знань – програма, що дозволяє інженерові створювати, редагувати й доповнювати базу знань у діалоговому режимі.

Відмінність експертних систем від інших комп'ютерних систем:

- маніпулювання знаннями, а не даними;
- як правило, одержання ефективних оптимальних рішень;
- одержання іноді помилкових рішень;
- потенційна здатність до навчання, у тому числі на своїх помилках.

Переваги експертних систем:

- компетентність не втрачається, може нарощуватися, передаватися, документуватися, відтворюватися, копіюватися;
- досягаються більш стійкі результати, тому що відсутні емоційні й інші людські фактори;
- дешевше висококваліфікованих фахівців, висока вартість розробки врівноважується низькою вартістю експлуатації;

Недолік експертних систем – менша пристосованість до навчання, творчості й винахідливості, у порівнянні з висококваліфікованим фахівцем. Штучна компетентність експертних систем не може повністю замінити людину.

Моделювання мови необхідно для повноцінного спілкування в людино-комп'ютерному середовищі, а моделювання логічних міркувань - для рішення різних задач інтелектуальною системою (ІС). Система штучного інтелекту повинна вирішувати творчі задачі і вести діалог з людиною на основі закладених у неї знань (процедурних і декларативних). Прикладом інтелектуальної штучної системи є експертна система (ЕС). Якість ЕС визначається в першу чергу тим, наскільки природним є спілкування з нею людини під час розв'язування задач.

II. Мови представлення знань як варіант штучних мов. У штучному інтелекті важливим є поняття «знання», а при побудові інтелектуальних систем – мови представлення знань. **Знання** – це досвід, навички й уміння.

Знання бувають **декларативними** («знаю що») і **процедурними** («знаю як»). Знання в інтелектуальній системі представляють спеціальними моделями штучними мовами лінгвістичного забезпечення цієї ІС. Для прикладної лінгвістики особливо важливі **мови представлення знань (МПЗ)**, що використовують для опису знань, МПЗ – це штучні мови, побудовані за законами штучних мов. Найбільш відомі МПЗ – це логічні, мережні, фреймові і продукційні мови.

Логічні МПЗ представляють знання у вигляді синтаксично правильних формул якої-небудь формальної логічної системи. Логічних систем дуже багато, але найчастіше використовують логіку предикатів 1-го порядку і її мову.

Приклад:

Василь любить Марійку. – речення ПМ.

L (V, M) – речення мовою логіки предикатів 1-го порядку

Можна написати і по-іншому – **Любити (Василь, Марійка),**

Рішення задач у логіці предикатів 1-го порядку на логічній МПЗ – це логічний (дедуктивний) висновок за правилами цієї логічної системи.

Мережні МПЗ як моделі використовують семантичні мережі, де вузли мережі – це певні інформаційні одиниці – поняття, факти, процеси, імена, а дужки – відносини між ними. Відносини можуть бути будь-якими (тимчасові, причинно-наслідковими, «менше-більше» і т. д.). Мережні МПЗ часто використовують для явного опису відносин у тій чи іншій ситуації. Рішення задач у мережних моделях зводиться до пошуку фрагмента мережі, що збігається з даним зразком і до організації логічного висновку у семантичній мережі. Найбільша проблема мереж – їхня розмірність.

Фреймові МПЗ. Основні характеристики цих МПЗ – компактність і вкладеність рівнів. Фрейм («frame» – каркас, кістяк, рамка) – це шаблон типової ситуації чи певна структура, що містить дані про визначений об'єкт, його характеристики або їхні значення, і виступає як цілісна одиниця знань. Поняття «фрейму» увів М.Мінський. У класичному розумінні це **мінімальний опис чи факту явища, у якого не можна видалити жодну з частин опису без того, щоб не втратити повноту цього опису.** Фрейм може бути **рольовий і структурний.** Структурний фрейм переважно відбиває декларативні знання, тобто описує структуру якого-небудь поняття. Рольовий фрейм, на відміну від структурного, представляє процедурні знання, наприклад сценарії яких-небудь робіт. Сфера використання рольових фреймів – робототехніка.

Продукційні МПЗ представляють в основному процедурні знання, причому формула проста – **«Якщо ..., то ...».**

Приклад:

«Якщо температура 40 С°, то викликай швидку допомогу за телефоном 103».

Задачі для продукційної моделі формулюють як задачі пошуку потрібної послідовності продукцій, при якому досягається необхідна мета. Продукції – це найпопулярніша форма представлення знань в експертних системах.

Питання для самоперевірки:

- I. Наведіть визначення моделі, мовної моделі.*
- II. Дайте визначення експертної системи. Наведіть компоненти ЕС, її структуру.*
- III. Перерахуйте переваги та недоліки ЕС.*
- IV. Наведіть приклади практичного застосування ЕС у сучасному житті.*

ЛЕКЦІЯ 5

ЛІНГВІСТИКА У ЗАДАЧАХ ІНФОРМАЦІЙНОГО ПОШУКУ.

План

- I. *Лінгвістика у задачах інформаційного пошуку.*
 1. *Інформаційний пошук.*
 2. *Види пошуку.*
 3. *Характеристики інформаційного пошуку.*
- II. *Інформаційно-пошукові мови як штучні мови. Класифікація інформаційно-пошукових мов.*
- III. *Формула семантичного пошуку.*

I. Лінгвістика у задачах інформаційного пошуку. Пошук інформації – це процес відшукування у деякій системі збереження інформації таких документів (текстів, записів і т.д.), котрі відповідають запиту, що надійшов.

Засобами збереження і пошуку інформації виступають **інформаційно-пошукові системи (ІПС)**, елементами яких є **масив документів** (база даних), що виступають як об'єкт пошуку, **інформаційно-пошукова мова (ІПМ)**, **правила індексування, правила пошуку, технічні засоби** (для автоматизованих ІПС).

При уведенні документа у базу даних ІПС його **індексують**, тобто визначають ключові слова (тему, терміни) і виражають їх інформаційно-пошуковою мовою, прийнятою у цій системі, та записують у вигляді **пошукового образу (ПО)**. Так, база даних ІПС складається з пошукових образів (формальна модель для пошуку). Під час пошуку відбувається зіставлення вашого запиту, тобто того, що ви запитали мовою ІПС, із пошуковим образом, тобто тим, що у ній зберігається. **Пошукове розпорядження** (запит на пошук) також повинне бути обмежене за мовою і мати визначене формальне представлення.

У залежності від того, яка модель знаходиться у формі (in form), розрізняють наступні **види автоматизованого пошуку**:

- 1) **бібліографічний** (запис у каталогах, типу Record);
- 2) **бібліотечний** (тексти, документи-записи у вигляді файлів);
- 3) **фактографічний** (табличні дані у базі даних);
- 4) **продукційний** (процедурні правила виду "якщо..., то...").

Характеристики інформаційного пошуку – це такі його семантичні показники, як:

- a) **повнота видачі/втрати інформації;**
- б) **точність її видачі;**
- в) **інформаційний шум.**

II. Інформаційно-пошукові мови як штучні мови. Класифікація інформаційно-пошукових мов. Із визначення інформаційно-пошукової системи видно, що її лінгвістичними засобами є **спеціалізована штучна мова (інформаційно-пошукова мова)**, а також **методи індексування та пошуку документів**. Розглянемо основний лінгвістичний елемент ІПС – її мову.

Інформаційно-пошукова мова (ІПМ) – це спеціалізована штучна мова, призначена для індексування (опису) формальних характеристик документів у вигляді пошукового

образу, а також для опису інформаційних запитів (пошукових розпоряджень) до інформаційно-пошукових систем.

Інформаційно-пошукова мова, як і будь-яка мова, складається з **алфавіту, словника** (терміни чи спеціальні символічні коди), **граматики** і є **штучною мовою**, тобто обмеженою за своєю формою і структурою. Найчастіше для побудови пошукового образу і розпорядження прийнято використовувати **переліки слів**, згруповані за **визначеними правилами** (методом упорядкування за допомогою логічних зв'язувань).

У класичній літературі з інформаційного пошуку звичайно виділяють наступні види інформаційно-пошукових мов:

- a) ієрархічні;
- b) алфавітно-предметні;
- c) фасетні;
- d) дескрипторні;
- e) семантичні коди.

Ієрархічні ІПМ («ієрархія» – класифікація від загального до частки). Така організація використовується для пошуку книг у бібліотеці (наприклад, «Універсальна десяткова класифікація» – УДК).

Алфавітно-предметні ІПМ представлені у вигляді алфавітного списку ключових слів якого-небудь документа (наприклад, алфавітно-предметний покажчик наприкінці книг). Використовуються для побудови різних покажчиків, каталогів, картотек.

Фасетні мови також використовують ієрархічне дерево класів, але групують його у блоки (фасети). Індексування у фасетних мовах ефективніше за рахунок формул-шаблон.

Дескрипторні мови описового характеру використовують списки слів (дескриптори), організовані у спеціальні словники (тезауруси).

Типова словникова стаття має вигляд:

- ключові слова (синоніми);
- родові слова (дескриптор, що підпорядковує поданий);
- видові слова (дескриптор, що підпорядковуються поданому);
- асоціації (відносини).

У таких мовах використовується принцип координатного індексування тобто, перелічуються ключові слова або слова-дескриптори, що виражають центральну тему чи цілісну характеристику даного об'єкту.

Дескрипторні мови – найбільш природна і популярна форма для вираження **пошукових образів (ПО) і пошукових розпоряджень (ПР)** і в даний час широко використовується у сучасних пошукових системах (зокрема в багатьох пошукових системах мережі Internet).

III. Формула семантичного пошуку. В ІПС семантика являє собою значні труднощі, тобто, для такого роду систем основна проблема – це **проблема моделювання суті, чи змісту документа**, що зберігається, а також **формулювання запиту, і їхнє зіставлення**. За допомогою семантичних засобів ІПС здійснюються значеннєві операції обробки документів і запитів. Простіше послідовність цих операцій відобразити формулою:



де

Ci – план змісту запиту (значення запиту);

Li – план вираження запиту природною мовою;

Si – пошуковий запит (представлення запиту пошуковою мовою);

Sd – представлення документа пошуковою мовою (пошуковий образ);

Ld – текст документу природною мовою (план вираження документа);

Cd – поняття у документі, що відшукувався;

→ – індексування;

↔ – порівняння (зіставлення).

Часто для індексування використовуються *тезауруси*, спеціальні словники для інформаційного пошуку, організовані за принципом зіставлення слів з їхніми поняттями.

Питання для самоперевірки:

- I. Дайте визначення інформаційного пошуку, інформаційно-пошукової системи.*
- II. Перерахуйте компоненти ІПС. Перерахуйте різні типи ІПС.*
- III. Наведіть характеристики інформаційного пошуку. Які з них, на вашу думку, найбільше впливають на ефективність пошуку інформації?*
- IV. Прокоментуйте кожен елемент флормули семантичного пошуку. Поясніть, у чому полягає складність організації ефективного автоматичного пошуку.*

ЛЕКЦІЯ 6

ЛЕКСИКОГРАФІЯ, ЗАДАЧІ ТА СУЧАСНІ ТЕНДЕНЦІЇ РОЗВИТКУ.

План

- I. Лексикографія як один з важливих напрямків прикладної лінгвістики. Словник як об'єкт лексикографії. Класифікація й організація словників.*
- II. Традиційна і машинна лексикографія. Основні напрямки машинної лексикографії. Деякі особливості автоматичних словників.*
- III. Ідеографічні словники та тезауруси.*

I. Лексикографія як один з важливих напрямків прикладної лінгвістики. Словник як об'єкт лексикографії. Класифікація та організація словників. Лексикографія, як один із найважливіших напрямків прикладної лінгвістики, займається теорією і практикою укладання словників. Під словником розуміють:

- повний словниковий склад мови,
- довідкову книгу слів, розташованих у визначеному порядку, що дає кому-небудь інформацію про те чи інше слово.

Основна задача словника – представлення або опис лексики мови. Це надзвичайно складне завдання, тому що лексика мови має тенденцію збільшуватися і якісно змінюватися.

Основною структурною одиницею словника (як книги) є словникова стаття. Словник словника – це перелік термінів словника без їхніх тлумачень.

Класифікація словників:

1. лінгвістичні
 - 1.1. спеціальні
 - 1.2. загальні
2. нелінгвістичні (енциклопедії і т. д.)
3. термінологічні
4. ідеографічні (тезауруси і т. д.)

Примітка: загальні лінгвістичні словники (розумні, перекладні, орфографічні, частотні і др.), спеціальні (синонімів, антонімів, словник прислів'їв та приказок, сленга, словники іноземних слів та ін.)

Найважливішим питанням при укладанні словників є порядок розташування словникових статей, найчастіше це алфавітний або предметний (тематичний) порядок, згідно якого слова групуються по темах (наприклад, словник у картинках).

II. Традиційна і машинна лексикографія. Основні напрямки машинної лексикографії. Деякі особливості автоматичних словників. У лексикографії існує два науково-практичних напрямки: **традиційна та машинна лексикографія**. Традиційна лексикографія має глибокі історичні корені й у більшій мірі займається теорією та практикою укладання «традиційних» словників. На відміну від першої, машинна (комп'ютерна) лексикографія – молода наука, що реалізує традиційні напрацювання у технічних середовищах і створює різноманітні автоматичні словники.

Машинна лексикографія займається автоматизацією підготовки словників та перетворює словники у комп'ютерну форму. У даний час виділяють три основних напрямки машинної лексикографії:

- автоматичне укладання різних словників з тексту за допомогою комп'ютера (наприклад, частотні словники, словники конкордансів, термінологічні);
- розробка теоретичних та практичних аспектів укладання комп'ютерних словників для машинного перекладу, штучного інтелекту і т. д. (наприклад, словники коренів для морфологічного аналізу);
- створення словників, що є машинними версіями традиційних словників.

Автоматичні словники – це складний комплекс комп'ютерних програм. Відзначимо **деякі особливості автоматичних словників:**

- крім словникової бази даних (перелік слів за алфавітом) для роботи автоматичного словника необхідні спеціальні алгоритми, наприклад, алгоритм лематизації. **Лематизація** – це приведення різних форм слова до його канонічної (вихідної) форми (подібно до морфологічного аналізу слів);
- у машинних словниках присутні не тільки переліки окремих слів, але і до 50% словосполучень (особливо у термінологічних словниках, що дуже важливо для перекладу технічних текстів);
- автоматичні словники мають у даний час велике розповсюдження у машинному перекладі. Ефективність перекладу набагато вища при використанні

комп'ютерного словника, тому що пошук здійснюється у багато разів швидше. Найголовніше у таких словниках – це представлення ВСІХ значень слова.

Примітка: варто розрізняти автоматичні словники і системи машинного перекладу. Останні лише включають перші і є значно складніше.

Комп'ютерна лексикографія має набір інструментальних засобів для реалізації своїх задач, тобто бази даних, комп'ютерні каталоги, програми обробки текстів, які дозволяють у автоматичному режимі формувати словникові статті, зберігати словникову інформацію та здійснювати її обробку. Лексикографічні програми можна поділити на дві складові: **програми підтримки лексикографічних робіт та автоматичні словники різних типів, що містять лексикографічні бази даних.** Така рівнева структура інструментального апарату комп'ютерної лексикографії вимагає адекватного підходу до його вивчення. Для успішної реалізації практичних завдань лексикографії майбутнім фахівцям необхідно розуміти, що узгодженість роботи інструментальних засобів на різних рівнях забезпечує успішну реалізацію завдань цієї галузі. Важливо також усвідомлювати, що для розробки і підтримки програм здійснення лексикографічних робіт фахівцям необхідні знання принципів програмування та конфігурування словників, які їм забезпечує цикл дисциплін з інформатики. Створення автоматичних словників для кінцевого користувача – також робота спеціаліста-лексикографа із знаннями програмування, в той час як лексичне наповнення, оптимізація структури та обсягу словників вимагає, перш за все, знань лінгвістичних дисциплін: лексикології та лексикографії. Отже, лише успішне оволодіння основами суміжних галузей – інформатики, лексикології та лексикографії – здатне забезпечити відповідні професійні уміння. З цієї точки зору, завданнями курсу прикладної лінгвістики є, в першу чергу, наголосити на потребі у навичках суміжних дисциплін, ознайомити з основними програмами, які функціонують як програми підтримки словників, а також зосередити увагу на вивченні базових принципів укладання словників, їх типів та з'ясувати, які лексикографічні параметри (обсяг, повнота опису, лексична база тощо) є вирішальними для ефективної роботи словника, й електронного словника, зокрема.

Термін „електронний словник” існує вже досить давно і звично використовується на позначення „паперового” словника, що представлений у комп'ютерному варіанті і має більш зручні у порівнянні з останнім засоби відображення та пошуку. Насправді електронний словник – це особливий лексикографічний об'єкт, який відрізняється від класичного „паперового” словника не лише за принципами відображення; він дозволяє реалізацію багатьох ідей, спрямованих на розширення і удосконалення його можливостей.

Принципові відмінності між традиційним „паперовим” та „електронним словниками” спостерігаються на різних рівнях. Так, на рівні створення словника, електронні словники мають очевидні переваги, оскільки цей процес потребує набагато менше етапів, оминаючи такі процедури як: складання рукопису словника, передрукування і редагування рукопису, авторське доопрацювання, верстування і друк. Очевидно, що суттєве зменшення кількості етапів означає економію матеріалу, часу і витрат, проте найбільшим недоліком паперових словників можна вважати те, що за час тривалого процесу створення, „картина світу”, яку словники повинні фіксувати у системі своїх значень, прикладів та перекладів „старішає” і не відбиває об'єктивну реальність. Ще суттєвіші відмінності спостерігаємо на іншому рівні – рівні реалізації сутнісних характеристик і можливостей. Так, наприклад, великий обсяг паперового словника може викликати труднощі під час користування, а специфічна концепція (наприклад, створення багатомовних, комбінаторних словників), яка вимагає застосування інтегрованих засобів опису лексичних значень, відчутно звужує його лексичну базу. І навіть електронні версії паперових словників, що вже існують, дозволяють подолати згадані проблеми. Окрім того, доступ до електронних словників може реалізовуватися за допомогою різних лінгвістичних

технологій (розпізнавання і синтезування звуку, пошук по тексту, здійснення морфологічного та синтаксичного варіантів аналізу слів у тексті тощо).

Отже для майбутнього спеціаліста у галузі прикладної лінгвістики необхідно усвідомити, що два типи словників – паперові та електронні – розрізняються не лише за типом носія; різняться як практичні етапи їх створення, так і принципи, за якими вони побудовані, і можливості, які вони надають. Для ефективної реалізації цього завдання у процесі навчання доцільним є виконання практичних завдань із застосуванням різних видів паперових та електронних словників. Ефективним також є порівняння процесів роботи з ідентичними словниками у різних їх версіях (електронній та друкованій).

Можливо найбільш важливим аспектом сучасної комп'ютерної лексикографії є **лексикографічне моделювання**, яке можна розглядати на різних рівнях: від створення унікальної словникової статті до побудови концептуальної „картини світу”, що реалізується за допомогою „фреймів” (структур представлення інформації) та „сценаріїв” (процедур представлення знань). Реалізація можливостей лексикографічного моделювання тісно пов'язана з поняттям **гіпертексту** як основного способу організації віртуального текстового простору.

У звичайному паперовому словнику мінімальною одиницею доступу є лексема, ім'я словникової статті, а структура словникової статті є лінійною. Тому для визначення наявності відповіді на запит користувача, необхідно прочитати всю статтю, кількість значень і „підзначень” в якій може бути величезною. З точки зору структури електронний словник є аналогом друкованого і, якщо розглядати її лінійно, має аналогічні складові (наприклад, передмову, список скорочень, правила користування, граматичний довідник). Однак взаємодія між частинами словникової статті електронного словника не є лінійною і підпорядкована законам гіпертексту. Словникова стаття має чітку логічну структуру з ієрархічними зв'язками між елементами. Основним способом реалізації гіпертексту в електронних словниках є можливість користування посиланнями. Користувач, що здійснює запит, вводить певний параметр і активізує не всю словникову статтю, а лише ті елементи, які його цікавлять, і йому не потрібно проглядати всю статтю. Далі, за допомогою посилань користувач послідовно створює свою унікальну інформаційну структуру. Як бачимо, електронний словник здійснює не лише інформативну функцію, а і діє як пошуковий механізм: знаходить необхідну інформацію і представляє її у зручній для користувача формі.

Внутрішня структура електронного словника також відрізняється від друкованого. Вона побудована за принципами логіки мови програми підтримки словника і не обов'язково побудована за алфавітним чи гніздовим принципами. Словникова стаття запиту може бути навіть не знайденою, а синтезованою програмою у відповідності із запитом користувача. Такий результат є можливим завдяки тому, що інформаційні елементи, потенційно закладені у словникову статтю, є незалежними, але водночас здатні доповнювати один одного. Усвідомлення специфіки структури електронної словникової статті дає можливість говорити про варіанти лексикографічного моделювання. На рівні конкретного запиту воно проявляється у побудові індивідуальної словникової статті відповідно до потреб користувача, на рівні ж словника – це можливість концептуальної побудови „моделі світу”.

З методологічної точки зору зауважимо, що поняття „гіпертексту” як специфічно організованого, „багатовимірною” тексту повинно засвоюватися в контексті принципу „гіпертекстуальності”, який є невід'ємною і суттєвою рисою сучасного інформаційного комп'ютерного простору. Необхідним є розуміння того, що гіпертекстовість реалізується у більшій мірі завдяки системі посилань, які не лише дають можливість проникати „вглиб” тексту, а й реалізують потенціал лексикографічного моделювання, створюючи унікальні

„моделі світу”. Практично, вивчення механізмів гіпертексту реалізується через роботу в мережі Інтернет, пошукових системах та створення власних електронних проєктів (презантацій, блогів), що ілюструють принцип гіпертекстуальності.

III. Ідеографічні словники та тезауруси. Клас ідеографічних словників (предметні, тематичні), до яких відносяться і тезауруси, – це особливого роду словники, організовані, по перше, за тематичним принципом, і, по друге, за принципом «від змісту до слова», тобто ідеографічні словники орієнтовані на семантику мови, і кожен такий словник – це певна семантична модель лексики, побудована на ієрархічних відносинах типу «рід – вид», «частина – ціле», «синоніми» і т. д. Найвідоміший класичний ідеографічний словник – словник Пітера Роже, у якому міститься 1000 основних тем.

Тезаурус – це також ідеографічний словник, але він має чітку ієрархію відносин. Словник тезауруса – це перелік дескрипторів («дескриптор» – однозначний стандартний термін). У словникових статтях тезауруса відбиті істотні для даного терміна зв'язки з іншими поняттями, іншими словами – це, фактично, маленька енциклопедія. Тезауруси у машинній лексикографії – дуже точні інструменти, тому що вони призначені для інформаційного пошуку, а характеристики інформаційного пошуку (інформаційний шум, точність видачі та ін.) безпосередньо залежать від структури тезауруса, тому існують чіткі стандарти для зіставлення інформаційно-пошукових тезаурусів.

Питання для самоперевірки:

- I. Дайте визначення лексикографії, словника (2 різних тлумачення).*
- II. Наведіть класифікацію словників та приклади різних типів словників. У чому полягають особливості кожного типу словників?*
- III. Дайте визначення машинної лексикографії, окресліть її задачі.*
- IV. Перерахуйте особливості машинних словників, наведіть відмінності між паперовими та електронними словниками, переваги та недоліки кожного типу.*
- V. У чому полягає специфіка тезаурусів? Які вони мають переваги та недоліки у порівнянні з іншими видами словників?*

ЛЕКЦІЯ 7

МАШИННИЙ ПЕРЕКЛАД, ЙОГО ВИДИ І ОСОБЛИВОСТІ. МАШИННОГО ПЕРЕКЛАДУ В ЗАДАЧАХ ПРИКЛАДНОЇ ЛІНГВІСТИКИ.

План

- I. Машинний переклад (МП). Види МП.*
- II. Задача машинного перекладу як одна з найважливіших задач прикладної лінгвістики. Підходи до моделювання МП.*
- III. Основні етапи МП.*

I. Машинний переклад. Види МП. Переклад з однієї мови на іншу у загальному вигляді полягає у зміні алфавіту, лексики та синтаксису мови зі збереженням її семантики.

Переклад – це вид інформаційної діяльності, потреба в якій ніколи не скорочується, а навпаки зростає. Дослідження ринку перекладів показали, що цей вид діяльності збільшується на 15% на рік. У складі перекладів переважають науково-технічні переклади – це 40% від загального обсягу, далі – юридичний, бухгалтерський, усний, синхронний, навчальний, ... художній – 0,3%.

Проблема моделювання перекладу як додатка у комп'ютерній техніці є центральною проблемою як для прикладної лінгвістики, так і для штучного інтелекту. Очевидно, що автоматизація перекладу дозволить підвищити його ефективність, а також розширить межі міжлюдської комунікації.

Машинний переклад (МП) – це перетворення комп'ютером тексту однією природною мовою в еквівалентний за змістом текст іншою природною мовою. **Системи МП** – це комплекс спеціальних найскладніших програм й алгоритмів плюс спеціальні автоматичні словники вхідної та вихідної мов.

Історію МП відраховують з 1946 року, з виникнення такої науки, як **кібернетика** («науки про керування»). У 1954 році був проведений так названий «Джоржтаунський експеримент»: перекладався текст (250 слів) з російської мови англійською. Перші ж досвіди з МП показали величезні труднощі під час моделюванні семантики (змісту). Ці проблеми не вирішені і дотепер. Проте, у даний час актуальність МП дуже висока, що пов'язано із зростанням інформації, комунікації, комп'ютеризації суспільства.

Види сучасного МП:

- **інформативний** – грубий послівний переклад без участі людини, достатній для поверхневого ознайомлення зі змістом тексту (по-іншому – **автоматичний**);
- **професійний** – більш якісний переклад з наступним редагуванням людиною (**автоматизований**);
- **персональний** – вважається дуже професійним перекладом. Цей тип перекладу проходить у режимі діалогу людини з комп'ютером (системи типу **TRANSLATION MEMORY**).

Відомими програмними продуктами для автоматичного та автоматизованого перекладу довгий час були Промт (Stylus, Промт, Magic Gooddy), Socrate, SYSTRAN (амер.), LOGOVISTA (амер.), EUROLANG (франц.), LOGOS (амер.). Усі вони пропонували цілком задовільну якість інформативного й автоматизованого перекладу. Системи для персонального перекладу на порядок складніше і дорожче і є прерогативою професійних перекладачів художньої та технічної літератури. Якість МП багато в чому залежить від кількості та якості автоматичних словників, що дозволяють перекласти усі відтінки значень слова, підібрати потрібний термін у залежності від конкретної предметної області перекладного тексту.

II. Задача машинного перекладу як одна з найважливіших задач прикладної лінгвістики. Підходи до моделювання МП. Основною задачею машинного перекладу є аналіз та синтез тексту природною мовою, тобто задача **лінгвістичної трансляції**. Вона вимагає:

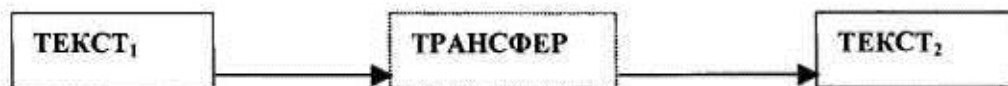
- розпізнавання символів (графіки) на етапі введення у комп'ютер;
- морфологічного і лексичного аналізу;
- синтаксичного аналізу та синтезу;
- семантичних перетворень.

Під час МП машина переглядає всі рівні мовної структури, переходячи від однієї підзадачі до іншої. Для моделювання МП необхідне моделювання його підзавдань, розробка алгоритмів та комп'ютерних програм для роботи цих алгоритмів. Якщо уважно

придивитися до етапів МП, можна побачити, що все це окремі і складні задачі прикладної лінгвістики.

За історію машинного перекладу виділилися **три підходи до його моделювання**:

- «**прямі**» системи МП дають послівний переклад (слово в слово);
- «**текст – зміст – текст**», тут метою є розробити значущу мову-посередник;
- «**текст – текст**». Переклад відбувається на рівні перекладних відповідностей мовної пари. Одиницею перекладу виступає перекладна відповідність, тісно пов'язана з лексичною одиницею (словосполучення, слово).



У третьої моделі з'являється проміжний етап – **трансфер**, на якому відбувається встановлення перекладних відповідностей (узгодження мов). Така модель перекладу найбільш вдала і в даний час переважає над першими двома, котрі більш обмежені у своїх можливостях.

III. Основні етапи МП. Класично МП розбивається на **етапи аналізу вхідного тексту та синтезу вихідного тексту**.

Перетворення повинне починатися із попередньої підготовки – аналізу тексту. Наприклад, якщо на вхід англо-української системи МП надходить текст китайською мовою, те найперший етап графічного та морфологічного аналізу повідомить про неможливість перекладу. Розрізняють наступні види аналізу: **морфологічний, лексичний, синтаксичний, семантичний**. Метою кожного етапу аналізу є побудова внутрішнього представлення вхідного речення.

Так, метою **етапу трансферу** є перетворення структури вхідного речення у внутрішню структуру вихідного за допомогою спеціальних формальних правил трансферу.

Мета **етапу синтезу** – на основі отриманої в результаті аналізу структури побудувати (синтезувати) правильне речення вихідною мовою.

Прикладом машинного перекладу може служити операція **трансляції** в мовах програмування, коли введена у комп'ютер програма (наприклад, програма мовою Pascal) перекладається мовою машини (двійкові коди 1/0).

Трансляцію здійснює складна програма – **компілятор**. Але мова програмування – штучна мова, де формальною граматикою задані тверді правила, тому сам процес перекладу простіший, ніж у природній мові.

Питання для самоперевірки:

- I. Дайте визначення машинного перекладу*
- II. Перерахуйте види сучасного машинного перекладу. Наведіть приклади програм перекладу кожного виду. У чому полягають переваги та недоліки цих видів?*
- III. Перерахуйте задачі машинного перекладу.*
- IV. Назвіть основні етапи машинного перекладу. Які етапи, на вашу думку, є найскладнішими для практичної реалізації у системах МП? Чому?*

ЛЕКЦІЯ 8

АВТОМАТИЧНА ОБРОБКА ТЕКСТУ. ЛІНГВОСТАТИСТИКА. ЛІНГВОДИДАКТИКА.

План

- I. Автоматична обробка тексту. Текст та гіпертекст.
- II. Розпізнавання усної і письмової мови.
- III. Лінгвостатистика.
- IV. Автоматизоване навчання іноземним мовам (лінгводидактика).

У даній лекції ми коротко розглянемо інші прикладні задачі лінгвістики, що з'явилися або отримали новий розвиток із виникненням комп'ютерної техніки.

I. Автоматична обробка тексту. Текст та гіпертекст. Автоматична обробка тексту – це будь-яке перетворення тексту природною чи штучною мовами за допомогою комп'ютера.

На початку розвитку комп'ютерних технологій користувачеві були доступні лише примітивні операції з обробки тексту у простих комп'ютерних програмах-редакторах (Lexicon, NC), згодом вимоги до редагування та представлення документів зростали, що призвело до створення удосконалених систем типу WORD Processors (Microsoft Word та ін.). З'явилися настільні видавничі системи для професійної верстки документів, книг (наприклад, PageMaker). Таким чином, досить розвиненою на дійсний час операцією з перетворення тексту є автоматизоване редагування – коректування введеного в комп'ютер текстового матеріалу людиною-редактором (наприклад, форматування тексту).

Наступним етапом стали комп'ютерні системи типу TEXT Processors для автоматичного реферування та анотування текстів, машинного перекладу та лексикографічної обробки текстів.

Системами найвищої складності стали програми для обробки природної мови типу **NLP (Natural Language Processors)** для складних систем штучного інтелекту, експертних та професійних систем машинного перекладу.

Останнім часом з розвитком комп'ютерних мереж і нових інформаційних технологій значно трансформувалися традиційні лінгвістичні поняття, зокрема поняття тексту.

Під **текстом** (лат. «textus») – зв'язність, матеріал, сплетення), письмовим чи усним, прийнято розуміти логічно зв'язану послідовність лінгвістичних знаків.

Основні характеристики тексту:

- зв'язність;
- усвідомленість;
- цілісність (текст повинен бути закінченим за змістом).

Крім того, класично текст має лінійну структуру, але із середини ХХ століття з'явилося нове розуміння тексту, коли стали вважати, що нелінійний характер тексту може бути ефективнішим. Із середини 80-х років, коли швидкими темпами стали розвиватися комп'ютерні та телекомунікаційні мережі (WWW, Internet) і разом з тим міжлюдська комунікація, актуальним стало нове поняття гіпертексту (hypertext).

Гіпертекст – особливим образом структурований текст, розбитий на окремі блоки, що має нелінійне представлення для ефективної презентації інформації у комп'ютерних середовищах. Гіпертекст є ієрархічною структурою (перехід від посилання до посилання з поверненням на початкову сторінку). Поряд зі зручним представленням тексту, існують і обмеження на гіпертекст:

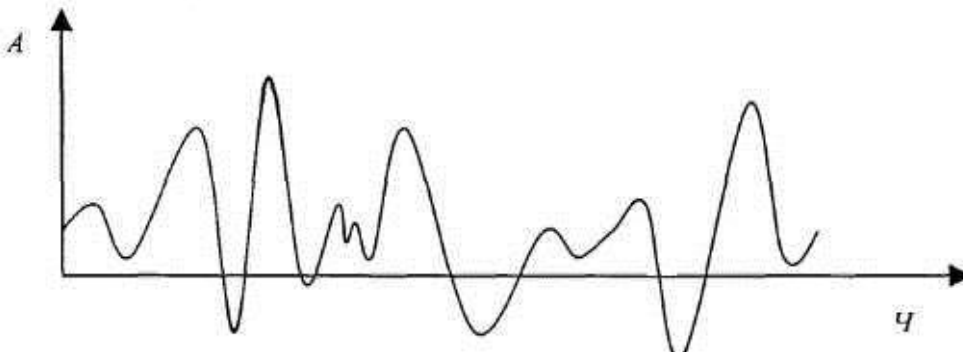
- за обсягом екрана (1-2 екрана на інформативний вузол);
- за глибиною (вкладеність інформації не повинна перевищувати 5 рівнів);
- проблеми навігації (руху по мережі при її великій глибині).

Для стандарту представлення інформації в мережах розробляються спеціалізовані мови. Наприклад, у мережі Internet ця мова називається **HTML – HyperText Markup Language (мова гіпертекстової розмітки документів)**.

II. Розпізнавання усної та письмової мови. Розпізнавання комп'ютером усної і письмової мови також відноситься до задач прикладної лінгвістики. Моделювання мовних сутностей та їхнє використання можливе тільки тоді, коли мовна інформація вводиться у комп'ютер у вигляді письмових текстів чи текстів усного мовлення.

а) **Розпізнавання мови (Voice/Speech Recognition).**

Комунікація людини і машини за допомогою мови зручна й ефективна, тому що усне мовлення одна з найшвидших форм спілкування. Із середини 50-х років ХХ століття цей напрямок прикладної лінгвістики одержав розвиток з винаходом спеціального фізичного приладу – **спектрографа**. Під час розпізнавання голосу існує тісний зв'язок лінгвістики та фізики (акустики), тому що кожен звук має свої акустичні характеристики, що відображаються на спеціальних графіках (спектрограмах) у координатах «час – частота – амплітуда». У даний час спектрограми (див. мал.) є найбільш доступною формою фіксації усного мовлення, на основі яких будується автоматичне пофонемне розпізнавання мови (у вигляді фонологічної транскрипції).



За результатами численних досліджень було встановлено, що автоматичне розпізнавання мови є дуже складним, за винятком того випадку, коли словник розпізнаваних слів обмежений. У системах з розпізнаванням мови розробники програмного забезпечення поки не вийшли за рівень лексики мови. Незважаючи на свою складність, цей напрямок розвивається швидкими темпами, і вже в 90-х роках ХХст. з'явилися комерційні системи розпізнавання мови для різних цілей.

б) **Розпізнавання графем (письмової мови).**

Букви алфавіту, пунктуаційні знаки й інші символи служать у якості складових будь-якого природно-мовного тексту, що вводиться в пам'ять комп'ютера. Розпізнавання символів реалізується ефективніше для друкованого тексту, ніж для рукописного; і у даний час задача розпізнавання друкованого тексту практично вирішена. Для рукописного тексту розпізнавання графем набагато складніше, тому що потрібні більш складні алгоритми розпізнавання образів.

III. Лінгвостатистика. Основною задачею статистичної лінгвістики (лінгвостатистики) є застосування кількісних математичних методів для розкриття закономірностей функціонування одиниць мови у мові, а також установа закономірностей побудови тексту. Колективні дослідження в області лінгвістики ведуться у таких основних напрямках:

- розробка лінгвостатистичних моделей на матеріалі різних мов;
- пошук певної універсальної математичної загальмовної моделі мови;

– статистична лексикографія.

Актуальною проблемою статистичної лексикографії є створення частотних словників, що дають матеріал для вивчення структури текстів різних предметних областей науки і техніки.

IV. Автоматизоване навчання іноземним мовам (лінгводидактика). Одним з напрямків прикладної лінгвістики є можливість автоматизованого навчання іноземним мовам (комп'ютерна лінгводидактика). За кордоном цей напрямок, відомий як **Computer Assisted Language Learning (CALL)**, є перспективним і викладається як спеціальна дисципліна прикладного мовознавства на лінгвістичних факультетах коледжів і університетів.

Застосування комп'ютерних технологій (особливо технології «мультимедіа») у процесі навчання іноземним мовам позитивно зарекомендувало себе в останні кілька десятиліть. Мультимедіа технологія дозволяє моделювати середовище, яке імітує лінгвістичну та комунікативну реальність, що дуже важливо для мовного навчання, а також активізувати основні методичні принципи навчання іноземним мовам – розвиток навичок: аудіювання, говоріння, читання та письма. Невід'ємною частиною комп'ютерного розвитку з упровадженням нових інформаційних і мережних технологій стали такі електронні засоби його підтримки, як машинні перекладачі та словники, що автоматизують переклад необхідної інформації, одержуваної в WWW. Величезний потенціал у даний час має використання можливостей Internet для CALL.

Важливою умовою успішного застосування комп'ютера при мовному навчанні є створення потужних середовищ підтримки CALL, що, крім навчальних комп'ютерних курсів, складються з текстових процесорів, електронних словників, машинних перекладачів, різного виду довідників, тестових програм, орфографічних систем і т. д. Сполучною ланкою для організації CALL-середовищ слугує гіпертекстова технологія.

Питання для самоперевірки:

- I. Що таке «автоматична обробка тексту»? Дайте визначення тексту та гіпертексту, наведіть їх основні спільні риси та відмінності.*
- II. У чому полягають основні проблеми реалізації процесів автоматичного розпізнавання усного та писемного мовлення?*
- III. Дайте визначення лінгвостатистики. Наведіть приклади практичного застосування методів лінгвостатистики у прикладній лінгвістиці.*
- IV. Дайте визначення лінгво дидактики. Прокоментуйте два основних напрямки лінгводидактики у прикладній лінгвістиці. Наведіть приклади програм, що навчають мовам.*

БІБЛІОГРАФІЯ

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие / [Большакова Е. И., Клышинский Э. С., Ландэ Д. В., Носков А. А. и др.]. – М. : МИЭМ, 2011. – 272 с.
2. Баранов А. Н. Введение в прикладную лингвистику : [учебное пособие] / Анатолий Николаевич Баранов. – М. : Едиториал УРСС, 2003. – 360 с.
3. Беляева Л. Н. Лингвистические автоматы в современных гуманитарных технологиях / Лариса Николаевна Беляева. – Спб. : ООО “Книжный Дом”, 2007. – 192 с.
4. Білецький А.О. Про мову і мовознавство. – К., 1996. – 224 с.
5. Блехман М. С. Комп'ютерна лінгвістика / Михайло Самойлович Блехман. – Х. : ХГУ, 1997. – 153 с.
6. Бурдаєв В. П. Системи штучного інтелекту : конспект лекцій / Володимир Петрович Бурдаєв. – Х. : ХНЕУ, 2008. – 156 с.
7. Виноградов В. С. Введение в переводоведение (общие и лексические вопросы) / Венедикт Степанович Виноградов. – М. : Издательство института общего среднего образования РАО, 2001. – 224 с.
8. Височанський В.С., Кардаш А.І., Костів О.В., Черняхівський В.В. Елементи інформатики. – Львів, 1990. – 190 с.
9. Вінцюк Т. Комп'ютерні автоматичні системи розпізнавання та синтезу українського мовлення // Проблеми українізації комп'ютерів. – К., 1993. – С.21-32.
10. Гайна Г. А. Системи штучного інтелекту / Георгій Анатолійович Гайна. – К. : КНУБА, 2007. – 207 с.
11. Глибовець М. М. Штучний інтелект : [підручник для студентів вищих навчальних закладів] / М. М. Глибовець, О. В. Олецкий. – К. : Видавничий дім “КМ Академія”, 2002. – 366 с.
12. Городецкий Б.Ю. Компьютерная лингвистика: моделирование языкового общения // Компьютерная лингвистика. Вып. XXIV. – Новое в зарубеж. лингвистике. – М., 1989. – 432 с
13. Дарчук Н. П. Комп'ютерна лінгвістика (автоматичне опрацювання тексту) : підручник / Наталія Петрівна Дарчук. – К. : Видавничо-поліграфічний центр “Київський університет”, 2008. – 351 с.
14. Доля В. Г. Комп'ютерні системи штучного інтелекту : навчальний посібник для студентів вищих навчальних закладів / Віктор Гнатович Доля. – К : Університет “Україна”, 2011. – 296с.
15. Зубов А. В. Основы искусственного интеллекта для лингвистов : [учеб. пособие] / А. В. Зубов, И. И. Зубова. – М. : Университетская книга; Логос, 2007. – 320 с.
16. Інформаційні технології в перекладі : [навч. посібник] / [за заг. ред. Т. І. Коваль]. – К. : Видавничий центр КНЛУ, 2010. – 261 с.
17. Історія машинного перекладу: стислий огляд : [навч. посіб.] / [уклад. П. О. Бочан]; Чернів. нац. ун-т ім. Ю.Федьковича. – Чернівці : Рута, 2010. – 47 с.
18. Карпіловська Є. А. Вступ до прикладної лінгвістики : комп'ютерна лінгвістика : [підручник] / Євгенія Анатоліївна Карпіловська. – Донецьк : ТОВ “Юго-Восток, Лтд”, 2006. – 188 с.
19. Клименко Н.Ф. Нові підходи до укладання комп'ютерних словників // Мовознавство. – 1996. - № 4-5. – С.11-15.

20. Клименко Н.Ф., Карпіловська Є.А., Комарова Л.І. та ін. Морфемнословотвірний фонд української мови як дослідницька та інформаційнодовідкова система // Мовознавство. – 1990. - № 6. – С.41-50.
21. Клименко Н.Ф., Русанівський В.М. Від універсальної бази лінгвістичних знань до комп'ютерного укладання словників // Мовознавство. – 1995. - № 4-5. – С.3-10. 16.
22. Леонтьева Н. Н. Автоматическое понимание текстов: системы, модели, ресурсы : [учеб. пособие для студ. лингв. фак. вузов] / Нина Николаевна Леонтьева. – М. : Издательский центр “Академия”, 2006. – 304 с.
23. Мамедова М. Г. Машинный перевод : эволюция и основные аспекты моделирования / М. Г. Мамедова, З. Ю. Мамедова. – Баку : Издательство “INFORMASIYA TEXNOLOGIYALARI”, 2005. – 156 с.
24. Партико З. В. Прикладна і комп'ютерна лінгвістика : Вступ до спеціальності : [навчальний посібник] / Зіновій Васильович Партико. – Львів : Афіша, 2008. – 224 с.
25. Перебийніс В.І. Статистичні методи для лінгвістів: Навч. посібник. – Вінниця, 2002. – 176 с.
26. Перебийніс В.С., Муравицька М.П., Дарчук Н.П. Частотні словники та їх використання. – К., 1985. – 204 с.
27. Пещак М.М. Нариси з комп'ютерної лінгвістики. – Ужгород, 1999. – 200с.
28. Сніжко М.Д., Сніжко Н.В. Автоматизована система класифікації лінгвістичних об'єктів // Проблеми українізації комп'ютерів. – К., 1993. – С.69-72.
29. Сніжко Н.В. Идеографічний тезаурус як модель лексико-семантичної системи (за наслідками автоматизованого аналізу українських іменників) // Мовознавство. – 1995. - № 6. – С.28-35.
30. Сніжко Н.В., Сніжко М.Д. “Идеографічний тезаурус” як інформаційнодовідкова система при вивченні закономірностей структурнофункціональної організації лексики // Мовознавство. - 1996. - № 4-5. – С.23-28.
31. Стахмич Ю. С. Прикладна лінгвістика та інформатика : лабораторний практикум / Ю. С. Стахмич, В. Д. Мельник. – Івано-Франківськ : ІФНТУНГ, 2017. – 96 с.
32. Українська мова: Енциклопедія. – К., 2000. – Статті: Автоматизація лінгвістичних досліджень. Автоматичне оброблення тексту. Автоматичний аналіз. Автоматичний словник. Аналіз за безпосередніми складниками. Аналіз у термінах залежностей. Граматики формальні. Конкорданс. Математична лінгвістика. Машинна мова. Машинний переклад. Машинний фонд української мови. Модель. Прикладна лінгвістика. Прогнозування лінгвістичне. Формалізація в лінгвістиці. Частотний словник
33. Шевченко І.В. Алгоритмічна словозмінна класифікація української лексики // Мовознавство. – 1996. - № 4-5. – С.40-44.
34. Широков В.А. Інформаційна теорія лексикографічних систем. – К., 1998. – 331 с.