

МІНІСТЕРСТВО КУЛЬТУРИ УКРАЇНИ  
ХАРКІВСЬКИЙ ДЕРЖАВНИЙ ІНСТИТУТ КУЛЬТУРИ

На правах рукопису

ШАРОНОВА Наталія Валеріївна

КОМПАРАТОРНА ІДЕНТИФІКАЦІЯ  
ЛІНГВІСТИЧНИХ ОБ'ЄКТІВ

Спеціальність 05.25.05 – інформаційні  
системи і процеси

А В Т О Р Е Ф Е Р А Т  
дисертації на здобуття наукового  
ступеня доктора технічних наук

*Шаронова*

Харків - 1994



00778379 (+)

Дисертацією є рукопис.

Робота виконана на кафедрі Проєкту  
Харківського державного технічногоНауковий консультант – Заслужений діяч науки і техніки України,  
доктор технічних наук, професор  
ШАБАНОВ-КУШНАРЕНКО Ю.П.

Офіційні опоненти:

1. Член-кореспондент АН України,  
доктор фізико-математичних наук,  
професор ЮЦЕНКО Катерина Логвинівна
2. Доктор технічних наук, професор  
СЕНЧЕНКО Микола Іванович.
3. Доктор технічних наук, професор  
РОСЬ Анатолій Олександрович.

Провідна установа – Київський університет імені Тараса Шевченка  
Міністерства освіти України, м.КиївЗахист відбудеться "18" листопада 1994 р. о 13 годині  
у конференцзалі Харківського державного інституту культури  
на засіданні спеціалізованої вченої ради Д 02.021.02  
за адресою: ЗІ0001, Харків, Бурсацький узвіз, 4.З дисертацією можна ознайомитися у бібліотеці Харківського  
державного інституту культури за адресою: ЗІ0001, Харків,  
Бурсацький узвіз, 4.Автороферат розіслано "10" листопада 1994 р.Вчений секретар  
спеціалізованої вченої ради,  
доктор техн. наук

Є.В.Боднарський

ЛНБ ім. В. Стефаника  
АН України

Актуальність теми. Розробка і використання нових інформаційних технологій в умовах всезростаючого потоку інформації та безупинного ускладнення керованих об'єктів немислимі зараз без включення до них елементів штучного інтелекту, забезпечення їх системами, які розуміють природну мову та реалізують функції інтелектуальної діяльності людини. Надія на вирішення основних проблем інформатики за рахунок нарощування та удосконалення технічних засобів, яка існувала раніше, поступово змінилась розумінням значущості семантичних методів, розробка яких вимагає серйозних досліджень в галузі формалізації інтелектуальних функцій людини, перш за все таких як розуміння, усвідомлення й сприйняття.

На шляху широкого розповсюдження нових інформаційних технологій та обчислювальної техніки взагалі стоїть проблема створення інтелектуального інтерфейсу, який забезпечує для користувача максимально комфортні умови спілкування з комп'ютером. Використання природної мови (ПМ) в інформаційних системах набуває надзвичайного значення в світлі сучасного уявлення про інтелектуальні прикладні програми, у широкому сенсі цього слову, як людинно-машинні системи, засновані на базах знань (БЗ), одним з компонентів яких є знання про ПМ: всілякі словники, граматики, правила побудови семантичних конструкцій і т.і. Основний внесок в розробку теоретичних та прикладних питань автоматизації обробки ПМ як універсального засобу опису дійсності й комунікації з обчислювальною системою зробили вчені Глушков В.М., Ляпунов О.О., Піотровський Р.Г., Перевозчикова О.Л., Поспелов Д.О., Попов Е.В., Виноград Т., Філлмор К., Хомський Н., Шенк Р., Ющенко К.Л. та багато ін.

Незважаючи на досягнені серйозні успіхи в теорії й практиці інформаційних систем (ІС), вони, проте, не задовольняють повністю вимог широкого кола користувачів через те, що, точно кажучи, не забезпечують йому дійсно дружнього інтерфейсу та глибокого смислового пророблення інформації, яка вимагає моделювання здібностей до міркування, розуміння, мислення. Наука, яка вивчає механізми природного інтелекту з метою використання набутих знань для створення систем штучного інтелекту (ШІ), розробляється декілька останніх десятиріч науковою школою професора Ю.П.Шабанова-Кушнарєнка й називається теорією інтелекту.

Однією з основоположних властивостей інтелектуальної діяльності є здатність людини порівнювати одні відчуття з іншими, сприйняття, думки, колірні відчуття, емоції й т.і., відрізняючи їх одне від одного, або ототожнюючи. Здатність людини порівнювати між собою суб'єктивні відчуття і установлювати їх рівність чи нерівність є глибинною якістю особи, яка забезпечує її єдність. Це в повній мірі стосується здібності людини володіти мовою, сприймати та обробляти текстову інформацію. Установлення збіжності чи різниці будь-яких двох суб'єктивних станів людини дає реальну можливість переходу від словесних описів до фізично вірогідного об'єктивного математичного опису інтелектуальних функцій випробуваного /далі іспитника/ під час проведення психофізичного експерименту, що становить суть методу компараторної ідентифікації.

Метою дисертаційної роботи є дослідження, розробка й розвиток методу компараторної ідентифікації в застосуванні до лінгвістичних об'єктів різного рівня на основі текстів російської і української мов як бази знань про природу мову та їх реалізація у вигляді прикладних програмно-інформаційних систем автоматизованої обробки текстової інформації.

Для досягнення поставленої мети необхідно вирішити такі завдання:

1. Дослідити особливості лінгвістичних об'єктів як об'єкту моделювання й сформулювати основні принципи їх ідентифікації.

2. Розробити інструментальні засоби компараторної ідентифікації, які необхідні для моделювання лінгвістичних відносин. Розробити математичні засоби опису ситуаційно-текстового предикату, його декомпозиції, опису функцій сприйняття ситуації та розуміння тексту.

3. Побудувати математичні моделі семантичних відносин морфологічних й деривативних утворень, дослідити структуру міжморфемних семантичних відносин.

4. Розробити систему одержання попередньої оцінки труднощів машинного перекладу (МП), використовуючи вимір послівної несхожості паралельних текстів. Запропонувати методику побудови системи морфологічного МП з російської мови на українську й навпаки.

5. Сформулювати й дослідити найбільш важливі задачі автоматизованої обробки текстів російської та української мов, розробити оптимальні методи й алгоритми вирішення цих задач і реалізувати їх у вигляді комплексу автоматизованих інформаційних систем.

Наукова новизна та основні положення, які виносяться на захист. В роботі обґрунтовано та вперше сформульовано концепцію нового наукового напрямку в галузі комп'ютерної лінгвістики і розробки інтелектуальних систем, об'єктами дослідження якого є розуміння й відтворення лінгвістичних конструкцій різного рівня, основний метод математичного опису – метод компараторної ідентифікації. Особливістю даного підходу є те, що будь-які

відомості про мовну поведінку людини ґрунтуються тільки на фізичному, тобто суто об'єктивному, дослідженні поведінки іспитника під час проведення лінгвістичного психофізичного експерименту.

Розроблено інструментальні засоби компараторної ідентифікації: операції над предикатами, алгебри предикатних операцій, досліджено властивості предиката еквівалентності, пророблено подання загального виду бінарних предикатів, запропоновано кванторну алгебру скінчених предикатів довільного порядку. Введення ситуаційно-текстового предикату та обґрунтування об'єктивності підходу до опису семантики ПМ дозволило зробити математичний опис фізично спостережених функцій сприйняття ситуації й розуміння тексту. Вперше досліджено попередні оцінки труднощів машинного перекладу, запропоновано методику виміру послівної несхожості паралельних текстів.

#### На захист виносяться:

- концепція застосування методу компараторної ідентифікації при вирішенні задач обробки текстової інформації й дослідження мовної поведінки людини;
- класифікація й узагальнення лексико-граматичних і семантичних відносин російської та української мов на морфемному й міжморфемному рівнях з метою побудови математичних моделей морфемних відносин;
- інструментарій компараторної ідентифікації: алгебри предикатних операцій, підстановочних операцій, кванторна алгебра довільного порядку, засоби представлення загального виду бінарних предикатів, декомпозиції ситуаційно-текстового предикату;
- принципи, оптимальні методи й алгоритми вирішення ряду важливих завдань автоматичної переробки тексту: машинний пере-

клад із попередньою оцінкою труднощів, виявлення та корекція помилок, аналіз і синтез слівформ і т.і.;

- комплекс програмних засобів вирішення логічних рівнянь, програмні засоби корекції помилок у текстах, програмний комплекс російсько-українського морфологічного перекладу.

Особистий внесок. Дисертантові належить обґрунтування й формулювання нового наукового напрямку в галузі комп'ютерної лінгвістики, дослідження мовної поведінки людини на основі методу компараторної ідентифікації, розробка математичних моделей мовних відносин різного рівня та впровадження розроблених моделей при вирішенні важливих завдань народного господарства, які пов'язані зі створенням інформаційного, математичного, програмного та лінгвістичного забезпечення автоматизованих інформаційних і інтелектуальних систем.

Методи дослідження. В роботі використані методи теорії інформації, інженерної лінгвістики, теорії множин, теорії графів, оптимізації, структур даних, подання знань, дискретної математики, алгебри логіки та системного програмування.

Вірогідність і обґрунтованість теоретичних положень, висновків і отриманих результатів визначаються відповідністю та узгодженістю з результатами інших учених: філологів, психологів, психофізиків, математиків, - в рамках проведених психофізичних експериментів при дослідженні лінгвістичної поведінки людини.

Практична новизна та реалізація результатів роботи. На основі теорії інтелекту, концепції та методів, запропонованих і обґрунтованих в роботі, визначені перспективні напрями в галузі комп'ютерної переробки текстової інформації машинних систем подання знань про мову та різних систем автоматичної обробки знань. Розроблено програмні комплекси АССАЗ /Автоматизована

система смислового аналізу зображень/, комп'ютерного перекладу з російської мови на українську й навпаки, корекції помилок у текстах, система формування ділових документів українською мовою, інструментально-програмний комплекс для розв'язання рівнянь алгебри предикатних відносин.

Перелічені системи впроваджені та використовуються для автоматизації наукових досліджень і навчання мові, для формування ділової документації в Харківському технічному університеті радіоелектроніки, Одеському інституті народного господарства, Київському державному економічному університеті, в Управлінні по соціально-економічному розвитку м.Одеси, в НВО "Хар-трон" м.Харкова, у впровадницькій науково-виробничій фірмі "Прогресивні технології", м.Харків. Отримані в дисертації результати використовуються в учбовому процесі ХТУРЕ в курсі "Логічний аналіз" і в Харківському державному університеті в курсі "Теорія інтелекту".

Результати, викладені в дисертації, одержано в рамках досліджень і розробок, здійснюваних на кафедрі Програмного забезпечення ЕОМ факультету Обчислювальної техніки Харківського державного технічного університету радіоелектроніки відповідно до плану науково-дослідних робіт: держбюджетні НДР "Розробка теорії інтелекту та створення на її основі програмно-технічного забезпечення ЕОМ нових поколінь" /№ ДР ІІ02І362/, НДР "Розробка автоматизованої системи смислової обробки текстової інформації для розпізнавання образів і аналізу зображень" /1991-1993 рр./, НДР "Розробка методів ідентифікації процесів інтелектуальної діяльності та їх застосування для автоматизації інформаційних процесів" /Державна програма, п.6 "Інформатика, автоматизація, приладобудування", підрозділ 6.2.2. Перспективні інформаційні технології та системи, постанова Державного Комітету з питань

науки та технологій України від 04.05.92 р., № І2/, а також за госпдоговірними НДР із НВО "Хартрон", ІНВФ "Прогресивні технології", м.Харків.

Апробація роботи. Основні положення і результати дисертаційної роботи були представлені й розглянуті на: Другому міжнародному семінарі "Теорія та застосування штучного інтелекту", Болгарія, Созопол, 1989 р.; XI Всесоюзній нараді "Проблеми управління 89", Ташкент, 1989 р.; IX симпозиумі "Ефективність, якість і надійність систем "людина - техніка", Воронеж, 1990; IV Всесоюзній конференції "Математичні методи розпізнавання образів", Рига, 1989; Міжнародній конференції з нових інформаційних технологій у навчанні, Київ, 1993; Міжнародній конференції Схід-Захід з нових інформаційних технологій в освіті, Москва, 1992 р.; I Міжнародній конференції з інформаційних технологій і систем, Львів, 1993; Першій Всеукраїнській конференції "Обробка сигналів і зображень та розпізнавання образів", Київ, 1992, а також на Всесоюзній школі-семінарі "Психологічна біоніка" /Харків, 1986, 1988/ і "Біоніка інтелекту" /Харків, 1987, 1989, 1991/, на інших республіканських і зональних форумах.

Публікації. За темою дисертації опубліковано 52 праці, з них монографій - І, статей - 20, депонованих рукописів - І2, тез доповідей - І9.

Структура роботи. Дисертаційна робота складається з вступу, 7 розділів, закінчення, списку літератури з І37 назв і додатків. Загальний обсяг роботи становить 240 сторінок.

## ЗМІСТ РОБОТИ

У вступові обґрунтовано актуальність досліджуваної проблеми, сформульовано мету і завдання роботи, наведено основні

положення, які виносяться на захист, показано наукову новизну та практична цінність отриманих результатів.

У першому розділі досліджено основні підходи до вирішення задач ідентифікації лінгвістичних об'єктів, визначено клас задач автоматизованої обробки текстової інформації, проведений аналіз їх розв'язання, досліджено найбільш важливі особливості природної мови як об'єкту моделювання. Викладено концепцію ідентифікації лінгвістичних об'єктів на основі методу компараторної ідентифікації.

Показано, що центральною проблемою ШІ є побудова формальних моделей інтелектуальної діяльності людини, де одним з найважливіших напрямків є дослідження її мовної поведінки. Аналіз літературних даних дозволяє зробити висновок, що при побудові комп'ютерних лінгвістичних моделей розробники здебільшого орієнтуються на індуктивні й індуктивно-дедуктивні методи, засновані на ретельному та цілеспрямованому вивченні статистичної й інформаційно-сислової структури реальних текстів ПМ, що складає суть відтворюючого моделювання лінгвістичних об'єктів і процесів.

Аналіз стану досліджень в галузі моделювання семантики показав, що ключова проблема в галузі моделювання ПМ і використання її в машинних системах – це розробка методів, які б дозволили машині "розуміти" смисл тексту. Це завдання дуже важке тому, що смисл тексту не є простим об'єднанням смислів, які можна витягти з окремих частин тексту, оскільки, зокрема, між реченнями, віддаленими в тексті невизначено далеко, можуть існувати зв'язки, які змінюють їх ізольований смисл. Основні труднощі, які виникають при машинному аналізі смислу окремої фрази, значною мірою полягають в тому, що разом із словами програмі треба подати велику кількість інформації про їх зна-

чення. Трудність полягає не тільки в ретельному формальному описі значень слів, але й в умінні обробляти цю інформацію.

Огляд різних підходів до побудови систем семантичного аналізу привів до висновку, що здебільшого аналіз в них засновується на особливостях текстів, які розглядаються, і в той же час використовується деяка кількість інформації про зовнішню дійсність, що міститься в семантичних словниках і таблицях. Останні, як правило, будуються шляхом визначення деякого універсального семантичного алфавіту, набору елементарних об'єктів і пов'язуючих їх відносин так, що значення мовних одиниць можна подати за допомогою абстрактних відносин між комбінаціями об'єктів. На цей час інтенсивні роботи в галузі розуміння текстів продовжуються. Перспективними є різні схеми для пояснення семантики ПМ, які включають інтенціональну логіку, засновані на теорії ситуаційної семантики, на моделюванні зв'язувального дискурсу та діалогу, особливо на прагматичних проблемах, таких як структура події, фокус, анафора та багато ін.

Проте слід визнати, що реалізовані за останні десять років проекти продемонстрували неадекватність моделей ПМ, обмежених в основному структурою ізольованого речення. Виявилось, що для побудови ефективних семантичних систем потрібне більш тонке вивчення суті розуміння. Крім того, фундаментальною й недостатньо повно дослідженою проблемою, безпосередньо пов'язаною з розробкою знань про ПМ та до вирішення якої до цього часу немає єдиного системного підходу, є проблема організації лінгвістичного експерименту. Основною трудностю тут є одержання об'єктивної інформації про суб'єктивні відчуття людини при дослідженні її інтелектуальної діяльності, зокрема, при дослідженні її мовної поведінки.

У першому розділі сформульовано основні завдання дослідження, які передбачають дослідження особливостей лінгвістичних об'єктів з метою їх ідентифікації методом компарації дедуктивним способом, виходячи виключно з фізично спостережасмих фактів. Для цього ставляться завдання розробки математичних засобів опису ситуаційно-текстового предикату, його декомпозиції, опису функцій сприйняття ситуації й розуміння тексту.

Другий розділ присвячено розробці проблемно-орієнтованого інструментарію компараторної ідентифікації. Метод порівняння, або метод компараторної ідентифікації, є основним у теорії інтелекту та дає можливість викладати основні положення теорії інтелекту дедуктивним способом, виходячи з фізично спостережасмих фактів. Суть методу полягає в тому, що іспитник у спеціально поставлених експериментах своїми відповідями формує значення деяких предикатів  $P_1, P_2, \dots, P_2$ . У цих експериментах виявляються властивості предикатів  $P_1, P_2, \dots, P_2$ , які формально записуються у вигляді логічних рівнянь, зв'язуючих предикатні змінні  $X_1, X_2, \dots, X_2$ . Деякі з цих рівнянь використовуються як аксіоми або вихідні постулати теорії інтелекту. З аксіом, як із рівнянь, знаходяться значення предикатних змінних  $X_1, X_2, \dots, X_2$ , які є відповідно предикатами  $P_1, P_2, \dots, P_2$ . Внутрішня структура знайдених предикатів відповідає тим чи іншим механізмам інтелекту людини.

При використанні методу компараторної ідентифікації дослідник впливає на органи чуттів іспитника фізичними сигналами /стимулами/  $X_1, X_2, \dots, X_n$ , які викликають в його свідомості певні суб'єктивні становища  $Y_1, Y_2, \dots, Y_n$ . Якщо припустити, що становища  $Y_1, Y_2, \dots, Y_n$  однозначно залежать від відповідних їм стимулів  $X_1, X_2, \dots, X_n$ , то існують функції  $Y_1 = f_1(X_1)$ ,  $Y_2 = f_2(X_2)$ ,  $\dots$ ,  $Y_n = f_n(X_n)$ . В експериментах стимули вибираються

з визначених дослідником множин  $A_1, A_2, \dots, A_n$ , так що завжди  $x_1 \in A_1, x_2 \in A_2, \dots, x_n \in A_n$ . Множини  $A_i$  ( $i \in \{1, 2, \dots, n\}$ ) відповідають завданням, які ставить перед собою дослідник. Передбачається, що кожен зі стимулів  $x_i \in A_i$  викликає досить певний стан  $y_i$ . Множина всіх значень функції  $y_i = f_i(x_i)$ , яка задана на множині  $A_i$ , утворює множину  $B_i$ . Таким чином, кожна з функцій  $f_i$  уявляє собою сюр'єкцію, яка відображує  $A_i$  на множину  $B_i$ . Функції  $f_i$  характеризують здатність іспитника реагувати на зовнішні предмети відповідними до них суб'єктивними становищами.

У ході лінгвістичного експерименту дослідник формулює іспитникові завдання, яке той повинен виконати. У завданні показується деяке відношення  $L$ , яке зв'язує становища  $y_1, y_2, \dots, y_n$ , виникаючі в свідомості іспитника внаслідок пред'явлення йому відповідних стимулів  $x_1, x_2, \dots, x_n$ . Якщо для цих становищ відношення  $L$  виконується, то іспитник повинен відповісти  $\xi = 1$ , якщо не виконується, то  $\xi = 0$ . Під час виконання даного завдання іспитник реалізує предикат  $\xi = L(y_1, y_2, \dots, y_n)$ , який відповідає відношенню  $L$ . Предикат  $\xi$  характеризує дію механізму свідомості іспитника, порівнюючого стани:

$y_1, y_2, \dots, y_n$  відповідно до одержаного завдання.

Предикат  $P(x_1, x_2, \dots, x_n) = L(f_1(x_1), f_2(x_2), \dots, f_n(x_n))$  характеризує фізично спостережену лінгвістичну поведінку іспитника, що виконує завдання дослідника та реагує на стимули  $x_1, x_2, \dots, x_n$  відповідно  $\xi = P(x_1, x_2, \dots, x_n)$ . Задача моделювання того чи іншого процесу полягає в тому, що з властивостей предикату  $P$ , виявлених в експериментах з іспитником, дістати внутрішню структуру сигналів  $x_1, x_2, \dots, x_n$ ;  $y_1, y_2, \dots, y_n$ ; вид функцій  $f_1, f_2, \dots, f_n$  та вид предикату  $L$ . Ця задача допускає узагальнення на випадок  $\mathcal{E}$  предикатів  $P_1, P_2,$

...,  $P_2$ . У загальному випадку при одержанні іспитником запитань він їх виконує по черзі для різних наборів вхідних сигналів.

Знайдені закономірності лінгвістичної поведінки іспитника записуються у вигляді системи логічних рівнянь

$$\begin{aligned} R_1(X_1, X_2, \dots, X_n) &= I, \\ R_2(X_1, X_2, \dots, X_n) &= I, \\ &\dots\dots\dots, \\ R_e(X_1, X_2, \dots, X_n) &= I, \end{aligned} \quad (I)$$

які зв'язують між собою предикатні змінні  $X_1, X_2, \dots, X_n$ . Тут  $R_1, R_2, \dots, R_e$  - предикати від предикатів  $X_1, X_2, \dots, X_n$ . Предикат  $X_j$  ( $X_1, X_2, \dots, X_n$ ) ( $j \in \{1, 2, \dots, n\}$ ) заданий на декартовому добутку  $A_{1j} \times A_{2j} \times \dots \times A_{nj}$ . Мається на увазі, що рішення  $X_1 = P_1, X_2 = P_2, \dots, X_n = P_n$  задовольняє системі рівнянь (I).

Значення аргументів  $X_1, X_2, \dots, X_n$  предикатів  $P_1, P_2, \dots, P_n$  у експериментах виступають спочатку як абстрактні елементи, внутрішня структура яких невідома, вона може бути одержана дедуктивними прийомами з умов (I). З цих же рівнянь з'являється внутрішня структура предикатів  $P_1, P_2, \dots, P_n$ , яка складається із внутрішньої структури сигналів  $Y_{1j}, Y_{2j}, \dots, Y_{nj}$ , функцій  $f_{1j}, f_{2j}, \dots, f_{nj}$  і предикату  $L_j$  для кожного з предикатів  $P_j$  ( $X_1, X_2, \dots, X_n$ ) =  $L_j(f_{1j}(X_1), f_{2j}(X_2), \dots, f_{nj}(X_n)) = L_j(Y_{1j}, Y_{2j}, \dots, Y_{nj})$ .

Задачі у такій постановці неможливо розв'язувати без достатньо розвинутої математичної мови. Насамперед, необхідні формалізми для запису предикатів, які реалізує іспитник у ході експерименту. Далі, треба мати мову для запису рівнянь, які описують властивості цих предикатів. Крім того, необхідно мати формальні засоби для опису внутрішньої структури стимулів, що

пред'являються іспитникові, і станів, що викликаються ними, а також внутрішньої структури предикатів, які реалізує іспитник. Нарешті, необхідно мати математичні засоби одержання із властивостей предикатів їх внутрішньої структури.

Для математичного вираження предикатів, які реалізуються іспитником у ході експерименту, в роботі використовується спеціальна формальна мова – алгебра предикатних операцій. Мовою алгебри предикатів можна записати будь-яке скінченне відношення і будь-яку скінченну функцію. Для введення алгебри предикатів використовується алгебра логіки, яка є різновидом бульової алгебри. Предикатом  $P$ , заданим на  $U^n$ , називається будь-яка функція  $\xi = P(x_1, x_2, \dots, x_n)$ , яка відображує множину  $U^n$  у множину  $\Sigma = \{0, 1\}$ . Змінні  $x_1, x_2, \dots, x_n$  при цьому називаються предметними, а їх значення – предметами. Множина усіх  $n$ -арних предикатів, заданих на  $U^n$ , на якій визначені операції диз'юнкції, кон'юнкції та заперечення предикатів, називається алгеброю  $n$ -арних предикатів на  $U$ . При цьому операції диз'юнкції, кон'юнкції та заперечення є базисними для алгебри предикатів. Базисними предикатами для алгебри предикатів є предикати виду

$$x_i^a = \begin{cases} 1, & \text{якщо } x_i = a, \\ 0, & \text{якщо } x_i \neq a, \end{cases} \text{ де } i \in \{1, 2, \dots, n\}, \quad (2)$$

що називаються предикатами пізнання предмету  $a$  за змінною  $x_i$ . Операція пізнання предикату  $A$  за змінною  $P_j$  ( $j \in \{1, 2, \dots, t\}$ ) визначається таким чином:

$$P_j^A = \begin{cases} 1, & \text{якщо } P_j = A, \\ 0, & \text{якщо } P_j \neq A, \end{cases} \quad (3)$$

а будь-яку  $t$ -місцеву операцію  $F$  над предикатами  $P_1, P_2, \dots, P_t$  можна так виразити у вигляді суперпозиції операцій, які уже уведені:

$$F(P_1, P_2, \dots, P_t) = \bigvee_{A_1, A_2, \dots, A_t \subseteq U^n} B(A_1, A_2, \dots, A_t) P_1^{A_1} P_2^{A_2} \dots P_t^{A_t}, \quad (4)$$

де  $B(A_1, A_2, \dots, A_t) = F(A_1, A_2, \dots, A_t)$  – фіксований предикат, який являє собою значення операції  $F$  на наборі предикатів  $A_1, A_2, \dots, A_t$ . Таким чином, система операцій, яка складається з диз'юнкції, кон'юнкції, пізнавань всіляких предикатів за змінними  $P_1, P_2, \dots, P_t$  і усіх константних операцій, повна. Пізнавання кожного предикату  $A$  за будь-якою з предикатних змінних  $P_j$  ( $j \in 1, 2, \dots, t$ ) виражається у вигляді ось такої суперпозиції кванторів спільності, операції рівнозначності предикатів і фіксованого предикату  $P_j$ :

$$P_j^A = \forall x_1 \forall x_2 \dots \forall x_n (P_j(x_1, x_2, \dots, x_n) \sim A(x_1, x_2, \dots, x_n)). \quad (5)$$

Багато уваги приділено використанню кванторів спільності та існування для предметних і предикатних змінних, введення яких у роботу є корисним фактором у арсеналі виразних засобів математичного апарату, що використовується.

Алгебри предикатних операцій уводяться в роботу для опису властивостей предикатів, які реалізує іспитник у ході експерименту. Властивості предикатів, з яких виводиться загальний вид перетворення  $f$  сигналу в об'єкт дослідження, називаються аксіомами предикату. Система аксіом – це визначальні властивості предикату. Якщо зі системи аксіом можна вивести загальний вид предикату, вона є повною. Система аксіом предикату є нескоротною, якщо жодну з них неможливо вивести з решти аксіом системи. Сукупність тверджень про предикат, які можна вивести з аксіом предикату, створюють теорію даного предикату.

Нехай  $\mathcal{T}$  – теорія,  $U$  – універсум предметів цієї теорії,  $x_1, x_2, \dots, x_m$  – предметні змінні теорії  $\mathcal{T}$ ,  $\mathcal{M}$  – множина усіх предикатів  $P(x_1, x_2, \dots, x_m)$ , визначених на просторі  $U^m$ . Таким чином,  $\mathcal{M}$  – це універсум предикатів. Уведемо на

$\mathcal{M}$  предикатні змінні  $X_1, X_2, \dots, X_n$  теорії  $\mathcal{T}$ , значеннями яких є предикати, визначені на  $U^m$ . Множина  $\mathcal{M}^n$  – це предикатний простір,  $U^m$  – предметний простір. Елементами множини  $\mathcal{M}^n$  є предикатні вектори /набори/, елементами  $U^n$  – предметні вектори. Будь-яка функція  $Y = \mathcal{F}(X_1, X_2, \dots, X_n)$ , яка відображує множину  $\mathcal{M}^n$  у множину  $\mathcal{M}$ , називається предикатною операцією. Якщо створити множину  $Q$  усіх предикатних операцій, то можна увести алгебру предикатних операцій над  $Q$ .

У роботі доведено ствердження, що бульова алгебра предикатних операцій з базисом елементів, який складається з усяких констант і усяких операцій пізнавання предикату, повна. Алгебра, базис операцій якої складається з диз'юнкції та кон'юнкції, а базис елементів утворений з усяких операцій пізнавань предикату, є повною, і називається диз'юнктивно-кон'юнктивною алгеброю предикатних операцій. Доведено теорему про те, що будь-яка предикатна операція  $\mathcal{F}$  може бути виражена досконалою диз'юнктивною нормальною формою (ДНФ) за формулою

$$\mathcal{F}(X_1, X_2, \dots, X_n) = \bigvee \mathcal{F}(P_1, P_2, \dots, P_n) X_1^{P_1} X_2^{P_2} \dots X_n^{P_n}. \quad (6)$$

$$(P_1, P_2, \dots, P_n) \in \mathcal{M}^n$$

Доведено теоремами про диз'юнктивно-кон'юнктивну алгебру предикатних операцій, досліджено їх властивості, перелічено її основні тотожності.

У третьому розділі продовжено розробку математичних засобів компараторної ідентифікації. Перший параграф присвячений опису основних властивостей предикату еквівалентності, який відіграє важливу роль при використанні методу порівняння. Уводиться визначення предикату еквівалентності через його властивості рефлексивності, симетричності й транзитивності. Доведено ряд теорем про предикат еквівалентності  $E$ , які допомагають розкрити його основні властивості. З теоремами про предикат

еквівалентності тісно пов'язане поняття ізоморфізму еквівалентностей. Вивчено основні особливості ізоморфізму предикатів, доведено ряд теорем про ізоморфні предикати. Дано визначення функціонального предикату, сформульовано і доведено теорему про функціональний предикат, яка дозволяє переходити до предикату рівності. Це є важливим результатом, оскільки дана теорема дозволяє переходити від ізоморфних предикатів до співпадаючих, що значно спрощує опис.

Другий параграф третього розділу присвячений одержанню загального виду довільного бінарного предикату. Для довільного бінарного предикату  $P$ , який заданий на множині  $M_1 \times M_2$ , одержане таке уявлення, в якому порівняння значень двох відповідних до нього функцій  $f_1$  та  $f_2$  здійснюється за допомогою найпростішого в якомусь смислі предикату, замінюючого предикат рівності у формулі:

$$E(x, y) = D(f(x), f(y)), \quad (7)$$

де множина  $M_2$ , сюр'єктивна функція  $f: M_1 \rightarrow M_2$ , предикат рівності  $D$  визначаються з точністю до ізоморфних до них об'єктів.

З цією метою вводяться визначення порядку та узгодження предикатів. Доведено такі ствердження. Відношення порядку  $>$ , заданого на множині предикатів еквівалентності, які діють на  $M_1 \times M_2$ , є відношенням часткового порядку. Те ж саме відношення, задане на множині  $\mathcal{R}_P$  предикатів, узгоджених з предикатом  $P$ , який визначений на  $M_1 \times M_2$ , є відношенням часткового порядку. Доведено теорему про те, що предикат  $\langle G^m, N_1 \times N_2 \rangle$ , визначений за допомогою формул (7) та поданих нижче:

$$\forall x_1, x_2 \in M_1 \quad E_1^m(x_1, x_2) = 1 \Leftrightarrow (\forall y \in M_2 : P(x_1, y) = P(x_2, y)); \quad (8)$$

$$\forall y_1, y_2 \in M_2 \quad E_2^m(y_1, y_2) = 1 \Leftrightarrow (\forall x \in M_1 : P(x, y_1) = P(x, y_2)), \quad (9)$$

узгоджений з предикатом  $\langle P, M_1 \times M_2 \rangle$ .

Доведено таку теорему. Для будь-якого предикату  $< P, M_1 \times M_2 >$  предикат  $G^m$ , визначений на  $N_1 \times N_2$  за допомогою формул (8), (9) та нижчеподаною:

$$\forall n_1 \in N_1, \forall n_2 \in N_2 \quad G^m(n_1, n_2) = P(g_1^m(n_1), g_2^m(n_2)), \quad (I0)$$

с максимальним серед усіх узгоджених з  $P$ , тобто  $\forall G \in \mathcal{R}_P$   $G^m > G$ . Наслідком даної теореми є те, що з антисиметричності відношення  $>$  на  $\mathcal{R}_P$  виходить, що коли  $G > G^m$ , то  $G_P$  ізоморфний до  $G^m$ . Тобто на множині  $\mathcal{R}_P$  предикатів, узгоджених із  $P$ , максимальний елемент визначений з точністю до ізоморфних  $G^m$  їх  $P$ -обмежень. Таким чином, довільний бінарний предикат  $P$ , визначений на множині  $M_1 \times M_2$ , можна представити у вигляді

$$\forall x \in M_1, \forall y \in M_2 \quad P(x, y) = G^m(f_1(x), f_2(y)), \quad (II)$$

де  $f_i: M_i \rightarrow N_i$  - сюр'єктивні функції,  $N_i$  - деякі множини,  $G^m$  - бінарний предикат, визначений з точністю до ізоморфізму на  $N_1 \times N_2$ .

У психофізичних дослідженнях, зокрема, у пов'язаних з природною мовою, часто доводиться розглядати предикат  $P$  на менших множинах  $M'_1$  й  $M'_2$  за ті, на яких він реально визначений. Тому природно ставити питання про те, наскільки вичерпну інформацію про предикат можна одержати, досліджуючи його на множині  $M'_1 \times M'_2$ , і які достатньо вибрати множини  $M'_1 \subset M_1, M'_2 \subset M_2$ , щоб мати достатню або всю інформацію про предикат  $P$ .

Одержано представлення для бінарного предикату  $P$  розпізнавання деяких фізичних об'єктів  $X_i$ , яке моделює процес з'явлення інтелектуальних образів  $f_i(X_i)$  і функцію їх розпізнавання  $L$ . Дане представлення має властивість максимальності в тому сенсі, що недосяжна більш детальна ідентифікація фізичних об'єктів  $X_i$ . Розглянуто достатню умову для опису властивостей предикату  $P$ , коли дослідження проводиться на

обмеженій області фізичних об'єктів, меншій за область визначення предикату  $P$ .

Розглянуто загальні види класів рефлексивних, симетричних, антирефлексивних, антисиметричних і асиметричних предикатів, представлені з використанням вільного предикатного параметру. За допомогою предикатів довільного порядку та алгебраїчної системи для їх формального опису подані деякі логічні поняття й поняття теорії відношень. Розроблено кванторну алгебру скінченних предикатів другого порядку, досліджено її основні властивості.

У скінченній алгебрі квантор спільності  $\forall x P(x)$  і квантор існування  $\exists x P(x)$  для предикату  $P(x)$  записується так:

$$\forall x P(x) = P(a_1) \wedge P(a_2) \wedge \dots \wedge P(a_k), \quad (I2)$$

$$\exists x P(x) = P(a_1) \vee P(a_2) \vee \dots \vee P(a_k). \quad (I3)$$

При запису цих формул прийнято, що змінна  $x$  задана у скінченній алгебрі на множині  $A = \{a_1, a_2, \dots, a_k\}$ . Якщо скінченна алгебра розуміється як універсальна, це визначення кванторів стає неефективним, оскільки області завдань для змінних універсальної алгебри чітко не окреслені. Звідси праві частини формул (I2) і (I3) повинні містити невизначено велику кількість кон'юнктивних і диз'юнктивних членів. Ефективне визначення кванторів для універсальної алгебри досягається введенням фіксованої області  $M = \{\sigma_1, \sigma_2, \dots, \sigma_p\}$  для зв'язаної змінної  $x$ .

З визначення приймаємо

$$\forall x (x \in M \supset P(x)) = P(\sigma_1) \wedge P(\sigma_2) \wedge \dots \wedge P(\sigma_p), \quad (I4)$$

$$\exists x (x \in M \supset P(x)) = P(\sigma_1) \vee P(\sigma_2) \vee \dots \vee P(\sigma_p). \quad (I5)$$

Мовою універсальної алгебри та з використанням уведених визначень описані поняття належності елемента множині, перетин, об'єднання, різниця і симетрична різниця множин; відношення еквівалентності виражається через породжуюче її розбивання. Роз-

глянуто предикати другого порядку простішого виду  $P(X(x))$  з областю задання  $M$ , що є сімейством усіх одномісних предикатів, заданих на скінченній множині літер  $M = \{a_1, a_2, \dots, a_k\}$ . Відповідно до формули для ДДНФ, будь-який предикат можна подати у вигляді

$$P(X(x)) = \bigvee X(x)^{\Sigma(x)}, \\ P(\Sigma(x)) = 1 \quad (I6)$$

В рамках диз'юнктивно-кон'юнктивної алгебри предикати виду  $X(x)^{\Sigma(x)}$  є елементарними. Кванторною алгеброю скінченних предикатів другого порядку буде називатись алгебраїчна система, в ролі елементарних предикатів якої виступають предикати другого порядку  $X(a_1), X(a_2), \dots, X(a_k)$ , а в ролі елементарних операцій – заперечення, кон'юнкція та диз'юнкція предикатів.

Останній параграф третього розділу присвячений удосконаленню алгоритмів мінімізації формул логічної алгебри. В універсальній алгебрі скінченних предикатів довільного порядку ми позбавлені можливості використовувати принцип подвійності, діючий в алгебрі логіки, тому операції диз'юнкції та кон'юнкції в універсальній алгебрі не є рівноправними. Через це методи кон'юнктивної мінімізації повинні розглядатись спеціальним чином.

У четвертому розділі розглянуте коло питань, пов'язаних з уведенням ситуаційно-текстового предикату. Основною ідеєю першого параграфу є обґрунтування об'єктивності підходу до вивчення семантики природної мови з використанням методу компараторної ідентифікації. Уводяться робочі визначення понять ситуації й тексту. Під ситуацією розуміється будь-який не дуже протяжний у просторі й часі фрагмент об'єктивної реальності. Під текстом в роботі мається на увазі правильно побудоване "осмислене" розповідне речення. Будь-який текст повинен виражати цілком певну думку. Думка, яка є в тексті, називається смислом цього тексту.

Під текстом розуміється якийсь фізичний об'єкт, а не суб'єктивний результат його сприймання або розуміння іспитником. Множина текстів, сформована для проведення експериментів з іспитником, є якась, достатньо чітко окреслена, сукупність текстів, сформована дослідником згідно з задачами вивчення інтелекту іспитника в даному експерименті.

Таким чином, є множина ситуацій  $S = \{s_i\}$ ,  $j = \overline{1, n}$  і множина текстів  $T = \{t_k\}$ ,  $k = \overline{1, m}$ . Ці дві множини є входними для експерименту, в ході якого дослідник пред'являє іспитникові різні пари  $(s, t)$ ,  $s \in S$ ,  $t \in T$ . Відповідь іспитника на кожне запитання дослідника можна трактувати як предикат  $P_i(s, t) = \varepsilon$ , де  $s \in S$ ,  $t \in T$ ,  $i$  - номер іспитника,  $\varepsilon = \{0, 1\}$ . В роботі вводяться визначення еквівалентності ситуації відносно іспитника  $i$ . В результаті експерименту з різними іспитниками  $i$  та  $i'$  може статися, що  $P_i(s, t) \neq P_{i'}(s, t)$ . Ця нерівність відображає індивідуальні відмінності іспитників, оскільки зрозуміло, що, формулюючи предикат  $P_i$ , іспитник  $i$  оперує не з ситуацією  $s$  як такою, а з її образом  $\sigma(s, i)$  у своїй свідомості, і не з текстом  $t$ , а з його образом  $\tau(t, i)$ , так що в дійсності реалізується предикат  $\Pi(\sigma, i), \tau(t, i) = P_i(s, t)$ . Доведено теорему про те, що відношення еквівалентності  $s \sim s'$  між ситуаціями рефлексивне, симетричне й транзитивне. Доведено аналогічну теорему про відношення  $t \sim t'$  між текстами.

В результаті впливу ситуації  $s$  на органи почуття іспитника в його свідомості виникає суб'єктивний образ  $\sigma$  цієї ситуації, який називається сприйняттям ситуації  $s$ . Покладаємо, що сприйняття  $\sigma$  однозначно визначається ситуацією, яка його викликала. Функція  $\sigma = f(s)$  називається функцією сприйняття ситуації,  $f: S \rightarrow M$ , де  $M$  - сукупність усіх сприймань,

породжуваних ситуаціями з множини  $S$ . Сприймаючи текст  $t$ , розуміючи його, іспитник дістає з нього цілком певну думку  $\tau$ , яка є суб'єктивним образом тексту. Будемо вважати, що думка  $\tau$  однозначно визначається текстом  $t$ , який її викликав. Функцію залежності думки  $\tau$  від тексту  $t$   $\tau = g(t)$  назвемо функцією розуміння тексту,  $g: T \rightarrow N$ , де  $N$  - множина значень функції  $g$ , тобто сукупність усіх думок, які породжені текстами з множини  $T$ .

Покладасмо, що відповідь іспитника  $\varepsilon = P(s, t)$  повністю визначається сприйняттям  $\sigma = f(s)$  ситуації  $s$  і смислом  $\tau = g(t)$  тексту  $t$ . Звідси виходить, що існує предикат  $\varepsilon = L(\sigma, \tau)$ , реалізуємий іспитником, який зветься предикатом усвідомлення. Вибір цього терміну обумовлений тим, що іспитник формує значення  $\varepsilon \in \{0, 1\}$  предикату  $L(\sigma, \tau)$  в результаті усвідомлення відповідності ( $\varepsilon = 1$ ) або невідповідності ( $\varepsilon = 0$ ) думки  $\tau$  сприйняттю  $\sigma$ . Предикати  $P$  і  $L$ , функції  $f$  і  $g$  та змінні  $s, t, \sigma, \tau$  пов'язані залежністю

$$P(s, t) = L(f(s), g(t)) = L(\sigma, \tau). \quad (I7)$$

Існування предикату  $L$ , функцій  $f$  і  $g$ , їх конкретний вигляд і взаємозв'язок, виражений формулою (I7), можна установити, базуючись виключно на фізичному спостереженні поведінки іспитника, яка характеризується предикатом  $P(s, t)$ .

Для знаходження структури предикату  $P$  вводяться предикати  $E_1(s_1, s_2) = \forall t \in T (P(s_1, t) \sim P(s_2, t))$ ,  $E_2(t_1, t_2) = \forall s \in S (P(s, t_1) \sim P(s, t_2))$ , які однозначно визначаються предикатом  $P$ . Предикат  $E_1$  заданий на множині  $S \times S$ , предикат  $E_2$  - на  $T \times T$ . Предикат  $E_1$  назвемо предикатом метамірності ситуацій, предикат  $E_2$  - тотожності текстів. Предикати  $E_1$  і  $E_2$  рефлексивні, симетричні й транзитивні, що означає, що це предикати еквівалентності. Предикат  $E_1$  визначає розбивання  $\mathcal{R}$  множини  $S$

на шари ситуацій. Усі ситуації, які належать одному шарові розбиванні  $\mathcal{R}$ , метамірні. Предикат  $E_2$  визначає розбивання  $\mathcal{G}$  множини  $\mathcal{T}$  на шари текстів. Усі тексти, які належать одному шарові розбиванні, тотожні.

Предикати  $E_1$  і  $E_2$  можна подати у вигляді

$$\begin{aligned} E_1(s_1, s_2) &= D_1(f(s_1), f(s_2)); \\ E_2(t_1, t_2) &= D_2(g(t_1), g(t_2)), \end{aligned} \quad (18)$$

де  $f$  - канонічне відображення множини  $\mathcal{S}$  на розбивання  $\mathcal{R}$ ,  $g$  - канонічне відображення множини  $\mathcal{T}$  на розбивання  $\mathcal{G}$ ;  $D_1$  - предикат рівності на  $\mathcal{S} \times \mathcal{S}$ ,  $D_2$  - предикат рівності на  $\mathcal{T} \times \mathcal{T}$ . Класові  $K_\alpha$  усіх ситуацій  $s \in \mathcal{S}$ , метамірних ситуації  $\alpha \in \mathcal{S}$ , тобто сприйняттю, яке породжується у свідомості іспитника ситуацією  $\alpha \in \mathcal{S}$ , відповідає предикат  $K_\alpha(s) = E_1(s, \alpha)$ . Класові  $Q_\beta(t)$  усіх текстів  $t \in \mathcal{T}$ , тотожних текстові  $\beta \in \mathcal{T}$ , тобто думці, яка виникає у свідомості іспитника у відповідь на пред'явлення йому текста  $\beta$ , відповідає предикат  $Q_\beta(t) = E_2(t, \beta)$ .

З урахуванням залежностей (18), отримаємо формули:

$$\begin{aligned} K_\alpha(s) &= \forall t \in \mathcal{T} (P(s, t) \sim P(\alpha, t)); \\ Q_\beta(t) &= \forall s \in \mathcal{S} (P(s, t) \sim P(s, \beta)), \end{aligned} \quad (19)$$

які виражають суб'єктивні за своєю природою сприйняття й думки іспитника через предикат  $P$ , який відображує його об'єктивно спостережену поведінку.

Розглянутий спосіб визначення предикату усвідомлення  $L$ . Такий предикат існує для будь-якого  $P$ . Предикат  $L$  можна обчислити за відомим предикатом  $P$  та відомими функціями  $f$  і  $g$  за такою формулою:

$$L(\sigma, \tau) = \exists s \in \mathcal{S} \exists t \in \mathcal{T} (P(s, t) f(s, \sigma) g(t, \tau)). \quad (20)$$

З рівності (20) випливає також залежність

$$P(s, t) = \exists \sigma \in \mathcal{M} \exists \tau \in \mathcal{N} (L(\sigma, \tau) f(s, \sigma) g(t, \tau)), \quad (21)$$

за допомогою якої предикат  $P$  може бути обчислений за відо-

ним предикатом  $L$  і відомими функціями  $f$  і  $g$ .

П'ятий розділ дисертації присвячений компараторній ідентифікації лінгвістичних об'єктів морфемного рівня. Відомо, що смисл похідного слова в загальному випадку неможливо описати як просту суму смислів складаючих його морфів. Для кращого розуміння механізму формування смислу деривату досліджені та описані математично міжморфемні семантичні зв'язки, які є в процесі деривації між префіксальними й кореневими морфемами, кореновими й суфіксальними морфемами, а також між основами й закінченнями. Розглянуто основні проблеми математичного моделювання міжморфемних семантичних зв'язків і можливість застосування основних засобів методу компараторної ідентифікації для подолання цих проблем.

Нехай  $\epsilon$  множина похідних слів, кожне з яких можна подати набором його морфів

$$\Pi_1 * \Pi_2 * \dots * \Pi_5 * K * C_1 * C_2 * \dots * C_5 * Ok, \quad (22)$$

де  $\Pi_i$  - префіксальні морфи ( $i = \overline{1,5}$ ),  $K$  - коренева морфема,  $C_j$  - суфіксальні морфи у суфіксальному ланцюжку ( $j = \overline{1,5}$ ),  $Ok$  - закінчення. Знак  $*$  позначає, що між цими морфами установлені певні семантичні зв'язки, які описуються математичними засобами. Нехай  $M$  - множина морфем одного типу. На цій множині уведемо систему предикатів  $\{P(\varphi)\}$  так, щоб будь-який предикат  $P(\varphi_i)$  обернувся в 1 на множині морфем з якоюсь семантичною роллю  $\varphi_i$  і дорівнював 0 у протилежному разі. Таким чином, множину предикатів  $P(\varphi)$  можна зіставити з множиною семантичних ролей морфем даного типу.

Кожній морфемі  $m_i \in M$  відповідає деякий предикат  $P(\varphi)$ , який дорівнює 1 при підстановці семантичних ролей морфемі  $m_i$ . Отже, кожному елементові  $m_i$  взаємно однозначно відповідає певний одномісний підстановочний предикат, який задає множину

семантичних ролей даної морфєми, поданої у різних словах своїми морфємами. Операція з'єднання двох морфєм  $m_1$  і  $m_2$  у вигляді композиції двох предикатів  $P_1(q_1) \in M_1$  і  $P_2(q_2) \in M_2$  буде характеризуватись узгодженням певних /не усіх/ семантичних ролей цих морфєм. Насправді, частина семантичних ролей морфєм, які стоять поряд, не узгоджуються, тому операція з'єднання двох морфєм запишеться таким чином:

$$P_1(q_1) * P_2(q_2) = \lambda(q_1, q_2) \cdot P_1(q_1) \cdot P_2(q_2), \quad (23)$$

де логічний добуток  $P_1(q_1) \cdot P_2(q_2)$  описує усі можливі зв'язки між семантичними ролями морфєм  $m_1$  і  $m_2$ , а предикат  $\lambda(q_1, q_2)$  виключає ті зв'язки, які в мові не реалізуються.

Для зручності математичного опису можна без огрублення задачі вважати, що множини  $M_1$  і  $M_2$  співпадають із множинами усіх предикатів, визначених відповідно на множинах  $Q_1$  і  $Q_2$ . Тоді бульові алгебри предикатів із множин  $M_1$  і  $M_2$  є підалгебрами бульової алгебри бінарних предикатів, заданих на  $Q_1 \times Q_2$ . Доведено теорему, яка дозволяє аксіоматично увести операцію з'єднання морфєм. Нехай є скінченна бульова алгебра  $M$  і дві підалгебри  $M_1 \subseteq M$ ,  $M_2 \subseteq M$ .

Теорема. Нехай функція  $f: M_1 \times M_2 \rightarrow M$  задовольняє властивостям:

$$\begin{aligned} f(x, y) \wedge y &= f(x, y); & f(x, y \vee z_2) &= f(x, y) \vee f(x, z_2); \\ f(x, y_1 \vee y_2) &= f(x, y_1) \vee f(x, y_2), \end{aligned} \quad (24)$$

тоді  $f(x, y)$  можна представити у вигляді

$$f(x, y) = \lambda \wedge y, \quad (25)$$

де  $\lambda$  - фіксований елемент з  $M$ . Навпаки, якщо  $f(x, y)$  можна представити у вигляді (25), то виконуються властивості (24).

У роботі побудовано математичні моделі семантичної взаємодії префіксів у складі префіксального ланцюжка /мал. I/, префіксів і коренів, суфіксальних морфєм у суфіксальних ланцюжках

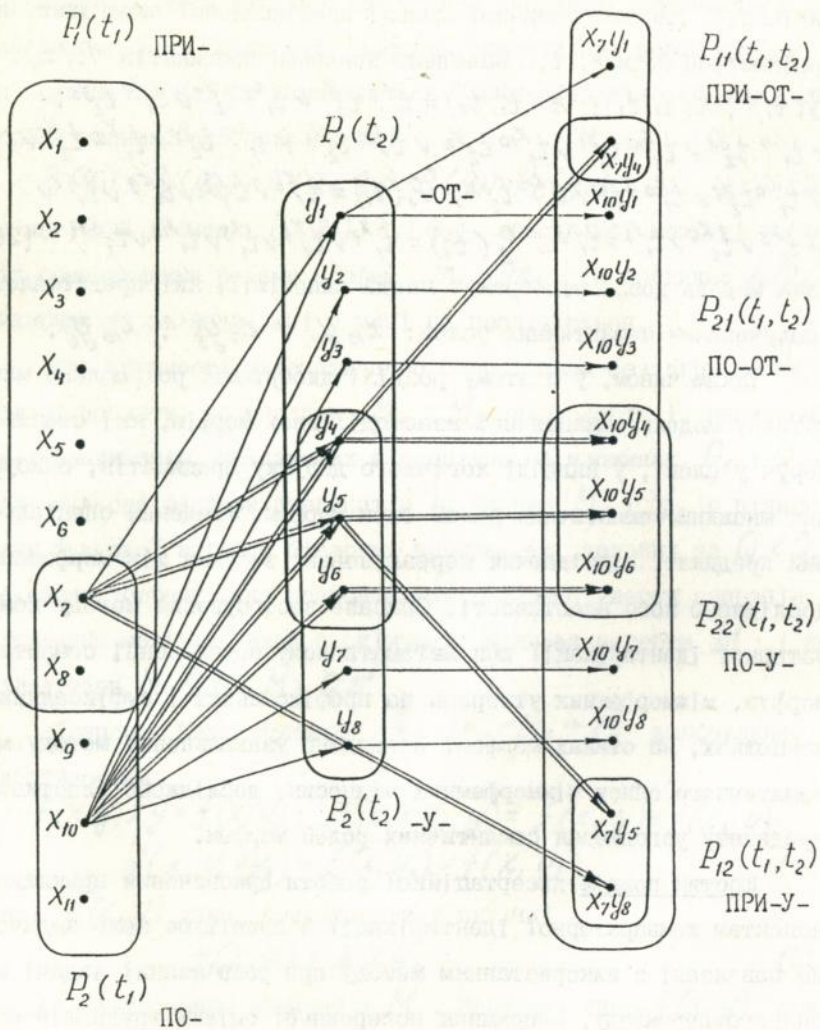
і взаємодії суфіксів і закінчень у слові. Наявність або відсутність узгодження між морфемами визначається під час експерименту над іспитником з використанням методу компараторної ідентифікації. Для математичної моделі префіксальної семантики, яка

проілюстрована мал. 1, наведемо значення предикатів  $P_1(t_1)$ ,  $P_2(t_1)$ ,  $\lambda(t_1, t_2)$ :  $\lambda(t_1, t_2) = t_1^{x_7} t_2^{y_1} \vee t_1^{x_7} t_2^{y_4} \vee t_1^{x_7} t_2^{y_5} \vee t_1^{x_7} t_2^{y_8} \vee t_1^{x_{10}} t_2^{y_1} \vee t_1^{x_{10}} t_2^{y_2} \vee t_1^{x_{10}} t_2^{y_3} \vee t_1^{x_{10}} t_2^{y_4} \vee t_1^{x_{10}} t_2^{y_6} \vee t_1^{x_{10}} t_2^{y_8} \vee t_1^{x_{10}} t_2^{y_9} \vee t_1^{x_{10}} t_2^{y_{10}} \vee t_1^{x_{10}} t_2^{y_{11}} \vee t_1^{x_5} \vee t_1^{x_6} \vee t_1^{x_7} \vee t_1^{x_8}$ ;  $P_1(t_1) = t_1^{x_1} \vee t_1^{x_2} \vee t_1^{x_3} \vee t_1^{x_4} \vee t_1^{x_5} \vee t_1^{x_6} \vee t_1^{x_7} \vee t_1^{x_8}$ ;  $P_2(t_1) = t_1^{x_7} \vee t_1^{x_8} \vee t_1^{x_9} \vee t_1^{x_{10}} \vee t_1^{x_{11}}$  (26)

Така модель добре відображає явища синонімії, які представлені сполученнями семантичних ролей:  $x_{10} y_4$ ,  $x_{10} y_5$ ,  $x_{10} y_6$ .

Таким чином, у п'ятому розділі дисертації розроблено математичну модель семантичної взаємодії двох морфів, які стоять поруч у слові, у вигляді логічного добутку предикатів, описуючих множини семантичних ролей пари морфем. Уведений спеціальний предикат, відсіваючий нереалізовані зв'язки між морфемами, досліджено його властивості. Описане застосування методу компараторної ідентифікації при математичному моделюванні семантики морфів, міжморфемних утворень на префіксальних і суфіксальних ланцюжках, на стиках морфем. Проведено узагальнення методу математичного опису міжморфемних відносин, досліджено властивості предикату узгодження семантичних ролей морфем.

Шостий розділ дисертаційної роботи присвячений прикладним аспектам компараторної ідентифікації й висвітлює коло задач, що пов'язані з використанням методу при розв'язанні задачі машинного перекладу, одержання попередньої оцінки труднощів машинного перекладу, виміру послівної несхожості паралельних текстів при перекладі, побудови морфологічної системи для російсько-українського і українсько-російського комп'ютерного перекладу, моделювання деривативних відносин для української мови,



Мал. I. Семантична сполучуваність  
префіксальних морфем

а також описано побудову математичних моделей семантичних закономірностей процесу термінологізації в кольоропозначенні англійської мови. Таким чином, шостий розділ ілюструє використання методу, який пропонується, у розв'язанні різних лінгвістичних задач.

Машинний переклад сьогодення в більшій мірі відповідає інформаційній реальності, ніж колишні проекти високоякісного МП будь-яких текстів без редагування, оскільки стало очевидним, що різні тексти в різній мірі придатні для МП, і економічна доцільність МП залежить як від типу текста, так і від типу перекладу. Виходячи з цього, критерії ефективності МП і оцінки складності, які уводяться, в більшій мірі залежать від зазначених задач. В роботі проаналізовано переклади простої художньої прози на шість європейських мов з точки зору виміру послівної несхожості перекладів. Було запропоновано таку методику виміру послівної несхожості перекладів двох паралельних текстів  $T_1$  і  $T_2$ : тексти пред'являються експертові, який переглядає їх речення, ігноруючи відмінності, які виникли з причини граматичних відмін або через використання стандартної фразеології, пропускаючи послівні збіжності і накопичуючи число значущих слів у лічильниках  $c_1$  і  $c_2$  для текстів  $T_1$  і  $T_2$  відповідно. По закінченні текстів обчислюється міра послівної несхожості  $d$  по формулі

$$d = \frac{c_1 + c_2}{|T_1| + |T_2|}, \quad (27)$$

де  $|T_1|$  і  $|T_2|$  означають число значущих слів у текстах  $T_1$  і  $T_2$ .

З використанням описаної методики підраховано щільність послівних розбіжностей  $d$  для усіх пар з текстів сьома мовами: А - англійською, Р - російською, П - польською, І - іспанською, У - українською, Н - німецькою, Ф - французькою, для результатів підрахунку побудовано симетричну матрицю. З аналізу одержан-

них даних видно, що щільність послівних розбіжностей в середньому дорівнює 0,20 і дисперсія не дуже велика. У генетично близько споріднених мовах щільність розбіжностей не менш середньої, а сама більша різниця  $d = 0,41$  спостерігається на російсько-польській парі.

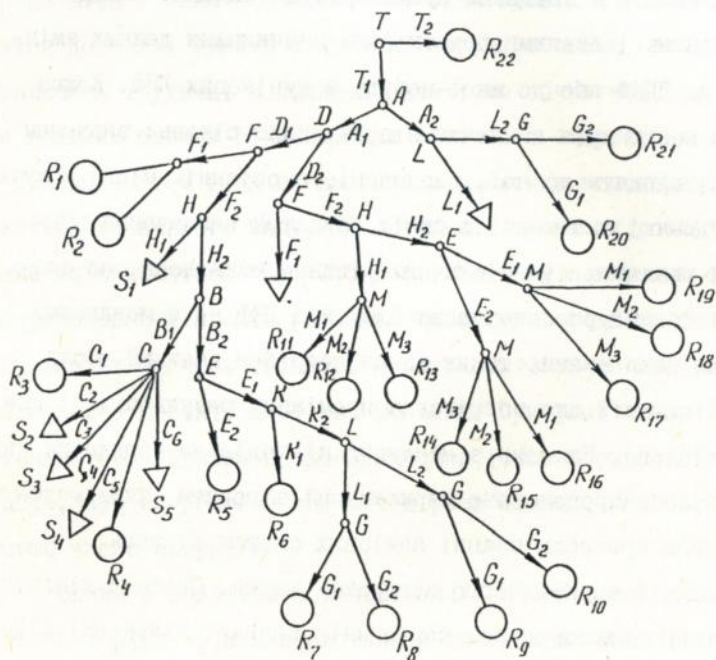
Факт, що міра послівної несхожості у споріднених мовах така велика, підтверджується дослідженнями, проведеними на паралельних текстах однієї й тієї ж мови, тобто у випадку, коли розглядаються переклади різних перекладачів з оригіналу на одну мову. В цьому випадку виявилось, що міра  $d$  значно більша за середню. Одержані результати дозволяють провести попередню перевірку складності МП залежно від типу тексту і задачі, яка вирішується.

Побудовано програмну систему для морфологічного перекладу з російської мови на українську й навпаки. Система морфологічного МП не претендує на повноту й високоякісність перекладу в цілому, однак має переваги перед існуючими подібними системами російсько-українського перекладу в області морфологічної обробки текстів. Розв'язання задач морфологічного аналізу та синтезу базується на понятті словозмінної парадигми, представленій таблицями і рівняннями, які пов'язують значення перемінних словозмінних ознак. Сукупність таких таблиць і рівнянь утворює морфологічний словник. Використання методу компараторної ідентифікації на етапі формування морфологічного словника дозволяє записати інформацію в найбільш компактній формі.

В роботі проведено математичне моделювання деривативних відносин для української мови, дано порівняльний аналіз деривативних систем двох мов на матеріалі іменників і дієслів української і російської мов. Побудовано дерево семантичних ролей суфіксальних морфем української мови, виділено істотні ознаки,

відмінні от ознак, які уведені для російської деривації.

Розроблено математичні моделі семантичних закономірностей процесу термінологізації в кольоропозначенні англійської мови. Під процесом термінологізації розуміється сукупність семантичних і структурних механізмів, які сприяють цілеспрямованій зміні, в першу чергу, плану змісту лексичної одиниці, її семантичної структури. Уведено й обґрунтовано повну, несуперечливу і нескоротну систему ознак, адекватно відображуючу зв'язки й відношення у досліджуваному процесі. Побудовано математичну модель процесу термінологізації у кольоропозначенні англійської мови у вигляді системи рівнянь алгебри предикатів і предикатних операцій /мал. 2/.



Мал. 2. Фрагмент семантичного дерева для моделі процесу термінологізації

У цьому розділі розглянуто практичні застосування результатів дослідження та дано рекомендації у напрямках можливого використання розроблених методів, моделей й алгоритмів. Математичні моделі різних природномовних процесів являють собою системи рівнянь, розв'язати які можна при підстановці різних початкових умов. Під час розвитку основних положень теорії інтелекту та їх практичного використання було розроблено декілька вирішувачів логічних рівнянь, кожен з яких мав безперечні переваги над попередніми, проте мав і свої недоліки. У роботі дано порівняльний аналіз розроблених програмних систем розв'язання логічних рівнянь, в розробці яких автор брав участь.

Універсальною процедурою при розв'язанні систем рівнянь алгебри предикатів є зведення предикату, що заданий вихідною системою рівнянь і певними початковими значеннями деяких змінних ознак, до ДДНФ або до якої-небудь з тупікових ДНФ. Кожна елементарна кон'юнкція визначає відразу клас рішень: значення змінних, які входять до неї, визначені, а останні змінні можуть приймати будь-які значення із своїх областей визначення. При розв'язанні реальних лінгвістичних рівнянь виявилось, що скористуватись процедурою одержання ДДНФ або ДНФ не є можливим, оскільки для розв'язання таких задач потрібен великий обсяг оперативної пам'яті для зберігання проміжних результатів і великий час лічіння. Як один з можливих підходів до подолання цих труднощів було запропоновано евристичний алгоритм, який моделює дії людини при розв'язанні подібних систем рівнянь.

Принцип дії евристичного алгоритму такий. Системи лінгвістичних рівнянь являють собою кон'юнкцію великої кількості відносно коротких рівнянь, кожне з яких має невелике число змінних. Якщо одне з рівнянь системи після підстановки знайдених проміжних значень змінних і виключення констант обертається в  $I$ , його

можна виключити при подальшому розв'язанні системи. Якщо рівняння обертається у бульов 0, це значить, що система несумісна з початковими умовами, і розв'язання не існує. Процес розв'язання продовжується до того часу, коли після підстановки у рівняння системи проміжних значень деякої змінної нових значень знайти не вдасться.

Наступний програмно реалізований метод розв'язання систем рівнянь, в розробці якого автор брав участь в рамках виконання держбюджетної тематики, — це метод ітераційного знаходження ДНФ через логічний добуток рівнянь системи. При цьому будується деревовидний граф рішення, у вузлах якого розташовані предикати, а дуги позначають входження більш віддалених від кореня дерева предикатів у менш віддалені. На відміну від евристичного, даний метод дозволить знайти будь-який корінь системи, але на практиці виникають труднощі з великою кількістю обчислень для розв'язання великих систем рівнянь. Для істотного скорочення часу розв'язання систем рівнянь виявилось ефективним попереднє зведення системи до певного виду, що дозволило одержати розв'язання системи за прийнятний час  $/0,5-1$  с/. За таку попередню обробку був запропонований процес розв'язання системи рівнянь без задання початкових значень змінних, в результаті якого знаходиться деяка множина коренів системи. Оскільки при розв'язанні систем рівнянь без задання початкових умов істотно зростає число операцій, був запропонований метод розв'язання, який реалізує найменше число операцій.

Одним з найбільш перспективних способів розв'язання систем логічних рівнянь з використанням програмних засобів є спроба розпаралелювання розв'язання систем предикатних рівнянь методами повного перебору й виключення змінних. При проведенні аналізу ефективності методів рішення уведено такі кількісні характе-

ристики /часові й просторові/:  $T_1$  – час обчислення при виконанні усіх дій послідовно, дорівнює об'єму обчислень методу або загальному числу операцій для знаходження рішення;  $T_w$  – час обчислення при максимально паралельному виконанні усіх дій на заданому рівні паралелізму, завжди дорівнює об'ємові обчислень в одній з виділених паралельних віток;  $W_w$  – максимальне число паралельних віток при часу обчислення  $T_w$ ;  $S_w$  – прискорення обчислень при  $W_w$  паралельних вітках у порівнянні з послідовним обчисленням:  $S_w = \frac{T_1}{T_w}$ ;  $v_w$  – щільність обчислень у  $W_w$  вітках:  $v_w = \frac{T_1}{W_w T_w} = \frac{S_w}{W_w} \cdot v_w$  характеризує можливу ефективність використання  $W_w$  паралельних процесів для реалізації обчислень у паралельних вітках.

Розглянуті методи й алгоритми розв'язання систем лінгвістичних рівнянь дозволяють підібрати необхідні програмні засоби для розв'язання конкретних задач, оскільки володіють як перевагами, так і недоліками, які корелюють із складністю й типом задачі.

Іншим цікавим додатком, описаним у сьомому розділі, є розробка семантичної структури найменувань хімічних сполучень, що є дуже важливим при формуванні бази знань експертних систем у хімічній галузі. Для побудови математичної моделі були досліджені відношення на множинах:  $M_1$  – назви сполучень,  $M_2$  – структура назв, тобто складові елементи хімічних назв: суфікси, префікси, корені, цифрові префікси;  $M_3$  – множина хімічних формул /їх структурне представлення, яке прийняте у хімії/;  $M_4$  – множина структурних складових хімічних формул. Досліджено семантичні ролі префіксів, суфіксів і коренів, які беруть участь в утворенні назв хімічних сполучень. Наприклад, рівняння

$$R_{1,2}(t_1, t_2) = t_1^{d_1} (t_2^{z_1} \vee t_2^{z_2} \vee t_2^{z_3} \vee t_2^{z_4}) \vee t_1^{d_2} (t_2^{z_1} \vee t_2^{z_2})$$

описує органічне сполучення метан, яке належить до групи най-

простіших органічних речовин, володіє властивостями насиченості або граничності, складається з одного атому вуглецю і 4 атомів водню, предикат  $R_{1,2}$  описує відношення між коренем мет- і суфіксом -ан.

Розглянуто зв'язки між префіксальною, кореневою і суфіксальною морфемами. Наприклад, зв'язки між префіксальною морфемою ді-, коренем мета-, коренем ета-, а також коренем пропа- і суфіксом -іл запишуться рівнянням:

$$R_{16, 1, 15} (t_{16}, t_1, t_{15}) = t_{16}^{n_1} t_1^{k_1} t_{15}^{m_1} \vee t_{16}^{n_1} t_1^{k_1} t_{15}^{m_2} \vee t_{16}^{n_1} t_1^{k_1} t_{15}^{m_3} \vee t_{16}^{n_1} t_1^{k_2} t_{15}^{m_1} \vee t_{16}^{n_1} t_1^{k_2} t_{15}^{m_2} \vee t_{16}^{n_1} t_1^{k_2} t_{15}^{m_3} \vee t_{16}^{n_2} t_1^{k_1} t_{15}^{m_1} \vee t_{16}^{n_1} t_1^{k_1} t_{15}^{m_3}$$

Задаючи ті чи інші початкові умови в подібних рівняннях, ми можемо одержати різну інформацію, необхідну при розв'язанні конкретної задачі: шукання потрібного хімічного найменування, побудова відповідної хімічної формули, аналіз структурних складових формули або шукання хімічного еквіваленту.

Важливим практичним результатом роботи є програмна реалізація двохарової декомпозиції ситуаційно-текстового предикату. Оскільки будь-які текстові відношення можна представити у вигляді відповідних предикатів, а природна мова є незвичайно складним об'єктом, текстові відношення описуються складного виду предикатами, виникає необхідність спростити опис, тобто представляти /бажано автоматично/ у вигляді кон'юнкції більш простих предикатів, тобто утворювати декомпозицію предикатів. Нехай деяке бінарне відношення  $xPy$  описується предикатом  $P(x, y)$ . Його можна представити у вигляді

$$P(x, y) = Q(f(x), g(y)), \quad (28)$$

де  $u = f(x)$ ,  $v = g(y)$  - деякі класифікуючі функції, предикат  $Q(u, v)$  описує зв'язок між цими функціями. Метод двохарової декомпозиції предикатів розділяє їх як би на два шари - нижній, представлений функціями  $f(x)$  і  $g(y)$ , і верхній, представле-

ний предикатом  $Q(u, v)$ . Виконання методу двошарової декомпозиції предикатів дає можливість здійснювати схемну реалізацію системи рівнянь.

Розроблена програма *PRED* призначена для автоматичного здійснення двошарової декомпозиції бінарних предикатів, складається з основного модулю і 9 процедур, реалізована на ПЕОМ типу IBM PC/AT під керуванням операційної системи *MS DOS 5.0*. Потрібний об'єм оперативної пам'яті – 64 К, час лічби контрольного прикладу – 1,5 с, написана програма на мові *Pascal*. На вихідні дані накладаються такі обмеження: число змінних предикатів не повинно перевищувати 2, потужність області визначення кожної змінної  $\leq 80$ . В роботі наведений загальний алгоритм програми *PRED*; результати її роботи.

У дисертаційній роботі проведено аналіз можливих напрямків практичного використання розроблених моделей і алгоритмів, позначене коло задач, де використовується або може використовуватись метод компараторної ідентифікації, вироблено практичні рекомендації з використання одержаних результатів. Найбільш цікавими задачами автоматизованої обробки текстової інформації є задачі, пов'язані з розумінням тексту: анотування й реферування, машинний переклад, розпізнавання образів за текстом, формування баз знань із семантичною інтерпретацією, робота з гіпертекстами.

У додатках наведено документи, які підтверджують впровадження й використання результатів дисертаційної роботи, а також приклади роботи програм і результати обчислювального експерименту.

## ОСНОВНІ РЕЗУЛЬТАТИ І ВИСНОВКИ

Головним результатом дисертаційної роботи є розробка основ методології компараторної ідентифікації як нового напрямку в дослідженні та моделюванні мовної поведінки людини з метою використання розроблених математичних моделей при побудові інтелектуальних інформаційних систем. За матеріалами виконаної дисертації можна сформулювати такі основні висновки й результати.

1. Сформульовано концепцію застосування методу компарації при дослідженні та моделюванні мовної поведінки людини. Установлення збіжності або різниці будь-яких двох суб'єктивних станів людини дає реальну можливість переходу від словесних описів до фізично вірогідного об'єктивного математичного опису інтелектуальних функцій іспитника під час проведення психофізичного лінгвістичного експерименту.

2. Досліджено особливості лінгвістичних об'єктів різного рівня і сформульовано основні принципи ідентифікації морфем і міжморфемних відносин у слові на базі української та російської морфології й деривації, англійської термінологізації у кольоропозначенні.

3. Розроблено інструментальні засоби компараторної ідентифікації, необхідні для моделювання лінгвістичних відносин: алгебри предикатних операцій, підстановочних операцій, кванторна алгебра скінченних предикатів довільного порядку, засоби представлення загального виду бінарних предикатів, методи мінімізації формул алгебри скінченних предикатів довільного порядку, спосіб двошарової декомпозиції бінарних предикатів.

4. Побудовано математичні моделі семантичних відносин морфологічних і деривативних утворень, досліджено структуру міжморфемних семантичних відносин.

5. Запропоновано методику побудови ефективної системи морфологічного машинного перекладу з російської мови на українську й навпаки. Уперше розроблено методику одержання попередньої оцінки складності машинного перекладу, проведено вимір послівної несхожості паралельних текстів для основних європейських мов.

6. Проведений порівняльний аналіз розроблених за участю автора програмних систем для розв'язання рівнянь алгебри скінченних предикатів різних типів, описані можливі методи розв'язання спеціальних лінгвістичних рівнянь.

7. Ефективність використання методу компараторної ідентифікації підтверджується вирішенням ряду важливих задач автоматизованої обробки текстової інформації. Побудовано й програмно реалізовано математичну модель семантичної структури найменувань хімічних сполучень. Здійснено програмну реалізацію методу двохарової декомпозиції ситуаційно-текстового предикату. Розглянуто коло задач, пов'язаних з розумінням тексту.

8. Сформульовані та досліджені найбільш важливі задачі автоматизованої обробки текстової інформації, пов'язані з розумінням тексту: анотування, реферування, пошук інформації в базах даних, корекція помилок тощо. Розглянуто основні проблеми створення бібліотечних інформаційних систем, вироблено практичні рекомендації з використання одержаних результатів.

#### ОСНОВНІ ПРАЦІ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Шабанов-Кушнарєнко Ю.П., Шаронова Н.В. Компараторная идентификация лингвистических объектов. - Монография. - Киев: институт системных исследований образования Украины, 1993.-116с.

2. Бондаренко М.Ф., Шаронова Н.В. О математическом описании процессов словообразования. Проблемы бионики, Харьков, Вища школа, 1981, вып. 27, с. 83-88.

3. Бондаренко М.Ф., Шаронова Н.В. Задача фрагментации суффиксов имен существительных. Проблемы бионики, Харьков, Вища школа, 1981, вып. 27, с. 79-83.
4. Бондаренко М.Ф., Буркова Н.М., Шаронова Н.В. Математическое описание суффиксального словообразования глаголов. Проблемы бионики, Харьков, Вища школа, 1982, вып. 28, с. 8-13.
5. Осыка А.Ф., Буркова Н.М., Шаронова Н.В. Математическая модель фрагментированных суффиксов глаголов русского языка. - Проблемы бионики, Харьков, Вища школа, вып. 29, 1982, с. 12-19.
6. Шаронова Н.В. Формализация распределения морфов в морфеме. Проблемы бионики, Харьков, Вища школа, вып. 30, с. 29-33.
7. Шаронова Н.В., Бузницкая Э.М. О структуре системы признаков при моделировании словообразования. Проблемы бионики, Харьков, Вища школа, вып. 31, 1983, с. 12-19.
8. Иванова О.А., Шаронова Н.В. Моделирование семантических связей в русских дееминутивах. Проблемы бионики, Харьков, Вища школа, вып. 39, 1987, с. 34-39.
9. Левицкий А.С., Шаронова Н.В., Бузницкая Э.М. Использование лингвистического регистра при решении задач анализа и синтеза русского текста. Проблемы бионики, Харьков, Вища школа, вып. 38, 1987, с. 12-19.
10. Рябова Н.В., Шаронова Н.В. О математическом моделировании отсубстантивных имен существительных русского языка. Проблемы бионики, Харьков, Вища школа, вып. 34, 1985, с. 30-34.
11. Шабанов-Кушнарченко С.Ю., Шаронова Н.В. О кванторной алгебре конечных предикатов произвольного порядка. Проблемы бионики, Харьков, Вища школа, вып. 40, 1988, с. 35-41.
12. Романовская Н.Н., Шаронова Н.В. Математическое моделиро-

вание семантических закономерностей процесса терминологизации /на материале лексических единиц цветообозначения английского языка/. - Проблемы бионики, Харьков, Вища школа, вып. 40, 1988, с.31-35.

13. Шабанов-Кушнаренко С.Ю., Шаронова Н.В., Шубин И.Ю. О некоторых алгоритмах канонической конъюнктивной минимизации формул алгебры конечных предикатов. Проблемы бионики, Харьков, Вища школа, вып. 42, 1989, с. 50-55.

14. Шабанов-Кушнаренко С.Ю., Шаронова Н.В., Шубин И.Ю., Алгоритмы канонической дизъюнктивной минимизации формул алгебры конечных предикатов. Проблемы бионики, Харьков, Вища школа, вып. 43, 1989, с. 39-44.

15. *Yu. P. Shabanov-Kushnarenko, N. Sharonova. Development of Intelligence Theory for Systems of Semantic Image Processing. - Pattern Recognition and Image Analysis, Vol. 2, No 2, 1992, p. 220-222.*

16. Бондаренко М.Ф., Ситников Д.Э., Шаронова Н.В., Явтушенко Е.В. О математическом описании межморфемных отношений. - Проблемы бионики, Харьков, "Основа", вып. 47, 1991, с. 135-139.

17. Булкин В.И., Ситников Д.Э., Шабанов-Кушнаренко Ю.П., Шаронова Н.В. Математические модели межморфемных связей на множестве полисемантических производящих основ и словообразовательных суффиксов. - Проблемы бионики, Харьков, "Основа", вып. 47, 1991, с. 3-8.

18. Рублинцкий В.И., Шабанов-Кушнаренко Ю.П., Шаронова Н.В. Объективный подход к изучению семантики естественного языка. Сообщение 1. - "АСУ и приборы автоматки", Харьков, "Основа", вып. 99, 1993, с. 35-39, Сообщение 2. - с. 40-43.

19. Ицков Ф.Э., Шабанов-Кушнарченко Ю.П., Шаронова Н.В. Об основных свойствах предиката эквивалентности и его использовании в теории интеллекта. – Научно-практический сборник "Методы анализа и синтеза систем", Северодонецк, 1992, с. 22–27.

20. Булкин В.И., Кравец О.А., Шаронова Н.В. Математическое моделирование межбуквенных отношений. – "АСУ и приборы автоматки", Харьков, "Основа", 1994, вып. 104, с. 18–22.

21. Ковалев Ю.В., Лихачева О.А., Шаронова Н.В. Вычленение простейших отношений в алгебре конечных предикатов. – Деп. в УкрНИИТИ, 1988, № 1019 – Ук88, 23 с.

22. Ковалев Ю.В., Лихачева О.А., Шаронова Н.В. Об одном методе декомпозиции уравнений алгебры конечных предикатов. – Деп. в УкрНИИТИ, 1988, № 515 – Ук88, 11 с.

23. Левицкий А.С., Рябова Н.В., Шаронова Н.В. О математическом моделировании деривативного анализа суффиксальных существительных со значением лица. – Деп. в УкрНИИТИ, 1988, № 1574–Ук88, 16 с.

24. Булкин В.И., Кравец О.А., Ситников Д.Э., Шаронова Н.В. Разработка математических моделей межморфемных семантических отношений на множестве префиксальных морфем. – Деп. в УкрИНТЭИ, № 1742 – Ук92, 1992, 10 с.

25. Дюкарев М.Ю., Шабанов-Кушнарченко Ю.П., Шаронова Н.В., Бузницкая Э.М. Формальное описание типов графов, маршрутов и связности на языке алгебры идей. – Деп. в УкрИНТЭИ, № 1488 – Ук92, 1992, 8 с.

26. Дюкарев М.Ю., Шабанов-Кушнарченко Ю.П., Шаронова Н.В. Использование алгебры идей для формального представления понятий теории графов. – Деп. в УкрИНТЭИ, № 159 – Ук93, 7 с.

27. Дюкарев М.Ю., Шабанов-Кушнарченко Ю.П., Шаронова Н.В. Использование алгебры конечных предикатов в формальном представ-

лении графов и семантических сетей. - Деп. в УкрИНТЭИ, № 160 - Ук93, 1993, 10 с.

28. Ситников Д.Э., Шабанов-Кушнаренко Ю.П., Шаронова Н.В.  
О представлении общих видов классов бинарных предикатов с помощью свободного параметра. - Деп. в УкрИНТЭИ, № 138 - Ук92, 6 с.

29. Ситников Д.Э., Шабанов-Кушнаренко Ю.П., Шаронова Н.В.  
О решении логических уравнений с двумя переменными. - Деп. в УкрИНТЭИ, № 134 - Ук92, 1992, 12 с.

#### Sharonova N. V. Comparator Identification of Linguistic Objects

The present thesis is a manuscript to compete for earning a doctoral degree in technical sciences, the speciality 05.25.05 - the Information Systems and Processes, Kharkov State Institute of Culture, Kharkov, 1994.

52 scientific works and 1 monography, containing a theoretical generalization of the comparator identification methodology when studying and simulating the human linguistic behaviour with a view to apply the models in the intelligence systems creation, are defended. It has been established that the given method gives the possibility to create the natural language relations models, it is efficient in solving the most important problems of the computer-aided processing of the texts especially those connected with the problem of the texts' sense comprehension. The program complexes for solving a member of Russian, Ukrainian, English texts processing problems are offered, the data on efficiency of their introduction are cited.

Шаронова Н.В. Компаративная идентификация лингвистических объектов.

Диссертация является рукописью на соискание ученой степени доктора технических наук по специальности 05.25.05 – информационные системы и процессы, Харьковский государственный институт культуры, Харьков, 1994 .

Защищается монография и 52 научные работы, которые содержат теоретическое обобщение методологии компаративной идентификации при исследовании и моделировании языкового поведения человека с целью использования моделей при построении интеллектуальных систем. Установлено, что предлагаемый метод дает возможность построения адекватных моделей естественных языковых отношений, эффективен при решении важных задач автоматизированной обработки текстов, особенно связанных с проблемой понимания смысла текстов. Предложены программные комплексы для решения ряда задач обработки текстов русского, украинского, английского языков, приводятся данные об эффективности их внедрения.

Ключові слова:

компаративна ідентифікація, моделювання мовної поведінки людини, природномовні відношення, автоматизована обробка текстів, розуміння.

Відповідальний за випуск

А.Ф.Воловик

Підписано до друку 3.10.94. Формат 60x84 1/16.Папір для мн.ап.

Друк офс. Ум. арк. 2,56. Облік.-вид. арк. 2,0. Зам. № 8

Тираж 100 прим.

Ротапринт Харківського державного інституту культури

310003, Харків – 3, Бурсацький узвіз, 4

AB 31.137

**AB 31.137**