

МІНІСТЕРСТВО ОСВІТИ І НАУКИ
УКРАЇНИ

Національний авіаційний університет

А.І. Вавіленкова

**ТЕОРЕТИЧНІ ОСНОВИ АНАЛІЗУ ЕЛЕКТРОННИХ
ТЕКСТІВ**

КИЇВ 2016

УДК 510.635:004.891(02)

ББК Ч 800.5

В 121

Рецензенти:

О.Г. Додонов – д-р техн. наук, професор, заслужений діяч науки і техніки України (Інститут проблем реєстрації інформації НАН України);

Бісікало О.В. – д-р техн. наук, професор (Вінницький національний технічний університет).

Рекомендовано до видання вченою радою Національного авіаційного університету (протокол №2 від 23.03.2016 р.).

Вавіленкова А.І.

В121 Теоретичні основи аналізу електронних текстів: монографія. – К.: ТОВ «СІК ГРУПІ УКРАЇНА», 2016. – 192 с. ISBN 978-617-7092-79-6

Викладено теоретичні основи комп'ютерної обробки текстової інформації у розрізі моделей та алгоритмів формалізації, аналізу і порівняння електронних документів. Запропоновано методологію автоматизованої побудови логіко-лінгвістичних моделей текстової інформації як основного засобу екстракції знань з речень природної мови. Розкрито основні принципи порівняльного аналізу логіко-лінгвістичних моделей речень природної мови. Розроблено концепцію автоматизованого лінгвістичного аналізу електронних текстів.

Усі наведені результати належать автору особисто.

Для студентів та науковців, які займаються аналітичною обробкою електронних документів, спеціалізуються у галузі інформаційних технологій, комп'ютерної та структурної лінгвістики.

ISBN 978-617-7092-79-6

© А.І.Вавіленкова, 2016

ЗМІСТ

ВСТУП	5
I. ЗАСОБИ АНАЛІТИЧНОЇ ОБРОБКИ ТЕКСТОВОЇ ІНФОРМАЦІЇ	9
1.1 Проблема обробки електронних текстів.....	9
1.2 Формальні засоби представлення текстової інформації.....	18
1.3 Технології попередньої обробки електронних текстів.....	32
II. ФОРМАЛІЗАЦІЯ ТЕКСТОВОЇ ІНФОРМАЦІЇ ЗА ДОПОМОГОЮ ЛОГІКО-ЛІНГВІСТИЧНИХ МОДЕЛЕЙ	45
2.1 Семантико-синтаксична структура речень.....	45
2.2 Принципи побудови логіко-лінгвістичних моделей текстової інформації.....	50
2.3 Формалізовані правила визначення структурних компонентів логіко-лінгвістичних моделей.....	64
III. МЕТОДОЛОГІЯ АВТОМАТИЗОВАНОЇ ПОБУДОВИ ЛОГІКО-ЛІНГВІСТИЧНИХ МОДЕЛЕЙ ТЕКСТОВОЇ ІНФОРМАЦІЇ	79
3.1 Загальний алгоритм методу автоматизованого формування логіко-лінгвістичних моделей текстової інформації.....	82
3.2 Ідентифікація вхідного тексту.....	86
3.3 Концептуалізація, визначення характеристик елементів формальної системи.....	88
3.4 Синтаксичний аналізатор, визначення ролей.....	95
3.5 Формалізація процесу побудови логіко-лінгвістичних моделей.....	103
IV. ПРИНЦИПИ ПОРІВНЯННЯ ЛОГІКО-ЛІНГВІСТИЧНИХ МОДЕЛЕЙ РЕЧЕНЬ ПРИРОДНОЇ МОВИ	117
4.1 Протиріччя і тавтології у реченнях природної мови.....	118
4.2 Вплив функціонально-істинних операцій на зміст речень природної мови.....	123
4.3 Особливості методу резолюцій при пошуку збігів у реченнях природної мови.....	129

4.4 Алгоритм порівняння логіко-лінгвістичних моделей речень природної мови.....135

V. АВТОМАТИЗОВАНИЙ ЛІНГВІСТИЧНИЙ АНАЛІЗ ЕЛЕКТРОННИХ ТЕКСТІВ.....144

5.1 Характеристики семантичного рівня.....144

5.2 Когезія як основний засіб формування семантичних зв'язків у тексті.....150

5.3 Теоретичні основи лінгвістичного аналізу тексту.....163

5.4 Концепція автоматизованого лінгвістичного аналізу тексту.....175

СПИСОК ЛІТЕРАТУРНИХ ДЖЕРЕЛ.....182

*Для будь-якої істини існує лише одна
міль свята між нескінченістю, коли
її вважають хибною, та
нескінченістю, коли її вважають
тривіальною.*

Ж.А. Пуанкаре

ВСТУП

Розвиток індустрії систем електронного документообігу супроводжується зростанням масивів повнотекстових документів і вимагає нових засобів обробки електронних текстів. Системи, що автоматизують обробку електронних документів з метою виявлення їх змісту та структури, системи документообігу та машинного перекладу, інформаційно-пошукові системи, а також різноманітні інструментальні засоби, на зразок текстових процесорів, відзначаються складністю об'єкту моделювання та слабкою теоретичною базою.

Останнім часом аналіз тексту привертає все більшу увагу в областях безпеки, комерції, науки. Сучасний рівень розвитку апаратних та програмних засобів дає можливість постійно працювати з базами даних оперативної інформації на різних рівнях управління. Тому інтелектуальні інформаційні технології використовуються для розширення кола задач, які вирішуються за допомогою комп'ютерів у слабо структурованих областях, а також для підвищення рівня інтелектуальної інформаційної підтримки.

Постійне зростання кількості електронних документів спричиняє дублювання матеріалів в інформаційних мережах. У зв'язку з цим, у всіх сферах соціального середовища виникає проблема появи ідентичних за змістом копій документів. Так, в освіті – це плагіат курсових, дипломних та дисертаційних робіт, у політичній сфері – створення суперечливих законопроектів, в інформаційних технологіях – відсутність можливості здійснення коректної пошукової оптимізації, в економіці – створення однотипних проектів та ін. Усе це загрожує дезінформацією, виникненням логічних протиріч і неточністю прийняття рішень.

Тому сьогодні існує суспільна потреба у розробці ефективних лінгвістичних технологій, на яких базуватимуться технології

оперування знаннями. Зокрема, не вирішеною залишається задача створення сучасних засобів автоматичної екстракції знань з електронних текстів, а також проблема порівняльного аналізу електронних документів за змістом.

Сучасні автоматизовані системи, які аналізують семантичну структуру тексту на рівні вище за речення, знаходяться на експериментальній стадії. Їх основний недолік – те, що вони працюють зі змістом окремих слів та словосполучень, рідше – із зв'язками між ключовими словами, але структура речення і тексту в цілому не аналізується.

Дослідження та розробки таких великих комерційних компаній, як Microsoft, Google, Apple, Abby та IBM використовують технології аналізу текстів з метою подальшої автоматизації процесів аналізу та вилучення даних. На сьогодні найбільш результативною технологією роботи зі знаннями вважається Data Mining, що об'єднує у собі широкий математичний інструментарій та останні досягнення у сфері інформаційних технологій. В основу Data Mining покладена концепція шаблонів, що відображають фрагменти багатоаспектних відношень в даних. Такі шаблони представляють собою закономірності, пошук яких обмежений певними наборами розподілу значень показників, що аналізуються. Темі обробки текстової інформації присвячено чимало праць вітчизняних та зарубіжних вчених у сфері інформаційних технологій (Ланде Д.В., Маккарти Д., Питтса У., Ньюелла А., Люгера Д.Ф., Рубашкіна В.Ш., Хопкрофта Дж., Бродера А., Винограда Т.В., Поспелова Д.А., Попова Е.В., Осуга С., Лакоффа Дж.) [3,6,17,19,31] та комп'ютерної лінгвістики (Широкова В.А., Леонтьєва Н.Н., Шемакіна Ю.І., Гальперіна І.Р., Лайонза Дж., Кобозевої І.М., Мельчука І.А., Лайонза Д., Кронгауза М.А., Греймаса А.Ж., Фостера Д.М.) [7,45, 52, 75]. Проте питання про створення єдиної методики автоматичного аналізу текстових документів досі залишається відкритим.

Отже, вирішення проблем аналітичної обробки текстової інформації на даний момент неможливе без застосування нових алгоритмів та методів змістовної формалізації електронних текстів. Адже, лінгвістичний аналіз будь-якої текстової інформації передбачає встановлення семантичних зв'язків між окремими її частинами, тобто екстракцію знань. Це неможливо

без здійснення семантичного аналізу мінімальних комунікативних одиниць мови – речень, що повинні бути цілісними і передавати інформацію в усій складності залежностей і зв'язків.

Метою даної монографії є викладення результатів, отриманих автором в процесі дослідження перерахованих вище проблем обробки електронних текстів. До таких результатів можна віднести:

- уніфіковану форму логіко-лінгвістичної моделі речення природної мови як основного засобу формалізації текстової інформації;

- методологію автоматизованої побудови логіко-лінгвістичних моделей текстової інформації як основного засобу вилучення знань з речень природної мови;

- основні принципи порівняльного аналізу речень природної мови на базі використання логіко-лінгвістичних моделей;

- алгоритм порівняння логіко-лінгвістичних моделей речень природної мови;

- принципи формування змістовних моделей електронних текстових документів;

- концепція лінгвістичного аналізу електронних текстів;

- алгоритм автоматизованого лінгвістичного аналізу електронних документів, що базується на використанні морфологічного, синтаксичного та семантичного аналізу, формалізації зв'язків між структурними одиницями текстів.

Основний зміст книги висвітлено у п'яти розділах.

Перший розділ присвячено розкриттю основних проблем обробки текстової інформації. Проаналізовано існуючі засоби формалізації текстової інформації, продемонстровано їх недосконалість щодо обробки природно мовних текстів.

У другому розділі запропоновано уніфіковану форму логіко-лінгвістичної моделі як основного засобу формалізації текстової інформації. Викладено принципи побудови таких моделей, описано правила створення їх структурних компонентів. Формалізовані правила утворено як інтерпретацію відношення прилягання, узгодження та керування у словосполученнях природної мови.

У третьому розділі монографії представлена методологія автоматизованої побудови логіко-лінгвістичних моделей

текстової інформації. Описано метод автоматизованого формування логіко-лінгвістичних моделей, запропоновано правила побудови таких моделей для різних типів речень. Формалізовано неявні закономірності побудови логічних зв'язків у природно мовних текстах.

У четвертому розділі описано алгоритм порівняння логіко-лінгвістичних моделей речень природної мови, що стало можливим завдяки дослідженню структури речень, а також врахуванню основних елементів когезії.

П'ятий розділ акумулює у собі результати досліджень чотирьох попередніх розділів. Тут запропоновано концепцію автоматизованого лінгвістичного аналізу електронних текстів. Вона базується на побудові змістовної моделі тексту, дослідженні логічних зв'язків між окремими частинами текстів та виявленні засобів забезпечення континууму тексту.

Монографія буде корисною для студентів та науковців, які займаються аналітичною обробкою електронних документів, спеціалізуються у галузі інформаційних технологій, комп'ютерної та структурної лінгвістики.

Написання цієї монографії було б не можливим без підтримки та терпіння моїх близьких.

Окремо висловлюю величезну подяку моєму науковому консультанту, доктору технічних наук, професору, завідувачу кафедри Комп'ютеризованих систем управління НН ІКІТ Національного авіаційного університету Литвиненко Олександровичу Євгенійовичу за підхід дослідження текстової інформації шляхом логіко-лінгвістичного моделювання, а також за ідеї формування окремих форм логіко-лінгвістичних моделей.

Окрема подяка моїй доньці Марисі за витримку при навчанні за методикою Глена Домана, що дало можливість зробити висновки про те, як людина мислить, сприймає та вчиться формулювати речення природної мови.

I. ЗАСОБИ АНАЛІТИЧНОЇ ОБРОБКИ ТЕКСТОВОЇ ІНФОРМАЦІЇ

1.1 Проблема обробки електронних текстів

Стрімке поширення інформаційно-комунікаційних технологій, економічне зростання, нова економічна та суспільна парадигми, перехід від інформаційного суспільства до суспільства знань – все це спричинило появу величезної кількості інформаційних потоків, внаслідок чого людина у багатьох випадках уже не в змозі самотійно опрацювати всю необхідну для її діяльності інформацію [1]. У зв'язку з цим широкого застосування набули комп'ютерні системи обробки електронних документів.

Сьогодні основним способом зберігання та передачі інформації служать електронні тексти на природній мові. Створення великих сховищ даних, розвиток мережових комунікацій та індустрії електронного документообігу, що супроводжується зростанням масивів повнотекстових документів, робить актуальними задачі екстракції знань, формування коректно побудованих текстових документів і порівняння електронних текстів за змістом, вимагає нових засобів організації доступу до інформації, релевантної конкретному запиту користувача.

Одна з причин автоматизації обробки текстової інформації – час, необхідний для написання програм. Його можна було б значно скоротити, якби обчислювальні машини були здатні розуміти прості висловлювання на природній мові [2].

Ще однією областю, що потребує лінгвістичного аналізу даних, є автоматичний машинний переклад. Основний недолік сучасних перекладачів – відсутність змістовного аналізу речень природної мови, через що перекладені слова формують неправильні синтаксичні конструкції.

Отже, комп'ютер не може вирішувати поставлені задачі обробки текстової інформації повністю, оскільки поки що не створено адекватних формалізованих моделей природно мовних об'єктів, а розв'язання відповідних завдань містить неформальні, творчі елементи, властиві лише людині [3].

Задачі обробки природно мовних текстів, незважаючи на півстолітню історію досліджень у цій галузі штучного інтелекту

(це роботи таких радянських, вітчизняних та зарубіжних вчених, як Попов Е.В., Поспелов Д.А., Мельников Г.П., Гладкий А.В., Рубашкін В.Л., Башмаков А.І., Кобозева І.М., Большакова Є.І., Широков В.А., Грязнухіна Т.О., Звягінцев В.А., Карнап Р., Вольф Є.М., Т.Виноград, Осуга С., Уено Х., Ісідзука М., Хомський Н., Дж. Лакофф, Греймас А.Ж., Фостер Д.М., Кронгауз М.А., Лайонз Дж.) [4], розвиток інформаційних технологій та суміжних дисциплін, задовільного розв'язання більшості практичних задач аналітичної обробки текстової інформації поки що не існує [5]. Основною задачею, що вирішується сучасними системами обробки електронних текстів, є формування інформаційного портрету тексту в термінах ключових понять, виявлення змістовних зв'язків між поняттями, автоматичне реферування. На початку досліджень аналітики прагнуть до повноти запиту, а не до його точності, об'єм релевантних відповідей складає сотні або тисячі одиниць. Флективні мови погано піддаються опису формалізмами, тому подальше дослідження проблеми відбувається уже після отримання підібраних документів за допомогою кластерних, семантичних карт та інших методів.

Існуючі засоби автоматичної обробки тексту можуть виділяти частини мови і відмічати граматичні зв'язки, наприклад, дієслівні групи, суб'єкти та об'єкти, зв'язки займенників та фраз, які їх описують [6]. Більш глибока лінгвістична обробка ґрунтується на розв'язанні проблеми усунення лінгвістичних неоднозначностей, що зустрічаються в текстах. Подібні інструменти розглядаються як компоненти загальної системи розуміння природної мови. З їх допомогою тексти конвертуються в джерела інформації, доступні для опрацювання комп'ютером, що відкриває можливість для подальшої машинної обробки.

Отже, суспільна потреба у розробці ефективних лінгвістичних технологій, на яких базуватимуться технології оперування знаннями, вимагає створення універсальної системи підтримки лінгвістичних досліджень та розробок [7]. З точки зору конструктивної семантики (науки, що займається вивченням і формуванням законів синтезу знакових систем, які забезпечують процес відображення об'єктивного світу в його моделях [8]), формалізація передбачає перехід до оперування символами, при якому не потрібно додаткового аналізу речей об'єктивного світу і вся теорія розвивається у знаковій області, без звернення до

досвіду, емпірії. Мова розробника моделі орієнтована на взаємодію конструктивної семантики з інженером по знаннях у процесі формування моделі.

В абсолютній більшості, саме на користувача перекладається екстракція з отриманих документів корисної для нього інформації, тобто вилучення необхідних знань. Для вирішення цієї задачі потрібні інструменти, що ґрунтуються на принципах розуміння природної мови, а це, у свою чергу, вимагає розроблення методів та засобів глибокої формалізації природно мовних структур.

Зокрема, не вирішена проблема порівняльного аналізу електронних текстів, яка виникає щоразу, коли з'являється потреба у визначенні збігів або виявленні логічних протиріч у текстових документах. З проблемою визначення збігів зіштовхуються у тих сферах діяльності, де кінцевим результатом є текстовий документ. Це, в першу чергу, освіта, наука, законотворчість, патентування, інноваційна та інша діяльність, пов'язана із захистом інтелектуальної власності.

Друга проблема – виявлення логічних протиріч у текстових документах – лежить, головним чином, у площині професійних інтересів різних юридичних та інформаційно-аналітичних організацій і підрозділів. Особливу актуальність останнім часом вона набула у зв'язку з перспективою вступу України до Європейського співтовариства, що вимагає гармонізації українського законодавства та забезпечення його адекватності відносно загальноєвропейських нормативних актів [4].

Системи, що автоматизують обробку електронних документів з метою виявлення їх змісту та структури, системи документообігу, інформаційно-пошукові системи і системи машинного перекладу, а також різноманітні інструментальні засоби типу текстових процесорів відзначаються складністю об'єкту моделювання та слабкою теоретичною базою.

Основний недолік подібних систем – вони працюють зі змістом окремих слів та словосполучень, рідше – із зв'язками між ключовими словами, але структура речення в цілому не аналізується. Сучасні автоматизовані системи, які аналізують семантичну структуру тексту на рівні вище за речення, знаходяться на експериментальній стадії. На сьогодні жодна з теорій не може претендувати на повноту, хоча найбільш досконалі з них досягли задовільних теоретичних результатів [9].

Існує багато програм, які здійснюють лінгвістичну обробку тексту, проте жодна з них не використовує в роботі механізми вилучення змісту із текстової інформації. У табл.1.1 перераховані найвідоміші з таких програм.

Таблиця 1.1

Програми аналізу та лінгвістичної обробки текстів

№	Назва системи	Опис, задачі
1.	Cognitive Dwarf	Програма для аналізу та лінгвістичної обробки тексту. Програмний пакет містить синтаксичний аналізатор для російської та англійської мови і систему автоматичного перекладу.
2.	AOT	Програма функціонує для російської та англійської мови, здійснює графематичний (виділяє слова, цифрові комплекси, формули), морфологічний (будує морфологічну інтерпретацію слів вхідного тексту), синтаксичний (будує дерево залежностей всього речення) та семантичний (будує семантичний граф тексту) аналіз.
3.	Core Language Engine	Використовує для представлення знань квазі-логічні форми, знання тут представляються у вигляді набору форм, що слабо залежать від контексту. Система використовується для перекладу текстів, для управління базами даних, інтерпретації пошукових запитів на природній мові. Дозволяє обмежити область можливих інтерпретацій синтаксичної структури речення, відсікаючи варіанти, що не мають квазі-логічної форми для перетворення.
4.	Link Grammar Parser	Синтаксичний парсер для англійської, російської, арабської та перської мови. В результаті розбору речення система визначає його синтаксичну структуру, що складається з набору помічених зв'язків, які об'єднують пари слів.
5.	Cibola/Oleada	Проекти Cibola/Oleada реалізують системи лінгвістичного аналізу текстів, включають засоби роботи з мультимовними текстами,

№	Назва системи	Опис, задачі
		здійснюють статистичний аналіз та автоматичний переклад.
6.	Лингвоанализатор	Програма математичного аналізу структури тексту, результатом є визначення близькості довільного тексту, що вводиться користувачем до одного з авторських еталонів, які визначені завчасно, видаючи на виході імена трьох письменників, що могли б бути найбільш вірогідними авторами.
7.	Text Analyst	Програма дозволяє побудувати семантичну мережу понять з посиланнями на контекст, здійснює змістовний пошук фрагментів тексту з урахуванням прихованих у тексті зв'язків за словами запиту. Текст аналізується шляхом побудови ієрархічного дерева тем.
8.	Langsoft	Програмне забезпечення для обробки природної мови, здійснює граматичний розбір речень, перевірку орфографії та граматики, логічний вивід, аудіо- та відео-переклад речень.
9.	Shingles Expert Pro, Compare It!, IsEqual, Text.ru	Відкриті програми для порівняльного аналізу електронних документів, що використовують статистичні методи, проте не аналізують тексти за змістом.
10.	AskNet	Програма містить інструментарій розробника, що реалізує повний лінгвістичний аналіз текстів на російській та англійській мовах. Модулі лінгвістичного аналізу включають в себе морфологію, синтаксис та семантику.
11.	Quintura Searchcrystal	Метапошукові системи, що представляють результати пошуку у графічному вигляді. Результати кластеризуються за статичними критеріями. Для побудови візуального кластера використовується морфологічний аналіз, синтаксичний та семантичний аналіз не реалізовано.

№	Назва системи	Опис, задачі
12.	Арион SyTech	Система забезпечує автоматизацію аналітичної обробки однотипної текстової інформації та представлення звітів у різноманітному вигляді. Однак, вона не забезпечує високі показники з достовірності виявлення фактографічної інформації та наочності їх представлення у зв'язку з відсутністю реалізації синтаксичного та семантичного аналізу текстів. Програма вимагає великого об'єму ручної праці для налаштування системи на нові прикладні області.
13.	Vivisimo Nigma	Метапошукові системи з кластеризацією результатів пошуку забезпечують можливість виділення слів, що часто зустрічаються зі словами пошукового запиту. Системи здійснюють графематичний та морфологічний аналіз.
14.	Google Mini, Google Desktop	Програмно-апаратний комплекс для пошуку інформації на комп'ютерах корпоративної локальної мережі, реалізує пошук за ключовими словами з урахуванням морфології російської та англійської мови.
15.	Яндекс.Server	Програма спеціалізованого HTTP- сервера, що дозволяє індексувати сайти. Пошук працює з урахуванням морфології російської, англійської та української мов. Результатом пошуку є список документів, упорядкованих за релевантністю та датою. Синтаксичний та семантичний аналізу текстів не реалізовано.
16.	Система «Рефератор»	Призначена для обробки текстів на природній мови, засоби автоматичного реферування дозволяють розбити текст на семантично цілісні фрагменти, які відображають основні теми документу, та виділити найбільш інформативні. За допомогою даної системи можна створювати реферати текстів та проводити їх індексацію. У системі застосовуються оригінальні евристичні алгоритми

№	Назва системи	Опис, задачі
		морфологічного аналізу, який надає можливість визначати граматичні характеристики нових слів. Підтримка системи каталогізації надає можливість зберігати результати роботи в базі даних.
17.	Система «VitaminE»	Система за допомогою семантичного аналізу змістовного контексту речення обирає з можливих альтернатив коректний варіант перекладу, чим поліпшує якість тексту машинного перекладу.
18.	Oracle Text	Програмний комплекс, інтегрований в СУБД, дозволяє ефективно працювати з запитамі, що відносяться до неструктурованих текстів, дозволяє здійснювати пошук, класифікацію та кластеризацію документів, вилучати ключові поняття, робити автоматичне анотування, шукати в документах асоціативні зв'язки.
19.	ADVEGO	Здійснює семантичний аналіз тексту шляхом розрахунку відношення не важливих слів у документі до загальної кількості слів, порівняльний аналіз текстів з використанням методу «шинглів».
20.	IBM Watson	Суперкомп'ютер, основна задача якого розуміти питання, сформульовані на природній мові та знаходити на них відповіді у базі даних.
21.	IBM Intelligent Miner for Text	Представляє собою набір окремих утиліт, що запускаються не залежно одна від одної. Так, утиліта Language Identification Tool – автоматично визначає мову, на якій написано документ; Categorization Tool - автоматично відносить текст до певної категорії; Clusterization Tool - розбиває велику множину документів на групи в залежності від близькості стилю, форми або частотних характеристик; Feature Extraction Tool – виявляє у документів ключові слова на основі аналізу наперед заданого словника

№	Назва системи	Опис, задачі
22.	Технологія ABBY Y Compreno	Система розуміння, аналізу та перекладу текстів, що розробляється компанією ABBY (2012-2013 рр.). Програма здійснює письмовий та усний переклад з однієї мови на іншу, робить інтелектуальний пошук, лексичний, морфологічний та семантико-синтаксичний аналіз, здійснює синтез текстів (із внутрішнього представлення), масштабний лексико-семантичний опис природної мови, класифікацію та фільтрацію документів, захист від несанкціонованого використання інформації, автоматичне реферування та анотування документів.

Як видно з табл.1.1, системи обробки природної мови поділяються на декілька категорій:

- системи аналізу та лінгвістичної обробки текстів;
- програми перетворення текстів;
- програми-генератори: системи створення онтологій, електронних корпусів;
- системи автоматичного перекладу та вилучення фактів із текстів;
- психолінгвістичні програми;
- системи обробки природної мови (NL-Processing): аналізатори (морфологічні, синтаксичні, семантичні), системи автореферування та анотування;
- системи автоматичного розпізнавання та синтезу мови;
- словники та тезауруси;
- системи порівняльного аналізу електронних текстів.

Системи аналітичної обробки текстової інформації застосовуються у найрізноманітніших галузях людської діяльності. Так, сьогодні у роздрібній торгівлі відслідковують детальну інформацію про кожну покупку, використовуючи кредитні картки та комп'ютеризовані системи контролю. Проте для збільшення продажів, покращення реклами та напрацювання стратегії створення запасів товарів у цій сфері потрібні системи, що аналізуватимуть споживчий кошик та виявлятимуть товари, які купляють разом. Створення прогнозуючих моделей дає можливість торгівельним підприємствам дізнаватися характер

потреб різних категорій клієнтів з певною поведінкою [10]. Аналогічно у банківській справі важливо мати системи аналізу транзакцій для виявлення крадіжок, здійснювати сегментацію клієнтів для цілеспрямованої та результативної роботи, будувати прогнози зміни клієнтури. У сфері бізнесу необхідні системи прогнозування популярності характеристик тих чи інших продуктів (машин, літаків, засобів телекомунікації, побутової техніки та ін.), системи, що відслідковуватимуть політику гарантій та постійних клієнтів. Все це не можливо без розробки нових методів глибокої інтелектуальної обробки даних.

Використання технологій обробки текстової інформації є перспективним напрямком для розв'язання задач оцінки відповідності програмного забезпечення вимогам безпеки. Це дозволить підвищити ефективність проведення, скоротити час та вартість сертифікаційних досліджень, покращить процедуру сертифікації [11].

Основною проблемою систем машинного перекладу типу Prompt, Lingvo, Stylus, Pragma є низька якість автоматично переведених текстів. Це результат того, що системи автоматичного перекладу природно мовних текстів обробляють тексти на рівні морфологічного та синтаксичного аналізу, не маючи можливості працювати зі змістом – на семантичному та прагматичному рівнях. У випадках, коли системи не можуть визначити однозначно правильний переклад, видається список альтернатив можливого перекладу вхідного слова. Функції реферування та індексування використовуються для побудови загального реферату за ключовими словами, розгортання та просування web-сайтів. Система фільтрації та перехоплення Internet-повідомлень із застосуванням лінгвістичних методів аналізу дасть можливість створювати спеціалізовані засоби мережових та протокольних аналізаторів, програми моніторингу мереж, засоби збору трафіка, системи виявлення мережових атак, мережеві сканери, систем захисту та аналізу інформації, якою обмінюються користувачі мереж з метою збору статистичних даних, розподілу доступу і т.д. [12].

Здійснення аналітичної обробки текстової інформації необхідне в медицині для постановки діагнозів. Такі системи будуються в основному за правилами, що описують сукупність симптомів різних захворювань. Сформовані правила дозволяють

автоматично визначити засоби медикаментозного лікування, виявляти протипоказання та передбачити результат курсу лікування [11].

Тобто, майже всі області знань, науки і техніки через складність своєї системної організації потребують інтелектуальної обробки даних. Зрозуміло, що для вирішення проблеми коректної обробки електронних документів, необхідне об'єднання зусиль спеціалістів не лише в області інформаційних технологій, але й лінгвістики, філології, філософії, штучного інтелекту, програмування та математичного моделювання. Такими синтетичними дисциплінами є комп'ютерна лінгвістика, структурна і комп'ютерна семантика та ін. Із розвитком цих дисциплін відбувається перенесення центру тяжіння з машинного представлення процедур до машинного представлення знань. Проблема представлення знань вимагає створення моделей, що базуються на більш глибокому розумінні людської пам'яті. Велика увага приділяється розробці програм автономних об'єктів, що рухаються, здатних в режимі реального часу створювати сценарії та прогнози виходу з різноманітних конфліктних ситуацій. Частковим випадком цієї проблеми є задача редукції та синтезу програм на основі специфікації початкових і кінцевих умов. А сильна метафоричність мов є одним з головних факторів складності формалізації [13].

Отже, обробка електронних документів потребує сьогодні потужних механізмів формалізації та глибинного лінгвістичного аналізу текстової інформації, які б забезпечили повний та коректний перехід від природної мови до математичних моделей з метою застосування до них відомих алгоритмів математичної логіки.

1.2 Формальні засоби представлення текстової інформації

Першими формальними засобами, які найбільш досліджені теоретично та мають широке застосування на практиці, є контекстно-вільні граматики Хомського. Фактично, формальні граматики представляють собою правила, що дозволяють строго та експліцитно описувати деякі граматичні закономірності природної мови [14].

Граматика G – це четвірка $\langle N, T, S, P \rangle$, де N – множина не термінальних символів або змінних; T – множина термінальних символів, що не перетинаються з N , фактично, це множина вхідних символів, а об'єднання $V = N \cup T$ – словник граматики G ; S – виділений символ у множині не термінальних, початковий; P – кінцева множина продукцій або правил типу $x \rightarrow y$, де x та y – ланцюжки з V [15].

Мова $L(G) = L$, **визначена граматикою** G – множина ланцюжків термінальних символів, що виводяться з S .

Якщо до граматики G додати продукцію $AB \rightarrow BA$, то нова граматика G' породжує якісно іншу мову, оскільки змінює порядок символів у ланцюжку без розгортання не термінального символу. Ця продукція називається **трансформаційною**, а граматики, що містять такі продукції, – **трансформаційними граматиками**.

Формальні граматики дають природний спосіб впорядкування мов за допомогою обмежень, які можна ввести в продукції. За Хомським розрізняють чотири типи формальних граматик [16].

Тип 0 – не обмежені граматики або граматики з фазовою структурою без обмежень, породжуються граматикою, у якій всі продукції мають вигляд $x \rightarrow y$, де x та y – ланцюжки з V . Через складність цей тип граматики не знайшов практичного застосування.

Тип 1 – контекстно-залежні, породжуються граматикою, у якій всі продукції мають вигляд:

- $xAu \rightarrow xwu$, де w , x та y – ланцюжки з V , A – змінна (не термінальна), тобто ланцюжок w можна вивести з A , якщо A з'являється в контексті x та y ;

- $x \rightarrow y$, де x та y – ланцюжки з V , а число $|y|$ символів в y не менше числа символів в x .

Ці класи граматик еквівалентні, можуть використовуватися при аналізі текстів на природній мові, однак при побудові компіляторів практично не використовуються в силу своєї складності.

Тип 2 – контекстно-вільні граматики, породжуються граматиною, в якій всі продукції мають вигляд $A \rightarrow x$, де A – не термінальна змінна, x – ланцюжок з V . Тобто граMATика допускає появу в лівій частині лише не термінального символу.

Контекстно-вільні граматики широко застосовуються для опису синтаксису комп'ютерних мов.

Тип 3 – регулярні граматики – найпростіші з формальних грамаTIK, породжуються граматиною, продукції якої мають вигляд: $A \rightarrow aB$ або $A \rightarrow a$, де a – термінальний символ, B – не термінальна змінна.

Регулярні граматики застосовуються для опису простіших конструкцій: ідентифікаторів, рядків, констант, а також мов асемблера, командних процесорів.

Можна простежити, що всі мови типу 3 є мовами типу 2, мови типу 2 – мовами типу 1, мови типу 1 – мовами типу 0 [17].

Контекстно-вільні граматики, породжуючи різноманітні правильні синтаксичні конструкції, за потужністю дуже великі для аналізу синтаксису і в той же час недостатні для визначення змісту фрази на природній мові. У зв'язку з цим з'явилися граматики для аналізу текстів, які є модифікацією або розширенням контекстно-вільних грамаTIK, що дозволяє враховувати семантичні ознаки. Проте універсальність семантичних ознак, їх мовна орієнтація та недостатній рівень розробки роблять даний підхід важким для використання. Спроба змодельовати розуміння людиною семантично зв'язаних текстів приводить до питання про семантичні структури та рівні, на яких описується значення слів.

Спеціалісти в області семантики виділяють два рівні представлення висловлювань [18].

Поверхневий – поверхневі структури відображають будову речення у формі узгодження грамаTIчних одиниць природної мови (наприклад, дієслова, іменники і т.д.), це передбачає виокремлення елементарних ситуацій та формування великої кількості шаблонів, що вимагає тонкої диференціації семантичних ознак у предметній області. Р. Шенк запропонував підхід, який базується на представленні ідентичних речень, виражених різними поверхневими структурами, єдиною концептуальною конструкцією.

Семантичний або глибинно-синтаксичний – глибинні структури, що виражають семантичні відношення між змістом дієслова та змістом тієї чи іншої з його іменних груп, незалежні від мови, універсальні, так як відображають логічну, а не граматичну будову речень природної мови. Тобто глибинний рівень – це опис текстів за допомогою предикатно-аргументних структур. Мову для запису глибинної структури та деякі правила перекладу глибинних структур в поверхневі запропонував Ч. Філмор, встановлюючи одиниці глибинних структур або семантичні ролі аргументів.

Така обробка текстів базується на спрощеній моделі мови і дозволяє ігнорувати складність синтаксису. Система семантичних ролей, запропонована Ч. Філмором, та модель контекстно-вільних граматик, запропонована Р. Шенком, сильно залежать від контексту та прив'язані до дієслова [17].

Отже, традиційні засоби аналізу за ключовими параметрами та еталонами відповідей, а також універсальні методи контролю знань не дозволяють повністю аналізувати природно мовні тексти і не сприяють змістовому аналізу тексту. Виникнення понять автоматної мови, регулярної множини, ланцюга Маркова було пов'язане з статистичними дослідженнями послідовності голосних та приголосних в літературних текстах. Введене Хомським поняття формальної граматики виникло в результаті спроби створення чітких описів граматичних закономірностей природної мови. Сьогодні основна мета дослідження природної мови – побудова таких описів мови, які б могли бути повністю формалізовані за допомогою певної алгебраїчної конструкції. Окрім важливого теоретичного значення, ці роботи мають практичну цінність для автоматизації аналізу та синтезу текстів за допомогою комп'ютерних технологій.

Знання – це ознака об'єму інформації, що визначає її статус та відокремлює від всієї іншої інформації за критерієм здатності рішення поставленої задачі [18].

Знання – сукупність даних, фактів та правил виводу про світ, що включають в себе інформацію про властивості об'єктів, закономірності процесів, явищ, а також правил, що використовуються цією інформацією для прийняття рішень [14]. Головна відмінність знань від даних полягає в їх активності, тобто

появі в базі нових фактів або встановлення нових зв'язків, що може стати джерелом змін у прийнятті рішень.

Обробка знань у комп'ютері представляє собою обробку їх змісту правилами перетворення таких форм, якими описуються знання в машині. Відповідно, при обробці знань найбільш фундаментальною та важливою проблемою є перш за все опис змісту проблем різного діапазону, а також наявність такої форми опису знань, яка гарантує, що обробка їх змісту формальними правилами перетворення буде здійснюватися правильно. Ця проблема називається проблемою представлення знань, а самі знання описуються формалізмами [19].

Однією із найважливіших проблем, з якими зіштовхуються сьогодні при обробці текстової інформації, є представлення знань. Це пов'язане з тим, що результат представлення знань визначає характеристики систем. При правильному виборі способу представлення знань можна уникнути ускладнення системи. Однак вибір оптимального способу представлення знань залежить від характеру та складності задач, що вирішуються.

Представлення знань – один із основних напрямків робіт в області штучного інтелекту, когнітології (представлення знань пов'язане з тим, як людина зберігає та обробляє інформацію) та інформатики (основна ціль – підбір представлення конкретних та узагальнених знань, відомостей і фактів для накопичення та осмислення обробки інформації за допомогою комп'ютера). В штучному інтелекті мета представлення знань – навчитися обробляти інформацію, досягти аналогу людського мислення, інтелекту [2].

Предметна область – це частина реального світу, що розглядається в межах даного контексту (області дослідження, що є об'єктом деякої діяльності). Під предметною областю розуміють також множину всіх предметів, властивості яких та співвідношення між якими розглядаються у науковій теорії [20].

Однорідне представлення приводить до спрощення механізму управління логічним виводом та спрощення управління знаннями. Показником інтелектуальності системи, з точки зору представлення знань, вважається здатність системи використовувати у потрібний момент необхідні (релевантні) знання. У проблемі доступу до знань можна виділити три аспекти [21].

1) *Зв'язність (агрегація) знань* є основним способом, що забезпечує прискорення пошуку релевантних знань. Знання організовують навколо найбільш важливих об'єктів (сутностей) предметної області.

2) *Механізм доступу до знань* – деякий опис сутності, призначений для відшукування в базі знань об'єктів, що відповідають цьому опису. Очевидно, що впорядкування та структурування знань можуть значно пришвидшити процес пошуку.

3) *Способи співставлення*: синтаксичні, параметричні, семантичні та примусові. У випадку синтаксичного співставлення співвідносяться форми (зразки), а не зміст об'єктів. Успішним вважається співставлення, у результаті якого зразки ідентичні. Результат синтаксичного співставлення є бінарним. У параметричному співставленні вводиться параметр, що визначає ступінь співставлення. У випадку семантичного співставлення співвідносяться не зразки об'єктів, а їх функції.

Розвиток комп'ютерної семантики тісно пов'язаний з двома формами представлення знань – процедурною та декларативною, відносною вагою процедурних та декларативних знань у відображенні предметної області, їх взаємозв'язками та роллю в процесі обробки інформації [10].

Алгоритмічні моделі представлення знань є замкнутими моделями, в рамках яких розглядається повне знання про проблемну область. Проте на практиці моделі представлення знань повинні бути орієнтованими на принципову неповноту даних.

Процес внутрішньої інтерпретації даних та знань шляхом встановлення внутрішніх та зовнішніх структурних зв'язків називається *структуризацією*. Зовнішня структура зв'язків відображає організацію даних та знань або відношення між класами однотипних понять. У комп'ютерній семантиці структури даних розуміються як декларативні знання, що несуть лише функцію відображення проблемної області. Над певними структурами даних здійснюється впорядкована послідовність операцій – програма, процес, що реалізує певний алгоритм. Результатом роботи програми завжди є нове декларативне знання. А програма представляє собою процедурне знання.

Семантично традиційна обробка інформації орієнтована на процес – на задану послідовність обробки даних з метою отримання бажаного результату, а її розвиток, у значній мірі, пов'язаний із структуризацією даних. Подальший розвиток структур даних співвідноситься з дослідженнями в області штучного інтелекту, які привели до появи спеціальних структур – моделей представлення знань.

Модель представлення знань – це формалізм, призначений для відображення статичних та динамічних властивостей предметної області [22].

Знання як форма представлення інформації в комп'ютері володіє рядом особливостей:

1) **внутрішня інтерпретованість** – кожна інформаційна одиниця повинна мати унікальне ім'я, за яким інформаційна система знаходить її, а також відповідає на запити, в яких це ім'я згадувалося;

2) **структурованість** – інформаційні одиниці повинні володіти гнучкою структурою, для них повинна виконуватися рекурсивна вкладеність одних інформаційних одиниць в інші;

3) **зв'язність** – в інформаційній базі між інформаційними одиницями повинна бути передбачена можливість встановлення зв'язків різного типу;

4) **семантична метрика** – відношення, яке характеризує ситуаційну близькість інформаційних одиниць, тобто силу асоціативного зв'язку між інформаційними одиницями;

5) **активність** – всі процеси, що протікають у комп'ютері, ініціюються командами, а дані використовуються цими командами лише у випадку необхідності.

Усі вище перераховані особливості інформаційних одиниць визначають ту грань, за якою дані перетворюються у знання, а бази даних переростають у бази знань. В області штучного інтелекту розв'язок задач можна спростити правильним вибором методу представлення знань. Не всі методи представлення знань однаково ефективні для конкретної області. Неправильний метод представлення уповільнює обробку. Тобто не існує такого способу представлення знань, який можна використовувати для всіх задач або зробити всі задачі однаково простими. Саме тому важливим кроком при проектуванні системи обробки текстової інформації є правильний вибір моделі представлення знань.

Розрізняють п'ять основних моделей представлення знань.

Семантична мережа представляє собою інформаційну модель предметної області (сукупність фактів і тверджень із бази даних) і має вигляд графа, вершини якого відповідають об'єктам (поняттям) предметної області, а дуги – відношенням між ними. Відношення мають для семантичних мереж виключне значення, так як надають базову структуру для організації знань [23,24,25]. Важливою технікою, що використовується у семантичних мережах, є ієрархія або система класифікації, на основі якої об'єкти, що відносяться до проблемної області, класифікуються на деяку кількість категорій або класів на основі їх загальних властивостей [26].

Використаємо семантичну мережу для представлення знань, що несе в собі наступне речення: «Якщо на підсистему здійснюється керуючий вплив, то таку систему називають керованою» (рис. 1.1).

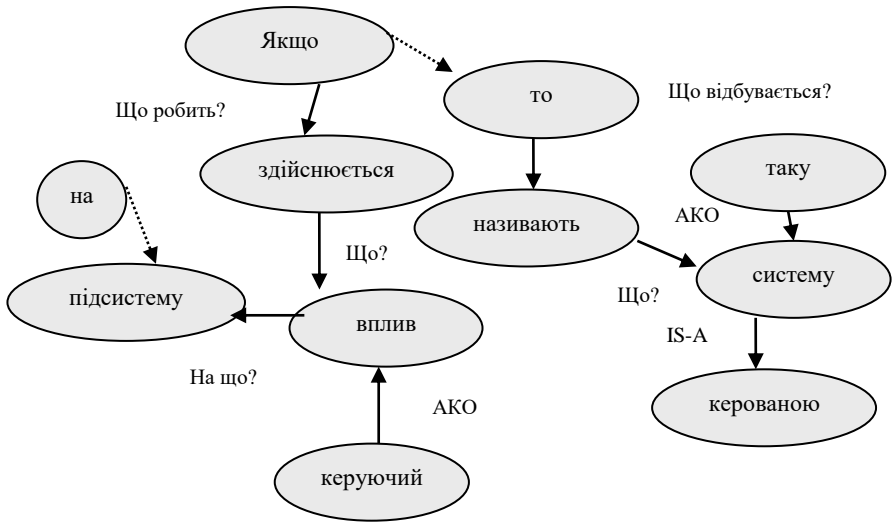


Рис. 1.1. Екстенціональна семантична мережа речення природної мови

Тут вершини семантичної мережі – це слова-поняття, а дуги – це логічні зв'язки, за допомогою яких слова зв'язані у реченні.

З рис. 1.1 видно, що збільшення кількості слів у реченні підвищує нагромадження зв'язків у семантичній мережі. Крім того, різноманіття запитань, які можна ставити до членів речення, не дозволяє обмежитися трьома характерними для семантичної мережі типами зв'язків. Також неможливо поставити запитання до та від службових частин мови. Із цього зрозуміло, що речення природної мови не можна представити у вигляді семантичної мережі з типовими зв'язками [27]. Отже, семантична мережа придатна для семантичної обробки на основі змістовних зв'язків між прототипами та застосовується для представлення категоричних знань [23].

Фреймові структури. В основі теорії фреймів лежить сприйняття фактів засобами співставлення отриманої інформації з конкретними елементами та значеннями, рамками, визначеними для кожного концептуального об'єкта [28]. На верхньому рівні фрейму представлена фіксована інформація: факт, що стосується стану об'єкта, який вважається істинним. На наступних рівнях фрейму представлена множина термінальних слотів (терміналів), які заповнені значеннями і даними. У кожному слоті задається умова, що виконується при встановленні відповідності між значеннями. Ця властивість фреймів часто застосовується для вирішення логічних задач (рис. 1.2) [29].

Опишемо ситуацію «Якщо на підсистему здійснюється керуючий вплив, то таку систему називають керованою» у вигляді фрейму:

«Фрейм-зразок: ПІДСИСТЕМА<що робить?, на що?, що?, який?>;

Фрейм-зразок: СИСТЕМА<що роблять?, що?, яку?, якою?>;

Фрейм-екземпляр: ПІДСИСТЕМА<здійснюється, підсистему, вплив, керуючий>;

Фрейм-екземпляр: СИСТЕМА<називають, систему, таку, керованою>.»

З прикладу видно, що теорію фреймів при описі природно мовного речення можна скоріше використати для постановки задачі. Фрейм може надавати знання з вузьких тематик, більшість яких задають завчасно, також у фреймових структурах допускається необмежена модифікація або знищення слотів [30].

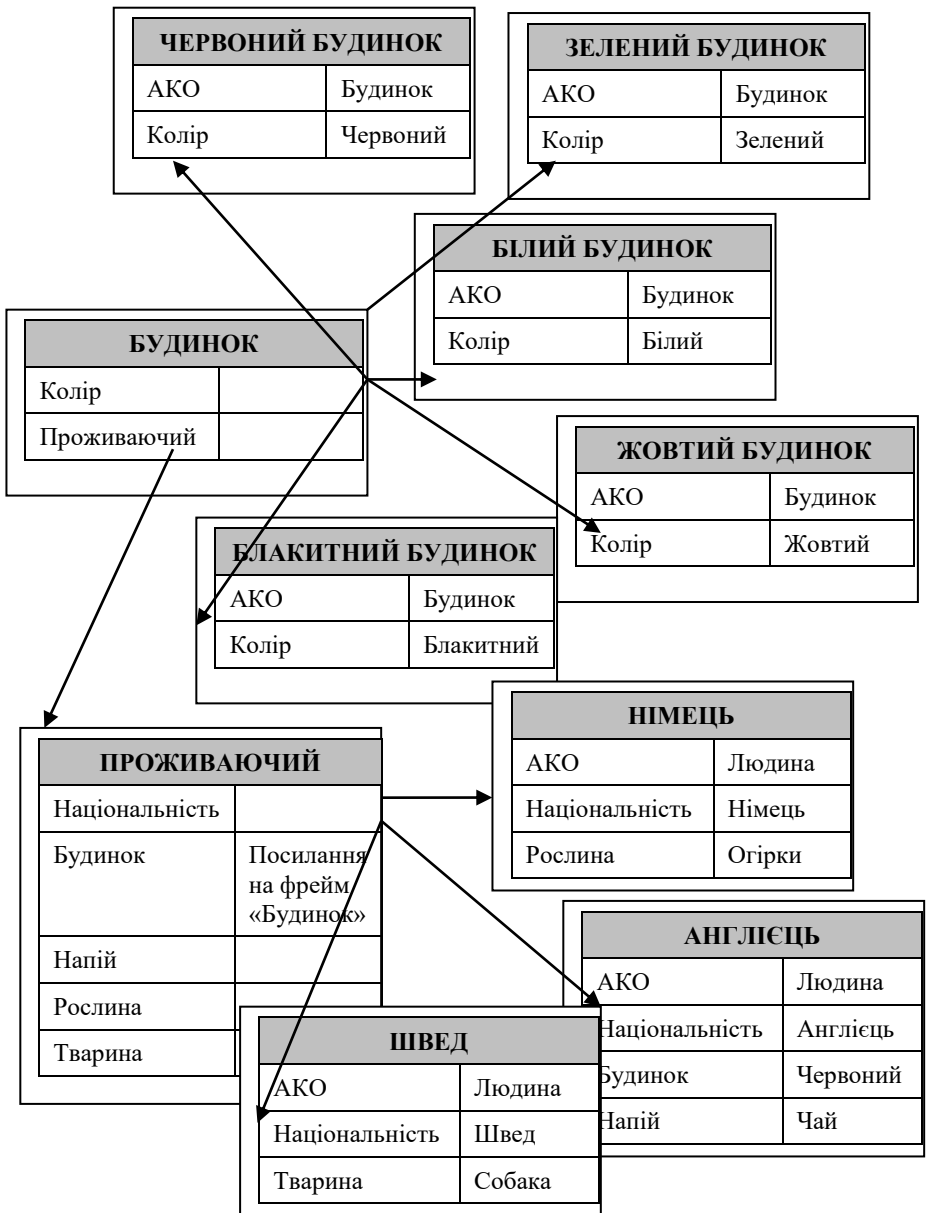


Рис. 1.2. Приклад застосування фреймової структури для вирішення логічних задач

Продукційна модель базується на правилах, дозволяє представити знання у вигляді структури типу «Якщо..., то». Таке представлення дозволяє при формуванні баз знань впорядкувати опис об'єктів, зберігаючи їх ієрархію. Якщо до деяких впорядкованих об'єктів у процесі логічного виводу застосовувати правила, то можна організувати звернення окремо до об'єкта, окремо до атрибуту та окремо до значення [31].

Якщо у пам'яті системи зберігається певний набір продукцій, то вони утворюють систему продукцій. У такій системі повинні бути задані спеціальні процедури управління продукціями, за допомогою яких відбувається їх актуалізація та вибір для виконання однієї з них із числа актуалізованих [32]. У робочій пам'яті системи, що базується на продукційних моделях, зберігаються пари атрибут-значення, істинність яких встановлена в процесі вирішення конкретної задачі на поточний момент. Це відбувається по мірі спрацювання правил.

Для даного речення природної мови *«Якщо на підсистему здійснюється керуючий вплив, то таку систему називають керованою»* продукційна модель буде мати такий вигляд:

«Посилання: *на підсистему здійснюється керуючий вплив*

Результат: *таку систему називають керованою».*

У даній моделі зміст речення не враховується, має значення тільки форма і зовнішнє представлення. Тобто початковий рядок (посилання) – набір символів, що представляють вхідні дані, згідно з продукційним правилом, закладеним у експертній системі, перетворюється на вихідний рядок символів (результат).

Застосування продукційної моделі для представлення речення природної мови має ряд недоліків: по-перше, не виокремлює з даного речення знання; по-друге, отриманий результат залежить лише від наперед заданих правил, і якщо система не знайде відповідного посилання, то результат не буде отримано; по-третє, відсутність стратегії управління, тобто хаотичне розміщення правил у системі, уповільнює роботу під час пошуку необхідного посилання. Застосування продукційної моделі не для складнопідрядного умовного речення, а, наприклад, для складносурядного чи простого, неможливе, так як структура цих речень не дозволяє сформулювати посилання і результат [27].

Логічна модель. Основна ідея при побудові логічних моделей знань полягає у тому, що вся інформація, необхідна для вирішення прикладних задач, розглядається як сукупність фактів та тверджень, що представляються як формули деякої логіки. Знання відображаються сукупністю формул, а отримання нових знань зводиться до реалізації процедур логічного виводу. Для представлення знання в математичній логіці використовуються логічні формалізми – обчислення предикатів, що містить формальну семантику та операційну підтримку, так як для нього вже розроблені механізми виводу [33].

Опис предметних областей, виконаний на логічній мові, називають *логічними (формальними) моделями*. Знання, які можуть бути представлені за допомогою логіки предикатів, є фактами, які описуються логічними формулами. Для представлення будь-якої області у вигляді логічних формул, перш за все, необхідно обрати константи, які визначають об'єкти даної області, а також функціональні та предикатні символи, що визначають відповідну функціональну залежність та відношення між об'єктами [14]. Основною задачею, що вирішується у рамках логіки предикатів, є визначення істинності або хибності заданої формули на деякій області інтерпретації [34].

Логіко-лінгвістична модель представляє собою вираз або вирази у вигляді термінів природної мови та лінгвістичних змінних (які можуть бути не лише числами, але й словами і словосполученнями природної або штучної мови). Основою для логіко-лінгвістичних моделей служить логіка предикатів, де предикат – це функція P , яка приймає значення 0 або 1 , і аргументи якої пробігають значення із довільної множини M : $P(x_1, x_2, \dots, x_n)$ [35].

Для заданого речення «Якщо на підсистему здійснюється керуючий вплив, то таку систему називають керованою» логіко-лінгвістична модель матиме вигляд:

$$\forall x \forall y \forall z [ЗДІЙСН(x, y) \rightarrow НАЗИВ(x, z)],$$

де x – підсистема, система;

y – керуючий вплив;

z – керована;

ЗДІЙСН – предикат, що означає можливість здійснення дії;

НАЗИВ – предикат, що означає процес формулювання назви [24].

Логіко-лінгвістична модель – це засіб відображення відповідності між поняттями, що описуються термінами природної мови, та предикатними формулами. Недоліком логіко-лінгвістичних моделей є те, що знання важко структурувати, проте для кінцевого користувача ця модель є найбільш зрозумілою при розв’язанні задачі представлення знань складного речення [24].

Онтологія. Онтологія – це детальний опис деякої проблемної області, що використовується для формального та декларативного визначення її концептуалізації. На формальному рівні онтологія – це система, яка містить концепти (поняття, класи), властивості концептів (атрибути, ролі), відношення між ними та обмеження, які визначаються аксіомами. Аналіз онтологій дає можливість зробити висновки про доцільність їх застосування для автоматичного вилучення знань з текстової інформації, пошуку знань в Інтернет, інтелектуального анотування, створення інтелектуальних систем у різних галузях. Основну частину знань аналітики отримують у результаті порівняння, аналізу та синтезу інформації з різних фактів, розміщених у текстах. Таким чином, при роботі з великими потоками документів процес автоматичної структуризації текстової інформації заміняє експертний процес виділення фактографічної інформації та об’єктів, що виконується вручну.

Побудова онтологій передбачає виконання наступних задач [36]:

- визначення основних понять та їх взаємозв’язків у обраній проблемній області;
- створення точних, не суперечливих визначень для кожного поняття та відношення;
- створення та документування словника термінів;
- опис правил та обмежень, згідно з якими на базі введеної термінології формуються істинні твердження, що описують стани системи;
- розбиття сукупності основних термінів, що використовують в онтології, на класи;
- вибір та розробка спеціальної мови для представлення знань;

- формування концептуалізації в рамках обраної мови представлення знань [37]. Фактично, онтологія є одним із різновидів семантичної мережі.

Основні задачі, що вирішуються за допомогою онтологій [38]: створення та використання баз знань, організація ефективного пошуку в базах даних та знань, створення систем, що реалізують механізми розмірковувань, організація пошуку за змістом у текстовій інформації, семантичний пошук в Інтернет. Онтологія є своєрідним узагальненим описом інформації, що стосується певної предметної області.

Побудуємо онтологію для заданого речення «Якщо на підсистему здійснюється керуючий вплив, то таку систему називають керованою» (рис. 1.3). Із рис.1.3 видно, що дослідження довільного речення природної мови за допомогою онтологій потребує наявності вичерпних знань про кожен суб'єкт та його дію, що вимагає наявності великих і повних баз даних.

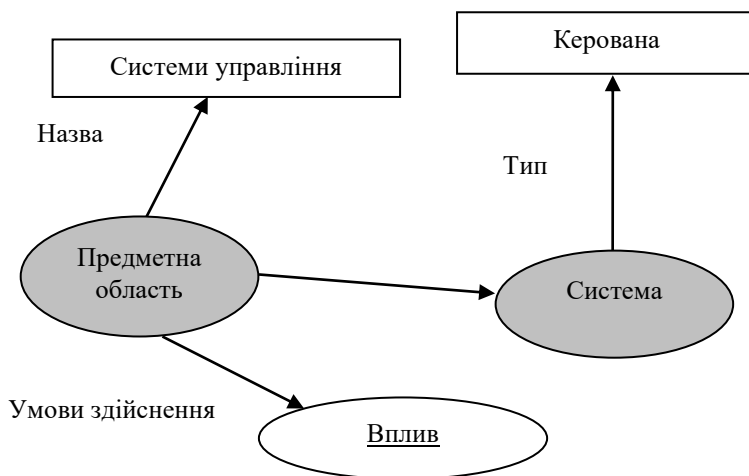


Рис. 1.3. Орієнтовна схема побудови онтології

Недоліком онтологій є те, що існує лише один стандарт формування онтологій, весь їх аналіз базується на теорії, рівень практичного застосування онтологій дуже низький. Фактично, онтологія представляє собою моделювання певної предметної області за допомогою різних мов програмування.

1.3 Технології попередньої обробки електронних текстів

Засоби сучасних ЕОМ, що використовуються для обробки електронних текстів, дозволяють задавати різні обмеження на шукані комбінації слів у тексті, визначаючи обов'язковість або необов'язковість присутності тих чи інших слів, допустиму відстань між ними та порядок їх знаходження в тексті, а також дозволяючи проводити аналіз слова у всіх граматичних формах. Це дає можливість формулювати висновки, точно і повно описуючи можливі способи представлення змісту тексту.

Для підвищення точності аналізу текстів розробляються методи, що базуються на їх попередньому лінгвістичному розборі. Для ефективного практичного застосування такі методи вимагають збереження отриманих описів граматичної або семантичної структури у вигляді спеціальних індексів, які потім повинні використовуватися для порівняння із структурою самого речення, що отримується лінгвістичним аналізатором.

Автоматизоване вилучення знань із тексту є однією з основних задач штучного інтелекту і безпосередньо пов'язане із розумінням текстів на природній мові.

Сучасні технології попередньої обробки електронних текстів передбачають багаторівневе представлення природної мови. Загально прийнята схема аналізу тексту на природній мові зображена на рис.1.4.

Компоненти, що складають мовну модель, – лінгвістичні процесори, які один за одним обробляють вихідний текст. Вхід одного процесора є виходом іншого.

Завданням *графематичного аналізу* є поділ вихідного тексту на слова, виокремлення речень з вихідного тексту. На етапі графематичного аналізу формується таблиця усіх слів тексту з вказівкою порядкового номера речення, в якому зустрічається слово, і порядковим номером слова у вказаному реченні [39,40].

Графематичний аналіз є початковим етапом обробки тексту, в ході якого визначаються елементи граматичної структури (слова, знаки пунктуації, числа, скорочення тощо). Для цього слід визначити поняття графематичного аналізу з погляду теорії формальних мов.

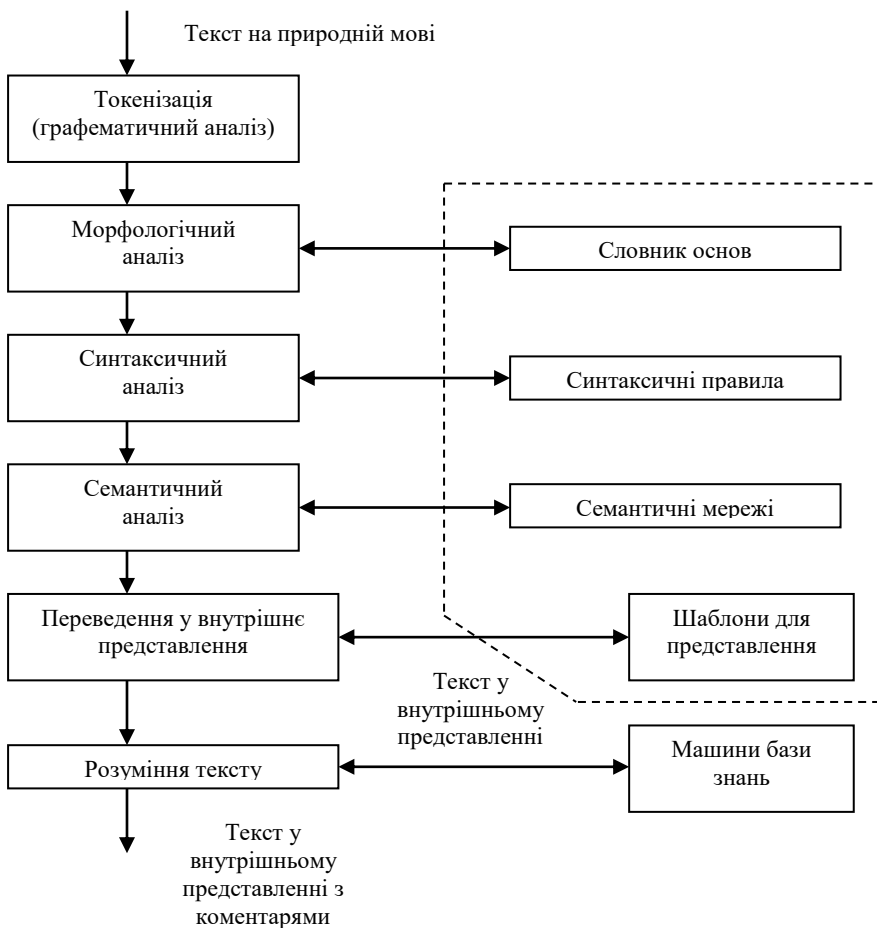


Рис. 1.4. Схема аналізу тексту

На етапі **морфологічного аналізу** в словах тексту виділяються незмінні частини (основи), і словам ставиться у відповідність ряд граматичних характеристик (частина мови, рід, число, відмінок і т.д.) [41,42].

Морфологічний розбір – це визначення характеристик слова як частини мови з урахуванням особливостей його контекстуального використання [43].

Загальна схема морфологічного аналізу включає в себе:

- визначення категоріального значення, початкової форми;

- визначення морфологічної характеристики слова: лексико-граматичні розряди або семантико-функціональні групування, конкретизуються семантичні ознаки та показники; морфологічні категорії (категоріальна семантика); словозмінний тип (особливості парадигми);

- визначення морфологічних характеристик словоформ: словозмінні морфологічні категорії, показники форми.

Після того, як був здійснений морфологічний аналіз тексту, може бути проведений так званий пост морфологічний аналіз. Він дозволяє зняти частину граматичної омонімії, що залишилася після автоматичного морфологічного аналізу, при цьому враховується найближчий контекст слова, повторні вживання слів у тексті та ін. Граматична омонімія, що не здійснена в ході пост морфологічного аналізу, може бути знята за допомогою подальшого автоматичного синтаксичного аналізу.

Головна особливість всіх систем, що здійснюють автоматичний синтаксичний розбір, полягає у побудові елементарних змістовних дерев речень. До цих дерев можна застосувати операції приєднання та підстановки.

Синтаксичний розбір речення відбувається шляхом набору послідовних перетворень:

- пошук граматичних ідіом;
- лексико-граматичний аналіз речення з усуненням неоднозначності у визначенні частин мови;
- знаходження іменної групи об'єкта і суб'єкта;
- знаходження дієслівної групи;
- виділення головних та залежних речень.

Граматики, які використовуються для опису мов програмування, більш прості, ніж граматики, потрібні для природних мов.

Проблеми комп'ютерної лінгвістики, які стосуються, перш за все, граматичного розбору тексту природною мовою спонукають на створення якісного синтаксичного аналізатора, що спрямований на ефективне розв'язання задачі пошуку інформації. Складність практичної реалізації можливого аналізатора обумовлена наявністю тісного зв'язку між синтаксисом та семантикою. Для розв'язання подібних задач, що називається синтаксичною омонімією, необхідно створення спеціального

тлумачно-комбінаторного словника, який буде включати у себе синтаксичну та семантичну інформацію про узгодженість слів.

Синтаксичний аналіз (парсинг) – це процес співставлення лінійної послідовності лексем мови з його формальною граматику. Результатом зазвичай є дерево розбору. Синтаксичний парсинг здійснюється паралельно з лексичним аналізом. При парсингу початковий текст перетворюється в структуру даних, що відображає синтаксичну структуру вхідної послідовності з метою подальшої обробки.

Тобто, формально, метою синтаксичного розбору є побудова дерева залежностей між словами у фразі. У разі успіху речення згортається в повно зв'язне дерево з єдиною кореневою вершиною. Оскільки одна словоформа може відповідати декільком граматичним формам слова, зокрема для різних слів, в процесі аналізу необхідно проводити згортання речення для всіх можливих варіантів. Ті ж з них, які призводять до максимального згортання фрази (із мінімальним числом висячих вершин), пропонуються вважати найбільш достовірними при розборі речення [44].

Синтаксичний розбір можна здійснювати шляхом заповнення внутрішньої частини трикутника деревом виводу (рис. 1.5).

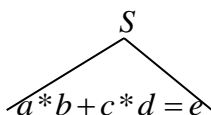


Рис. 1.5. Трикутник залежностей

Структура речення задається однією із контекстно-вільних грамастик. Так, граматика включає в себе певні правила, що задаються завчасно, наприклад [45]:

$S \rightarrow NP VP$;

$NP \rightarrow \text{іменник}$;

$VP \rightarrow \text{дієслово } NP$;

$\text{іменник} \rightarrow \text{директор договір}$;

$VP \rightarrow \text{підписав}$;

Речення: «Директор підписав договір».

Вважається, що запропоноване речення породжене перерахованими вище правилами підстановки даної граматики. У результаті підстановки отримуємо набір слів природної мови (рис. 1.6):

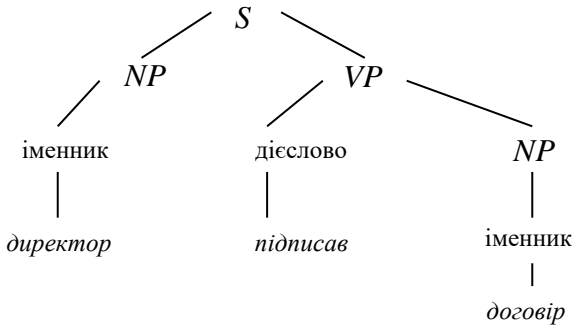


Рис. 1.6. Дерево залежності між словами у фразі

Проте синтаксичний аналіз речень не дає відповіді на те, чи відповідає речення змістовому навантаженню, яке хотів у нього вкласти автор. Синтаксична правильність речення виключає його перефразування через неправильну форму.

Семантичний аналіз – аналіз тексту, що полягає в остаточному формуванні внутрішнього формалізованого представлення тексту шляхом співставлення знань з імітаційної моделі з фактами, які знаходяться безпосередньо у тексті. Тобто все відображається семантичним графом тексту. Семантичний аналіз проводиться різними методами для різних мов та для різних підмножин одного і того ж тексту.

Семантичний аналіз тексту базується на результатах синтаксичного аналізу, отримуючи на вході уже не набір слів, розбитих на речення, а набір дерев, що відображають синтаксичну структуру кожного речення. Так як методи синтаксичного аналізу поки що не дають бажаних результатів, рішення цілого ряду задач семантичної обробки тексту базується на результатах аналізу окремих слів, і, замість синтаксичної структури речення, аналізуються набори слів, які стоять поряд.

Загальною базою методів семантичного аналізу, яка дозволяє виявити семантичні відношення між словами, є тезаурус мови. На математичному рівні він представляє собою орієнтований граф,

вузлами якого є слова у їх основній словоформі, а дуги задають відношення між словами і можуть відображати ряд ознак. Таким чином, тезаурус задає набір бінарних відношень на множині слів природної мови.

Фактично, семантичний аналіз повинен пояснити, як речення природної мови розуміються, інтерпретуються та співвідносяться зі станами, процесами та об'єктами зовнішнього світу. Щоб зрозуміти значення речення та його семантичні відношення з іншими виразами, необхідно знати не лише значення його лексичних елементів, але й те, як вони співвідносять один з одним. Це, у свою чергу, залежить від синтаксичної структури речення. Синтаксичні властивості, релевантні для семантики, включають досить абстрактні граматичні відношення [46].

Метою семантичного аналізу є визначення для кожного слова та фрази в цілому змістовних характеристик. Методи семантичного пошуку базуються на чотирьох механізмах, пов'язаних із застосуванням:

- словники синонімів;
- ієрархії понять;
- бази лінгвістичних правил для граматичного аналізу тексту;
- спеціальні семантичні мережі.

Семантична інтерпретація не обов'язково повинна здійснюватися лише після завершення синтаксичного аналізу. Вона може розпочатися раніше на основі неповної інформації про поверхневі синтаксичні структури. Тоді під час синтаксичного аналізу може використовуватися інформація семантичного та прагматичного рівня [48].

Отримане на етапі семантичного аналізу представлення речення називається *семантичним графом речення*. Його можна інтерпретувати як експліцитне задання змісту вхідного речення. Семантичний граф речення, як і знання системи, виражається за допомогою мови розширених семантичних мереж. Побудову такого графу можна розділити на дві частини: побудова неструктурованого семантичного графу речення та структурування семантичного графу.

Побудова неструктурованого семантичного графу речення. Даний етап полягає у виборі із узагальненої синтаксичної структури такої її частини, яка не суперечить абстрактним знанням системи; побудові для нього неструктурованого

семантичного графу та виділенні квантифікованих слів. Тобто на даному етапі відбувається обробка слів, які мають модель управління, та обробка іменних груп, тобто слів, що не мають моделей управління, так як вносять у речення різне семантичне навантаження і потребують застосування різноманітних методів. Семантика слів з моделлю управління визначається не контекстом, а із словника. Інформація з абстрактної мережі використовується лише для підтвердження правильності вибору, зробленого семантичним аналізом. Семантичні відношення, що пов'язують слова в іменній групі, визначаються в основному контекстом. Фактично, необхідно привести у відповідність число і зміст описів, що характеризують вхідне речення, та описи, що відображають деякий факт у базі даних (рис. 1.7, рис. 1.8) [47].

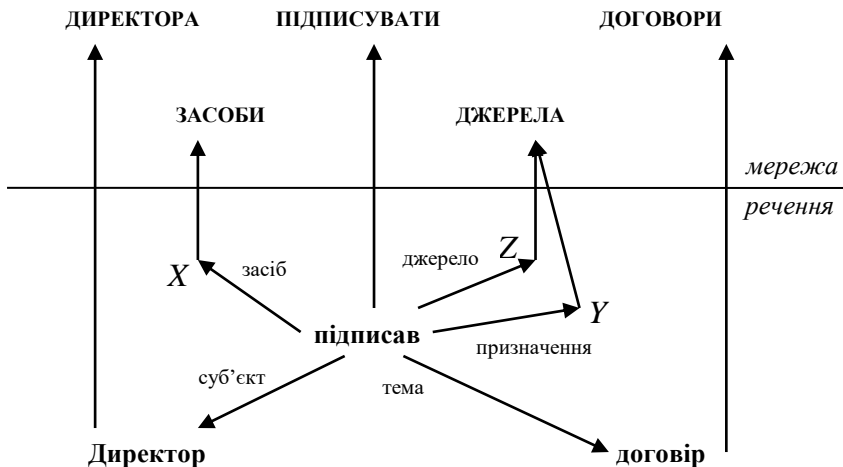


Рис. 1.7. Приклад семантичного представлення речення

При обробці іменних груп виконуються наступні дії:

- за абстрактною семантичною мережею виявляються характеристики, не вказані у реченні безпосередньо;
- обирається омонім поняття за виділеним фрагментом семантичної мережі;

- синтаксичні відношення замінюються на семантичні шляхом переводу іменної групи у внутрішнє представлення, ідентичне семантичній мережі (рис. 1.9).

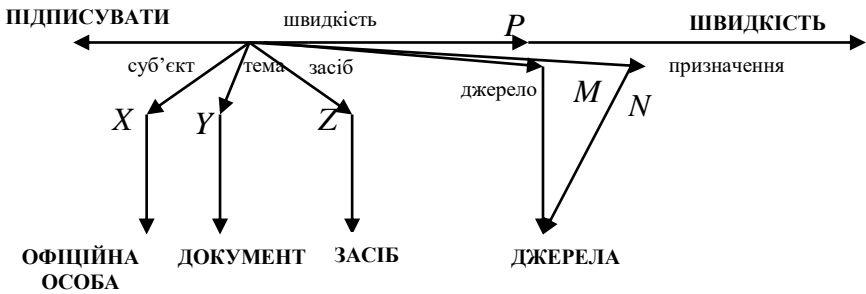


Рис. 1.8. Представлення правила-обмеження для предиката

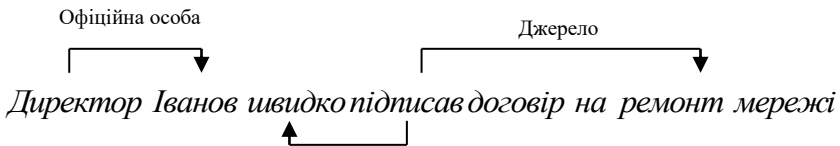


Рис. 1.9. Розбір іменної групи

Структурування семантичного графу. На даному кроці відбувається структурування семантичного графу, отриманого на попередньому етапі. Структурування полягає у введенні до семантичного графу предикатів більш високих порядків, визначення їх вкладеності та областей дії. Більшість цих предикатів можуть розглядатися як квантори.

Отже, задача квантифікації полягає у визначенні для речення типів кванторів, порядку їх розгляду та вкладеності областей дії. Тип кванторів визначається тільки кванторними словами. Порядок, у якому розглядаються квантори, визначається постійною вагою, що приписується квантору, та синтаксичною структурою речення. Визначення області дії кванторів важко здійснити у процесі обробки іменних груп до семантичної обробки всього речення, так як вплив кванторів, що знаходяться

всередині деякої іменної групи, зазвичай не обмежується областю такої групи.

Нехай сформовано неструктурований семантичний граф (рис. 1.10) для запитального речення «Чи швидко підписано договір на ремонт мережі Івановим?»

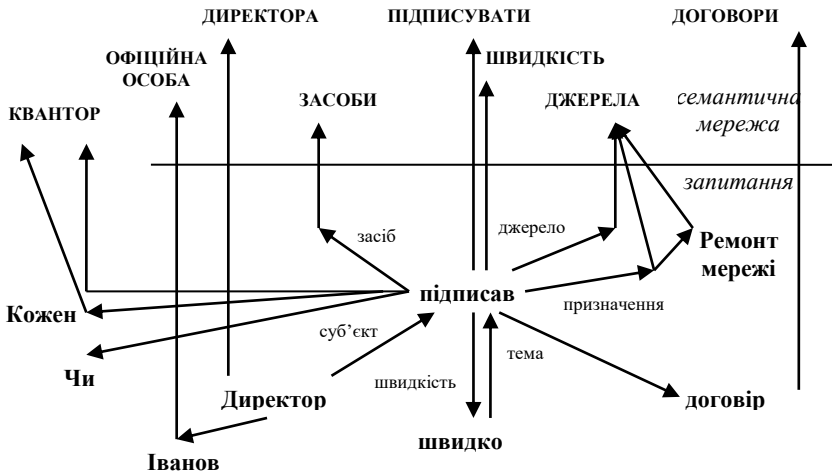


Рис. 1.10. Представлення неструктурованого семантичного графу

Запит має форму «ТАК–НІ». Так, для довільного y , що є елементом множини ШВИДКІСТЬ, наприклад, «швидко» за допомогою структурованого семантичного графу можна визначити його відношення до договору на ремонт мережі, тобто:

$$(\forall y)((y \in \text{ШВИДКО}) \wedge (\text{Договір}(y, \text{ремонт})) \rightarrow$$

$$\text{Підписано}(\text{Договір}, \text{Івановим}, y)).$$

Квантифікований семантичний граф представлено на рис. 1.11.

Порядок обробки кванторів включає у себе наступні етапи:

- обробка конструкції (іменна група, предикат), яка має квантор із більшою вагою;
- за наявності декількох кванторів у іменній групі першим обробляється зовнішній квантор;

- якщо вага кванторів однакова, у іменних групах першим обробляється квантор більш лівий, ближчий до початку.

Отже, на етапі семантичного аналізу формується семантична структура вхідного речення, яка описує змістовні зв'язки між словами.

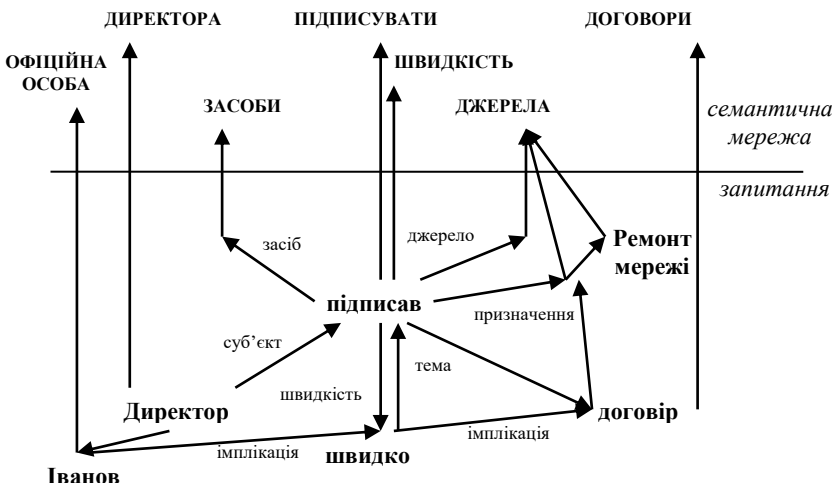


Рис. 1.11. Результат квантифікації

У відповідність глибинній структурі довільного речення природної мови ставиться її семантична інтерпретація [49]. Людина, яка говорить, розуміє зміст довільного речення із довільної множини речень, виконуючи операцію об'єднання змісту слів у зміст словосполучень та речень. Саме цю операцію – побудову змісту складного цілого із змістів його складових частин – повинні здійснювати правила семантичної компоненти.

Семантичний аналіз передбачає здійснення процедур, направлених на автоматичну змістовну обробку вихідного тексту та створення на його основі нових лінгвістичних об'єктів. **Семантичний аналіз** – це алгоритм, що дозволяє представити семантичну (змістову) структуру речення та тексту у вигляді строгої формальної системи, шляхом аналітичного дослідження взаємозв'язків між окремими об'єктами та подіями з предметної області. Семантичну структуру тексту можна описати за допомогою семантичних зв'язків та семантичних відношень.

Семантичний вузол – будь-який вузол семантичного графу. Як і на всіх етапах аналізу природно мовних текстів, семантичні вузли утворюються зі слів вихідного речення. Головним джерелом гіпотез про склад семантичного вузла є синтаксичний аналіз. Багато синтаксичних груп можуть перейти до семантичних вузлів, інші повинні перетворитися на атрибути вузлів. Також джерелами можуть бути тезауруси.

У семантичного відношення можуть бути наступні параметри:

- посилання на словникову статтю, звідки було взято це відношення;
- перелік слів, які є лексичними реалізаціями цього відношення у вихідному тексті;
- синтаксичне відношення;
- умовна довжина.

Семантичний підхід до вилучення знань з природно мовного тексту спирається на лінгвістичну модель. Відповідно до цієї моделі основу семантичної структури висловлювання представляє пропозиційна компонента плану змісту. Ця компонента відображає мовну ситуацію, що описується реченням. Вона характеризує його об'єктивний зміст, на відміну від інших компонент (модальної, комунікативної), які так чи інакше характеризують або відношення того, хто говорить, або співвіднесеність ситуації з деяким моментом часу, умовами її реалізації, і тому відносяться до сфери суб'єктивного змісту.

Таким чином, семантичний підхід до вилучення знань передбачає виділення зі структури фрази її **семантичного ядра** – об'єктивного опису ситуації і абстрагування від несуттєвих, суб'єктивних компонентів плану змісту. Результати семантичного аналізу тексту відкривають широкі можливості для використання виявлених знань в інформаційно-пошукових системах, наприклад, для пошуку в тексті описів ситуацій заданого класу і виявлення їх учасників [37].

Додатково семантичний аналіз тексту дозволяє сформувати змістовний портрет документа, виділяючи його ключові поняття і ранжуючи їх за значущістю у документі. Застосування засобів синтаксичного синтезу, трансформацій відношень і тезаурусів забезпечує перетворення різних семантико-синтаксичних конструкцій до загального уніфікованого вигляду, що дозволяє

ототожнювати однакові елементи змісту, інтерпретовані у вигляді різних виразів. Ця інформація, разом із статистикою вживання та іншими чинниками, дозволяє виділити ключові елементи тексту для порівняння документів при пошуку і класифікації.

Семантична інтерпретація глибинної структури речення – це повний набір пар, що складаються з вузла глибинної структури заданого речення та максимальної множини тлумачень, що притаманні ланцюжку слів, які помічені даним вузлом. Набір пар вважається повним, якщо кожному вузлу глибинної структури поставлена у відповідність максимальна множина тлумачень.

Для подальшого аналізу алгоритмів семантичної структуризації необхідно знати властивості основних елементів семантичної компоненти тексту.

Множина C **аномальна**, тоді і тільки тоді, коли множина тлумачень, приписаних C , – пуста.

Множина C **семантично однозначна** тоді і тільки тоді, коли множина тлумачень, приписаних C містить лише один елемент.

Множина C **семантично n -значна** тоді і тільки тоді, коли множина тлумачень, приписаних C містить n елементів для довільного $n > 1$.

Множини C_1 та C_2 **частково синонімічні** тоді і тільки тоді, коли множина тлумачень, що приписуються C_1 , та множина тлумачень, що приписуються C_2 , мають хоча б один спільний елемент [50].

Через те, що не існує єдиного принципу здійснення семантичної структуризації електронних документів, а кожна нова програма удосконалює попередні, використовуючи нові методики та працюючи, в основному, з конкретною предметною областю, алгоритми семантичної структуризації розглядаються у розрізі систем, у яких вони використовуються.

Окремі речення і ситуації, з яких необхідно вилучити знання для подальшої обробки системою, повинні бути подані на вхід у певному впорядкованому вигляді. Більшість інтелектуальних систем працюють із заздалегідь заданими шаблонами або зразками, а на основі введеного слова (або декількох слів) виводять наперед заданий шаблон з підстановкою у нього конкретних слів.

Взагалі, зародження семантичної теорії пов'язане із зауваженнями Шеннона про те, що зміст повідомлень не має ніякого відношення до теорії інформації. Карнап та Бар-Хіллел дали визначення змісту, що міститься в деякому реченні S . Вони визначили $cont(S)$ – **семантичний зміст речення**, як [50]:

$$cont(S) = p(\sim S) = 1 - p(S), \quad (1.1)$$

де $p(S)$ – логічна ймовірність істинності речення S .

Згодом також була запропонована аналогічна функція:

$$inf(S) = -\log_2 p(S) \text{ або } inf(S) = -\log_2 \frac{1}{1 - cont(S)}. \quad (1.2)$$

Двом приведеним визначенням (1.1) – (1.2) семантичного змісту речення властиві аналогічні обмеження:

- якщо вираз $S_1 > S_2$ логічно істинний, то для речень S_1 та S_2 справедлива нерівність $I(S_1) \leq I(S_2)$;
- функція $I(S) \geq 0$;
- якщо речення логічно істинне, то $I(S) = 0$.

Для функції $cont(S)$ справедлива наступна умова адитивності: $I(S_1 - S_2) = I(S_1) + I(S_2)$ у тому і лише у тому випадку, коли вираз $(S_1 \vee S_2)$ логічно істинний. З іншого боку, функція $inf(S)$ задовольняє наступній умові: $I(S_1 \cdot S_2) = I(S_1) + I(S_2)$ лише у тому випадку, коли $p(S_1 \cdot S_2) = p(S_1) \cdot p(S_2)$.

Введення таких функцій є спробою визначити абсолютний інформаційний зміст речення, спираючись на його логічну істинність, причому інформаційний зміст речення тим більший, чим більше можливостей воно виключає. Такий підхід застосовується для доведення істинності теорій.

II. ФОРМАЛІЗАЦІЯ ТЕКСТОВОЇ ІНФОРМАЦІЇ ЗА ДОПОМОГОЮ ЛОГІКО-ЛІНГВІСТИЧНИХ МОДЕЛЕЙ

Принципово нові можливості комп'ютерів та мереж зумовлюють необхідність якісно нових засобів опрацювання інформації, які розвиватимуться, насамперед, у напрямку інтелектуалізації. Суспільна потреба у розробці ефективних лінгвістичних технологій, на яких базуватимуться технології оперування знаннями, вимагає створення універсальної системи підтримки лінгвістичних досліджень та розробок. Зокрема, один із підходів представлення текстової інформації у формалізованій формі базується на побудові логіко-лінгвістичної моделі тексту. Основною проблемою на цьому шляху є автоматизація процесу побудови такої моделі, що дозволить надалі застосовувати до логіко-лінгвістичних моделей алгоритми доведення логічних протиріч як для формальних моделей, представлених предикатами першого та вищого порядків.

Формалізація передбачає посилення ролі формальної логіки, оскільки у випадку формальних теорій вже не можна задовольнятися інтуїтивними твердженнями, так як та чи інша аргументація узгоджена з логічними правилами, набутими завдяки здатності до правильного мислення. Існує твердження, що повністю можуть бути формалізовані лише елементарні теорії з простою логічною структурою і невеликим запасом понять. Розробка методу автоматизованого перетворення речення природної мови у логіко-лінгвістичну модель спростує це твердження і доведе, що формалізувати можна будь-яку текстову інформацію.

2.1 Семантико-синтаксична структура речень

Будь-яка флективна мова як система складається з трьох рівнів: фонетичного, лексико-граматичного (формального) та семантичного [51]. Всі ці рівні взаємодоповнюють один одного і взаємопов'язані. Основною одиницею фонетичного рівня є звук, лексико-граматичного – словоформа, семантичного – семема. Структурні компоненти кожного з ярусів представляють собою:

- інтегральні та диференціальні фонетичні ознаки для фонетичного рівня;

- морфеми та граматичні множники (складові компоненти словоформи, що відображають її словотвірні та парадигматичні зв'язки) для лексико-граматичного рівня;

- семи (елементарні, нероздільні компоненти) та семантичні множники (складові компоненти семеми, що виражають її дериваційні зв'язки) для семантичного рівня.

Для роботи з електронними текстовими документами важливими є два останні рівні – лексико-граматичний та семантичний. Побудова логіко-лінгвістичних моделей текстової інформації передбачає використання основних характеристик лексико-граматичного рівня.

Значення складного виразу – це функція значень його частин та синтаксичних правил, що їх поєднують. В лінгвістиці слово вважається основною одиницею мови і знаходиться в центрі лінгвістичних описів.

Слово – найменша самостійна і вільно відтворювана в мовленні, відокремлено оформлена, значима одиниця мови, яка співвідноситься з окремим елементом дійсності (предметом, явищем, ознакою, процесом, відношенням та ін.). **Слово** — це послідовність морфем, об'єднаних за граматичними правилами певної мови.

Основними одиницями лексико-граматичного рівня мови є **словоформи** – це форми або одна із форм конкретного слова, здатна бути елементом тексту. Словоформа є мінімальною самостійною одиницею мови, а її мінімальними несамостійними структурними компонентами є конкретні морфеми: корінь, префікс, суфікс, закінчення.

Лексема – це клас (парадигма словоформ), ідентичних за лексичним значенням та приналежністю до однієї і тієї ж частини мови, тобто тотожних за своєю синтаксичною функцією. Лексема по суті є множиною словоформ, володіє лексичним значенням. Лексичне значення характеризує дану лексеми, а різні граматичні значення характеризують цілі класи слів, до яких дана лексема належить і які виділені за деяким граматичним принципом [52].

Вважається, що словоформа успадковує всі семантичні характеристики лексеми, до яких вона відноситься. Проте при цьому словоформа характеризується ще одним типом граматичного значення – словозмінним, наприклад, значенням категорії відмінка та числа для іменника. Різні словоформи однієї

і тієї ж лексеми мають різні словозмінні характеристики. Носіями таких значень є спеціальні граматичні морфеми – флексії (закінчення). Наприклад, для слова «будинків» лексемою буде «будинок», а конкретна словоформа «будинків» буде характеризуватися родовим відмінком множини.

Типоформа – це клас словоформ з одним і тим самим формантом (афіксом) [53]. Формант, як ідентифікатор тип оформи, повинен бути тотожним до самого себе не лише субстанційно, але й функціонально. Тому, наприклад, слова «холодно» та «вино» не можна вважати типоформами.

Матеріальна спільність формантів для всіх типоформ граматичної форми не обов'язкова, необхідна лише спільність їх семантико-синтаксичних функцій. Узгодження простих одиниць утворюють різні типи лексико-граматичних синтагм (простих та складних словосполучень).

Словосполучення – це синтаксична одиниця, що утворюється поєднанням двох або більше повнозначних слів, одне з яких головне, а друге залежне. Кожне речення представляє собою окреме висловлювання, яке характеризується смисловою завершеністю і цілісністю.

Слово у складі словосполучення втрачає у тій чи іншій мірі свій узагальнюючий характер, набуваючи конкретного лексичного значення.

Отже, словосполучення розглядається як граматично оформлене з'єднання слів, що повинно відповідати двом умовам: граматичній правильності (передбачає формування довільного словосполучення за морфологічними та синтаксичними нормами мови) та номінативній самодостатності (вимагає, щоб словосполучення мало зміст).

При цьому словосполучення – це перша сходинка конкретизації слова. Словосполучення входить у речення як його складова частина. Тому наступний крок конкретизації та актуалізації значення слова є речення, а потім текст.

Кожне речення природної мови характеризується граматичною організацією: будовою та граматичним вираженням (оформленням) членів речення. Різні типи синтаксичних структур дають надзвичайно багаті можливості для якнайточнішого вираження змісту текстів та передачі різноманітних логічних відношень. Складні синтаксичні структури відрізняються від

простих речень своєю будовою: членами простого речення є слова або словосполучення, а членами складних конструкцій є кілька моделей речень (прості речення поєднуються і стають частинами складних).

Речення – це мінімальна і основна комунікативна одиниця мови. Речення має бути цілісним і передавати інформацію в усій складності залежностей і зв'язків. Синтаксичні зв'язки у реченнях називають за функцією залежного члена речення: означальні (між означенням і підметом, додатком, обставиною), обставинні (між присудком і обставинами), додаткові (між присудком і додатком), присудкові (між підметом і присудком). Саме граматичне оформлення членів речення надалі буде називатися синтаксичними ролями. Основною синтаксичною одиницею є речення, оскільки воно виражає повідомлення [54].

На конструктивно-синтаксичному рівні речення, як і словосполучення, залишається номінативною одиницею, що виконує певну комунікативну задачу, відображає певний зміст.

Поняття синтаксичної предикативності як істотної ознаки речення пов'язано з граматичними категоріями особи, часу, модальності, що виражаються особовими формами дієслова. У побудові комунікативної синтаксичної конструкції визначальна роль належить дієслову, оскільки саме дієслово наповнює речення часовим значенням, указує на певний спосіб дії, граматичний вид, у дієслово транспонується й морфологічно закріплюється властиве іменникові граматичне значення особи, числа й роду. Основним же показником центральної ролі дієслова у реченні є його семантико-синтаксична валентність, яка «прогнозує» семантико-синтаксичну структуру речення, визначає семантичну роль не предикатних знаків.

У кожного речення є структурний мінімум, вихідна структура, яка може зводитися до: підметово-присудкової структури; конструкції, яка включає до свого складу необхідні члени речення; наявності предикатного ядра з його валентностями, які підлягають заповненню відповідними іменами. Поняття валентності (від лат. *valentia* – сила) можна трактувати як у широкому, так і у вузькому розумінні. У широкому розумінні йдеться про загальне з'єднання слів та інших мовних рівнів (фонем, морфем, усі повнозначні слова, елементарні речення у структурі складних). У вузькому розумінні валентність є

семантико-синтаксичною категорією, що проявляється у властивості предикатного слова речення зумовлювати своєю семантикою певну кількість синтаксем і формувати семантико-синтаксичну структуру простого елементарного речення [55].

Предикат – це термін логіки та мовознавства, що означає конститутивний член висловлювання, – те, що говорить про суб'єкт. Предикат знаходиться в предикативному відношенні до суб'єкта, здатного набувати різні модальні значення. Предикат – це змістовний аспект, тому існують не лише формальні типи присудка речення, але й семантичні типи предикату.

Грамматика флективної мови складається з двох частин: морфології та синтаксису. Синтаксис (від грецького «побудова», «об'єднання», «порядок») вивчає будову словосполучень та речень, змістові відносини в них та способи їх вираження. Синтаксис тісно пов'язаний з морфологією. В ньому морфологічні одиниці (частини мови, словоформи) вступають у відношення та зв'язки з іншими словами, і в результаті будуються словосполучення і речення. У морфології слова розглядаються в аспекті їх внутрішньої структури, як вони класифікуються між собою, створюють морфологічні форми. В синтаксисі ж ці слова розглядаються зі сторони їх участі у побудові словосполучень та речень. Таким чином, словоформи самостійних частин мови стають членами речення, службові слова виконують обслуговуючу роль при членах та частинах речення [56].

Здійснивши синтаксичний аналіз, тобто визначивши зв'язки між всіма словами вхідної текстової інформації, а також визначивши, якими членами речення виступають слова та словосполучення, можна добратися до змісту висловлювання. Морфологічний розбір забезпечує розуміння того, які об'єкти, суб'єкти та дії приймають участь у висловлюванні, їх кількість, час, рід, форму. А синтаксичний аналіз дає можливість зрозуміти саму суть виловлювання: з яким об'єктом відбувається дія, яка дія відбувається над об'єктом, який суб'єкт виконує цю дію, місце виконання, обставини, ознаки. Усе це допомагає встановити зв'язки між словами речення.

2.2 Принципи побудови логіко-лінгвістичних моделей текстової інформації

Процес розуміння змісту тексту на природній мові можна розглядати як обчислення деякої функції з перетворення фрагмента реальності – тексту – у внутрішню модель – ситуацію, а надалі у зміст тексту.

В алгебраїчній логіці висувається гіпотеза про те, що речення природної мови представляють собою формули деякої алгебри. Математичні тексти теж пишуться у вигляді окремих речень і називаються висловлюваннями [57]. Якщо встановити однозначну відповідність між реченнями природної мови та логічними формулами, то стане можливим обробляти природну мову так само, як і математичні вирази. У математиці абсолютно точно встановлено, що зміст висловлювання – це предикат, який виражається цим висловлюванням, аргументами такого предикату є змінні, присутні у даному висловлюванні.

Отже, текст на природній мові можна ототожнити з математичною формулою, записаною за певними правилами, а змістом речення природної мови вважати деякий предикат. Тоді в математиці, до якої належить формула, повинні бути присутні такі елементи, як змінні, операції над змінними, операції над операціями. Послідовність операцій над змінними та операціями буде записана у вигляді формули [58]. З точки зору конструктивної семантики формалізація передбачає перехід до оперування символами, під час якого не потрібно додаткового аналізу речей об'єктивного світу і вся теорія розвивається у знаковій області.

Визначення логічної формули в логіці предикатів – це правила побудови складних речень природної мови з простих речень та синтаксичних одиниць, які включають опис відношень між основними поняттями атомарних предикатів. Ці правила відповідають синтаксичним правилам природної мови.

Таким чином формули логіки предикатів будуються як символічні системи безвідносно до понять зовнішнього світу. Це передбачає встановлення відповідності між константами логіки предикатів та об'єктами зовнішнього світу, формулами логіки предикатів та функціональними відношеннями між атомарними

предикатами і концептуальними відношеннями, про які йдеться у тексті.

Логічна формула описує поняття одного глобального відношення і приймає значення «істина» або «хибність» залежно від наявності або відсутності концептуальних відношень, описаних цією формулою. Також обов'язковою є відповідність між формулами логіки предикатів та концептами реального світу, їх кількістю, завдяки чому в мову привноситься конкретний зміст, що називається семантикою логіки предикатів.

Отже, знання, які можуть бути представлені за допомогою логіки предикатів, виступають фактами, що відображаються логічними формулами. Автоматичне перетворення речень, написаних на природній мові, на мову формальних систем типу логіки предикатів, називається розумінням природної мови. Логіці предикатів властивий високий рівень модульності знань, і разом з цим вона дозволяє отримати єдину систему представлення, в якій логічно роз'яснюються властивості знань як одного цілого. Першочерговим завданням логіки предикатів є пояснення логічних основ природної мови. Оскільки будь-яка флективна мова надзвичайно складна, то об'єкти, якими оперує логіка предикатів, обмежені тими компонентами, елементи яких легко піддаються формалізації. Тому вводяться базові поняття для систематизації цієї логіки, що дасть змогу вилучати знання із різних структур текстової інформації.

Логіко-лінгвістична модель речення природної мови формується із логічних виразів, атомарних предикатів, кожному з яких ставиться у відповідність просте речення природної мови, а кожній логічній операції відповідає певне концептуальне відношення. Фактично, логіко-лінгвістична модель є зв'язкою між елементами формальної логіки та синтаксичною структурою речень природної мови.

Так як речення – це слово або група слів, за допомогою яких відображається думка, і воно характеризується смисловою завершеністю, граматичною організацією та предикативністю, то кожна логічна формула є інтерпретацією певного речення природної мови.

Просте речення у формальній логіці – це атомарний предикат, складне речення – складне логічне висловлювання, сукупність атомарних предикатів поєднаних логічними зв'язками.

Уніфікована форма логіко-лінгвістичної моделі довільного речення природної мови має вигляд [59]:

$$L^S = \bigwedge_{p \in P^S} \bigwedge_{h \in H_p^S} L_p^S(h), \quad (2.1)$$

$$L_p^S(h) = \bigwedge_{x \in X_p^S(h)} \bigwedge_{g \in G_p^S(x,h)} L_p^S(x, g, h), \quad (2.2)$$

$$L_p^S(x, g, h) = \bigwedge_{y \in Y_p^S(x,g,h)} \bigwedge_{q \in Q_p^S(x,g,y,h)} L_p^S(x, g, y, q, h), \quad (2.3)$$

$$L_p^S(x, g, y, q, h) = \bigwedge_{z \in Z_p^S(x,g,y,q,h)} \bigwedge_{r \in R_p^S(x,g,y,q,z,h)} L_p^S(x, g, y, q, z, r, h), \quad (2.4)$$

де S – речення природної мови;

p – відношення, що пов’язує суб’єкти, об’єкти та предмети відношень у реченні S , $p \in P^S$ – множина відношень, що входять до речення S ;

h – характеристика p -го відношення речення S , $h \in H_p^S$ – множина характеристик p -го відношення у реченні S ;

$L_p^S(h)$ – предикат (предикативний вираз) [60], який описує p -е відношення з h -ю характеристикою і пов’язує суб’єкти, об’єкти та предмети відношення p у реченні S ;

x – суб’єкт речення S ,

$x \in X_p^S(h)$ – множина суб’єктів, що пов’язані з об’єктами речення S p -м відношенням, яке володіє h -ю характеристикою;

g – характеристика суб’єкта x речення S , $g \in G_p^S(x, h)$ – множина характеристик суб’єкта $x \in X_p^S(h)$;

$L_p^S(x, g, h)$ – предикат (предикативний вираз), який описує p -е відношення з h -ю характеристикою між суб’єктом $x \in X_p^S(h)$ з характеристикою $g \in G_p^S(x, h)$, об’єктами та предметами p -го відношення у реченні S ;

y – об’єкт речення S , $y \in Y_p^S(x, g, h)$ – множина об’єктів, що пов’язані з суб’єктами речення S p -м відношенням, яке володіє h -ю характеристикою;

q – характеристика об’єкта y речення S , $q \in Q_p^S(x, g, y, h)$ – множина характеристик об’єкта $y \in Y_p^S(x, g, h)$;

$L_p^S(x, g, y, q, h)$ – предикат (предикативний вираз), який описує p -е відношення з h -ю характеристикою між суб’єктом $x \in X_p^S(h)$ з характеристикою $g \in G_p^S(x, h)$ і об’єктом $y \in Y_p^S(x, g, h)$ з характеристикою $q \in Q_p^S(x, g, y, h)$ та предмети p -го відношення у реченні S ;

z – предмет p -го відношення речення S , $z \in Z_p^S(x, g, y, q, h)$ – множина предметів p -го відношення, яке володіє h -ю характеристикою, між суб’єктом $x \in X_p^S(h)$ з характеристикою $g \in G_p^S(x, h)$ та об’єктом $y \in Y_p^S(x, g, h)$ з характеристикою $q \in Q_p^S(x, g, y, h)$;

r – характеристика предмета p -го відношення речення S , $r \in R_p^S(x, g, y, q, z, h)$ – множина характеристик предмета $z \in Z_p^S(x, g, y, q, h)$;

$L_p^S(x, g, y, q, z, r, h)$ – простий, нероздільний предикат, який описує частину речення, що має закінчений зміст та відображає у реченні S p -е відношення з h -ю характеристикою між суб’єктом $x \in X_p^S(h)$ з характеристикою $g \in G_p^S(x, h)$ і об’єктом $y \in Y_p^S(x, g, h)$ з характеристикою $q \in Q_p^S(x, g, y, h)$, предмет якого $z \in Z_p^S(x, g, y, q, h)$ володіє характеристикою $r \in R_p^S(x, g, y, q, z, h)$.

Кількість простих предикатів у формулі (2.1) дорівнює сумі можливих комбінацій узгодження суб'єктів, об'єктів, предметів, відношень між ними та їх характеристик.

Логіко-лінгвістична модель L^S речення S відображається сукупністю формул (2.1) – (2.4) та формально представляє собою послідовність восьми кон'юнкцій, що входять до цих формул. Тобто логіко-лінгвістична модель L^S є множиною простих предикатів $L_p^S(x, g, y, q, z, r, h)$, кількість яких дорівнює v^S .

Перехід від загальної формули (2.1) до предикату $L_p^S(x, g, y, q, z, r, h)$ є декомпозицією проблеми формального опису довільного речення природної мови та відображає системний підхід до її вирішення. Тому складний вираз L^S істинний тоді і тільки тоді, коли істинними є всі елементарні предикати типу $L_p^S(x, g, y, q, z, r, h)$, що входять до нього.

Основою принципів побудови логіко-лінгвістичної моделі (2.1) – (2.4) служить синтаксичний розбір речення, тобто визначення зв'язків між усіма його словами та встановлення їх синтаксичних ролей. Формула (2.4) є інтерпретацією синтаксичної структури простого речення природної мови з урахуванням семантичних зв'язків, що є формальним засобом відображення змісту текстової інформації. Логічні формули (2.2) – (2.4) є інтерпретацією семантико-синтаксичної структури речення природної мови, так як будуються на основі наявних у реченні означальних (між означенням і підметом, означенням і додатком), обставинних (між присудком і обставинами), додаткових (між присудком і додатком), присудкових (між підметом і присудком) зв'язків.

Принципи побудови логіко-лінгвістичних моделей:

- 1) визначити кількість простих атомарних предикатів v^S .
- 2) зафіксувати значення елементів кортежу логічних операцій $O(S)$ речення S :

$$O(S) = [o_e(S) \mid e = \overline{1, h}],$$

де h – загальна кількість логічних операцій, наявних у реченні S ;

3) визначити множину відношень P^S між суб'єктами та об'єктами речення S , що фактично містить дії, стани або властивості суб'єкта;

4) зафіксувати множини предикатних констант H_p^S , що містять характеристики (параметри) відношень між суб'єктами та об'єктами у реченні S ;

5) визначити множини суб'єктів $X_p^S(h)$, що входять до речення S та пов'язані з об'єктами речення S p -м відношенням, яке володіє h -ю характеристикою;

6) зафіксувати множини предикатних констант $G_p^S(x, h)$, що містять характеристики (параметри) суб'єктів, які входять до речення S ;

7) визначити множини об'єктів $Y_p^S(x, g, h)$, що входять до речення S та пов'язані з суб'єктами речення $x \in X_p^S(h)$ з характеристиками $g \in G_p^S(x, h)$ p -м відношенням, яке володіє h -ю характеристикою;

8) зафіксувати множини предикатних констант $Q_p^S(x, g, y, h)$, що містять характеристики (параметри) об'єктів, які входять до речення S ;

9) визначити множини предметів $Z_p^S(x, g, y, q, h)$ p -го відношення, яке володіє h -ю характеристикою, між суб'єктами $x \in X_p^S(h)$ з характеристиками $g \in G_p^S(x, h)$ та об'єктами $y \in Y_p^S(x, g, h)$ з характеристиками $q \in Q_p^S(x, g, y, h)$;

10) зафіксувати множину предикатних констант $R_p^S(x, g, y, q, z, h)$, що містить характеристики (параметри) предметів p -го відношення, яке володіє h -ю характеристикою, між суб'єктами $x \in X_p^S(h)$ з характеристиками $g \in G_p^S(x, h)$ та об'єктами $y \in Y_p^S(x, g, h)$ з характеристиками $q \in Q_p^S(x, g, y, h)$ у реченні S ;

11) сформувати один логічний вираз (2.1) для L^S , шляхом підстановки у нього змінних, отриманих за формулами (2.2) – (2.4) відповідно до кількості елементів зафіксованих множин згідно принципів 3)-10).

Налаштуванням логіко-лінгвістичної моделі на ситуацію називається приведення її загальної форми (2.1) – (2.4) до вигляду, адекватного змісту конкретного типу речення.

Залежно від типу речення природної мови, уніфікована форма логіко-лінгвістичної моделі (2.1) – (2.4) може приймати такі окремі форми [61,62].

1) Просте речення природної мови містить лише одну частину, що має закінчений зміст, тому:

$$\begin{aligned} |P^S| = 1; |H_p^S| = 1; |X_p^S(h)| = 1; |G_p^S(x, h)| = 1; |Y_p^S(x, g, h)| = 1; \\ |Q_p^S(x, g, y, h)| = 1; |Z_p^S(x, g, y, q, h)| = 1; |R_p^S(x, g, y, q, z, h)| = 1. \end{aligned}$$

Внаслідок цього логіко-лінгвістична модель (2.1) – (2.4) зводиться до простого предикату $L^S = L_p^S(x, g, y, q, z, r, h)$, а кількість частин речення S , що мають закінчений зміст $v^S = |R_p^S(x, g, y, q, z, h)| = 1$.

Наприклад, для простого речення природної мови «*Цікава лекція заохочує студентів до подальшого навчання*» логіко-лінгвістична модель буде мати вигляд:

$$L(S) = L^{S_1} = p_1(x_1, g_1, y_1, q_1, z_1, r_1, h_1) = p_1(x_1, g_1, y_1, 0, z_1, r_1, 0),$$

$L(S) = L^{S_1} = \text{заохочує (лекція, цікава, студентів, 0, навчання, подальшого, 0)}$.

Для простого речення природної мови «*Гори востаннє засвітилися ажурним золотом*» логіко-лінгвістична модель матиме вигляд:

$$L(S) = L^{S_2} = p_2(x_2, 0, y_2, q_2, 0, 0, h_2),$$

$L(S) = L^{S_2} = \text{засвітилися(гори, 0, золотом, ажурним, 0, 0, востаннє)}$.

2) Логіко-лінгвістична модель простого речення, ускладненого однорідними членами, приймає різні форми залежно від синтаксичних ролей, які виконують у ньому однорідні члени.

А) Якщо множина відношень P^S , що пов'язують суб'єкти, об'єкти та предмети відношень у реченні S містить декілька елементів, тоді:

$$\begin{aligned} |H_p^S| = 1; |X_p^S(h)| = 1; |G_p^S(x, h)| = 1; |Y_p^S(x, g, h)| = 1; \\ |Q_p^S(x, g, y, h)| = 1; |Z_p^S(x, g, y, q, h)| = 1; |R_p^S(x, g, y, q, z, h)| = 1. \end{aligned}$$

Для такого випадку логіко-лінгвістична модель (2.1) – (2.4) набуває вигляду: $L^S = \bigwedge_{p \in P^S} L_p^S(x, g, y, q, z, r, h)$, кількість яких визначається потужністю множини відношень:

$$v^S = \sum_{p \in P^S} |R_p^S(x, g, y, q, z, h)| = |P^S|.$$

Наприклад, для простого речення природної мови, ускладненого однорідними присудками «Цікава лекція заохочує та спонукає студентів до подальшого навчання» типова форма логіко-лінгвістичної моделі (2.1) – (2.4) буде мати такий вигляд:

$$\begin{aligned} L(S) = L^{S_1} = p_{11}(x_1, g_1, y_1, q_1, z_1, r_1, h_1) \& p_{12}(x_1, g_1, y_1, q_1, z_1, r_1, h_1) = \\ = p_{11}(x_1, g_1, y_1, 0, z_1, r_1, 0) \& p_{12}(x_1, g_1, y_1, 0, z_1, r_1, 0), \\ L(S) = L^{S_1} = \text{заохочує (лекція, цікава, студентів, 0, навчання,} \\ \text{подальшого, 0)} \& \text{спонукає (лекція, цікава, студентів, 0, навчання,} \\ \text{подальшого, 0)}. \end{aligned}$$

Б) Якщо множина $G_p^S(x, h)$ характеристик (параметрів) хоча б одного із суб'єктів x містить більше одного елемента, то

$$\begin{aligned} |P^S| = 1; |H_p^S| = 1; |Y_p^S(x, g, h)| = 1; |Q_p^S(x, g, y, h)| = 1; \\ |Z_p^S(x, g, y, q, h)| = 1; |R_p^S(x, g, y, q, z, h)| = 1. \end{aligned}$$

Логіко-лінгвістична модель (2.1) – (2.4) приймає такий вигляд:

$$L^S = \bigwedge_{x \in X_p^S(h)} \bigwedge_{g \in G_p^S(x, h)} L_p^S(x, g, y, q, z, r, h),$$

кількість яких визначається за формулою:

$$v^S = \sum_{x \in X_p^S(h)} \sum_{g \in G_p^S(x, h)} |R_p^S(x, g, y, q, z, h)| = \sum_{x \in X_p^S(h)} |G_p^S(x, h)|.$$

Наприклад, для простого речення природної мови, ускладненого однорідними означеннями для підмета «Цікава та інформативна лекція заохочує студентів до подальшого навчання» типова форма логіко-лінгвістичної моделі (2.1) – (2.4) матиме такий вигляд:

$$L(S) = L^S = p_1(x_1, g_{11}, y_1, q_1, z_1, r_1, h_1) \& p_1(x_1, g_{12}, y_1, q_1, z_1, r_1, h_1) = \\ = p_1(x_1, g_{11}, y_1, 0, z_1, r_1, 0) \& p_1(x_1, g_{12}, y_1, 0, z_1, r_1, 0),$$

$L(S) = L^S =$ заохочує (лекція, цікава, студентів, 0, навчання, подальшого, 0) & заохочує (лекція, інформативна, студентів, 0, навчання, подальшого, 0).

В) Якщо множина $Q_p^S(x, g, y, h)$ характеристик (параметрів) хоча б одного із об'єктів y містить більше одного елемента, то:

$$\left|P^S\right|=1; \left|H_p^S\right|=1; \left|X_p^S(h)\right|=1; \left|G_p^S(x, h)\right|=1; \\ \left|Z_p^S(x, g, y, q, h)\right|=1; \left|R_p^S(x, g, y, q, z, h)\right|=1.$$

Логіко-лінгвістична модель (2.1) – (2.4) буде мати такий узагальнений вигляд:

$$L^S = L_p^S(h), \\ L_p^S(h) = L_p^S(x, g, h), \\ L_p^S(x, g, h) = \bigwedge_{y \in Y_p^S(x, g, h)} \bigwedge_{q \in Q_p^S(x, g, y, h)} L_p^S(x, g, y, q, h), \\ L_p^S(x, g, y, q, h) = L_p^S(x, g, y, q, z, r, h).$$

Тобто модель (2.1) – (2.4) зводиться до кон'юнкції простих предикатів $L_p^S(x, g, y, q, z, r, h)$:

$$L^S = \bigwedge_{y \in Y_p^S(x, g, h)} \bigwedge_{q \in Q_p^S(x, g, y, h)} L_p^S(x, g, y, q, z, r, h),$$

кількість яких визначається:

$$v^S = \sum_{y \in Y_p^S(x, g, h)} \sum_{q \in Q_p^S(x, g, y, h)} \left| R_p^S(x, g, y, q, z, h) \right| = \sum_{y \in Y_p^S(x, g, h)} \left| Q_p^S(x, g, y, h) \right|.$$

Наприклад, для простого речення природної мови, ускладненого однорідними означеннями та додатками «Цікава

лекція заохочує лінивих і працюючих студентів та практикантів до подальшого навчання» типова форма логіко-лінгвістичної моделі (2.1) – (2.4) буде мати такий вигляд:

$$L(S) = L^{S_1} = p_1(x_1, g_1, y_{11}, q_{11}, z_1, r_1, h_1) \& p_1(x_1, g_1, y_{12}, q_{11}, z_1, r_1, h_1) \& \\ p_1(x_1, g_1, y_{11}, q_{12}, z_1, r_1, h_1) \& p_1(x_1, g_1, y_{12}, q_{12}, z_1, r_1, h_1) = \\ = p_1(x_1, g_1, y_{11}, q_{11}, z_1, r_1, 0) \& p_1(x_1, g_1, y_{12}, q_{11}, z_1, r_1, 0) \& \\ p_1(x_1, g_1, y_{11}, q_{12}, z_1, r_1, 0) \& p_1(x_1, g_1, y_{12}, q_{12}, z_1, r_1, 0),$$

$L(S) = L^{S_1} =$ заохочує (лекція, цікава, студентів, лінивих, навчання, подальшого, 0) & заохочує (лекція, цікава, практикантів, лінивих, навчання, подальшого, 0) & заохочує (лекція, цікава, студентів, працюючих, навчання, подальшого, 0) & заохочує (лекція, цікава, практикантів, працюючих, навчання, подальшого, 0).

Для простого речення, ускладненого однорідними означеннями додатка «Гори востаннє засвітилися легким, ажурним золотом» використовується форма:

$$L(S) = L^{S_2} = p_2(x_2, g_2, y_2, q_{21}, z_2, r_2, h_2) \& p_2(x_2, g_2, y_2, q_{22}, z_2, r_2, h_2) = \\ p_2(x_2, 0, y_2, q_{21}, 0, 0, h_2) \& p_2(x_2, 0, y_2, q_{22}, 0, 0, h_2),$$

$L(S) = L^{S_2} =$ засвітилися(гори, 0, золотом, легким, 0,0,востаннє) & засвітилися(гори, 0, золотом, ажурним, 0,0,востаннє).

Г) Якщо множина $R_p^S(x, g, y, q, z, h)$ характеристик (параметрів) хоча б одного із предметів p -го відношення містить більше одного елемента, то:

$$\left| P^S \right| = 1; \left| H_p^S \right| = 1; \left| X_p^S(h) \right| = 1; \left| G_p^S(x, h) \right| = 1; \\ \left| Y_p^S(x, g, h) \right| = 1; \left| Q_p^S(x, g, y, h) \right| = 1,$$

При цьому логіко-лінгвістична модель (2.1) – (2.4) набуде такого узагальненого вигляду:

$$L^S = L_p^S(h), \\ L_p^S(h) = L_p^S(x, g, h), \\ L_p^S(x, g, h) = L_p^S(x, g, y, q, h),$$

$$L_p^S(x, g, y, q, h) = \bigwedge_{z \in Z_p^S(x, g, y, q, h)} \bigwedge_{r \in R_p^S(x, g, y, q, z, h)} L_p^S(x, g, y, q, z, r, h).$$

Логіко-лінгвістична модель (2.1) – (2.4) зводиться до кон'юнкції простих предикатів $L_p^S(x, g, y, q, z, r, h)$:

$$L^S = \bigwedge_{z \in Z_p^S(x, g, y, q, h)} \bigwedge_{r \in R_p^S(x, g, y, q, z, h)} L_p^S(x, g, y, q, z, r, h),$$

кількість яких визначається за формулою:

$$v^S = \sum_{z \in Z_p^S(x, g, y, q, h)} |R_p^S(x, g, y, q, z, h)|.$$

Наприклад, для простого речення природної мови, ускладненого однорідними додатками «Цікава лекція заохочує студентів до подальшого навчання та роботи» типова форма логіко-лінгвістичної моделі (2.1) – (2.4) буде мати такий вигляд:

$$\begin{aligned} L(S) &= L^{S_1} = p_1(x_1, g_1, y_1, q_1, z_{11}, r_1, h_1) \& p_1(x_1, g_1, y_1, q_1, z_{12}, r_1, h_1) = \\ &= p_1(x_1, g_1, y_1, 0, z_{11}, r_1, 0) \& p_1(x_1, g_1, y_1, 0, z_{12}, 0, 0), \end{aligned}$$

$L(S) = L^{S_1} = \text{заохочує (лекція, цікава, студентів, 0, навчання, подальшого, 0)} \& \text{заохочує (лекція, цікава, студентів, 0, роботи, 0, 0)}$.

3) Складні речення природної мови описуються логіко-лінгвістичною моделлю (2.1) – (2.4), яка, залежно від типу речення, буде містити декілька простих предикатів, між якими можуть вживатися різні логічні операції. Для складних речень v^S завжди більша одиниці, а потужність множин $|P^S| \geq 2$;
 $|X_p^S(h)| \geq 2$.

А) Для складносурядного та безсполучникового речення природної мови, частини якого рівноправні за змістом та пов'язані сурядним зв'язком, логіко-лінгвістична модель (2.1) – (2.4) набуває такого вигляду:

$$L^S = L_{p^{(1)}}^{S^{(1)}}(x^{(1)}, g^{(1)}, y^{(1)}, q^{(1)}, z^{(1)}, r^{(1)}, h^{(1)}) \& \dots \&$$

$$L_{p^{(\lambda)}}^{S^{(\lambda)}}(x^{(\lambda)}, g^{(\lambda)}, y^{(\lambda)}, q^{(\lambda)}, z^{(\lambda)}, r^{(\lambda)}, h^{(\lambda)}) \& \dots \&$$

$$L_{p^{(\varphi)}}^{S^{(\varphi)}}(x^{(\varphi)}, g^{(\varphi)}, y^{(\varphi)}, q^{(\varphi)}, z^{(\varphi)}, r^{(\varphi)}, h^{(\varphi)}),$$

де $L_{p^{(1)}}^{S^{(1)}}(x^{(1)}, g^{(1)}, y^{(1)}, q^{(1)}, z^{(1)}, r^{(1)}, h^{(1)})$ – предикатний вираз, що описує частину складного речення S , яка відображає закінчений зміст і виражається за допомогою простого речення першого рівня вкладеності;

$$L_{p^{(\lambda)}}^{S^{(\lambda)}}(x^{(\lambda)}, g^{(\lambda)}, y^{(\lambda)}, q^{(\lambda)}, z^{(\lambda)}, r^{(\lambda)}, h^{(\lambda)})$$
 – предикатний вираз, що

описує частину складного речення S , яка відображає закінчений зміст і виражається за допомогою простого речення рівня вкладеності $\lambda = \overline{1, \varphi}$, φ – кількість простих речень;

$$L_{p^{(\varphi)}}^{S^{(\varphi)}}(x^{(\varphi)}, g^{(\varphi)}, y^{(\varphi)}, q^{(\varphi)}, z^{(\varphi)}, r^{(\varphi)}, h^{(\varphi)})$$
 – предикатний вираз, що

описує частину складного речення S , яка відображає закінчений зміст і виражається за допомогою простого речення рівня вкладеності φ .

Якщо частини безсполучникового або складносурядного речення протиставляються або зіставляються, взаємно виключаються або чергуються, то у логіко-лінгвістичній моделі логічну операцію кон'юнкції буде замінено на операцію диз'юнкції.

Складне безсполучникове речення *«Інформаційні технології посідають важливу роль у житті людей, суспільство просувається вперед»* має таку логіко-лінгвістичну модель:

$$\begin{aligned} L(S) &= L'(S_1) \& L''(S_1) = \\ &= p'_1(x'_1, g'_1, y'_1, q'_1, z'_1, r'_1, h'_1) \& p''_1(x''_1, g''_1, y''_1, q''_1, z''_1, r''_1, h''_1) = \\ &= p'_1(x'_1, g'_1, y'_1, q'_1, z'_1, r'_1, 0) \& p''_1(x''_1, 0, 0, 0, 0, 0, h''_1), \end{aligned}$$

$L(S) = L'(S_1) \& L''(S_1) =$ *посідають (технології, інформаційні, роль, важливу, житті, людей, 0) & просувається (суспільство, 0, 0, 0, 0, 0, вперед).*

Б) Зміст складнопідрядного або безсполучникового речення природної мови, що складається з декількох простих речень, одне

з яких уточнює, конкретизує інші, буде описано логіко-лінгвістичною моделлю вигляду:

$$\begin{aligned}
 L^S = & [L_{p^{(1)}}^{S(1)}(x^{(1)}, g^{(1)}, y^{(1)}, q^{(1)}, z^{(1)}, r^{(1)}, h^{(1)}) \rightarrow \\
 & (L_{p^{(\lambda)}}^{S(\lambda)}(x^{(\lambda)}, g^{(\lambda)}, y^{(\lambda)}, q^{(\lambda)}, z^{(\lambda)}, r^{(\lambda)}, h^{(\lambda)}) \& \dots \& \\
 & L_{p^{(\phi)}}^{S(\phi)}(x^{(\phi)}, g^{(\phi)}, y^{(\phi)}, q^{(\phi)}, z^{(\phi)}, r^{(\phi)}, h^{(\phi)}))] \vee \\
 & [L_{p^{(1)}}^{S(1)}(x^{(1)}, g^{(1)}, y^{(1)}, q^{(1)}, z^{(1)}, r^{(1)}, h^{(1)}) \rightarrow \\
 & (L_{p^{(\lambda)}}^{S(\lambda)}(x^{(\lambda)}, g^{(\lambda)}, y^{(\lambda)}, q^{(\lambda)}, z^{(\lambda)}, r^{(\lambda)}, h^{(\lambda)}) \vee \dots \vee \\
 & L_{p^{(\phi)}}^{S(\phi)}(x^{(\phi)}, g^{(\phi)}, y^{(\phi)}, q^{(\phi)}, z^{(\phi)}, r^{(\phi)}, h^{(\phi)}))] .
 \end{aligned}$$

Як правило, складнопідрядне речення описується двома простими предикатами, один із яких описує головне речення, а другий – підрядне, яке може виражати ознаку суб'єкта x , об'єкта y , предмета відношення z ; розкривати зміст пояснювального слова u головному реченні; виражати просторові, часові, причинні, цільові та інші відношення.

Предикатний вираз $L_{p^{(1)}}^{S(1)}(x^{(1)}, g^{(1)}, y^{(1)}, q^{(1)}, z^{(1)}, r^{(1)}, h^{(1)})$ описує закінчений зміст головного речення, що входить до складнопідрядного, у свою чергу підрядне речення може складатися із декількох простих, представлених простими предикатами, що описують рівноправність:

$$\begin{aligned}
 & (L_{p^{(\lambda)}}^{S(\lambda)}(x^{(\lambda)}, g^{(\lambda)}, y^{(\lambda)}, q^{(\lambda)}, z^{(\lambda)}, r^{(\lambda)}, h^{(\lambda)}) \& \dots \& \\
 & L_{p^{(\phi)}}^{S(\phi)}(x^{(\phi)}, g^{(\phi)}, y^{(\phi)}, q^{(\phi)}, z^{(\phi)}, r^{(\phi)}, h^{(\phi)})), \\
 & (L_{p^{(\lambda)}}^{S(\lambda)}(x^{(\lambda)}, g^{(\lambda)}, y^{(\lambda)}, q^{(\lambda)}, z^{(\lambda)}, r^{(\lambda)}, h^{(\lambda)}) \vee \dots \vee \\
 \text{та} & L_{p^{(\phi)}}^{S(\phi)}(x^{(\phi)}, g^{(\phi)}, y^{(\phi)}, q^{(\phi)}, z^{(\phi)}, r^{(\phi)}, h^{(\phi)})),
 \end{aligned}$$

якщо підрядні речення протиставляються один одному.

Логіко-лінгвістична модель для складнопідрядного речення причини *«Експерт дав негативну рецензію, щоб автори виправили свої помилки»* має вигляд:

$$\begin{aligned}
L(S) &= L'_1(S_1) \rightarrow L''_1(S_1) = \\
&= p'_1(x'_1, g'_1, y'_1, q'_1, z'_1, r'_1, h'_1) \rightarrow p''_1(x''_1, g''_1, y''_1, q''_1, z''_1, r''_1, h''_1) = \\
&= p'_1(x'_1, 0, y'_1, q'_1, 0, 0, 0) \rightarrow p''_1(x''_1, 0, y''_1, q''_1, 0, 0, 0), \\
L(S) &= L'_1(S_1) \rightarrow L''_1(S_1) = \text{дав (експерт, 0, рецензію, негативну,} \\
&0, 0, 0) \rightarrow \text{виправили (автори, 0, помилки, свої, 0, 0, 0)}.
\end{aligned}$$

Складнопідрядне речення з послідовною підрядністю «Як я заздрила тим людям, що не мали відпочинку, поки їх нелюдська втома не валила на часинку» матиме таку логіко-лінгвістичну модель:

$$\begin{aligned}
L(S) &= L'(S_2) \& L''(S_2) \rightarrow L'''(S_2) = \\
&= p'_2(x'_2, 0, y'_2, q'_2, 0, 0, 0) \& \neg p''_2(y'_2, 0, y''_2, 0, 0, 0, 0) \rightarrow \\
&\neg p'''_2(x'''_2, g'''_2, y'''_2, 0, 0, 0, h'''_2), \\
L(S) &= L'(S_2) \& L''(S_2) \rightarrow L'''(S_2) = \\
&= \text{заздрила (я, 0, людям, тим, 0, 0, 0) \&} \\
&\neg \text{мали (люди, 0, відпочинку, 0, 0, 0, 0) \rightarrow} \\
&\neg \text{валила (втома, людська, їх, 0, 0, 0, часинку)}.
\end{aligned}$$

Залежно від концептуальних зв'язків між простими реченнями, які входять до складного, частини логіко-лінгвістичної моделі (2.1) – (2.4) інтерпретуються, як складні або прості висловлювання, що дозволяє рекурсивно повертатися до загального вигляду формули (2.1) – (2.4) і накладати її на конкретну ситуацію доти, доки не будуть однозначно інтерпретовані всі частини речення природної мови S , що відображають закінчений зміст.

Із виведених окремих форм видно, що формування логіко-лінгвістичних моделей для різних типів речень природної мови відбувається за єдиним принципом: уніфікована форма (шаблон) заповнюється для кожного речення предикатними змінними та константами у відповідності з тим, яку синтаксичну роль виконує те чи інше слово. Тобто зміст кожного речення природної мови фактично представляє собою визначення відношень між підметом (суб'єктом) та присудком (предикатом), а також їх спільне відношення до того, що вони виражають в дійсності, та формують

основну граматичну властивість речення – його суть – предикативність [63].

Обґрунтуванням можливості побудови логіко-лінгвістичної моделі (2.1) – (2.4), а також підтвердженням принципів її формування є метод автоматизованого формування логіко-лінгвістичних моделей текстової інформації.

2.3 Формалізовані правила визначення структурних компонентів логіко-лінгвістичних моделей

Для того, щоб аналізувати зміст текстів, необхідно виявити зв'язки між лексико-граматичним та семантичним рівнями флективної мови, встановити, які саме когнітивні процеси до цього приводять. Це дасть можливість пояснити, як синтаксичні ланцюжки передають різний семантичний зміст. Саме синтаксичні ланцюжки є структурними компонентами логіко-лінгвістичних моделей текстової інформації. Зокрема, правильне визначення синтаксичних ролей у реченні дає можливість встановити в логіко-лінгвістичній моделі предикатні змінні та константи, що їх описують.

З точки зору прагматики можна говорити про те, що предметом синтаксису та семантики служить граматична структура речень. Це і пропонується довести за допомогою подальших досліджень.

Вважається, що синтаксис формалізований, якщо існують правила, які рекурсивно визначають, що таке правильно побудований ланцюжок. Множина правил, що виражають граматичні відношення між членами речень, дають можливість встановити зміст окремих структурних одиниць логіко-лінгвістичних моделей.

Поставимо у відповідність кожному слову речення S_i ($i = \overline{1, r}$, де r – кількість слів у реченні) набір характеристик [62, 64]:

$$Z_i(S_i) = \{cm_i, g_i, n_i, k2_i, t_i, h_i, l_i, ch_i\}, \quad (2.5)$$

де $cm_i = \overline{1, 11}$ – граматична характеристика, що позначає частину мови, кожному цифровому значенню якої відповідає іменник, прикметник, числівник, займенник, дієслово, дієприкметник, дієприслівник, прислівник, прийменник, сполучник або частка відповідно;

$g_i = \overline{1,7}$ – морфологічна ознака, що відповідає за відмінок;

$n_i = \overline{1,2}$ – граматичний параметр, що означає число;

$k2_i = \overline{1,4}$ – граматичний параметр, що означає рід;

$t_i = \overline{0,3}$ – граматичний параметр, що означає час;

$h_i = \overline{1,3}$ – граматичний параметр, що означає спосіб;

$l_i = \overline{1,3}$ – граматичний параметр, що означає особу;

$ch_i = \overline{1,5}$ – параметр, що означає синтаксичну роль (підмет, присудок, додаток, означення, обставина).

Тобто у формалізованому вигляді характеристики кожного слова можна представити як одновимірний масив. Отже, кожне слово речення буде характеризуватися набором цифр, наприклад, характеристики $Z_i(S_i) = \{1,1,1,1,0,0,0,0\}$ означають, що слово S_i – іменник у називному відмінку, чоловічого роду однини.

Для всіх флективних форм характерно, що наступний за словом рівень синтаксичних конструкцій – це словосполучення. У кожній природній мові слова у словосполученнях пов'язані за певними законами, наприклад, в українській мові – відношення прилягання, узгодження, керування. Аналізуючи ці відношення, можна сформулювати ряд правил, за якими неявно формуються словосполучення, та формалізувати їх, базуючись на тому, що слова пов'язані за рахунок знаходження та узгодження певних граматичних ознак. Всі сформульовані правила об'єднаємо у систему продукцій. В основі створеної системи лежить продукційна модель представлення знань, в якій правила виступають засобами прямого опису способів логічного виводу. При цьому характерною рисою представлення знань з високою модульністю є простота доповнення, модифікації та анулювання даних.

В системах, які базуються на правилах, знання представлені не за допомогою відносно декларативного, статичного способу, а у формі багато чисельних правил, що вказують, які висновки повинні бути зроблені в залежності від ситуації. Одним із найбільш часто вживаних методів формування логічних виводів є дедуктивна логіка, яка використовується для визначення обґрунтувань доведень. Логічне доведення – це група тверджень, в якій останні розглядаються як обґрунтовані з використанням

попередніх в ланцюгу розмірковувань. Одним із типів логічного доведення є силогізм, який складається з двох посилянь та одного висновку, який впливає з посилянь.

У доведенні посилення використовуються як докази, що дозволяють довести істинність висновків, відповідно з істинних посилянь впливає істинний висновок.

Використовуючи два основних припущення логіки предикатів «будь-яке висловлювання або істинне або хибне (принцип двозначності)» та «істинне значення складного висловлювання залежить лише від істинності значень простих висловлювань, що в нього входять, та характеру їх зв'язку», сформовано правила, що відображають можливі семантичні зв'язки між словами у словосполученні.

Кожне правило складається з двох частин. Перша з них – **антецедент** або посилення правила, складається з елементарних частин, з'єднаних логічними зв'язками «і», «або» і т.д. Друга частина, що називається **консеквентом** або висновком, складається з однієї або декількох частин, які утворюють рішення або вказують на дію, яку необхідно виконати. Антецедент представляє собою зразок правила, призначений для розпізнавання ситуації, коли це правило має спрацювати.

Дія компоненти виводу базується на правилі **modus ponens**. Слово «модус» в логіці означає різновид деякої загальної форми міркувань. «*Modus ponens*» – термін середньовічної логіки, що означає певне правило виводу і відповідний йому логічний закон. Правило виводу *modus ponens* (ще називається правило визначення або гіпотетичний силогізм) дозволяє від твердження умовного висловлювання та твердження на його основі (антецедента) перейти до твердження висновку (консеквента).

Такі правила дають можливість встановити парадигматичні відношення між словами у реченнях природної мови. Таким чином, словосполучення формуються через відповідність граматичних форм кожного слова. Для української мови сформована система продукцій, що містить тридцять два правила формування словосполучень [4].

1) Правило трактується наступним чином: якщо частина мови для першого слова прикметник, а для другого – іменник, відмінки, число та рід обох слів співпадають, то слова утворюють

словосполучення. Наприклад, словосполучення «інформаційна технологія».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 2) \text{and} (\text{cm}(S_{i+1}) = 1)) \text{and} (g(S_i) = g(S_{i+1})) \\ & \text{and} (n(S_i) = n(S_{i+1})) \text{and} (k2(S_i) = k2(S_{i+1})) \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned} .$$

Тобто, якщо у реченні зустрінуться два слова, набір характеристик кожного з яких відповідає одновимірному масиву $Z_i(S_i) = \{2,2,1,2,0,0,0,0\}$ та $Z_i(S_i) = \{1,2,1,2,0,0,0,0\}$ відповідно, то такі два слова є словосполученням. Прикладом для такого набору значень граматичних характеристик є словосполучення «інформаційної технології» - прикметник та іменник знаходяться в родовому відмінку однини, жіночого роду.

Відповідно міркування за правилом «*modus ponens*» можна відобразити такою схемою:

Антецедент1:

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 2) \text{and} (\text{cm}(S_{i+1}) = 1)) \text{and} (g(S_i) = g(S_{i+1})) \\ & \text{and} (n(S_i) = n(S_{i+1})) \text{and} (k2(S_i) = k2(S_{i+1})) \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

Антецедент2:

$$\begin{aligned} & ((\text{cm}(S_i) = 2) \text{and} (\text{cm}(S_{i+1}) = 1)) \text{and} (g(S_i) = g(S_{i+1})) \\ & \text{and} (n(S_i) = n(S_{i+1})) \text{and} (k2(S_i) = k2(S_{i+1})) \end{aligned}$$

$$\text{Консеквент}::: (S_j = S_i \cup S_{i+1})$$

2) Якщо два слова, що розглядаються, вжиті в одному відмінку, роді та числі, при чому перше з них – порядковий числівник, а друге – іменник, то слова утворюють словосполучення. Наприклад, «перша група».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 3) \text{and} (\text{cm}(S_{i+1}) = 1)) \text{and} (g(S_i) = g(S_{i+1})) \\ & \text{and} (n(S_i) = n(S_{i+1})) \text{and} (k2(S_i) = k2(S_{i+1})) \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned} .$$

3) Якщо два слова, що розглядаються, вжиті в одному відмінку, роді та числі, при чому перше з них – займенник, а друге – іменник, то слова утворюють словосполучення. Наприклад, «наш друг».

$$\begin{aligned} & \text{if}((cm(S_i) = 4) \text{ and } (cm(S_{i+1}) = 1)) \text{ and } (g(S_i) = g(S_{i+1})) \\ & \text{and } (n(S_i) = n(S_{i+1})) \text{ and } (k2(S_i) = k2(S_{i+1})) \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

4) Якщо два слова, що розглядаються, вжиті в одному відмінку, роді та числі, при чому перше з них – дієприкметник, а друге – іменник, то слова утворюють словосполучення. Наприклад, «*вирішена задача*».

$$\begin{aligned} & \text{if}((cm(S_i) = 6) \text{ and } (cm(S_{i+1}) = 1)) \text{ and } (g(S_i) = g(S_{i+1})) \\ & \text{and } (n(S_i) = n(S_{i+1})) \text{ and } (k2(S_i) = k2(S_{i+1})) \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

5) Якщо два слова, що розглядаються, вжиті в одному відмінку та числі, проте різного роду, при чому перше з них – іменник, а друге – іменник власна назва, то слова утворюють словосполучення. Наприклад, «*село Піски*».

$$\begin{aligned} & \text{if}((cm(S_i) = 1) \text{ and } (cm(S_{i+1}) = 1)) \text{ and } (g(S_i) = g(S_{i+1})) \\ & \text{and } (n(S_i) = n(S_{i+1})) \text{ then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

6) Якщо перше слово – дієслово, а друге – іменник (як правило) у родовому або знахідному (обов'язково не в називному) відмінку, то слова утворюють словосполучення. Наприклад, «*написати підручник*».

$$\begin{aligned} & \text{if}((cm(S_i) = 5) \text{ and } (cm(S_{i+1}) = 1)) \text{ and} \\ & ((g(S_{i+1}) = 2) \vee (g(S_{i+1}) = 4)) \vee (g(S_{i+1}) \neq 1). \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

7) Якщо перше слово – дієслово, друге – прийменник, а третє – іменник не в називному відмінку, то слова утворюють словосполучення. Наприклад, «*працює в університеті*».

$$\begin{aligned} & \text{if}((cm(S_i) = 5) \text{ and } (cm(S_{i+1}) = 9) \text{ and } (cm(S_{i+2}) = 1)) \text{ and} \\ & (g(S_{i+2}) \neq 1) \text{ then}(S_j = S_i \cup S_{i+1} \cup S_{i+2}) \end{aligned}$$

8) Якщо перше слово – дієслово, друге – прийменник, а третє – особовий займенник не в називному відмінку, то слова утворюють словосполучення. Наприклад, «*прийшов до нас*».

$$\begin{aligned} & \text{if}((cm(S_i) = 5) \text{ and } (cm(S_{i+1}) = 9) \text{ and } (cm(S_{i+2}) = 4)) \text{ and} \\ & (g(S_{i+2}) \neq 1) \text{ then}(S_j = S_i \cup S_{i+1} \cup S_{i+2}) \end{aligned}$$

9) Якщо перше слово – дієслово, а друге – особовий займенник не в називному відмінку, то слова утворюють словосполучення. Наприклад, «*пізнати себе*».

$$\begin{aligned} & \text{if}((cm(S_i) = 5) \text{ and } (cm(S_{i+1}) = 4)) \text{ and} \\ & (g(S_{i+1}) \neq 1) \text{ then } (S_j = S_i \cup S_{i+1}) \end{aligned} \quad \cdot$$

10) Якщо перше слово – дієприслівник, а друге – іменник (як правило) у родовому або знахідному (обов'язково не в називному) відмінку, то слова утворюють словосполучення. Наприклад, «*прочитавши книгу*».

$$\begin{aligned} & \text{if}((cm(S_i) = 7) \text{ and } (cm(S_{i+1}) = 1)) \text{ and} \\ & ((g(S_{i+1}) = 2) \vee (g(S_{i+1}) = 4)) \vee (g(S_{i+1}) \neq 1). \\ & \text{then } (S_j = S_i \cup S_{i+1}) \end{aligned}$$

11) Якщо перше слово – дієприслівник, а друге – особовий займенник не в називному відмінку, то слова утворюють словосполучення. Наприклад, «*запросивши декого*».

$$\begin{aligned} & \text{if}((cm(S_i) = 7) \text{ and } (cm(S_{i+1}) = 4)) \text{ and} \\ & (g(S_{i+1}) \neq 1) \text{ then } (S_j = S_i \cup S_{i+1}) \end{aligned} \quad \cdot$$

12) Якщо перше слово – іменник, а друге – іменник (як правило) у родовому або знахідному (обов'язково не в називному) відмінку, то слова утворюють словосполучення. Наприклад, «*кадр фільму*».

$$\begin{aligned} & \text{if}((cm(S_i) = 1) \text{ and } (cm(S_{i+1}) = 1)) \text{ and} \\ & ((g(S_{i+1}) = 2) \vee (g(S_{i+1}) = 4)) \vee (g(S_{i+1}) \neq 1). \\ & \text{then } (S_j = S_i \cup S_{i+1}) \end{aligned}$$

13) Якщо перше слово – іменник, друге – приєдник, а третє – іменник (як правило) у родовому або знахідному (обов'язково не в називному) відмінку, то слова утворюють словосполучення. Наприклад, «*лист з дому*».

$$\begin{aligned} & \text{if}((cm(S_i) = 1) \text{ and } (cm(S_{i+1}) = 9) \text{ and } (cm(S_{i+2}) = 1)) \text{ and} \\ & ((g(S_{i+2}) = 2) \vee (g(S_{i+2}) = 4)) \vee (g(S_{i+2}) \neq 1) \\ & \text{then } (S_j = S_i \cup S_{i+1} \cup S_{i+2}) \end{aligned} \quad \cdot$$

14) Якщо перше слово – числівник, а друге – іменник (як правило) у родовому або знахідному (обов'язково не в називному)

відмінку, то слова утворюють словосполучення. Наприклад, «п'ять книг».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 3) \text{ and } (\text{cm}(S_{i+1}) = 1)) \text{ and} \\ & ((g(S_{i+1}) = 2) \vee (g(S_{i+1}) = 4)) \vee (g(S_{i+1}) \neq 1). \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

15) Якщо перше слово – числівник, друге – числівник, а третє – іменник (як правило) у родовому або знахідному (обов'язково не в називному) відмінку, то слова утворюють словосполучення. Наприклад, «одна четверта семестру».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 3) \text{ and } (\text{cm}(S_{i+1}) = 3) \text{ and } (\text{cm}(S_{i+2}) = 1)) \text{ and} \\ & ((g(S_{i+2}) = 2) \vee (g(S_{i+2}) = 4)) \vee (g(S_{i+2}) \neq 1) \\ & \text{then}(S_j = S_i \cup S_{i+1} \cup S_{i+2}) \end{aligned}$$

16) Якщо перше слово – прикметник, друге – прийменник, а третє – іменник (як правило) у родовому або знахідному (обов'язково не в називному) відмінку, то слова утворюють словосполучення. Наприклад, «широкий у плечах».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 2) \text{ and } (\text{cm}(S_{i+1}) = 9) \text{ and } (\text{cm}(S_{i+2}) = 1)) \text{ and} \\ & ((g(S_{i+2}) = 2) \vee (g(S_{i+2}) = 4)) \vee (g(S_{i+2}) \neq 1) \\ & \text{then}(S_j = S_i \cup S_{i+1} \cup S_{i+2}) \end{aligned}$$

17) Якщо перше слово – прикметник, друге – прийменник, а третє – особовий займенник (як правило) у родовому або знахідному (обов'язково не в називному) відмінку, то слова утворюють словосполучення. Наприклад, «кращий за інших».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 2) \text{ and } (\text{cm}(S_{i+1}) = 9) \text{ and } (\text{cm}(S_{i+2}) = 4)) \text{ and} \\ & ((g(S_{i+2}) = 2) \vee (g(S_{i+2}) = 4)) \vee (g(S_{i+2}) \neq 1) \\ & \text{then}(S_j = S_i \cup S_{i+1} \cup S_{i+2}) \end{aligned}$$

18) Якщо перше слово – числівник, друге – прийменник, а третє – іменник (як правило) у родовому або знахідному (обов'язково не в називному) відмінку, то слова утворюють словосполучення. Наприклад, «троє зі студентів».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 3) \text{ and } (\text{cm}(S_{i+1}) = 9) \text{ and } (\text{cm}(S_{i+2}) = 1)) \text{ and} \\ & ((g(S_{i+2}) = 2) \vee (g(S_{i+2}) = 4)) \vee (g(S_{i+2}) \neq 1) \\ & \text{then}(S_j = S_i \cup S_{i+1} \cup S_{i+2}) \end{aligned}$$

19) Якщо перше слово – числівник, друге – прийменник, а третє – особовий займенник (як правило) у родовому або знахідному (обов'язково не в називному) відмінку, то слова утворюють словосполучення. Наприклад, «*двоє з нас*».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 3) \text{ and } (\text{cm}(S_{i+1}) = 9) \text{ and } (\text{cm}(S_{i+2}) = 4)) \text{ and} \\ & ((g(S_{i+2}) = 2) \vee (g(S_{i+2}) = 4)) \vee (g(S_{i+2}) \neq 1) \\ & \text{then}(S_j = S_i \cup S_{i+1} \cup S_{i+2}) \end{aligned}$$

20) Якщо перше слово – дієслово, а друге – прислівник, то вони утворюють словосполучення. Наприклад, «*робити добре*».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 5) \text{ and } (\text{cm}(S_{i+1}) = 8)) \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

21) Якщо перше слово – іменник, а друге – дієслово в неозначеній формі, то вони утворюють словосполучення. Наприклад, «*бажання вчитися*».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 1) \text{ and } (\text{cm}(S_{i+1}) = 5)) \text{ and } (h(S_{i+1}) = 0) \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

22) Якщо перше слово – дієслово, а друге – дієприслівник, то вони утворюють словосполучення. Наприклад, «*сидів думаючи*».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 5) \text{ and } (\text{cm}(S_{i+1}) = 7)) \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

23) Якщо перше слово – прислівник, а друге – прикметник, то вони утворюють словосполучення. Наприклад, «*дуже гарний*».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 8) \text{ and } (\text{cm}(S_{i+1}) = 2)) \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

24) Якщо перше слово – дієприкметник, а друге – прислівник, то вони утворюють словосполучення. Наприклад, «*дозволений недавно*».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 6) \text{ and } (\text{cm}(S_{i+1}) = 8)) \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

25) Якщо перше слово – дієприслівник, а друге – прислівник, то вони утворюють словосполучення. Наприклад, «*піднявши догори*».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 7) \text{ and } (\text{cm}(S_{i+1}) = 8)) \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

26) Якщо перше слово – прислівник, а друге – також прислівник, то вони утворюють словосполучення. Наприклад, «*дуже добре*».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 8) \text{ and } (\text{cm}(S_{i+1}) = 8)) \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

27) Якщо перше слово – дієприкметник, друге – прийменник, а третє – іменник не в називному відмінку, то слова утворюють словосполучення. Наприклад, «*втілений у життя*».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 6) \text{ and } (\text{cm}(S_{i+1}) = 9) \text{ and } (\text{cm}(S_{i+2}) = 1)) \text{ and} \\ & (g(S_{i+2}) \neq 1) \text{ then}(S_j = S_i \cup S_{i+1} \cup S_{i+2}) \end{aligned}$$

28) Якщо перше слово – дієприкметник, друге – прийменник, а третє – особовий займенник, то слова утворюють словосполучення. Наприклад, «*звернений до всіх*».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 6) \text{ and } (\text{cm}(S_{i+1}) = 9) \text{ and } (\text{cm}(S_{i+2}) = 4)) \text{ and} \\ & (g(S_{i+2}) \neq 1) \text{ then}(S_j = S_i \cup S_{i+1} \cup S_{i+2}) \end{aligned}$$

29) Якщо перше слово – дієприслівник, друге – прийменник, а третє – іменник не в називному відмінку, то слова утворюють словосполучення. Наприклад, «*їдучи до столиці*».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 7) \text{ and } (\text{cm}(S_{i+1}) = 9) \text{ and } (\text{cm}(S_{i+2}) = 1)) \text{ and} \\ & (g(S_{i+2}) \neq 1) \text{ then}(S_j = S_i \cup S_{i+1} \cup S_{i+2}) \end{aligned}$$

30) Якщо перше слово – дієслово, а друге – дієслово в неозначеній формі, то вони утворюють словосполучення. Наприклад, «*сів відпочити*».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 5) \text{ and } (\text{cm}(S_{i+1}) = 5)) \text{ and } (h(S_{i+1}) = 0) \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

31) Якщо перше слово – дієприкметник, а друге – дієслово в неозначеній формі, то вони утворюють словосполучення. Наприклад, «*запрограмований розпізнавати*».

$$\begin{aligned} & \text{if}((\text{cm}(S_i) = 6) \text{ and } (\text{cm}(S_{i+1}) = 5)) \text{ and } (h(S_{i+1}) = 0) \\ & \text{then}(S_j = S_i \cup S_{i+1}) \end{aligned}$$

32) Якщо перше слово – прислівник, друге – прийменник, а третє – іменник не в називному відмінку, то слова утворюють словосполучення. Наприклад, «далеко від університету».

$$\text{if}((\text{cm}(S_i) = 8) \text{ and } (\text{cm}(S_{i+1}) = 9) \text{ and } (\text{cm}(S_{i+2}) = 1)) \text{ and} \\ (\text{g}(S_{i+2}) \neq 1) \text{ then } (S_j = S_i \cup S_{i+1} \cup S_{i+2})$$

Також для української мови створена система продукцій, що містить двадцять два правила визначення граматичної основи речень [4].

1) Правило трактується наступним чином: якщо слово – іменник у називному відмінку, то це слово – підмет. Наприклад, у реченні «**Людина** владною рукою стукала в двері космосу».

$$\text{if}((\text{cm}(S_i) = 1) \text{ and } (\text{g}(S_i) = 1) \\ \text{then } (\text{ch}(S_i) = 1)$$

Тобто, якщо у реченні слово має такий набір характеристик $Z_i(S_i) = \{1, 1, n_i, k2_i, t_i, h_i, l_i, ch_i\}$, то останній елемент одномірного масиву буде рівний одиниці, тобто $Z_i(S_i) = \{1, 1, n_i, k2_i, t_i, h_i, l_i, 1\}$. Відповідно міркування за правилом «*modus ponens*» можна відобразити такою схемою:

Антецедент1:

$$\text{if}((\text{cm}(S_i) = 1) \text{ and } (\text{g}(S_i) = 1) \\ \text{then } (\text{ch}(S_i) = 1)$$

Антецедент2:

$$((\text{cm}(S_i) = 1) \text{ and } (\text{g}(S_i) = 1)$$

$$\text{Консеквент}:: (\text{ch}(S_i) = 1)$$

2) Якщо слово – особовий займенник у називному відмінку, то це слово – підмет. Наприклад, у реченні «**Він** владною рукою стукав у двері космосу».

$$\text{if}((\text{cm}(S_i) = 4) \text{ and } (\text{g}(S_i) = 1) \\ \text{then } (\text{ch}(S_i) = 1)$$

3) Якщо слово – прикметник у називному відмінку, то це слово – підмет. Наприклад, у реченні «У парламенті перемагають **праві**».

$$\begin{aligned} & \text{if}((cm(S_i) = 2) \text{ and } (g(S_i) = 1)) \\ & \text{then}(ch(S_i) = 1) \end{aligned}$$

4) Якщо слово – числівник у називному відмінку, то це слово – підмет. Наприклад, у реченні «**Обоє** мовчки озируються».

$$\begin{aligned} & \text{if}((cm(S_i) = 3) \text{ and } (g(S_i) = 1)) \\ & \text{then}(ch(S_i) = 1) \end{aligned}$$

5) Якщо слово – неозначена форма дієслова, наступний за ним символ у реченні – тире, а після нього у реченні зустрічається дієслово, то це слово – підмет. Наприклад, у реченні «**Жути** – значить працювати».

$$\begin{aligned} & \text{if}((cm(S_i) = 5) \text{ and } (h(S_i) = 0) \text{ and } (S_{i+1} = "-")) \\ & \text{and } (cm(S_{i+k}) = 5) \text{ then}(ch(S_i) = 1) \end{aligned}$$

де k – кількість слів між тире та знайденим дієсловом.

6) Якщо одне слово у реченні – іменник у називному відмінку, друге – прийменник «з», а третє – іменник в орудному відмінку, до того ж ці три слова є словосполученням, то таке словосполучення виступає підметом. Наприклад, у реченні «**Експерти з інженерами** вирішили проводити більше експериментів».

$$\begin{aligned} & \text{if}((cm(S_i) = 1) \text{ and } (cm(S_{i+1}) = 9) \text{ and } (cm(S_{i+2}) = 1)) \text{ and} \\ & (g(S_i) = 1) \text{ and } (g(S_{i+2}) = 5) \text{ and } (S_{i+1} = "з") \text{ and} \\ & (S_j = S_i \cup S_{i+1} \cup S_{i+2}) \\ & \text{then}((ch(S_i) = 1) \text{ and } (ch(S_{i+1}) = 1) \text{ and } (ch(S_{i+2}) = 1)) \end{aligned}$$

7) Якщо одне слово у реченні – особовий займенник у називному відмінку, друге – прийменник «з», а третє – іменник в орудному відмінку, до того ж ці три слова є словосполученням, то таке словосполучення виступає підметом. Наприклад, у реченні «**Вони з інженерами** вирішили проводити більше експериментів».

$$\begin{aligned} & \text{if}((cm(S_i) = 3) \text{ and } (cm(S_{i+1}) = 9) \text{ and } (cm(S_{i+2}) = 1)) \text{ and} \\ & (g(S_i) = 1) \text{ and } (g(S_{i+2}) = 5) \text{ and } (S_{i+1} = "з") \text{ and} \\ & (S_j = S_i \cup S_{i+1} \cup S_{i+2}) \\ & \text{then}((ch(S_i) = 1) \text{ and } (ch(S_{i+1}) = 1) \text{ and } (ch(S_{i+2}) = 1)) \end{aligned}$$

8) Якщо одне слово у реченні – числівник у називному відмінку або одне із прислівників «мало», «чимало», «багато», друге – іменник у родовому відмінку множини, то такі два слова є підметом. Наприклад, у реченні «**Десятеро студентів** пройшли строем», «**Багато студентів** отримали двійки».

$$\begin{aligned} & \text{if} \left((cm(S_i) = 4) \vee (S_i = \text{"мало"}) \vee (S_i = \text{"чимало"}) \right) \text{and} \\ & \left(\vee (S_i = \text{"багато"}) \text{and} (cm(S_{i+1}) = 1) \right) \\ & (g(S_i) = 1) \text{and} (g(S_{i+1}) = 2) \text{and} (n(S_{i+1}) = 2) \\ & \text{then} ((ch(S_i) = 1) \text{and} (ch(S_{i+1}) = 1)) \end{aligned}$$

9) Якщо одне слово у реченні – числівник у називному відмінку або одне із прислівників «мало», «чимало», «багато», друге – займенник у родовому відмінку множини, то такі два слова є підметом. Наприклад, у реченні «**П'ятеро їх** не подужали б жодного з нас», «**Чимало наших** пішли до кафе».

$$\begin{aligned} & \text{if} \left((cm(S_i) = 4) \vee (S_i = \text{"мало"}) \vee (S_i = \text{"чимало"}) \right) \text{and} \\ & \left(\vee (S_i = \text{"багато"}) \text{and} (cm(S_{i+1}) = 3) \right) \\ & (g(S_i) = 1) \text{and} (g(S_{i+1}) = 2) \text{and} (n(S_{i+1}) = 2) \\ & \text{then} ((ch(S_i) = 1) \text{and} (ch(S_{i+1}) = 1)) \end{aligned}$$

10) Якщо одне слово у реченні – числівник у називному відмінку або одне із прислівників «мало», «чимало», «багато», друге – прикметник у родовому відмінку множини, то такі два слова є підметом. Наприклад, у реченні «**П'ятеро сильних** не подужали б жодного з нас», «**Мало хоробрих** виявилось серед присутніх».

$$\begin{aligned} & \text{if} \left((cm(S_i) = 4) \vee (S_i = \text{"мало"}) \vee (S_i = \text{"чимало"}) \right) \text{and} \\ & \left(\vee (S_i = \text{"багато"}) \text{and} (cm(S_{i+1}) = 2) \right) \\ & (g(S_i) = 1) \text{and} (g(S_{i+1}) = 2) \text{and} (n(S_{i+1}) = 2) \\ & \text{then} ((ch(S_i) = 1) \text{and} (ch(S_{i+1}) = 1)) \end{aligned}$$

11) Якщо одне слово у реченні – іменник у називному відмінку, друге – іменник у родовому відмінку множини, то такі два слова є підметом. Наприклад, у реченні «**Половина присутніх** погодилася на екскурсію».

$$\begin{aligned} & \text{if}((cm(S_i) = 1) \text{ and } (cm(S_{i+1}) = 1)) \text{ and} \\ & (g(S_i) = 1) \text{ and } (g(S_{i+1}) = 2) \\ & \text{then} ((ch(S_i) = 1) \text{ and } (ch(S_{i+1}) = 1)) \end{aligned}$$

12) Якщо одне слово у реченні – іменник у називному відмінку, друге – займенник у родовому відмінку множини, то такі два слова є підметом. Наприклад, у реченні «**Чверть наших** залишилися до завтра».

$$\begin{aligned} & \text{if}((cm(S_i) = 1) \text{ and } (cm(S_{i+1}) = 3)) \text{ and} \\ & (g(S_i) = 1) \text{ and } (g(S_{i+1}) = 2) \text{ and } (n(S_{i+1}) = 2) . \\ & \text{then} ((ch(S_i) = 1) \text{ and } (ch(S_{i+1}) = 1)) \end{aligned}$$

13) Якщо одне слово у реченні – іменник у називному відмінку, друге – прикметник у родовому відмінку множини, то такі два слова є підметом. Наприклад, у реченні «**Половина сміливих** не пішли б до того лісу».

$$\begin{aligned} & \text{if}((cm(S_i) = 1) \text{ and } (cm(S_{i+1}) = 2)) \text{ and} \\ & (g(S_i) = 1) \text{ and } (g(S_{i+1}) = 2) \text{ and } (n(S_{i+1}) = 2) . \\ & \text{then} ((ch(S_i) = 1) \text{ and } (ch(S_{i+1}) = 1)) \end{aligned}$$

14) Якщо одне слово у реченні – займенник «кожен», «дехто», «хтось» і т.д. у називному відмінку, друге – прийменник, а третє – іменник у родовому відмінку множини, то ці три слова є підметом. Наприклад, у реченні «**Хто з експертів** не знає цього».

$$\begin{aligned} & \text{if}((cm(S_i) = 3) \text{ and } (cm(S_{i+1}) = 9) \text{ and } (cm(S_{i+2}) = 1)) \text{ and} \\ & (g(S_i) = 1) \text{ and } (g(S_{i+2}) = 2) \text{ and } (S_j = S_i \cup S_{i+1} \cup S_{i+2}) . \\ & \text{then} ((ch(S_i) = 1) \text{ and } (ch(S_{i+1}) = 1) \text{ and } (ch(S_{i+2}) = 1)) \end{aligned}$$

15) Якщо одне слово у реченні – займенник «кожен», «дехто», «хтось» і т.д. у називному відмінку, друге – прийменник, а третє – займенник у родовому відмінку множини, то ці три слова є підметом. Наприклад, у реченні «**Хто з нас** не знає цього».

$$\begin{aligned} & \text{if}((cm(S_i) = 3) \text{ and } (cm(S_{i+1}) = 9) \text{ and } (cm(S_{i+2}) = 1)) \text{ and} \\ & (g(S_i) = 1) \text{ and } (g(S_{i+2}) = 2) \text{ and } (S_j = S_i \cup S_{i+1} \cup S_{i+2}) . \\ & \text{then} ((ch(S_i) = 1) \text{ and } (ch(S_{i+1}) = 1) \text{ and } (ch(S_{i+2}) = 1)) \end{aligned}$$

16) Якщо слово – дієслово у дійсному (теперішній, минулий або майбутній час), умовному або наказовому способі, то це

слово – присудок. Наприклад, у реченні «Усе навкруги **дзвеніло і мінлося барвами**».

$$\text{if}((\text{cm}(S_i) = 5)) \text{and}((h(S_i) = 1) \vee (h(S_i) = 2) \vee (h(S_i) = 3)) \\ \text{then}(\text{ch}(S_i) = 2)$$

17) Якщо одне слово у реченні – дієслово, а друге дієслово у неозначеній формі, то такі слова виступають у реченні присудком. Наприклад, у реченні «Він **вміє розв'язувати і першу і другу задачі**», «Багато добра **повинен давати** маленькій людині педагог».

$$\text{if}((\text{cm}(S_i) = 5) \text{and}(\text{cm}(S_{i+1}) = 5)) \text{and} (h(S_{i+1}) = 0) \\ \text{then} ((\text{ch}(S_i) = 2) \text{and}(\text{ch}(S_{i+1}) = 2))$$

18) Якщо у реченні зустрілося тире або дієслово «бути» у всіх способах, а наступне після нього слово – іменник у називному або орудному відмінку, то це слово – присудок. Наприклад, у реченні «Всесвіт наш – **казка**».

$$\text{if}(((S_i = "-") \vee (S_i = "бути"))) \text{and}(\text{cm}(S_{i+1}) = 1)) \text{and} \\ (g(S_{i+1}) = 1) \vee (g(S_{i+1}) = 5) \\ \text{then} ((\text{ch}(S_i) = 2) \text{and}(\text{ch}(S_{i+1}) = 2))$$

19) Якщо у реченні зустрілося тире або дієслово «бути» у всіх способах, а наступне після нього слово – прикметник у називному або орудному відмінку, то це слово – присудок. Наприклад, у реченні «Руки **білі, а сумління – чорне**».

$$\text{if}(((S_i = "-") \vee (S_i = "бути"))) \text{and}(\text{cm}(S_{i+1}) = 2)) \text{and} \\ (g(S_{i+1}) = 1) \vee (g(S_{i+1}) = 5) \\ \text{then} ((\text{ch}(S_i) = 2) \text{and}(\text{ch}(S_{i+1}) = 2))$$

20) Якщо у реченні зустрілося тире або дієслово «бути» у всіх способах, а наступне після нього слово – займенник у називному або орудному відмінку, то це слово – присудок. Наприклад, у реченні «А ви **хто?**».

$$\text{if}(((S_i = "-") \vee (S_i = "бути"))) \text{and}(\text{cm}(S_{i+1}) = 3)) \text{and} \\ (g(S_{i+1}) = 1) \vee (g(S_{i+1}) = 5) \\ \text{then} ((\text{ch}(S_i) = 2) \text{and}(\text{ch}(S_{i+1}) = 2))$$

21) Якщо у реченні зустрілося тире або дієслово «бути» у всіх способах, а наступне після нього слово – числівник у називному

або орудному відмінку, то це слово – присудок. Наприклад, у реченні «Він був **п'ятим** у черзі» .

$$\begin{aligned} & \text{if}(((S_i = \text{"-"}) \vee (S_i = \text{"бути"})) \text{and} (cm(S_{i+1}) = 4)) \text{and} \\ & (g(S_{i+1}) = 1) \vee (g(S_{i+1}) = 5) \\ & \text{then} ((ch(S_i) = 2) \text{and} (ch(S_{i+1}) = 2)) \end{aligned}$$

22) Якщо у реченні зустрілося тире або дієслово «бути» у всіх способах, а наступне після нього слово – дієприкметник у називному або орудному відмінку, то це слово – присудок. Наприклад, у реченні «Тепер їхні імена **знані**» .

$$\begin{aligned} & \text{if}(((S_i = \text{"-"}) \vee (S_i = \text{"бути"})) \text{and} (cm(S_{i+1}) = 6)) \text{and} \\ & (g(S_{i+1}) = 1) \vee (g(S_{i+1}) = 5) \\ & \text{then} ((ch(S_i) = 2) \text{and} (ch(S_{i+1}) = 2)) \end{aligned}$$

Таким чином, виявлення граматичних та синтаксичних зв'язків між усіма словами речення дає можливість визначити семантичні відношення між ними. Для простеження семантичних зв'язків між реченнями природної мови, що входять до певного текстового фрагменту, на кожне з вище перерахованих правил бази знань накладаються обмеження у вигляді додаткових умов. Тобто за рахунок збільшення кількості антецедентів у правилах можна виявити логічні зв'язки між однорідними членами речення, знайти взаємопов'язані слова у складних реченнях та ін.

ІІІ. МЕТОДОЛОГІЯ АВТОМАТИЗОВАНОЇ ПОБУДОВИ ЛОГІКО-ЛІНГВІСТИЧНИХ МОДЕЛЕЙ ТЕКСТОВОЇ ІНФОРМАЦІЇ

Формальні системи представляють собою системи операцій над об'єктами, що розуміються як послідовності символів (тобто як слова в фіксованих алфавітах), самі операції також є операціями над символами. Термін «формальні» підкреслює, що об'єкти та операції над ними розглядаються формально, без змістовної інтерпретації символів. Припускається, що між символами не існує ніяких зв'язків та відношень крім тих, що явно описані засобами самої формальної системи [65].

Мова символічної логіки – це універсальний, обов'язковий інструмент систематизації та аналізу змісту тексту, представленого природною мовою. Проте символічна логіка створювалася математиками і була орієнтована на аналіз математичних понять та тверджень. Складності, які було виявлено при зіткненні з реальним мовним матеріалом, привели до того, що прикладні логічні дослідження (тобто дослідження апарату логіки до вирішення мовних проблем поза математикою) у минулі десятиріччя були спрямовані переважно на розробку вузьких систем понять: модальна логіка, логіка оцінок, логіка причинності і т.д. Проте основне для практичних досліджень питання про формальне представлення мови як цілого – залишається за рамками локальних логічних систем [66].

Тому для того, щоб автоматизувати процес створення логіко-лінгвістичної моделі (2.1) – (2.4), тобто досягти можливості перетворення довільних речень природної мови у формули логіки предикатів, необхідно представити речення у вигляді формальної системи, яка дасть змогу сприймати будь-яку флективну мову як єдине ціле не залежно від предметної області, до якої належить той чи інший текст.

Мислення не можливе поза мовою. Всі явища зовнішнього світу знаходять відображення у вигляді категорій представлення. Здійснюючи роботу з мовними структурами, можна змінювати поточний стан відображеного зовнішнього світу. Успіх у формалізації розуміння змісту природної мови досягнутий на шляху використання аналогії між математичними та природними мовами [24].

Внутрішнє представлення фрагменту реальності представляє собою ситуацію, яка виступає у ролі імітаційної моделі і є суб'єктивною. Множина можливих фрагментів реальності відповідає множині внутрішніх представлень – ситуації. Тому можливо говорити про існування деякої функції, яка перетворює фрагмент реальності у ситуацію, тобто процес сприйняття. Вхідні аргументи цієї функції мають область визначення у вигляді множини фрагментів реальності і називаються ситуаційними змінними. Як значення змінних виступає інформація, що вилучається з фрагменту дійсності. Вихідні значення функції сприймання мають область визначення у вигляді множини різноманітних ситуацій [56].

Процес розуміння змісту тексту на природній мові також можна розглядати як обчислення деякої функції перетворення фрагменту реальності – тексту – у внутрішню модель – ситуацію, а надалі у зміст тексту.

В алгебраїчній логіці висувається гіпотеза про те, що речення природної мови представляють собою формули деякої алгебри. Математичні тексти теж пишуться у вигляді окремих речень і називаються висловлюваннями. Якщо така гіпотеза підтвердиться, то стане можливим обробляти природну мову так само, як і математичні вирази. У математиці абсолютно точно встановлено, що зміст висловлювання – це предикат, який виражається цим висловлюванням, аргументами такого предикату є змінні, що присутні у даному висловлюванні [67].

Отже, текст на природній мові можна ототожнити з математичною формулою, записаною за певними правилами, а змістом речення природної мови є деякий предикат. Тоді у математиці, до якої належить формула, повинні бути присутні такі елементи, як змінні, операції над змінними, операції над операціями. Послідовність операцій над змінними та операціями буде записана у вигляді формули.

Структура речення виражається структурою формули предиката. Зміст речення виражається функцією, що реалізується за допомогою предикату. Два тексти різної структури зазвичай мають різну структуру у формулах предикатів, однак, якщо ці предикати виражають тотожно рівні функції, то і тексти виражають однаковий зміст. Тотожність функцій двох різних предикатів означає, що за будь-якої ситуації ці два предикати

мають тотожні значення. Речення складаються із слів, а слова з літер, тому з'являється можливість оперувати даними на різних лінгвістичних рівнях, використовуючи один математичний апарат. За допомогою алгебри кінцевих предикатів довільного порядку можна побудувати моделі словотворення і словозміни будь-якої флективної мови.

Таким чином, сприйняття – це функція, що полягає у перетворенні фрагменту дійсності до ситуаційного предикату, а розуміння – функція, що полягає в обчисленні значення змістовного предикату на основі його аргументу – ситуаційного предикату. Вхідним аргументом функції мислення є суб'єктивна модель зовнішнього світу – ситуація та змістовий предикат – зміст тексту, результатом – значення змістового предикату в даній ситуації. Синтаксис речення – структура формули предиката, а функція, що реалізується предикатом, – зміст цього речення [68].

Засобом формалізації текстової інформації, тобто математичною формулою за допомогою якої можна представити речення природної мови і вилучити з неї зміст, є логіко-лінгвістична модель, структура формули якої базується на синтаксичній структурі речення. Саме на твердженні «синтаксис речення – це структура формули предикату, а функція, що реалізується предикатом – зміст цього речення» ґрунтується ідея методу автоматизованого формування логіко-лінгвістичної моделі текстової інформації.

Метод автоматизованого формування логіко-лінгвістичних моделей полягає в автоматизації процесу перетворення синтаксичних конструкцій (речень) природної мови у формули логіки предикатів першого порядку. Метод дозволяє вилучити з вхідної текстової інформації знання, формалізовані у відповідності з певними структурними правилами, і які комп'ютер зможе автономно використовувати при вирішенні задач за наперед обраними алгоритмами.

Запропонована методологія дає можливість обробляти текстову інформацію не залежно від предметної області, що аналізується. Саме це забезпечує особливість методу автоматизованого формування логіко-лінгвістичних моделей і здійснює зв'язок методологічної бази лінгвістичних досліджень та принципів формування автоматизованих систем управління.

3.1 Загальний алгоритм методу автоматизованого формування логіко-лінгвістичних моделей текстової інформації

Метод автоматизованого формування логіко-лінгвістичних моделей включає у себе декілька етапів (рис. 3.1), кожен з яких представляє собою складний механізм роботи формальної системи, а її елементи відіграють важливу роль для вилучення знань із тестової інформації. В основу методу покладено відповідність між формулами логіки предикатів та концептами, що належать реальному світу. Ототожнивши вирази логіки предикатів зі словами природної мови згідно з формалізованими правилами визначення синтаксичних ролей (детально описані у розділі II), можна засобами мови предикатів відображати зміст тексту.

Механізмом для здійснення процесу перетворення речення у формули логіки предикатів служить автоматизований синтаксичний аналізатор, на основі результатів роботи якого формується логіко-лінгвістична модель.

Першим етапом методу є розбиття вхідної текстової інформації на словоформи і ототожнення їх з елементами формальної системи, що дозволить розглядати речення як систему, взаємозв'язаними елементами якої є слова та словосполучення, які, фактично, представляють собою набір символів, розділених пробілами та знаками пунктуації.

Розбиття тексту на словоформи та ідентифікація його з формальною системою надасть можливість розглядати речення як систему взаємозв'язаних елементів, правила взаємодії яких служать основою формування логіко-лінгвістичної моделі, а отже визначають зміст введеної текстової інформації.

Для того, щоб прослідкувати зв'язки між елементами формальної системи, необхідно володіти інформацією про властивості цих елементів, що і забезпечує етап концептуалізації. На даному кроці здійснюється морфологічний аналіз слів речення, тобто визначається, до якої частини мови належить те чи інше слово і якими граматичними значеннями воно володіє [69].

Морфологічний аналіз є підготовчим етапом перед роботою механізму синтаксичного аналізатору – засобу визначення структури речення.

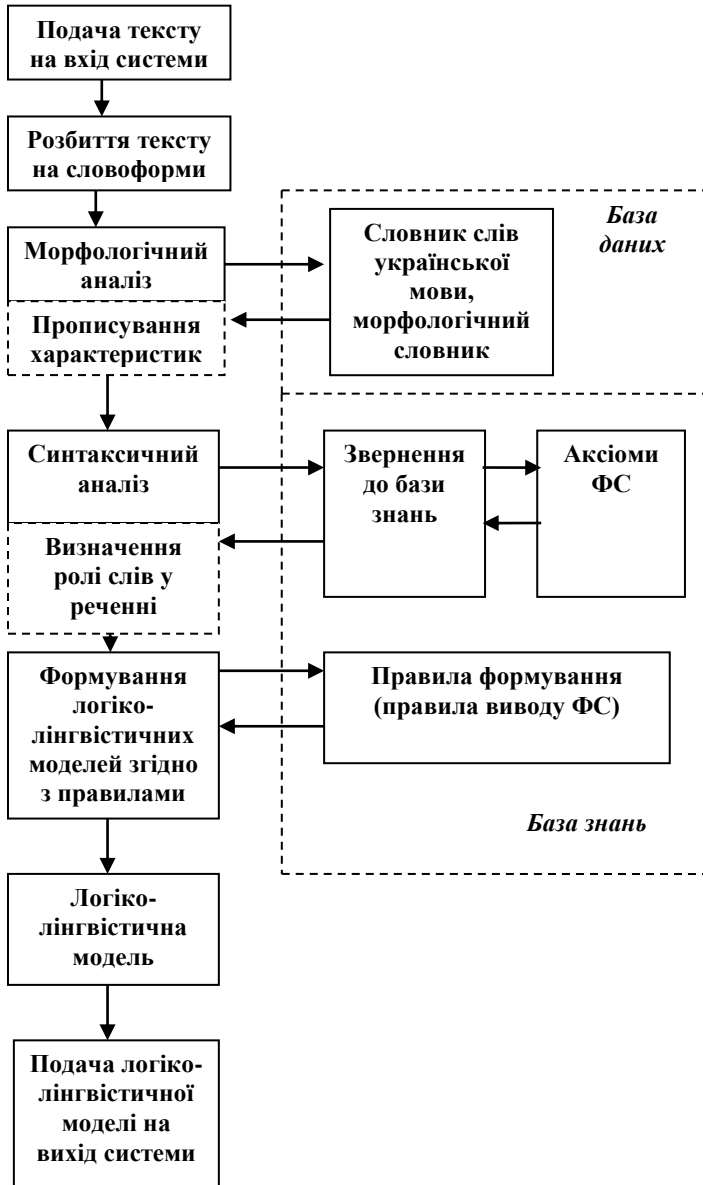


Рис. 3.1. Загальний алгоритм методу автоматизованого формування логіко-лінгвістичних моделей

Для визначення морфологічної та синтаксичної будови речення використовується база даних, що містить граматичний і морфологічний словники флективної мови, представлені у вигляді реляційних таблиць, та базу знань, яка інтерпретується як набір аксіом формальної системи і містить правила визначення синтаксичних ролей слів, правила пунктуації та правила формування логіко-лінгвістичних моделей.

Логіко-лінгвістична модель є засобом відображення відповідності між поняттями, що описуються термінами природної мови, та предикатними формулами. Ця відповідність базується на правилах формування логіко-лінгвістичних моделей згідно тих синтаксичних ролей та синтаксичних правил флективної мови, які були визначені на попередніх етапах методу.

За допомогою механізму логічного виводу в логіко-лінгвістичну модель підставляються слова вхідної текстової інформації відповідно до того, яку синтаксичну роль виконує дане слово у реченні (дія, об'єкт, суб'єкт, обставина, характеристика).

Таким чином, шляхом аналізу слів речення, знаків пунктуації, що до нього входять, та визначення його синтаксичної структури, визначається зміст речення, тобто те, про що у ньому йдеться, про які дії, суб'єкти та об'єкти, засобом вираження чого є логіко-лінгвістична модель, здатна відображати зміст тексту [44].

Фактична реалізація методу автоматизованого формування логіко-лінгвістичних моделей текстової інформації розуміє під собою сортування масиву слів конкретного речення у відповідності з чітко сформованими правилами.

Основною ідеєю цього перетворення є визначення відношень між підметом (суб'єктом) та присудком (предикатом), а також їх загальне відношення до того, що вони виражають в дійсності.

Проведення паралелі між структурою речення та структурою логічного висловлювання, визначення правил формування логіко-лінгвістичної моделі на основі синтаксичної структури речення, відображення в наборі аксіом, що складають логіко-лінгвістичну модель, всіх зв'язків між словоформами, дозволяє надалі аналізувати логіко-лінгвістичні моделі з метою відшукування протиріч у різноманітних текстах. Загальна схема (рис. 3.2) дає

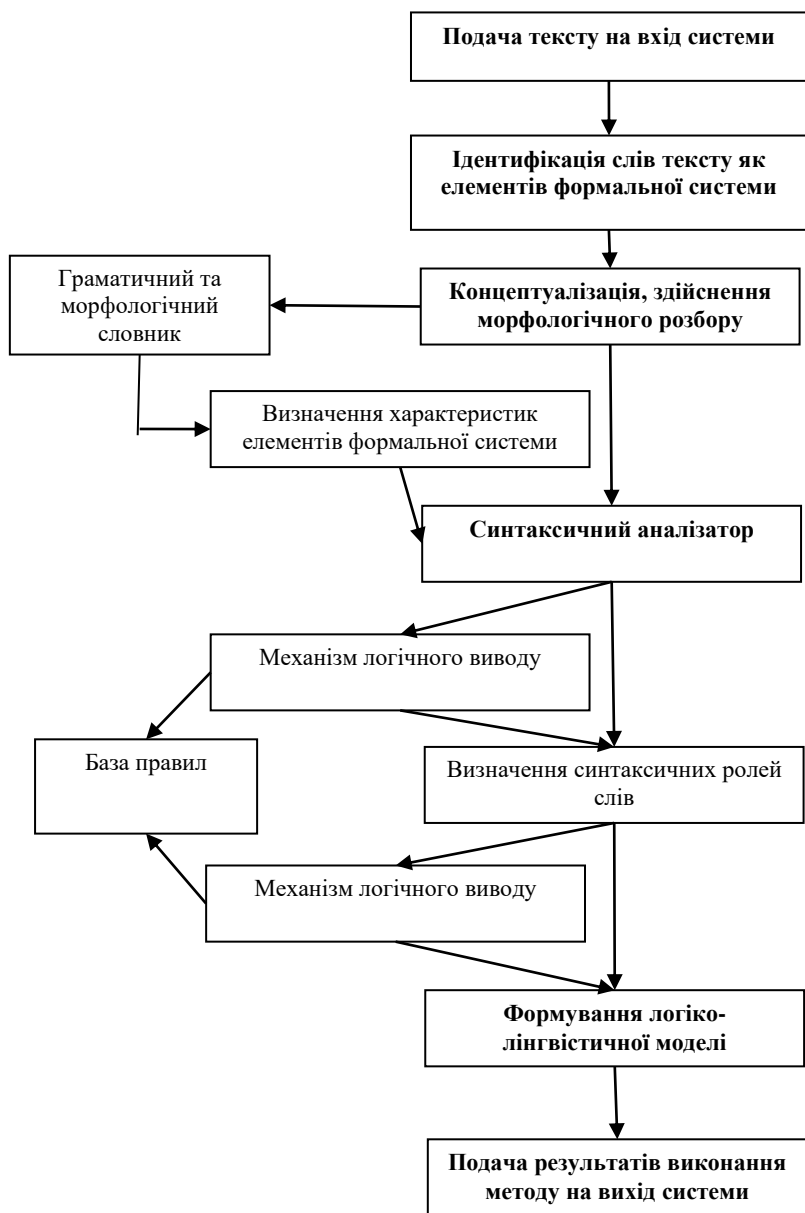


Рис. 3.2. Загальна схема виконання етапів методу

змогу ознайомитися з усіма етапами формування логіко-лінгвістичних моделей текстової інформації.

3.2 Ідентифікація вхідного тексту

Для вилучення знань із текстової інформації необхідно застосувати до вхідного тексту метод, який дозволив би розглядати кожне введене речення як систему взаємозв'язаних елементів з метою дослідити зв'язки між ними. Таким методом є аналіз. У логіці цей термін понятійний, тобто робить можливим утворення понять, розклад єдиного на множини, цілого – на частини, складного – на його компоненти, події – на її окремі ступені, змісту – на його елементи, поняття – на його ознаки.

Розбиття вхідного тексту S на словоформи відбувається за допомогою застосування логічного прийому розкладу за ознаками на складові частини, щоб дослідити текстову інформацію (речення) та зв'язки між складовими словоформами, а також вилучити знання (суть), які на виході системи будуть перетворені в логіко-лінгвістичні моделі.

Нехай речення представляє собою систему, що складається з елементів (словоформ): простих (слів) та складних (словосполучень, які автоматично включають у себе слова). Тоді об'єктом управління є складна організаційна система, що складається з r простих взаємодіючих елементів $S_i, i = \overline{1, r}$ та з m складних взаємодіючих елементів $S_j^i, i = \overline{1, r}, j = \overline{1, m}$ (рис.3.3). Зв'язки між елементами системи можна прослідкувати тільки після виконання наступних кроків алгоритму, так як для кожної конкретної системи зв'язки будуть різні [70].

$S_i, i = \overline{1, r}$ – прості елементи системи (слова), які можуть входити або не входити до структури складних елементів.

$S_j^i, i = \overline{1, r}, j = \overline{1, m}$ – складні елементи системи (словосполучення), головною складовою якого є простий елемент $S_i, i = \overline{1, r}$.

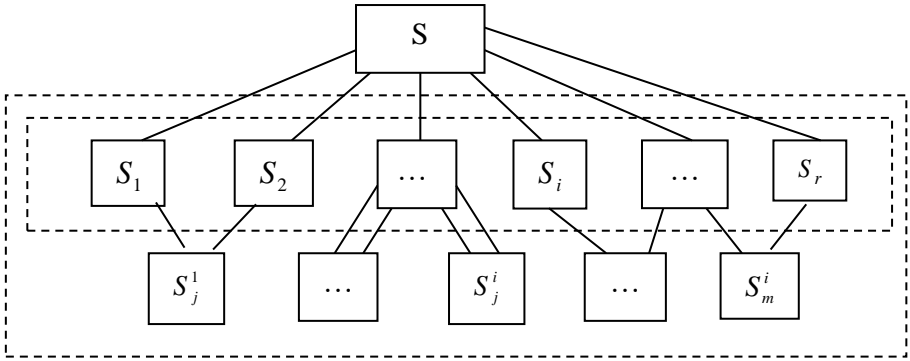


Рис. 3.3. Узагальнена структура формальної системи

Тобто на етапі ідентифікації вхідна текстова інформація, яка представляє собою сукупність синтаксичних одиниць природної мови, ототожнюється з формальною системою $S = \bigwedge_{i \in r} S_i = \bigwedge_{i \in r} \bigwedge_{j \in m} S_j^i$.

У даному випадку використана логічна зв'язка «кон'юнкція», яка для конкретного речення означає «сукупність слів», їх об'єднання. Речення ж розглядається як система із сукупності простих елементів (слів) та складних елементів (словосполучень), при чому простий елемент S_i може входити до складу як одного, так і декількох складних.

Таке перетворення дає змогу розглядати будь-яке речення (як завгодно складної структури) як систему операцій над послідовностями символів, а самі операції – формально, без змістовної інтерпретації.

Ідентифікація вхідного тексту, фактично, є тотожністю «розбиття вхідного тексту на словоформи». Цей етап необхідний для того, щоб ідентифікувати слова речення з простими елементами формальної системи. Здійснюється це шляхом перетворення речення в одновимірний масив, елементами якого є слова.

3.3 Концептуалізація, визначення характеристик елементів формальної системи

На даному етапі експліціюються ключові поняття, відношення і зв'язки між елементами, про які йшлося на етапі ідентифікації, а також визначаються характеристики цих елементів, необхідні для опису подальшого процесу розв'язання поставленої задачі [44].

Концептуалізація даних представляє собою процедуру введення онтологічних представлень у накопичений масив емпіричних даних; первинну теоретичну форму, що забезпечує теоретичну організацію матеріалу; схему зв'язку понять, що дозволяє продукувати гіпотези про природу та характер взаємозв'язків. На цьому етапі визначається спосіб організації подальшої роботи, що дозволить рухатися від первинних теоретичних концептів до більш абстрактних конструктів.

Кожен простий елемент системи описується вектором значень характеристик $z_i = (z_{k1}(S_i), k1 = \overline{1, w})$, де w – кількість граматичних характеристик i -го елемента системи, $i = \overline{1, r}$. Величину w визначено як сталу для всіх елементів формальної системи, так як кожне слово в українській мові має ряд морфологічних ознак, характерних для інших слів, наприклад, частина мови, відмінок, рід, число і т.д. Під час морфологічного розбору кожного слова, за відсутності певних морфологічних ознак у того чи іншого елемента, такі характеристики будуть замінені нулями, що дасть більше теоретичної інформації для подальшого встановлення взаємозв'язків між елементами у складних синтаксичних конструкціях. Тобто визначення характеристик елементів системи дає первинне концептуальне пояснення, що носить інтерпретаційний (розпізнавальний) характер.

Процес визначення характеристик кожного елемента формальної системи представляє собою концептуальну схему, що задає множину можливих способів роботи на теоретичному рівні, припущень про природу і властивості елементів, які досліджуються (рис. 3.4), в основу якого покладено механізм роботи морфологічного аналізатора.

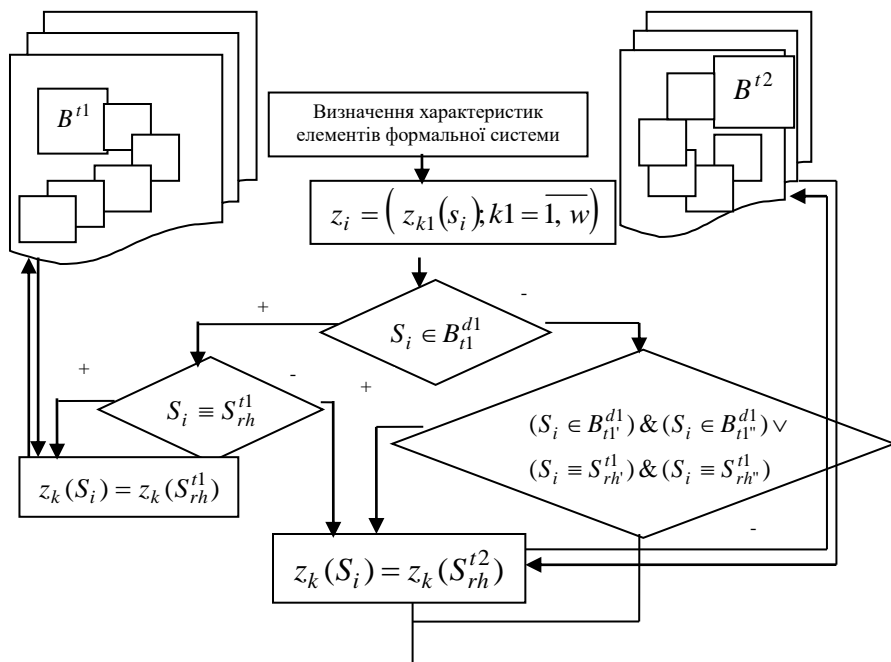


Рис. 3.4. Концептуальна схема для визначення характеристик кожного елементу системи

Де B^{d1} – множина слів бази даних;

B_{t1}^{d1} - множина слів таблиці бази даних, $B^{d1} \in B_{t1}^{d1}$;

B^{t1} - множина таблиць бази даних;

S_{rh}^{t1} - слово t_1 -ї таблиці із множини таблиць бази даних B^{t1} ;

B^{d2} - множина слів бази даних;

B_{t2}^{d2} - множина слів таблиці бази даних, $B^{d2} \in B_{t2}^{d2}$;

B^{t2} - множина таблиць бази даних;

S_{rh}^{t2} - слово t_2 -ї таблиці із множини таблиць бази даних B^{t2} .

B^{d1} представляє собою словозмінний словник, який організований у вигляді реляційної таблиці бази даних. Умова $S_i \in B_{t1}^{d1}$ перевіряє, чи зустрічається конкретний елемент серед

множини значень B_{rl}^{d1} . Основою побудови такого граматичного словника B^{d1} є наявність формальної моделі словозміни, що означає установлення та формалізацію лінгвістичних критеріїв, відповідно до яких вся множина слів природної мови розбивається на певні підмножини. Взаємний перетин підмножин є порожнім, в середині нього словозміна відбувається за єдиним правилом (алгоритмом).

Отже, якщо елемент системи $S_i, i = \overline{1, r}$ однозначно ідентифіковано зі словом із граматичного словника, тобто ідентичність кількості та порядку символів у слові S_i та S_{rh}^{t1} відповідно, що і перевіряє умова $S_i = S_{rh}^{t1}$, то всі граматичні характеристики слова $z_k(S_{rh}^{t2})$ бази даних B^{d1} присвоюються характеристикам елементу системи $z_k(S_i)$, тобто $z_k(S_i) = z_k(S_{rh}^{t1})$.

Якщо слово S_i не ідентифіковано, тобто не вдалося точно визначити, в якому стовпчику і полі граматичного словника знаходиться конкретний елемент формальної системи, то перевіряється умова

$$(S_i \in B_{rl}^{d1}) \& (S_i \in B_{rl}^{d1}) \vee (S_i \equiv S_{rh}^{t1}) \& (S_i \equiv S_{rh}^{t1}).$$

Ідентичність $S_i \equiv S_{rh}^{t1}$ може бути не встановлена через ряд причин.

Перша частина умови $(S_i \in B_{rl}^{d1}) \& (S_i \in B_{rl}^{d1})$ перевіряє, чи належить слово декільком таблицям бази даних B^{d1} . Це можливо у зв'язку з омонімією флективної мови, тому таке слово неможливо одразу ідентифікувати, а необхідно прослідкувати, у контексті з якими словами воно вживається, дослідити більш детально його граматичні значення.

Друга частина умови $(S_i \equiv S_{rh}^{t1}) \& (S_i \equiv S_{rh}^{t1})$ перевіряє, чи не зустрічається декілька ідентичних слів в одній і тій самій таблиці B^{t1} . Наприклад, слово «науковий» та «наукове» у родовому відмінку множини будуть мати однакову форму «наукового» та «наукового». Так як обидва слова є прикметниками, то вони будуть знаходитися в одній таблиці prk , але у різних стовпчиках і полях.

Для вирішення цієї проблеми відбувається звернення до бази даних V^{d2} , яка представляє собою морфологічний словник флективної мови, в якому перераховані всі формальні ознаки, за допомогою яких весь масив лексики природної мови поділяється на підмножини слів з відповідними властивостями – ці підмножини носять назву словозмінних парадигматичних класів.

Граматики будь-якої природної мови вивчає закономірності розвитку та функціонування граматичних одиниць, формує систему правил про те, як необхідно ними користуватися і як з морфем формувати слова та змінювати їх, шукати необхідні форми та об'єднувати їх в словосполучення, як знаходити для слів та словосполучень єдиний правильний порядок формування речень для вираження думки. Мовні одиниці характеризуються у граматиці тим, що мають граматичне значення. Для того, щоб зрозуміти природу граматичного значення потрібно розглядати його у співставленні з предметами або явищами реального світу, з поняттями та їх назвами. Граматичне значення слів виражається у їх граматичних формах, які будуть називатися граматичними характеристиками елемента формальної системи.

$z_i = (z_{k1}(S_i), k1 = \overline{1, w})$. Між граматичним значенням та граматичною формою існує діалектичний взаємозумовлений зв'язок, як між змістом і формою [44].

Фактично, на етапі концептуалізації здійснюється морфологічний розбір слів введеної текстової інформації. У результаті морфологічного аналізу «прописуються» характеристики кожного елемента формальної системи (кожного слова), змінна $z_{k1}(S_i)$ є граматичною категорією слова флективної мови. Морфологічний розбір передбачає визначення характеристик слова як частини мови.

Не всі слова природної мови мають однакові граматичні характеристики, але для виконання методу обраний такий стандартний набір граматичних значень, що якщо певному елементу не властива конкретна граматична форма, то її значення вважається рівним нулю (ці дані необхідні для точної ідентифікації зв'язків між елементами формальної системи). Для виконання етапу концептуалізації методу автоматизованого формування логіко-лінгвістичних моделей аналізується набір характеристик (2.5).

Тобто внаслідок здійснення морфологічного розбору на етапі концептуалізації створюється двовимірний масив, у якому рядки відповідають за номер елемента в системі, а стовпчики – за характеристики цих елементів:

$$Z = \begin{vmatrix} cm_{11} & g_{12} & n_{13} & k2_{14} & t_{15} & h_{16} & l_{17} & ch_{18} \\ cm_{21} & g_{22} & n_{23} & k2_{24} & t_{25} & h_{26} & l_{27} & ch_{28} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ cm_{i1} & g_{i2} & n_{i3} & k2_{i4} & t_{i5} & h_{i6} & l_{i7} & ch_{i8} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ cm_{r1} & g_{r2} & n_{r3} & k2_{r4} & t_{r5} & h_{r6} & l_{r7} & ch_{r8} \end{vmatrix}.$$

Основну роль при визначенні характеристик елементів формальної системи відіграє база даних, яка представляє собою словник слів української мови та словник морфем. Це дає можливість отримати інформацію про конкретний елемент системи $S_i, i = \overline{1, r}$.

База даних організована за допомогою сукупності таблиць, реалізованих у вигляді реляційної моделі, завдяки чому забезпечується структурний аспект – дані представляють собою набір відношень. Відношення відповідають умовам цілісності та аспект обробки – можливість маніпулювання даними. Все інформаційне наповнення бази даних представлено одним способом, а саме через явні значення атрибутів у кортежах відношень. Зокрема, по тому, якому значенню атрибута відповідає елемент системи $S_i, i = \overline{1, r}$, у якому стовпчику таблиці бази даних він знаходиться, визначаються його граматичні характеристики: відмінок, рід, число, форма і т.д.

База даних, що використовується у концептуальній схемі (рис. 3.4), має вигляд рис. 3.5, рис. 3.6.

Якщо конкретне слово (елемент) не знайдено у базі даних B_{t1}^{d1} або знайдено декілька атрибутів з однаковим значенням і не зрозуміло, як правильно ідентифікувати елемент, відбувається звернення до другої частини бази даних B_{t2}^{d2} , яка представляє собою морфологічний словник, побудований у вигляді реляційної

моделі, значеннями атрибутів якої є морфеми: закінчення, суфікси, корені, префікси (рис. 3.7).

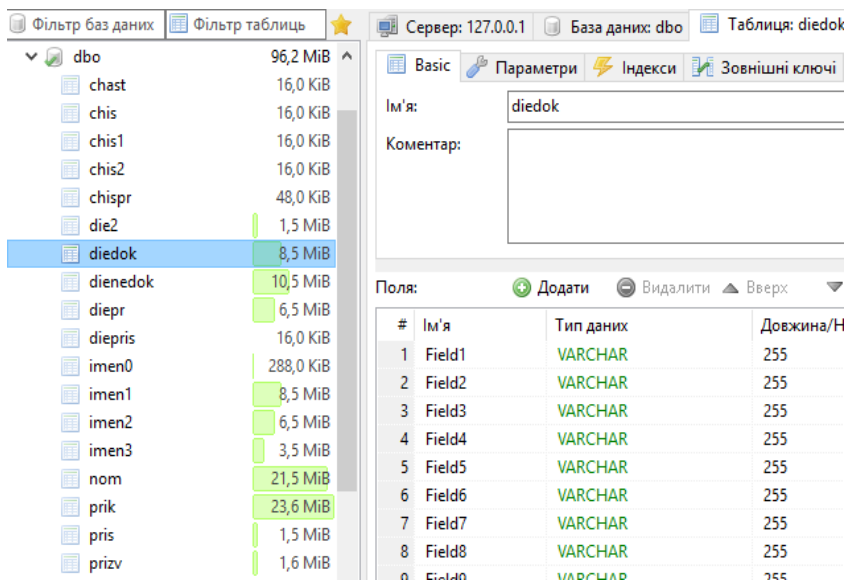


Рис. 3.5. База даних B_{11}^{d1}

Field1	Field2	Field3	Field4	Field5	Field6
іюекційний	іюекційного	іюекційному	іюекційний	іюекційним	іюекційному, іюек
іюекнерний	іюекнерного	іюекнерному	іюекнерний	іюекнерним	іюекнерному, іюек
іюекнерно-будівельний	іюекнерно-будівельного	іюекнерно-будівельному	іюекнерно-будівельний	іюекнерно-будівельним	іюекнерно-будівель
іюекнерно-механічний	іюекнерно-механічного	іюекнерно-механічному	іюекнерно-механічний	іюекнерно-механічним	іюекнерно-механіч
іюекнерно-технічний	іюекнерно-технічного	іюекнерно-технічному	іюекнерно-технічний	іюекнерно-технічним	іюекнерно-технічн
іюекнерно-фізичний	іюекнерно-фізичного	іюекнерно-фізичному	іюекнерно-фізичний	іюекнерно-фізичним	іюекнерно-фізичн
іюекнерський	іюекнерського	іюекнерському	іюекнерський, іюекнерського	іюекнерським	іюекнерському, іюек
іюекни	іюекниого	іюекниному	іюекни, іюекниого	іюекниим, іюекни	іюекниному, іюекни
іюекнинговий	іюекнингового	іюекнинговому	іюекнинговий	іюекнинговим	іюекнинговому, ію
іюекнирний	іюекнирного	іюекнирному	іюекнирний	іюекнирним	іюекнирному, іюекни
іюекнировий	іюекнирового	іюекнировому	іюекнировий	іюекнировим	іюекнировому, іюекни
іюекнистий	іюекнистого	іюекнистому	іюекнистий	іюекнистим	іюекнистому, іюекни
іюекниціальний	іюекниціального	іюекниціальному	іюекниціальний	іюекниціальним	іюекниціальному, іюекни
іюекниціально-цифровий	іюекниціально-цифрового	іюекниціально-цифровому	іюекниціально-цифровий	іюекниціально-цифровим	іюекниціально-цифров
іюекниціативний	іюекниціативного	іюекниціативному	іюекниціативний, іюекниціативного	іюекниціативним	іюекниціативному, іюекни
іюекниціаторський	іюекниціаторського	іюекниціаторському	іюекниціаторський, іюекниціаторського	іюекниціаторським	іюекниціаторському, іюекни
іюекниціовальний	іюекниціовального	іюекниціовальному	іюекниціовальний, іюекниціовального	іюекниціовальним	іюекниціовальному, іюекни
іюекниціаторський	іюекниціаторського	іюекниціаторському	іюекниціаторський, іюекниціаторського	іюекниціаторським	іюекниціаторському, іюекни
іюекнисовий	іюекнисового	іюекнисовому	іюекнисовий	іюекнисовим	іюекнисовому, іюекни
іюекниціаторський	іюекниціаторського	іюекниціаторському	іюекниціаторський, іюекниціаторського	іюекниціаторським	іюекниціаторському, іюекни
іюекницізційний	іюекницізційного	іюекницізційному	іюекницізційний	іюекницізційним	іюекницізційному, іюекни
іюекницізовий	іюекницізового	іюекницізовому	іюекницізовий, іюекницізового	іюекницізовим	іюекницізовому, іюекни
іюекниціреторний	іюекниціреторного	іюекниціреторному	іюекниціреторний, іюекниціреторного	іюекниціреторним	іюекниціреторному, іюекни
іюекницірустийний	іюекницірустийного	іюекницірустийному	іюекницірустийний	іюекницірустийним	іюекницірустийному, іюекни

Рис. 3.6. Внутрішня організація таблиць бази даних B_{11}^{d1}

	A	B	C	D	E
25	єць	0	іменник	ч.р.	2
26	ар	0	іменник	ч.р.	2
27	яр	0	іменник	ч.р.	2
28	ур	0	іменник	ч.р.	2
29	ер	0	іменник	ч.р.	2
30	ач	0	іменник	ч.р.	2
31	яч	0	іменник	ч.р.	2
32	ч	0	іменник	ч.р.	2
33	ист	0	іменник	ч.р.	2
34	іст	0	іменник	ч.р.	2
35	тель	0	іменник	ч.р.	2
36	ак	0	іменник	ч.р.	2
37	их	а	іменник	ж.р.	2
38	ес	а	іменник	ж.р.	2
39	ш	а	іменник	ж.р.	2
40	анин	0	іменник	ч.р.	2
41	янин	0	іменник	ч.р.	2
42	чанин	0	іменник	ч.р.	2
43	ин	0	іменник	ч.р.	2

Рис. 3.7. Таблиця реляційної бази даних B_{12}^{d2}

Використання реляційної бази даних на етапі концептуалізації надає методу автоматизованого формування логіко-лінгвістичних моделей ряд переваг у процесі його реалізації: єдиною інформаційною конструкцією є таблиця, що дозволяє узагальнити структуру запиту до бази даних і прорахувати час реакції на запит. Адже, при проектуванні реляційної бази даних застосовуються строгі правила, що базуються на математичному апараті. Також для побудови запитів немає необхідності знати конкретну організацію бази даних у зовнішній пам'яті, через великий об'єм якої значно скорочується швидкість доступу до таблиць бази даних.

Таким чином, морфологічний аналіз здійснюється для визначення граматичних форм кожного елементу системи шляхом дослідження структури слова, способу його творення, особливостей зміни та вираження властивих слову морфологічних значень. Умовою закінчення етапу концептуалізації є визначення всіх характеристик (властивостей) $z_i = (z_{k1}(S_i), k1 = \overline{1, w})$ кожного елементу системи $S_i, i = \overline{1, r}$ і створення двовимірного масиву SZ ,

елементами якого є безпосередньо прості елементи формальної системи та їх характеристики:

$$S \Rightarrow S = \begin{array}{|c|} \hline S_1 \\ \hline S_2 \\ \hline \dots \\ \hline S_i \\ \hline \dots \\ \hline S_r \\ \hline \end{array} \Rightarrow SZ = \begin{array}{|c|} \hline S_{10} \quad cm_{11} \quad g_{12} \quad n_{13} \quad k2_{14} \quad t_{15} \quad h_{16} \quad l_{17} \quad ch_{18} \\ \hline S_{20} \quad cm_{21} \quad g_{22} \quad n_{23} \quad k2_{24} \quad t_{25} \quad h_{26} \quad l_{27} \quad ch_{28} \\ \hline \dots \\ \hline S_{i0} \quad cm_{i1} \quad g_{i2} \quad n_{i3} \quad k2_{i4} \quad t_{i5} \quad h_{i6} \quad l_{i7} \quad ch_{i8} \\ \hline \dots \\ \hline S_{r0} \quad cm_{r1} \quad g_{r2} \quad n_{r3} \quad k2_{r4} \quad t_{r5} \quad h_{r6} \quad l_{r7} \quad ch_{r8} \\ \hline \end{array}$$

3.4 Синтаксичний аналізатор, визначення ролей

Вся задача розуміння висловлювання поділяється на два етапи: аналіз та інтерпретацію. На етапі концептуалізації було проаналізовано кожний елемент системи і визначено його характеристики, тобто граматичні категорії для кожної із словоформ висловлювання. Наступним кроком для формування логіко-лінгвістичної моделі і відповідно розуміння змісту висловлювання є інтерпретація всіх простих елементів формальної системи в масив синтаксичних ролей.

Грамматика флективної мови складається з двох частин: морфології та синтаксису. Синтаксис (від грецького «побудова», «об'єднання», «порядок») вивчає будову словосполучень та речень, змістові відношення у них та способи їх вираження. Синтаксис тісно пов'язаний з морфологією. В ньому морфологічні одиниці (частини мови, словоформи) вступають у відношення та зв'язки з іншими словами і в результаті будуються словосполучення і речення. У розрізі морфології слова розглядаються в аспекті їх внутрішньої структури, як вони класифікуються між собою, створюють морфологічні форми. У розрізі ж синтаксису ці слова розглядаються зі сторони їх участі у побудові словосполучень та речень. Таким чином, словоформи самостійних частин мови стають членами речення, службові слова виконують обслуговуючу роль при членах та частинах речення [71].

Здійснивши синтаксичний аналіз, тобто визначивши зв'язки між усіма словами вхідної текстової інформації, а також визначивши, якими членами речення виступають слова та словосполучення, тобто визначивши синтаксичну роль кожного елемента системи, можна добратися до змісту висловлювання. Морфологічний розбір забезпечує розуміння того, які об'єкти, суб'єкти та дії беруть участь у висловлюванні, їх кількість, час, рід, форму. А синтаксичний аналіз дає можливість зрозуміти саму суть виловлювання: з яким об'єктом відбувається дія, яка дія відбувається над об'єктом, який суб'єкт виконує цю дію, місце виконання, обставини, ознаки, що забезпечує встановлення зв'язків між словами речення. Для комп'ютера неможливо виконати синтаксичний аналіз, не виконавши попередньо морфологічний, що і було здійснено на етапі концептуалізації.

Кожне речення природної мови характеризується граматичною організацією: будовою речення та граматичним вираженням (оформленням) членів речення. Саме граматичне оформлення членів речення надалі буде називатися синтаксичними ролями.

Таким чином, на основі отриманих даних на етапі інтерпретації передається запит до бази знань, яка представлена у вигляді аксіом формальної системи, внаслідок чого кожному елементу формальної системи S_i ставиться у відповідність синтаксична роль G_i . Якщо розглядати речення, як масив $S = \bigwedge_{i \in \Gamma} S_i = \bigwedge_{i \in \Gamma} \bigwedge_{j \in \Pi} S_j^i$, то і по відношенню до ролей буде справедлива формула $G = \bigwedge_{i \in \Gamma} G_i$.

Для здійснення синтаксичного аналізу використовується аналізатор співставлення за зразками. Аналізатори даного класу базуються на тому, що у найпростішому випадку аналіз зводиться до співставлення речення з деякою множиною зразків, які представляють собою послідовності з одного або декількох слів.

На етапі здійснення синтаксичного аналізу на вхід аналізатора надходить масив простих елементів формальної системи та їх граматичних характеристик SZ . Кожен елемент цього масиву по черзі заноситься до робочої пам'яті бази знань, де за допомогою механізму логічного виводу зіставляється зі зразками бази правил.

База знань формальної системи описується у формі конкретних фактів і правил логічного виводу над базами даних та процедурами обробки інформації, що представляють собою

відомості про синтаксичну будову речень природної мови у логічній формі. Фактично, у базі знань B^{t3} зберігається розв'язок задачі ініціалізації та ідентифікації масиву ролей, а B^{t4} – розв'язок задачі формування логіко-лінгвістичних моделей.

Згідно зі схемою побудови бази знань (рис. 3.8), система аналізу задачі та зміни знань забезпечує формування моделі знань у відповідності з поставленими цілями. Завдяки незалежності форми представлення моделі знань від типу задачі виникає можливість створення єдиних проблемно-незалежних механізмів формування моделі знань.

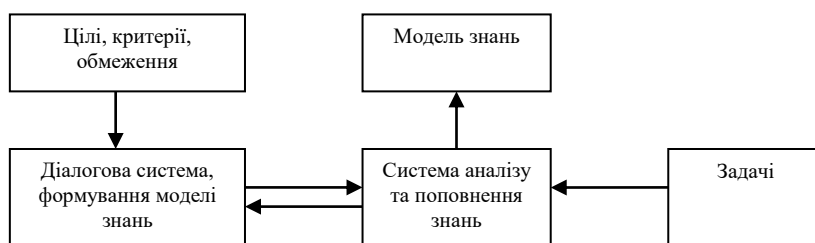


Рис. 3.8. Структурна схема формування бази знань

За основу створення обох баз B^{t3} та B^{t4} обрана продукційна модель представлення знань, в якій правила виступають засобами прямого опису способів логічного виводу. При цьому характерною рисою представлення знань з високою модульністю є простота доповнення, модифікації та анулювання даних.

Крім того, для комп'ютера таким чином буде визначено простий і точний механізм використання знань з високою однорідністю, описаних згідно з одним синтаксисом.

Продукційна система складається з трьох основних компонентів [20] (рис. 3.9): набору правил, робочої пам'яті та результатів виводу в сукупності з механізмом логічного виводу, що використовує правила у відповідності з вмістом робочої пам'яті.

Таким чином, система, що базується на знаннях, як один із компонентів, включає пам'ять для зберігання правил, яка містить набір правил, що спрацьовують у певних ситуаціях і мають форму ЯКЩО–ТО – продукції.

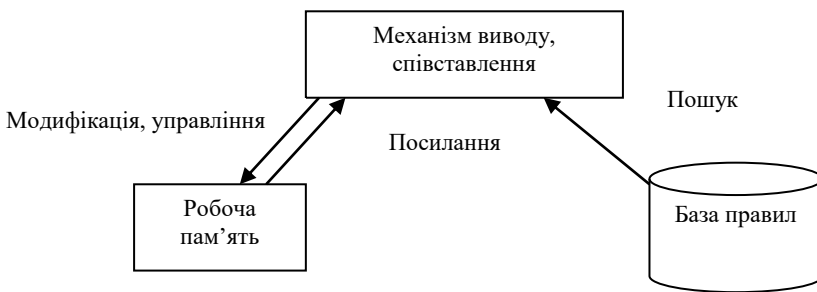


Рис. 3.9. Конфігурація системи продукцій

Дана система відноситься до систем з монотонним виводом, тобто факти, що зберігаються у робочій пам'яті, статичні, вони не змінюються в процесі рішення задачі.

Дані, які записуються у робочу пам'ять представляють собою зразки у вигляді набору символів. При зверненні до бази знань B^{t3} до робочої пам'яті надходять прості елементи формальної системи $S_i, i = \overline{1, r}$ разом з характеристиками, отриманими на етапі концептуалізації. У цілому в робочу пам'ять надходять дані у вигляді одновимірного масиву елементів формальної системи та двовимірного масиву характеристик цих елементів (рис. 3.10).

Таким чином, до робочої пам'яті бази знань заноситься масив елементів з граматичними характеристиками кожного із них $Z_i = \{cm_i, g_i, n_i, k2_i, t_i, h_i, l_i, ch_i\}$, а внаслідок дії механізму логічного виводу отримується одновимірний масив ролей:

$$S \Rightarrow S = \begin{pmatrix} S_1 \\ S_2 \\ \dots \\ S_i \\ \dots \\ S_r \end{pmatrix} \Rightarrow SZ = \begin{pmatrix} S_{10} & cm_{11} & g_{12} & n_{13} & k2_{14} & t_{15} & h_{16} & l_{17} & ch_{18} \\ S_{20} & cm_{21} & g_{22} & n_{23} & k2_{24} & t_{25} & h_{26} & l_{27} & ch_{28} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ S_{i0} & cm_{i1} & g_{i2} & n_{i3} & k2_{i4} & t_{i5} & h_{i6} & l_{i7} & ch_{i8} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ S_{r0} & cm_{r1} & g_{r2} & n_{r3} & k2_{r4} & t_{r5} & h_{r6} & l_{r7} & ch_{r8} \end{pmatrix} \Rightarrow G = \begin{pmatrix} G_1 \\ G_2 \\ \dots \\ G_i \\ \dots \\ G_r \end{pmatrix} .$$

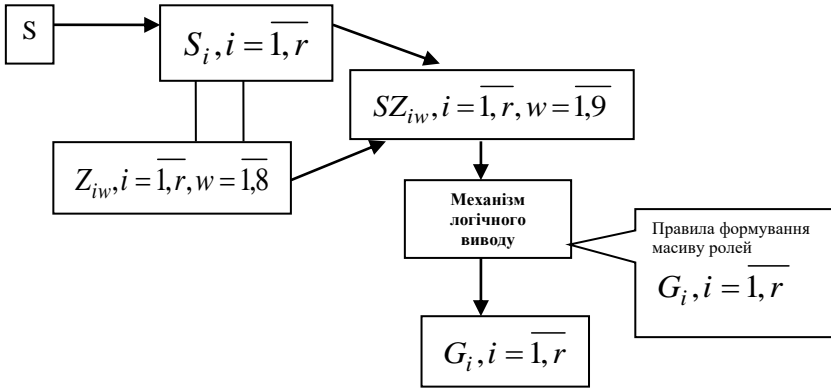


Рис. 3.10. Загальна схема перетворення $S_i, i = \overline{1, r}$ в $G = \bigwedge_{i \in r} G_i$

Використовуючи два основних припущення логіки предикатів про те, «будь-яке висловлювання або істинне або хибне (принцип двозначності)» та «істинне значення складного висловлювання залежить лише від істинності значень простих висловлювань, що в нього входять, та характеру їх зв'язку», сформовано правила бази знань, характерні для флективної мови. Для полегшення подальшої роботи з базою правил їх розділено на декілька категорій.

Набір аксіом формальної системи представляє собою сукупність правил природної мови [72], згідно яких слова об'єднуються в словосполучення. Правила утворення словосполучень інтерпретовані у формули логіки предикатів першого порядку [73].

Кожне правило складається з двох частин. Перша з них – антецедент або посилення правила, складається з елементарних частин, з'єднаних логічними зв'язками «і», «або» і т.д. Друга частина, що називається консеквентом або висновком, складається з однієї або декількох частин, які утворюють рішення або вказують на дію, яку необхідно виконати. Антецедент представляє собою зразок правила, призначений для розпізнавання ситуації, коли це правило має спрацювати.

1. Правила, що вказують на зв'язки між простими елементами та визначають структуру складних елементів формальної системи

$PR_1(B^{r3})$. У флективній мові ці правила ототожнюються з типами зв'язку в словосполученнях (керування, узгодження, прилягання), де S_j^i – складний елемент формальної системи, S_i – головний елемент словосполучення, $i = \overline{1, r}, j = \overline{1, m}$ (формалізовані правила формування словосполучень, описані у другому розділі).

2. Правила визначення граматичної основи (головних членів) речення $PR_2(B^{r3})$. У флективній мові це правила визначення головних членів речення, де x – підмет, P – присудок (формалізовані правила визначення граматичної основи речення, описані у другому розділі).

Механізм логічного виводу або інтерпретатор правил забезпечує перегляд існуючих зразків з робочої пам'яті (масив SZ) та правил із бази знань B^{r3} . Цей механізм включає в себе дві компоненти: реалізація логічного виводу та компонента управління цим процесом.

Управляюча компонента визначає порядок застосування правил і виконує чотири функції [68]:

1) *зіставлення* – зразок правила зіставляється з реальними даними (зразками з масиву SGZ);

2) *вибір*, якщо у конкретній ситуації можна застосувати одразу декілька правил, то серед них обирається одне, що найбільш підходить до заданого критерію (розв'язання конфлікту);

3) *спрацювання* – якщо зразок правила при співставленні збігається зі зразками робочої пам'яті, то правило спрацює;

4) *дія* – робоча пам'ять піддається зміні шляхом додавання до неї висновку із правила, яке спрацювало.

Інтерпретатор продукцій працює циклічно. У кожному циклі він переглядає всі правила, щоб з'ясувати ті, посилення яких співпадають з відомими на даний момент фактами із робочої пам'яті. Після вибору правило спрацює, його висновок заноситься до робочої пам'яті, і потім цикл повторюється спочатку.

В одному циклі може спрацювати лише одне правило. Якщо декілька правил успішно зіставлені з фактичними зразками, то інтерпретатор здійснює вибір за визначеним критерієм єдиного правила, яке спрацює у даному циклі [20] (рис. 3.11).

Інформація з робочої пам'яті послідовно зіставляється з посиланнями правил для виявлення успішного співставлення.

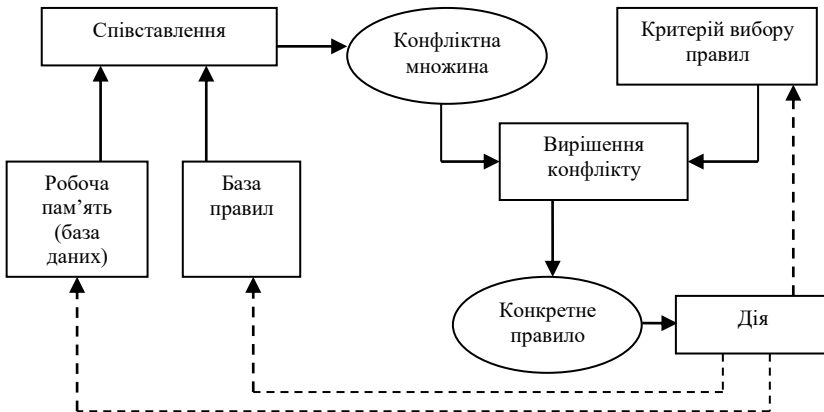


Рис. 3.11. Цикл роботи інтерпретатора

Сукупність обернених правил складає так звану конфліктну множину. Для вирішення конфлікту інтерпретатор має критерій, за допомогою якого він вибирає одне єдине правило, після чого воно спрацьовує. Це виражається у занесенні фактів, що утворюють висновки правила, до робочої пам'яті або у зміні критерію вибору конфліктуєчих правил. Робота машини виводу залежить тільки від стану робочої пам'яті та від складу бази знань [20] (рис. 3.12).

Після того, як до робочої пам'яті записуються зразки (масив з елементів системи та їх характеристик), розглядається можливість застосування правил. Спочатку механізм виводу співставляє зразки із умовної частини правил зі зразками, що зберігаються у робочій пам'яті. Якщо такі зразки існують в робочій пам'яті, умовна частина вважається істинною, в інакшому випадку – не вірною.

Для баз знань існуючої формальної системи застосовано механізм прямого логічного виводу, тобто тут за відомими фактами відшукуються висновки, які з цих фактів впливають. Такий вивід часто називають виводом, що управляє антецедентами, так як для отримання висновку проводиться

робота з вилучення попередньо записаного вмісту робочої пам'яті, застосування правил і доповнення даних робочої пам'яті.

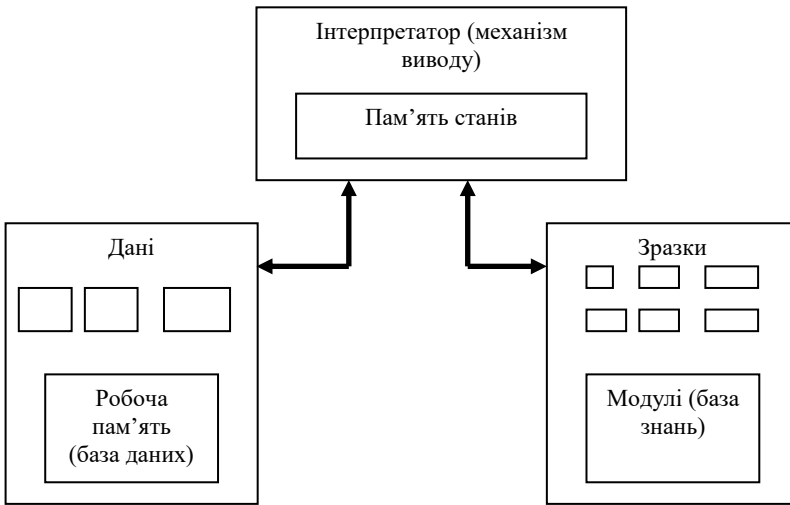


Рис. 3.12. Схема функціонування інтерпретатора

Визначення ролей відбувається на основі інформації, отриманої з бази даних та бази знань.

Етап концептуалізації служить своєрідною підготовкою для здійснення синтаксичного аналізу формальної системи і створення нового масиву даних, розмірність якого відповідає кількості елементів системи r , а значення елементів є назвами синтаксичних ролей кожного з простих або (якщо на етапі інтерпретації це стає можливим) складних елементів [73].

Синтаксичну роль деяких елементів можна визначити ще на етапі концептуалізації, отримавши результат запити до бази даних V^{t1} . Проте основна складність методу, що розглядається, полягає в тому, що текстова інформація, яка перетворюється на логіко-лінгвістичну модель, може відноситися до різних предметних областей, відповідно типи речень і взаємозв'язки між ними можуть бути різноманітними. Тому для універсального визначення синтаксичних ролей елементів використовуються аксіоми бази знань V^{d3} , що представляють собою правила

природної мови, згідно з якими визначається, яким членом речення виступає те чи інше слово або словосполучення.

На основі отриманих даних формується масив ролей $G = \bigwedge_{i \in r} G_i$, де кожному елементу масиву G_i ставиться у відповідність елемент формальної системи $S_i, i = \overline{1, r}$.

Після визначення синтаксичних ролей необхідно звернути увагу на знаки пунктуації, присутні у вхідній текстовій інформації, а також сполучники та сполучні слова. Тут відбувається звернення до бази даних B^{d1} і бази знань B^{d4} , що представляє собою множину правил виводу.

3.5 Формалізація процесу побудови логіко-лінгвістичних моделей

Знання, які можуть бути представлені за допомогою логіки предикатів, є фактами, що відображаються логічними формулами. Автоматичне перетворення речень, написаних на природній мові, у мову формальних систем типу логіки предикатів, називається розумінням природної мови. Логіці предикатів властивий високий рівень модульності знань, і разом з цим вона дозволяє отримати єдину систему представлення, в якій логічно роз'яснюються властивості знань як одного цілого. Першочерговим завданням логіки предикатів є пояснення логічних основ природної мови. Оскільки будь-яка флективна мова надзвичайно складна, то об'єкти, якими оперує логіка предикатів, обмежені тими компонентами, елементи яких легко піддаються формалізації. Тому введемо базові поняття для систематизації цієї логіки, що дасть змогу вилучати знання з як завгодно складних структур текстової інформації.

Логіко-лінгвістична модель представляє собою правило, яке об'єднує елементи пропозиційної логіки та синтаксичну структуру речень природної мови (2.1) – (2.4).

Механізм виводу або інтерпретатор правил забезпечує перегляд існуючих зразків з робочої пам'яті (масив SZ) та правил із бази знань B^{t4} . Реалізація механізму логічного виводу здійснюється, як і для бази знань B^{t3} на основі правила *modus ponens*.

У базі знаходиться набір правил, згідно з якими здійснюється логічний вивід. База правил заповнюється відповідно до формалізованих синтаксичних правил флективної мови.

Для української мови сформована система продукцій, що містить сімнадцять правил визначення компонентів логіко-лінгвістичних моделей [4].

3) Правило трактується наступним чином: якщо у реченні відсутні розділові знаки, воно є простим, у логіко-лінгвістичній моделі йому відповідає просте висловлювання L^S . Наприклад, речення «Я н'ю невимовну радість життя».

$$\begin{aligned} & \text{if } ((S_i \neq ",") \text{ and } (S_i \neq ";") \text{ and } (S_i \neq "-") \text{ and } (S_i \neq ":")) \\ & \text{then } (L(S) = L^S). \end{aligned}$$

Відповідно, міркування за правилом *modus ponens* можна відобразити такою схемою:

Антецедент 1:

$$\begin{aligned} & \text{if } ((S_i \neq ",") \text{ and } (S_i \neq ";") \text{ and } (S_i \neq "-") \text{ and } (S_i \neq ":")) \\ & \text{then } (L(S) = L^S) \end{aligned}$$

Антецедент 2:

$$((S_i \neq ",") \text{ and } (S_i \neq ";") \text{ and } (S_i \neq "-") \text{ and } (S_i \neq ":"))$$

$$\text{Консеквент} :: (L(S) = L^S).$$

Логіко-лінгвістична модель речення буде мати вигляд:

$$L^{S_1} = p_1(x_1, g_1, y_1, q_1, z_1, r_1, h_1) = p_1(x_1, 0, y_1, q_1, z_1, 0, 0),$$

$$L^{S_1} = n \text{ 'ю } (я, 0, \text{радість, невимовну, життя}, 0, 0).$$

2) Якщо у реченні зустрівся знак пунктуації «,», а серед слів речення, які стоять до та після коми, присутні підмет і присудок, дане речення є складним, і у ньому йдеться про одночасність виконання певних дій, станів, характеристик.

$$\begin{aligned} & \text{if } \left((S_i = ",") \text{ and } ((ch(S_k) = 1) \vee (ch(S_{k1}) = 2)) \text{ and } \right. \\ & \left. ((ch(S_{j1}) = 1) \vee (ch(S_{j2}) = 2)) \right) \\ & \text{then } (L(S) = L'(S_1) \& L''(S_1)). \end{aligned}$$

Тут $S_k, k = \overline{1, i-1}$ – одне із слів, що стоїть до коми;

$S_{k1}, k1 = \overline{1, i-1}$ – одне із слів, що стоїть до коми;

$S_{j_1}, j_1 = \overline{i+1, r}$ – одне із слів, що стоїть після коми;

$S_{j_2}, j_2 = \overline{i+1, r}$ – одне із слів, що стоїть після коми.

Наприклад, речення «Світить з блакиті сонце, зеленіють гори, джерельною свіжістю дихає вода».

Логіко-лінгвістична модель речення буде мати вигляд:

$$\begin{aligned} L(S) &= L'(S_1) \& L''(S_1) \& L'''(S_1) = \\ &= p'_1(x'_1, g'_1, y'_1, q'_1, z'_1, r'_1, h'_1) \& p''_1(x''_1, g''_1, y''_1, q''_1, z''_1, r''_1, h''_1) \& \\ & p'''_1(x'''_1, g'''_1, y'''_1, q'''_1, z'''_1, r'''_1, h'''_1) = p'_1(x'_1, 0, 0, 0, 0, 0, h'_1) \& \\ & p''_1(x''_1, 0, 0, 0, 0, 0, 0) \& p'''_1(x'''_1, 0, y'''_1, q'''_1, 0, 0, 0), \\ L'(S_1) \& L''(S_1) \& L'''(S_1) &= \text{світить (сонце, } 0, 0, 0, 0, 0, \text{блакиті)} \& \\ \text{зеленіють (гори, } 0, 0, 0, 0, 0, 0) \& \\ \text{дихає (вода, } 0, \text{ свіжістю, джерельною, } 0, 0, 0). \end{aligned}$$

3) Якщо у реченні зустрівся знак пунктуації «,», а серед слів речення, які стоять до та після коми, присутні підмети (присудки) з однаковими граматичними характеристиками, дане речення є простим і містить однорідні члени у вигляді підметів (присудків).

$$\text{if} \left(\begin{array}{l} (S_i = ",") \text{ and } ((ch(S_k) = 1) \text{ and } (ch(S_{j_1}) = 1) \text{ and } (z_k = z_{j_1})) \vee \\ \vee ((ch(S_{k_1}) = 2) \text{ and } (ch(S_{j_2}) = 2) \text{ and } (z_{k_1} = z_{j_2})) \end{array} \right) \\ \text{then} (L(S) = L^S).$$

Наприклад, речення «І радість, і смуток, і срібний передзвін огортають мене своїм снуванням».

Логіко-лінгвістична модель речення буде мати вигляд:

$$\begin{aligned} L(S) &= L^{S^1} = p_1(x_{11}, 0, y_1, 0, z_1, r_1, 0) \& \\ & p_1(x_{12}, 0, y_1, 0, z_1, r_1, 0) \& p_1(x_{13}, g_1, y_1, 0, z_1, r_1, 0), \\ L(S) &= L^{S^1} = \text{огортає (радість, } 0, \text{ мене, } 0, \text{ снуванням, своїм, } 0) \& \\ \text{огортає (смуток, } 0, \text{ мене, } 0, \text{ снуванням, своїм, } 0) \& \\ \text{огортає (передзвін, срібний, мене, } 0, \text{ снуванням, своїм, } 0). \end{aligned}$$

4) Якщо у реченні зустрівся знак пунктуації «;», а серед слів речення, які стоять до та після коми, присутні підмет і присудок, дане речення є складним, і в ньому йдеться про одночасність виконання певних дій, станів, характеристик.

$$\text{if} \left((S_i = ";") \text{ and } ((ch(S_k) = 1) \vee (ch(S_{k_1}) = 2)) \text{ and} \right. \\ \left. ((ch(S_{j_1}) = 1) \vee (ch(S_{j_2}) = 2)) \right) \\ \text{then} (L(S) = L'(S_1) \& L''(S_1)).$$

Наприклад, речення «Сонце заходить; гори чорніють».

Логіко-лінгвістична модель речення буде мати вигляд:

$$L(S) = L'(S_1) \& L''(S_1) = p'_1(x'_1, 0, 0, 0, 0, 0, 0) \&$$

$$p''_1(x''_1, 0, 0, 0, 0, 0, 0),$$

$$L(S) = L'(S_1) \& L''(S_1) = \text{заходить (сонце, } 0, 0, 0, 0, 0, 0) \& \\ \text{чорніють(гори, } 0, 0, 0, 0, 0, 0).$$

5) Якщо у реченні зустрівся знак пунктуації «,», а серед слів речення, які стоять до та після коми, присутні слова з однаковими граматичними характеристиками і відносяться до одного й того ж слова, дане речення є простим і містить однорідні другорядні члени.

$$\text{if} \left((S_i = ",") \text{ and } ((ch(S_k) = 3) \text{ and } (ch(S_{j_1}) = 3) \text{ and } (z_k = z_{j_1})) \vee \right. \\ \left. \vee ((ch(S_k) = 4) \text{ and } (ch(S_{j_1}) = 4) \text{ and } (z_k = z_{j_1})) \vee \right. \\ \left. \vee ((ch(S_k) = 5) \text{ and } (ch(S_{j_1}) = 5) \text{ and } (z_k = z_{j_1})) \text{ and} \right. \\ \left. (S_j = S_{k_1} \cup S_k) \& (S_{j+1} = S_{k_1} \cup S_{j_1}) \right)$$

$$\text{then} (L(S) = L^S).$$

Наприклад, речення «Вчені розробили новий, фундаментальний підхід до лінгвістичного аналізу».

Логіко-лінгвістична модель речення буде мати вигляд:

$$L(S) = L^S = p_1(x_1, 0, y_1, q_{11}, z_1, r_1, 0) \& p_1(x_1, 0, y_1, q_{12}, z_1, r_1, 0),$$

$$L(S) = L^S = \text{розробили (вчені, } 0, \text{ підхід, новий, аналізу,} \\ \text{лінгвістичного, } 0) \& \text{розробили (вчені, } 0, \text{ підхід, фундаментальний,} \\ \text{аналізу, лінгвістичного, } 0).$$

6) Якщо у реченні зустрівся знак пунктуації «:», а серед слів речення, які стоять до та після двокрапки, присутні підмет і присудок, то речення є складним, і у ньому йдеться про еквівалентність виконання тих чи інших дій, станів, характеристик.

$$\text{if} \left(\begin{array}{l} (S_i = " : ") \text{ and } ((ch(S_k) = 1) \vee (ch(S_{k1}) = 2)) \text{ and} \\ ((ch(S_{j1}) = 1) \vee (ch(S_{j2}) = 2)) \end{array} \right) \\ \text{then} (L(S) = L'(S_1) \& L''(S_1)).$$

Наприклад, речення *«Кажуть: хліб усьому голова»*.

Логіко-лінгвістична модель речення буде мати вигляд:

$$L(S) = L'(S_1) \& L''(S_1) = p'_1(0, 0, 0, 0, 0, 0, 0) \& p''_1(x''_1, 0, y''_1, 0, 0, 0, 0),$$

$$L(S) = L'(S_1) \& L''(S_1) = \text{кажуть} (0, 0, 0, 0, 0, 0, 0) \& \text{голова (хліб, 0, усьому, 0, 0, 0, 0)}.$$

7) Якщо у реченні зустрівся знак пунктуації «←», а серед слів речення, які стоять до та після тире, присутні підмет і присудок, то речення є складним, і у ньому йдеться про певні наслідки виконаних дій, станів, характеристик.

$$\text{if} \left(\begin{array}{l} (S_i = " - ") \text{ and } ((ch(S_k) = 1) \vee (ch(S_{k1}) = 2)) \text{ and} \\ ((ch(S_{j1}) = 1) \vee (ch(S_{j2}) = 2)) \end{array} \right) \\ \text{then} (L(S) = L'(S_1) \rightarrow L''(S_1)).$$

Наприклад, речення *«З'ясувальні відношення виражаються інтонацією і смисловим зв'язком – друга частина пояснює зміст першої»*.

Логіко-лінгвістична модель речення буде мати вигляд:

$$L(S) = L'(S_1) \rightarrow L''(S_1) = (p'_1(x'_1, g'_1, y'_{11}, 0, 0, 0, 0) \&$$

$$p'_1(x'_1, g'_1, y'_{12}, q'_{12}, 0, 0, 0)) \rightarrow p''_1(x''_1, g''_1, y''_1, 0, z''_1, 0, 0),$$

$$L(S) = L'(S_1) \rightarrow L''(S_1) = (\text{виражаються (відношення, з'ясувальні, інтонацією, 0, 0, 0, 0)} \& \text{виражаються (відношення, з'ясувальні, зв'язком, смисловим, 0, 0, 0)}) \rightarrow \text{пояснює(частина, друга, зміст, 0, першої, 0, 0)}.$$

8) Якщо у реченні зустрівся знак пунктуації «:», а слово перед нею має такі ж граматичні характеристики, як і слова, що стоять через кому за двокрапкою, то речення просте і містить однорідні другорядні члени та узагальнююче слово.

$$\text{if } \left(\begin{array}{l} (S_i = ":" \text{ and } ((z_{i-1} = z_k / n(S_{i-1}) \neq n(S_k)) \text{ and } (S_{k+1} = ",") \text{ and} \\ ((z_{i-1} = z_{k1} / n(S_{i-1}) \neq n(S_{k1})) \text{ and } (S_{k1+1} = ",") \text{ and} \\ ((z_{i-1} = z_{j1} / n(S_{i-1}) \neq n(S_{j1})) \text{ and } (S_{j1+1} = ",") \end{array} \right)$$

$$\text{then } (L(S) = L^S).$$

Наприклад, речення «*Експерт перевірів наявність всіх документів: актів, анкети, бланків звітності*».

Логіко-лінгвістична модель речення буде мати вигляд:

$$L(S) = L^S = p_1(x_1, 0, y_1, 0, z_{11}, r_{11}, 0) \& p_1(x_1, 0, y_1, 0, z_{12}, 0, 0) \& \\ p_1(x_1, 0, y_1, 0, z_{13}, 0, 0) \& p_1(x_1, 0, y_1, 0, z_{14}, r_{14}, 0),$$

$L(S) = L^S = \text{перевірів (експерт, 0, наявність, 0, документів, всіх, 0)} \& \text{перевірів (експерт, 0, наявність, 0, актів, 0, 0)} \& \\ \text{перевірів (експерт, 0, наявність, 0, анкети, 0, 0)} \& \\ \text{перевірів (експерт, 0, наявність, 0, бланків, звітності, 0)}.$

9) Якщо у реченні зустрівся знак пунктуації «–», перед ним присутні однорідні члени речення, а після нього – узагальнююче слово, то речення просте .

$$\text{if } \left(\begin{array}{l} (S_i = "-" \text{ and } ((z_{i+1} = z_k) \text{ and } (S_{k+1} = ",") \text{ and} \\ ((z_{i+1} = z_{k1}) \text{ and } (S_{k1+1} = ",") \text{ and} \\ ((z_{i+1} = z_{j1}) \text{ and } (S_{j1+1} = ",") \text{ and } (cm(S_{i+1})) = 4 \end{array} \right)$$

$$\text{then } (L(S) = L^S).$$

Наприклад, речення «*Вишні, черешні, яблуні, ранні груші – все стоїть у цвіту*».

Логіко-лінгвістична модель речення буде мати вигляд:

$$L(S) = L^S = p_1(x_{11}, 0, y_1, 0, 0, 0, 0) \& p_1(x_{12}, 0, y_1, 0, 0, 0, 0) \& \\ p_1(x_{13}, 0, y_1, 0, 0, 0, 0) \& p_1(x_{14}, g_{14}, y_1, 0, 0, 0, 0) \& p_1(x_{15}, 0, y_1, 0, 0, 0, 0),$$

$L(S) = L^S = \text{стоять (вишні, 0, цвіту, 0, 0, 0, 0)} \& \text{стоять (черешні, 0, цвіту, 0, 0, 0, 0)} \& \\ \text{стоять (яблуні, 0, цвіту, 0, 0, 0, 0)} \& \\ \text{стоять (груші, ранні, цвіту, 0, 0, 0, 0)} \& \text{стоять (все, 0, цвіту, 0, 0, 0, 0)}.$

10) Якщо у реченні зустрівся знак пунктуації «–», у групі слів перед ним знаходиться один із головних членів речення, а у групі слів після тире – другий, то речення просте.

$$\text{if } (S_i = "-" \text{ and } (ch(S_k) = 1) \text{ and } (ch(S_{j_1}) = 2))$$

$$\text{then } (L(S) = L^S).$$

Наприклад, речення «Наша мелодійна мова – співучий струмок».

Логіко-лінгвістична модель речення буде мати вигляд:

$$L(S) = L^S = p_1(x_1, g_{11}, 0, 0, 0, 0, h_1) \& p_1(x_1, g_{12}, 0, 0, 0, 0, h_1),$$

$$L(S) = L^S = \text{струмок (мова, наша, 0, 0, 0, 0, співучий)} \& \\ \text{струмок (мова, мелодійна, 0, 0, 0, 0, співучий)}.$$

11) Якщо у реченні зустрівся знак пунктуації «,», наступне слово – сполучник з ознакою Sp_1 , а слова, що стоять до і після коми мають однакові граматичні характеристики, то речення просте з однорідними членами.

$$\text{if } (S_i = "," \text{ and } (cm(S_{i+1}) = 10) \text{ and } (z_{i-1} = z_{i+2}))$$

$$\text{then } (L(S) = L^S)$$

Наприклад, речення «Для вирішення цієї проблеми вчені не шкодували ні сил, ні здоров'я, ні коштів».

Логіко-лінгвістична модель речення буде мати вигляд:

$$L(S) = L^S = \neg p_1(x_1, 0, y_1, 0, z_1, r_1, \neg h_{11}) \&$$

$$\neg p_1(x_1, 0, y_1, 0, z_1, r_1, \neg h_{12}) \& \neg p_1(x_1, 0, y_1, 0, z_1, r_1, \neg h_{13}),$$

$$L(S) = L^S = \neg \text{шкодували (вчені, 0, вирішення, 0, проблеми, цієї,} \\ \neg \text{сил)} \& \neg \text{шкодували (вчені, 0, вирішення, 0, проблеми, цієї,} \\ \neg \text{здоров'я)} \& \neg \text{шкодували (вчені, 0, вирішення, 0, проблеми, цієї,} \\ \neg \text{коштів)}.$$

12) Якщо в реченні зустрівся знак пунктуації «,», наступне слово – сполучник з ознакою Sp_2 , а серед групи слів, що стоять до та після коми є граматична основа, то речення є складним, і в ньому йдеться про альтернативу виконання певних дій, станів, характеристик.

$$\text{if } \left((S_i = ", ") \text{ and } (cm(S_i) = 10) \text{ and } \right. \\ \left. ((ch(S_k) = 1) \vee (ch(S_{k1}) = 2)) \text{ and } ((ch(S_{j1}) = 1) \vee (ch(S_{j2}) = 2)) \right) \\ \text{then } (L(S) = L'(S_1) \vee L''(S_1)).$$

Наприклад, речення «Мокрий сніг засипав дороги, але ніхто не звертав на це уваги».

Логіко-лінгвістична модель речення буде мати вигляд:

$$L(S) = L'(S_1) \vee L''(S_1) = p'_1(x'_1, g'_1, y'_1, 0, 0, 0, 0) \vee \neg p''_1(x''_1, 0, y''_1, 0, z''_1, 0, 0),$$

$$L(S) = L'(S_1) \vee L''(S_1) = \text{засипав (сніг, мокрий, дороги,} \\ 0, 0, 0, 0) \vee \neg \text{звертав (ніхто, 0, уваги, 0, це, 0, 0)}.$$

13) Якщо у реченні зустрівся знак пунктуації «,», наступне слово – сполучник з ознакою Sp_1 , а серед групи слів, що стоять до та після коми є граматична основа, то речення є складним, і в ньому йдеться про одночасність виконання певних дій, станів, характеристик.

$$\text{if } \left((S_i = ", ") \text{ and } (cm(S_i) = 10) \text{ and } \right. \\ \left. ((ch(S_k) = 1) \vee (ch(S_{k1}) = 2)) \text{ and } ((ch(S_{j1}) = 1) \vee (ch(S_{j2}) = 2)) \right) \\ \text{then } (L(S) = L'(S_1) \& L''(S_1)).$$

Наприклад, речення «Вчені розробили новий метод, і студенти написали його програмну верифікацію».

Логіко-лінгвістична модель речення буде мати вигляд:

$$L(S) = L'(S_1) \& L''(S_1) = p'_1(x'_1, 0, y'_1, q'_1, 0, 0, 0) \&$$

$$(p''_1(x''_1, 0, y''_1, q''_{11}, 0, 0, 0) \& p''_1(x''_1, 0, y''_1, q''_{12}, 0, 0, 0)),$$

$$L(S) = L'(S_1) \& L''(S_1) = \text{розробили (вчені, 0, метод, новий,} \\ 0, 0, 0) \& \text{написали (студенти, 0, верифікацію, його, 0, 0, 0) \&} \\ \text{написали (студенти, 0, верифікацію, програмну, 0, 0, 0)}).$$

14) Якщо у реченні зустрівся знак пунктуації «-», наступне слово – сполучник з ознакою Sp_1 , а серед групи слів, що стоять до та після коми є граматична основа, то речення є складним, і в ньому йдеться про певні наслідки виконаних дій, станів, характеристик.

if $\left((S_i = "-") \text{ and } (cm(S_i) = 10) \text{ and } \right. \\ \left. ((ch(S_k) = 1) \vee (ch(S_{k1}) = 2)) \text{ and } ((ch(S_{j1}) = 1) \vee (ch(S_{j2}) = 2)) \right)$
 then $(L(S) = L'(S_1) \rightarrow L''(S_1))$.

Наприклад, речення «Розсуваєш чіпке листя чагарників, – і перед тобою несподівано виблискує вода».

Логіко-лінгвістична модель речення буде мати вигляд:

$L(S) = L'(S_1) \rightarrow L''(S_1) = p'_1(0, 0, y'_1, q'_1, z'_1, 0, 0) \rightarrow p''_1(x''_1, 0, y''_1, q''_1, 0, 0, h''_1)$,

$L(S) = L'(S_1) \rightarrow L''(S_1) = \text{розсуваєш } (0, 0, \text{листя, чіпке, чагарників, } 0, 0) \rightarrow \text{виблискує (вода, } 0, \text{ перед, тобою, } 0, \text{ несподівано)}$.

15) Якщо у реченні зустрівся знак пунктуації «,», перед нею стоїть дієприкметник або дієприслівник, то речення просте, ускладнене дієприкметниковим або дієприслівниковим зворотом.

if $((S_i = ",") \text{ and } ((cm(S_k) = 6) \vee (cm(S_k) = 7)))$

then $(L(S) = L^S)$.

Наприклад, речення «Здійснивши перевірку, експерти залишилися задоволені».

Логіко-лінгвістична модель речення буде мати вигляд:

$L(S) = L^S = p_1(x_1, g_1, 0, 0, 0, 0, 0) \& p_2(x_1, 0, y_1, 0, 0, 0, 0)$,

$L(S) = L^S = \text{залишилися (експерти, задоволені, } 0, 0, 0, 0, 0) \& \text{здійснили (експерти, } 0, \text{ перевірку, } 0, 0, 0, 0)$.

16) Якщо в реченні зустрівся знак пунктуації «,», а після неї стоїть дієприкметник або дієприслівник, то речення просте, ускладнене дієприкметниковим або дієприслівниковим зворотом.

if $((S_i = ",") \text{ and } ((cm(S_{j1}) = 6) \vee (cm(S_{j1}) = 7)))$

then $(L(S) = L^S)$.

Наприклад, речення «В основу принципу покладено метод, розроблений статистами».

Логіко-лінгвістична модель речення буде мати вигляд:

$L(S) = L_1(S) \& L_2(S) = p_1(x_1, 0, y_1, 0, z_1, 0, 0) \&$

$p_2(x_1, 0, y_2, 0, 0, 0, 0)$,

$L(S) = L_1(S) \& L_2(S) = \text{покладено (метод, } \theta, \text{ основу, } \theta, \text{ принципу, } \theta, \theta) \& \text{ розроблений (метод, } \theta, \text{ статистами, } \theta, \theta, \theta, \theta).$

17) Якщо у реченні зустрівся знак пунктуації «,», наступне слово – сполучник з ознакою Sp_3 , а серед групи слів, що стоять до та після коми є граматична основа, то речення є складним, і в ньому йдеться про наслідки або причини виконання певних дій, станів, характеристик.

$$\text{if } \left((S_i = ",") \text{ and } (cm(S_i) = 10) \text{ and } \right. \\ \left. ((ch(S_k) = 1) \vee (ch(S_{k1}) = 2)) \text{ and } ((ch(S_{j1}) = 1) \vee (ch(S_{j2}) = 2)) \right) \\ \text{then } (L(S) = L'(S_1) \rightarrow L''(S_1)).$$

Наприклад, речення «Розробники написали детальну інструкцію з користування програмою, щоб уникнути маси запитань».

Логіко-лінгвістична модель речення буде мати вигляд:

$$L(S) = L'(S_1) \rightarrow L''(S_1) = p'_1(x'_1, 0, y'_1, q'_1, z'_1, r'_1, 0) \rightarrow$$

$$p''_1(0, 0, y''_1, q''_1, 0, 0, 0),$$

$L(S) = L'(S_1) \rightarrow L''(S_1) = \text{написали (розробники, } \theta, \text{ інструкцію, } \theta, \text{ детальну, користування, програмою, } \theta) \rightarrow \text{уникнути } (\theta, \theta, \text{ маси, запитань, } \theta, \theta, \theta).$

Налаштувавши формулу для конкретної формальної системи, елементами якої є $S_i, i=1, r$, можна отримати чітку схему перетворень текстової інформації в логіко-лінгвістичну модель у загальному вигляді для будь-якого висловлювання типу L^S (рис. 3.13). За таким перетворенням логіко-лінгвістичну модель можна розглядати як одновимірний масив, значеннями елементів якого є прості елементи формальної системи, упорядковані згідно з формулою (2.4). Тобто вхідна текстова інформація спочатку перетворюється на одновимірний масив S , елементами якого є слова, потім на базі цього масиву формується масив ролей G , елементами якого є назви членів речення, якими виступають слова (розмірність масивів однакова).

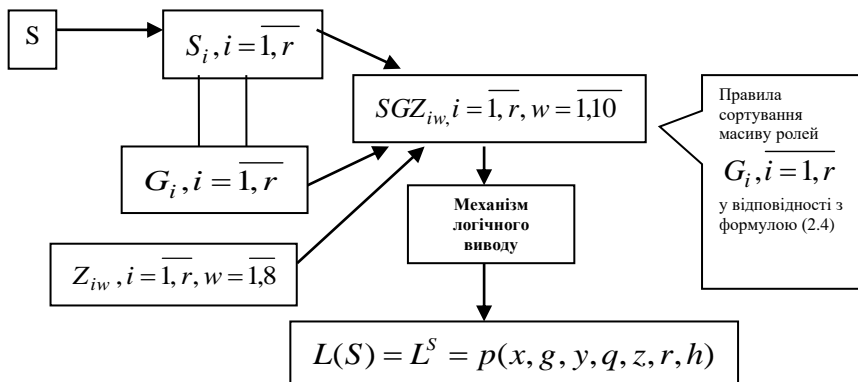


Рис. 3.13. Загальна схема перетворення текстової інформації в логіко-лінгвістичну модель

На базі двох існуючих одновимірних масивів, а також двовимірного масиву граматичних характеристик простих елементів формальної системи Z формується двовимірний масив SGZ , який представляє собою остаточний набір даних, що розміщується в робочій пам'яті бази знань і на основі якого механізм логічного виводу видає результат – одновимірний масив L , що фактично містить всі елементи початкового масиву слів S , відсортованих у відповідності з формулою (2.4), яка являється однією з аксіом логіко-лінгвістичної моделі та інтерпретує прості висловлювання типу L^S [74].

$$S \Rightarrow S = \begin{pmatrix} S_1 \\ S_2 \\ \dots \\ S_i \\ \dots \\ S_r \end{pmatrix} \Rightarrow G = \begin{pmatrix} G_1 \\ G_2 \\ \dots \\ G_i \\ \dots \\ G_r \end{pmatrix} \Rightarrow SGZ = \begin{pmatrix} S_{10} & G_{i0} & cm_{11} & g_{12} & n_{13} & k2_{14} & t_{15} & h_{16} & l_{17} & ch_{18} \\ S_{20} & G_{20} & cm_{21} & g_{22} & n_{23} & k2_{24} & t_{25} & h_{26} & l_{27} & ch_{28} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ S_{i0} & G_{i0} & cm_{i1} & g_{i2} & n_{i3} & k2_{i4} & t_{i5} & h_{i6} & l_{i7} & ch_{i8} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ S_{r0} & G_{r0} & cm_{r1} & g_{r2} & n_{r3} & k2_{r4} & t_{r5} & h_{r6} & l_{r7} & ch_{r8} \end{pmatrix} \Rightarrow L(S).$$

Для повного розуміння змісту речення необхідно прослідкувати зв'язки між простими елементами формальної

системи (словами) та принципом їх об'єднання в складні елементи словосполучення.

Формування складних елементів формальної системи відбувається шляхом звернення до бази знань B^{t3} . До робочої пам'яті бази знань послідовно заносяться прості елементи формальної системи $S_i, i = \overline{1, r}$. Використовуються формалізовані правила формування словосполучень, описані у другому розділі.

Результатом роботи механізму логічного виводу буде двовимірний масив SJ , кількість рядочків якого рівна кількості словосполучень в реченні, тобто $j = \overline{1, m}$, а значення елементів масиву представляють собою словосполучення, його, головний елемент, властивість залежного слова в словосполученні (узгодження, прилягання, керування), та безпосередньо залежну частину складного елемента формальної системи (рис. 3.14).

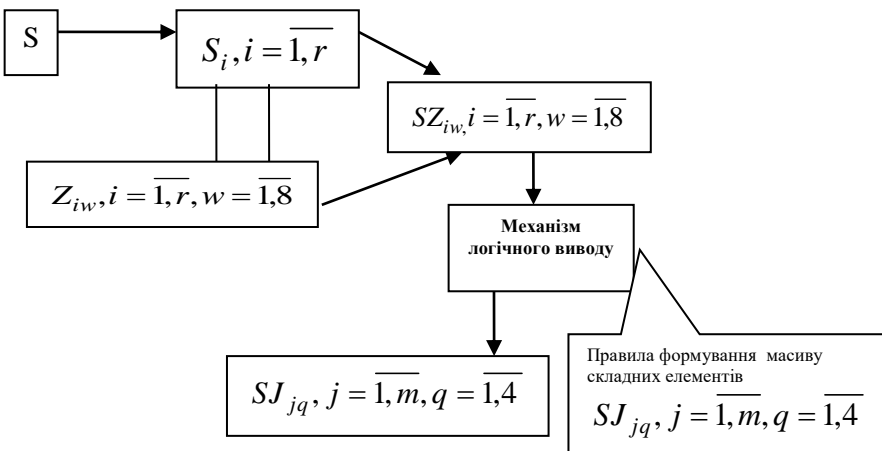


Рис. 3.14. Загальна схема формування аксіом

Кількість рядочків двовимірного масиву SJ дорівнює кількості аксіом (правил визначення компонентів логіко-лінгвістичної моделі). Узагальнивши інформацію щодо способів формування логіко-лінгвістичних моделей та засобів, які при цьому використовуються, можна представити загальну схему формування логіко-лінгвістичної моделі (рис. 3.15). Загальний алгоритм представленого методу (рис. 3.16).

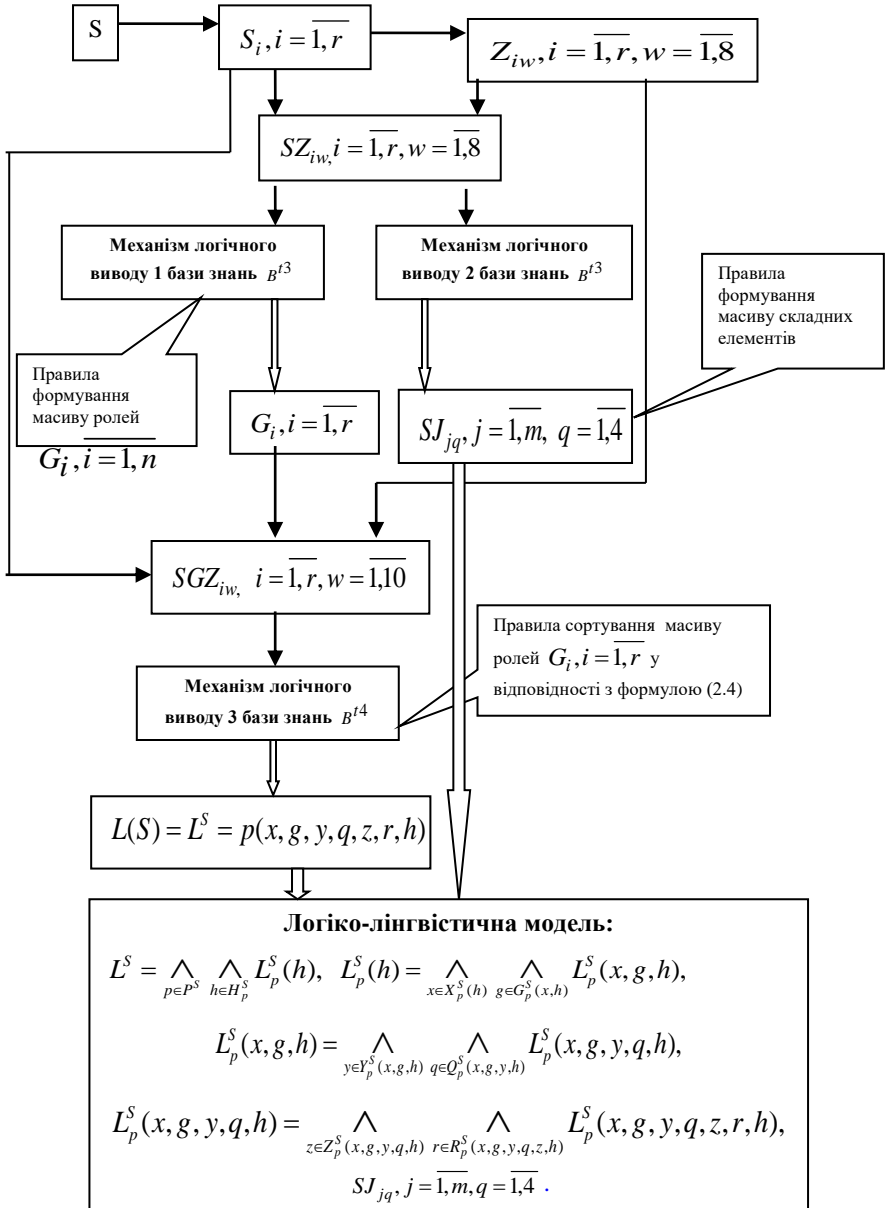


Рис. 3.15. Загальна схема формування логіко-лінгвістичної моделі

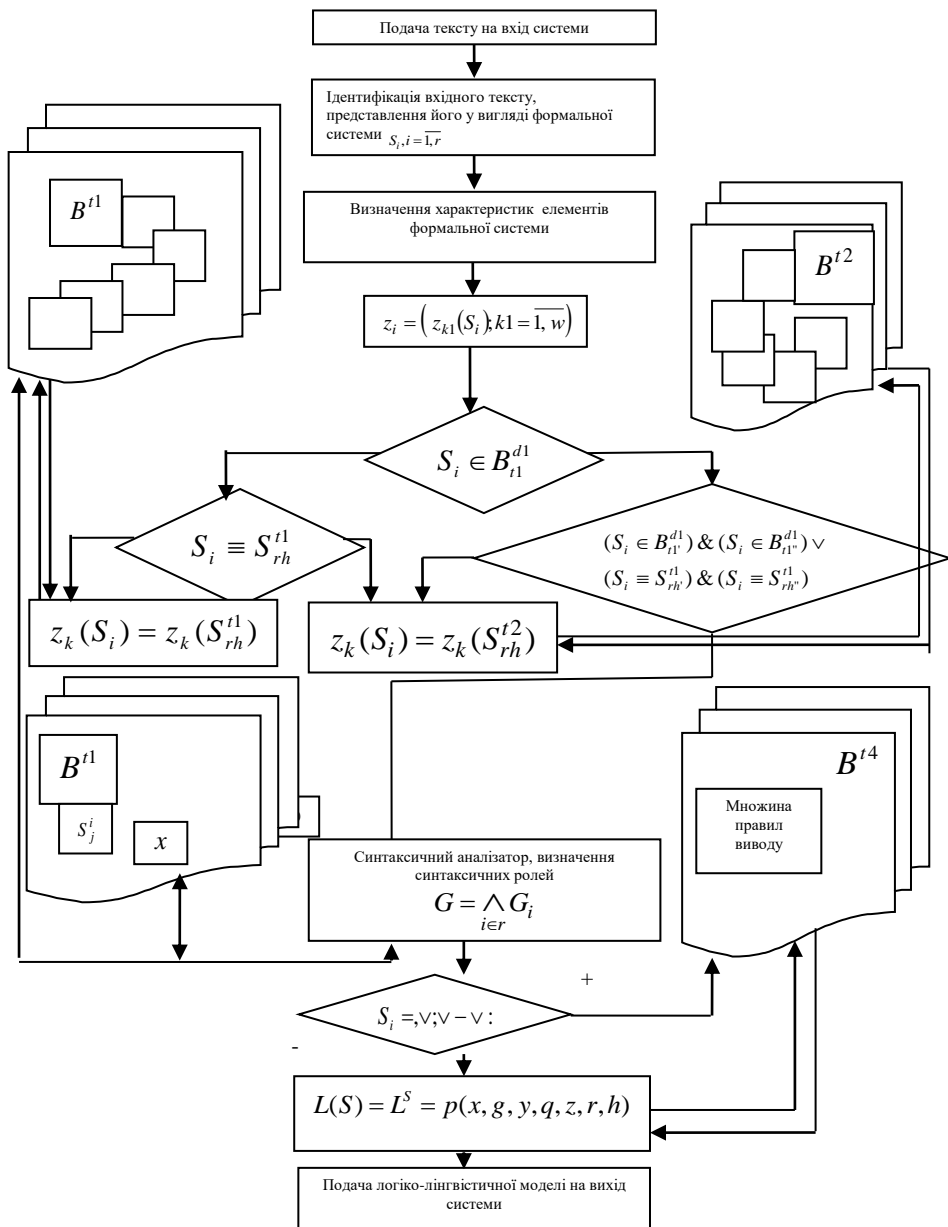


Рис. 3.16. Загальна схема алгоритму автоматизованого формування логіко-лінгвістичної моделі

IV. ПРИНЦИПИ ПОРІВНЯННЯ ЛОГІКО-ЛІНГВІСТИЧНИХ МОДЕЛЕЙ РЕЧЕНЬ ПРИРОДНОЇ МОВИ

Описати значення речення – значить визначити умови його істинності або хибності для тієї ситуації або стану, який воно має на меті описати [75]. Тобто знання значення речення є знання умов, за яких воно є істинним або хибним. Наприклад, речення «*У нас ніколи не було сніжної зими*» може бути істинним для країн Африки, проте буде хибним, якщо у тексті йтиметься про країни помірнього поясу.

Жодне з вище наведених формулювань не ототожнює значення речення з умовами його істинності, а також залишає відкритим питання про те, що саме мається на увазі під знанням істинності речення.

Таким чином, оцінка тверджень, що стосуються осмислення та семантичної еквівалентності, передбачає маніпулювання нормальними онтологічними пропозиціями, характерними для конкретного суспільства. Тут пропозиція – це дані, які виражені у вигляді речень, і є основою змістовного повідомлення, що міститься у тексті [76].

Тому велика увага повинна приділятися умовно-істинній еквівалентності як важливого і єдиному компоненту семантичної еквівалентності речень природної мови.

Принцип умовно-істинної еквівалентності діє не залежно від фактів, що описуються у конкретних ситуаціях: ***речення мають однаковий пропозиційний зміст тоді і лише тоді, коли вони мають однакові умови істинності*** [75].

Це означає, що говорити про істинність двох речень природної мови можна лише тоді, коли читає та сприймає їх людина за одних і тих самих умов (ситуацій), і яка достеменно знає факти, про які іде мова в реченнях, що порівнюються. Тобто, для ситуації, коли викладач оголошує оцінки за модуль, на якому не були присутні 5% студентів, речення «*За результатами модульної контрольної всі студенти засвоїли матеріал*» буде хибним, а речення «*За результатами модульної контрольної всі присутні на модулі студенти засвоїли матеріал*» – істинним.

Теорія умовно-істинної семантики сформульована так, що вона виключає випадки невизначеності у семантичній структурі природних мов, і їх можна одразу ж відкидати з розгляду [77].

4.1 Протиріччя і тавтології у реченнях природної мови

Для того, щоб надалі працювати з формулами логіки предикатів та навчитися порівнювати логіко-лінгвістичні моделі за змістом, необхідно чітко уявляти, що представляють собою тавтології та протиріччя у реченнях природної мови.

Тавтології – це пропозиції, які є необхідно істинними в силу їх логічної форми. Наприклад, речення «Або ця інформація цікава, або не цікава» містить тавтологію.

Протиріччя – це пропозиції, які обов'язково хибні в силу своєї логічної форми. Так, речення «Ця інформація цікава, і ця інформація не цікава» містить протиріччя внаслідок об'єднання сурядним зв'язком двох простих речень, що суперечать один одному.

Логічне протиріччя – це твердження, в якому одночасно стверджується наявність певної ситуації A та її відсутність $\neg A$. Два вирази, в яких містяться протиріччя, не можуть бути одночасно ні істинними, ні хибними. Це підтверджується другим законом логіки про не протиріччя. Саме логічні закони дають змогу зафіксувати властивості та ознаки, що постійно відбуваються у світі (визначеність, унікальність, стабільність), та простежити правильність побудови логічних ланцюжків [78].

Закон тотожності – довільне поняття або висловлювання тотожне само собі. Якщо керуватися цим законом, то в процесі дослідження не можна замінити конкретні поняття іншими поняттями. Проте на практиці, а особливо при написанні текстів на природній мові, заміни відбуваються дуже часто, наприклад, явища синонімії та омонімії, що само по собі є генератором помилок, пов'язаних із застосуванням закону тотожності. Таким чином, закон тотожності *забезпечує визначеність логічного мислення* і розуміє під собою те, що не можна ототожнювати різні думки, а тотожні думки сприймати як не тотожні.

Закон протиріччя – два протилежні висловлювання не можуть бути одночасно істинними, принаймні одне з них має бути хибним. Якщо перший закон виражає відношення логічної однозначності, то другий закон – відношення логічної несумісності. Математичний запис закону протиріччя: $\neg(A \ \& \ \neg A) = 1$. Цей закон є фундаментальним логічним законом, на якому побудована вся сучасна математика. Закон протиріччя діє у

відношенні довільних контрарних та контрадикторних висловлювань.

Контрарні висловлювання – висловлювання, в одному з яких говориться про ствердження деякого предмету із певної множини, а у іншому – заперечується деяка частина цієї множини. Ці висловлювання не можуть бути одночасно ні істинними, ні хибними. Наприклад, висловлювання «*всі громадяни України приймають участь у виборах*» та «*деякі громадяни України не приймають участь у виборах*» – контрарні.

Контрадикторні висловлювання – висловлювання про предмет, в одному з яких дещо стверджується, а в іншому те ж саме заперечується. Наприклад, висловлювання «*всі громадяни України приймають участь у виборах*» та «*не всі громадяни України приймають участь у виборах*» – контрадикторні.

В основі закону протиріччя лежить якісна визначеність речей та явищ, відносна стійкість їх властивостей. Проте потрібно чітко розуміти, що другий закон заперечує тільки логічні протиріччя. Це протиріччя неправильного мислення, а не об'єктивної реальності. Протиріччя зникає одразу, щойно встановлюється, що твердження відноситься до одного об'єкту, а заперечення – до іншого. Тобто закон заперечення *забезпечує не протиріччя і послідовність мислення*, здатність фіксувати та виправляти довільні протиріччя.

Закон вилучення третього – два протилежні висловлювання не можуть бути одночасно хибними, одне з них обов'язково істинне, інше – обов'язково хибне, третє висловлювання – виключено. Оскільки закон вилучення третього діє лише по відношенню до протилежних висловлювань, то його формулювання ведеться за формою «АБО-АБО». Цей закон задає напрямок мислення у пошуку істини про те, що можливі лише два варіанти рішення проблеми, при чому одне з них обов'язково істинне. Доведення цього закону здійснюють методом доведення від протилежного. Важливе застосування закон має в юридичній практиці, де потребується категоричне вирішення питання. В результаті застосування закону вилучення третього *досягається однозначність логічного мислення*.

Закон достатнього обґрунтування – довільне висловлювання має достатньо підстав. Цей закон фактично говорить про те, що всі думки, які можна пояснити, вважаються істинними, а ті, що не

можна пояснити, – хибними. Закон носить змістовний характер, виступає як методологічний принцип, що забезпечує здатність мислення порівнювати підстави з подальшими доведеннями. Усі види протиріч здійснюються та вирішуються, знімаються та створюються. Протиріччя як джерело розвитку дійсне тільки разом з його вирішенням. Тобто достатньою підставою для довільного висловлювання є інше висловлювання, уже перевірене та визнане істинним. Закон достатнього обґрунтування *забезпечує обґрунтованість мислення* [79].

Без логічного закону неможливо зрозуміти, що таке логічний вивід та що таке логічне доведення. Правильне мислення – це мислення за законами логіки, за абстрактними схемами, що ними фіксуються. Закони логіки утворюють каркас, на якому тримаються послідовні розмірковування і без якого вони перетворюються на хаотичну мову [80].

Довести відсутність внутрішнього протиріччя висловлювання означає довести, що у ньому не існує такої формули ψ , такої, що і вона, і її заперечення $\bar{\psi}$ виводяться з цього висловлювання. Для того, щоб довести істинність певного висловлювання, необхідно металогічними засобами довести неможливість формального.

Нехай потрібно знайти логічне протиріччя між двома логічними виразами:

$$p_1(x_1, g_1, y_1, 0, z_1, r_1, 0) \rightarrow U,$$

$$\neg p_1(x_1, g_1, y_1, 0, z_1, r_1, 0) \rightarrow U.$$

Зробимо заміну $x_1, g_1 = Q, y_1, 0, z_1 = R, r_1, 0 = T$, тоді матимемо:

$$p_1(Q, R, T) \rightarrow U,$$

$$\neg p_1(Q, R, T) \rightarrow U.$$

Якщо буде доведено істинність одного із заданих висловлювань, то інше буде хибним, відповідно обидва висловлювання будуть містити логічне протиріччя [81].

Для доведення істинності логічного висловлювання методом резолюцій, необхідно заперечити той консеквент, який впливає із заданого логічного висловлювання, тобто $\neg U$. Це заперечення додається до множини $S = \{p_1(Q, R, T), U, \neg U, \neg p_1(Q, R, T)\}$, і

всі члени розширеної множини перетворюються у послідовність диз'юнктивів: $\{\neg U, \neg p_1(Q, R, T) \vee U, \neg p_1(Q, R, T), p_1(Q, R, T) \vee U\}$.

Здійснюємо резолюцію заперечення припущення $\neg U$ та послідовності диз'юнктивів, отриманих у результаті заміни $\neg p_1(Q, R, T) \vee U$, отримуємо резольвенту $\neg p_1(Q, R, T)$. Аналогічні дії необхідно зробити і для другого виразу, внаслідок чого отримаємо резольвенту $p_1(Q, R, T)$. Резольвентою отриманих двох логічних виразів $\neg p_1(Q, R, T)$ та $p_1(Q, R, T)$ є пуста множина, це означає, що вирази містять логічні протиріччя.

У стандартній пропозиційній логіці тавтологія та протиріччя відображаються за допомогою логічного «не», кон'юнкції та диз'юнкції.

Логічні істини, або тавтології, є підкласом аналітичних істин, тобто пропозицій, істинність яких повністю визначається їх значенням.

Для того, щоб сформулювати принципи порівняння логіко-лінгвістичних моделей речень природної мови, необхідно зрозуміти, що вкладається у зміст цих речень, а також те, коли текстові одиниці вважаються однаковими за змістом (тавтологіями) і коли суперечать одне одному.

Речення мають один і той самий пропозиційний зміст тоді і лише тоді, коли вони мають ті ж умови істинності. Існує таке поняття, як тематичне значення, яке може бути різним для декількох речень, що мають однакові умови істинності і, відповідно, однаковий пропозиційний зміст [75]. Наприклад, наступні речення відрізняються за тематичним значенням, але мають одні і ті самі умови істинності та пропозиційний зміст.

Перше речення природної мови: *«Ці речення відрізняються за своїм тематичним значенням».*

Логіко-лінгвістична модель даного речення:

$$L(S) = L^S = p_1(x_1, g_1, y_1, q_{11}, 0, 0, 0) \& p_1(x_1, g_1, y_1, q_{12}, 0, 0, 0),$$

L(S) = L^S = відрізняються (речення, ці, значенням, своїм, 0, 0, 0) & відрізняються (речення, ці, значенням, тематичним, 0, 0, 0).

Друге речення природної мови *«За своїм тематичним значенням ці речення відрізняються».*

Логіко-лінгвістична модель даного речення:

$L(S) = L^S = p_1(x_1, g_1, y_1, q_{11}, 0, 0, 0) \& p_1(x_1, g_1, y_1, q_{12}, 0, 0, 0),$

$L(S) = L^S =$ відрізняються (речення, ці, значенням, своїм, 0, 0, 0) & відрізняються (речення, ці, значенням, тематичним, 0, 0, 0).

Третє речення природної мови «Це саме ті речення, які відрізняються за своїм тематичним значенням».

Логіко-лінгвістична модель даного речення:

$L(S) = L^S = p_1(x_1, g_1^*, y_1, q_{11}, 0, 0, 0) \& p_1(x_1, g_1^*, y_1, q_{12}, 0, 0, 0),$

$L(S) = L^S =$ відрізняються (речення, саме ті, значенням, своїм, 0, 0, 0) & відрізняються (речення, саме ті, значенням, тематичним, 0, 0, 0).

Із наведеного прикладу видно, що логіко-лінгвістичні моделі цих речень однакові.

Ще декілька нюансів, які потрібно враховувати при оцінці ідентичності змісту речень природної мови, – це порядок слів, граматична структура, а також пасивний та активний стан.

Наприклад, речення «Експерти висунули ряд гіпотез».

Логіко-лінгвістична модель даного речення:

$L(S) = p_1(x_1, 0, y_1, 0, z_1, 0, 0),$

Висунули (експерти, 0, ряд, 0, гіпотез, 0, 0).

Інтерпретація тотожного за змістом до заданого речення: «Експертами висунуто ряд гіпотез».

Логіко-лінгвістична модель даного речення:

$L(S) = p_1(0, 0, y_1, 0, z_1, c_1(z_1), 0),$

Висунуто (0, 0, експертами, 0, ряд, гіпотез, 0).

Хоча дані речення описуються різними логіко-лінгвістичними моделями, проте мають однаковий пропозиційний зміст.

Таким чином, мова написання речення обмежується знанням структури речення, а значення речення є умовно-істинним. Тому важливо вміти правильно виділяти атомарні семантичні компоненти, за допомогою яких можна описувати значення слів та визначати істинність їх комбінацій. Якщо різні частини речення містять поняття з суперечливими властивостями, то таке висловлювання є хибним [82].

4.2 Вплив функціонально-істинних операцій на зміст речень природної мови

Існує гіпотеза «правило на правило», згідно з якою між *граматичною структурою речення та логічною формою існує повна відповідність*. Тобто для аналізу пропозиційного змісту будуть використовуватися традиційні поняття, що стосуються граматичної структури речень природної мови. А те, які операції логіки предикатів будуть використовуватися в логіко-лінгвістичних моделях, залежить від типів речень природної мови.

Кожне правило граматики пов'язане з семантичним правилом, що приписує інтерпретацію для складного виразу, який утворюється шляхом цього ж конкретного граматичного правила. Операції, за допомогою яких складні пропозиції утворюються з простих, є *функціонально-істинними*. Це означає, що істинне значення складної пропозиції повністю визначається виходячи з істинних значень пропозицій, які його утворюють, а також специфічного ефекту кожної операції [75].

Операція кон'юнкції утворює складну пропозицію $(L'(S_1) \& L''(S_1))$, яка є істинною тоді і лише тоді, коли обидві її частини істинні. Ця операція характерна для простих речень з однорідними членами та для складносурядних речень природної мови.

При порівнянні логіко-лінгвістичних моделей речень природної мови з присутньою у них операцією кон'юнкції необхідно враховувати наступні нюанси.

1. Між ситуаціями, які описуються пропозиціями-компонентами, є темпоральний або каузальний зв'язок, внаслідок чого важливим є порядок вживання речень.

Наприклад, речення «*Вчений не доповідався на конференції і не зміг захистити дисертацію*» та «*Вчений не зміг захистити дисертацію і не доповідався на конференції*».

Логіко-лінгвістична модель першого речення:

$$L(S) = L'(S_1) \rightarrow L''(S_1) = \neg p'_1(x'_1, 0, y'_1, 0, 0, 0, 0) \rightarrow \neg p''_1(x'_1, 0, y''_1, 0, 0, 0, 0),$$

$$L'(S_1) \rightarrow L''(S_1) = \neg \text{ доповідався (вчений, 0, конференції, 0, 0, 0, 0)} \rightarrow$$

$$\neg \text{ зміг \& захистити (вчений, 0, дисертацію, 0, 0, 0, 0)}.$$

Логіко-лінгвістична модель другого речення:

$L(S) = L''(S_1) \& L'(S_1) = \neg p_1''(x'_1, 0, y''_1, 0, 0, 0, 0) \& \neg p_1'(x'_1, 0, y'_1, 0, 0, 0, 0),$
 $L(S) = L''(S_1) \& L'(S_1) = \neg \text{ зміг \& захистити(вчений, 0,}$
дисертацію, 0,0,0,0) & $\neg \text{ доповідався(вчений, 0, конференції, 0,0,0,0)}.$

Ці моделі, як і речення природної мови, не є тавтологіями і повинні використовуватися за різних обставин. Сполучник «і» у першому реченні вживається у значенні «і тому», а в другому реченні – «разом з цим». Таким чином, у першому реченні сполучник «і» не еквівалентний показнику пропозиційної кон'юнкції, оскільки $(L'(S_1) \& L''(S_1))$ не має ті ж значення істинності, що й $(L''(S_1) \& L'(S_1))$.

2. Сполучник «і» може використовуватися у реченнях природної мови в значенні «проте» та «хоча». У цьому випадку теж не можна говорити про нейтральний показник кон'юнкції пропозиційного змісту. Тобто речення «Вчений не доповідався на конференції і не зміг захистити дисертацію» та «Вчений не зміг захистити дисертацію і не доповідався на конференції» з різною інтерпретацією сполучника «і» матимуть різне пропозиційне значення, а відповідно і зміст.

Аналогічна ситуація і для складносурядних речень. Наприклад, речення «Вчений опублікував понад 100 наукових статей – і експерти не призначили премію» та «Експерти не призначили премію – і вчений опублікував понад 100 наукових статей».

Логіко-лінгвістична модель першого речення:

$L(S) = L'(S_1) \& L''(S_1) = p_1'(x'_1, g'_1, y'_1, q'_1, 0, 0, h'_1) \& \neg p_1''(x''_1, 0, y''_1, 0, 0, 0, 0),$
 $L(S) = L'(S_1) \& L''(S_1) = \text{опублікував (вчений, 0, статей,}$
наукових, 0,0, понад \& 100) & $\neg \text{ призначили (експерти, 0, премію, 0,0,0,0)}.$

Логіко-лінгвістична модель другого речення:

$L(S) = L''(S_1) \rightarrow L'(S_1) = \neg p_1''(x''_1, 0, y''_1, 0, 0, 0, 0) \rightarrow p_1'(x'_1, g'_1, y'_1, q'_1, 0, 0, h'_1),$
 $L(S) = L''(S_1) \rightarrow L'(S_1) = \neg \text{ призначили (експерти, 0, премію, 0,0,0,0)}$
 $\rightarrow \text{ опублікував (вчений, 0, статей, наукових, 0,0, понад \& 100)}.$

Речення різні за змістом, а тому мають різні логіко-лінгвістичні моделі.

Операція диз'юнкції утворює складну пропозицію $(L'(S_1) \vee L''(S_1))$, яка є істинною тоді і тільки тоді, коли або висловлювання $L'(S_1)$, або $L''(S_1)$ істинне, або обидва висловлювання істинні.

Розрізняють інклюзивну та ексклюзивну диз'юнкцію. Залежно від того, який тип диз'юнкції використано у реченнях природної мови, їх можна вважати еквівалентними чи навпаки [75].

1. *Інклюзивна диз'юнкція* $(L'(S_1) \vee L''(S_1))$ істинна не тільки тоді, коли один із членів диз'юнкції або $L'(S_1)$, або $L''(S_1)$ істинний, а інший хибний, проте і тоді, коли як $L'(S_1)$, так і $L''(S_1)$ істинні.

Наприклад, речення «*Студенти, що не прийшли вчасно або не виконали завдання, не будуть допущені до екзамену*» трактується як таке, що не лише студенти, які не виконали одну із цих умов, але і студенти, які не виконали обидві умови, не будуть допущені до екзамену.

Логіко-лінгвістична модель для цього речення буде мати вигляд:

$$L(S) = (L'_{11}(S_1) \& L'_{12}(S_1)) \& L''(S_1) = (\neg p'_{11}(x'_1, 0, 0, 0, 0, 0, h'_1) \& \neg p'_{12}(x'_1, 0, y'_1, 0, 0, 0, 0)) \& \neg p''(x'_1, 0, y''_1, 0, 0, 0, 0),$$

$$L(S) = L'(S_1) \& L''(S_1) = (\neg \text{прийшли(студенти, } 0, 0, 0, 0, 0, \text{вчасно)}) \& \neg \text{виконали(студенти, } 0, \text{завдання, } 0, 0, 0, 0) \& \neg \text{будуть\& допущені(студенти, } 0, \text{екзамену, } 0, 0, 0, 0).$$

2. *Ексклюзивна диз'юнкція* істинна лише тоді, коли або $L'(S_1)$ істинно і $L''(S_1)$ хибне, або $L''(S_1)$ істинне і $L'(S_1)$ хибне. Тобто виключається можливість одночасної істинності $L'(S_1)$ і $L''(S_1)$.

На відміну від інклюзивної диз'юнкції, речення «*Студенти, що не прийшли вчасно або не виконали завдання, не будуть допущені до екзамену*» трактується як таке, що до екзамену не будуть допущені студенти, що не виконали одну із умов і не обов'язково студенти, які не виконали обидві умови.

Тоді логіко-лінгвістична модель для цього речення буде мати вигляд:

$$L(S) = (L'_{11}(S_1) \& L'_{12}(S_1)) \& L''(S_1) = (\neg p'_1(x'_1, 0, 0, 0, 0, 0, h'_1) \vee \neg p'_1(x'_1, 0, y'_2, 0, 0, 0, 0)) \& \neg p''_1(x'_1, 0, y''_1, 0, 0, 0, 0),$$

$$L(S) = (L'_{11}(S_1) \& L'_{12}(S_1)) \& L''(S_1) = (\neg \text{прийшли(студенти, 0, 0, 0, 0, 0, вчасно)} \vee \neg \text{виконали(студенти, 0, завдання, 0, 0, 0, 0)}) \& \neg \text{будуть \& допущені(студенти, 0, екзамену, 0, 0, 0, 0)}.$$

Таким чином, при інклюзивній та ексклюзивній диз'юнкції речення природної мови будуть представлені у вигляді логіко-лінгвістичних моделей, що відрізнятимуться логічними операціями.

Операція імплікації утворює складну пропозицію $(L'(S_1) \rightarrow L''(S_1))$, яка є істинною лише тоді, коли і $L'(S_1)$, і $L''(S_1)$ істинні.

За допомогою імплікації передаються складнопідрядні речення природної мови з підрядними з'ясувальними та підрядними обставинними реченнями. Наприклад, речення «*Вчені дуже засмутились, що не отримали премію за багаторічну роботу*».

Логіко-лінгвістична модель для цього речення буде мати вигляд:

$$L(S) = L'(S_1) \rightarrow L''(S_1) = \neg p'_1(x'_1, 0, y'_1, 0, z'_1, r'_1, 0) \rightarrow p''_1(x'_1, 0, 0, 0, 0, 0, h''_1),$$

$$L(S) = L'(S_1) \rightarrow L''(S_1) = \neg \text{отримали(вчені, 0, премію, 0, роботу, багаторічну, 0)} \rightarrow \text{засмутились(вчені, 0, 0, 0, 0, 0, дуже)}.$$

1. Пропозиція $(L'(S_1) \rightarrow L''(S_1))$ істинна не лише тоді, коли $L'(S_1)$ і $L''(S_1)$ мають одне і те саме значення істинності (обидва істинні або обидва хибні), але й також коли $L'(S_1)$ хибне, а $L''(S_1)$ істинне. Тобто «*вчені не отримали премію за багаторічну роботу*», але могли не засмутитися. У більшості випадків такі ситуації парадоксальні. Сам факт, що довільна хибна пропозиція імплікує кожен істинну пропозицію, розглядається як один із парадоксів імплікації [75].

2. Особливістю імплікації є те, що істинне значення $(L'(S_1) \rightarrow L''(S_1))$ імплікує, подібно значенню $(L'(S_1) \& L''(S_1))$ повністю незалежне від довільного каузального зв'язку між ситуаціями, описаними пропозиціями-компонентами. Наприклад,

речення «Якщо експерти запропонували нову модель оцінки, то вчені не отримали премію за багаторічну роботу».

Логіко-лінгвістична модель для цього речення буде мати вигляд:

$$L(S) = L'(S_1) \rightarrow L''(S_1) = p'_1(x'_1, 0, y'_1, q'_1, z'_1, 0, 0) \rightarrow \\ \neg p''_1(x''_1, 0, y''_1, 0, z''_1, r'_1, 0),$$

$L(S) = L'(S_1) \rightarrow L''(S_1) =$ запропонували (експерти, 0, модель, нову, оцінки, 0, 0) \rightarrow \neg отримали (вчені, 0, премію, 0, роботу, багаторічну, 0).

Пропозиція, виражена у цьому реченні, буде істинною (не залежно від того, чи «запропонували експерти нову модель оцінки»), якщо «вчені не отримують премію за багаторічну роботу». Тобто на функціонально-істинне значення імплікації не впливає відсутність причинно-наслідкового зв'язку між ситуаціями, позначеними двома пропозиціями-компонентами $L'(S_1)$ та $L''(S_1)$.

Операція заперечення утворює складну пропозицію $\neg L'(S_1)$ із простої пропозиції $L'(S_1)$ і визначається таким чином, що пропозиція $L'(S_1)$ є істинною тоді і лише тоді, коли $\neg L'(S_1)$ є хибною, і $L'(S_1)$ хибна, коли $\neg L'(S_1)$ істинна.

Часто вживається рекурсивне заперечення, тобто заперечення $\neg L'(S_1)$ дає $L'(S_1)$, що еквівалентно, так як подвійне заперечення дає ствердження. Прослідкуємо, як логічний опис заперечення співвідноситься зі значенням заперечних речень природної мови.

Речення природної мови, що містять пропозиції протилежної полярності [75], мають однакову структуру, але заперечення предиката еквівалентно запереченню пропозиції.

1. Існує такий вид предикатного заперечення, який не еквівалентний усій пропозиції. Наприклад, речення «Викладач не адекватний» та «Викладач неадекватний».

Логіко-лінгвістичні моделі даних речень будуть мати вигляд:

$$L(S) = L^S = \neg p_1(x_1, 0, 0, 0, 0, 0, 0),$$

$$L(S) = L^S = \neg \text{адекватний (викладач, 0, 0, 0, 0, 0, 0)},$$

$$L(S) = L^S = p_1(x_1, 0, 0, 0, 0, 0, 0),$$

$L(S) = L^S = \text{неадекватний (викладач, 0, 0, 0, 0, 0)}$.

Пропозиція є контрарною до іншої, якщо вони обидві не можуть бути істинними, однак обидві можуть бути хибними. Тому перша пропозиція є контрарною по відношенню до пропозиції «*Викладач адекватний*».

2. До складу речень природної мови можуть входити заперечні іменні вирази. Наприклад, речення «*Не бюджет платить за навчання повністю*» та «*Бюджет не платить за навчання повністю*».

Логіко-лінгвістичні моделі даних речень будуть мати вигляд:

$L(S) = L^S = p_1(\neg x_1, 0, y_1, 0, 0, 0, h_1)$,

$L(S) = L^S = \text{платить} (\neg \text{бюджет}, 0, \text{навчання}, 0, 0, 0, \text{повністю})$.

$L(S) = L^S = \neg p_1(x_1, 0, y_1, 0, 0, 0, h_1)$,

$L(S) = L^S = \neg \text{платить} (\text{бюджет}, 0, \text{навчання}, 0, 0, 0, \text{повністю})$.

Саме заперечення такого типу, як і предикатне заперечення, впливає на пропозиційний зміст речення та носить функціонально-істинний характер.

3. Заперечення може застосовуватися як до простого $\neg L'(S_1)$, так і до складного речення природної мови $\neg(L'(S_1) \rightarrow L''(S_1))$, тому існує суттєва різниця між виразами $\neg(L'(S_1) \rightarrow L''(S_1))$ і $(\neg L'(S_1) \rightarrow L''(S_1))$.

Наприклад, речення «*Студент не виконав домашнє завдання, тому що святкував день народження*» та «*Студент не виконав домашнє завдання, не тому що святкував день народження*».

Логіко-лінгвістичні моделі для таких складнопідрядних причинно-наслідкових речень будуть мати вигляд:

$L(S) = \neg L'(S_1) \rightarrow L''(S_1) = \neg p'_1(x'_1, 0, y'_1, q'_1, 0, 0, 0) \rightarrow p''_1(x''_1, 0, y''_1, 0, z''_1, 0, 0)$,

$L(S) = \neg L'(S_1) \rightarrow L''(S_1) = \neg \text{виконав} (\text{студент}, 0, \text{завдання}, \text{домашнє}, 0, 0, 0) \rightarrow \text{святкував} (\text{студент}, 0, \text{день}, 0, \text{народження}, 0, 0)$.

$L(S) = \neg(L'(S_1) \rightarrow L''(S_1)) = \neg p'_1(x'_1, 0, y'_1, q'_1, 0, 0, 0) \rightarrow$

$\neg(p''_1(x''_1, 0, y''_1, 0, z''_1, 0, 0))$,

$L(S) = \neg(L'(S_1) \rightarrow L''(S_1)) = \neg$ виконав (студент, 0, завдання, домашнє, 0, 0, 0) $\rightarrow \neg$ (святкував (студент, 0, день, 0, народження, 0, 0)).

У першому випадку заперечення відноситься до пропозиційного змісту головного речення, у другому – до складної пропозиції в цілому.

Отже, з вище наведених прикладів видно, що значення речення вичерпується його пропозиційним змістом і може бути експліційоване за допомогою умов істинності.

4.3 Особливості методу резолюцій при пошуку збігів у реченнях природної мови

Одним із основних методів пошуку протиріч у математичній логіці, що базується на логіці предикатів, є правило резолюцій. Якщо для двох диз'юнктивів існує атомарна форма, яка в один із диз'юнктивів входить додатною, а в іншій – від'ємною, то, викресливши відповідно з одного диз'юнкта додатне входження атомарної форми, а з другого – від'ємне та об'єднавши диз'юнкти, отримуємо резольвенту. Метод резолюцій базується на обчисленні резольвент. Існує теорема про те, що питання про доведення істинності довільної формули у логіці предикатів приводить до питання про доведення порожнього списку в обчисленні резольвент [83]. Тому доведення того, що список формул в обчисленні резольвент порожній, еквівалентне доведенню про хибність формули у логіці предикатів.

У *методі резолюцій* множина речень розглядається як складений предикат, що містить декілька простих предикатів, з'єднаних логічними функціями і кванторами існування та узагальнення [24]. Оскільки однакові за змістом предикати можуть мати різний вигляд, то речення перетворюються до клаузальної форми – різновиду кон'юнктивної нормальної форми, з якої видалені квантори існування, узагальнення, символи імплікації, рівнозначності і т.д.

Метод резолюцій розглядають [22-25,83], інтерпретуючи висловлювання S , як сукупність речень C_1, C_2, \dots, C_n . Вважається, що в S є такі речення, що якщо в одне з них входить

деякий літерал, то в інше входить його заперечення (контрарна пара):

$$C_1 = (L_1, L_2, L_4),$$

$$C_2 = (L_3, \neg L_4).$$

Їх викреслення формує новий диз'юнкт із складових частин диз'юнктів. Новий сформований диз'юнкт C_3 називається *резольвентою диз'юнктів* C_1 і C_2 :

$$C_3 = C_1 \cdot C_2 = (L_1, L_2, L_4) \cdot (L_3, \neg L_4) \rightarrow (L_1, L_2, L_3).$$

Тобто, резольвента C_3 , отримана з двох диз'юнктів C_1 і C_2 , є логічним наслідком цих двох диз'юнктів. Тоді, якщо інтерпретувати метод резолюцій на два довільних речення природної мови, можна стверджувати, що резольвентою двох протилежних за змістом речень буде пуста множина і навпаки.

Задачі, які вирішуються сьогодні методом резолюцій, здебільшого спрямовані на доказ того, що одне висловлювання впливає з іншого, або їх розв'язки використовуються для знаходження певного елемента згідно заданих умов.

Наприклад, вирішення такої задачі: «Якщо другий пілот завжди слідує за першим, а перший знаходиться в кабіні, то де другий пілот?» у правильно побудованій формі (ППФ) буде мати вигляд:

$$S = \begin{cases} (\forall x)\{AT(\text{перший_пілот}, x) \rightarrow AT(\text{другий_пілот}, x)\}; & (\text{аксіома}_1), \\ AT(\text{другий_пілот}, \text{кабіна}); & (\text{аксіома}_2), \end{cases}$$

де предикат AT означає, що дехто знаходиться у певному місці.

На запитання «Де другий пілот» можна дати відповідь, якщо довести, що ППФ $(\exists x)AT(\text{другий_пілот}, x)$ впливає із множини ППФ S , а потім знайти частковий випадок змінної x , який існує.

Основна ідея полягає у такому перетворенні запитання в ППФ, яка містить квантор існування. Якщо на основі даних фактів можна дати відповідь на запитання, то ППФ, побудована таким чином, буде логічно впливати з множини ППФ S . Після доведення залишається вилучити той частинний випадок лінгвістичної змінної, який відноситься до квантору існування і служить відповіддю на запитання.

Для визначення лінгвістичної змінної x із даного прикладу необхідно виконати такі дії:

1. Довести, що $(\exists x)AT(\text{другий_пілот}, x)$ випливає із множини ППФ S . Для цього ППФ, яку необхідно довести, заперечується, внаслідок чого отримано: $(\forall x)(\neg AT(\text{перший_пілот}, x))$.

2. Це заперечення додається до множини S , і всі члени розширеної множини перетворюються у форму речень (представляються у вигляді послідовності диз'юнкцій літералів).

Згідно з правилом виключення знаків імплікації отримано:

$$\neg AT(\text{перший_пілот}, x) \vee AT(\text{другий_пілот}, x).$$

Тоді послідовність диз'юнктивів літералів для даного прикладу буде мати вигляд:

$$\{\neg AT(\text{другий_пілот}, x),$$

$$\{\neg AT(\text{перший_пілот}, x) \vee AT(\text{другий_пілот}, x)\},$$

$$\{AT(\text{перший_пілот}, \text{кабіна})\}.$$

3. За допомогою методу резолюцій можна показати, що ця множина не задовільна (рис. 4.1).

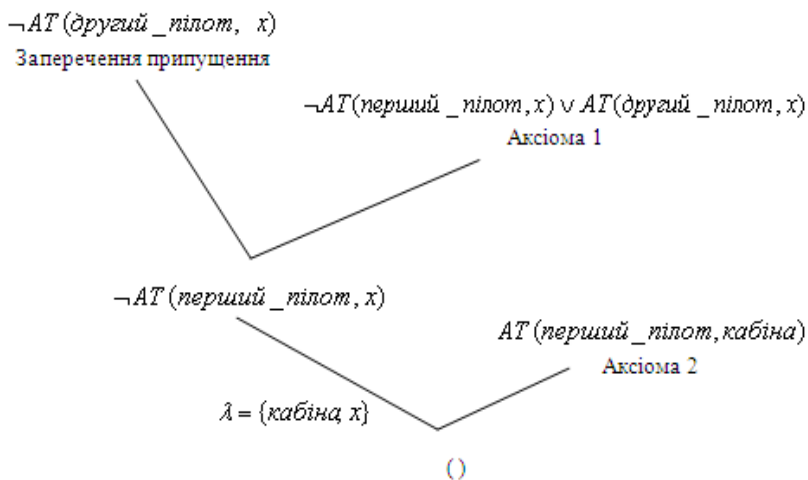


Рис. 4.1. Застосування методу резолюцій до послідовності диз'юнктивів

4. Із графа заперечення можна вилючити відповідь на запитання «Де другий пілот?» Для цього до кожного речення

додається його заперечення. Тоді заперечення « $\neg AT(\text{другий_пілот}, (x))$ » приймає вигляд:

$$\neg AT(\text{другий_пілот}, (x)) \vee AT(\text{другий_пілот}, (x)).$$

У відповідності зі структурою графа заперечення виконуються ті ж самі резолюції, що й раніше до того часу, поки в його корні не буде отримано деяке речення. Речення у кореневій вершині представляє собою відповідь-ствердження (рис. 4.2).

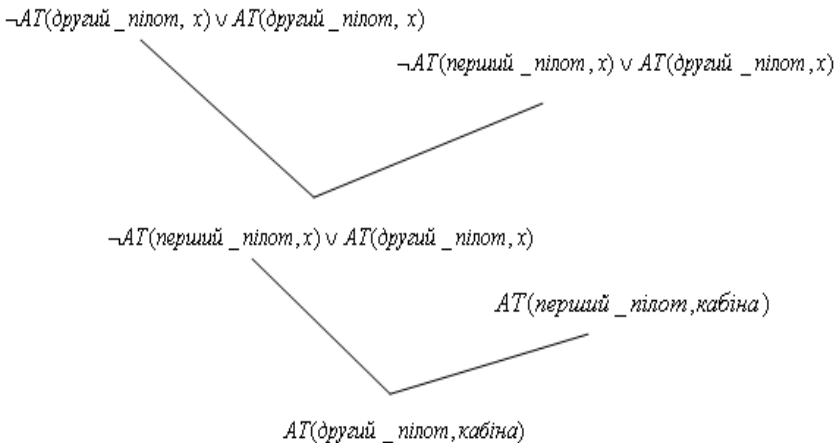


Рис. 4.2. Процес отримання відповіді-ствердження

Отже, метод резолюцій дає можливість знайти певний елемент із заданих висловлювань, в яких вказані певні початкові умови.

Якщо для логіки висловлювань знаходження контрарних пар не викликає труднощів, то для логіки предикатів це не так. Принцип резолюцій володіє важливою властивістю – повнотою, яка встановлюється наступною теоремою: множина диз'юнктивів S не виконується тоді і тільки тоді, коли існує вивід із S порожнього диз'юнкта.

У силу нерозв'язності логіки предикатів першого порядку для істинної множини диз'юнктивів S процедура, що базується на принципі резолюцій, буде працювати нескінченно довго.

Роботу методу резолюцій для логіки предикатів першого порядку можна продемонструвати на прикладі продукційної моделі представлення знань [2, 81], що відображає таку ситуацію.

Існують студенти, які цінують всіх викладачів. Жоден зі студентів не цінує не професіоналів. Відповідно, жоден із викладачів не являється не професіоналом.

Ці твердження на мові логіки предикатів будуть мати вигляд (4.1):

$$\begin{aligned} & \exists x(C(x) \ \& \ \forall y(P(y) \rightarrow L(x, y))) \\ & \forall x(C(x) \rightarrow \forall y(H(y) \rightarrow \neg L(x, y))) \end{aligned} \quad , \quad (4.1)$$

$$\forall y(P(y) \rightarrow \neg H(y))$$

де $C(x)$ – предикат, що означає « x є студент»;

$P(x)$ – предикат, що означає « x є викладач»;

$H(x)$ – предикат, що означає « x є професіонал»;

$L(x, y)$ – предикат, що означає « x цінує y ».

1. Нехай a – студент, тоді предикат, що означає студента буде $C(a)$;

2. За правилом виключення знаків імплікації перше та друге посилення набудуть вигляду:

$$\neg P(y) \vee L(a, y) \cdot (\text{друга} \cdot \text{частина} \cdot \text{першого} \cdot \text{посилання})$$

$$\neg C(x) \vee \neg(H(y) \vee \neg L(x, y)) \cdot (\text{друге} \cdot \text{посилання})$$

$$\neg \forall y(\neg P(y) \vee \neg H(y)) = \exists y(P(y) \ \& \ H(y)) \cdot (\text{консеквент})$$

3. Якщо b – викладач, то предикат, що означає викладача $P(b)$, а предикат, що означає викладач-професіонал – $H(b)$.

4. Принцип резолюцій дає змогу отримати $L(a, b)$ із $\neg P(y) \vee L(a, y)$ та $P(b)$ при $\sigma = \{b / y\}$.

5. Резольвента $\neg(H(y) \vee \neg L(a, y))$ отримана із $C(a)$ та $\neg C(x) \vee \neg(H(y) \vee \neg L(x, y))$ при $\sigma = \{a / x\}$.

6. Резольвентою $\neg(H(y) \vee \neg L(a, y))$ та $H(b)$ при $\sigma = \{b / y\}$ буде $\neg L(a, b)$.

7. Резольвентою $L(a, b)$ та $\neg L(a, b)$ є пуста множина, що говорить про те, що два посилення, які розглядалися – протилежні за змістом.

Такі результати отримано після проектування моделі (4.1) на конкретні значення x та y , а також після приведення її до стандартного вигляду.

На відміну від логіки висловлювань, у логіці предикатів першого порядку дуже велика кількість інтерпретацій формули, тому загальну значимість або заперечення цієї формули неможливо довести перевіркою її істинності при усіх можливих інтерпретаціях.

Проведемо дослідження, використовуючи уніфіковану форму логіко-лінгвістичної моделі, яка відображає зміст речень природної мови (2.1) – (2.4).

Нехай є два простих речення, кожному з яких поставлена у відповідність логіко-лінгвістична модель.

Перше речення: «Літак пролітає над нашим будинком», логіко-лінгвістична модель для нього представлена у вигляді:

$$L(S) = L^S = p_1(x_1, 0, y_1, q_1, 0, 0, 0),$$

Пролітає (літак, 0, будинком, нашим, 0, 0, 0).

Друге речення: «Літак пролітає над великим містом», логіко-лінгвістична модель для такого речення буде мати вигляд:

$$L(S) = L^S = p_1(x_1, 0, y'_1, q'_1, 0, 0, 0),$$

Пролітає (літак, 0, містом, великим, 0, 0, 0).

Множина атомів A буде містити літерали з однаковими предикатами:

$$A = p_1(x_1, 0, y_1, q_1, 0, 0, 0), p_1(x_1, 0, y'_1, q'_1, 0, 0, 0).$$

Згідно з алгоритмом уніфікації шукаємо підстановку $Q = \{y_1 / y'_1, q_1 / q'_1\}$. Тоді резольвента для цих двох речень буде мати вигляд

$$p_1(x_1, 0, y_1, q_1, 0, 0, 0), p_1(x_1, 0, y_1, q_1, 0, 0, 0).$$

Те, що резольвентою є не пуста множина, доводить, що речення не протилежні за змістом, а однакові. Проте насправді речення не однакові за змістом, у них лише однаковий предикат та предикатна змінна – суб'єкт, а предикатна змінна об'єкт та предикатна константа, яка його характеризує, – різні, що безпосередньо впливає на зміст речення.

Демонстрація конкретного випадку порівняння двох речень (власне, призначення методу), а також дослідження особливостей

здійснення алгоритму, дозволяють виявити ряд *недоліків методу резолюцій* [81].

1. Вибір літералів для множини атомів A відбувається тільки серед літералів із однаковими предикатами, проте при цьому не враховуються синоніми, рівнозначність активної та пасивної форми дієслів, що за певних умов також дають рівність предикатів за змістом.

2. Підстановка Q не бере до уваги зміст предикатних змінних та констант, що приводить до невірних результатів при порівнянні.

3. Велика кількість ітерацій для отримання кінцевого значення підстановки Q для складних та поширених речень.

4. Якщо предикати різні, до них не можна застосувати метод резолюцій.

4.4 Алгоритм порівняння логіко-лінгвістичних моделей речень природної мови

Спроби знайти ступінь зв'язку між різними парами слів у реченні природної мови робило багато вчених, зокрема, В. Левент [84] та А.Греймас [85], аналізуючи глибинну структуру речень природної мови. Наприклад, речення «*Студент бере підручники та йде до університету*» на етапі глибинної репрезентації розпадається на два елементарних висловлювання «*Студент бере підручники*» та «*Студент йде до університету*». Вчені відмічають, що зв'язок у словосполученні «*студент бере*» такий самий, як і в словосполученні «*студент йде*» [86]. При побудові логіко-лінгвістичної моделі даного речення зв'язки у наведених словосполученнях також однакові.

Існує припущення [87], що значимість слова залежить від того, яку кількість разів воно входить до глибинної структури речення (тобто кількість словосполучень, утворених із заданим словом).

Для здійснення семантичного аналізу речень природної мови застосовують різноманітні підходи. Один із них полягає у приписуванні всім словам речення деяких сутностей. Тобто, за теорією Фреге, поняття « t_1 співвідноситься з t_2 », де t_1 – структурний опис одиничного терму, а t_2 – сам терм. Така теорія дає можливість робити співставлення без звернення до певних

семантичних понять, крім базисного «співвідноситься з» [88]. Таким чином, буде ефективно виконуватися процедура співвіднесення довільного одиничного терму, що належить до певного універсуму, з тим, що він означає [89].

Відомо, що предикати виступають у ролі певних функціональних виразів, а речення – у ролі складних одиничних термів. Логічно еквівалентні одиничні терми мають один і той самий референт (об'єкт, про який говориться у контексті даного речення). Складний одиничний терм не змінює свій референт, якщо замінити одну із його складових, яка має те ж саме значення. Тоді, якщо t_1 і t_2 – це слова двох речень з однаковими істинними значеннями, то всі речення, в які будуть входити ці слова, повинні мати однакове істинне значення, а це не вірно. Зміст речення залежить, у першу чергу, від відношень предикату, суб'єкту та об'єкту.

Вимога ж, що ставиться до семантичної теорії формальної мови L , полягає у тому, щоб без звернення до семантичних понять накладати на предикат «є P » обмеження, достатні для отримання із схеми (4.2) всіх речень, у яких t_1 замінено структурним описом речення, а t_2 – самим реченням.

$$t_1 \in P \text{ тоді і тільки тоді, коли } t_2. \quad (4.2)$$

Довільні два предиката, що задовольняють цим умовам, мають один і той самий екстенціонал [90] (об'єм поняття, тобто множину об'єктів, що складають мовну одиницю). Речення, до яких застосовується цей предикат, будуть істинними реченнями формальної мови L , оскільки умова що накладається на семантичну теорію (4.2), фактично, повторює адекватність формального семантичного визначення істинності.

Отже, описана семантична теорія показує, яким чином значення речень залежать від значення слів за умови, що на предикат речення накладається ряд обмежень, а сам предикат співвідноситься з істинними реченнями.

Також усім реченням природної мови притаманний певний порядок слів: прямий (суб'єкт, предикат, аргументи і т.д.) або інверсний (аргументи, предикат, суб'єкт, означення і т.д.).

Логічна формула (2.4) є інтерпретацією синтаксичної структури речень з урахуванням семантичних зв'язків, що є формальним засобом відображення змісту текстової інформації.

Значення речення представляє собою не елементарну, складну структуру. Для порівняння речень природної мови, представлених у вигляді логіко-лінгвістичних моделей, необхідно враховувати відмінкову граматику – опис семантики речення без модальної, комунікативної та інших рамок, за допомогою поняття глибинного відмінку. Принциповою для відмінкової граматики є різниця декількох рівнів глибинних та поверхневих структур. У її основі лежить поняття відмінкової рамки, функція якої полягає у простеженні зв'язків між описом ситуації та глибинними синтаксичними структурами. Відмінкова рамка приписує семантико-синтаксичні ролі конкретним учасникам ситуації [90].

Аналіз логічної структури речення передбачає визначення пропозиції – відношень та зв'язків, що створюють певну ієрархію.

Ч.Філлмор ввів шість необхідних відмінків [91]:

- *агентив* – відмінок ініціатора дії, що ідентифікується дієсловом;
- *інструменталіс* – відмінок певного предмета, що впливає на дію або стан, позиціонується як причина;
- *датив* – відмінок певного об'єкта, на який впливає дія або стан;
- *фактив* – відмінок предмета або істоти, що виникає внаслідок здійснення дії або розуміється як частина дієслова;
- *локатив* – відмінок, яким характеризується місце знаходження або просторова орієнтація дії;
- *об'єктив* – семантично найбільш нейтральний відмінок, що визначається семантичною інтерпретацією самого дієслова.

Якщо розглянути два речення ідентичної будови «*Викладач відчинив двері*» та «*Ключ відчинив двері*», то, незважаючи на те, що структура речень однакова, слово «*викладач*» виступає у реченні агнетивом, а слово «*ключ*» – інструменталісом. При побудові логіко-лінгвістичних моделей цих двох речень і «*викладач*» і «*ключ*» позиціонуються як суб'єкти:

$$L(S) = L^S = p_1(x_1, 0, y_1, 0, 0, 0, 0),$$

$$L(S) = L^S = \text{відчинив (викладач, 0, двері, 0, 0, 0, 0)},$$

$$L(S) = L^S = p_1(x'_1, 0, y_1, 0, 0, 0, 0),$$

$L(S) = L^S = \text{відчинив (ключ, 0, двері, 0,0,0,0)}$, при чому $x_1 \neq x'_1$.

Тобто побудова логіко-лінгвістичної моделі згідно формули (2.4) дає змогу встановити відношення між структурними елементами речення, проте не бере до уваги їх зміст (вважається, що речення істинне і несе у собі певні знання).

Дослідити істинність того чи іншого речення природної мови можливо лише у контексті порівняння його з іншими реченнями. Саме у цьому випадку велику роль відіграє відмінкова граматики [92, 93].

Розглянемо декілька речень «*Викладач відчинив двері*», «*Ключ відчинив двері*», «*Викладач відчинив двері ключем*», «*Двері відчинили*», «*Двері було відчинено ключем*». Усі перераховані вище речення несуть знання про факт відкриття дверей, проте суб'єкти цих речень різні, що пояснюється вживанням різних відмінкових рамок. Флективна мова, крім використання великої кількості синонімів, передбачає різноманітність структурних засобів побудови речень, так, вживання активної та пасивної форми дієслів, займенників, прямий та інверсний порядок слів у реченні. Відповідно, для порівняння за змістом необхідно простежити всі можливі варіації побудови речень, створити систему продукцій, побудовану на основі правил природної мови та формалізувати процес відшукання збігів і протиріч.

Для цього необхідно побудувати логіко-лінгвістичні моделі кожного з речень.

1. $L(S) = L^S = p_1(x_1, 0, y_1, 0, 0, 0, 0),$

$L(S) = L^S = \text{відчинив (викладач, 0, двері, 0,0,0,0)}$.

2. $L(S) = L^S = p_1(x'_1, 0, y_1, 0, 0, 0, 0),$

$L(S) = L^S = \text{відчинив (ключ, 0, двері, 0,0,0,0)}$.

3. $L(S) = L^S = p_1(x_1, 0, y_1, 0, x'_1, 0, 0),$

$L(S) = L^S = \text{відчинив (викладач, 0, двері, 0, ключем,0,0)}$.

4. $L(S) = L^S = \bar{p}_1(0, 0, y_1, 0, 0, 0, 0),$

$L(S) = L^S = \text{відчинили (0, 0, двері, 0,0,0,0)}$.

5. $L(S) = L^S = \bar{p}'_1(0, 0, y_1, 0, x'_1, 0, 0),$

$L(S) = L^S = \text{було \& відчинено (0, 0, двері, 0, ключем, 0, 0)}$.

Аналіз утворених логіко-лінгвістичних моделей речень природної мови дає змогу зробити наступні висновки.

1. Предикати p_1 , \widehat{p}_1 та \widehat{p}'_1 є спільнокореневими словами, p_1 і \widehat{p}_1 – дієслова в активній формі, проте у різному числі, а \widehat{p}'_1 є їх пасивною формою, про що свідчить перша складова предиката «було». Це дає можливість говорити про те, що зміст p_1 , \widehat{p}_1 та \widehat{p}'_1 однаковий, тому у всіх логіко-лінгвістичних моделях можна записувати одну змінну p_1 .

2. Четверта логіко-лінгвістична модель $L(S) = L^S = \widehat{p}_1(0, 0, y_1, 0, 0, 0, 0)$, є істинною по відношенню до $L(S) = L^S = p_1(x_1, 0, y_1, 0, 0, 0, 0)$, $L(S) = L^S = p_1(x'_1, 0, y_1, 0, 0, 0, 0)$, так як немає суб'єкту, але об'єкт для всіх трьох логіко-лінгвістичних моделей однаковий.

3. Аналогічно логіко-лінгвістична модель $L(S) = L^S = \widehat{p}'_1(0, 0, y_1, 0, x'_1, 0, 0)$, є істинною по відношенню до $L(S) = L^S = p_1(x_1, 0, y_1, 0, 0, 0, 0)$, $L(S) = L^S = p_1(x'_1, 0, y_1, 0, 0, 0, 0)$, $L(S) = L^S = p_1(x_1, 0, y_1, 0, x'_1, 0, 0)$, $L(S) = L^S = \widehat{p}_1(0, 0, y_1, 0, 0, 0, 0)$, так як також не містить суб'єкта, а об'єкти у цих реченнях співпадають.

4. Не можна говорити про ідентичність за змістом висловлювань:

$L(S) = L^S = \text{відчинив (викладач, 0, двері, 0, 0, 0, 0)}$,

$L(S) = L^S = \text{відчинив (ключ, 0, двері, 0, 0, 0, 0)}$ та

$L(S) = L^S = \text{відчинили (0, 0, двері, 0, 0, 0, 0)}$,

$L(S) = L^S = \text{було \& відчинено (0, 0, двері, 0, ключем, 0, 0)}$.

Суб'єкт у наведених вище парах висловлювань різний, з чого можна припустити, що двері було відчинено не викладачем, а, наприклад, учнем.

5. Ідентифікація різних форм предиката, а також слова «ключ» у різних реченнях, дає змогу зробити заміни для спрощення процесу порівняння.

Після заміни змінних у логіко-лінгвістичних моделях будемо мати наступний набір формул:

$$1. L(S) = L^S = p_1(x_1, 0, y_1, 0, 0, 0, 0),$$

$L(S) = L^S = \text{відчинив (викладач, 0, двері, 0, 0, 0, 0)}$.

$$2. L(S) = L_1(S) = p_1(x'_1, 0, y_1, 0, 0, 0, 0),$$

$L(S) = L^S = \text{відчинив (ключ, 0, двері, 0, 0, 0, 0)}$.

$$3. L(S) = L_1(S) = p_1(x_1, 0, y_1, 0, x'_1, 0, 0),$$

$L(S) = L^S = \text{відчинив (викладач, 0, двері, 0, ключем, 0, 0)}$.

$$4. L(S) = L_1(S) = p_1(0, 0, y_1, 0, 0, 0, 0),$$

$L(S) = L^S = \text{відчинили (0, 0, двері, 0, 0, 0, 0)}$.

$$5. L(S) = L_1(S) = p_1(0, 0, y_1, 0, x'_1, 0, 0),$$

$L(S) = L^S = \text{було \& відчинено (0, 0, двері, 0, ключем, 0, 0)}$.

Таким чином, дослідження структури речень природної мови, а також врахування основних елементів когезії дає змогу сформулювати базові етапи алгоритму порівняння логіко-лінгвістичних моделей текстової інформації [94, 95].

1. Побудова логіко-лінгвістичних моделей. На цьому етапі кожному реченню природної мови ставиться у відповідність логічна формула, що представляє собою одновимірний масив слів, з яких складаються речення, упорядкованих у відповідності до того, яку синтаксичну роль вони виконують. Тобто вся текстова інформація буде формалізована і записана у вигляді моделей, створених за єдиним шаблоном (2.1) – (2.4).

2. Ідентифікація. Відбувається почерговий перегляд усіх елементів логіко-лінгвістичних моделей: предикатів, предикатних змінних (суб'єктів), предикатних змінних (аргументів), предикатних констант. Серед складових логіко-лінгвістичних моделей знаходяться спільнокореневі слова, синоніми, активні та пасивні форми спільнокореневих дієслів. Почерговий перегляд відбувається не лише між елементами, що виконують одну і ту саму синтаксичну роль, але й між усіма структурними складовими. Це дає змогу повністю ідентифікувати всі слова речень, навіть якщо вони мають різні графеми, проте відображають один і той самий зміст.

3. Заміна тотожних змінних. Якщо на етапі ідентифікації знайдено тотожні змінні, у всіх логіко-лінгвістичних моделях відбувається їх перепозначення, завдяки чому одні й ті самі слова (навіть якщо вони мають різні граматичні рамки) будуть позначатися однаково, і, відповідно, мати ідентичний зміст.

4. Логічний вивід. На останньому етапі алгоритму порівняння логіко-лінгвістичних моделей після ідентифікації та заміни тотожних змінних застосовується система продукцій, що містить правила порівняння логіко-лінгвістичних моделей.

Найбільшу цікавість представляє собою останній етап. Саме тут використовується база знань у вигляді продукційної моделі, сформована на основі правил формування речень флективних мов. Такі правила дозволяють через встановлені кореляції між словами переходити до представлення значень слів у вигляді комбінацій елементарних компонентів змісту. Тобто побудова таких правил передбачає компонентний аналіз лексичних значень.

Відношення, у яких знаходяться слова у флективних мовах, мають різний ступінь системності. Якщо розуміти під системністю організацію деякої множини об'єктів як можливість розміщення цих об'єктів за допомогою правил, кількість яких менша, ніж кількість об'єктів, то можна вважати, що системними є ті кореляції, які забезпечують системність організації множини слів [43]. Із цього випливає, що системні кореляції не повинні бути унікальними, повинні характеризувати не одну пару взаємопов'язаних слів, а цілий ряд таких пар. Саме на основі таких системних відношень між словами будуються пропорційні рівності, тобто семантичні пропорції.

Компонентний аналіз передбачає співставлення значень слів у декілька етапів [79]:

- визначення тієї одиниці змісту, яка включає значення слова (початкова форма та словосполучення, що дає пояснення компоненти змісту);
- знаходження синонімів та гіперонімів;
- дослідження одиниць одного ієрархічного рівня і знаходження їх перетину із шуканим словом;
- створення мінімального списку тих компонентів, що відрізняють шуканий елемент від усіх інших;
- формулювання дефініції слова на основі діагностичних компонент.

Порівняння логіко-лінгвістичних моделей речень природної мови дає можливість аналізувати текстові документи за змістом, знаходити у них збіги та протиріччя. Алгоритм порівняльного аналізу, запропонований автором, включає в себе базові етапи, кожен з яких представляє собою об'ємну процедуру використання логічних, лексичних, синтаксичних та семантичних правил формування текстів [96].

Нехай є два речення «*Стаття присвячена дослідженню існуючих методів виявлення логічних протиріч у текстових документах на основі використання формул логіки предикатів, зокрема дослідженню методу резолюцій*» та «*Статтю присвячено дослідженню існуючих методів виявлення логічних протиріч у текстових документах на основі використання формул логіки предикатів, зокрема дослідженню методу резолюцій*». Ці речення мають однаковий зміст, проте одне з них у активному стані (суб'єкт виконує дію), друге – у пасивному (суб'єкт відсутній).

Логіко-лінгвістична модель першого із заданих речень буде мати вигляд:

$$L(S) = L^S = p_{11}(x_{11}, 0, y_{11}, 0, z_{11}, r_{11}, 0) \&$$

$$p_{12}(z_{11}, r_{11}, y_{12}, q_{12}, z_{12}, r_{12}, h_{12}) \& p_{13}(h_{12}, 0, y_{13}, 0, z_{13}, r_{13}, 0) \&$$

$$p_{11}(x_{11}, 0, y_{11}, 0, z_{14}, r_{14}, 0),$$

$L(S) = L^S =$ *присвячена (стаття, 0, дослідженню, 0, методів, існуючих, 0) & виявлення (методів, існуючих, протиріч, логічних, документах, текстових, основі) & використання (основі, 0, формул, 0, логіки, предикатів, 0) & присвячена (стаття, 0, дослідженню, 0, методу, резолюцій, 0).*

Логіко-лінгвістична модель другого заданого речення має вигляд:

$$L(S) = L^S = \hat{p}'_{11}(0, 0, \hat{x}_{11}, 0, y_{11}, q'_{11}, 0) \&$$

$$p_{12}(z_{11}, r_{11}, y_{12}, q_{12}, z_{12}, r_{12}, h_{12}) \& p_{13}(h_{12}, 0, y_{13}, 0, z_{13}, r_{13}, 0) \&$$

$$\hat{p}'_{11}(0, 0, \hat{x}_{11}, 0, y_{11}, q''_{11}, 0),$$

$L(S) = L^S =$ *присвячено (0, 0, статтю, 0, дослідженню, існуючих методів, 0) & виявлення (методів, існуючих, протиріч, логічних, документах, текстових, основі) & використання (основі, 0,*

формул, 0, логіки, предикатів, 0) & присвячено (0, 0, статтю, 0, дослідженню, методу резолюцій, 0).

Порівняння речень різного стану відбувається за такими принципами.

1. Відбувається порівняння предикатів речень p_1 та \hat{p}'_1 . Якщо p_1 та \hat{p}'_1 – дієслова, їх час співпадає, \hat{p}'_1 – дієслово в пасивній формі, а p_1 – дієслово в активній формі, а також співпадають корені та суфікси предикатів, або p_1 та \hat{p}'_1 є синонімами, то предикати p_1 і \hat{p}'_1 можна вважати тотожними за змістом;

2. Якщо в одному із речень, що порівнюються, відсутній суб'єкт, а x_1 та y'_1 іменники у називному та знахідному відмінку відповідно, x_1 та y'_1 – спільнокореневі слова або є синонімами, то суб'єкти речень (віртуальний у випадку пасивного стану) можна вважати тотожними за змістом;

3. Якщо справджуються умови 1 та 2, то речення можна вважати однаковими за змістом.

На даному прикладі продемонстровано можливість здійснення порівняльного аналізу окремих речень природної мови різної структури за змістом за допомогою використання логіко-лінгвістичних моделей цих речень [52]. Як видно з прикладу, це дуже важка і громіздка процедура, в якій необхідно врахувати абсолютно всі (як синтаксичні, так і семантичні) особливості речень та зв'язки між словами і словосполученнями.

V. АВТОМАТИЗОВАНИЙ ЛІНГВІСТИЧНИЙ АНАЛІЗ ЕЛЕКТРОННИХ ТЕКСТІВ

5.1 Характеристики семантичного рівня

Розуміння – це багаторівневий процес. Відповідно, є спроби створення різноманітних систем, що відповідають кожному або декільком рівням розуміння. На першому, найпростішому рівні, всі дані про зміст тексту, що розглядається, вилучаються у результаті його аналізу без залучення додаткових знань, відомих системі. На другому рівні відбувається визначення часової, просторової та причинно-наслідкової структур подій за допомогою процедур логічного розуміння інформації. На третьому рівні до сформованого представлення змісту додається інформація, релевантна цьому змісту та відома системі. На четвертому рівні до нього приєднуються відомості, вилучені з бази знань та пов'язані з текстом, що аналізується, лише відношенням асоціації. На п'ятому рівні розуміння з тексту вилучається його прагматичний зміст. При цьому система виконує всі обумовлені цим текстом дії, наприклад, вирішує задачу, для якої є готова програма, а у початковому тексті вказані початкові дані для неї. Найбільший практичний інтерес представляють системи, які реалізують п'ятий рівень [37].

Застосування таких систем вимагає наявності компоненти, що здійснюватиме семантичний аналіз та працюватиме зі змістом заданого тексту. Метою семантичного аналізу є визначення змістовних характеристик для кожного слова та фрази в цілому. Складності виникають через семантичну неоднозначність. Часто для її зняття використовують семантичні статті, пов'язані одна з одною в рамках семантичної мережі [47]. Аналіз відношень у ній дозволяє отримати інформацію, що явно відсутня у фразі, проте без якої адекватне розуміння фрази неможливе. Складності такої реалізації пов'язані з великим розміром семантичних мереж та багатоваріантністю аналізу.

Сема – змістовний аналог, семантична функція словоформи як основної одиниці лексико-граматичного рівня [51].

Сема – одиниця змісту, що співвідноситься з морфемою як сукупність компонентів її змісту. Є мінімальною одиницею системи змісту. Що співвідноситься з елементом системи виразу.

Структурними компонентами семеми є семи та семантичні множники [53].

Семи – це елементарні, не роздільні для мовного змісту компоненти семем, це диференціальна семантична ознака, компонента значення, яка виявляється при співставленні значень різних слів, наприклад «бути», «стан», «дія».

Семантичні множники – це безпосередні складові семеми, можуть бути різними за своїм складом. Наприклад, складовими семеми «експерт» є такі семантичні множники, як «кваліфікований», «обізнаний», «розумний», «досвідчений», «володіє знаннями з предметної області» і т.д.

Лексичне значення – це загальне, інваріантне значення слів даної лексеми, опосередковане від її граматичних значень. Таким чином, одна й та сама лексема може виражати декілька лексичних значень. І навпаки, те саме лексичне значення може виражатися декількома лексемами.

Словотвірне значення – це значення словотвірного афікса, що є частиною лексичного значення похідного слова. Наприклад, префікс «без» має словотвірне значення відсутності будь-чого, суфікс «еньк» – означає зменшену форму. Тобто словотвірне значення відрізняється від лексичного тим, що його вираження пов'язане з афіксальними морфемами та відповідає дериваційним моделям.

Узгодження семем, що відбувається за семантичними правилами (а вони нерозривно пов'язані із синтаксичними правилами формування словосполучень флективних мов), представляє собою прості конкретні семантичні синтагми.

Синтагма – сукупність декількох слів, об'єднаних за принципом семантико-граматично-фонетичної узгодженості. Об'єм конкретної синтагми визначається не лише реальним вживанням слів у зв'язці, але й можливістю об'єднання предметів, ознак, процесів. Мінімальна одиниця синтагми – прості словосполучення, однак їх розмір інколи може сягати цілої фрази.

Семантична категорія – це клас виразів з однотипними предметними значеннями, що включає в себе всі вирази з предметним значенням даного типу.

На семантичному рівні між словами розрізняють два типи зв'язків: синтагматичні та парадигматичні.

Синтагматичні відношення об'єднують одиниці мови у їх одночасній послідовності. «Слова, об'єднуючись одне з одним, вступають між собою у відношення, що базуються на лінійному характері мови, а це виключає можливість вживання двох елементів одночасно. Ці елементи ідуть один за одним у потоці мови» [52]. На синтагматичних відношеннях будуються слова, словосполучення, прості та складні речення.

Парадигматичні відношення – це відношення, які об'єднують одиниці мови у групи, категорії, розряди. На парадигматичних відношеннях побудована система відмінків, синонімічний ряд. Парадигматичні відношення дозволяють обрати потрібну одиницю, а також утворювати форми та слова по аналогії.

Парадигма – відношення між одиницями, які можуть замінити одна одну в одній позиції.

У сучасній лінгвістиці синтагматичні відношення прирівнюються до чисто логічних відношень кон'юнкції, а парадигматичні відношення – до логічних відношень диз'юнкції або відношення по горизонталі та вертикалі відповідно. Різницю між синтагматичними та парадигматичними типами відношень можна простежити на прикладі слова «книга». Парадигматичні зв'язки для цього слова будують відмінковий («книжок», «книгою», «книгарня» і т.д.) та синонімічний ряди («журнал», «том», «підручник» і т.д.). Синтагматичні ж зв'язки викликають ряд асоціацій «товста книжка», «книжка в обкладинці», «дитяча», «художня» і т.д.

Синтагматичні зв'язки мають таку градацію:

- **сурядні** (координація);
- **підрядні** (субординація) – узгодження («зелена трава»), керування («бути здоровим»), прилягання («книжка викладача»);
- **конститутивно-інтегративні** – характерні для лінійних одиниць, текст складається з речень (висловлювань), речення – з лексико-граматичних синтагм більшої чи меншої складності, словоформи – з морфем; на семантичному рівні необхідно враховувати комплектистність, конфігурацію одиниць, лінійність, що розглядається як проекція на певну послідовність лексико-граматичних одиниць, а також те, що зміст одиниць вищого семантичного рівня не зводиться до змісту його складових;

- *контрастуючі* – можуть бути контактними (відношення суміжності у словосполученні) та перервними (відношення між словосполученнями, реченнями).

Парадигматичні відношення класифікують наступним чином:

- відношення повної невідповідності;
- відношення повної тотожності;
- відношення включення;
- відношення перетину.

Відношення включення та перетину ще називають опозиціями – суттєва лінгвістична різниця між одиницями виразу та одиницями змісту.

Привативна опозиція – виражає відношення включення: об'єм ознак (структурних компонентів) одного з членів опозиції повністю включається до складу іншого маркованого члена. Один член опозиції відрізняється від іншого наявністю або відсутністю корелятивної ознаки або марки кореляції. Наприклад, «книга»– «книгарня».

Градуальна або ступінчата опозиція – члени опозиції відрізняються один від одного різним ступенем однієї і тієї ж ознаки. Наприклад, «книга» – «книжечка».

Еквіполентна опозиція – виражає відношення логічного перетину, тобто рівноправність між членами опозиції, що мають спільні ознаки (основу опозиції), проте кожен з них характеризується (маркується) ще й специфічною ознакою. Наприклад, «художня книга» – «дитяча книга».

Дослідженням проблеми співвідношення мови та свідомості, ролі мови у концептуалізації та категоризації світу займається когнітивна лінгвістика. Об'єктом її вивчення є мова як механізм пізнання. Внаслідок чого для розуміння текстової інформації лінгвісти виділяють обов'язкові поняття, коректне визначення яких дає змогу зрозуміти зміст текстових документів.

Денотат – термін лінгвістики та екстенціональної логіки, що означає множину явищ дійсності, які цим іменем можуть називатися.

Референт – об'єкт, який мається на увазі у контексті конкретної мовної ситуації, предмет референції. Наприклад, у одному тексті один і той самий референт «експерт» може приймати значення «він», «фахівець», «досвідчений» і т.д.

Інтенціонал – термін семантики, що означає зміст поняття, тобто сукупність ознак, які позначаються поняттям предмета чи явища. До інтенціоналу поняття «книжка» входять всі властивості, які характерні для книжки: «папір», «зміст», «обкладинка», «сторінки» і т.д.

Екстенціонал – термін семантики, що означає об'єм поняття, тобто множину об'єктів, здатних називатися даною мовною одиницею. Наприклад, до екстенціоналу слова «книжка» входять всі об'єкти, що володіють властивістю книг, тобто «збірник», «підручник», «Енеїда» і т.д.

Десігнат – значення слова, протиставляється денотату як предметній області та визначається як суб'єктивний образ або концепт. Наприклад, «круглий квадрат».

Сигніфікат – понятійний зміст імені або знаку, ідентифікує певний об'єкт, розкриваючи різноманітні його ознаки. Наприклад, сигніфікатами слова «говорити» можуть бути «балакати», «базікати», «бовкнути», «розмовляти», «теревенити» і т.д.

Семантичне ядро значення – це об'єктивний компонент змісту, абстрагований від стилістичних, прагматичних, модальних, емоційних, суб'єктивних відтінків. Наприклад, вирази «бути людиною», «виступати експертом» є семантичними ядрами, а «бути прогресивним», «виступати розумним» – ні.

Узагальнена інформація щодо перерахованих основних елементів семантичного рівня відображена на рис. 5.1. Цей матеріал є теоретичною основою для розробки методів та алгоритмів змістовного аналізу електронних документів.

Проте контекст ще не створює текст, він необхідний для підключення мовних механізмів, актуалізації значень, запуску процесу референції. Паралельно з актуалізацією понять відбувається процес прагматичного усвідомлення, у результаті якого людина отримує непросту суму пропозицій. Такий процес ще називається **текстуальністю**. Це одно направлене прагматичне поле, що формується при читанні та дозволяє сприймати послідовність мовних одиниць як текст [97]. Таким чином, лінгвістичний аналіз тексту – це аналіз текстуальності.

За Р.-А. де Бограндом та В. Дресслером текстуальністю є сукупність властивостей, ознак, які притаманні тексту: когезія, когерентність, інтенціональність, сприйняття, інформативність, ситуативність та інтертекстуальність [98].

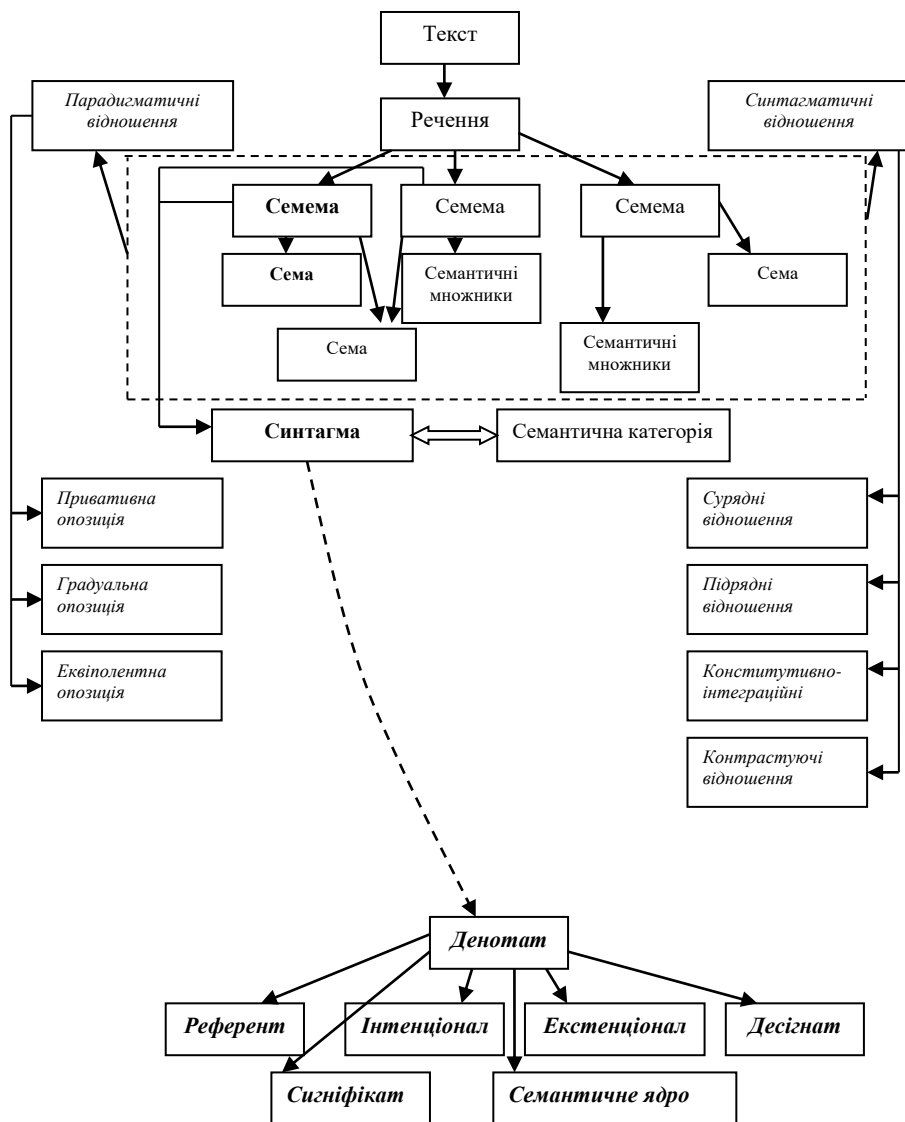


Рис. 5.1. Основні елементи семантичного рівня

5.2 Когезія як основний засіб формування семантичних зв'язків у тексті

Детальне дослідження текстових документів показало, що усі тексти, крім авторських, мають певну структуру, характерну для всіх стилів: вступ, зав'язка, постановка проблеми, розв'язка, висновки і т.д. По аналогії з синтаксичною та лексичною структурами, можна виділити основні змістовні (семантичні) елементи, що притаманні кожному тексту.

Когезія – це особливий вид зв'язку, що забезпечує континуум, тобто логічну послідовність, взаємозв'язок окремих повідомлень, фактів, дій і т.д. [99]. Це сукупність лексичних та граматичних засобів для вираження зв'язків між одиницями тексту [100].

Когезія або формально-граматична зв'язність тексту визначається різними мовними відношеннями між реченнями, що формують текст [101].

1. Семантичне повторення – дає змогу простежити ряд речень з однаковим змістом (синтаксична структура таких речень буде однаковою, проте суб'єкт або об'єкт може бути замінений синонімами).

Повне тотожне повторення – це найбільш простий механізм зв'язку, що полягає у повторенні однакових слів або словоформ, які мають один корінь.

Розглянемо фрагмент тексту, в якому використовується повне тотожне повторення та простежимо, яким чином такий засіб когезії можна виявити за допомогою логіко-лінгвістичних моделей.

Нехай є текст *«В основу практично всіх епідеміологічних моделей, розроблених до сьогодні, покладені класичні моделі поширення захворювань. Однак, ці моделі є повністю детерміновані і можуть бути застосовані тільки для оцінки поведінки великих популяцій. В традиційних моделях також передбачається, що всі індивіди безперервно і рівномірно перемішуються. У цих умовах класичні моделі будуть давати більш-менш реалістичні результати тільки при моделюванні на великих територіях»* [102].

Повне тотожне повторення для даного уривку полягає у використанні слова *«моделі»* у різних граматичних формах:

«моделей», «моделі», «моделях», а також спільнокореневого слова «моделюванні».

Побудуємо для цього фрагменту тексту логіко-лінгвістичні моделі речень.

$$L(S_1) = p'_1(0, 0, y'_1, q'_1, z'_1, r'_1, h'_1) \&$$

$$(p''_1(h'_1, 0, y''_1, q''_{11}, 0, 0, h''_1) \& p''_1(h'_1, 0, y''_1, q''_{12}, 0, 0, h''_1)),$$

$$L(S_2) = p'_2(x'_2, g'_2, y'_2, q'_2, 0, 0, 0) \& p''_2(x'_2, g'_2, y''_2, q''_2, z''_2, r''_2, h''_2),$$

$$L(S_3) = p'_3(0, 0, y'_3, q'_3, 0, 0, h'_3) \rightarrow (p''_3(x''_3, g''_3, 0, 0, 0, 0, h''_{31}) \&$$

$$p''_3(x''_3, g''_3, 0, 0, 0, 0, h''_3)),$$

$$L(S_4) = p_4(x_4, g_4, y_4, q_4, z_4, r_4, h_4),$$

$L(S_1)$ = *покладені (0,0, моделі, класичні, поширення, захворювань, основу) & (розроблених (основу, 0, моделей, всіх, 0,0, практично) & розроблених (основу, 0, моделей, епідеміологічних, 0,0, практично)),*

$L(S_2)$ = *ε (моделі, класичні, детерміновані, повністю, 0,0,0) & можуть & бути & застосовані (моделі, класичні, оцінки, поведінки, великих, популяцій, тільки),*

$L(S_3)$ = *передбачається (0,0,моделях,традиційних, 0,0,також) → (перемішуються (індивіди, всі,0,0,0,0,безперервно) & перемішуються (індивіди, всі,0,0,0,0,рівномірно)),*

$L(S_4)$ = *будуть & давати (моделі, класичні, результати, реалістичні, моделюванні, великих & територіях, цих & умовах).*

Для довільного тексту сукупність логіко-лінгвістичних моделей речень природної мови представляє собою двовимірний масив, елементами якого є структурні елементи логіко-лінгвістичних моделей, кількість рядочків у такому масиві дорівнює кількості речень у тексті.

Повне тотожне повторення дозволяє перебудувати логіко-лінгвістичні моделі наступним чином:

$$L(S_1) = p'_1(0, 0, y'_1, q'_1, z'_1, r'_1, h'_1) \&$$

$$(p''_1(h'_1, 0, y''_1, q''_{11}, 0, 0, h''_1) \& p''_1(h'_1, 0, y''_1, q''_{12}, 0, 0, h''_1)),$$

$$L(S_2) = p'_2(y'_1, g'_2, y'_2, q'_2, 0, 0, 0) \& p''_2(y'_1, g'_2, y''_2, q''_2, z''_2, r''_2, h''_2),$$

$L(S_3) = p'_3(0, 0, y'_1, q'_3, 0, 0, h'_3) \rightarrow (p''_3(x''_3, g''_3, 0, 0, 0, 0, h''_{31}) \& p''_3(x''_3, g''_3, 0, 0, 0, 0, h''_3)),$

$L(S_4) = p_4(y'_1, g_4, y_4, q_4, y'_1, r_4, h_4),$

Тобто слово «*моделі*», що в різних граматичних формах виступає у тексті як суб'єктом, так і об'єктом, дає змогу здійснити заміну: $y'_1 \equiv y''_1, x'_2 \equiv y'_1, y'_3 \equiv y'_1, x_4 \equiv y'_1, z_4 \equiv y'_1$. Таким чином, слово «*моделі*» є ключовим для уривку тексту, що розглядається.

Тематичне повторення – слова у складі тексту виявляють спільність лексичних значень, позначають різні сторони, компоненти, складові частини елементів однієї ситуації. Слова однієї тематичної групи утворюють всередині тексту парадигму, яка пов'язує різні частини тексту.

Нехай є текст «*Наукові експерименти в області епідеміології здійснити важко або практично не можливо. Місто суттєво впливає на поширення та еволюцію інфекційних захворювань. В основу практично всіх епідеміологічних моделей, розроблених до сьогодні, покладені класичні моделі поширення захворювань*» [103].

Логіко-лінгвістична модель даного уривку буде мати вигляд:

$L(S_1) = p'_1(0, 0, y'_1, q'_1, z'_1, r'_1, h'_{11}) \vee p'_1(0, 0, y'_1, q'_1, z'_1, r'_1, -h'_{12}),$

$L(S_2) = p'_2(x'_2, 0, y'_{21}, 0, z'_2, r'_2, 0) \& p'_2(x'_2, 0, y'_{22}, 0, z'_2, r'_2, 0),$

$L(S_3) = p'_3(0, 0, y'_3, q'_3, z'_3, r'_3, h'_3) \&$

$(p''_3(h'_3, 0, y'_3, q''_{31}, 0, 0, h''_3) \& p''_3(h'_3, 0, y'_3, q''_{32}, 0, 0, h''_3)),$

$L(S_1) = \text{здійснити}(0, 0, \text{експерименти, наукові, області, епідеміології, важко}) \vee \text{здійснити}(0, 0, \text{експерименти, наукові, області, епідеміології, практично} \& \neg \text{можливо}),$

$L(S_2) = \text{впливає (місто, 0, поширення, 0, захворювань, інфекційних, суттєво)} \& \text{впливає (місто, 0, еволюцію, 0, захворювань, інфекційних, суттєво)},$

$L(S_3) = \text{покладені (0, 0, моделі, класичні, поширення, захворювань, основу)} \& \text{(розроблених (основу, 0, моделей, всіх, 0, 0, практично)} \& \text{розроблених (основу, 0, моделей, епідеміологічних, 0, 0, практично)}).$

Такі слова та словосполучення як «епідеміологія», «інфекційні захворювання», «епідеміологічні моделі», «поширення захворювань» задають тематику тексту. Стає зрозуміло, що уривок взято із медичної статті. Тематичний повтор не дає змоги зробити заміну в логіко-лінгвістичних моделях, проте всі тематичні слова є предикатними змінними (аргументами). Це дає можливість поставити таким тематичним аргументам додаткові характеристики при здійсненні семантичного аналізу.

Експресивне повторення – вживання слів, які формують загальний план оцінки.

На прикладі попереднього уривку можна побачити, що такі слова як «важко», «не можливо», «практично» говорять про невпевненість автора та про песимістичний характер статті.

Отже, автор уривку використав одразу два засоби когезії: тематичне та експресивне повторення.

Синонімічне повторення – вживання синонімів для забезпечення міжфразового зв'язку, об'єднання змісту різних речень.

Нехай є уривок тексту «Місто суттєво впливає на поширення та еволюцію інфекційних захворювань. Міське середовище згладило процеси природного відбору, пом'яжило вплив зовнішніх чинників, змінило імунну систему людини» [104].

Слово «місто» та «міське середовище» є синонімами, тому логіко-лінгвістичні моделі речень заданого уривку будуть мати вигляд:

$$L(S_1) = p'_1(x'_1, 0, y'_{11}, 0, z'_1, r'_1, 0) \& p'_1(x'_1, 0, y'_{12}, 0, z'_1, r'_1, 0),$$

$$L(S_2) = p'_{21}(x'_1, 0, y'_{21}, 0, z'_{21}, r'_{21}, 0) \& p'_{22}(x'_1, 0, y'_{22}, 0, z'_{22}, r'_{22}, 0) \&$$

$$p'_{23}(x'_1, 0, y'_{23}, q'_{23}, z'_{22}, 0, 0),$$

$L(S_1) =$ впливає (місто, 0, поширення, 0, захворювань, інфекційних, суттєво) & впливає (місто, 0, еволюцію, 0, захворювань, інфекційних, суттєво),

$L(S_2) =$ згладило (місто, 0, процеси, 0, відбору, природного, 0) & пом'яжило (місто, 0, вплив, 0, чинників, зовнішніх, 0) & змінило (місто, 0, систему, імунну, людини, 0, 0).

Повторення різних стилістичних варіантів одного і того ж слова.

Розглянемо дещо змінений попередній уривок «Урбоекосистема суттєво впливає на поширення та еволюцію інфекційних захворювань. Міське середовище згладило процеси природного відбору, пом'якшило вплив зовнішніх чинників, змінило імунну систему людини».

Тут слово «урбоекосистема» та словосполучення «міське середовище» є різними стилістичними варіантами слова «місто». Тому логіко-лінгвістичні моделі даної фрази будуть тотожні попередньому запису при синонімічному повторенні:

$$L(S_1) = p'_1(x'_1, 0, y'_{11}, 0, z'_1, r'_1, 0) \& p'_1(x'_1, 0, y'_{12}, 0, z'_1, r'_1, 0),$$

$$L(S_2) = p'_{21}(x'_1, 0, y'_{21}, 0, z'_{21}, r'_{21}, 0) \& p'_{22}(x'_1, 0, y'_{22}, 0, z'_{22}, r'_{22}, 0) \&$$

$$p'_{23}(x'_1, 0, y'_{23}, q'_{23}, z'_{22}, 0, 0),$$

$L(S_1) =$ впливає (місто, 0, поширення, 0, захворювань, інфекційних, суттєво) & впливає (місто, 0, еволюцію, 0, захворювань, інфекційних, суттєво),

$L(S_2) =$ згладило (місто, 0, процеси, 0, відбору, природного, 0) & пом'якшило (місто, 0, вплив, 0, чинників, зовнішніх, 0) & змінило (місто, 0, систему, імунну, людини, 0, 0).

Зв'язність через відношення частина-ціле шляхом родово-видової заміни, тобто слова, не будучи синонімами, стають позначеннями одного і того ж предмета, тобто контекстуальними синонімами.

Нехай є текст «Агент – це деяка сутність, яка володіє активністю, автономною поведінкою, може приймати рішення відповідно до деякого набору правил. Агентні моделі використовуються для дослідження децентралізованих систем».

«Агент» та «агентні моделі» заходяться у відношенні частина-ціле, через те, що агентні моделі є конкретною реалізацією агента як такого. Тому при побудові логіко-лінгвістичних моделей даного фрагменту буде здійснена заміна:

$$L(S_1) = p'_1(x'_1, 0, 0, 0, 0, 0, h'_1) \& (p''_{11}(p'_1, h'_1, y''_{11}, 0, 0, 0, 0) \&$$

$$p''_{11}(p'_1, h'_1, y''_{12}, q''_{12}, 0, 0, 0) \& p''_{12}(p'_1, h'_1, y''_{13}, q''_{13}, z''_{13}, r''_{13}, h''_{13})),$$

$$L(S_2) = p_2(x'_1, 0, y_2, 0, z_2, r_2, 0),$$

$L(S_1) = \text{сутність (агент, 0, 0, 0, 0, 0, деяка)} \& \text{(володіє (сутність, деяка, активністю, 0, 0, 0, 0))} \& \text{володіє (сутність, деяка, поведінкою, автономною, 0, 0, 0, 0))} \& \text{може} \& \text{приймати (сутність, деяка, рішення, правил, набору, деякого, відповідно),}$

$L(S_2) = \text{використовуються (агенти, 0, дослідження, 0, систем, децентралізованих, 0)}.$

Зв'язність через колокацію або антонімічне повторення – дає змогу підкреслити протиріччя, конфліктність подій, станів, створити антитезу образів. Такі відношення можуть бути виявлені шляхом статистики частого вживання слів.

Наприклад, «Для знешкодження пожежі застосовуються сили з довколишніх міст, внаслідок чого зменшується рівень пожежного захисту. Використання автономних літальних апаратів здатне збільшити ефективність виявлення займань на ранніх стадіях».

У фразі вжито антоніми «зменшується» та «збільшується», що несе змістовне навантаження протиставлення.

Логіко-лінгвістичні моделі заданого уривку будуть мати вигляд:

$L(S_1) = p'_1(x'_1, g'_1, y'_1, 0, z'_1, 0, 0) \rightarrow p''_1(x''_1, 0, y''_1, z'_1, 0, 0, 0),$

$L(S_2) = \bar{p}''_1(x'_2, 0, y'_{21}, 0, z'_{21}, r'_{21}, h'_2) \& (x'_2(0, 0, y'_{22}, q'_{22}, 0, 0, 0) \& x'_2(0, 0, y'_{22}, q'_{23}, 0, 0, 0)),$

$L(S_1) = \text{застосовуються (сили, довколишніх} \& \text{міст, знешкодження, 0 пожежі, 0, 0)} \rightarrow \text{зменшується (рівень, 0, захисту, пожежного, 0, 0, 0)}.$

$L(S_2) = \text{здатне} \& \text{збільшити (використання, 0, ефективність, 0, виявлення, займань, ранніх} \& \text{стадіях)} \& \text{(використання (0, 0, апаратів, автономних, 0, 0, 0))} \& \text{використання (0, 0, апаратів, літальних, 0, 0, 0)}.$

Паралельно зі зв'язком через колокацію автор використовує тотожне повторення спільнокоренових слів «пожежі» та «пожежного» у першому реченні фрази.

Антонімічне повторення не дає змоги зробити заміну в логіко-лінгвістичних моделях, проте антоніми виступають протилежними за змістом предикатами. Виявлення такого виду

семантичного повторення дає можливість порівняти логіко-лінгвістичні моделі речень за їх відношеннями предикатами).

2. Засоби контактного та дистантного зв'язку. Зазвичай, змістовне наповнення речення пов'язують з пресупозицією – компонентою змісту речення, яка повинна бути істинна для того, щоб речення не сприймалося як аномальне. Пресупозиції, що визначають зміст висловлювання, можуть міститися у попередньому контексті [105].

Ланцюжки з власними назвами – власна назва набуває свого конкретного предметного значення, лише входячи до складу ланцюжка назв. Тобто це засоби, що підсилюють інформативну якість тексту та належать до формальних показників зв'язку текстових структурних одиниць [106].

Наприклад, *«Теорія взаємодіючих послідовних процесів запропонована відомим теоретиком інформатики Ч. Хоаром. Хоар вважає, що подія відбувається миттєво, тобто є елементарною дією»* [107].

У першому реченні тексту дається пояснення того, хто такий Хоар і яке відношення має до інформатики. Тоді друге речення сприймається з точки зору інформатики.

Робота з логіко-лінгвістичними моделями є аналогічною, як і у випадку повного тотожного повторення:

$$L(S_1) = p_1(x_1, g_1, \hat{x}_1, q_1, z_1, r_1, 0),$$

$$L(S_2) = p'_2(r'_1, 0, 0, 0, 0, 0, 0) \& (p''_{21}(x''_2, 0, 0, 0, 0, 0, h''_2) \& p''_{22}(x''_2, 0, y''_{22}, q''_{22}, 0, 0, 0)),$$

$L(S_1)$ = запропонована (теорія, взаємодіючих & послідовних & процесів, теоретиком, відомим, інформатики, Хоаром, 0),

$L(S_2)$ = вважає (Хоар, 0, 0, 0, 0, 0, 0) & (відбувається (подія, 0, 0, 0, 0, миттєво) & є (подія, 0, дією, елементарною, 0, 0, 0)).

Також у даному фрагменті використано тотожне повторення спільнокореневих слів *«теоретиком»* та *«теорія»*, що враховано у символічному записі логіко-лінгвістичних моделей.

Дейктичне повторення або анафоричні зв'язки вказують на вже названі предмети, ознаки та обставини за допомогою спеціальних слів: займенників, прислівників, семантично спустошених слів, які вживаються замість іншого слова або

словосполучення з попереднього контенту, чисельників у ролі підмета та ін.

Нехай є текст «В основу практично всіх епідеміологічних моделей, розроблених до сьогодні, покладені класичні моделі поширення захворювань. Вони базуються на використанні систем диференціальних рівнянь».

У другому реченні займенник вони вказує на «класичні моделі», про які йшлося у попередньому реченні. Тому дійктичне повторення дає змогу здійснити заміну в логіко-лінгвістичних моделях:

$$L(S_1) = p'_1(0, 0, y'_1, q'_1, z'_1, r'_1, h'_1) \& \\ (p''_1(h'_1, 0, y'_1, q''_{11}, 0, 0, h''_1) \& p''_1(h'_1, 0, y'_1, q''_{12}, 0, 0, h''_1)),$$

$$L(S_2) = p_2(y'_1, g_2, y_2, 0, z_2, r_2, 0),$$

$L(S_1)$ = покладені (0,0, моделі, класичні, поширення, захворювань, основу)& (розроблених (основу, 0, моделей, всіх, 0,0, практично)& розроблених (основу, 0, моделей, епідеміологічних, 0,0, практично)),

$L(S_2)$ = базуються (моделі, класичні, використанні, 0, систем, диференціальних & рівнянь, 0).

Ще один приклад використання дійктичного повторення в межах одного речення «Для знешкодження пожежі застосовуються сили з довколишніх міст, внаслідок чого зменшується рівень їх пожежного захисту».

Тут займенник «їх» вказує на словосполучення «довколишніх міст» із головної частини складнопідрядного речення:

$$L(S_1) = p'_1(x'_1, g'_1, y'_1, 0, z'_1, 0, 0) \rightarrow p''_1(x''_1, \bar{g}'_1, y''_1, q''_1, 0, 0, 0),$$

$L(S_1)$ = застосовуються (сили, довколишніх & міст, знешкодження, 0 пожежі, 0, 0) → зменшується (рівень, міст, захисту, пожежного, 0, 0, 0).

Дієприслівниковий зворот на початку речення – виконує функцію обставини із значенням часу або відтінком причини, що відноситься до всього речення в цілому. Дієприслівникові звороти встановлюють зв'язок однієї події з іншою, вказують на послідовність дій, можуть означати дію, яка відбувається за уже названою, проте передує подіям, що описуються далі по тексту.

Такий засіб когезії дозволяє зображувати події, не описуючи їх окремо, проте вважаючи точкою відліку для іншої дії.

Нехай є фрагмент тексту «Відвідавши таку величезну країну як Канада, людина розуміє всю силу природи».

Логіко-лінгвістична модель такого речення матиме вигляд:

$$L(S_1) = (p'_1(x'_1, 0, y'_1, q'_{11}, z'_1, 0, 0) \& p'_1(x'_1, 0_1, y'_1, q'_{12}, z'_1, 0, 0)) \rightarrow p''_1(x'_1, 0, y''_1, q''_1, z''_1, 0, 0),$$

$L(S_1) =$ (відвідавши (людина, 0, країну, таку, Канада, 0, 0) & відвідавши (людина, 0, країну, величезну, Канада, 0, 0)) → розуміє (людина, 0, силу, всю, природи, 0, 0).

Тобто візит до Канади є точкою відліку для розуміння всієї сили природи, тому у логіко-лінгвістичній моделі використано операцію імплікації.

Дискурсивні слова – сурядні та підрядні сполучники. Змістовні відношення між реченнями у тексті виражаються за допомогою сполучників, що стоять на початку підрядних речень. Вибір сполучника пов'язаний з характером змістовних зв'язків у тексті (одночасність чи послідовність подій, протиставлення, залежність, альтернативність вибору та ін.).

Наприклад, є текст «В основу практично всіх епідеміологічних моделей, розроблених до сьогодні, покладені класичні моделі поширення захворювань. В традиційних моделях також передбачається, що всі індивіди безперервно і рівномірно перемішуються. Проте природа епідемічних процесів носить імовірнісний характер».

У даному фрагменті сполучник «проте» носить характер протиставлення третього речення двом попереднім, а підрядний сполучник «що» дає роз'яснення того, що передбачається в традиційних моделях. Логіко-лінгвістична модель даного фрагменту матиме вигляд:

$$L(S_1) = p'_1(0, 0, y'_1, q'_1, z'_1, r'_1, h'_1) \&$$

$$(p''_1(h'_1, 0, y'_1, q''_{11}, 0, 0, h''_1) \& p''_1(h'_1, 0, y'_1, q''_{12}, 0, 0, h''_1)),$$

$$L(S_2) = p'_2(0, 0, y'_1, q'_2, 0, 0, h'_2) \rightarrow (p''_{21}(x''_2, g''_2, 0, 0, 0, 0, h''_{21}) \&$$

$$p''_{21}(x''_2, g''_2, 0, 0, 0, 0, h''_{22})),$$

$$L(S_3) = p_3(x_3, g_3, y_3, q_3, 0, 0, 0),$$

$L(S_1)$ = покладені (0,0, моделі, класичні, поширення, захворювань, основу) & (розроблених (основу, 0, моделей, всіх, 0,0, практично) & розроблених (основу, 0, моделей, епідеміологічних, 0,0, практично)),

$L(S_2)$ = передбачається (0,0, моделях, традиційних, 0,0, також) → (перемішуються (індивіди, всі, 0,0,0,0, безперервно) & перемішуються (індивіди, всі, 0,0,0,0, рівномірно)),

$L(S_3)$ = носить (природа, епідемічних & процесів, характер, імовірнісний, 0,0,0).

Як видно з прикладу, дискурсивні слова впливають на операції, що використовуються у логіко-лінгвістичних моделях.

Еліпсис або контекстуально неповні синтаксичні конструкції – засіб когезії, що означає пропуск у тексті певної мовної одиниці, яку можна відтворити за змістом попередніх речень. Семантична та структурна неповнота речень пов'язана з тим, що при описі нової ситуації окремі її елементи уже названі у попередньому контексті і не потребують повторного позначення. Неповні конструкції не можуть вживатися самостійно і реалізують свою комунікативну функцію лише разом з іншими реченнями тексту.

Наприклад, у реченні «*На основних усталених руках несуче тіло обертається навколо подовжньої осі, а на побічних – ні*» тире та слово «ні» заміняє «усталених руках не обертається» [108]. Це можна зрозуміти із контексту першої частини складного речення.

Логіко-лінгвістична модель даного речення:

$L(S_1) = (p_1(x_1, g_1, y_1, q_{11}, z_1, r_1, h_1) \& p_1(x_1, g_1, y_1, q_{12}, z_1, r_1, h_1)) \& \neg p_1(x_1, g_1, y_1, q_{13}, z_1, r_1, h_1),$

$L(S_1)$ = (обертається (тіло, несуче, руках, основних, осі, подовжньої, навколо) & обертається (тіло, несуче, руках, усталених, осі, подовжньої, навколо)) & – обертається (тіло, несуче, руках, побічних, осі, подовжньої, навколо).

Вставні слова – виконують функцію утворення тексту, вказуючи порядок викладення думки, наприклад, по-перше, по-друге, нарешті, отже, до речі, іншими словами та ін.

Вставні слова не відображаються в логіко-лінгвістичній моделі, проте мають суттєве значення для зв'язків між реченнями природної мови, а також для формування стилістичного забарвлення тексту.

Нехай є уривок тексту *«Нарешті, результати чисельного моделювання відокремлених внутрішніх хвиль узгоджуються з результатами лабораторних експериментів. Ми вважаємо, що результати, отримані для нашої ідеалізованої геометрії, можуть бути корисними для інтерпретації складних процесів взаємодії»* [109].

Вставні слова *«нарешті»* та *«ми вважаємо»* дають можливість зрозуміти, що уривок взято з певних висновків, а попередньо у тексті йшлося про чисельне моделювання, лабораторні експерименти, які проводилися для ідеалізованої геометрії.

Логіко-лінгвістична модель даного уривку:

$$L(S_1) = p'_1(x'_1, g'_1, y'_1, 0_1, z'_1, r'_1, 0_1) \&$$

$$(x'_1(g'_1, g''_1, y''_1, q''_{11}, 0, 0, 0) \& x'_1(g'_1, g''_1, y''_1, q''_{12}, 0, 0, 0)),$$

$$L(S_2) = p'_2(x'_2, 0, 0, 0, 0, 0, 0) \rightarrow (p''_2(x'_2, g''_2, y''_2, q''_2, z''_2, r''_2, 0) \&$$

$$(g''_2(x'_2, 0, y''_2, q''_{21}, 0, 0, 0) \& g''_2(x'_2, 0, y''_2, q''_{22}, 0, 0, 0))),$$

$L(S_1) =$ узгоджуються (результати, моделювання, результатами, 0, експериментів, лабораторних, 0) & (результати (моделювання, чисельного, хвиль, відокремлених, 0,0,0) & результати (моделювання, чисельного, хвиль, внутрішніх, 0,0,0)),

$L(S_2) =$ вважаємо (ми, 0,0,0,0,0,0) → (можуть & бути (результати, отримані, інтерпретації, процесів & взаємодії, складних, корисними, 0) & (отримані (результати, 0, геометрії, нашої, 0,0,0) & отримані (результати, 0, геометрії, ідеалізованої, 0,0,0))).

3. Граматичні зв'язки для утворення текстів. Одним із засобів зв'язності текстів є повторення граматичних значень.

Морфологічне вираження елементів тексту. Використання тієї чи іншої частини мови зумовлене задачами тексту та функціонально-змістовним типом тексту.

Наприклад, для тексту *«Наукові експерименти в області епідеміології здійснити важко або практично не можливо. Місто*

суттєво впливає на поширення та еволюцію інфекційних захворювань. В основу практично всіх епідеміологічних моделей, розроблених до сьогодні, покладені класичні моделі поширення захворювань», що уже наводився раніше, характерно вживання безособових дієслів «здійснити» та «покладені». Це свідчить про науковий стиль тексту. Також у фрагменті присутня значна кількість іменників, характеристики яких описуються не більше, ніж одним прикметником, що можна виявити за кількістю предикатних змінних та констант у логіко-лінгвістичній моделі речень.

Логіко-лінгвістична модель даного уривку буде мати вигляд:

$$L(S_1) = p_1(0, 0, y_1, q_1, z_1, r_1, h_{11}) \& p_1(0, 0, y_1, q_1, z_1, r_1, -h_{12}),$$

$$L(S_2) = p'_2(x'_2, 0, y'_{21}, 0, z'_2, r'_2, 0) \& p'_2(x'_2, 0, y'_{22}, 0, z'_2, r'_2, 0),$$

$$L(S_3) = p'_3(0, 0, y'_3, q'_3, z'_3, r'_3, h'_3) \&$$

$$(p''_3(h'_3, 0, y'_3, q''_{31}, 0, 0, h''_1) \& p''_3(h'_3, 0, y'_3, q''_{32}, 0, 0, h''_2)),$$

L(S₁) = здійснити (0,0,експерименти, наукові, області, епідеміології, важко) & здійснити (0,0,експерименти, наукові, області, епідеміології, ¬ можливо &практично),

L(S₂) = впливає (місто, 0, поширення, 0, захворювань, інфекційних, суттєво) & впливає (місто, 0, еволюцію, 0, захворювань, інфекційних, суттєво),

L(S₃) = покладені (0,0, моделі, класичні, поширення, захворювань, основу) & (розроблених (основу, 0, моделей, всіх, 0,0, практично) & розроблених (основу, 0, моделей, епідеміологічних, 0,0, практично)).

Єдність граматичних значень виду та часу – означає, що дієслівні форми об'єднують речення природної мови загальним часовим планом існування. Хронотипи можуть характеризувати як окремі відрізки текстів, так і тексти в цілому.

Розглянемо фрагмент тексту «У сучасному землеробстві все гостріше проявляються проблеми економічного та екологічного плану. Аналіз можливих шляхів подальшого підвищення ефективності рослинництва вказує на перспективність впровадження новітніх технологій на основі керованого розвитку рослин» [110].

Логіко-лінгвістична модель даного уривку:

$L(S_1) = p_1(x_1, 0, y_1, q_{11}, z_1, r_1, h_1) \& p_1(x_1, 0, y_1, q_{12}, z_1, r_1, h_1),$
 $L(S_2) = p'_2(x'_2, 0, y'_2, 0, z'_2, 0, h'_2) \& x'_2(0, 0, y''_2, 0, z''_2, r''_2, h''_2) \&$
 $z'_2(h'_1, 0, y''_2, q''_2, z''_2, r''_2, h''_2),$

$L(S_1)$ = проявляються (проблеми, 0, плану, економічного, землеробстві, сучасному, все & гостріше) & проявляються (проблеми, 0, плану, екологічного, землеробстві, сучасному, все & гостріше),

$L(S_2)$ = вказує (аналіз, 0, перспективність, 0, впровадження, 0, основі) & аналіз (0, 0, шляхів, 0, підвищення, подальшого, ефективності & рослинництва) & впровадження (основі, 0, технологій, новітніх, розвитку, керованого, рослин).

Предикати «*проявляються*» і «*вказує*» вжиті у теперішньому часі, що говорить про незавершеність дії.

Синтаксичний паралелізм – паралельна побудова речень вказує на те, що вони пов'язані за змістом. Цей засіб когезії характерний для опису явищ, які відбуваються одночасно або змінюють одне одного. Він проявляється у однаковому порядку розташування членів речення або його предикативної основи та у збігу видово-часових форм дієслів-присудків. Паралельні конструкції пов'язані відношеннями перерахування, співставлення або протиставлення. Синтаксичний паралелізм передбачає не лише структурну подібність речень, але й повторення граматичної семантики конструкцій.

Розглянемо фрагмент «*Наукові експерименти в області епідеміології здійснити важко або практично не можливо. В основу практично всіх епідеміологічних моделей, розроблених до сьогодні, покладені класичні моделі поширення захворювань*».

$L(S_1) = p_1(0, 0, y_1, q_1, z_1, r_1, h_{11}) \& p_1(0, 0, y_1, q_1, z_1, r_1, -h_{12}),$
 $L(S_2) = p'_2(0, 0, y'_2, q'_2, z'_2, r'_2, h'_2) \&$
 $(p''_2(h'_2, 0, y'_2, q''_{21}, 0, 0, h''_2) \& p''_2(h'_2, 0, y'_2, q''_{22}, 0, 0, h''_2)),$

$L(S_1)$ = здійснити (0, 0, експерименти, наукові, області, епідеміології, важко) & здійснити (0, 0, експерименти, наукові, області, епідеміології, \neg можливо & практично),

$L(S_2)$ = покладені (0,0, моделі, класичні, поширення, захворювань, основу)& (розроблених (основу, 0, моделей, всіх, 0,0, практично)& розроблених (основу, 0, моделей, епідеміологічних, 0,0, практично)).

Логіко-лінгвістичні моделі речень мають однакову структуру, зокрема, відсутня предикатна змінна (суб'єкт).

Кожен із розглянутих засобів когезії безпосередньо впливає на змістовні зв'язки у тексті, дає можливість надати інформації того чи іншого стилістичного забарвлення. Тому для здійснення лінгвістичного аналізу тексту важливо точно визначити семантичні зв'язки між його найменшими структурними одиницями (реченнями природної мови).

5.3 Теоретичні основи лінгвістичного аналізу тексту

Аналіз, в процесі якого текст розглядається крізь призму того чи іншого природно мовного поняття, називається понятійно-спрямованим. Тільки за допомогою такого аналізу можна сформулювати уявлення про структуру тексту. Кожен текст, записаний на природній мові можна розбити на типові фрагменти – групи речень, що мають спільне значення [111]. Зокрема:

- предмет та його ознаки (об'єкт та його ознаки – предикатна змінна-аргумент та її константи);

- місце та розташований на ньому предмет (обставина – предикатна константа);

- особу та її дії (предикатна змінна-суб'єкт та предикат);

- особу та її стан (предикатна змінна-суб'єкт та предикатні константи);

- середовище та його стан (обставина – предикатна константа).

Виділяють три види лінгвістичного аналізу:

- **змістовно-композиційний** – полягає у визначенні основного поняття тексту – його теми, основної ідеї, мікротеми, абзаців, плану;

- **стилістичний** – як основне поняття тексту розглядає функціональний стиль, жанрові різновиди та стилістичні ресурси мови;

- *типологічний* – основним поняттям вважається функціонально-змістовний тип мови, або типовий фрагмент тексту.

Повний аналіз тексту представляє собою сукупність перерахованих вище видів лінгвістичного аналізу і проводиться за наступним планом: визначення теми та основного змісту висловлювань, стилю та типу мови, виділення у тексті фрагментів з типовим значенням, визначення способів зв'язку речень, спостереження над функціональними мовними засобами.

Лінгвістичний аналіз тексту є однією з найскладніших методологічних задач, адже текст є засобом передачі складної суми змістів, один із варіантів нескінченної кількості комбінацій у мові. Проте контекст ще не створює текст, він необхідний для підключення мовних механізмів, для актуалізації значень, для запуску процесу референції. Паралельно з актуалізацією понять відбувається процес прагматичного усвідомлення, у результаті якого людина отримує непросту суму пропозицій. Такий процес ще називається *текстуальністю*.

Одна з теорій, що трактує онтологічний статус тексту, це концепція І.Р. Гальперіна, за якою текст – це деяке утворення, до якого можна застосувати методи та прийоми граматичних досліджень [99]. Метою цього є виявлення закономірностей, без яких неможливо досягнути природу створення текстових документів у цілому.

Будь-який текст володіє рядом властивостей.

Цілісність тексту проявляється у тому, що його властивості не можна звести до суми властивостей одиниць, які складають текст.

Зв'язність – одна із найважливіших ознак тексту, що визначає його цілісність. Існує глобальна зв'язність (когерентність), яка досягається завдяки зв'язкам між реченнями, та локальна зв'язність (когезія), засобами якої є повторення одиниць різних рівнів тексту.

Членування – це категорія тексту, протилежна до цілісності, здатність тексту розбиватися на більш мілкі одиниці.

Автосемантия відрізків тексту – властивість тексту, що обумовлена його членуванням. Відокремлені у тексті одиниці володіють відносною самостійністю.

Діалогічність – категорія тексту: внутрішня діалогічність проявляється у тексті у вигляді діалогу, а зовнішня – відображає

взаємодію різних текстів, що дозволяє оцінити текст як реакцію на інші тексти, визначає міжтекстову полеміку.

Модальність – це інтерпретація відношення ситуації та її елементів, що відображаються у тексті, до дійсності.

Текст – це комплекс взаємопов'язаних одне з одним речень, що володіє певною автономністю по відношенню до аналогічних комплексів, змістовною цінністю, яка забезпечується єдністю комунікативного напрямку. Цілісність тексту перетворює його на систему, в якій елементи залежать і передбачають один одного. Це проявляється у двох типах явищ: у пронизуючих та скріплюючих текст повторах деяких змістовних одиниць та в упущенні деяких окремих фрагментів тексту, які в ізольованому реченні були б необхідними, а в цілому тексті можуть бути відновлені з інших речень. А отже, текст можна формалізувати, створивши його змістовну модель.

Текст вважається універсальним засобом представлення, накопичення та передачі знань, тому технології роботи з природно мовними текстами завжди вважалися найважливішими для штучного інтелекту.

Змістова модель тексту – це абстрактна модель, що об'єднує у собі основні властивості тексту та його складових частин, відображає основні взаємозв'язки між структурними компонентами, представляє собою впорядковану четвірку [112]:

$$t = \langle C, F, B, A \rangle, \quad (5.1)$$

де $t \in T$ – конкретний електронний текст із всієї множини текстів;

$C = \{c_1, \dots, c_i, \dots, c_n\}$ – множина існуючих типів текстів, $i = \overline{1, n}$, n – кількість типів;

$F = \{f_1, \dots, f_j, \dots, f_m\}$ – множина складних синтаксичних частин тексту, $j = \overline{1, m}$, m – кількість складних синтаксичних частин;

B – текстова база, що складається з набору ключових слів тексту та взаємопов'язаних пропозицій;

$A = \{a_1, \dots, a_k, \dots, a_q\}$ – множина абзаців тексту, $k = \overline{1, q}$, q – кількість абзаців.

Тоді алгоритм побудови змістовної моделі тексту буде складатися з таких кроків.

1. Визначити тип зв'язного тексту $t \in T$, в межах якого формується глобальний зв'язок $C = \{c_1, \dots, c_i, \dots, c_n\}$ – множина існуючих типів текстів, $i = \overline{1, n}$, n – кількість типів. Кожен тип тексту має свої мовні та когнітивні особливості, на основі чого для кожного з них можна побудувати текстову базу та ситуаційну модель [113]. Властивості тексту як єдиного цілого визначаються особливими цілями, які переслідує його автор [105].

Текстова база включає в себе семантичне представлення тексту та фіксує його зміст у вигляді взаємопов'язаних пропозицій. Залежно від ключових слів, що входять до текстової бази, можна виділити тексти наукового, публіцистичного, художнього та офіційно-ділового стилів (табл.5.1) [114, 115].

Таблиця 5.1

Стилістичні, семантичні та синтаксичні особливості тексту

Тип тексту	Стилістичні, семантичні та синтаксичні особливості
Науковий	Точність, однозначність слова, часто повторюються ключові слова, стилістично нейтральні слова, загальнонаукові терміни; переважання іменників, вживання іменників середнього роду, використання дієслів узагальненого значення, використання недоконаного виду дієслів, вживання займенників третьої особи; прямий порядок слів, використання словосполучень «іменник + іменник у Р.в.», використання безособових або неозначено-особових речень, вживання складних речень, часте використання дієприкметникових та дієприслівникових зворотів.
Публіцистичний	Вживання метафор, риторичних запитань, звертань, фразеологічних зворотів, використання літературних цитат, засоби гумору, сатири; використання іменників у родовому відмінку у ролі неузгоджених означень, використання дієслів у наказовому способі; використання однорідних членів, вставних слів та речень, дієприкметникових та дієприслівникових зворотів, складних речень.

Тип тексту	Стилістичні, семантичні та синтаксичні особливості
Художній	Використовується як книжкова лексика, так і діалектизми, ділові звороти, синоніми, антоніми, пароніми, омоніми, неологізми і т.д.; насичення тексту дієсловами, велика кількість дієприкметників; використовуються всі типи складних речень, використання називних односкладних речень.
Офіційно-діловий	Стандартні звороти, спеціальна термінологія, стійкі словосполучення не емоційного характеру (як правило, це певні шаблони: заяви, пояснювальні записки, автобіографії, заявки); відсутність дієслів першої та другої особи, а також особових займенників першої та другої особи, використання узагальнюючих іменників, форми третьої особи дієслова та займенника – неозначені, використання прийменників та дієслівних іменників у родовому відмінку, іменників чоловічого роду для позначення жіночого роду; вживаються ускладнені прості речення.

2. Здійснити членування тексту та виявити типи зв'язків між структурними частинами тексту. Цей крок відповідає за розбиття всієї текстової інформації (електронного документу), на декілька рівнів. Членування тексту – функція загального композиційного плану документу, основну роль в якому відіграють розмір частин та змістовно-фактуальна інформація.

За Гальперіном можна виділити два типи членування текстів: об'ємно-прагматичне та контекстно-варіативне [99]. В основі першого членування лежить кількісний параметр (поділ на розділи, частини, абзаци і т.д.). Другий тип членування дає можливість встановити типи зв'язків між складними синтаксичними частинами, тобто побудувати схематичну структуру тексту, що аналізується (у тексті виділяють три основні частини: зачин, розгортання думки і висновок). Залежно від типу зв'язного тексту таких складових може бути більше, тому для

конкретної текстової бази буде свій набір складних синтаксичних частин F .

Аналіз складної синтаксичної частини $f_j \in F$ полягає у:

- визначенні структури міжфразової єдності (відіграє велику конструктивну роль, визначає тематичну перспективу всього міжфразового простору);

- виявленні співвідношень складної синтаксичної частини $f_j, j = \overline{1, m}$ та множини абзаців, які входять до неї

$A = \{a_1, \dots, a_k, \dots, a_q\}, k = \overline{1, q}, q$ – кількість абзаців;

- виділенні засобів міжфразового зв'язку (власне лексичні – синоніми, антоніми, лексичні повтори; лексико-граматичні – анафора, катафора; граматичні – порядок слів, сполучники, форми часу, модальні слова, обставини часу, причини, місця і т.д.);

- визначенні провідного зв'язку в складній синтаксичній частині.

3. Визначити тип кожного абзацу та зв'язки між реченнями у ньому.

Абзаци $a_k \in A$ можуть бути:

- **аналітико-синтетичні** – містять аналітичну частину в першій позиції, а узагальнюючу – у другій;

- **синтетико-аналітичні** – починаються з узагальнюючої, стрижневої фрази, зміст якої розкривається у наступних реченнях;

- **рамочні абзаци** мають суміщену структуру: зачин розпочинає тему, далі – пояснююча частина, і завершується абзац узагальнюючою фразою).

Визначення типів зв'язків між реченнями у кожному абзаці є одним із найважливіших етапів проведення лінгвістичного аналізу тексту. Зв'язок між реченнями у тексті буває ланцюговим (контактним) або паралельним (дистанційним).

Ланцюговий зв'язок – думка проходить від одного речення до іншого та здійснюється, зазвичай, при повторі думки з попереднього речення та розгортанні її у наступному реченні. При цьому виді зв'язку наступні речення використовують задане раніше позначення предмету для поширення відомостей про нього, містять у собі наслідок, причину, спираються на вказаний раніше час та місце. Досить часто залежність між реченнями для такого виду зв'язку настільки велика, що без попереднього

речення зміст наступних не зрозумілий. Засобами ланцюгового зв'язку може бути: лексичний повтор, синоніми, вказівні слова, особові займенники, займенникові прислівники, сполучники, словесний пропуск і т.д.

Паралельний зв'язок – полягає у вживанні речень, в яких однаковий порядок слів, однотипні граматичні форми вираження членів речення. Таким чином, речення при такому виді зв'язку рівноправні між собою, при цьому у них здійснюється перерахування чогось, вони зіставляються або порівнюються. Основним засобом реалізації паралельних зв'язків виступає синтаксичний паралелізм, тобто ідентична або схожа побудова речень, що найчастіше проявляється в однаковій послідовності слів та єдності часових форм дієслів-присудків (предикатів).

4. Визначити засоби побудови та відновлення логічних зв'язків, а також засоби когезії у кожному абзаці $a_k \in A$. Зв'язний текст представляє собою деяку закінчену послідовність речень, пов'язаних між собою за змістом у рамках задуму автора. Задача дослідження – встановити види такого зв'язку та визначити правила його передачі.

Речення у контексті розпадаються на дві частини: одна з них містить уже відому інформацію, а друга – нову. Виділення у висловлюванні даного та нового і є **актуальним членуванням**.

Засобами актуального членування є механізми виділення теми і реми. Актуальне членування можливе лише в тексті. Виявлення та перерахування тем відображає змістовну частину тексту. Структура тексту може бути представлена як ланцюжок тем і рем, що чергуються, пов'язуючи елементи тексту одне з одним.

За Ф.Данешом розрізняють п'ять видів **тематичних прогресій** (абстрактних моделей, що лежать в основі побудови текстів), які вживаються в текстах у вигляді комбінацій [98].

Для того, щоб прослідкувати концептуальні зв'язки у тексті, необхідно представити кожен з видів тематичних прогресій у вигляді мережі переходів з кінцевим числом станів.

Мережа переходів з кінцевим числом станів задається множиною вузлів та направлених дуг, що їх з'єднують. Ці вузли відповідають не термінальним символам, а дуги – термінальним [116].

Правила використання мережі переходів з кінцевим числом станів:

1. Обирати одну з направлених дуг, яка виходить з вузла, що розглядається, та пройти по ній.

2. Коли дуга пройдена, необхідно взяти один з термінальних символів підмножини, що відповідає цій дузі.

3. Продовжувати процес доти, доки не буде досягнутий вузол S^* .

4. A, B, C, D, E, F – прості речення природної мови, що описуються простими предикатами типу $L_p^S(x, g, y, q, z, r, h)$.

Проста лінійна прогресія – для неї характерне послідовне розгортання інформації, коли рема попереднього речення стає темою наступного речення. Тобто розгортання тексту відбувається від даного (теми) до нового (реми) (рис. 5.2):

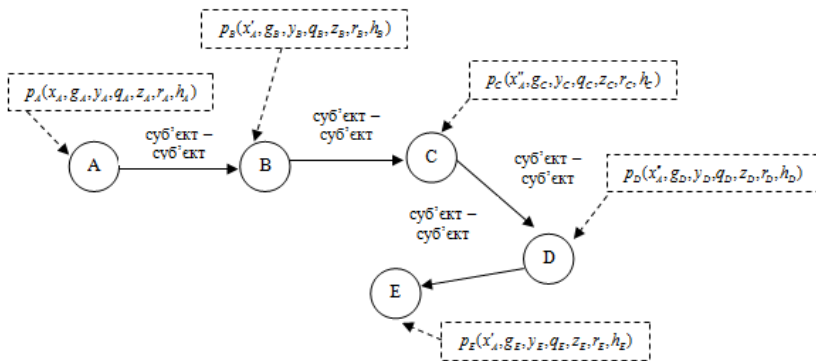


Рис. 5.2. Інтерпретація простої лінійної прогресії у вигляді мережі переходів з кінцевим числом станів

Логіко-лінгвістичні моделі речень природної мови, з'єднані за таким принципом, будуть пов'язані операцією кон'юнкції і матимуть вигляд:

$$L(S^{(1)}) = p^{(1)}(x^{(1)}, g^{(1)}, y^{(1)}, q^{(1)}, z^{(1)}, r^{(1)}, h^{(1)}) \&$$

$$L(S^{(2)}) = p^{(2)}(y^{(1)}, g^{(2)}, y^{(2)}, q^{(2)}, z^{(2)}, r^{(2)}, h^{(2)}) \&$$

.....

$$L(S^{(i)}) = p^{(i)}(y^{(i-1)}, g^{(i)}, y^{(i)}, q^{(i)}, z^{(i)}, r^{(i)}, h^{(i)}) \&$$

$$L(S^{(n)}) = p^{(n)}(y^{(n-1)}, g^{(n)}, y^{(n)}, q^{(n)}, z^{(n)}, r^{(n)}, h^{(n)}),$$

де $i = \overline{1, n}$ – номер речення у фрагментів тексту,

n – загальна кількість речень у фрагменті тексту.

Прогресія з наскрізною темою – характеризується наявністю однієї теми, що повторюється у кожному реченні тексту. Таким чином, одна і та сама тема пронизує весь текст. У моделі такої прогресії перший елемент тематичного ланцюга може бути факультативним (рис. 5.3):

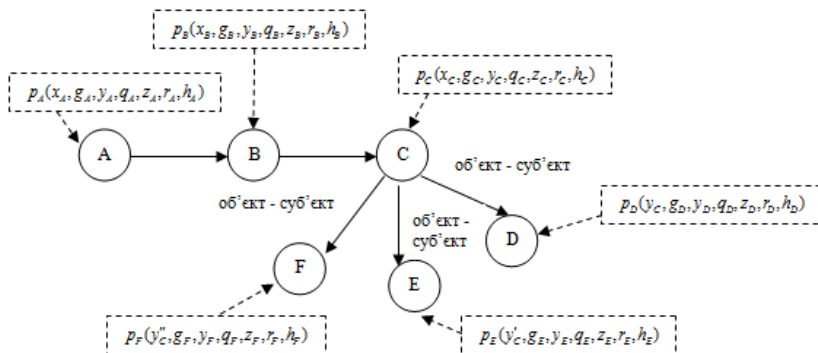


Рис. 5.3. Інтерпретація прогресії з наскрізною темою у вигляді мережі переходів з кінцевим числом станів

Логіко-лінгвістичні моделі речень природної мови, з'єднані за таким принципом, будуть пов'язані операцією кон'юнкції і матимуть вигляд:

$$L(S^{(1)}) = p^{(1)}(x^{(1)}, g^{(1)}, y^{(1)}, q^{(1)}, z^{(1)}, r^{(1)}, h^{(1)}) \&$$

$$L(S^{(2)}) = p^{(2)}(x^{(1)}, g^{(2)}, y^{(2)}, q^{(2)}, z^{(2)}, r^{(2)}, h^{(2)}) \&$$

$$L(S^{(i)}) = p^{(i)}(x^{(1)}, g^{(i)}, y^{(i)}, q^{(i)}, z^{(i)}, r^{(i)}, h^{(i)}) \&$$

$$L(S^{(n)}) = p^{(n)}(x^{(1)}, g^{(n)}, y^{(n)}, q^{(n)}, z^{(n)}, r^{(n)}, h^{(n)}).$$

Прогресія з похідними темами – кожне речення тексту, не маючи у своєму складі елементів послідовної лематизації (перший тип тематичних прогресій) або наскрізної тематизації (другий тип), використовується для вираження загальної тематичної направленості тексту. Тобто ряд окремих тем розкриває одну загальну гіпертему, яка може бути названа у тексті, а може бути відсутня (рис. 5.4):

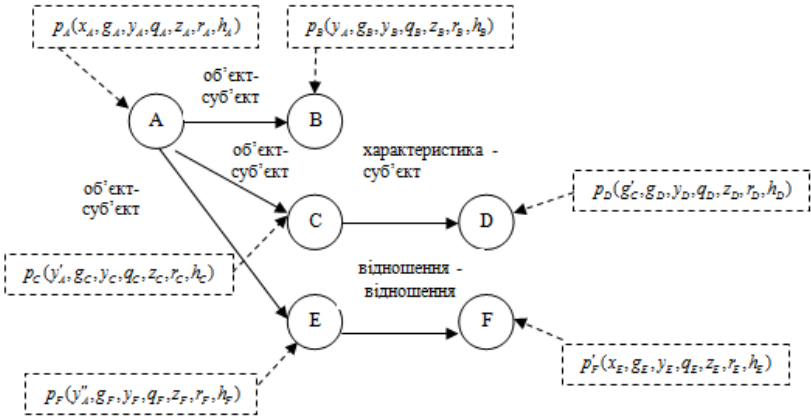


Рис. 5.5. Інтерпретація прогресії з похідними темами у вигляді мережі переходів з кінцевим числом станів

Логіко-лінгвістичні моделі природної мови будуть з'єднані логічною операцією кон'юнкції і матимуть наступний вигляд:

$$L(S^{(1)}) = p^{(1)}(x^{(1)}, g^{(1)}, y^{(1)}, q^{(1)}, z^{(1)}, r^{(1)}, h^{(1)}) \&$$

$$L(S^{(2)}) = p^{(2)}(\tilde{x}^{(1)}, g^{(2)}, y^{(2)}, q^{(2)}, z^{(2)}, r^{(2)}, h^{(2)}) \&$$

.....

$$L(S^{(i)}) = p^{(i)}(\hat{x}^{(1)}, g^{(i)}, y^{(i)}, q^{(i)}, z^{(i)}, r^{(i)}, h^{(i)}) \&$$

.....

$$L(S^{(n)}) = p^{(n)}(\tilde{x}^{(1)}, g^{(n)}, y^{(n)}, q^{(n)}, z^{(n)}, r^{(n)}, h^{(n)}),$$

якщо $\tilde{x}^{(1)}$ – синонім суб'єкта $x^{(1)}$, $\hat{x}^{(1)}$ – гіперонім суб'єкта $x^{(1)}$, $\tilde{x}^{(1)}$ – конверсив суб'єкта $x^{(1)}$.

Якщо загальна тематика фрагменту тексту стосується предмета відношень першого речення, тобто $\hat{z}^{(1)}$ – гіперонім предмета відношень $z^{(1)}$, $\tilde{z}^{(1)}$ – синонім предмета відношень $z^{(1)}$, $\tilde{z}^{(1)}$ –

конверсив предмета відношень $z^{(1)}$, то логіко-лінгвістичні моделі матимуть вигляд:

$$L(S^{(1)}) = p^{(1)}(x^{(1)}, g^{(1)}, y^{(1)}, q^{(1)}, z^{(1)}, r^{(1)}, h^{(1)}) \&$$

$$L(S^{(2)}) = p^{(2)}(\hat{z}^{(1)}, g^{(2)}, y^{(2)}, q^{(2)}, z^{(2)}, r^{(2)}, h^{(2)}) \&$$

.....

$$L(S^{(i)}) = p^{(i)}(\tilde{z}^{(1)}, g^{(i)}, y^{(i)}, q^{(i)}, z^{(i)}, r^{(i)}, h^{(i)}) \&$$

.....

$$L(S^{(n)}) = p^{(n)}(\tilde{z}^{(1)}, g^{(n)}, y^{(n)}, q^{(n)}, z^{(n)}, r^{(n)}, h^{(n)}).$$

Прогресія з розщепленою темою – основу складає подвійна рема, компоненти якої при тематизації утворюють початкові точки для розвитку окремих тематичних прогресій (рис. 5.6):

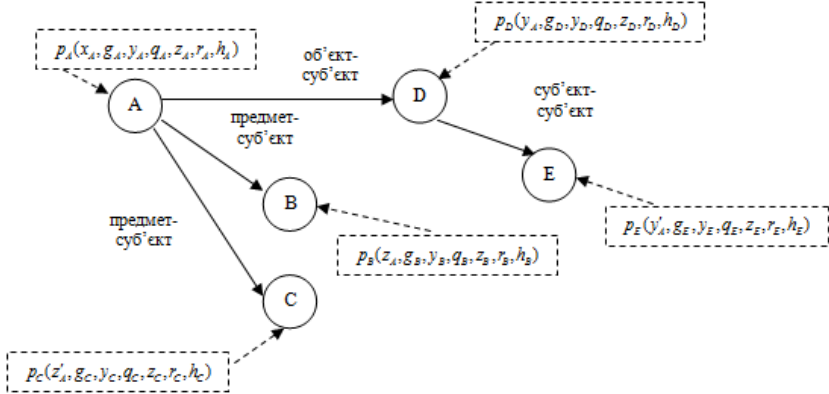


Рис. 5.6. Інтерпретація прогресії з розщепленою темою у вигляді мережі переходів з кінцевим числом станів

Логіко-лінгвістичні моделі речень природної мови, з'єднані за таким принципом, будуть пов'язані операцією кон'юнкції і матимуть вигляд:

$$L(S^{(1)}) = p^{(1)}(x_1^{(1)} \& \dots \& x_n^{(1)}, g^{(1)}, y^{(1)}, q^{(1)}, z^{(1)}, r^{(1)}, h^{(1)}) \&$$

$$L(S^{(2)}) = p^{(2)}(x_1^{(1)}, g^{(2)}, y^{(2)}, q^{(2)}, z^{(2)}, r^{(2)}, h^{(2)}) \&$$

.....

$$L(S^{(i)}) = p^{(i)}(x_i^{(1)}, g^{(i)}, y^{(i)}, q^{(i)}, z^{(i)}, r^{(i)}, h^{(i)}) \&$$

.....

$$L(S^{(n)}) = p^{(n)}(x_n^{(1)}, g^{(n)}, y^{(n)}, q^{(n)}, z^{(n)}, r^{(n)}, h^{(n)}).$$

Прогресія з тематичним стрибком – передбачає наявність розриву в тема-рема-тичному ланцюгу, який можна відновити з контексту (рис. 5.7):

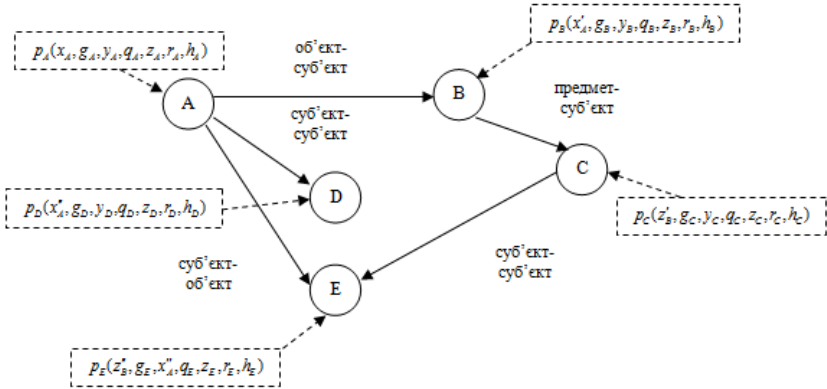


Рис. 5.7. Інтерпретація прогресії з тематичним стрибком у вигляді мережі переходів з кінцевим числом станів

Логіко-лінгвістичні моделі речень природної мови, з'єднані за таким принципом, будуть пов'язані операцією кон'юнкції і матимуть вигляд:

$$L(S^{(1)}) = p^{(1)}(x^{(1)}, g^{(1)}, y^{(1)}, q^{(1)}, z^{(1)}, r^{(1)}, h^{(1)}) \&$$

$$L(S^{(2)}) = p^{(2)}(y^{(1)}, g^{(2)}, y^{(2)}, q^{(2)}, z^{(2)}, r^{(2)}, h^{(2)}) \&$$

.....

$$L(S^{(i)}) = p^{(i)}(z^{(1)}, g^{(i)}, y^{(i)}, q^{(i)}, z^{(i)}, r^{(i)}, h^{(i)}) \&$$

.....

$$L(S^{(n)}) = p^{(n)}(x^{(1)}, g^{(n)}, y^{(n)}, q^{(n)}, z^{(n)}, r^{(n)}, h^{(n)}).$$

Отже, послідовність розгортання змісту текстового фрагменту можна визначити за компонентами логіко-лінгвістичних моделей, зокрема, за зв'язками між їх суб'єктами, об'єктами та предметами відношень.

У реальних текстах тематичні лінії переплітаються по-різному, тому вище перераховані тематичні прогресії не можуть охопити всі види зв'язків у тексті. Проте тематичні прогресії використовують для спостереження за рухом інформації у тексті.

Ще одним видом зв'язку між реченнями є *парцеляція* – розділення складних або поширених висловлювань на несамостійні у змістовому та структурному відношеннях частини. Парцеляція служить засобом змістовного виділення інформативно значимих компонентів речень. Парцеляти представляють собою окремі реми, зв'язка між якими носить особливий характер, відмінний від ланцюгового та паралельного типів зв'язку.

Залежно від лексико-граматичного представлення реми, розрізняють [118] шість рематичних домінант:

- *предметна* – використання у рематичній частині іменників та займенників, що дозволяє висунути на перший план предметний світ;

- *статальна* – містить лексику стану людини та оточуючого середовища, що описується, як правило, за допомогою великої кількості іменників, прикметників та дієслів;

- *динамічна* – у ремі переважають дієслова, спрямовані на конкретну фізичну дію, діяльність, рух;

- *якісна* – спостерігається у текстах із характеристикою предметів, переважають прикметники, іменники оцінювання, прислівники;

- *імпресивна* – у ремі виражається емоційне відношення, переважають дієслова наказового способу;

- *комбінована* – характеризується послідовним з'єднанням та чергуванням у тексті двох або більше рематичних домінант.

5.4 Концепція автоматизованого лінгвістичного аналізу тексту

Якщо розглядати текст як семантичну структуру, то головною проблемою, яку необхідно вирішити при здійсненні лінгвістичного аналізу – сегментувати зміст тексту, класифікувати кванти інформації, що утворюють зміст тексту. Адже, різні елементарні текстові фрагменти співвідносяться з ситуаціями, формуючи у тексті ієрархію семантичних складових. Таким

чином, у розрізі семантичного простору текст розглядається як сукупність лінійно розташованих знаків та як сукупність змістів [119].

Запропонована концепція автоматизованого лінгвістичного аналізу враховує обидві ці позиції. Фактично, результатом автоматизованого лінгвістичного аналізу є змістовна модель тексту (5.1), компоненти якої визначаються за допомогою аналізу логіко-лінгвістичних моделей речень, що формують текст.

Нехай потрібно зробити лінгвістичний аналіз такого тексту.

«Найновіші технічні досягнення дають нам можливість створювати роботизовані автомобілі, що рухатимуться дорогами без втручання людини. Завдяки величезним обчислювальним потужностям, якими володіють комп'ютери, така машина може значно швидше реагувати на непередбачувані ситуації, що виникають на дорогах, а отже, уникати аварій.

Проте міська дорога є досить простим середовищем. Якщо ми хочемо гарантувати безпеку в місті, ми маємо розробляти машини, котрі без проблем справляються із значно складнішими задачами.

Задача управління роботизованим автомобілем складається з цілого комплексу підзадач різного рівня».

Алгоритм автоматизованого лінгвістичного аналізу тексту складається із наступних кроків (рис. 5.8).

1. Членування тексту. На даному етапі відбувається побудова схематичної структури документу, тобто текст розбивається на розділи, частини, абзаци, а абзаци в свою чергу – на речення. Членування тексту – це суто технічний крок, на якому не враховуються синтаксичні та семантичні зв'язки (об'ємно-прагматичне членування).

Для заданого тексту членування відбувається на три абзаци, що в сукупності представляє собою п'ять речень природної мови.

2. Визначення ключових слів та словосполучень. Цей крок дасть змогу частково побудувати текстову базу та з'ясувати тематику тексту:

$$B = \langle K, SJ, D \rangle,$$

де K – множина ключових слів тексту;

SJ – множина ключових словосполучень тексту $S_j, j = \overline{1, m}$;

D – множина пропозицій.

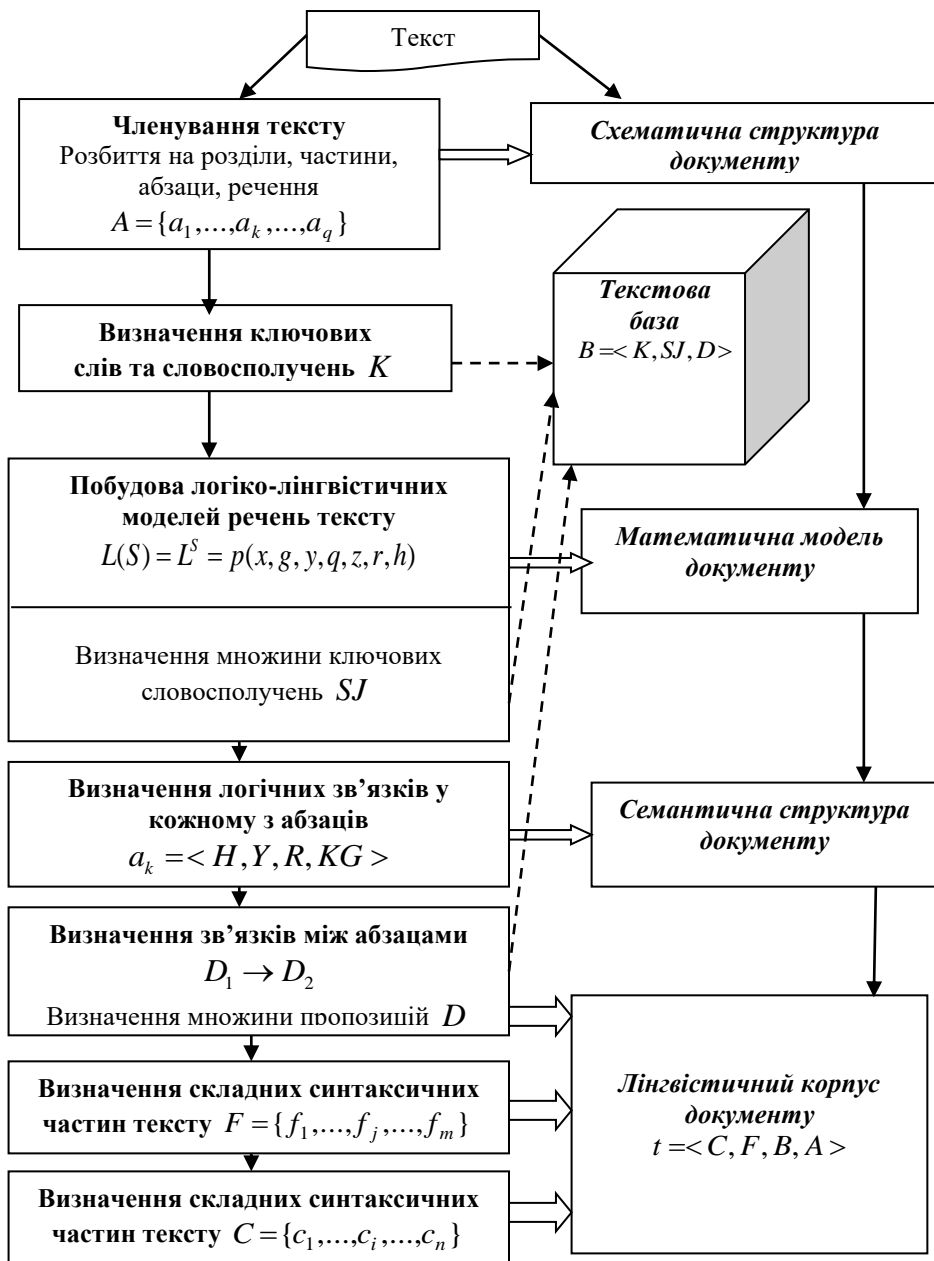


Рис. 5.8. Концепція автоматизованого лінгвістичного аналізу тексту

Ключовими словами даного тексту є «роботизовані», «автомобілі», «дорога», «безпека», «місто», «машина», «задача», «управління». Всі ці слова утворюють множину K .

3. Побудова логіко-лінгвістичних моделей речень тексту.

На даному етапі застосовується метод автоматизованої побудови логіко-лінгвістичних моделей текстової інформації, в результаті чого кожне речення перетвориться в логічну формулу (2.1) – (2.4).

Також продовжує формуватися текстова база, в яку входять синтаксичні конструкції, виявлені при формуванні логіко-лінгвістичних моделей (словосполучення).

Логіко-лінгвістичні моделі заданого тексту будуть мати вигляд:

$$L(S_1) = (p'_1(x'_1, g'_{11}, y'_1, 0, z'_1, r'_1, 0) \& p'_1(x'_1, g'_{12}, y'_1, 0, z'_1, r'_1, 0)) \& p''_1(x''_1, g''_1, 0, 0, z''_1, r''_1, h''_1),$$

$$L(S_2) = (p'_2(x'_2, 0, y'_2, q'_{21}, 0, 0, 0) \& p'_2(x'_2, 0, y'_2, q'_{22}, 0, 0, 0)) \rightarrow p''_{21}(x''_2, g''_2, y''_{21}, q''_2, 0, 0, h''_2) \rightarrow p''_{22}(x''_2, g''_2, y''_{22}, 0, 0, 0, 0),$$

$$L(S_3) = p_3(x_3, g_3, y_3, q_3, 0, 0, h_3),$$

$$L(S_4) = p'_4(x'_4, 0, y'_4, 0, z'_4, 0, 0) \rightarrow p''_4(x''_4, 0, y''_4, 0, 0, 0, 0) \& p'''_4(x'''_4, 0, y'''_4, 0, z'''_4, r'''_4, 0),$$

$$L(S_5) = p'_5(x'_5, 0, y'_5, 0, z'_5, r'_5, 0) \& p''_5(0, 0, y''_5, 0, z''_5, r''_5, 0),$$

$L(S_1)$ = (дають &можливість& створювати (досягнення, найновіші, нам, 0, автомобілі, роботизовані, 0) & дають & можливість & створювати (досягнення, технічні, нам, 0, автомобілі, роботизовані, 0)) & рухатимуться (автомобілі, роботизовані, 0, 0, втручання, людини, дорогами),

$L(S_2)$ = (володіють (комп'ютери, 0, потужностями, величезними, 0, 0, 0) & (володіють (комп'ютери, 0, потужностями, обчислювальними, 0, 0, 0)) → може & реагувати (машина, така, ситуації, непередбачувані, 0, 0, значно & швидше) → може & уникати (машина, така, аварій, 0, 0, 0, 0),

$L(S_3)$ = ϵ (дорога, середовищем, досить простим, 0, 0, протє),

$L(S_4)$ = хочемо & гарантувати (ми, 0, безпеку, 0, місті, 0, 0) → маємо & розробляти (ми, 0, машини, 0, 0, 0, 0) & справляються (машини, 0, проблем, 0, задачами, значно & складнішими, 0),

$L(S_5)$ = складається (задача, 0, комплексу, 0, підзадач, різного & рівня, 0) & задача (0,0, управління, 0, автомобілем, роботизованим, 0).

До множини ключових словосполучень SJ входять словосполучення: «технічні досягнення», «роботизовані автомобілі», «обчислювальні потужності», «непередбачувані ситуації», «виникають на дорогах», «міська дорога», «гарантувати безпеку», «розробляти машини», «задача управління».

4. Визначення логічних зв'язків у кожному з абзаців $a_k \in A$. На даному етапі аналізуються логіко-лінгвістичні моделі речень природної мови, побудовані на третьому кроці алгоритму та визначаються засоби когезії, що використані в абзаці $a_k \in A$, $k = \overline{1, q}$. Кожен абзац a_k представляє собою четвертку:

$$a_k = \langle H, Y, R, KG \rangle,$$

де $H = \{1, 2\}$ – множина типів зв'язків між реченнями (ланцюговий чи паралельний);

$Y = \{1, 2, 3, 4, 5\}$ – множина типів тематичних прогресій, що вжиті у абзаці $a_k \in A$;

$R = \{1, 2, 3, 4, 5, 6\}$ – множина рематичних доміант у абзаці $a_k \in A$;

KG – одновимірний масив засобів когезії, що використовуються у даному абзаці.

Отже, для тексту, що розглядається, множина абзаців $A = \{a_1, a_2, a_3\}$. У першому абзаці ланцюговий тип зв'язку, вжито просту лінійну прогресію, застосовано комбіновану рему (предметна та динамічна). Для зв'язку речень використано такі дискурсивні слова як «що», «якими», «завдяки», а також дійктичні зв'язки шляхом використання займенника «така».

Таким чином, перший абзац тексту характеризується четверкою

$$a_1 = \langle 1, 1, 6, \{ \text{дискурс} _ \text{слова}, \text{дейктичні} _ \text{зв} \} \rangle.$$

У другому абзаці використано паралельний тип зв'язку, використана прогресія з розщепленою темою, якісна рема, вжито дискурсивне слово «котрі», тому

$$a_2 = \langle 2, 4, 4, \{ \text{дискурс_слова} \} \rangle .$$

У третьому абзаці, що складається з одного речення, використано просту лінійну прогресію, предметну рему, тематичний повтор слова «задач», тому

$$a_3 = \langle 1, 1, \{ \text{тематичний_повтор} \} \rangle .$$

5. Визначення зв'язків між абзацами. Призначення реми у тексті полягає не лише у репрезентації нового змісту та актуалізації комунікативної значимості інформації, але і в організації семантики тексту [120]. За межами речення рема вступає в змістовні зв'язки з ремами сусідніх речень, при цьому вона створює рематичну домінанту текстового фрагменту, що сигналізує її семантичну єдність [121]. Аналогічно попередньому етапу уже не речення, а абзаци піддаються перевірці на засоби когезії.

Відбувається остаточне формування текстової бази, в яку входять пропозиції, сформовані на основі отриманих зв'язків між абзацами тексту:

$$D_1 \rightarrow D_2,$$

тут D_1 – частину тексту з абзацу $a_k \in A$, що за змістом передуює або з якої впливає частина тексту D_2 з абзацу $a_{k+1} \in A$, $k = \overline{1, q}$.

Використані засоби когезії у тексті дають змогу здійснити заміни: $x_1'' \equiv z_1'$, $g_1'' \equiv r_1'$, $x_2'' \equiv z_1'$, $g_2'' \equiv r_1'$, $x_3 \equiv h''$, $x_4' \equiv y_1'$, $x_4'' \equiv y_1'$, $y_4'' \equiv z_1'$, $x_4''' \equiv z_1'$, $z_4''' \equiv x_5'$, $z_5' \equiv \widehat{x}_5'$, $p_5'' \equiv x_5'$, $z_5'' \equiv z_1'$, $r_5'' \equiv r_1'$ та привести логіко-лінгвістичні моделі речень до більш загального вигляду:

$$L(S_1) = (p_1'(x_1', g_{11}', y_1', 0, z_1', r_1', 0) \& p_1'(x_1', g_{12}', y_1', 0, z_1', r_1', 0)) \& p_1''(z_1', r_1', 0, 0, z_1'', r_1'', h_1''),$$

$$L(S_2) = (p_2'(x_2', 0, y_2', q_{21}', 0, 0, 0) \& p_2'(x_2', 0, y_2', q_{22}', 0, 0, 0)) \rightarrow p_{21}''(z_1', r_1', y_{21}'', q_2'', 0, 0, h_2'') \rightarrow p_{22}''(z_1', r_1', y_{22}'', 0, 0, 0, 0),$$

$$L(S_3) = p_3(x_3, g_3, y_3, q_3, 0, 0, h_3),$$

$$L(S_4) = p'_4(y'_1, 0, y'_4, 0, z'_4, 0, 0) \rightarrow p''_4(y'_1, 0, z'_1, 0, 0, 0, 0) \& \\ p'''_4(z'_1, 0, y'''_4, 0, x'_5, r'_4, 0),$$

$$L(S_5) = p'_5(x'_5, 0, y'_5, 0, \bar{x}'_5, r'_5, 0) \& x'_5(0, 0, y''_5, 0, z'_1, r'_1, 0),$$

Саме замінені слова та словосполучення є зв'язками між абзацами конкретного тексту.

6. Визначення складних синтаксичних частин тексту
 $f_j \in F$, $F = \{f_1, \dots, f_j, \dots, f_m\}$. У результаті отриманих зв'язків між абзацами на попередньому етапі алгоритму текст ділиться на складні синтаксичні частини (контекстно-варіативне членування).

Увесь текст складається із однієї складної синтаксичної частини f_1 .

7. Визначення типу тексту c_i . На основі визначених на попередніх етапах типах змістовних зв'язків між складними та простими синтаксичними конструкціями, визначається тип тексту $c_i \in C$, $C = \{c_1, \dots, c_i, \dots, c_n\}$ – множина існуючих типів текстів, $i = \overline{1, n}$, n – кількість типів.

Для даного тексту характерна точність, однозначність слова, у різних абзацах часто повторюються ключові слова, у тексті переважають іменники, а також сполучення «іменник + іменник у Р.в.» та «дієслово + іменник у Р.в.», тому даний текст наукового стилю, $c_1 = 1$.

Результати досліджень, висвітлених у монографії, базуються на одному із підходів до вирішення основної проблеми автоматичної обробки текстової інформації (відсутності формального апарату, який дозволив би аналізувати текстову інформацію за єдиним принципом, систематизувати процес змістовного пошуку та уникнути неоднозначності при аналізі текстових документів). Зокрема, підхід заснований на формуванні змістовних моделей текстових документів на базі використання математичного апарату логіки предикатів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Клыков Ю.И. Банки данных для принятия решений / Ю.И. Клыков, Л.Н. Горьков. – М.: Сов. радио, 1980. – 208 с.
2. Хант Э. Искусственный интеллект / Хант Э. – М.: Мир, 1978. – 560 с.
3. Люгер Д.Ф. Искусственный интеллект: стратегии и методы решения сложных проблем / Д.Ф. Люгер. – М.: ООО «Вильямс», 2005. – 4-тое издание. – 864 с.
4. Вавіленкова А.І. Методи та алгоритми автоматизованого формування логіко-лінгвістичних моделей текстової інформації: дисертація на здобуття наукового ступеня кандидата технічних наук: 05.13.06 / НАН України, Інститут проблем математичних машин і систем. – Київ, 2011. – 155 с.
5. Таунсенд К. Проектирование и программная реализация экспертных систем на персональных ЭВМ / К. Таунсенд, Д. Фохт. – М.: Финансы и статистика, 1990. – 320 с.
6. Рубашкин В.Ш. Представление и анализ смысла в интеллектуальных информационных системах / Рубашкин В.Ш. – М.: Наука. Гл. ред. физ.-мат. лит., 1989. – 192 с.
7. Корпусна лінгвістика / [Широков В.А., Бугаков О.В., Грязнухіна Т.О. та ін.]. – К.: Довіра, 2005. – 471 с.
8. Звегинцев В.А. Новое в зарубежной лингвистике. Новое в лингвистике: вып. 2 / В.А. Звегинцев. – М.: Из-во «Прогресс», 1962. – 686 с.
9. Звіт про науково-дослідну роботу «Комп'ютерна технологія порівняльного аналізу електронних текстів (№589-ДБ09)». Розробка методів та алгоритмів попередньої обробки електронних текстів, що передують процедурі їх порівняльного аналізу, 2009. – 50 с.
10. Дюк В.А. Data Mining – интеллектуальный анализ данных / В.А. Дюк. – Санкт-Петербургский институт информатики и автоматизации РАН, 2015, http://iteam.ru/publications/it/section_92/article_1448.
11. Беляков И. А. Применение интеллектуальных технологий в процессе сертификации программного обеспечения / И. А.

Беляков, М. А. Еремеев // Молодой ученый. — 2011. — №11. Т.1. — С. 23-31.

12. Орган управления и координатор ГНТП «Образный компьютер», http://www.obrazcomp.irtc.org.ua/Rus/Filtr2_r.html.

13. Шемакин Ю.И. Компьютерная семантика / Ю.И. Шемакин, А.А. Романов – М.: НОЦ «Школа Китайгородской», 1995. – 344 с.

14. Словарь по кибернетике / [ред. академик В.М. Глушков]. – К.: Главная редакция украинской советской энциклопедии, 1979. – 624 с.

15. Хомский Н. Три модели языка / Н. Хомский // Кибернетический сборник. – 1961. – Вып. 2. – С. 81–92.

16. Хомский Н. О понятии «правило грамматики» / Н. Хомский. – С. 115-146.

17. Хопкрофт Дж. Введение в теорию автоматов, языков и вычислений / Дж. Хопкрофт, Р. Мотвани, Дж. Ульман. – М.: ООО «Вильямс», 2002. – 528 с.

18. Бухараев Р.Г. Семантический анализ в вопросно-ответных системах / Р.Г. Бухараев, Д.Ш. Сулейманов. – Казань: Издательство Казанского университета, 1990. – 124 с.

19. Осуга С. Приобретение знаний / С.Осуга, Ю.Сазки. – М.: Мир, 1990. – 304 с.

20. Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – Спб.: Питер, 2000. – 384 с.

21. Искусственный интеллект: в 3-х кн. / [ред. Э.В. Попов]. – Кн.1: Системы общения и экспертные системы: справочник. – М.: Радио и связь, 1990. – 464 с.

22. Попов Э.В. Статические и динамические экспертные системы / Э.В. Попов, И.Б. Фоминых, Е.Б. Кисель, М.Д. Шапот. – М.: Финансы и статистика, 1996. – 320 с.

23. Поспелов Д.А. Моделирование рассуждений. Опыт анализа мыслительных актов / Поспелов Д.А. – М.: Радио и связь, 1989. – 114 с.

24. Вагин В.Н. Дедукция и обобщение в системе принятия решений / Вагин В.Н. – М.: Наука, 1988. – 384 с.

25. Джарратано Д. Экспертные системы: принципы разработки и программирование / Д. Джарратано, Г. Райли; пер. с англ. К.А. Птицына. – 4-е издание. – М.: ООО «Вильямс», 2007. – 1152 с.

26. Портал искусственного интеллекта, <http://www.aiportal.ru/articles/knowledge-models/1/>.

27. Вавіленкова А.І. Аналіз моделей представлення знань в експертних системах управління / А.І. Вавіленкова // Проблеми інформатизації та управління: зб. наук. праць. – К.: НАУ, 2007. – Вип. 4 (22). – С. 14–17.

28. Эндрю А. Искусственный интеллект / Эндрю А.; пер. с англ. Д.А. Поспелова. – М.: Мир, 1985. – 264 с.

29. Рогозин О.В. Применение фреймовой модели представления знаний на примере решения задачи Эйнштейна / О.В. Рогозин // Средства и методы повышения качества подготовки специалистов. – 2008. – №4. – С. 8–14.

30. Уэно Х. Представление и использование знаний / Х. Уэно, М. Исидзука. – М.: Мир, 1989. – 220 с.

31. Попов Э.В. Экспертные системы: решение неформализованных задач в диалоге с ЭВМ / Попов Э.В. – М.: Наука. Гл. ред. физ.-мат. лит., 1987. – 288 с.

32. Осуга С. Обработка знаний / Осуга С. – М.: Мир, 1989. – 293 с.

33. Орехов Ю.В. Математическая логика: учебное пособие / Ю.В. Орехов, Э.Ю. Орехов. – Уфа: УГАТУ, 2006. – 161 с.

34. Гладкий А.В. Синтаксические структуры естественного языка в автоматизированных системах общения / Гладкий А.В. – М.: Наука. Гл. ред. физ.-мат. лит., 1985. – 144 с.

35. Поспелов Д. А. Логико-лингвистические модели в системах управления / Поспелов Д.А. – М.: Энергоиздат., 1981. – 232 с.

36. Вавіленкова А.І. Аналіз онтології як моделі представлення знань текстової інформації / А.І. Вавіленкова // Міжнар. наук.-техн. конф. «Інтелектуальні технології лінгвістичного аналізу»: тези доповідей. – К.: НАУ- друк, 2010. – С. 16.

37. Башмаков А.И. Интеллектуальные информационные технологии: Учеб. пособие. / А.И. Башмаков, И.А. Башмаков – М.:Изд-во МГТУ им. Баумана, 2005. – 304 с.

38. Онтологии и представление знаний Б. Конев, Ливерпульский университет, 2010, <http://logic.pdmi.ras.ru>.

39. Сокирко А.В. Графематический анализ // АОТ: Технологии: Графематический анализ, 2005г., <http://www.aot.ru/docs/graphan.html>.

40. Седунов А.А. Модель графематического анализа в системе обработки естественного языка. – Вестник ВГУ, серия:

Системный анализ и информационные технологии, 2007, № 2. – С. 69-77.

41. Соколовский В.В. Обзор задач и методов смысловой обработки электронных данных, роль метаданных для практических задач смысловой обработки // ГПНТБ России: http://ellib.gpntb.ru/doc/11/d11_12.htm

42. Грамматический разбор. Словообразовательный, морфемный и морфологический анализ слова, gumfak.ru

43. Плющ М.Я. Українська мова. Довідник / М.Я. Плющ, Н.Я. Грипас. – К.: Радянська школа, 1990. – 255 с.

44. Попов Э.В. Общение с ЭВМ на естественном языке / Попов Э.В. – М.: Наука. Главная редакция физико-математической литературы, 1982. – 360 с.

45. Фостер Дж. Автоматический синтаксический анализ/ Форстер Дж: пер. с англ. В.В. Мартынюка. – М.: Мир, 1975. – 70 с.

46. Звегинцев В.А. Новое в зарубежной лингвистике. Прикладная лингвистика: вып. 12 / Звегинцев В.А. – М.: Из-во «Прогресс», 1983. – 460 с.

47. Звегинцев В.А. Новое в зарубежной лингвистике. Лингвистическая семантика: вып. 10 / Звегинцев В.А. – М.: Из-во «Прогресс», 1981. – 568 с.

48. Звегинцев В.А. Новое в зарубежной лингвистике. Когнитивные аспекты языка: вып. 23 / Звегинцев В.А. – М.: Из-во «Прогресс», 1988. – 320 с.

49. Вавіленкова А.І. Обробка текстової інформації через призму аналізу та інтерпретації елементів формальної системи /А.І. Вавіленкова // Системи підтримки прийняття рішень. Теорія і практика: зб. доп. наук.-практ. конф. з міжнар. участю. – Київ: ІПММС НАНУ, 2009. – С. 198–201.

50. Джордж Ф. Основы кибернетики / Ф. Джордж. – М.: Радио и связь, 1984. – 272 с.

51. Васильев Л.М. Современная лингвистическая семантика / Л.М. Васильев. – М.: Высш. шк., 1990. – 176с.

52. Кронгауз М.А. Семантика / М.А. Кронгауз. – М.: Издательский центр «Академия», 2005. – 352 с.

53. Звегинцев В.А. Новое в зарубежной лингвистике. Логический анализ естественного языка: вып. 18 / Звегинцев В.А. – М.: Из-во «Прогресс», 1986. – 392 с.

54. Волкова И.А. Введение в компьютерную лингвистику. Практические аспекты создания лингвистических процессоров / Волкова И.А. – М.: Издательский отдел факультета ВМиК МГУ, 2006. – 43 с.

55. Ф. де Соссюр Труды по языкознанию / Ф. де Соссюр. – М.: Прогресс, 1977. – 696с.

56. Вавіленкова А.І. Логіко-лінгвістична модель як засіб відображення синтаксичних особливостей текстової інформації /А.І. Вавіленкова // Математичні машини та системи. – 2010. – № 2. – С. 134–137.

57. Чери С. Логическое программирование и базы данных / С. Чери, Г. Готлоб, Л. Танка. – М.: Мир, 1992. – 352 с.

58. Лавров И.А. Задачи по теории множеств, математической логике и теории алгоритмов / И.А. Лавров, Л.Л. Максимова. – [3-е изд.]. – М.: Физматлит, 1995. – 250 с.

59. Вавіленкова А.І. Методологічні основи автоматичного аналізу логіко-лінгвістичних моделей текстових документів /А.І. Вавіленкова // Математичні машини та системи. – 2015. – № 1. – С. 65–71.

60. Марчук Ю.Н. Компьютерная лингвистика / Ю.Н. Марчук. – М.: АСТ: Восток – Запад, 2007. – 317с.

61. Вавіленкова А.І. Формування логіко-лінгвістичних моделей типових природно мовних конструкцій /А.І. Вавіленкова // Східно-європейський журнал передових технологій. – 2015. – № 3(75). – С. 35–46.

62. Литвиненко А. Е. Метод направленного перебора в системах управления и диагностирования / Литвиненко А.Е. – К.: Национальная библиотека Украины имени В.И. Вернадского, 2007. – 328 с.

63. Мацько Л.І. Українська мова. Посібник / Л.І. Мацько, О.М, Сидоренко. – К.: Либідь, 1996. – 432с.

64. Вавіленкова А.І. Извлечение смысла из предложений естественного языка // Программные продукты и системы. – Тверь: Главная редакция международного журнала НИИ "Центрпрограммсистем". – 2012. – № 4(100). – С. 87–90.

65. Вавіленкова А.І. Методика перетворення речення в логіко-лінгвістичну модель /А.І. Вавіленкова // Міжнар. наук.-техн. конф. «Комп'ютерні системи та мережні технології»: зб. тез. – К.: НАУ, 2008. – С. 49-53.

66. Кузнецов О.П. Дискретная математика для инженера / О.П. Кузнецов, Г.М. Адельсон-Вельский. – М.: Энергия, 1980 – 344 с.
67. Барсегян А.А. Методы и модели анализа данных: OLAP и Data Mining / А.А. Барсегян, М.С. Куприянов., В.В. Степаненко, И.И. Холод. – С.-Пб.: БХВ-Петербург, 2007. – 384с.
68. Современное состояние проблемы распознавания: Некоторые аспекты / [А.Л. Горелик, И.Б. Гуревич, В.А. Скрипкин и др.]. – М.: Радио и связь, 1985. – 160с.
69. Санников В.З. Русский синтаксис в семантико-прагматическом пространстве / Санников В.З. – М.: Языки славянских культур, 2008. – 625 с.
70. Нейман Дж. Вероятностная логика и синтез надежных организмов из ненадежных компонент / Дж. Нейман // Автоматы. – М.: ИЛ, 1956. – С. 68-139.
71. Искусственный интеллект: в 3-х кн. / [ред. Д.А. Поспелов]. – Кн. 2: Модели и методы: Справочник. – М.: Радио и связь, 1990. – 304 с.
72. Вавіленкова А.І. Обробка текстової інформації через призму аналізу та інтерпретації елементів формальної системи / А.І. Вавіленкова // Системи підтримки прийняття рішень. Теорія і практика: зб. доп. наук.-практ. конф. з міжнар. участю.–Київ: ІПММС НАНУ, 2009. – С. 198–201.
73. Никитин М.В. Курс лингвистической семантики / М.В. Никитин. – СПб.: Из-во РГПУ им. Герцена, 2007. – 819с.
74. Смальян Р. Теория формальных систем / Р. Смальян. – М.: «Энергия», 1981. – 208 с.
75. Лайонз Дж. Лингвистическая семантика. Монография. - М.: Языки славянской культуры, 2003. — 400 с.
76. Гальперин Р.И К проблеме зависимости предложения от контента / Р.И. Гальперин, 1977.
77. Левент Поверхностные и глубинные структуры языка, <http://bibl.tikva.ru/base/B1706/B1706Part28-80.php>.
78. Вавіленкова А.І. Виявлення логічних протиріч у текстовій інформації / А.І. Вавіленкова// Вісник НТУ "ХПІ". Збірник наукових праць. Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2012. – № 38. – С. 32 – 37.
79. Кобозева И.М. Лингвистическая семантика / Кобозева И.М. – М.: Эдитореал УРСС, 2000. – 352 с.

80. Новое в зарубежной лингвистике. Вып.18. Логический анализ естественного языка: Пер. с англ. Петрова В.В. – М.: Прогресс, 1986. – 392с.

81. Вавіленкова А.І. Особливості методу резолюцій під час здійснення порівняльного аналізу текстової інформації /А.І. Вавіленкова // Вісник Національного університету «Львівська політехніка» – 2011. – № 699. – С. 292–302.

82. Величковский Б.М. Значение в глобальных когнитивных моделях / Б.М. Величковский. Современная когнитивная психология. – М.: МГУ, 1982. – С. 206–221.

83. Вывод в логических моделях. Метод резолюций, <http://www.aiportal.ru/articles/knowledge-models/method-resolution.html>.

84. Левент Поверхностные и глубинные структуры языка <http://bibl.tikva.ru/base/B1706/B1706Part28-80.php>.

85. Греймас А.-Ж. Структурная семантика. Поиск метода / А.-Ж. Греймас. – М.: Академический проект, 2004. – 368с.

86. Dan Jurafsky, Professor of Linguistics, Christopher Manning, Associate Professor of Computer Science Natural Language Processing, Stanford University, 2012, <https://www.coursera.org/course/nlp>.

87. Ф.де Соссюр Курс общей лингвистики / Ф.де Соссюр. – Париж, 1991. – 284с.

88. Вавіленкова А.І. Алгоритм порівняння логіко-лінгвістичних моделей речень природної мови. Системи підтримки прийняття рішень. Теорія і практика: зб. доп. наук.-практ. конф. з міжнар. участю. – Київ: ІПММС НАНУ, 2012. – С. 132–135.

89. Вавіленкова А.І. Визначення семантичної істинності речень природної мови /А.І. Вавіленкова // Збірник тез V Міжнар. наук.-техн. конф. «Комп'ютерні системи та мережні технології» (CSNT–2012). – К.: НАУ, 2012. – С. 25.

90. Кузин Л.Т. Основы кибернетики: В 2-ч т. Т.2. Основы кибернетических моделей. Учеб. Пособие для вузов. – М.: Энергия, 1979. – 584с.

91. М.Руссо Чарльз Филмор: семантические роли и ногою другое, 2014, http://polit.ru/article/2014/02/21/ps_fillmore/.

92. Мельчук И.А. Опыт теории лингвистических моделей СМЫСЛ-ТЕКСТ. – М.: Языки русской культуры, 1999. – 346с.

93. Вавіленкова А.І. Структура лінгвістичного процесору системи порівняльного аналізу текстів за змістом /А.І. Вавіленкова // Системи підтримки прийняття рішень. Теорія і практика: зб. доп. наук.-практ. конф. з міжнар. участю.– Київ: ІПММС НАНУ, 2011. – С. 153–156.

94. Вавіленкова А.І. Виявлення компонентів змісту в реченнях природної мови /А.І. Вавіленкова // Сьома наук.-практ. конф. з міжнар. участю «Математичне та імітаційне моделювання систем. МОДС 2012»: тези доповідей. – Чернігів, 2012. – С. 240–242.

95. Ермаков А.Е. Эксплицирование элементов смысла текста средствами синтаксического анализа-синтеза // А.Е. Ермаков Компьютерная лингвистика и интеллектуальные технологии: труды Международной конференции Диалог'2003. – Москва, Наука, 2003.

96. Вавіленкова А.І. Логіко-лінгвістичні моделі речень як засіб порівняння текстових документів за змістом /А.І. Вавіленкова // Математичні машини та системи. – 2012. – № 1. – С. 166–173.

97. Глазков А. Текст и текстуальность/ А. Глазков. – М., 2010.

98. Філіппов К.А. Лингвистика текста. Курс лекцій. – Спб.: Издательство С.-Петербургского университета, 2008. – 336с.

99. Гальперин И.Р. Текст как объект лингвистического исследования. Изд. 5-тое, стереотипное / И.Р. Гальперин. – М: КомКнига, 2007. – 144с.

100. Лукашевич, 2011, Когезия как структурная связность текста, <http://like-money.ru/stati/271-kogeziya-kak-strukturnaya-svyaznost-teksta>.

101. Кибрик А.А. Фундаментальные направления современной американской лингвистики / А.А. Кибрик, В.А. Плунгян. – В сб.: под ред. А.А. Кибрика, И.М. Кобозевой и И.А. Секериной, М., 1997.

102. Заїка С.О. Агентне моделювання циркуляції респіраторних вірусів у міських екосистемах /С.О. Заїка, О.Л. Ляхов, А.Т. Лобурець, Ю.В. Величко // Восьма наук.-практ. конф. з міжнар. участю «Математичне та імітаційне моделювання систем. МОДС 2012»: тези доповідей. – Чернігів, 2013. – С. 38–41.

103. Рагозін Д.В. Дворівневі сенсорні мережі з мобільними датчиками /Д.В. Рагозін, // Восьма наук.-практ. конф. з міжнар. участю «Математичне та імітаційне моделювання систем. МОДС 2012»: тези доповідей. – Чернігів, 2013. – С. 114–118.

104. Обламський В.В. Створення стратегії ефективної боротьби з лісовими пожежами шляхом їхнього моделювання /В.В. Обламський // Восьма наук.-практ. конф. з міжнар. участю «Математичне та імітаційне моделювання систем. МОДС 2012»: тези доповідей. – Чернігів, 2013. – С. 31–33.

105. Головкина С.Х. Лингвистический анализ текста //С.Х. Головкина, С.Н. Смольников. – Вологда: Издательский центр ВИРО, 2006. – 124с.

106. Валгина Н.С. Теория текста. – М.: Логос, 2003.

107. Гайтан О.М. Математичне моделювання систем за допомогою теорії взаємодіючих послідовних процесів, часовий аспект /О.М. Гайтан // Восьма наук.-практ. конф. з міжнар. участю «Математичне та імітаційне моделювання систем. МОДС 2012»: тези доповідей. – Чернігів, 2013. – С. 287–289.

108. Філімоніхіна І.І. Комп'ютерне моделювання динаміки обертового несучого тіла з двох маятниковим автобалансиром /І.І. Філімоніхіна // Восьма наук.-практ. конф. з міжнар. участю «Математичне та імітаційне моделювання систем. МОДС 2012»: тези доповідей. – Чернігів, 2013. – С. 176–177.

109. Терлецька К.В. Трансформація відокремленої внутрішньої хвилі на сходинці в рельєфі дна: втрата енергії /К.В. Терлецька, В.С. Мадерич, І.О. Бровченко, Т. Таліпова, Ю. Пелінковський // Восьма наук.-практ. конф. з міжнар. участю «Математичне та імітаційне моделювання систем. МОДС 2012»: тези доповідей. – Чернігів, 2013. – С. 75–77.

110. Адамчук В.В. Перспективи розвитку механізованого рослинництва /В.В. Адамчук, В.Г. Мироненко, П.О. Косик // Восьма наук.-практ. конф. з міжнар. участю «Математичне та імітаційне моделювання систем. МОДС 2012»: тези доповідей. – Чернігів, 2013. – С. 240–243.

111. Фоменко Ю.В. Мифы современной лингвистики: монография/ Ю. В. Фоменко. – Новосибирск: Изд. НГПУ, 2010. – 176 с.

112. Вавіленкова А.І. Синтез логіко-лінгвістичних моделей речень природної мови як засіб побудови змістовної моделі тексту /А.І. Вавіленкова // Системи підтримки прийняття рішень. Теорія і практика: зб. доп. наук.-практ. конф. з міжнар. участю. – Київ: ПММС НАНУ, 2013. – С. 49–51.

113. Пономаренко И.Н. Симметрия и асимметрия в лингвистике текста/ И. Н. Пономаренко. – Краснодар: Кубанский государственный университет, 2005. – 322 с.

114. Кутузова А.Б. Методики определения сложности текста в рамках переводческого анализа // А.Б. Кутузова. – Вестник Нижегородского государственного лингвистического университета им. Н.А. Добролюбова. – Вып.4. – Лингвистика и межкультурная коммуникация. – Нижний Новгород: НГЛУ, 2009 г.

115. Схема лингвистического анализа текста, <http://www.novsh9.ru /ruskiy-yazyk/2013-05-10/novost-2>.

116. Vavilenkova A.I. Semantic structuring of chaotic-standing sentences of natural language /A.I. Vavilenkova // 6th international conference «Chaotic Modeling and Simulation (CHAOS 2013)»: abstracts. – Istanbul (Turkey), 2013. – P.134.

117. Золотова Г.А. Коммуникативная грамматика русского языка// Г.А. Золотова, Н.К. Онипенко, М.Ю. Сидорова. – М., 1988.

118. Вавіленкова А.І. Застосування методу розширених мереж переходів для здійснення семантичного аналізу складних речень природної мови /А.І. Вавіленкова // Восьма наук.-практ. конф. з міжнар. участю «Математичне та імітаційне моделювання систем. МОДС 2012»: тези доповідей. – Чернігів, 2013. – С. 274–277.

119. Лагута О.Н. Логика и лингвистика - Новосибирск, 2000, 116 с.

120. Evans, V. Lexical concepts, cognitive models and meaning-construction / V. Evans // Journal of Cognitive semiotics. – 2006. – P. 73–107.

121. Gries, S. Th. Corpus-based methods and cognitive semantics: the many meanings of to run / S. Th. Gries // Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis. – 2006. – P. 57–99.

