

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

ФЕДОРОВА ТЕТЯНА МИКОЛАЇВНА

УДК 519.7:007.52; 519.711.3

АЛГЕБРО – ЛОГІЧНІ МОДЕЛІ ТА МЕТОД ПОБУДОВИ ЛАНЦЮГІВ
ЛЕКСИЧНИХ ОДИНИЦЬ В СИСТЕМАХ ШТУЧНОГО ІНТЕЛЕКТУ

05.13.23 – системи та засоби штучного інтелекту

Автореферат дисертації на здобуття наукового ступеня
кандидата технічних наук

Харків – 2013

Дисертацією є рукопис.

Робота виконана в Харківському національному університеті радіоелектроніки Міністерства освіти і науки України.

Науковий керівник: доктор технічних наук, професор
Четвериков Григорій Григорович, Харківський національний університет радіоелектроніки, професор кафедри програмної інженерії.

Офіційні опоненти: доктор технічних наук, професор,
Асєєв Георгій Георгійович, Харківська державна академія культури, завідувач кафедри інформаційних технологій;

доктор технічних наук, професор,
Гороховатський Володимир Олексійович, Харківський інститут банківської справи Університету банківської справи Національного банку України, завідувач кафедри інформаційних технологій.

Захист відбудеться «___» _____ 2013 р. о ___ годині на засіданні спеціалізованої вченої ради Д 64.052.01 в Харківському національному університеті радіоелектроніки за адресою: 61166, м. Харків, пр. Леніна, 14.

З дисертацією можна ознайомитися у бібліотеці Харківського національного університету радіоелектроніки за адресою: 61166, м. Харків, пр. Леніна, 14.

Автореферат розісланий «___» _____ 2013 р.

Вчений секретар
спеціалізованої вченої ради

Є. І. Литвинова

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. З появою інформаційно – інтелектуальних комп'ютерних технологій виник феномен спілкування людини з неживою істотою безпосередньо засобами природної людської мови. Яким би скептичним не було ставлення до «розумових» можливостей обчислювальної машини, не можна не зважати на факт, що мовні реакції сучасного комп'ютера у деяких випадках уже неможливо відрізнити від реакцій людини. Можна прогнозувати, що через деякий час комп'ютерам буде передано істотну частину мовних функцій людини, що створить реальні передумови для побудови природномовних інформаційно – інтелектуальних комп'ютерних систем. У сучасних умовах постійного зростання інформаційних потоків проблема автоматичного оброблення мовної інформації набуває все більшого значення.

Моделювання природної мови людини є важливою проблемою теорії інтелекту. Центральними об'єктами опису природномовного тексту є слова. Для мов з розвиненою словозміною завдання граматичної ідентифікації є складним і трудомістким, а тому потребує автоматизації словозмінних процесів.

У зв'язку з цим необхідно являється розробка комп'ютерних програм автоматичної граматичної ідентифікації лексичних одиниць, здатних самостійно, без участі людини, приписувати текстовим словоформам параметри граматичного значення. Створення таких засобів неможливе без максимально формалізованих словозмінних класифікацій.

На сьогоднішній день процес обробки даних в інтелектуальних системах ґрунтується на конкретних знаннях про предметну область. Однак самі системи подання знань, які моделюють діяльність людини, недостатньо формалізовані. Чим більш узагальнений характер носить мова подання знань, тим на більшій різноманітності предметних областей вона може бути реалізована. Такою універсальною мовою представлення знань можна вважати мову алгебри скінченних предикатів, оскільки вона дозволяє формалізувати найрізноманітніші відношення. Розвиток засобів цієї алгебри привів до появи нового виду мереж, які отримали назву логічних.

Останнім часом поряд з традиційними лексико – граматичними дослідженнями в лінгвістиці все більшої актуальності набуває питання побудови формалізованих словозмінних класифікацій. Це пов'язано як з теоретичними задачами граматики по різноплановому та глибокому опису мовних систем, так і з необхідністю побудови ефективних лінгвістичних технологій. Основною вимогою до опису мовної системи, орієнтованої на застосування в лінгвістичних технологіях, є максимальна формалізація цього опису, доведення результатів відповідних досліджень до рівня формальних моделей і лінгвістичних алгоритмів, адаптованих до використання у сучасних мовно – інформаційних системах.

В рамках запропонованої тематики дисертаційного дослідження передбачається розробка алгебро – логічних моделей лексикографічних систем та їх комп'ютерна реалізація. Проведений аналіз засвідчує **актуальність** застосування отриманих моделей та методу для розробки систем спілкування з комп'ютером природною мовою, а також для підвищення ефективності виводу в автоматизованих

інформаційних системах. Дослідження, проведені в роботі, є продовженням робіт авторів Апресяна Ю. Д., Бондаренка М. Ф., Палагіна О. В., Поспелова Д. О., Шабанова-Кушнарєнка Ю. П., Широкова В. А., Рассела С., Норвіга П., Люгер Джорджа Ф. та багатьох інших вітчизняних і закордонних вчених.

Зв'язок роботи з науковими програмами, планами, темами. Розробка основних положень дисертації здійснювалась відповідно до планів НДР, що виконувались на кафедрі програмної інженерії Харківського національного університету радіоелектроніки: госпдоговірна тема № 10-37 «Розроблення лексикографічної системи українсько-російсько-англійського словника з електроніки та інформатики» (№ ДР 0111U002636) та держбюджетна НДР № 264 «Розробка теорії та принципів побудови мозкоподібних структур» (№ ДР 0112U000205), розділ 264-1 «Моделювання мозкоподібних структур».

Мета і завдання дослідження. Метою дисертаційної роботи є розробка математичних моделей у вигляді логічних мереж словозміни української та італійської мов, математичної моделі словозмінної класифікації частин мови італійської мови та подальший розвиток методу побудови ланцюгів лексичних одиниць української мови для підвищення ефективності обробки інформації в природномовних системах.

Для досягнення цієї мети в межах дисертаційної роботи необхідно вирішити наступні задачі:

— проаналізувати наукові досягнення в області формалізації структур природної мови;

— розробити математичну модель словозмінної класифікації частин мови на основі аналізу системи ознак, що визначають словозміни зазначених класів слів та їх значення в італійській мові;

— розробити модель словозміни дієслів української мови у вигляді логічної мережі з урахуванням виявлених граматичних ознак;

— розробити модель словозміни іменників італійської мови у вигляді логічної мережі з урахуванням виявлених граматичних особливостей;

— розвинути метод знаходження n -го лінійного логічного перетворення для побудови ланцюгів лексичних одиниць української мови;

— здійснити програмно-алгоритмічну реалізацію розроблених моделей та розвинутого методу.

Об'єкт дослідження — процеси обробки інформації в природномовних системах, зокрема, процедура формального опису природної мови.

Предмет дослідження — алгебро-логічні моделі та метод побудови ланцюгів лексичних одиниць в системах штучного інтелекту.

Методи дослідження: алгебра скінченних предикатів, алгебра предикатних операцій та теорія моделей — для отримання формального опису систем граматичних ознак української та італійської мов; теорія множин, лінгвістичний та морфологічний аналізи, комп'ютерне моделювання мовних процесів — для побудови математичної моделі словозмінної класифікації частин мови італійської мови, для подальшого розвитку методу знаходження n -го лінійного логічного перетворення для побудови ланцюгів лексичних одиниць.

Наукова новизна отриманих результатів:

1. Вперше запропоновано математичну модель словозмінної класифікації частин мови, яка характеризується системою ознак (граматичних категорій), які визначають словозміни зазначених класів слів та їх значення в італійській мові, що дає можливість уникнути неоднозначності в результатах морфологічного аналізу за рахунок введення додаткових класів слів при побудові автоматизованих електронних перекладних словників.

2. Вперше запропоновано модель флексійної обробки дієслів української мови логічною мережею, яка характеризується системою бінарних відношень та формулами відповідних предикатів, що дає можливість реалізовувати її на програмно – алгоритмічному рівні та використовувати для автоматичного вирішення задач аналізу, синтезу та нормалізації словоформи.

3. Набув подальшого розвитку метод знаходження n -го лінійного логічного перетворення для побудови ланцюгів лексичних одиниць в лексикографічній системі електронних тлумачних словників шляхом визначення початкової семантичної залежності на кожному етапі обчислення, що дає можливість уникнути тавтології та циклічних посилань при корегуванні словникових статей в лексикографічній системі тлумачного типу.

Практичне значення отриманих результатів.

1. Запропоновані модель словозмінної класифікації частин мови італійської мови та модель флексійної обробки дієслів української мови логічною мережею, а також метод знаходження n -го лінійного логічного перетворення для побудови ланцюгів лексичних одиниць в лексикографічній системі електронних тлумачних словників доведено до програмної реалізації, що дозволило експериментально підтвердити ефективність запропонованих підходів.

2. Розроблені в дисертаційній роботі моделі призначені для формалізації структури інформаційних скінченних об'єктів незалежно від їх природи. Моделі у вигляді логічної мережі можуть бути використані в системах штучного інтелекту, що, зокрема, відкриває нові перспективи в моделюванні процесів. Ці процеси відбуваються в системах, які мають мережеву структуру. Мережа виконує функції бази даних. У ролі атрибутів бази даних в мережі виступають предметні змінні, представлені полюсами мережі. У ролі доменів атрибутів бази даних в мережі виступають області зміни предметних змінних мережі. У ролі таблиць, що зберігаються в базі даних, виступають бінарні відношення, представлені гілками мережі. Результати використовуються при розробці математичних моделей мережі системи водопостачання та застосовуються у роботі аналітичного відділу в ПАТ АК «Київводоканал».

3. Набув подальшого розвитку метод знаходження n -го лінійного логічного перетворення для побудови ланцюгів в лексикографічній системі електронних тлумачних словників шляхом завдання початкової семантичної залежності на кожному етапі обчислення. На базі цього методу було розроблено програмний додаток «Побудова гіперланцюгів», призначений для опису семантичних відношень між лексичними одиницями природної мови, впроваджений в Українському мовно-інформаційному фонді НАН України. Метою створення програми є удосконалення процесу автоматизованої екстракції та формалізації обробки природномовних

текстів за допомогою аналізу тлумачних відношень та побудови ланцюгів між лексичними одиницями української мови. Використання програмного додатку дозволяє автоматизувати обробку словникових статей за рахунок введення відношень тлумачення, уникати повторів та знаходити в тексті або фрагменті тексту слово не тільки за конкретно заданим значенням, але й за його змістом.

4. Отримані в процесі досліджень наукові висновки та положення дисертації є обґрунтованими і достовірними. Обґрунтованість підтверджується використанням в навчальному процесі Харківського національного університету радіоелектроніки в рамках дисциплін «Високорівневі технології обробки мовної інформації» та «Дискретні структури» для студентів напрямку «Програмна інженерія» та в рамках дисциплін «Інформатика» та «Основи дискретної математики» для студентів напрямку «Видавничо-поліграфічна справа» (акт впровадження від 31.08.2012 р.)

5. Результати дисертації у складі моделей, методу і програмних додатків впроваджено в організаціях: ПАТ АК «Київводоканал» (акт впровадження від 17.08.2012 р.), Український мовно-інформаційний фонд НАН України (акт впровадження від 11.09.2012 р.).

Особистий внесок здобувача. Основні результати теоретичних і експериментальних досліджень отримані здобувачем особисто. У роботах, опублікованих зі співавторами, здобувачеві належать: [1] – пошук сталого режиму при додаванні в мережу активного елемента і опис роботи активних елементів в системі з кількома входами; [2] – інтерпретація методу розшарування предиката, що описує відношення між першою, другою і третьою буквами закінчення; [3] – досліджено методи реалізації морфологічного аналізу: словниковий (декларативний), що використовується для аналізу мов із нерозвинутим відмінюванням слів (англійська, французька, тощо), та алгоритмічний (процедурний) – для аналізу мов із розвинутим відмінюванням слів (наприклад, українська); [4] – побудова функціональних моделей для опису задач та процесів обробки мовної інформації в корпусній лінгвістиці, машинному перекладі та системах пошуку та класифікації; приклади автоматичного розпізнавання символів, автоматичного реферування текстів, створення електронних словників, тезаурусів, онтологій; [5] – модель флексійної обробки логічною мережею дієслів української мови, яка характеризується системою бінарних відношень та формулами відповідних предикатів (мережа призначена для розв'язання системи рівнянь, що задаються відповідною моделлю); [6] – постановка задачі побудови формалізованих словозмінних класифікацій на прикладі італійської мови; [7] – подальший розвиток методу знаходження n -го лінійного логічного перетворення для побудови семантичних ланцюгів між лексичними одиницями української мови в лексикографічній системі електронних тлумачних словників; наведено опис програмного засобу «Побудова гіперланцюгів», який дозволяє будувати, редагувати та аналізувати ланцюги лексичних одиниць; [8] – дослідження математичного опису вхідних/вихідних даних та їх проміжних перетворень відповідною АСП-структурою третього роду; [9] – аналіз структурних елементів комп'ютерної лінгвістики та шляхи побудови функціональних моделей для опису задач та процесів обробки мовної інформації в корпусній лінгвістиці, машинному перекладі та системах пошуку та класифікації.

Апробація результатів дисертації. Основні результати дисертаційної роботи доповідалися, обговорювалися і були схвалені на Міжнародних наукових конференціях «Горизонти прикладної лінгвістики і лінгвістичних технологій» MegaLing'2009, MegaLing'2010, MegaLing'2011 (Київ, 2009 р.; Крим, Партеніт, 2010 р.; Київ, 2011 р.); Міжнародній науковій молодіжній школі «Системи та засоби штучного інтелекту (СЗШІ-2009)» (Донецьк, 2009 р.); 2-й факультетській науково-практичній молодіжній школі – семінарі «Інформаційні інтелектуальні системи» (Харків, 2009 р.); Всеукраїнській науково – методичній конференції студентів та молодих науковців «Прикладна лінгвістика 2010: проблеми і рішення» (Миколаїв, 2010 р.); Міжнародній науково – технічній конференції SAIT 2011 (Київ, 2011 р.), а також на Міжнародній науково – технічній конференції «Искусственный интеллект. Интеллектуальные системы ИИ-2012» (Донецьк, 2012 р.).

Публікації. Основні результати дисертаційної роботи надруковано у 17 наукових працях, зокрема, 7 статей у виданнях, що входять до переліку наукових фахових видань України (з них 2 – одноосібно), 10 публікацій у матеріалах і тезах доповідей міжнародних наукових конференцій та збірнику наукових праць «Вісник ХНАДУ» (з них 3 – одноосібно).

Структура дисертації. Дисертація складається із вступу, п'ятих розділів, висновків, списку використаних джерел і двох додатків. Повний обсяг дисертації складає 178 сторінок, обсяг основного тексту 144 сторінки, 62 рисунка, 9 таблиць; список використаних джерел, що містить 147 найменувань та займає 15 сторінок; два додатки на 5 сторінках.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** обґрунтовано вибір та актуальність теми дисертаційної роботи, сформульовано мету та завдання дослідження, наведено відомості про зв'язки обраного напрямку досліджень із планами організації, де виконана робота. Надано загальну характеристику роботи, сформульовано основні положення, які винесено автором на захист, визначено їх практичну значущість.

У **першому розділі** проаналізовано сучасний стан розвитку систем опрацювання природної мови. Запити інтелектуалізації потребують від комп'ютера здатності розуміти людську мову, достатньої «кмітливості», здатності мислення й творчості. Відтак висувається задача забезпечення обчислювальної техніки достатньо «високоорганізованим інтелектом», що базується на апаратних і програмних засобах, наближених за своїми можливостями до природної мови людини.

Ця задача не передбачає простого й швидкого вирішення. Виникають суттєві складнощі. Через це не дають бажаних результатів роботи з автоматизації програмування, створення високоякісного перекладу з однієї мови на іншу, побудови систем для розпізнавання досить складних об'єктів, формування здатності машин зрозуміти побачене й почуте, спрямоване на вирішення аналогічних питань.

Наведено огляд методів та систем морфологічного аналізу в системах обробки природної мови. Для мов флективного типу, які мають багату словозмінну парадигму, серед задач граматичної ідентифікації складним і працемістким

завданням є формалізація процесів лексичної парадигматизації та отримання повного списку словоформ певної лексеми в усіх граматичних значеннях і виведення вихідної словоформи з будь-якої текстової словоформи.

При здійсненні морфологічного аналізу виділяють кілька напрямів. Один з них моделює класичну схему аналізу шляхом поділу словоформи на основу і можливе закінчення з подальшою перевіркою на сумісність закінчення з основою, яка залишається. Інший напрям використовує інформацію, що міститься в кінцевих буквосполученнях. Ця інформація отримується в результаті попередньої статистичної обробки словника. На третьому напрямі створюються універсальні математичні моделі морфології у формі відкритих систем рівнянь, які дозволяють шляхом їх рішення здійснювати нормалізацію словоформ, отримання граматичної інформації та синтез словоформ.

У результаті проведеного аналізу сформульовано основні задачі дисертаційних досліджень.

У **другому розділі** наведено систему ознак (граматичних категорій), що визначають словозміну зазначених класів слів італійської мови, та значення, які ці ознаки (категорії) можуть приймати.

Алгебра скінченних предикатів дає можливість формального опису інформаційних процесів у вигляді рівнянь. Рівняння описують відношення між змінними. Усі змінні у рівнянні рівноправні, будь-які з них можуть виступати як у ролі незалежних, так і у ролі залежних. Перевага рівняння полягає у можливості передбачення реакції системи навіть при неповній визначеності вхідних сигналів.

Мова – це система знаків, де мовний знак часто відображають у вигляді трикутника Фреге. Таким чином, об'єктами мови виступають певні психофізичні стани та процеси, які відбуваються у мовнорозумовому апараті людини, а усна та писемна її репрезентації слугують елементами інфраструктури мовного процесу.

Наявність формальної моделі словозміни флективної мови є визначальним фактором для побудови граматичного словника, що означає встановлення та формалізацію лінгвістичних критеріїв, згідно з якими уся множина слів мови розбивається на певні підмножини, взаємний перетин яких є порожнім і всередині кожної з них словозміна відбувається за одним правилом. Підмножини слів з такими властивостями називаються словозмінними парадигматичними класами.

За ознакою приналежності до певної частини мови та за додатковими ознаками, які є класифікуючими в межах певної частини мови, множина слів розподіляється на граматичні класи.

За словозмінними категоріями, що визначають словозмінну парадигму конкретних слів (сукупність граматичних значень та відповідних граматичних форм), вводяться парадигматичні типи. Парадигматичні типи можуть бути притаманними декільком граматичним класам, усередині кожного з яких виділяються парадигматичні класи. До одного парадигматичного класу входять тільки ті слова, які мають однакові набори змінних складових для всіх граматичних форм, а відрізняються один від одного лише незмінною складовою. Слова з одного класу еквівалентності, визначеного в такий спосіб, мають і однакові правила словозміни.

Таким чином, для кожного з граматичних класів (парадигматичних типів) будується розбиття на множини слів, що не перетинаються, і які є парадигматичними класами, всередині кожного з яких діють єдині правила словозміни. Для мов флективного типу це означає однаковість флексій граматичних форм та збіг характеру чергування в основі.

Парадигматичні класи італійської мови визначаються всередині кожного граматичного класу слів. Це викликано тим, що словозмінні процеси для слів різних класів визначаються своїми, притаманними саме для цього класу, словозмінними параметрами.

Словниковий склад ділиться за належністю до певної частини мови на такі класи слів – іменники (ознаки: рід, число), артиклі (ознаки: вид, рід, число), прикметники (ознаки: тип, рід, число), числівники (ознаки: тип, рід, число), займенники (ознаки: тип, рід, число, особа), дієслова (ознаки: особа, число, спосіб, час, стан), прийменники (ознака: тип), прислівники (ознаки: тип, вид), сполучники (ознака: тип), вигуки (ознака: вид).

За значеннями своїх класифікаційних категорій деякі граматичні класи розподіляються на підкласи. Кількість підкласів залежить від того, скільки значень може приймати дана класифікаційна категорія. Наприклад, клас артиклів розподіляється за ознакою *вид* на 2 підкласи; клас прикметників за *типом* поділяється на 6 підкласів; клас числівників за *типом* поділяється на 7 підкласів; клас займенників за *типом* поділяється на 5 підкласів.

У **третьому розділі** побудовано модель лексикографічної системи граматичного типу на прикладі словозміни італійської мови за допомогою алгебри предикатів.

Нехай L – множина усіх слів італійської мови. Розіб'ємо множину L на підмножини (класи слів), що взаємно не перетинаються:

$$L(x) = I_1 \vee I_2 \vee I_3 \vee I_4 \vee I_5 \vee I_6 \vee I_7 \vee I_8 \vee I_9 \vee I_{10}, \quad (1)$$

де I_1 – іменники, I_2 – артиклі, I_3 – прикметники, I_4 – числівники, I_5 – займенники, I_6 – дієслова, I_7 – прийменники, I_8 – прислівники, I_9 – сполучники, I_{10} – вигуки.

Омонімію будемо вважати знятою, а омоніми промаркованими, так що $I_{n_i} \wedge I_{n_j} = \emptyset$ при $n_i \neq n_j$, $i, j = \overline{1,10}$.

Деякі класи слів за своїми класифікаційними ознаками поділяються на підкласи. Підкласи будемо позначати x_n^k , де n – номер класу слів, k – індекс підкласу слів.

Позначимо $x_1^i = I_1$, $x_{22}^d = I_6$, $x_{23}^{ppm} = I_7$, $x_{24}^{ppc} = I_8$, $x_{25}^c = I_9$, $x_{26}^b = I_{10}$.

Артиклі за значенням граматичної ознаки *вид* поділяються на два підкласи:

$I_2^b \vee I_2^h = I_2$, де I_2^b – визначений артикль, I_2^h – невизначений артикль.

Підкласи $x_2^b = I_2^b$, $x_3^h = I_2^h$ будемо називати граматичними підмножинами артиклів.

Прикметники за значенням граматичної ознаки *тип* поділяються на наступні підкласи:

$1_3^{\text{як}} \vee 1_3^{\text{пр}} \vee 1_3^{\text{вк}} \vee 1_3^{\text{н}} \vee 1_3^{\text{п}} \vee 1_3^{\text{окл}} = 1$, де $1_3^{\text{як}}$ – якісні прикметники, $1_3^{\text{пр}}$ – присвійні прикметники, $1_3^{\text{вк}}$ – вказівні прикметники, $1_3^{\text{н}}$ – невизначені прикметники, $1_3^{\text{п}}$ – питальні прикметники, $1_3^{\text{окл}}$ – окличні прикметники.

Підкласи $x_4^{\text{як}} = 1_3^{\text{як}}$, $x_5^{\text{пр}} = 1_3^{\text{пр}}$, $x_6^{\text{вк}} = 1_3^{\text{вк}}$, $x_7^{\text{н}} = 1_3^{\text{н}}$, $x_8^{\text{п}} = 1_3^{\text{п}}$, $x_9^{\text{окл}} = 1_3^{\text{окл}}$, будемо називати граматичними підмножинами прикметників.

Числівники за значенням граматичної ознаки *тип* поділяються на наступні підкласи:

$1_4^{\text{к}} \vee 1_4^{\text{п}} \vee 1_4^{\text{інк}} \vee 1_4^{\text{м}} \vee 1_4^{\text{др}} \vee 1_4^{\text{кол}} \vee 1_4^{\text{дистр}} = 1$, де $1_4^{\text{к}}$ – кількісні числівники, $1_4^{\text{п}}$ – порядкові числівники, $1_4^{\text{інк}}$ – іменники неточної кількості (*una decina* (десяток), *una dozzina* (дюжина)), $1_4^{\text{м}}$ – мультиплікативні числівники (*il doppio* (в два рази), *il quadruplo* (чотири рази)), $1_4^{\text{др}}$ – дробові числівники (*un mezzo* (половина), *due terzi* (дві треті)), $1_4^{\text{кол}}$ – колективні числівники (*coppia* (пара), *trimestre* (триместр)), $1_4^{\text{дистр}}$ – дистрибутивні числівники (*due a due* (по двоє), *sei per volta* (по шість разів)).

Підкласи $x_{10}^{\text{к}} = 1_4^{\text{к}}$, $x_{11}^{\text{п}} = 1_4^{\text{п}}$, $x_{12}^{\text{інк}} = 1_4^{\text{інк}}$, $x_{13}^{\text{м}} = 1_4^{\text{м}}$, $x_{14}^{\text{др}} = 1_4^{\text{др}}$, $x_{15}^{\text{кол}} = 1_4^{\text{кол}}$, $x_{16}^{\text{дистр}} = 1_4^{\text{дистр}}$ будемо називати граматичними підмножинами числівників.

Займенники за значенням граматичної ознаки *тип* поділяються на наступні підкласи:

$1_5^{\text{о}} \vee 1_5^{\text{вк}} \vee 1_5^{\text{н}} \vee 1_5^{\text{в}} \vee 1_5^{\text{п}} = 1$, де $1_5^{\text{о}}$ – особисті займенники, $1_5^{\text{вк}}$ – вказівні займенники, $1_5^{\text{н}}$ – невизначені займенники, $1_5^{\text{в}}$ – відносні займенники, $1_5^{\text{п}}$ – питальні займенники.

Підкласи $x_{17}^{\text{о}} = 1_5^{\text{о}}$, $x_{18}^{\text{вк}} = 1_5^{\text{вк}}$, $x_{19}^{\text{н}} = 1_5^{\text{н}}$, $x_{20}^{\text{в}} = 1_5^{\text{в}}$, $x_{21}^{\text{п}} = 1_5^{\text{п}}$ будемо називати граматичними підмножинами прикметників.

В результаті для кожного слова $y \in L$ визначається приналежність до певного граматичного класу $F_m(x) = x_1 \vee \dots \vee x_m$, $m = \overline{1,26}$.

Всього маємо 26 граматичних класів:

$$L = \bigcup_{m=1}^{26} F_m; F_{m_i} \wedge F_{m_j} = \emptyset, \quad m_i \neq m_j, \quad i, j = \overline{1,26}.$$

У середині граматичних класів виділяємо парадигматичні класи. Це означає, що довільну лексему y представлено у вигляді комбінації незмінної та змінної складових:

$$y = c(y) * f(y), \quad (2)$$

де $c(y)$ – незмінна складова, $f(y)$ – змінна складова, $*$ – конкатенація.

Змінна складова крім суто флексії може включати частину основи, в якій відбувається чергування.

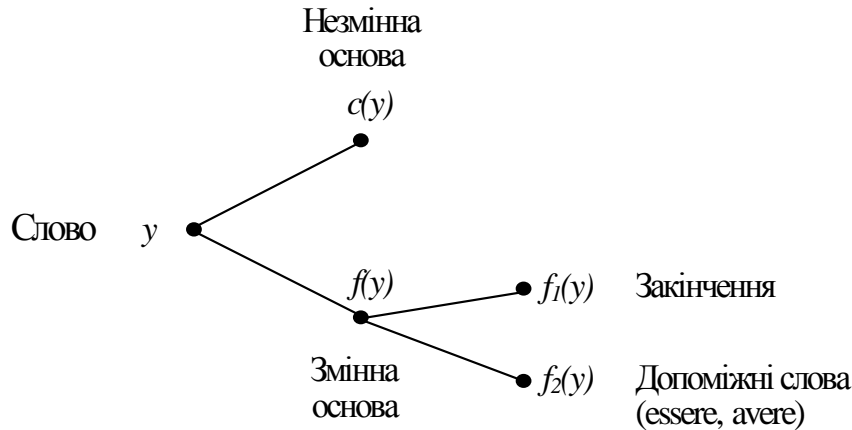
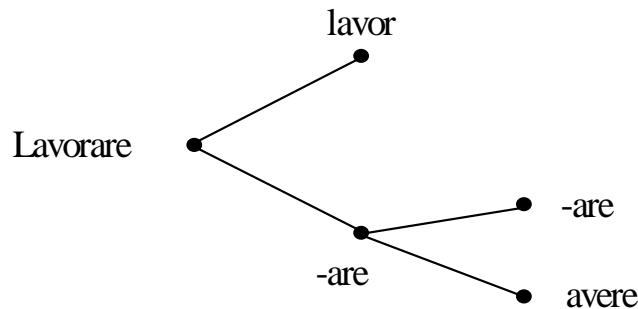


Рис. 1 – Модель словозмінної системи для італійської мови

Рис. 2 – Модель словозміни для дієслова «*lavorare*» – працювати

Також побудовано математичну модель іменників італійської мови. Вона характеризується системою бінарних відношень P_1 – P_4 , що задаються формулами відповідних предикатів. Утворюючи кон'юнкцію усіх цих предикатів, отримуємо предикат моделі:

$$P(x_1, x_2, q, z, y) = P_1(x_1, q) \wedge P_2(x_2, q) \wedge P_3(z, q) \wedge P_4(y, z). \quad (3)$$

Нехай $X = \{x_1, x_2\}$ – набір граматичних ознак, які характеризують вплив дальнього тексту на закінчення (x_1 – число зі значеннями одиниця (о), множина (м); x_2 – рід зі значеннями чоловічий (ч), жіночий (ж)); $Z = \{z\}$ – закінчення іменника, z може приймати наступні значення: z^o – іменники з закінченням -o, z^e – іменники з закінченням -e, z^a – іменники з закінченням -a, z^i – іменники з закінченням -i; $Y = \{y\}$ – останні літери основи, y може приймати наступні значення: y^{ist} – іменники з суфіксом -ist, y^{in} – іменники з суфіксом -ion, y^0 – всі інші іменники; q – номер впливу контексту.

Відношення $P_1(x_1, q)$ зв'язує змінні x_1 і q ; відношення $P_2(x_2, q)$ – змінні x_2 і q ; відношення $P_3(z, q)$ зв'язує номер впливу контексту закінчення z зі змінною q ; відношення $P_4(y, z)$ зв'язує змінну основи слова y зі змінною закінчення слова z .

Предикату моделі відповідає відношення моделі P , що зв'язує між собою предметні змінні x_1, x_2, q, z, y . На рис. 3 зображено логічну мережу для відношення моделі P .

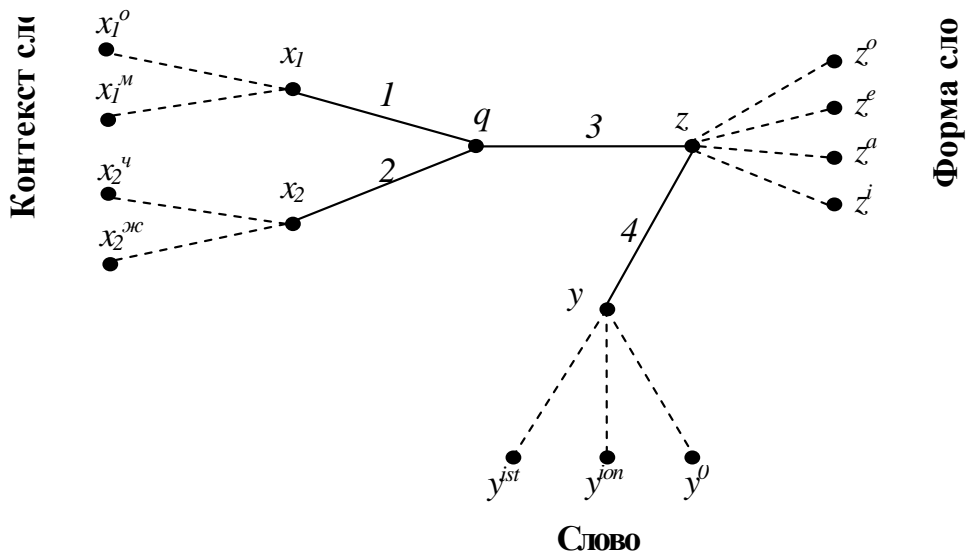


Рис. 3 – Логічна мережа словозміни іменників італійської мови

У побудованій мережі 5 полюсів і 4 гілки. Мережа має зовнішні і внутрішній полюси. До зовнішніх відносяться x_1, x_2, z, y , до внутрішнього – q . Зовнішні полюси можна об'єднати в три групи: 1) x_1 з значеннями $\{o, m\}$, x_2 з значеннями $\{ч, ж\}$, 2) y з значеннями $\{ist, ion, o\}$, 3) z з значеннями $\{o, e, a, i\}$. До першої групи входять полюси, значення яких характеризують контекст, що оточує слово. До другої групи входить полюс, що характеризує саме слово. До третьої групи входить полюс, що характеризує словоформу, відповідну до даного контексту і даного слова.

Кожну логічну мережу можна перетворити в електронну схему для автоматичного розв'язання деякого класу задач, обумовленого тією моделлю, для якої було побудовано дану мережу.

Електронну схему, яка відповідає мережі, що зображена на рис. 2, можна використовувати для автоматичного вирішення класу задач, що відносяться до відмінювання іменників італійської мови. Прикладами задач цього класу можуть бути наступні задачі:

- задача синтезу словоформи: задані слово і оточуючий його контекст, потрібно визначити відповідну їм словоформу;
- задача аналізу словоформи: задані форма слова і оточуючий її контекст, потрібно визначити граматичні ознаки, відповідні цій словоформі;
- задача завдання нормалізації словоформи: задані форма слова і оточуючий її контекст, потрібно відшукати слово, відповідне цій словоформі.

У четвертому розділі побудовано модель флексійної обробки дієслів української мови у вигляді логічної мережі.

Під час побудови математичної моделі флексійного оброблення мови описується фрагментне морфологічне відношення $L(X, Y)$, що відображає зв'язок між змістом X та змінним фрагментом (флексійним морфом) Y . У якості фрагменту із усіх морфем слова (префікс, корінь, суфікс, закінчення, постфікс) вибране закінчення. Основним типом граматичних афіксів в українській мові є флексія (закінчення), що відображає відношення слова до інших слів у словосполученні та реченні. Нехай $X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$ – набір граматичних ознак, які характеризують вплив дальнього тексту на закінчення; $Z = \{z_1, z_2, z_3\}$ – дієслівне закінчення, яке складається з першої, другої, третьої букв – z_1 , z_2 та z_3 відповідно; $Y = \{y_1, y_2, y_3, y_4\}$ – набір лексичних ознак; r – номер впливу контексту.

Для виявлення впливу контексту слова необхідно визначити значення граматичних ознак $x_1 - x_7$. Позначення граматичних ознак та їх значення наступні: x_1 – число зі значеннями одина (о), множина (м); x_2 – особа зі значеннями перша (1), друга (2), третя (3); x_3 – час зі значеннями минулий (м), неминулий (н); x_4 – рід зі значеннями чоловічій (ч), жіночій (ж), середній (с); x_5 – репрезентація зі значеннями особова форма (о), інфінітив (і); x_6 – спосіб зі значеннями дійсний (д), наказовий (н); x_7 – дієвідміна зі значеннями перша (1), друга (2).

Побудовано математичну модель для дієслів української мови. Вона характеризується системою бінарних відношень $P_1 - P_{16}$, що задаються формулами відповідних предикатів. Утворюючи кон'юнкцію всіх цих предикатів, отримуємо предикат моделі:

$$\begin{aligned} &P(x_1, x_2, x_3, x_4, x_5, x_6, x_7, r, z, z_1, z_2, z_3, y, y_1, y_2, y_3, y_4) = \\ &= P_1(x_1, r) \wedge P_2(x_2, r) \wedge P_3(x_3, r) \wedge P_4(x_4, r) \wedge P_5(x_5, r) \wedge P_6(x_6, r) \wedge \\ &\wedge P_7(x_7, r) \wedge P_8(z, r) \wedge P_9(z_1, r) \wedge P_{10}(z_2, r) \wedge P_{11}(z_3, r) \wedge \\ &\wedge P_{12}(y, z) \wedge P_{13}(y_1, z_1) \wedge P_{14}(y_2, z) \wedge P_{15}(y_3, z) \wedge P_{16}(y_4, z_2). \end{aligned} \quad (4)$$

Предикату моделі відповідає відношення моделі P , що зв'язує між собою предметні змінні $x_1, x_2, x_3, x_4, x_5, x_6, x_7, r, z, z_1, z_2, z_3, y, y_1, y_2, y_3, y_4$. Відношення P_1 зв'язує змінні x_1 і r ; відношення P_2 – змінні x_2 і r ; відношення P_3 – змінні x_3 і r ; відношення P_4 – змінні x_4 і r ; відношення P_5 – змінні x_5 і r ; відношення P_6 – змінні x_6 і r ; відношення P_7 – змінні x_7 і r ; відношення P_8 – номер впливу контексту r з закінченням z ; відношення P_9 – номер впливу контексту r з першою буквою закінчення z_1 ; відношення P_{10} – номер впливу контексту r з другою буквою закінчення z_2 ; відношення P_{11} – номер впливу контексту r з третьою буквою закінчення z_3 ; відношення P_{12} – закінчення z з впливом ближнього тексту y ; відношення P_{13} – першу букву закінчення z_1 з впливом ближнього тексту y_1 ;

відношення P_{14} – змінну z з впливом ближнього тексту y_2 ; відношення P_{15} – змінну z з впливом ближнього тексту y_3 ; відношення P_{16} – другу букву закінчення z_2 з впливом ближнього тексту y_4 .

Відношення моделі P можна наочно зобразити у вигляді мережі, яка називається логічною. Логічну мережу можна побудувати для будь-якої моделі. У нашому конкретному випадку отримуємо логічну мережу дієслів, що зображено на рис. 4.

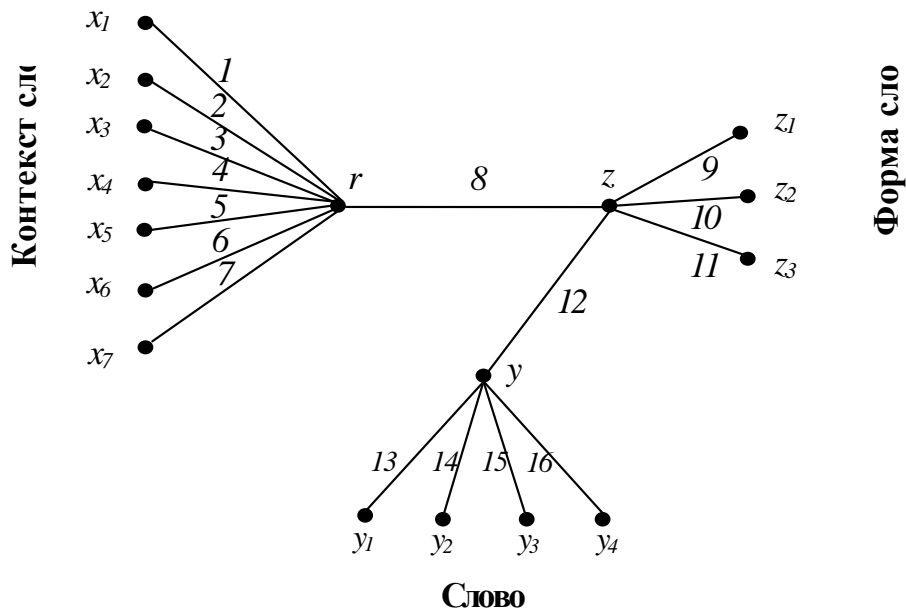


Рис. 4 – Логічна мережа словозміни дієслів української мови

Кожній гілці логічної мережі ставиться у відповідність своє бінарне відношення моделі, яке називається відношенням цієї гілки. Кожна гілка позначається номером свого відношення. Вона з'єднує два полюси, що відповідають тим предметним змінним, які зв'язуються відношенням, відповідним даній гілці. В побудованій мережі 17 полюсів і 16 гілок.

Полюси мережі діляться на два класи – зовнішні і внутрішні. Кожен зовнішній полюс з'єднаний лише з однією гілкою, кожен внутрішній – більш, ніж з однією. У побудованій мережі до зовнішніх відносяться полюси, помічені предметними змінними $x_1, x_2, x_3, x_4, x_5, x_6, x_7, z_1, z_2, z_3, y_1, y_2, y_3, y_4$, до внутрішніх – r, z, y . Зовнішні полюси мережі об'єднують в три групи: 1) $x_1, x_2, x_3, x_4, x_5, x_6, x_7$, 2) z_1, z_2, z_3 , 3) y_1, y_2, y_3, y_4 . У першу групу входять полюси, помічені тими предметними змінними, значення яких характеризують контекст, який оточує слово. У другу групу входять полюси, які характеризують словоформу, що відповідає даному контексту і даному слову. У третю групу входять полюси, що характеризують саме відмінювання слова.

Запропонована математична модель не орієнтована на рішення будь-якого одного виду граматичної обробки слова, а дозволяє розв'язувати як традиційні задачі аналізу, синтезу та нормалізації, так і різноманітні комбіновані задачі.

Наприклад, задачу з частково заданими вихідними даними, задачу знаходження та виправлення помилки в закінчені слова, задачу переходу від довільно взятої словоформи до іншої форми того ж слова за заданим набором значень граматичних ознак. Аналітично (тобто оперуючи з формулами) можна вирішити будь-яку систему рівнянь алгебри предикатів, що характеризує дану модель.

У п'ятому розділі набув подальшого розвитку метод знаходження n -го лінійного логічного перетворення для побудови ланцюгів лексичних одиниць в лексикографічній системі електронних тлумачних словників.

Розглянуто в загальному вигляді метод знаходження n -го лінійного логічного перетворення. Нехай $P(x)$, $Q(y)$ – предикати, $K(x, y)$ – бінарний предикат, що задає ядро лінійного логічного перетворення, M – множина, елементи якої є логічними векторами. Лінійні логічні перетворення можна представити у вигляді

$$Q(y) \Leftrightarrow \exists x \in M \left(P(x) \wedge K(x, y) \right). \quad (5)$$

При знаходженні n -го лінійного логічного перетворення, якщо виконується умова, що n та $n+1$ перетворення збігаються, то n -е логічне перетворення не зміниться далі. Таким чином, воно стабілізується на n -му кроці. Крім того, ядра логічних перетворень відповідних ступенів співпадають $K^{(n-1)} = K^{(n)} = K^{(n+1)}$.

Алгебра скінченних предикатів дозволяє формалізувати підхід до побудови ланцюгів лексичних одиниць таким чином: нехай $P(x)$ – слово, $K(x_{n-1}, x_n)$ – семантична залежність, що визначає функцію тлумачення, M – множина всіх слів в словарних статтях.

На першому кроці виділяємо слово $P(x_1)$ та формуємо ядро семантичної залежності $K_1(x_1, x_2)$. Отримуємо слово $P(x_2)$, яке пов'язане з вихідним словом $P(x_1)$ формулою $P(x_2) \Leftrightarrow \exists x_1 P(x_1) K_1(x_1, x_2)$.

На другому кроці задаємо $K_2(x_2, x_3)$ та отримуємо слово $P(x_3)$ за формулою $P(x_3) \Leftrightarrow \exists x_2 P(x_2) K_2(x_2, x_3)$.

На третьому кроці задаємо $K_3(x_3, x_4)$ та отримуємо слово $P(x_4)$ за формулою $P(x_4) \Leftrightarrow \exists x_3 P(x_3) K_3(x_3, x_4)$ і т.д. Обчислення триватиме до тих пір, поки не виконається умова закінчення побудови ланцюга, тобто поки в ланцюзі не з'являться два співпадаючих елементи:

$$x_n = x_i, \text{ якщо } \exists K_i(x_i, x_n), i = \overline{1, n-1}. \quad (6)$$

Наведемо в загальному вигляді формулу обчислення перетворення $P(x_n)$:

$$P(x_n) \Leftrightarrow \exists x_{n-1} P(x_{n-1}) K_{n-1}(x_{n-1}, x_n). \quad (7)$$

На рис. 5 зображено графічне представлення методу побудови ланцюгів лексичних одиниць, де x – слово, $x(i, j)$, i – номер рівня, j – індекс слова на рівні i .

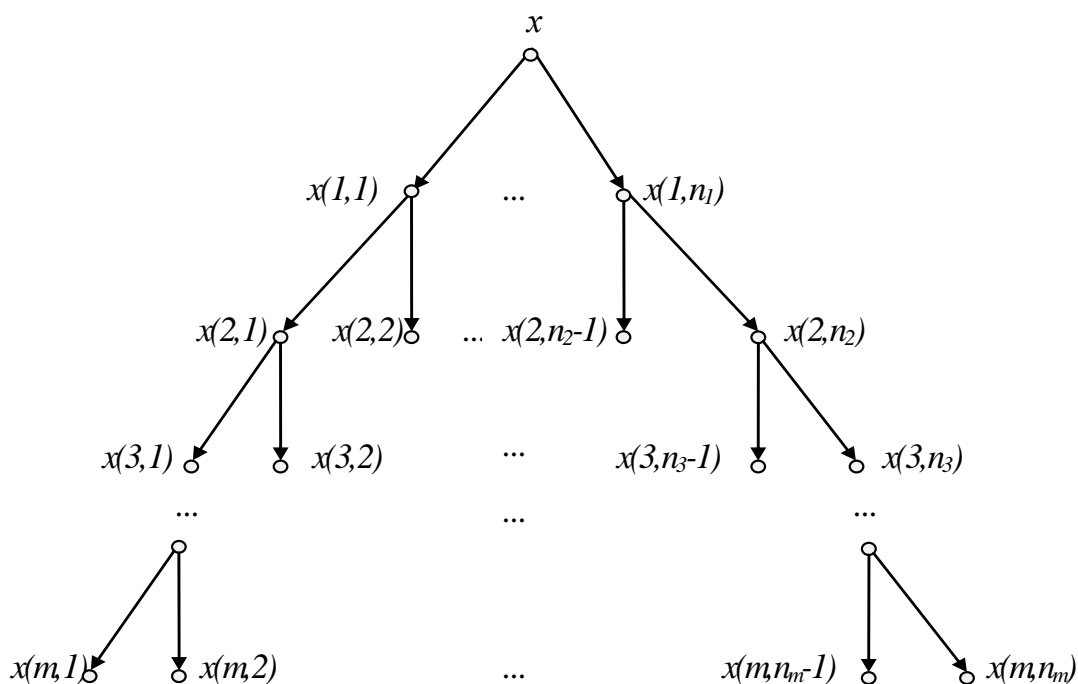


Рис. 5 – Графічне представлення методу побудови ланцюгів лексичних одиниць

На базі тлумачного словника української мови побудовано семантичну мережу, яка навчається. Тобто виконується автоматизована обробка словникових статей за допомогою аналізу відношень та побудови гіперланцюгів між лексичними одиницями української мови. Елементи ланцюгів зв'язують відношення двох типів: «тлумачиться через» або «за ознакою синонімії».

Розроблено програмний додаток «Побудова гіперланцюгів», призначений для опису семантичних відношень між лексичними одиницями української мови.

Для побудови гіперланцюгів, що зв'язуються відношенням «тлумачиться через», використано тлумачний словник «Віртуальна лексикографічна лабораторія Українського мовно-інформаційного фонду». Метою системи є автоматизація обробки текстів за допомогою аналізу відношень та побудови ланцюгів між лексичними одиницями української мови. Програма дозволяє будувати, редагувати та аналізувати ланцюги лексичних одиниць.

Програмний засіб використовує дві бази гіперланцюгів: в одній зберігаються ланцюги, розроблені дослідниками, а в іншій – ланцюги, які будуються автоматично за базою електронного тлумачного словника. Якщо для слова раніше вже було побудовано ланцюг за допомогою електронного словника, програма буде виводити результат, який раніше був збережений в базі електронного тлумачного словника, що дозволяє скоротити час побудови, особливо для слів, у яких кілька значень. Якщо ж користувачеві необхідно побудувати гіперланцюг знову, то можна задати параметр «примусовий пошук», і програма побудує новий ланцюг тільки за словником. Одне і те ж слово можна одночасно шукати в двох базах.

Використання програмного додатку «Побудова гіперланцюгів» дозволило автоматизувати обробку словникових статей за рахунок введення відношень

тлумачення, уникати повторів та знаходити в тексті або фрагменті тексту слово не тільки за конкретно заданим значенням, але і за його тлумаченням, характеристиками.

У першому додатку наведено акти впровадження результатів дисертаційної роботи.

У другому додатку наведено приклад побудови ланцюгів лексичних одиниць, елементи яких зв'язані відношенням «тлумачиться через».

ВИСНОВКИ

У дисертаційній роботі наведено результати, які відповідно до мети дослідження у сукупності є розв'язанням актуальної науково-практичної задачі підвищення ефективності обробки інформації в природномовних системах за рахунок використання моделі словозмінної класифікації частин мови італійської мови, логічних мереж словозміни дієслів української мови, іменників італійської мови та побудови ланцюгів лексичних одиниць української мови, що дозволяє зняти неоднозначність в тлумачних словниках. Автором одержано такі наукові та практичні результати:

1. Нова математична модель словозмінної класифікації частин мови італійської мови, яка характеризується системою ознак. Система ознак (граматичні категорії) визначає словозміни зазначених класів слів та їх значення. За рахунок введення додаткових класів слів модель дозволяє уникнути неоднозначності в результатах морфологічного аналізу при побудові автоматизованих електронних перекладних словників.

2. Нова модель флексійної обробки іменників італійської мови у вигляді логічної мережі, яка характеризується системою бінарних відношень, що задаються формулами відповідних предикатів. Модель описує фрагментне морфологічне відношення, що відображає зв'язок між змістом та змінним фрагментом тексту. Це дозволяє розширити клас задач, які розв'язуються за допомогою логічних мереж.

3. Нова модель флексійної обробки дієслів української мови у вигляді логічної мережі, що характеризується системою бінарних відношень та формулами відповідних предикатів. Це дозволяє реалізовувати дану модель на програмно – алгоритмічному рівні та використовувати її для автоматичного вирішення задач аналізу, синтезу та нормалізації словоформи. Були виділені наступні складові: набір граматичних ознак, дієслівні закінчення, набір лексичних ознак. Математичним апаратом моделювання є алгебра скінченних предикатів.

4. Набув подальшого розвитку метод знаходження n -го лінійного логічного перетворення для побудови ланцюгів в лексикографічній системі електронних тлумачних словників. Цей метод характеризується завданням початкової семантичної залежності на кожному етапі обчислення. Це дозволяє уникнути тавтології та циклічних посилань для корегування словникових статей в лексикографічній системі тлумачного типу. Застосування математичного апарату алгебри скінченних предикатів дозволило представити побудову ланцюгів лексичних одиниць не тільки в формульному, але і в більш наглядному, схематичному вигляді.

5. Результаты диссертационной работы нашли применения в Мовно-інформаційному фонді НАН України для автоматизації процесу обробки текстів за рахунок аналізу відношень тлумачення й побудови ланцюгів між лексичними одиницями української мови та в ПАТ АК «Київводоканал» при розробці математичних моделей мережі системи водопостачання в роботі аналітичного відділу. Результати дисертації використано в навчальному процесі кафедри програмної інженерії Харківського національного університету радіоелектроніки в рамках дисциплін «Високорівневі технології обробки мовної інформації» та «Дискретні структури» для студентів напрямку «Програмна інженерія» та в рамках дисциплін «Інформатика» та «Основи дискретної математики» для студентів напрямку «Видавничо-поліграфічна справа».

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

1. Четвериков, Г. Г. Моделирование технологических процессов управления активными элементами [Текст] / Г. Г. Четвериков, Т. Н. Федорова // Бионика интеллекта. – 2008. – №2 (69). – С. 84–88.

2. Четвериков, Г. Г. Расслоение предикатов на примере словоизменения прилагательных русского языка [Текст] / Г. Г. Четвериков, И. Д. Вечирская, Т. Н. Федорова // Искусственный интеллект. – 2009. – №3. – С. 170–177.

3. Моделювання фрагментів природної мови алгебро-логічними засобами (на прикладі k-значної реалізації морфологічного аналізу) [Текст] / Г. Г. Четвериков, Т. М. Федорова, І. Д. Вечірська, Д. Ф. Богданов // Бионика интеллекта. – 2009. – №1 (70). – С. 55–59.

4. Четвериков, Г. Г. Побудова функціональних моделей елементів комп'ютерної лінгвістики [Текст] / Г. Г. Четвериков, І. Д. Вечірська, Т. М. Федорова // Бионика интеллекта. – 2010. – №1 (72). – С. 83–88.

5. Федорова, Т. М. Побудова логічної мережі для флективної обробки дієслів української мови [Текст] / Т. М. Федорова // Бионика интеллекта. – 2012. – №1 (78). – С. 111–115.

6. Четвериков, Г. Г. Математическое описание словоизменительной классификации грамматических категорий в итальянском языке [Текст] / Г. Г. Четвериков, И. Д. Вечирская, Т. Н. Федорова // Искусственный интеллект. – 2012. – №4. – С. 264–271.

7. Федорова, Т. Н. О подходе к построению цепочек лексических единиц украинского языка в лексикографической системе электронного толкового словаря [Текст] / Т. Н. Федорова // АСУ и приборы автоматики : сб. науч. тр. – Х. : ХНУРЕ, 2012. – Вып. 159. – С. 33–40.

8. Chetverikov, G. G. Formal models of language systems [Text] / G. G. Chetverikov, I. D. Vechirskaya, T. M. Fedorova // Вестн. ХНАДУ : сб. науч. тр. – 2009. – Вып. 25. – С. 234–237.

9. Четвериков, Г. Г. Структура элементов в компьютерной лингвистики и пути ее моделирования [Текст] / Г. Г. Четвериков, И. Д. Вечирская, Т. Н. Федорова //

Прикладна лінгвістика та лінгвістичні технології (MegaLing'2010) : зб. наук. праць. – К. : Довіра, 2010. – С. 206–214.

10. Четвериков, Г. Г. Концептуально-методологічний підхід до моделювання природної мови алгебро-логічними засобами [Текст] / Г. Г. Четвериков, Т. М. Федорова, І. Д. Вечірська та ін. // Горизонты прикладной лингвистики и лингвистических технологий (MegaLing'2009) : материалы междунар. науч. конф., 20–27 сент. 2009 г., Киев. – К., 2009. – С. 234–237.

11. Четвериков, Г. Г. Побудова ієрархії дескрипторів на прикладі іменників української мови [Текст] / Г. Г. Четвериков, Т. М. Федорова // Системы и средства искусственного интеллекта (ССИИ-2009) : материалы междунар. науч. молодеж. шк., 28 сент. – 3 окт. 2009 г., пос. Дивноморское, Россия. – Донецк, 2009. – С. 180–182.

12. Четвериков, Г. Г. Побудова гіперланцюгів іменників української мови за допомогою інтегрованої лексикографічної системи [Текст] / Г. Г. Четвериков, Т. М. Федорова [Текст] // Информационные интеллектуальные системы (ИИС'2009) : материалы 2-й факультетской науч.-практ. молодежной школы-семинара, 8–9 дек. 2009 г. – Х. : ХНУРЭ, 2009. – С. 210–213.

13. Федорова, Т. Н. Построение цепочек синонимии для лексикографической системы [Текст] / Т. Н. Федорова // Прикладна лінгвістика 2010: проблеми і рішення : тези Всеукр. наук.-метод. конф. студ. та молодих науковців, 20–21 трав. 2010 р., м. Миколаїв. – Миколаїв : НУК, 2010. – С. 60–63.

14. Концепція моделювання елементів комп'ютерної лінгвістики [Електронний ресурс] / Г. Г. Четвериков, Т. М. Федорова, І. Д. Вечірська, В. В. Токарєв // Горизонты прикладной лингвистики и лингвистических технологий (MegaLing'2010) : материалы междунар. науч. конф., 1–7 окт. 2010 г., Партенит, АР Крым, Украина. – Режим доступа : [www. URL: http://megaling.crimea.edu/publications/program2010.pdf](http://www.megaling.crimea.edu/publications/program2010.pdf).

15. Федорова, Т. Н. Построение реляционной сети для глаголов украинского языка [Текст] / Т. Н. Федорова // Системный анализ и информационные технологии (SAIT 2011) : материалы Междунар. науч.-техн. конф., 23–28 мая 2011 г., Киев. – К. : УНК «ИПСА» НТУУ «КПИ», 2011. – С. 329.

16. Федорова, Т. Н. Автоматизация построения цепочек лексических единиц на примере украинских народных сказок [Электронный ресурс] / Т. Н. Федорова // Горизонты прикладной лингвистики и лингвистических технологий (MegaLing'2011) : материалы междунар. науч. конф., 20–26 сент. 2011 г., Киев. – Режим доступа : [www. URL: http://megaling.ulif.org.ua/arch-v](http://www.megaling.ulif.org.ua/arch-v).

17. Моделирование словоизменения на примере итальянского языка [Текст] / Г. Г. Четвериков, И. Д. Вечирская, Т. Н. Федорова, Н. Е. Русакова // Искусственный интеллект. Интеллектуальные системы (ИИ-2012) : материалы междунар. науч.-техн. конф., 1–5 окт. 2012 г., пос. Кацивели, АР Крым. – Донецк : Наука і освіта, 2012. – С. 130–132.

АНОТАЦІЯ

Федорова Т.М. Алгебро–логічні моделі та метод побудови ланцюгів лексичних одиниць в системах штучного інтелекту. – На правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту. – Харківський національний університет радіоелектроніки, Міністерство освіти і науки України, Харків, 2013.

Дисертацію присвячено розробці математичної моделі словозмінної класифікації частин мови італійської мови та математичних моделей у вигляді логічних мереж словозміни дієслів української мови та іменників італійської мови для підвищення ефективності обробки інформації в природномовних системах; крім того – подальшому розвитку методу знаходження n -го лінійного логічного перетворення для побудови ланцюгів в лексикографічній системі електронних тлумачних словників.

У роботі вперше побудовано математичну модель словозмінної класифікації частин мови, яка характеризується системою ознак (граматичних категорій), що визначають словозміни зазначених класів слів та їх значення в італійській мові. За рахунок введення додаткових класів слів модель дозволяє уникнути неоднозначності в результатах морфологічного аналізу при побудові автоматизованих електронних перекладних словників. Також у роботі вперше запропоновано і досліджено модель флексійної обробки дієслів української мови у вигляді логічної мережі, яка характеризується системою бінарних відношень та формулами відповідних предикатів, що дало можливість реалізовувати їх на програмно – алгоритмічному рівні та використовувати для автоматичного вирішення задач аналізу, синтезу та нормалізації словоформи. Також набув подальшого розвитку метод знаходження n -го лінійного логічного перетворення для побудови ланцюгів в лексикографічній системі електронних тлумачних словників. Цей метод характеризується завданням початкової семантичної залежності на кожному етапі обчислення, що дало можливість уникнути тавтології та циклічних посилань для корегування словникових статей в лексикографічній системі тлумачного типу.

Ключові слова: логічна мережа, модель, флексійна обробка, лексичні одиниці, ланцюг, лінійне логічне перетворення, тлумачний словник.

АННОТАЦИЯ

Фёдорова Т.Н. Алгебро–логические модели и метод построения цепочек лексических единиц в системах искусственного интеллекта. – На правах рукописи.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.23 – системы и средства искусственного интеллекта. – Харьковский национальный университет радиоэлектроники, Министерство образования и науки Украины, Харьков, 2013.

Диссертация посвящена разработке математической модели словоизменительной классификации частей речи итальянского языка и

математических моделей словоизменения глаголов украинского языка и существительных итальянского языка в виде логических сетей для повышения эффективности обработки информации в естественных языковых системах; кроме того – построению цепочек в лексикографической системе электронных толковых словарей с помощью развития метода нахождения n -го линейного логического преобразования на основе алгебры конечных предикатов. Проведен анализ состояния развития систем обработки естественного языка. Проанализированы основные результаты в области моделирования словоизменения для группы флективных языков и выявлены нерешенные задачи. Исследованы существующие средства моделирования фрагментов естественного языка с помощью алгебро–логических средств и лексикографических структур.

В работе построена математическая модель словоизменительной классификации частей речи итальянского языка, которая характеризуется системой признаков (грамматических категорий), определяющих словоизменения указанных классов слов и их значения. Парадигматические классы итальянского языка определяются внутри каждого грамматического класса слов. Это обусловлено тем, что словоизменительные процессы для слов разных классов определяются своими, присущими именно этому классу, словоизменительными параметрами. Данная модель позволяет избежать неоднозначностей в результатах морфологического анализа за счет введения дополнительных классов слов для построения автоматизированных электронных переводных словарей.

Также в диссертационной работе построена модель флексийной обработки глаголов украинского языка в виде логической сети, которая характеризуется системой бинарных отношений и формулами соответствующих предикатов, что позволило реализовывать ее на программно – алгоритмическом уровне и использовать для автоматического решения задач анализа, синтеза и нормализации словоформы. При построении математической модели флексийной обработки языка описывается фрагментное отношение, которое отражает зависимость фрагмента от его смысла. Из морфем слова в качестве фрагмента выбирается окончание.

Разработана модель флексийной обработки существительных итальянского языка в виде логической сети. Предикату модели соответствует отношение модели, которое связывает между собой предметные переменные. Каждой ветви логической сети ставится в соответствие свое бинарное отношение модели. Каждая ветвь обозначается номером своего отношения. Она соединяет два полюса, соответствующие тем предметным переменным, которые связываются отношением, соответствующим данной ветви. Это позволило расширить класс задач, которые решаются с помощью логических сетей.

В работе получил дальнейшее развитие метод нахождения n -го линейного логического преобразования для построения цепочек в лексикографической системе электронных толковых словарей путем задания начальной семантической зависимости на каждом этапе вычисления. Был использован критерий нахождения n -го линейного логического преобразования для нахождения условия окончания построения цепочек лексических единиц: если на двух последующих шагах значение преобразования повторяется, то такое линейное преобразование будет

искомым. Это позволило избежать тавтологии и циклических ссылок для корректировки словарных статей в лексикографической системе толкового типа.

Результаты практической реализации данного метода позволили автоматизировать обработку словарных статей за счет введения отношений толкования, избегать повторов и находить в тексте или фрагменте текста слово не только по конкретно заданным значениям, но и по его содержанию, описанию.

Теоретические результаты исследований внедрены в учебный процесс в рамках дисциплин «Высокоуровневые технологии обработки языковой информации» и «Дискретные структуры» для студентов направления подготовки «Программная инженерия» и в рамках дисциплин «Информатика» и «Основы дискретной математики» для студентов направления подготовки «Издательско – полиграфическое дело».

Ключевые слова: логическая сеть, модель, флексийная обработка, лексические единицы, цепочка, линейное логическое преобразование, толковый словарь.

ABSTRACT

Fyodorova T.N. Algebra – logical model and the method for constructing chains of lexical units in artificial intelligence systems. – Manuscript.

The thesis for the candidate degree in technical sciences on the specialty 05.13.23 – Artificial Intelligence Systems and Tools. – Kharkiv National University of Radio Electronics, Ministry of Education and Science of Ukraine, Kharkiv, 2013.

Dissertation is devoted to the development of the mathematical models of inflection classification of speech parts Italian language and of mathematical models of inflection Ukrainian verbs and Italian nouns in the form of logical networks to efficiency increase of information processing in natural language systems. Besides, it is devoted to the improvement of finding the n -th linear logical transformation method to chains construction in lexicographical system of electronic explanatory dictionaries.

The mathematical model of inflection classification of speech parts Italian language is constructed for the first time. It is characterized by the system of criteria (grammatical categories) that determine the inflection of these words classes and their meanings. The model allows obviating ambiguity in the results of the morphological analysis in construction of automated electronic translation dictionaries through the insertion of additional words classes. Also the model of inflectional processing Ukrainian verbs in the form of logic networks is proposed and explored for the first time. It is characterized by the system of binary relations and formulas of corresponding predicates. This allowed to realizing them on the program – algorithmic level and using to automatically solve the problems of analysis, synthesis and normalization of word forms. The method of n -th linear logical transformation to chains construction in lexicographical system of electronic explanatory dictionaries was further developed. This method is characterized by giving the initial semantic depending on each stage of calculation. This allowed to avoiding of tautology and circular references to correcting dictionaries articles in lexicographical system of explanatory type.

Keywords: logical network, model, inflection processing, lexical units, chain, linear logical conversion, explanatory dictionary.

Відповідальний випусковий **В. П. Машгалір**

Підп. до друку 02.04.13.
Умов. друк. арк. 0,9.
Зам. №

Формат 60x84 1/16.
Облік. вид. арк. 1,2.
Ціна договірна.

Спосіб друку – ризографія.
Тираж 100 прим.

ХНУРЕ, 61166, Харків, просп. Леніна, 14

Віддруковано в навчально-науковому
видавничо-поліграфічному центрі ХНУРЕ
61166, Харків, просп. Леніна, 14

