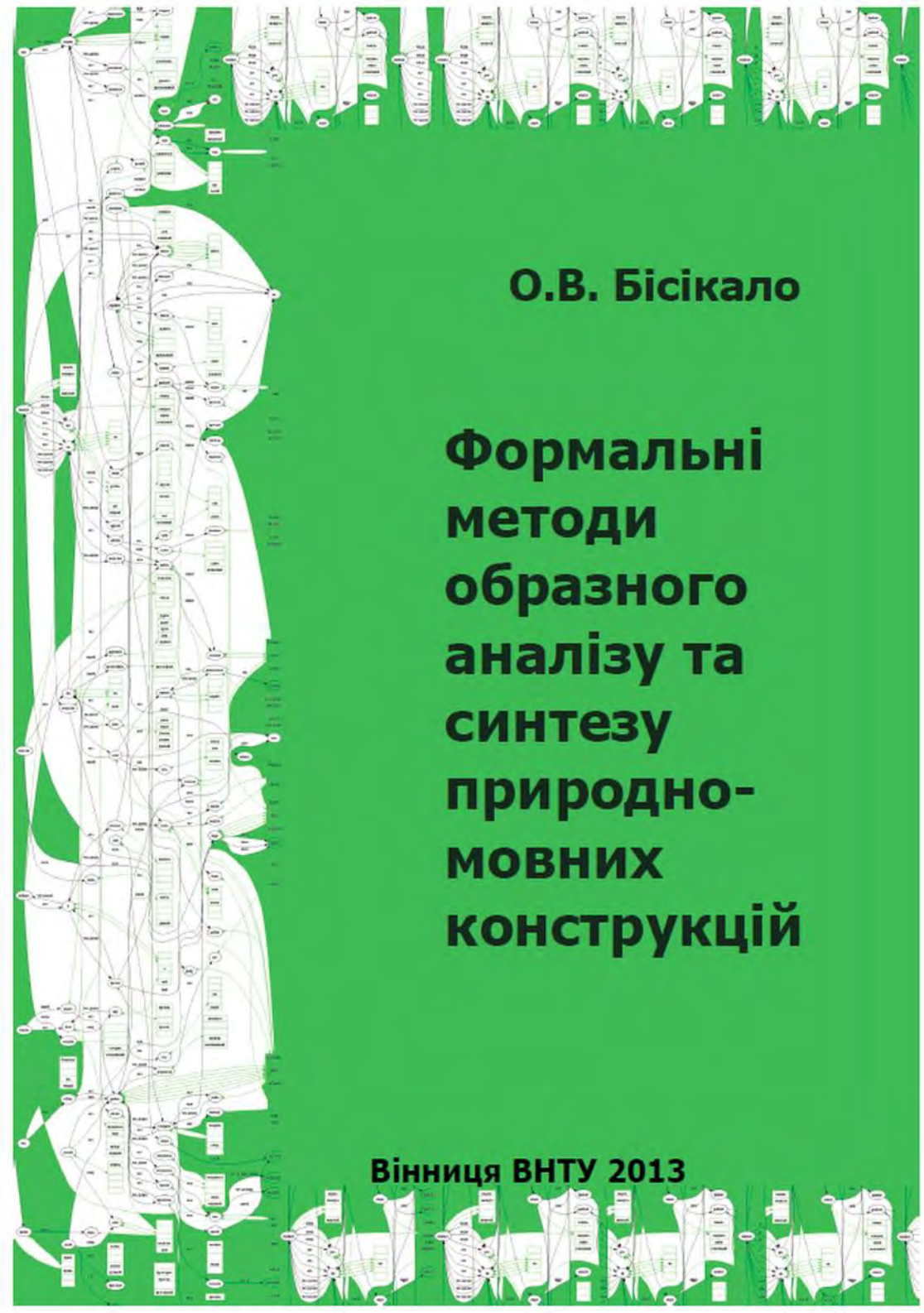

О.В. Бісікало

Формальні методи образного аналізу та синтезу природно-мовних конструкцій

Вінниця ВНТУ 2013

Міністерство освіти і науки України
Вінницький національний технічний університет

О. В. Бісікало

ФОРМАЛЬНІ МЕТОДИ ОБРАЗНОГО АНАЛІЗУ ТА СИНТЕЗУ ПРИРОДНО-МОВНИХ КОНСТРУКЦІЙ

Монографія

Вінниця
ВНТУ
2013

УДК 004.93:159.95
ББК 32.97
Б65

Рекомендовано до друку Вченою радою Вінницького національного технічного університету Міністерства освіти і науки України (протокол № 7 від 6 березня 2013 р.)

Рецензенти:

В. А. Широков, доктор технічних наук, професор
А. М. Пєтух, доктор технічних наук, професор

Бісікало, О. В.

Б65 Формальні методи образного аналізу та синтезу природномовних конструкцій : монографія / О. В. Бісікало. – Вінниця : ВНТУ, 2013. – 316 с.

ISBN 978-966-641-528-1

В монографії розглянуто теоретичні основи образного аналізу текстової інформації у відповідності до ідеї формалізації поняття образного сенсу через визначення його властивості та параметра. Запропоновано методи синтезу структурно-функціональних моделей системи образної обробки природномовного контенту. У межах функцій єдиної онтогенетичної системи з властивістю до самовдосконалення бази загальних знань образного сенсу отримано корисні моделі на рівні алгебраїчних операцій з мовними образами. Розроблено інформаційну технологію образного аналізу та синтезу природно-мовних конструкцій, що дозволило отримати нові розв'язки семантико-залежних задач.

УДК 004.93:159.95
ББК 32.97

ISBN 978-966-641-528-1

ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ І ПОЗНАЧЕНЬ

АВМ – асоціативно-вербальна мережа

АМО – асоціативна мережа образів

АО – ансамбль образів

БАС – булева алгебра сенсу

ВЕ – вектор емоцій

ЕК – електронний контент

ІС – інфологічна система

КМО – конструкція мовних образів

МО – мовний образ

ОК – образна конструкція

ПМК – природно-мовна конструкція

ППЗ – програмно-педагогічний засіб

СОПМК – система обробки природно-мовного контенту

Концепти мовних образів:

I (Image) – образ

N (Notion) – поняття

O (Object) – об'єкт

M (Method) – метод

Q – якість

ON – поняття об'єкта

QN – поняття якості

MN – поняття метода

OQ (Object-Quality) – якість об'єкта

MQ (Method-Quality) – якість методу

H – власне обставина (відповідь на питання як?)

T – обставина часу (відповідь на питання коли?)

L – обставина місця (відповідь на питання де?)

Focus – мовний образ у фокусі уваги

Типи асоціативних зв'язків:

P_x – асоціативний зв'язок синтагматичного або невідомого походження

$P_в$ – внутрішньообразна асоціація

$P_ч$ – асоціація типу окреме–загальне (гіпо–гіперонімія)

$P_ц$ – асоціація типу частина–ціле (меронімія)

$P_с$ – синонімічна асоціація

$P_о$ – омонімічна асоціація (окремий випадок паронімії)

$P_а$ – антонімічна асоціація

$P_р$ – асоціація за типом «у риму»

Операції формальної теорії:

$\oplus$ – операція об'єднання образних конструкцій «PLUS ОК».

$\backslash$ – зв'язок типу «головний–підлеглий» в асоціативній парі МО.

$\times$ – зв'язок типу «підмет–присудок» в асоціативній парі МО.

Інші формальні операції:

$\cup$ – об'єднання множин.

$\cap$ – перетин множин.

$\neg$ – доповнення множини.

$\vee$ – логічна диз'юнкція.

$\wedge$ – логічна кон'юнкція.

ПЕРЕДМОВА

Безпрецедентний розвиток новітніх інформаційних технологій та розбудова всесвітньої мережі Інтернет привели до появи загального інформаційного простору планетарного масштабу, що має ознаки абсолютно нового соціально-технічного утворення. Вражаючі можливості прозорого обігу інформації не тільки розмивають кордони між державами й скорочують відстань між людьми, але й відкривають шляхи до переходу людства у суспільство знань. Окреслюються нові перспективні підходи до підтримки вільного спілкування, швидкого доступу до інформації та неперервного навчання людини упродовж всього її життя. Щойно усталене поняття електронного контенту набуває помітного соціального й економічного значення.

Вибух технологічних досягнень на перетині тисячоліть став можливим завдяки основоположним фундаментальним працям Н. Вінера, К. Шеннона, Ф. де Соссюра, Б. де Куртене, А. Тьюринга, Д. Маккарті, А. М. Колмогорова, В. М. Глушкова, Д. А. Поспєлова, Т. А. Гаврилової, Т. Бернерс-Лі тощо. Проте можливості сучасних технологічних засобів оброблення контенту не відповідають вимогам міжнародних стандартів до його семантичних властивостей. Всупереч очікуванням Інтернет-спільноти, які постійно зростають, проблемним залишається розв'язання класу семантико-залежних задач обробки природномовної інформації, що потребують залучення експертних знань для оцінки отриманих розв'язків. Моделі та методи найбільш відомих підходів – лінгвістичного, статистичного та логічного – відчутно програють інтелектуальним можливостям людини-експерта, що виразно відображається в семантико-залежних задачах пошуку, перекладу, анотування тощо. Використання надвеликих обчислювальних ресурсів для вилучення знань з неструктурованих масивів інформації не може забезпечити самовдосконалення загальної бази знань.

Значний вклад у розвиток моделей та методів автоматизованого управління інформаційними системами, аналізу природної мови та когнітивних процесів людини внесли вітчизняні дослідники О. В. Палагін, В. А. Широков, М. І. Шлезінгер, А. В. Анісімов, Ю. П. Шабанов-Кушнаренко, С. Л. Кривий, М. Ф. Бондаренко, Ю. Р. Валькман, Н. В. Шаронова, В. П. Широчин та ін. Проте відчутна парадоксальність проблеми розв'язання семантико-залежних задач вимагає засто-

сування нових підходів і методів. Значна частина моделей лінгвістичної та когнітивної семантики будується на основі знань кваліфікованих експертів, але існуючі формальні засоби не досягають того рівня розуміння тексту, що демонструє звичайна дитина, яка ще не навчилася цей текст читати.

Недостатньо враховується в сучасних підходах до семантичного аналізу текстової інформації природний шлях отримання людиною більшості знань про навколишній світ за рахунок феноменів образного мислення. Відомі інтроспективні спроби логічного узагальнення інформаційних ознак психічних процесів людини, як правило, не відзначаються належним науковим обґрунтуванням. Теоретичного та експериментального дослідження потребує підтверджена результатами наук когнітивного напряму гіпотеза про витоки прагматичних і семантичних аспектів сенсу з понять образу, асоціації, потреб і емоцій [31]. Проте відсутність чіткої та зрозумілої картини процесів, що відбуваються у головному мозку людини, змушує дослідників користуватися моделями непрямої аналогії, які забезпечують лише окремі результати для окремих задач, які прийнято відносити до інтелектуальної діяльності.

Структурно-функціональний підхід до образного аналізу та синтезу природно-мовних конструкцій, що пропонується, також є спробою отримати нову модель непрямої аналогії для розв'язання широкого кола семантико-залежних задач обробки природно-мовного контенту. Саме така мета дослідження змусила автора зайвий раз прискіпливо проаналізувати відповідні результати наук когнітивного блоку – філософії, фізіології вищої нервової діяльності, психології, нейропсихології та лінгвістики. Насамперед для того, щоб отримані формальні конструкції не суперечили загальновизнаним на міждисциплінарному рівні положенням щодо образних підвалин розуміння сенсу природномовної інформації людиною.

Монографія, матеріал якої базується на роботах [1–75], складається з п'яти розділів. У першому з них окрім вже згаданого аналізу мультидисциплінарних основ моделювання когнітивної сфери людини для обґрунтування структурно-функціонального підходу проведено аналіз відомих методів моделювання процесів обробки природномовної інформації. Визначені проблема, мета і задачі дослідження, а також його провідна ідея – отримати розв'язки актуальних семантико-

залежних задач у вигляді комплексу функцій інформаційної системи обробки природно-мовного контенту.

У другому розділі розглянуто формалізацію образного аналізу текстової інформації на основі понять інфологічної системи та онтогенетичного принципу, змістовного поєднання образних та природномовних концептів. За допомогою формальної теорії визначено поняття образної конструкції, отримано її кількісні оцінки на основі нечіткої міри та одиниці образного сенсу з урахуванням понять ентропії та кількості інформації. Обґрунтовано методологію синтезу структурно-функціональних моделей інфологічної системи шляхом кібернетичної інтерпретації процесів інтелектуальної діяльності людини.

У третьому та четвертому розділах у розвиток запропонованого підходу розроблено методи моделювання процесів образної обробки природно-мовного контенту та синтезовано базові функції відповідної інформаційної системи, що забезпечують розв'язання семантико-залежних задач на основі образного пошуку та самовдосконалення бази знань. З цією метою застосовано математичний апарат теорій множин, графів, алгебраїчних систем і алгоритмів.

У п'ятому розділі розглянуто практичну реалізацію отриманої інформаційної технології та розв'язання актуальних семантико-залежних задач, визначених вимогами до інфологічної системи. За допомогою розробленого програмного забезпечення проведено експериментальні дослідження, що демонструють переваги отриманих та впроваджених розв'язків у порівнянні з існуючими аналогами.

Автор буде дуже вдячний за відгуки на цю книгу, які можна надсилати за адресою: кафедра АІВТ, ВНТУ, Хмельницьке шосе, 95, м. Вінниця, Україна, 21021 або на E-mail: obisikalo@gmail.com.

РОЗДІЛ 1
АНАЛІЗ МЕТОДІВ МОДЕЛЮВАННЯ ПРОЦЕСІВ ОБРОБКИ ПРИРОДНО-МОВНОЇ ІНФОРМАЦІЇ ТА ОБҐРУНТУВАННЯ СТРУКТУРНО-ФУНКЦІОНАЛЬНОГО ПІДХОДУ

На основі аналізу відомих методів моделювання процесів обробки природно-мовної інформації та загальновизнаних результатів наук когнітивного напряму визначаються проблема, мета і задачі дослідження. Обґрунтовуються характерні особливості структурно-функціонального підходу, що пропонується, та провідна ідея дослідження – отримати розв'язки актуальних семантико-залежних задач у вигляді комплексу функцій інформаційної системи обробки природно-мовного контенту (СОПМК).

1.1 Аналіз існуючих методів моделювання процесів обробки природно-мовної інформації та побудови баз знань

Актуальним завданням більшості задач штучного інтелекту та інтелектуальних інформаційних технологій є визначення семантичних характеристик процесів, що моделюються. Проте, не дивлячись на значні зусилля дослідників, семантичний аналіз як природно-мовних конструкцій (ПМК), так і інших продуктів інтелектуальної діяльності людини лишається найбільш проблемним і до цього часу. Корінь цієї надзвичайно складної проблеми насамперед пов'язаний з важкою формалізованістю предметної області внаслідок відсутності науково обґрунтованого спільного погляду на семантику в науках, які відносять до когнітивного напряму досліджень [31].

Дослідження в галузі оброблення ПМК проводяться вітчизняними науковцями Українського мовно-інформаційного фонду НАН України [76], Інституту кібернетики та Інституту проблем математичних машин і систем, Міжнародного науково-навчального центру інформаційних технологій і систем, Національного університету ім. Т. Шевченка, Національного технічного університету України «Київський політехнічний інститут», Національного університету «Львівська політехніка», Харківського національного університету радіоелектроніки, Донецького інституту проблем штучного інтелекту тощо.

Серед провідних закордонних наукових закладів, що приділяють велику увагу проблемі розпізнавання природної мови, можна виділити й такі відомі, як Інститут проблем інформатики РАН, РосНДІ штучного інтелекту, Санкт-Петербурзький державний політехнічний університет, Ульяновський державний технічний університет, Massachusetts Institute of Technologies, Carnegie Mellon University, Stanford University, Princeton University, компанії IBM, Xerox, Sun, Microsoft, Google та багато інших.

Проблема розуміння сенсу природної мови тісно пов'язана з когнітивним напрямом штучного інтелекту та моделюванням пізнавальної діяльності людини. Побудова експертних систем, у тому числі інтелектуальних електронних підручників, базується на використанні уніфікованих структур бази знань у вигляді семантичних мереж, фреймів, онтологій, матриць семантичних ознак, реляційних моделей тощо. Використання сучасних баз знань все ще обмежено вузькими предметними областями, оскільки ускладнюється питаннями «вилучення знань» з експерта і відсутністю в системі загальних представлень «здорового глузду» [77]. В той же час мовленнєва діяльність людини природним чином забезпечує навчання та отримання нових знань на основі вербальної інформації у вигляді тексту [78].

Сучасний електронний контент, семантика якого лежить у площині проблеми дослідження, на даний час, безумовно, став мультимедійним. До складу контенту входять не тільки текст і окремі ПМК, але й, згідно зі стандартами W3C, графічні, аудіо-, відео- та анімаційні об'єкти різних типів (форматів). ПМК потрібно вважати провідною складовою контенту хоча б тому, що в метадані всіх об'єктів записуються слова, словосполучення або речення. Отже, інтегральним поняттям для аналізу семантики залишається образ – мовний, графічний, музичний, анімаційний тощо.

Починаючи з моменту свого зародження у штучному інтелекті історично склалися два підходи до моделювання інтелектуальних процесів людини. Так, на основі досягнень фізіології і генетики виник так званий висхідний або м'який напрям (нейронні мережі, методи розпізнавання образів і генетичні алгоритми) [79]. Дослідники низхідного або жорсткого напряму в штучному інтелекті запропонували математичну інтерпретацію усвідомлених процесів абстрактного мислення (алгебра логіки та нечітка логіка, продукційні моделі, семантичні ме-

режі, фрейми, онтології, об'єктно-орієнтоване програмування, формальні граматики у лінгвістиці, евристичні моделі тощо) [80]. Дещо осторонь, як на наш погляд, розташовані в другому напрямі інфологічне моделювання та нормування відношень у базах даних, а також *Data mining* – пошук семантичної інформації у великих масивах даних. Окрім цього, порівняно молодим і ще до кінця несформованим напрямом досліджень можна вважати моделювання образного мислення людини [81].

Розглянемо існуючі підходи до моделювання семантики в окреслених вище напрямах та формальних методах. Якщо генетичні алгоритми частіше використовуються для розв'язання оптимізаційних задач, то розпізнавання образів за допомогою штучних нейронних систем найбільш тісно пов'язане з первинним семантичним аналізом навколишньої інформації. З математичної точки зору коло задач розпізнавання образів перетинається з задачами кластеризації, статистичної ідентифікації, побудови аксонометрій, факторним аналізом тощо [82]. У нейронних мережах традиційно моделюються принципи асоціативного сприйняття інформації людиною, проте, як правило, поняття «образ» застосовується у вузькому обмеженому сенсі, наприклад, візуальних об'єктів. Такий підхід не дозволяє поєднати основи низхідного та висхідного напрямів штучного інтелекту з точки зору когнітивної семантики.

Предметна область більшості розробок в галузі інженерії знань, як вже було зазначено раніше, має безпосереднє відношення до парадигматичної будови мови. До цього напряму досліджень належать логічні та евристичні моделі представлення знань [81]. У групу логічних моделей включають числення предикатів, псевдофізичні та багатомодальні логіки, реляційні моделі, а евристичними вважаються мережеві, фреймові і сценарні моделі. До останньої групи відносять також такі популярні останнім часом напрями, як онтології та об'єктно-орієнтоване програмування (ООП) [80,83].

Терміном «семантична мережа», що має безпосереднє відношення до предмету дослідження, позначається множина представлень, побудованих, як правило, на графах. Такі представлення відрізняються, головним чином, іменами вузлів, зв'язків та висновками, які можна робити в таких структурах [84]. З самого початку семантичні мережі моделювали поняття, в основному, абстрактного характеру та відношен-

ня між ними. Цей формалізм є візуально виразним і дозволяє представити будь-який вид знань, проте збільшення типів відношень викликає необхідність програмування зростаючої у геометричній прогресії кількості фактів та правил. Тому в сучасних розробках мереж намагаються досягти стандартизації відношень шляхом вибору мінімального набору семантичних примітивів.

Сценарії [85] та фрейми [86] являють собою оригінальні евристичні моделі, що мають явні витоки з результатів когнітивної семантики. Обравши за основу психологічні дослідження процесів мислення та пам'яті людини, означені моделі дають непогані результати розуміння семантики у вузьких предметних областях, проте поняття сенсу в них має чітко виражений функціональний характер. Варто відмітити, що для всіх трьох розглянутих напрямів евристичного моделювання на сьогоднішній день немає прикладів успішного застосування у випадках розуміння довільного тексту природної мови.

Одним з найбільш відомих результатів мінімізації семантичної мережі можна вважати парадигму об'єктно-орієнтованого програмування (ООП), що передбачає використання таких концептів, як об'єкт, метод та якість. Дуже важливе практичне значення з точки зору автоматизації та візуалізації програмування має використання трьох головних принципів ООП: наслідування, інкапсуляції та поліморфізму [87]. Ці принципи можна вважати такими, що відображають особливості парадигматичного мислення людини. Але важливу також роль грає абстрактне поняття, що узагальнює інші концепти, а якість властива як для об'єктів, так і для методів. Об'єднавчим концептом для поняття, об'єкта, методу та якості (об'єкта і методу) можна вважати образ.

Виявлення семантики у великих масивах інформації від самого початку супроводжує розвиток популярних в програмних технологіях реляційних моделей даних. З цією метою в практику створення та експлуатації СУБД разом з реляційною алгеброю та реляційним численням було впроваджено теорію нормалізації відношень, що застосовується на етапі інфологічного моделювання предметної області бази даних [88]. Проте в цьому випадку знаходять формальне представлення лише знання щодо структури сутностей предметної області та відношень між ними [89]. Зрештою всі інші семантичні властивості, у

т. ч. статистичні характеристики наявних зв'язків потребують додаткового моделювання засобами реляційної алгебри [90, 91].

Проблема розуміння природних мов як одна з ключових проблем штучного інтелекту завжди була рушійною силою досліджень в області представлення знань [92, 93, 94]. На відміну від формалізації синтаксису та морфології природно-мовних конструкцій, моделювання семантики у комп'ютерній лінгвістиці має значно скромніші досягнення. Ця ситуація є наслідком того, що основними моделями цього напряму досліджень є ті ж самі семантичні мережі у найбільш складному своєму варіанті [95], оскільки природна мова вільна від будь-яких обмежень, що можуть накладатися на штучні системи на зразок [96]. Окрім цього, існуючі системи орієнтуються на статистичний аналіз ключових слів тексту, а не асоціативних зв'язків між ними.

Розв'язання проблеми явно потребує нових підходів, що моделюють природний процес поступового накопичення знань про навколишній світ [97, 98]. На відміну від традиційних лінгвістичних методів сучасні інформаційні технології, які підтримуються такими велетнями комп'ютерної індустрії як Google, започаткували розвиток альтернативного статистичного підходу до аналізу ПМК [99]. Отримані результати розв'язку задач комп'ютерної лінгвістики, наприклад, у перекладі наближаються до існуючих досягнень, проте, ще далекі до рівня людини-експерта [77].

Принципово суб'єктивний характер будь-якого мовного висловлювання спирається на семантичні засоби лексики, що відображається у його змісті (значенні). З іншого боку, висловлювання завжди спрямоване на досягнення певної внутрішньої мети (мотиву, потреби) суб'єкта, тобто має прагматичну характеристику у вигляді сенсу цього акту мовлення. І хоча семантика та прагматика мовного висловлювання тісно пов'язані між собою, історична традиція математичної лінгвістики віддає безумовну перевагу першій складовій [100, 101]. Чи не вперше поняття змісту з'явилося в роботах Московської семантичної школи, де закладено витоки моделі «Зміст–текст». Під змістом висловлювання розумілося послідовне накопичування лінгвістичних відношень для ланцюга морфологія–синтаксис–семантика в умовах інтегрального опису словарної та граматичної компонент мови [102, 103]. Проте формальне визначення змісту і на сьогодні значно поступається природному практично в усіх семантико-залежних задачах

комп'ютерної лінгвістики. Це яскраво ілюструється парадоксом розуміння «смысла речи» неосвіченими людьми, які зовсім не тямлять у лінгвістиці, але чудово орієнтуються в бажаннях співрозмовника.

Розглянемо термінологію лінгвістичної семантики згідно з [104]. Основними поняттями, що характеризують предмет цієї дисципліни є зміст (*содержание*), значення (*значение*) та смисл/сенс (*смысл*). Оскільки предмет дослідження охоплює як семантику, так і прагматику природно-мовних виразів, то будемо притримуватися в роботі терміну сенс для перекладу російського *смысл*. Окрім цього, покажемо відмінності понять смисл та сенс.

Для предметної області «СЛОВО» [109] поняття смислу пов'язано з екстенсіоналом слова – множиною вказівників на ті сутності, які позначаються цим словом та асоціаціями і конотаціями – асоційованими у свідомості того, хто застосовує слово представленнями як фактичного, так оцінювального характеру. В асоціативному розумінні смисл отримує всю гаму визначень, що характеризують відчуття та емоції суб'єкта, викликані усвідомленням асоціативно пов'язаної з словом інформації [104].

На відміну від смислу поняття значення закріплює за певною одиницею мови відносно стабільний у часі стійкий зміст, інваріантний для всіх носіїв мови. Отже, значення *X*-а – це інформація, пов'язана з *X*-ом конвенційно, згідно з загальноприйнятими правилами використання *X*-а як засобу передачі інформації. В той же час смисл *X*-а для *Y-а* – це інформація, пов'язана з *X*-ом у свідомості *Y-а* в період часу *T*, коли *Y* застосовує або сприймає *X* як засіб передачі інформації [104]. Виходячи з цього та задач дослідження будемо вважати, що сенс *X*-а для *Y-а* – це інформація, пов'язана з *X*-ом у когнітивній сфері *Y-а* (не тільки у свідомості, але й на рівні рефлексів та підсвідомості) в період часу *T*, коли *Y* застосовує або сприймає *X* як засіб передачі інформації.

Прийняті визначення через екстралінгвістичні знання розширюють кордони предмета дослідження з «чистої» лінгвістики до близьких наук когнітивного напряму – філософії, психології, фізіології, педагогіки та прагматики. Такий підхід добре ілюструється цитатою з [104]: «…численні напрямки сучасної семантики можна звести до двох концепцій, що протистоять одна одній, існування яких об'єктивно обумовлене двоїстістю подвійністю) предмета семантики. Ці дві концепції семантики можна умовно назвати вузькою і широ-

кою. Вузька концепція семантики робить своїм предметом значення одиниць мови і побудованих із них мовних виразів. В широкій концепції семантики її предметом, крім того, є і сенс мовних виразів в конкретних умовах їхнього вживання.».

Проблема дослідження також тісно пов'язана з теорією інформації та спорідненими з нею теоріями ймовірностей і нечітких множин. Цінність інформації за Харкевичем пропорційна збільшенню імовірності досягнення системою мети [105], проте цей показник відображає, насамперед, прагматичну сторону повідомлення і лише опосередковано враховує семантичну. Функція належності в нечіткій логіці акумулює суб'єктивне бачення експерта відносно належності певного елемента до деякого класу елементів [106, 107], але будується на описовому принципі і не супроводжується конструктивним алгоритмом свого визначення поза участю експертів. Окрім цього, відомі нечіткі міри [108] характеризують саме ступінь належності, можливості, достатності, необхідності тощо, але при цьому лише опосередковано відображають той зростаючий обсяг знань, який накопичується у людини і, зрештою, дає підстави зробити той чи інший логічний висновок. З іншого боку, існуючі технології побудови словників на основі лексикографічної системи вводять поняття псевдотопології, за рахунок якого встановлюється ступінь близькості слів та/або лексем, а також обґрунтовують необхідність побудови специфічної формальної метамови для опису лексико-семантичних відношень природно-мовних конструкцій [109].

Формальною структурою для оцінки змісту речення на природній мові може служити штучна нейронна мережа, при цьому нейрон застосовується як математична основа кванта змісту, що масштабується для символу, частини слова, слова, словосполучення, речення, абзацу, всього тексту [110, 111]. Такий підхід передбачає, що зміст тексту або іншого мовного поняття закладено у предикат, що приймає значення «істина» чи «хибність», проте окрім верифікації повідомлень залишаються відкритими питання синтезу, наприклад, генерація адекватного відгуку системи на зовнішні впливи.

В роботі [112] запропоновано використати для вивчення природної мови таку комп'ютерну модель «дитини», в яку не закладено лінгвістичні знання. Опосередковано ця ідея використовується і в статистичному підході до аналізу ПМК [99], хоча варто зазначити, що схожі думки висловлювалися А. Тьюрингом ще на початку становлення

штучного інтелекту як наукового напряму [226]. Опубліковані підходи до теоретико-модельної формалізації процесів мислення та рефлексії, де враховуються поняття свідомості, представлення, образу, онтології, але лишається поза розглядом онтогенез психічних функцій людини [113]. В літературі висуваються ідеї моделювати комп'ютерну особистість шляхом використання спеціальних математичних методів для генерації нових знань та реалізації цілеспрямованої поведінки [114, 115]. Проте багато авторів відзначають, що існуючі підходи до моделювання операцій образного мислення мають поки що скоріше концептуальний, ніж практичний характер [116].

Не можна не відмітити, що навіть за наявного теоретично недостатнього рівня формалізації процесів семантичного аналізу вже існують приклади масштабних та амбітних технологічних проектів. Актуальність цього напряму досліджень підтверджують не тільки академічні розробки [117], але й фінансово та ресурсно вражаючі зусилля лідерів комп'ютерної індустрії IBM, ABBYY, GOOGLE [118, 119, 120].

Отже, загальний аналіз існуючих підходів до моделювання процесів обробки природно-мовної інформації та побудови баз знань дозволяє зробити такі висновки:

1. Не дивлячись на надзвичайну актуальність семантичного аналізу, в технологіях обробки тексту відсутні концептуальні підходи та відповідні теорії, за головну мету яких покладено визначення сенсу як прагматичної першооснови змісту.
2. Відсутнє поняття одиниці сенсу, не розроблено критерії для кількісної оцінки сенсу в ПМК та інших продуктах когнітивної діяльності людини.
3. Моделювання образного мислення на відміну від логічних та евристичних моделей має поки що, переважно, концептуальний характер і не доведено до практичної реалізації у комп'ютерній лінгвістиці.
4. Ідею Тьюринга про онтогенетичний характер моделювання продуктів інтелектуальної діяльності практично не розвинуто у низхідному (жорсткому) напряму штучного інтелекту.
5. Відомі формальні методи не забезпечують прийнятного рівня розв'язку широкого кола актуальних семантико-залежних задач, а тому потребують розробки нові комплексні підходи та методи, що враховують недоліки існуючих.

1.2 Мультидисциплінарні основи моделювання когнітивної сфери людини

Дослідження когнітивної сфери людини мають глибокі історичні корені та пов'язані з видатними діячами науки. Ще давньогрецький вчений Платон, зацікавлений феноменологічним характером процесів мислення, розглядав асоціацію як основу людської пам'яті і поведінки. Його співвітчизника Аристотеля вважають автором найпершої класифікації асоціацій за схожістю (червоне – пурпурне або кішка – тигр), за часовою послідовністю (день – ніч) та за контрастом (велике – маленьке або холодне – гаряче), яка стала основою для наступних численних класифікацій і типологій [121].

Протягом віків окремими філософськими системами у поняття асоціації вкладався різний зміст. Це поняття по-різному трактували та вивчали відомі дослідники. Так, Р. Декарт використовував асоціацію для розуміння процесів оволодіння власними пристрастями, а Б. Спіноза пояснював асоціаціями певні особливості «руху думок». Т. Гоббс, у свою чергу, створив першу систему механістичної психології, де елементи свідомості (відчуття та представлення) взаємодіють на основі механістичних за своєю суттю зв'язків за суміжністю відчуттів у просторі та часі [122].

Власне сам термін «асоціація» було введено в науковий світ в XVII сторіччі Дж. Локком з метою пояснення причин виникнення забобонів та «помилкових ідей». Якщо Дж. Берклі пояснював за допомогою асоціацій сприйняття простору, то у Д. Юма асоціація стає наріжним каменем всієї пізнавальної сфери психіки. Юм також розрізняв три типи переходів від однієї ідеї до іншої – подібність, суміжність у просторі (шия – голова) та часі, причинно-наслідковий зв'язок.

XVIII та перша половина XIX століть вважаються періодом класичного асоціанізму, який супроводжувався низкою гучних імен Д. Гартлі, Д. Прістлі, Джеймса Мілля і Т. Брауна. З другої половини XIX сторіччя предмет дослідження цікавив Джона Стюарта Мілля, А. Бена, Г. Спенсера, Г. Еббінгауза, представників англійської школи В. Вундта, Т. Цігена, Г. Мюллера [123, 124]. Протягом перших десятиліть XX сторіччя гостра криза асоціанізму привела до його остаточного зникнення як цілісного напряму психології та асиміляції його ідей в різних галузях психологічної теорії та практики. Важкий час

кризи сприяв появі нових гучних імен і напрямів в психології – Зиґмунда Фрейда, Карла Г. Юнга, Альфреда Адлера (психоаналіз), німецької школи гештальтпсихології, американської школи біхевіоризму, Ж. Піаже та Л. С. Виготського, російської школи фізіології І. П. Павлова та багатьох інших [122, 125–128].

Сучасний період розвитку вільного асоціативного експерименту пов'язаний з виникненням психолінгвістики та зміщенням фокуса досліджень на проблеми мовленнєвої діяльності людини і формування її мовної здатності. Визначення асоціативного значення слова ввів у сучасну наукову парадигму Дж. Діз [129], а Ч. Осгуд застосовував метод шкал для вимірювання смислових полів і показав, що афективне значення слова представляє собою координати в багатовимірному просторі [130]. За образним виразом Х. Хермана, «значення не є асоціація, але знання асоціації» [131]. Радянська школа психолінгвістики була створена зусиллями вітчизняних дослідників А. А. Залевської, І. Г. Овчиннікової, Н. О. Золотової, Ю. М. Караулова [132–136].

До наук когнітивного спрямування належить філософія. Практично всі філософи погоджуються з тим, що психічне – якісно своєрідна форма буття [137–142]. Життєздатність суб'єкта забезпечується властивостями живої системи використовувати носіїв інформації, що входять до складу суб'єкта, про стани зовнішнього і внутрішнього середовища при регуляції поведінки. Тоді психіку можна вважати саме такою властивістю суб'єкта. Аналогічної точки зору дотримувався російський математик О. Ляпунов, який стверджував, що життя – це стійкий стан речовини, який використовує для вироблення реакцій самозбереження інформацію, що кодується складом елементів цієї речовини [140].

Відомо, що людина відрізняється від тварин наявністю мови як системи кодів, що позначають предмети і такі відносини, за допомогою яких предмети вводяться у певні системи і категорії. Ця система кодів веде до формування абстрактного логічного мислення і формування «категоріальної» свідомості [137]. Здатність людини переходити за межі наочного, безпосереднього досвіду до відвернутого, раціонального досвіду є фундаментальною особливістю її свідомості. Інтроспективно або самоспостереженням суб'єкт принципово може визначити такі психологічні феномени, як: відчуття, сприйняття образу,

Шановний читачу!

Умови придбання надрукованих примірників монографії наведені на сайті видавництва http://publish.vntu.edu.ua/get/?isbn=978-966-641-528-1

Уважаемый читатель!

Условия приобретения печатных экземпляров монографии приведены на сайте издательства http://publish.vntu.edu.ua/get/?isbn=978-966-641-528-1

Dear reader!

You may order this monograph at the Web page
http://publish.vntu.edu.ua/get/?isbn=978-966-641-528-1

Наукове видання

Бісікало Олег Володимирович

ФОРМАЛЬНІ МЕТОДИ ОБРАЗНОГО АНАЛІЗУ ТА СИНТЕЗУ ПРИРОДНО-МОВНИХ КОНСТРУКЦІЙ

Монографія

Редактор С. А. Малішевська
Оригінал-макет підготовлено О. В. Бісікало

Підписано до друку 29.05.2013 р.
Формат 29,7×42¼. Папір офсетний.
Гарнітура Times New Roman.
Друк різографічний. Ум. др. арк. 18,25
Наклад 300 (1-й запуск 1–75) Зам № 06-04

Вінницький національний технічний університет,
КІВЦ ВНТУ,
21021, м. Вінниця, Хмельницьке шосе, 95,
ВНТУ, ГНК, к. 114.
Тел. (0432) 59-85-32.
Свідоцтво суб'єкта видавничої справи
серія ДК № 3516 від 01.07.2009 р.

Віддруковано ФОП Барановська Т. П.
21021, м. Вінниця, вул. Порика, 7.
Свідоцтво суб'єкта видавничої справи
серія ДК № 4377 від 31.07.2012 р.