



А. І. Воржевітіна<sup>1</sup>, В. Д. Шанідзе<sup>2</sup>

<sup>1</sup>НТУ «ХПІ», м. Харків, Україна, Anzhelika.Vorzhevitina@sgt.khpi.edu.ua,  
ORCID iD: 0009-0001-0562-0191

<sup>2</sup>НТУ «ХПІ», м. Харків, Україна, vladyslavashanidze@gmail.com,  
ORCID iD: 0009-0002-9809-5572

## ФОРМУВАННЯ СИСТЕМИ ОЗНАК ДЛЯ МОДЕЛЮВАННЯ ЛІНГВІСТИЧНИХ СТРУКТУР НА ОСНОВІ МЕТОДУ КОМПАРАЦІЇ ЗА ЗАСОБІВ ТЕОРІЇ ІНТЕЛЕКТУ

У статті розглядається проблема формалізації лінгвістичних структур в рамках використання основного методу теорії інтелекту – методу компараторної ідентифікації та математичного інструментарію – апарату алгебри скінченних предикатів (АСП). Автори обґрунтовують перехід від імовірнісних методів аналізу тексту до детермінованих логіко-математичних моделей, що забезпечує високий ступінь інтерпретованості результатів. Запропоновано метод ідентифікації деструктивних і маніпулятивних лінгвістичних структур на основі компараторного підходу, що дозволяє системі виконувати абдуктивний висновок і відновлювати пропущені ознаки (лакуни) в умовах невизначеності даних. У роботі детально описано алфавіт предикатних змінних (інтенція, суб'єктні відносини, лексичні маркери) і наведено результати верифікації моделі на контрольних вибірках текстів кримінальної та маніпулятивної спрямованості. Дослідження робить внесок у вирішення проблеми «заземлення символів» (Symbol Grounding) і створення прозорих когнітивних архітектур для систем інформаційної безпеки.

ЛІНГВІСТИЧНІ СТРУКТУРИ, АЛГЕБРА СКІНЧЕННИХ ПРЕДИКАТИВ, ТЕОРІЯ ІНТЕЛЕКТУ, КОМПАРАТОРНА ІДЕНТИФІКАЦІЯ, АБДУКТИВНИЙ ВИСНОВОК, ІНФОРМАЦІЙНА БЕЗПЕКА, ІНТЕРПРЕТОВАНИЙ ШІ, РЕАВТОМАТИЗАЦІЯ МОВИ

**A. I. Vorzhevitina, V. D. Shanidze. Formation of a system of features for modelling linguistic structures based on the method of comparison using the tools of the theory of intelligence.** This article examines the problem of formalizing linguistic structures within the framework of the main method of intelligence theory—the method of comparative identification and mathematical tools—the apparatus of finite predicate algebra (FPA). The authors justify the transition from probabilistic methods of text analysis to deterministic logical-mathematical models, which ensures a high degree of interpretability of the results. A method for identifying destructive and manipulative linguistic structures based on a comparator approach is proposed, enabling the system to perform abductive inference and restore missing features (lacunae) in the face of data uncertainty. The paper describes in detail the alphabet of predicate variables (intent, subject relations, lexical markers) and presents the results of model verification on control samples of criminal and manipulative texts. The study contributes to solving the problem of symbol grounding and creating transparent cognitive architectures for information security systems.

LINGUISTIC STRUCTURES, FINITE PREDICATE ALGEBRA, INTELLIGENCE THEORY, COMPARATOR IDENTIFICATION, ABDUCTIVE INFERENCE, INFORMATION SECURITY, INTERPRETED AI, LANGUAGE RE-AUTOMATION.

### Вступ

Сучасний етап розвитку міждисциплінарних досліджень характеризується інтеграцією когнітивних наук і лінгвістики, що зумовлює необхідність переосмислення фундаментальних механізмів обробки вербальної інформації. Вважається, що алгебра скінченних предикатів (АСП) – це універсальний математичний апарат для формального опису детермінованих, дискретних, скінченних об'єктів [1-5]. У роботах [1-3] доведено, що для того, щоб мати можливість розвивати теорію інтелекту, перш за все необхідно мати формальну мову, якою можна було б математично описувати структури і функції людського інтелекту, особливо ті, що стосуються накопичення та інтерпретації знань людиною. У якості такої мови у науковій школі професора Ю.П. Шабанова-Кушнаренка та його послідовників використовується АСП. У рамках теорії інтелекту мова розглядається не просто як інструмент комунікації, а як складна

когнітивна система, що оперує специфічними конфігураціями даних. Ключовим об'єктом аналізу в даному контексті виступають лінгвістичні структури (linguistic structures) – організовані елементи і зв'язки всередині мови, що забезпечують зберігання, передачу і генерацію смислів [6-8].

Варто зауважити, що сучасні дослідження багатомовності дедалі частіше зосереджуються на механізмах автоматизації та контролю мовленнєвої діяльності. У цьому контексті особливий інтерес становлять зовнішні прояви когнітивних процесів, одним із яких є почерк – складна психомоторна діяльність людини, що згодом переходить у високоавтоматизований режим. Це дозволяє розглядати почерк як детермінований слід автоматизованих лінгвістичних і моторних програм, чутливий до змін мовної домінанти та процесів її де- і реавтоматизації.

Питання про правомірність і межі використання терміну «лінгвістичні структури» залишається фундаментальним для сучасної науки. Незважаючи

на високий ступінь абстракції, дане поняття є загальноновизнаним науковим конструктом, що виступає базисом для таких шкіл, як структуралізм, генеративна лінгвістика Н. Хомського і когнітивізм [9]. Лінгвістичні структури охоплюють багаторівневу організацію мови: від фонетичних і морфологічних одиниць до складних синтаксичних схем, дискурсивних моделей і лінгвокультурних констант. У прикладному аспекті, особливо в завданнях моделювання інтелектуальних систем, заснованих на базах знань, та комп'ютерній лінгвістики, цей термін використовується для опису формальних алгоритмів та ієрархічних зв'язків (наприклад, синтаксичних дерев), що відображають внутрішню логіку мовної здатності людини [6-8, 10].

Для глибокого розуміння процесів обробки інформації необхідно розмежувати суміжні, але не тотожні поняття: «мовну систему», «структуру мови» і «лінгвістичні структури». В академічному дискурсі ці терміни утворюють сувору ієрархічну залежність:

1. Мовна система (language system) є найширшим поняттям – цілісним комплексом усіх підсистем і правил їх взаємодії, що має динаміку та здатність до розвитку [6-8, 11].

2. Структура мови (language structure) виступає як архітектурний каркас цієї системи, що визначає спосіб внутрішньої організації її елементів [6-8, 11].

3. Лінгвістичні структури (linguistic structures) є конкретними «будівельними блоками» і стійкими конфігураціями (схеми слів, речень, когнітивні схеми), які реалізують загальну структуру всередині живого функціонування системи [6-8, 11].

Таким чином, дослідження обробки лінгвістичних структур на основі теорії інтелекту дозволяє не тільки описувати формальну сторону мови, але й виявляти механізми того, як абстрактні ментальні категорії трансформуються в конкретні мовні одиниці. У даній статті запропоновано підхід до формалізації цих процесів за допомогою апарату алгебри скінченних предикатів, що дозволяє побудувати детерміновану модель ідентифікації смислових лінгвістичних структур. Основна увага приділяється розробці та верифікації алгоритму компараторної ідентифікації, який забезпечує розпізнавання деструктивних та маніпулятивних мовних стратегій навіть в умовах неповних або завуальованих даних. Це відкриває нові можливості для створення інтелектуальних систем із високим рівнем інтерпретованості (Explainable AI), здатних до глибокого аналізу когнітивних намірів автора тексту.

### **1. Огляд існуючих рішень з обробки лінгвістичних структур на основі теорії інтелекту**

Сучасний стан області обробки лінгвістичних структур характеризується переходом від суто емпіричних методів до спроб інтеграції глибокого

навчання з когнітивними та формально-логічними моделями інтелекту. Незважаючи на відсутність єдиної «універсальної теорії мовного інтелекту», аналіз існуючих рішень дозволяє виділити кілька ключових напрямків, що визначають поточний ландшафт досліджень.

Найбільш помітним досягненням останніх років стало використання великих мовних моделей (LLM), таких як GPT-4, Claude і LLaMA. Дослідження показують, що ці моделі здатні виконувати семантично складну бінарну класифікацію (наприклад, виявлення порушень прав людини в соціальних мережах за аналізом текстових повідомлень) з якістю, порівняною з людським експертним рівнем [12].

З точки зору оцінювання роботи нейромереж (LLM) не просто як технічної програми чи калькулятора, а як когнітивної системи, яка намагається відтворити інтелектуальні функції людини (такі як сприйняття, усвідомлення, розуміння текстових повідомлень, мовної ієрархії, логіку побудови речень, виявлення зв'язків, тощо), LLM демонструють феномен «емерджентного синтаксису» – здатність спонтанно засвоювати ієрархічні структури мови в процесі навчання на передбачення наступного токена. Однак архітектура цих рішень позбавлена прозорості та прямої інтерпретованості.

На противагу суто статистичним методам, когнітивно-семіотичні та гібридні підходи базуються на логічних і смислових принципах людського мислення, інтегруючи когнітивно-лінгвістичні алгоритми імітації механізмів засвоєння граматичних структур із семіотичним моделюванням мов зі складною морфологією, де рекурсія та еліпсис виступають фундаментом для інтелектуальних технологій представлення знань. Особливе місце в цьому ряді посідають логіко-алгебраїчні методи на основі алгебри скінченних предикатів і теорії категорій [13-15], що дозволяють формалізувати вилучення фактів і будувати прозорі детерміновані системи, здатні до логічного висновку в умовах невизначеності. Такий синтез підходів дозволяє подолати глибокий розрив між високою ефективністю нейромережових моделей у відтворенні синтаксичної ієрархії та їхньою неспроможністю забезпечити прозору інтерпретацію когнітивних намірів автора. Отже, критичний дефіцит інтерпретованості (ефект «чорної скриньки») та відсутність механізмів обґрунтованого висновку в умовах неповних даних роблять актуальним пошук альтернативних методів моделювання мовних знань, що поєднують математичну строгість із когнітивною глибиною.

### **2. Постановка завдань і цілей дослідження**

Незважаючи на істотний прогрес у галузі практичної обробки природної мови, відсутність єдиної теорії інтелектуальної обробки лінгвістичних структур

породжує низку критичних розривів між когнітивними здібностями людини та алгоритмічними рішеннями. Однією з центральних проблем залишається зведення воедино розрізнених рівнів організації мови, від морфології та синтаксису до дискурсу та прагматики, в рамках єдиної когнітивної архітектури [16]. Існуюча невизначеність у питаннях формального зіставлення людського лінгвістичного інтелекту та властивостей сучасних нейромережових моделей вимагає пошуку нових, інтерпретованих підходів до аналізу внутрішньої структури знань. Проблеми обґрунтованості символів і семантики, а також інтеграція фактичних, концептуальних і метакогнітивних рівнів знань залишаються відкритими, заважаючи створенню прозорих механізмів вилучення правил зі статистичних даних.

Метою даного дослідження є розробка сучасної моделі обробки лінгвістичних структур, яка базується на використанні методу компараторної ідентифікації теорії інтелекту та засобів алгебри скінченних предикатів для забезпечення глибокої формалізації переходу від низькорівневих ознак тексту до високорівневих когнітивних схем [17].

В рамках досягнення поставленої мети планується вирішити завдання побудови моделі обробки лінгвістичних структур у вигляді системи рівнянь алгебри скінченних предикатів, що описують класи еквівалентності текстових об'єктів на основі їх семантико-синтаксичних характеристик [15]. Особлива увага приділяється підвищенню ефективності аналізу текстів у специфічних доменах, таких як кримінальна лінгвістика, за рахунок врахування тонких структурних особливостей і механізмів операціоналізації знань [18]. Дослідження спрямоване на подолання проблеми опису слабоформалізованої інформації, шляхом впровадження процедур компараторної ідентифікації, що дозволяють відновлювати відсутні атрибути лінгвістичних об'єктів і забезпечувати обґрунтованість і пояснюваність рішень, що приймаються інтелектуальною системою в умовах невизначеності та суб'єктивності вихідних даних [17].

### 3. Формалізація лінгвістичних ознак у базисі алгебри скінченних предикатів

Для переходу від неструктурованого текстового масиву до формальної інтелектуальної моделі ми виділяємо скінченну множину об'єктів  $U$  (текстових фрагментів) і шляхом формування системи ознак та їх значень визначаємо класи еквівалентності  $K$  на  $U$ . В рамках апарату алгебри скінченних предикатів (АСП) кожен лінгвістичний об'єкт характеризується вектором ознак:

$$X = (x_1, x_2, x_3, \dots, x_n), \quad (1)$$

де  $n$  – кількість ознак, значення яких вибираються зі скінченних алфавітів.

Спираючись на методологію класифікації кримінальних текстів [18] і принципи компараторного аналізу [17], ми виділяємо три домінуючі атрибути, що формують «інтелектуальний портрет» повідомлення. Методологічною основою такого підходу є компараторний принцип, що передбачає зіставлення систем ознак. Цікавою в рамках цього підходу є також задача дослідження почерку людини при переході від однієї мови до іншої (рідної чи іноземної). Почерк у такому підході описується як система формалізованих параметрів (розмірні, просторові, варіативні ознаки), які утворюють структурно впорядкований опис письмової мови. Порівняння таких систем дозволяє виявляти ступінь подібності або розбіжності автоматизованих програм без суб'єктивної інтерпретації. Аналогічно, у даній роботі лінгвістичні ознаки (варіативність синтаксису, щільність маркерів) розглядаються як індикатори когнітивного навантаження: зростання нестабільності параметрів інтерпретується як ознака зниження автоматизації або активної перебування відповідних навичок.

#### 3.1. Визначення алфавіту ознак і предметної області

Система ознак будується на основі якісних характеристик тексту, які перетворюються в дискретні значення  $a_i^j$ . Тут перший (нижній) індекс  $i$  вказує на номер змінної, а другий (верхній) індекс  $j$  – на конкретне значення даної змінної. Така індексація дозволяє створити унікальну координатну адресу для кожного лінгвістичного значення в базі знань системи.

Атрибутивна змінна  $x_1$  – тип комунікативної інтенції. Даний параметр визначає прагматичну мету повідомлення. Множина значень:

$$M_1 = \{a_1^1, a_1^2, a_1^3\}, \quad (2)$$

де  $a_1^1$  (дескриптивний) – нейтральне оповідання про факти або звіт про події;  $a_1^2$  (індуктивний) – маніпулятивне спонукування до дії, характерне для вербування або шахрайських схем;  $a_1^3$  (агресивний) – виражена психологічна атака, пряма погроза або спроба придушення волі адресата.

Атрибутивна змінна  $x_2$  – структура суб'єктно-об'єктних відношень. Описує характер взаємодії між учасниками комунікації. Множина значень:

$$M_2 = \{a_2^1, a_2^2, a_2^3\}, \quad (3)$$

де  $a_2^1$  (персоніфікований) – жорстка зв'язка «Я – Ти», що вказує на адресність впливу;  $a_2^2$  (інституційний) – звернення від імені групи, організації або анонімної спільноти;  $a_2^3$  (невизначений) – повна анонімізація або розмитість суб'єкта дії.

Атрибутивна змінна  $x_3$  – Домінуючий лексичний маркер. Фіксує ключовий об'єкт інтересу в тексті.

Множина значень:

$$M_3 = \{a_3^1, a_3^2, a_3^3\}, \quad (4)$$

де  $a_3^1$  (ресурсний) – фокус на фінансах, персональних даних, криптоактивах або паролях;  $a_3^2$  (ідеологічний) – використання термінів боротьби, ворожнечі, помсти або політичних гасел;  $a_3^3$  (побутовий) – використання загальноживаної лексики без явних деструктивних маркерів.

Формальна коректність моделі забезпечується дотриманням закону несумісності знань у межах кожного алфавіту. Оскільки алфавіти ознак  $A_1$  (інтенція),  $A_2$  (відношення) та  $A_3$  (маркери) побудовані за принципом взаємовиключних категорій, виконується умова  $x_i^{a_i^j} \wedge x_i^{a_i^k} = \emptyset$ , що унеможливорює неоднозначність при ідентифікації. Наприклад, інтенція не може бути одночасно суто дескриптивний ( $a_1^1$ ) та агресивний ( $a_1^3$ ), оскільки наявність маркерів агресії автоматично змінює статус об'єкта в ієрархії класифікації

### 3.2. Побудова предикатної моделі та приклад операціоналізації

Згідно з теорією [14], підсумкова лінгвістична структура тексту формалізується предикатом  $P(x_1, x_2, x_3, \dots, x_n)$ . Розглянемо випадок  $n = 3$ , цільова змінна  $y$  визначає належність об'єкта до одного з трьох класів: кримінально-деструктивного ( $y_1$ ), інформаційно-маніпулятивного ( $y_2$ ) або нейтрального ( $y_3$ ), кожен клас відповідає певному рівню загрози та комунікативній стратегії, закладеній у тексті. Походження цих класів базується на експертному аналізі ознак  $x_1, x_2, x_3$ .

Клас  $y_1$  – кримінально-деструктивна структура. Формується при поєднанні ознаки  $x_1$  у значенні прямої агресії ( $x_1^{a_1^1}$ ), адресності ( $x_2^{a_2^1}$ ) і корисливого інтересу до ресурсів ( $x_3^{a_3^1}$ ). Це найбільш небезпечний тип контенту (вимагання, шантаж).

Клас  $y_2$  – інформаційно-маніпулятивна структура. Походить від поєднання прихованого спонукання ( $x_1^{a_1^2}$ ), анонімності або групового впливу ( $x_2^{a_2^2}$ ) та ідеологічного навантаження ( $x_3^{a_3^2}$ ). Це контент, спрямований на зміну переконань (пропаганда, вербування).

Клас  $y_3$  – нейтрально-інформаційна структура. Заснована на описовому викладі фактів ( $x_1^{a_1^3}$ ) в рамках офіційних або побутових відносин ( $x_2^{a_2^3}$  або  $x_3^{a_3^3}$ ) з використанням загальноживаної лексики ( $x_3^{a_3^3}$ ). Це безпечний повсякденний контент.

Приклад №1.

Розглянемо текст – «Якщо ти не надішлеш пароль від пошти протягом години, твої дані будуть опубліковані».

Аналіз інтенції – спостерігається пряма загроза («будуть опубліковані») і обмеження часу ( $x_1^{a_1^3}$  (агресивний)). Аналіз відношень – використано адресне звернення «ти» ( $x_2^{a_2^3}$  (персоніфікований)). Аналіз

маркерів – ключовими словами є «пароль» і «дані»  $x_3^{a_3^3}$  (ресурсний)).

Система класифікує даний об'єкт як кримінально-деструктивну структуру, оскільки повністю відповідає моделі шантажу:

$$P_{y_1} = x_1^{a_1^3} \wedge x_2^{a_2^1} \wedge x_3^{a_3^1}. \quad (5)$$

Приклад №2.

Розглянемо текст – «Наші прихильники вже отримують доступ до секретних архівів. Приєднуйся до боротьби за справедливість, поки не пізно!»

Аналіз інтенції – прямої загрози немає, але присутнє явне маніпулятивне спонукання і заклик до дії ( $x_1^{a_1^2}$  (індуктивний)). Аналіз відношень – автор виступає від імені групи «Наші прихильники», звертаючись до невизначеного кола осіб ( $x_2^{a_2^2}$  (інституційний)). Аналіз маркерів – текст насичений ідеологічними конструктами «боротьба», «справедливість» ( $x_3^{a_3^2}$  (ідеологічний)).

Система класифікує даний об'єкт як інформаційно-маніпулятивну структуру, оскільки не містить прямої загрози, але має ознаки когнітивного впливу:

$$P_{y_2} = x_1^{a_1^2} \wedge x_2^{a_2^2} \wedge x_3^{a_3^2}, \quad (6)$$

Приклад №3.

Розглянемо текст – «Завтра в офісі відбудеться планове засідання за результатами кварталу. Прохання підготувати звіти».

Аналіз інтенції – повідомлення має суто інформаційний характер без ознак агресії або маніпуляції ( $x_1^{a_1^1}$  (дескриптивний)). Аналіз відношень – звернення має діловий, інституційний характер ( $x_2^{a_2^1}$  (інституційний)). Аналіз маркерів – використовується стандартна ділова лексика без акценту на критичних ресурсах або ідеології ( $x_3^{a_3^1}$  (побутовий)).

Система класифікує даний об'єкт як нейтрально-інформаційну структуру. Система ігнорує його як безпечний, що не вимагає оперативного втручання:

$$P_{y_3} = x_1^{a_1^1} \wedge x_2^{a_2^1} \wedge x_3^{a_3^1}, \quad (7)$$

У разі відсутності однієї з ознак (наприклад, при розмитості суб'єкта  $x_2$ ), модель переходить до алгоритму компараторної ідентифікації для зіставлення наявних даних з еталонними кортежами бази знань.

### 4. Реалізація компараторного підходу та верифікація моделі

На відміну від традиційних методів жорсткої класифікації, компараторний підхід, що базується на теоретичних положеннях теорії інтелекту, дозволяє імітувати когнітивну здатність експерта розпізнавати прихований зміст в умовах неповноти даних. Це досягається за рахунок обчислення логічної близькості вхідного вектора до еталонних кортежів, що зберігаються в базі знань системи.

Для функціонування лінгвістичного компаратора в системі задані еталонні кортежі  $Et_{y_i}$ , які представляють собою ідеальні комбінації ознак для кожного класу (див. 3.2).

Еталон кримінально-деструктивної структури:

$$Et_{y_1} = x_1^{a_1^1} \wedge x_2^{a_2^1} \wedge x_3^{a_3^1}, \quad (8)$$

Еталон інформаційно-маніпулятивної структури:

$$Et_{y_2} = x_1^{a_1^2} \wedge x_2^{a_2^2} \wedge x_3^{a_3^2}, \quad (9)$$

Еталон нейтрально-інформаційної структури:

$$Et_{y_3} = x_1^{a_1^3} \wedge x_2^{a_2^3} \wedge x_3^{a_3^3}. \quad (10)$$

Компаратор обчислює логічну відстань між витягнутим з тексту (можливо, неповним) вектором і даними еталонами, виконуючи процедуру абдуктивного відновлення відсутніх значень.

#### 4.1. Механізм абдуктивного висновку лінгвістичного компаратора

Механізм «заповнення» пустих клітинок працює за принципом мінімізації логічної розбіжності:

1. Аналіз контекстного оточення – якщо атрибут  $x_2$  (відношень) є невизначеним, компаратор аналізує предикати  $x_1$  (інтенція) та  $x_3$  (маркер);

2. Абдуктивне припущення – у випадку, якщо інтенція визначена як «агресивна» ( $a_1^3$ ), а маркер як «ресурсний» ( $a_3^1$ ), система через механізм АСП робить логічний висновок, що найбільш імовірним (і єдино несуперечливим для класу  $y_1$ ) значенням для  $x_2$  є «персоніфікований» ( $a_2^1$ );

3. Логічна верифікація – тільки після того, як лакуна заповнена значенням, що робить предикат класу істинним  $P = 1$ , система приймає рішення про класифікацію.

Цей підхід дозволяє системі не просто ігнорувати неповні дані, а інтелектуально реконструювати прихований намір автора, що забезпечує «заземлення символів» (Symbol Grounding) у формальних логічних структурах.

#### 4.2. Верифікація моделі на розрахункових прикладах з лакунами

Нижче наведено приклади роботи лінгвістичного компаратора, де через покомпонентний аналіз тексту встановлюються значення ознак і відновлюються приховані смисли.

Приклад №1.

Об'єкт – «...слухай, якщо завтра не буде всієї суми, нарікай на себе. Я знаю, де ти живеш».

Аналіз інтенції – використання фразеологізованого виразу погрози «нарікай на себе» і створення умов психологічного тиску через згадку приватного простору «знаю, де ти живеш» вказує на ознаку «агресивний». Це маркери прямого придушення волі  $x_1^{a_1^1}$ .

Аналіз відношень – особисті займенники «я» і «ти» вказують на «персоніфікований», а імператив «слухай» свідчить про спрямований вплив конкретного суб'єкта на конкретну жертву  $x_2^{a_2^1}$ . Аналіз маркерів – ознака явно не названа (лакуна). Однак слово «сума» в семантичному полі «ультиматуму» і «термінів» однозначно вказує на фінансовий або інший матеріальний ресурс. Компаратор зіставляє загрозу і вимогу «суми», відновлюючи ознаку як (ресурсний)  $x_3^{a_3^1}$ . Результат класифікації відповідає формулі 5.

Приклад №2.

Об'єкт – «Секретні архіви вже розкриті. Приєднуйся до боротьби за справедливість, поки не пізно!»

Аналіз інтенції – дієслово в наказовому способі «приєднуйся» і маніпулятивний дедлайн «поки не пізно» – це класичні прийоми залучення до дії, що вказують на «індуктивний»  $x_1^{a_1^2}$ . Аналіз відношень – суб'єкт прихований (безлике твердження «архіви розкриті»). Лакуна відновлюється через аналіз контексту заклику: заклики до «боротьби» типові для «інституційний» впливу, де автор виступає як рупор анонімною спільно  $x_2^{a_2^2}$ . Аналіз маркерів – абстрактні ціннісні категорії «боротьба» і «справедливість», які використовуються для емоційного тригера, а не для передачі фактів, вказують на «ідеологічний»  $x_3^{a_3^2}$ . Результат класифікації відповідає формулі 6.

Приклад №3.

Об'єкт – «...зібрання в офісі, звіти за квартал потрібно підготувати до завтрашнього дня».

Аналіз інтенції – намір автора обрізаний (лакуна), але відсутність знаків оклику, погроз або закликів дозволяє компаратору припустити нейтральний «deskриптивний» на основі оточення  $x_1^{a_1^1}$ . Аналіз відношень – контекст «в офісі» і згадка «звітів», що характерно для формальної комунікації всередині організації або трудового колективу, вказують на «інституційний»  $x_2^{a_2^1}$ . Аналіз маркерів – вузькопрофесійна, але нейтральна лексика: «збори», «квартал», «підготувати», слова, позбавлені конфліктогенного потенціалу, вказують на «побутовий»  $x_3^{a_3^1}$ . Результат класифікації відповідає формулі 7.

Проведена верифікація демонструє, що лінгвістичний компаратор ефективно оперує не словами, а їх функціонально-семантичними ролями. Здатність системи відновлювати значення з лакун (наприклад, інтерпретувати «суму» як ресурсний маркер тільки в зв'язку з агресією) доводить перевагу предикатного моделювання перед простим лінгвістичним аналізом. Це дозволяє перейти від формального розпізнавання символів до інтелектуального аналізу смислових структур тексту, що знаходить своє відображення у вирішенні фундаментальних проблем штучного інтелекту.

### 4.3. Математична верифікація коректності моделі

Для підтвердження математичної коректності запропонованої моделі в базисі АСП необхідно провести перевірку дотримання фундаментальних законів теорії інтелекту: законів виключеного третього (повнота) і логічного протиріччя (несумісність). Це гарантує, що система здатна класифікувати будь-який вхідний об'єкт і при цьому виключає можливість його одночасного віднесення до взаємовиключних класів загрози.

**4.3.1. Повнота (Exhaustiveness).** Повнота моделі гарантує, що будь-який лінгвістичний об'єкт з універсуму  $U$  у межах визначених алфавітів  $M_1, M_2, M_3$  буде класифікований системою. Оскільки множини значень атрибутивних змінних  $x_1, x_2, x_3$  охоплюють весь спектр комунікативних проявів (від прямої агресії до побутового нейтралітету), то диз'юнкція всіх предикатів класів утворює тотожну одиницю, що означає, що в системі не існує «мертвих зон»: будь-який вхідний вектор або відразу відповідає еталону, або обробляється алгоритмом компараторної ідентифікації для відновлення лакуни:

$$P_{total} = P_{y_1} \vee P_{y_2} \vee P_{y_3} = 1. \quad (11)$$

**4.3.2. Виключення (Exclusivity).** Виключність гарантує відсутність суперечностей: один текст не може бути одночасно ідентифікований як два різних класи. Предикати побудовані на основі законів істинності та хибності, тобто кон'юнкція значень, що є взаємовиключними в межах одного алфавіту (наприклад,  $x_1^{a_3}$  «агресивний» та  $x_1^{a_1}$  «дескриптивний» не можуть бути істинними одночасно для одного висловлювання), забезпечує детермінованість моделі та прозорість прийнятих рішень.

$$P_{y_i} \wedge P_{y_j} = 0, i \neq j. \quad (12)$$

**4.3.3. Істинність-хибність (Validity).** Істинність підтверджується збігом результатів обчислень з експертними оцінками на контрольних вибірках. Система видає  $P = 1$  тільки тоді, коли вхідні ознаки (або відновлені лакуни) відповідають еталонній структурі класу. Будь-яка критична зміна ознаки (наприклад, заміна маніпулятивного наміру на інформаційний) миттєво змінює значення предиката на 0, що доводить чутливість та істинність моделі.

## 5. Лінгво-математичний синтез у вирішенні проблем теорії інтелекту

Запропонований метод формалізації на базі алгебри скінченних предикатів і компараторної ідентифікації дозволяє підійти до вирішення низки фундаментальних проблем, що стоять перед сучасною теорією штучного інтелекту. Перш за все, розроблений підхід пропонує ефективний шлях подолання проблеми «Symbol Grounding» (заземлення символів). На відміну від стандартних великих мовних моделей

(LLM), де токени обробляються як статистичні одиниці без прив'язки до реального значення, наше рішення забезпечує детермінований зв'язок предикатних змінних з категоріями людського досвіду. Предикат  $x_1^{a_3}$  в даній моделі перестає бути просто цифровим кодом і стає формальним описом акту агресії, що є необхідною умовою для побудови систем «сильного» ШІ, здатних до глибокого розуміння контексту.

Об'єктивним корелятом автоматизації мовленнєвих програм виступає також почерк, який може розглядатися як інструмент непрямого доступу до прихованих когнітивно-моторних станів носія мови. Запропонований підхід органічно поєднується з логіко-математичним апаратом АСП, оскільки дозволяє будувати прозорі, детерміновані моделі без втрати пояснювальної сили, що відкриває нові перспективи для дослідження динаміки мовних навичок у дорослому віці.

Важливим аспектом наукової новизни є забезпечення повної прозорості та інтерпретованості прийнятих рішень. Розроблена модель повністю виключає ефект «чорної скриньки», характерний для нейромережових архітектур. Завдяки використанню предикатних рівнянь, експерт отримує можливість верифікувати кожен логічний крок системи. Якщо об'єкт класифікується як маніпулятивний, система надає доказову базу у вигляді істинності конкретної комбінації ознак, наприклад, поєднання анонімності та ідеологічного навантаження. Така прозорість висновків має критичне значення для застосування ШІ в юридичній, слідчій та експертній практиці, де ціна помилки є надзвичайно високою.

Крім того, запропонована архітектура забезпечує безшовну інтеграцію різних рівнів знань, об'єднуючи в рамках одного математичного апарату фактографічний, концептуальний і метакогнітивний шари інформації. Використання апарату АСП дозволяє створити міст між первинною обробкою тексту на рівні лексем і високорівневим аналізом комунікативних цілей автора. Це реалізує ідеї [2] про створення єдиних інтелектуальних архітектур обробки знань, здатних до абдуктивного висновку та адаптації в умовах неповноти вихідних даних. Таким чином, теоретична значимість роботи полягає в переході від імовірнісного вгадування смислів до їх логічно обґрунтованої ідентифікації.

## Висновки

У представленому дослідженні запропоновано та обґрунтовано інноваційний підхід до обробки лінгвістичних структур на основі використання основного методу теорії інтелекту (методу компараторної ідентифікації) та математичного інструментарію алгебри скінченних предикатів. Розроблена система ознак дозволила забезпечити формалізацію якісних

характеристик тексту, перенісши суб'єктивний аналіз інтенцій та смислів у площину строгих математичних моделей. Доведено, що використання апарату АСП у поєднанні з компараторною ідентифікацією забезпечує повну пояснюваність прийнятих рішень, що є критично важливим для систем забезпечення інформаційної безпеки в специфічних доменах, таких як кримінальна лінгвістика. Впроваджений механізм відновлення лакун наділив ШІ-агента здатністю до абдуктивного висновку, дозволяючи розпізнавати приховані деструктивні структури навіть в умовах високої невизначеності або неповноти вихідних даних. Отримані результати створюють міцний теоретичний фундамент для проектування гібридних інтелектуальних систем нового покоління, де гнучкість нейромережевого навчання поєднується з жорсткою логічною інтерпретованістю експертних знань [5].

Подальший розвиток дослідження передбачає уточнення системи ознак з метою переходу від описового аналізу до абдуктивного висновку щодо процесів автоматизації та реавтоматизації мови та мовлення, враховуючи не тільки опис та обробку різноманітних лінгвістичних структур, але й формалізацію типів мовної поведінки носіїв мови при переході з однієї мови на інші, що передбачає не лише аналіз статичних структур, а й формалізацію динамічних параметрів мовної біографії суб'єкта в умовах міжмовної інтерференції.

#### Список літератури:

- [1] Бондаренко М. Ф., Шабанов-Кушнарченко Ю. П., Шаронова Н. В. Інструментарій компараторної ідентифікації. Біоніка інтелекту. 2010. №2(73).
- [2] Бондаренко М. Ф., Шабанов-Кушнарченко Ю. П., Шаронова Н. В. Ситуаційно-текстовий предикат. Біоніка інтелекту. 2010. №2(73). С. 87-98.
- [3] Бондаренко М. Ф., Шабанов-Кушнарченко Ю. П., Шаронова Н. В. Булева структура тексту. Біоніка інтелекту. 2010. №2(73). С. 99-110.
- [4] Хайрова Н. Ф., Шаронова Н. В. Інформаційно-лінгвістичні технології екстракції і ідентифікації глибинних знань в текстах: монографія. Х.: ФЛП Коряк С.Ф., 2016. 205 с.
- [5] Плехова Г. А., Шаронова Н. В. Авторський курс з дисципліни «Методи наукових досліджень» / Г. А. Плехова, Н. В. Шаронова. – Харків: ФОВ Бровін О.В. 2025.–84 с. ISBN 978-617-8238-97-1
- [6] Cholan, V., & Ponomarova, V. (2021). LINGUOCULTURAL CONSTANTS OF SLAVONIC TYPICAL TEXTS. Scientific Journal of Polonia University, 43(6), 16–22. <https://doi.org/10.23856/4302>
- [7] Oleksenko, V. P. (2021). STRUCTURAL AND SEMANTIC FEATURES OF THE NEOLEXES ON THE DESIGNATION OF THE CULTURAL AND ARTISTIC SPHERE. Opera in linguistica ukrainiana, (28), 149–168.
- [8] Bondarchuk, C. S., & Chumachenko, O. A. (2021). FEATURES OF WRITTEN LANGUAGE COMMUNICATION BY MEANS OF THE UKRAINIAN LANGUAGE (on the example of official-business, epistolary and conversational-functional styles). Alfred Nobel University Journal of Philology, 1(21).
- [9] Juanda, J. (2024). Analysis of Language Structure and Its Implications in Modern Linguistics: A Study of the Understanding and Application of Structural Linguistic Concepts. Journal of Educational and Social Research, 14(1), 226. <https://doi.org/10.36941/jesr-2024-0019>
- [10] Martin, A. E. (2020). A Compositional Neural Architecture for Language. Journal of Cognitive Neuroscience, 32(8), 1407–1427.
- [11] Boyarova, L. (2021). LINGUISTIC GLOBALIZATION: PROBLEMS OF LANGUAGE SOVEREIGNTY. In Priority areas for development of scientific research: domestic and foreign experience. Publishing House “Baltija Publishing”. <https://doi.org/10.30525/978-9934-26-049-0-23>
- [12] Smyrnova-Trybulska, E., Morze, N., & Varchenko-Trotsenko, L. (2022). Adaptive learning in university students' opinions: Cross-border research. Education and Information Technologies. <https://doi.org/10.1007/s10639-021-10830-7>
- [13] Каратаєв, О. А. (2026). Моделі та методи розроблення програмного забезпечення систем штучного інтелекту на основі інтерпретації знань: дис. ... канд. техн. Наук : галузь знань 12 : спец. 121 Інженерія програмного забезпечення. — Харків, 2026.
- [14] Дорошенко А. Ю. Інформаційна технологія інтелектуального аналізу фактографічних текстових ресурсів [Електронний ресурс] : дис. ... канд. техн. наук : спец. 05.13.06 : галузь знань 12 / Анастасія Юріївна Дорошенко ; наук. керівник Шаронова Н. В. ; Нац. техн. ун-т «Харків. політехн. ін-т». — Харків, 2018. — 177 с. — Бібліогр.: с. 149-167. — укр.
- [15] Karataiev, O., Sitnikov, D., & Sharonova, N. (2023). A Method for Investigating Links between Discrete Data Features in Knowledge Bases in the Form of Predicate Equations. In COLINS (1) (pp. 224-235).
- [16] Hahn, M., Futrell, R., Levy, R., & Gibson, E. (2022). A resource-rational model of human processing of recursive linguistic structure. Proceedings of the National Academy of Sciences, 119(43).
- [17] Cherednichenko, O., Vovk, M., Sharonova, N., & Vorzhevitina, A. (2025). Comparator-based identification of food edibility from natural language description.
- [18] Khairova, N., Kupriianov, Y., Vorzhevitina, A., & Shanidze, O. (2024). Models for effective categorization and classification of texts into specific thematic groups (using gender and criminal themes as examples). In Computational Linguistics Workshop at CoLInS 2024. CoLInS. <https://doi.org/10.31110/colins/2024-4/004>

Received (Надійшла) 11.01.2026

Accepted for publication (Прийнята до друку) 01.02.2026

Publication date (Дата публікації) 27.03.2026