

УДК 007. 519.714:519.766



КОНЦЕПЦІЯ УНІФІКАЦІЇ МЕТОДІВ ТА ЗАСОБІВ ПОБУДОВИ ПРОСТОРОВИХ БАГАТОЗНАЧНИХ СТРУКТУР МОВНИХ СИСТЕМ

Г.Г. Четвериков

Харківський національний університет радіоелектроніки, м. Харків, chetvergg@kture.kharkov.ua

Стаття присвячена аналізу проблеми створення мовних систем штучного інтелекту. Проведено аналіз алгебро-логічної структури природної мови, що дозволяє моделювати її елементи на фонетичному та морфологічному рівнях на підставі уніфікації логічного та апаратного підходів.

АЛГЕБРА СКІНЧЕННИХ ПРЕДИКАТИВ, ЛОГІКА, ПРИРОДНА МОВА, ШТУЧНИЙ ІНТЕЛЕКТ,
k-ЗНАЧНА СТРУКТУРА, АСП-СТРУКТУРА

Вступ

В умовах роботи реальних систем із високим рівнем невизначеності задач (середовищ) та інформації для побудови інтелектуальних систем (ІС) неминуче використання нових інформаційних технологій, орієнтованих на потоки контекстно залежної інформації. Фактично необхідною є розробка саме природномовних принципів побудови інтелектуального інтерфейсу. Очевидно, що єдиним, відомим нам, об'єктивним носієм моралі і інтелекту є людина, а виразником, засобом до зовнішнього спілкування, є людська мова. Отже, єдиним можливим шляхом найбільш об'єктивного, повного і ефективного створення інтелектуальних систем та систем штучного інтелекту (ШтІ) є шлях аналізу, моделювання та синтезу мовного інтелектуального інтерфейсу [1,2].

На сьогоднішній день таким середовищем, можливо єдиним, є інтелектуальні системи, у яких, починаючи з рівня застосування природної мови, в повній мірі постає питання про необхідність застосування *k*-значної логіки, структур і кодування. Визначивши середовище, якому необхідні багатозначні структури (БС), легше визначити, які властивості та характеристиками вони повинні мати і за якими принципами бути побудовані. З іншого боку, бурхливий розвиток комп'ютеризації, проникнення обчислювальної техніки в усі сфери науки, промисловості, суспільного життя, її використання для розв'язання найскладніших задач сьогодення вступили у протиріччя з технологією обробки даних традиційними фон-нейманівськими комп'ютерами. Спостерігається криза її основ, пов'язана з специфікою архітектури та принципів дії фон-нейманівського процесора з застосуванням послідовних алгоритмів роботи та виключно двозначних елементів, структур і кодування. Перелічені проблеми вимагають вирішення задач розвитку нових принципів організації елементів, структур і методів кодування для обчислювальних систем, зокрема систем ШтІ [3,4].

Людський мозок зберігає об'єктивну модель світової реальності, інакше людина не змогла б

еволюціонувати та вдосконалюватись у процесі свого розвитку. Таким чином можна зробити висновки про доцільність моделювання принципів дії людського мозку для задач створення ШтІ. Оскільки усі розумові здібності, що необхідно передати машині із ШтІ, вже наявні в людини на достатньо високому рівні розвитку, і ніякий інший інтелект, крім людського, науці недоступний, то міркувати машини повинні за тими ж законами, що й людина. Нейрофізіологічні дослідження принципів дії природного інтелекту мозку людини виявили у ньому наявність дво- та багатозначного (*k*-значного) кодування, просторового характеру активності мереж нервових клітин і організації діяльності мозку.

Звідси випливають основні вимоги щодо властивостей базових елементів і структур для побудови новітніх високоефективних систем ШтІ. Вони повинні реалізувати функції багатозначної логіки та кодування, володіти властивостями універсальності, просторовості, гібридності, гнучкого переналагодження без зміни структури, ієрархічності, за складністю порівнюватись із складністю задач, що розв'язуються. Найближчими за вказаними властивостями є багатозначні універсальні просторові елементи та структури [2,4,5].

Можливо прогнозувати і подальший розвиток інтелектуальних систем. Це можна зробити виходячи всього з двох достатньо очевидних міркувань. По-перше, інтелектуальні системи повинні одержати свою власну "інформаційну машину" – машину, що пододала теоретичні обмеження кінцевих автоматів і практично нереалізованої машини Тюрінга. Тут необхідне наукове і технічне рішення, що долає бар'єр "необчислюваності". По-друге, сам факт побудови такої машини, практика спілкування з нею приведуть до остаточного переосмислення поняття і значення інформаційної науки – інформатики. Достатньо очевидний факт неможливості існування інформаційної машини такого роду інакше, як у безперервному режимі засвоєння і реструктуризації інформації, приведе до розуміння інтелектуальної системи як системи досить динамічної.

Отже, основні постулати теорії інтелектуальних систем [1, 4]:

- комп'ютер фон-нейманівської архітектури, кінцевий автомат за своєю суттю, не може бути інструментом створення інтелектуальних систем;
- підвищується інтелектуальність системи при зниженні її точності і навпаки;
- теорія повинна розглядати можливість побудови абстрактних конструкцій, що реалізують не обчислювані в звичному значенні об'єкти;
- необхідна розробка природномовних принципів побудови інтелектуальних систем та систем ШтІ.

На даний час проблеми у розвитку систем ШтІ ставлять питання про застосування орієнтації на досягнення високорівневої технології обробки інформації (тобто отримання нової якості), що виявляється у намаганнях реалізувати на фон-неймановських комп'ютерах системи ШтІ. Отже, сучасні, нові вимоги до технології обробки інформації можна звести до двох моментів: необхідність вирішення проблем, що погано формалізуються, та наявність кінцевого користувача, який за своєю професійною діяльністю не є програмістом [2,3].

Таким чином, приходимо до розуміння одного з варіантів шляхів створення систем ШтІ – це шлях аналізу, моделювання та синтезу мовного інтелектуального інтерфейсу за допомогою засобів k -значної логічної системи (зокрема алгебри скінченних предикатів (АСП)) та k -значних структур і кодування [2, 6-8].

1. Мета роботи

Основним завданням цієї роботи є викладення новоствореної концепції уніфікації методів та засобів побудови просторових багатозначних структур мовних систем. Предмет досліджень – апаратні методи розв'язання рівнянь алгебри предикатів та засоби побудови універсальних багатовходових k -значних просторових структур мовних систем штучного інтелекту. Методи дослідження ґрунтуються на використанні теорії моделювання та базуються на математичному апараті теорії інтелекту, зокрема АСП, для отримання формального опису системи фонетичних ознак, букво-фонемних відношень та морфології української мови; при синтезі швидкодіючих багатозначних структур мовних систем використано апарат теорії інформації, множин, інтелекту, кодування та лексикографічних систем і лексикографічних числень [9 – 11].

2. Вибір математичного апарату

У роботі [10] для задач моделювання режимів автоматизованої обробки мови розроблено та проаналізовано мінімальну систему аксіом АСП. Доведено: АСП є повною алгеброю у тому розумінні, що за допомогою її формул можна в аналітичному

вигляді записати довільний скінченний предикат. Задача інтелектуалізації та українізації цифрових мереж і структур не передбачає простого й швидкого вирішення. Головна перешкода, на яку наштовхуються всі спроби радикально підвищити інтелектуальні здібності ЕОМ, полягає у недосконалості машинних мов. Вони значно поступаються природним мовам. Збільшується кількість прихильників того, що створення «машинного інтелекту» вимагає вивчення й моделювання людського інтелекту та, у першу чергу, природної мови, що лежить у його основі. Стає все очевиднішим, що при подальшому вдосконаленні машинних мов необхідно ширше використовувати організацію людської мови. Механізм природної мови, не дивлячись на поверхневу легкість користування людиною, надзвичайно складний й до того ж мало вивчений. Отже, розробка систем обробки мовної інформації часто базується на недостатньо міцному лінгвістичному фундаменті. Багато хто з дослідників, що займалися автоматизацією мовної діяльності, перейшли до поглибленого вивчення, дослідження й моделювання окремих сторін механізму природної мови.

Наявність АСП відкриває можливість переходу від алгоритмічного опису інформаційних процесів до опису їх у вигляді рівнянь. А рівняння в свою чергу задають відношення між змінними. Усі змінні в рівнянні рівноправні, будь-які з них можуть виступати як у ролі незалежних, так і в ролі залежних. При цьому перевагою рівнянь перед алгоритмами є те, що можна розрахувати реакцію системи навіть при неповній визначеності вхідних сигналів, у той час як не повністю розроблений алгоритм є непрацездатним. Встановлено, що за умов зміни знань про об'єкт система рівнянь АСП, покладених на структуру системи, завжди готова до використання, а алгоритм часто вимагає докорінної зміни її структури. За допомогою формул АСП будуюмо АСП-структури, які реалізують відповідні скінченні предикати. Це є аналог побудови комбінаційних схем за формулами алгебри логіки. В залежності від рівня функціонально-структурної реалізації маємо АСП-структури першого, другого та третього роду.

3. Природномовні принципи побудови систем штучного інтелекту

Виходячи із поставленої мети та враховуючи основні аксіоми теорії інтелектуальних систем управління, інтегруємо необхідні та уже розроблені природномовні принципи побудови інтелектуального інтерфейсу і систем ШтІ. У цій роботі, перш за все, хотілося б показати необхідність і можливість розробки загальної теорії побудови інтелектуального інтерфейсу та систем штучного інтелекту, яка могла б стати методологічною основою цілеспрямованого створення нових інформаційних та інтелектуальних технологій.

4. Алгебро-логічні засоби моделювання природної мови

Розглянуто алгебро-логічні засоби моделювання природної мови, аналіз та дослідження її алгебро-логічної структури. Природна мова (ПМ), як явище інтелектуальної діяльності людини, є дуже складним об'єктом. Але маючи формальний опис природної мови, його можна реалізувати на ЕОМ і таким чином наділити машину здібністю володіти природною мовою. Необхідно, щоб машина пристосовувалась до людини та розмовляла її мовою, а не навпаки — людина прилаштовувалась до машини. Природна мова для людини найзручніша. Людина ніколи не проміняє її ні на мову математики, ні на мову програмування. Адже немає іншої галузі знань, яка б в більшій мірі сприяла підвищенню темпів комп'ютеризації та інформатизації суспільства. Тоді можна буде наділити машину здібністю до природної мови. Якщо з'явиться можливість виявити з природної мови шляхом її аналізу більш досконалу алгебраїчну мову, це дуже збільшить можливості розробників інформаційних систем та нових інформаційних технологій. Такий алгебро-логічний апарат дозволить розширити можливості розробників, які займаються створенням нових інформаційних технологій. Отже, концептуально-методологічний підхід до ПМ (з математичної точки зору) дозволяє сприймати її як деяку алгебру, а тексти — як формули цієї алгебри.

Таким чином, розглянемо перспективи побудови ЕОМ за принципами роботи мозку. Відправною точкою у цих дослідженнях буде аналогія між багатозначними структурами з просторовими принципами побудови та мікроструктурою кори мозку.

Під час опису лінгвістичних зв'язків природної мови використовується метод зрізу як основний методологічний спосіб [10]. Згідно з цим методом фрагмент тексту X , який нас цікавить, відокремлюють в оточуючому контексті, а самим контекстом нехтують. Дію контексту замінюють набором ознак Y . Потім вивчається і математично описується лінгвістичне відношення $L: X L Y$, що зв'язує фрагмент тексту X з набором його ознак Y . Метод зрізу дозволяє розбити складну задачу математичного опису лінгвістичних закономірностей на множину простіших задач без неприпустимого спрощення й огрублення загальної задачі.

Повний математичний опис фонетичних й морфологічних зв'язків можна досягти простим об'єднанням їх у єдину систему рівнянь, що описують лінгвістичні відношення. У цій системі ознаки відіграють тільки другорядну роль: вони використовуються як проміжні змінні.

Вибір кожної ознаки й області її значень можна здійснити довільним чином. Але довільний набір ознак не можна признати оптимальним. Критерієм оптимальності служить складність отриманої сис-

теми рівнянь. Проте складність системи рівнянь не завжди легко оцінити. Саме цим пояснюються труднощі, що виникають у лінгвістів під час вирішення питання вибору оптимальної серед можливих систем ознак. Викладений нижче підхід досить повно визначає теоретичний статус поняття ознаки.

5. Фонетичний аналіз

У цьому розділі об'єктом моделювання є відношення, що базуються на лінгвістичних зв'язках, які існують між різними елементами фонетичного рівня української мови. На підставі вивчення даних фонетики, математично описуються відношення, що зв'язують окремі фонемні з системою їх фонетичних ознак:

y_1 — ознака голосності звуку зі значеннями: **г** — голосна, **п** — приголосна;

y_2 — ознака шумності приголосних зі значеннями: **ш** — шумова, **нш** — нешумова;

y_3 — ознака вокалізації звуку зі значеннями: **дз** — дзвінка, **г** — глуха;

y_4 — ознака проривності звуку зі значеннями: **пр** — проривна, **щ** — щілинна;

y_5, y_6 — ознаки веляризації та лабілізації звуку зі значеннями: **м, с, в** — мала, середня, велика відповідно;

y_7 — ознака палаталізації зі значеннями: **т, м** — тверда, **м'** — м'яка відповідно;

y_8 — ознака назалізації зі значеннями: **р** — ротова, **н** — носова;

y_9 — ознака вібрантиності зі значеннями: **с** — спокійна, **т** — тремтяча;

y_{10} — ознака місця артикуляції звуку зі значеннями: **г** — губна, **п** — передньоязикова, **з** — задньоязикова;

y_{11} — ознака локалізації передньоязикової шумової фонемі зі значеннями: **з** — зубна, **п** — піднебінна;

y_{12} — ознака африкативності передньоязикових шумових фонем зі значеннями: **аф** — африката, **неаф** — неафриката.

Встановлено зв'язки між звуками Y мови у визначеному фонетичному контексті Z та відповідними їм буквами українського алфавіту X у вигляді тернарного відношення $\Phi(X, Y, Z)$, яке названо буквофонемним.

Математична модель лінгвістичних зв'язків, що регулюють процес переходу від фонетичного подання слів до їх графічного запису й навпаки, матиме вигляд:

$$\Phi = \Phi_1 \& \Phi_2 \& \dots \& \Phi_r,$$

де $\Phi_1 = 1; \Phi_2 = 1, \dots, \Phi_r = 1$ — різні, незалежні одне від одного зв'язки між X, Y, Z , що фактично існують в українській мові; r — загальна їх кількість.

Під час переходу від значень фонетичних ознак до фонетичного зображення звуку формується

фонема, тобто знак, що позначає клас звуків, які тяжіють до певної букви тексту. Зворотне перетворення полягає в заміні фонетичного знаку набором відповідних йому значень фонетичних ознак звуку.

Фонема як знак: $\text{ФОН}(X) = x^A \vee x^B \vee x^V \vee x^Г \vee x^Г \vee x^Д \vee x^E \vee x^E \vee x^Ж \vee x^З \vee x^И \vee x^I \vee x^I \vee x^Й \vee x^K \vee x^Л \vee x^М \vee x^Н \vee x^О \vee x^П \vee x^Р \vee x^С \vee x^Т \vee x^У \vee x^Ф \vee x^Х \vee x^Ц \vee x^Ч \vee x^Ш \vee x^Щ \vee x^Ю \vee x^Я$.

Ознака голосності звуку y_1 :

$$\text{Голосн}(X) = x^A \vee x^E \vee x^I \vee x^O \vee x^Y \vee x^И;$$

$$\text{Приголсн}(X) = x^B \vee x^B \vee x^Г \vee x^Г \vee x^Д \vee x^E \vee x^Ж \vee x^З \vee x^Й \vee x^K \vee x^Л \vee x^М \vee x^Н \vee x^П \vee x^Р \vee x^С \vee x^Т \vee x^Ф \vee x^Х \vee x^Ц \vee x^Ч \vee x^Ш \vee x^Щ$$

Ознака шумності звуку y_2 :

$$\text{Шум}(X) = x^B \vee x^B \vee x^Г \vee x^Д \vee x^Ж \vee x^З \vee x^K \vee x^С \vee x^П \vee x^Т \vee x^Ф \vee x^Х \vee x^Ц \vee x^Ч \vee x^Ш;$$

$$\text{Сон}(X) = x^Л \vee x^М \vee x^Н \vee x^Р \vee x^Й;$$

$$y_1 \text{ ПРИГ} \rightarrow y_2 \text{ НШУМ} \wedge \text{Сон}(X) \vee y_2 \text{ ШУМ} \wedge \text{Шум}(X) = 1.$$

Ознака вокалізації звуку y_3 :

$$\text{Глух}(X) = x^K \vee x^П \vee x^С \vee x^Т \vee x^Ф \vee x^Х \vee x^Ц \vee x^Ч \vee x^Ш;$$

$$\text{Дзвн}(X) = x^B \vee x^B \vee x^Г \vee x^Д \vee x^Ж \vee x^З \vee x^Ц \vee x^Ч. \quad y_1 \text{ ПРИГ} \rightarrow y_3 \text{ ДЗ} \wedge \text{Дзвн}(X) \vee y_3 \text{ ГЛУХ} \wedge \text{Глух}(X) = 1.$$

Ознака проривності звуку y_4 :

$$\text{Прор}(X) = x^B \vee x^Г \vee x^Д \vee x^K \vee x^П \vee x^Т \vee x^Х \vee x^Ц \vee x^Ч;$$

$$\text{Щіл}(X) = x^B \vee x^Г \vee x^Ж \vee x^З \vee x^С \vee x^Ф \vee x^Ш \vee x^Щ;$$

$$y_1 \text{ ПРИГ} \wedge y_2 \text{ ШУМ} \rightarrow y_4 \text{ ПРОР} \wedge \text{Прор}(X) \vee y_4 \text{ ЩІЛ} \wedge \text{Щіл}(X) = 1.$$

Ознака веляризації звуку y_5 :

$$\text{Мвл}(X) = x^A \vee x^I \vee x^Я; \quad \text{Свл}(X) = x^E \vee x^E \vee x^O;$$

$$\text{Ввл}(X) = x^I \vee x^Y \vee x^И \vee x^Ю;$$

$$z^1 \rightarrow y_5^M \wedge \text{Мвл}(X) \vee y_5^C \wedge \text{Свл}(X) \vee y_5^E \wedge \text{Ввл}(X) = 1.$$

Ознака лабілізації звуку y_6 :

$$\text{Млб}(X) = x^A \vee x^E \vee x^I \vee x^И \vee x^E \vee x^Я; \quad \text{Слб}(X) = x^E \vee x^O;$$

$$\text{Влб}(X) = x^Y \vee x^Ю; \quad z^1 \rightarrow y_6^M \wedge \text{Млб}(X) \vee y_6^C \wedge \text{Слб}(X) \vee y_6^B \wedge$$

$$\text{Влб}(X) = 1.$$

Ознака палаталізації звуку y_7 :

$$\text{Мгол}(X) = x^E \vee x^I \vee x^I \vee x^Ю \vee x^Я;$$

$$\text{Тгол}(X) = x^A \vee x^O \vee x^Y \vee x^И \vee x^E; \quad z^1 \rightarrow y_7^T \wedge \text{Тгол}(X) \vee y_7^M \wedge$$

$$\text{Мгол}(X) = 1.$$

Рівняння зв'язку фонем X з ознаками y_5, y_6, y_7 :

$$z^1 \rightarrow y_5^M (y_6^M (y_7^T x^A \vee y_7^M x^Я) \vee y_6^B \wedge (y_7^T x^И \vee y_7^M x^I)) \vee y_5^C ((y_7^T x^O \vee y_7^M x^I) \vee (y_7^T x^E \vee y_7^M x^E)) \vee y_5^B (y_7^T x^Y \vee y_7^M x^Ю) = 1.$$

Ознака назалізації звуку y_8 :

$$\text{Рсн}(X) = x^Й \vee x^Л \vee x^М; \quad \text{Нсн}(X) = x^М \vee x^Н;$$

$$z^2 \rightarrow y_8^P \wedge \text{Рсн}(X) \vee y_8^H \vee \text{Нсн}(X) = 1.$$

Ознака вібрантності звуку

$$y_9: \text{Ссн}(X) = x^Й \vee x^Л \vee x^М \vee x^Н \vee x^В;$$

$$z^2 \rightarrow y_9^C \wedge \text{Ссн}(X) \vee y_9^T \wedge x^T = 1$$

Ознака місця артикуляції звуку

$$y_{10}: \text{Псн}(X) = x^Л \vee x^Н \vee x^P;$$

$$z^2 \rightarrow y_{10}^Г \wedge x^М \vee y_{10}^П \wedge \text{Псн}(X) \vee y_{10}^З \wedge x^Й = 1.$$

Рівняння зв'язку фонем X з ознаками y_8, y_9, y_{10} :

$$z^2 \rightarrow y_{10}^З x^Й \vee y_8^P y_{10}^П x^Л \vee y_8^H (y_{10}^Г x^М \vee y_{10}^П x^Н) \vee y_9^T x^P = 1.$$

Індивідуальні предикати для передньоязикових, задньоязикових та губних шумових фонем:

$$\text{Пшф}(X) = x^Д \vee x^Ж \vee x^З \vee x^ДЖ \vee x^ДЗ \vee x^С \vee x^Т \vee x^Ц \vee x^Ч \vee x^Ш;$$

$$\text{Зшф}(X) = x^Г \vee x^K \vee x^Х; \quad \text{Гшф}(X) = x^Б \vee x^В \vee x^Г \vee x^Г \vee x^Ф;$$

$$y_2 \text{ Ш} \rightarrow y_{10}^Г \text{Гшф}(X) \vee y_{10}^П \text{Пшф}(X) \vee y_{10}^З \text{Зшф}(X) = 1.$$

Індивідуальні предикати для зубних та піднебінних передньоязикових шумових фонем:

$$\text{Зни}(X) = x^Д \vee x^З \vee x^ДЖ \vee x^ДЗ \vee x^С \vee x^М \vee x^Ц;$$

$$\text{Пни}(X) = x^Д \vee x^Ж \vee x^Т \vee x^Ч \vee x^Ш;$$

$$y_2 \text{ Ш} \wedge y_{10}^П \rightarrow y_{11}^З \wedge \text{Зни}(X) \vee y_{11}^ПН \wedge \text{Пни}(X) = 1.$$

Ознака афrikативності звуку y_{12} :

$$\text{Неафни}(X) = x^Д \vee x^Ж \vee x^ДЖ \vee x^ДЗ \vee x^З \vee x^С \vee x^Т \vee x^Ш;$$

$$\text{Афни}(X) = x^Ч \vee x^Ц;$$

$$y_2 \text{ Ш} \wedge y_{10}^П \rightarrow y_{12}^АФ \wedge \text{Афни}(X) \vee y_{12}^НЕАФ \wedge \text{Неафни}(X) = 1.$$

Рівняння зв'язку фонем X з ознаками y_{10}, y_{11}, y_{12} :

$$z^3 \rightarrow y_{10}^Г x^Б \vee y_{10}^З x^Г \vee y_{10}^П [y_{11}^З (y_{12}^НЕАФ x^Д \vee y_{12}^АФ x^Ц) \vee y_{11}^ПН x^Ч] = 1;$$

$$z^4 \rightarrow y_{10}^Г x^В \vee y_{10}^З x^Г \vee y_{10}^П (y_{11}^З x^З \vee y_{11}^ПН x^Ж) = 1;$$

$$z^5 \rightarrow y_{10}^Г x^П \vee y_{10}^З x^K \vee y_{10}^П [y_{11}^З (y_{12}^НЕАФ x^Т \vee y_{12}^АФ x^Ц) \vee y_{11}^ПН x^Ч] = 1;$$

$$z^6 \rightarrow y_{10}^Г x^Ф \vee y_{10}^З x^Х \vee y_{10}^П (y_{11}^З x^С \vee y_{11}^ПН x^Ш) = 1.$$

Створені моделі зв'язують фонему з набором ознак звуку й дозволяють визначити значення фонем в залежності від значень ознак. Уведено відношення, що пов'язує фонетичний запис тексту з його орфографічним записом, між фонемою та фонетичними ознаками звуку. Перехід від фонетичного запису тексту до орфографічного й навпаки здійснено на основі правил української мови, що регулюють вимову (фонетичні правила) та написання (орфографічні правила).

Формально описано відношення, які описують вплив фонетичного контексту на вибір букв для позначення голосних звуків.

Правила фонетичних позицій голосних звуків у слові –

– положення після різних, відносно категорії твердості – м'якості приголосних:

$$G_1^T = J_2(p_V) q_V^T; \quad G_2^M = J_2(p_V) q_V^M,$$

де $J_2(p_V) = p_V^B \vee p_V^V \vee p_V^D \vee p_V^3 \vee p_V^L \vee p_V^M \vee p_V^G \vee p_V^{ДЖ} \vee p_V^{ДЗ} \vee p_V^H \vee p_V^П \vee p_V^P \vee p_V^C \vee p_V^T \vee p_V^Ф$;

– на абсолютному початку слова; після голосних звуків:

$$G_3^H = p_V;$$

$$G_4^Г = p_V^A \vee p_V^Y \vee p_V^O \vee p_V^I \vee p_V^И \vee p_V^E;$$

– після [ц] та шиплячих приголосних:

$$G_5^Ц = p_V^Ц q_V^T;$$

$$G_6^Ш = (p_V^Ш \vee p_V^Ж) (q_V^M \vee q_V^T) \vee p_V^Ч q_V^M;$$

– положення після задньоязикових звуків та після [j]:

$$G_7^3 = p_V^Г \vee p_V^K \vee p_V^X;$$

$$G_8^J = p_V^J.$$

Запис правил рівняннями АСП:

$$\Phi_{11}^T = G_1^T \rightarrow (p^A A \vee p^Y Y \vee p^O O \vee p^I I \vee p^И И) q_V^T t^3;$$

$$\Phi_{12}^T = G_1^T \rightarrow (p^Y Y \vee p^И И) q_V^T (t^1 \vee t^2),$$

де Φ_{11}^T , Φ_{12}^T – буквофонемні відношення твердих наголошених голосних після твердих приголосних; t^1 , t^2 – ознаки ненаголошених та переднаголошених складів відповідно; t_3 – ознака наголошених складів.

$$\Phi_{13}^T = G_1^T p_V \rightarrow p^E q^T t^3 x^E;$$

$$\Phi_{14}^T = G_1^T A \rightarrow p^E q^T (t^1 \vee t^2) x^E;$$

$$\Phi_{15}^T = G_1^T J_2(p_U) \rightarrow p^E q^T t^3 x^E, (x^E \vee x^E),$$

де $A = (p_V^T \vee p_V^D \vee p_V^C \vee p_V^3 \vee p_V^P \vee p_V^H) q_V^T$;

$$\Phi_{21}^M = G_2^M \rightarrow q^M t^3 (p^A Y \vee p^Y Y \vee p^O I \vee p^E E \vee p^И И);$$

$$\Phi_{22}^M = G_2^M \rightarrow p^Y q_V^M x^O (t^2 \vee t^1);$$

$$\Phi_{31}^П = G_3^П \rightarrow (p^A A \vee p^O O) q^T t^3;$$

$$\Phi_{32}^П = G_3^П \rightarrow [(p^Y Y \vee p^E E) q^T \vee p^I q^M x^I] (t^2 \vee t^3);$$

$$\Phi_{41}^Г = G_4^Г \rightarrow [(p^A A \vee p^O O \vee p^Y Y) q^T \vee p^I q^M x^I] t^3;$$

$$\Phi_{42}^Г = G_4^Г \rightarrow p^E t^3 (p^I q^T x^E \vee p^И q^T x^E);$$

$$\Phi_{51}^Ш = G_5^Ш \rightarrow (p^A \sim x^A) (p^E \sim x^E) (p^I \sim x^I) t^3;$$

$$\Phi_{52}^Ш = G_5^Ш \rightarrow (p^Y \sim x^Y) (t^3 \vee t^2 \vee t^1);$$

$$\Phi_{53}^Ш = G_5^Ш p_V \rightarrow p^O x^O q^T t^1;$$

$$\Phi_{54}^Ш = G_5^Ш J^3(p_V) \rightarrow p^O x^O q^T t^3,$$

де предикат $J^3(p_U) = p_U^B \vee p_U^V \vee p_U^Г \vee \dots \vee p_U^Ш$ – формалізує поняття приголосного звуку;

$$\Phi_{61}^Ц = G_6^Ц \rightarrow (p^A A \vee p^O O \vee p^E E) q^T t^3 \vee p^Y q^T x^Y.$$

$$\Phi_{62}^Ц = G_6^Ц \rightarrow [p^I \sim (x^I \sim x^И)] (t^3 \vee t^2 \vee t^1);$$

$$\Phi_{71}^3 = G_7^3 \rightarrow \{ [p^A q^T x^A \vee q^M x^Y] \vee p^O (q^T x^O q^M x^И) \} t^3 \vee p^Y (q_V^T x^Y \vee q_V^M x^Ю) (t^3 \vee t^2 \vee t^1);$$

$$\Phi_{72}^3 = G_7^3 \rightarrow (p^E x^E \vee p^И x^И) q^T t^3; \quad \Phi_{81}^J = G_8^J \rightarrow (p^A Y \vee p^O I \vee p^E x^E) q^T t^3; \quad \Phi_{82}^J = G_8^J \rightarrow (p^Y \sim x^Ю) q^M.$$

Отримана система рівнянь є підмоделлю буквофонемних зв'язків, що охоплює клас голосних звуків української мови.

Формально описано відношення, які описують вплив фонетичного контексту на вибір букв для позначення приголосних звуків.

Предикати впливу фонетичного контексту: $G_9^{СП}$ – сильна позиція; $G_{10}^{КЦ}$ – кінець позиції; $G_{11}^{ГЛУХ}$ – позиція перед глухим приголосним; $G_{12}^{ДЗ}$ – позиція перед дзвінкими шумовими приголосними.

$$G_9^{СП} = p_U^J \vee J_4(p_U) \vee J_5(p_U) \vee p_U^B,$$

де $J_4(p_U)$ і $J_5(p_U)$ – предикати, що задають формальний опис понять «голосний звук» та «сонорний звук» відповідно й визначаються виразами

$$J_4(p_U) = p_U^A \vee p_U^E \vee p_U^I \vee p_U^O \vee p_U^Y \vee p_U^И;$$

$$J_5(p_U) = p_U^Л \vee p_U^М \vee p_U^Н \vee p_U^Р.$$

$$G_{10}^{КЦ} = p_U; \quad G_{11}^{ГЛУХ} = J_6(p_U); \quad G_{12}^{ДЗ} = J_7(p_U),$$

де $J_6(p_U)$ і $J_7(p_U)$ визначають класи дзвінких та глухих шумових приголосних:

$$J_6(p_U) = p_U^B \vee p_U^V \vee p_U^Г \vee p_U^Д \vee p_U^Ж \vee p_U^3 \vee p_U^{ДЖ} \vee p_U^{ДЗ};$$

$$J_7(p_U) = p_U^П \vee p_U^X \vee p_U^K \vee p_U^T \vee p_U^Ш \vee p_U^Ч \vee p_U^C \vee p_U^Ц \vee p_U^Ф.$$

Чотири групи ознак, які формально подамо предикатами:

– предикат $\Phi_1^{СП}$ фіксує залежність між значеннями різних ознак умови, що звук U перебуває у сильній позиції; $\Phi_2^{КЦ}$ – за умови, що Y звучить в кінці слова; $\Phi_3^{ГЛУХ}$ – перед глухими приголосними; $\Phi_4^{ДЗ}$ – перед дзвінкими шумовими приголосними.

Ознака дзвінкості – глухості:

– в сильній позиції регулюються предикатом:

$$\Phi_1^{СП} = G_9^{СП} \rightarrow (p^B x_1^B \vee p^V x_1^V \vee p^Г x_1^Г \vee p^Д x_1^Д \vee p^Ж x_1^Ж \vee p^3 x_1^3 \vee p^K x_1^K \vee p^П x_1^П \vee p^C x_1^C \vee p^T x_1^T \vee p^Ф x_1^Ф \vee p^Ш x_1^Ш);$$

– в положенні кінця слова:

$$\Phi_2^{КЦ} = G_{10}^{КЦ} \rightarrow A \vee (p^K \sim (x_1^K \vee x_1^Г)),$$

де $A = (p^П \sim (x_1^П \vee x_1^Б)) \wedge \wedge (p^Ф \sim (x_1^Ф \vee x_1^В)) (p^Т \sim (x_1^Т \vee x_1^Д)) (p^C \sim x_1^3 \vee x_1^C) (p^Ш \sim (x_1^Ш \vee x_1^Ж))$;

– група правил, що задає відношення $\Phi_{3i}^{ГЛУХ}$ (за умови $G_{11}^{ГЛУХ} = 1$):

$$\Phi_{31}^{ГЛУХ} = (p_U^T \vee p_U^C) q_U^M \rightarrow A q^M;$$

$$\Phi_{32}^{ГЛУХ} = G_{11}^{ГЛУХ} p_U \rightarrow (p^П q^T \sim (x_1^П \vee x_1^Б));$$

$$\Phi_{33}^{ГЛУХ} = G_{11}^{ГЛУХ} p_U \rightarrow p^Ф q^T \sim x_1^K \vee x_1^В \vee x_1^X \vee x_1^Ф;$$

$$\Phi_{34}^{ГЛУХ} = G_{11}^{ГЛУХ} p_U \rightarrow (p^П q^T \sim (x_1^Т \vee x_1^Д));$$

$$\Phi_{35}^{ГЛУХ} = (p_U^Т \vee p_U^Ч) q_U^M \rightarrow (p^Т q^M \sim (x_1^Т \vee x_1^Д));$$

$$\Phi_{36}^{ГЛУХ} = G_{11}^{ГЛУХ} p_U \rightarrow (p^K q^T \sim (x_1^K \vee x_1^Г));$$

$$\Phi_{37}^{ГЛУХ} = G_{11}^{ГЛУХ} p_U \rightarrow (p^Ш q^T \sim (x_1^Ш \vee x_1^Ж));$$

$$\begin{aligned} \Phi_{38} \text{ГЛУХ} &= (p_U^K \vee p_U^\Pi) q^T \rightarrow (p^C q^T (x_1^C \vee x_1^3)); \\ \Phi_{39} \text{ГЛУХ} &= (p_U^K q_U^T \vee p_U^\Pi q_U^M) \rightarrow (p^C q^M (x_1^C \vee x_1^3)); \\ \Phi_{310} \text{ГЛУХ} &= p_U^\Pi q_U^T \rightarrow p^C q^T x_1^\Pi; \\ \Phi_{311} \text{ГЛУХ} &= p_U^\Phi q_U^T \rightarrow p^C q^T x_1^\Phi; \\ \Phi_{312} \text{ГЛУХ} &= p_U^Д q_U^T \rightarrow p^C q^T x_1^Д; \\ \Phi_{313} \text{ГЛУХ} &= p_U^Т \rightarrow p^C q^T x_1^Т; \\ \Phi_{314} \text{ГЛУХ} &= (p_U^C \vee p_U^B \vee p_U^\Phi \vee p_U^\Pi \vee p_U^M \vee p_U^Т \vee p_U^X) \\ &\rightarrow p^C q^T x_1^C; \\ \Phi_{315} \text{ГЛУХ} &= (p_U^\Gamma \vee p_U^\Pi \vee p_U^C) \rightarrow p^C q^M x_1^C; \\ \Phi_{316} \text{ГЛУХ} &= p_U^Ч \rightarrow p^C q^T x_1^\Gamma; \\ \Phi_{317} \text{ГЛУХ} &= p_U^Ж \rightarrow p^C q^T x_1^Ж; \\ \Phi_{318} \text{ГЛУХ} &= p_U^Ш \rightarrow p^C q^T x_1^Ш; \end{aligned}$$

– склад предиката $\Phi_4^{ДЗ}(X, Y, Z)$ (звук [y] перебуває у фонетичній позиції перед дзвінками шумовими приголосними):

$$\begin{aligned} \Phi_{41}^{ДЗ} &= p^B q^T (p_U^\Gamma \sim x_1^\Phi) (p_U^\Gamma \sim x_1^B); \\ \Phi_{42}^{ДЗ} &= (p_U^B \vee p_U^B) q_U^M \rightarrow p^B q^M x_1^B; \\ \Phi_{43}^{ДЗ} &= (q^T G_{12}^{ДЗ} p_U^Д \vee q^M p_U^Д q_U^M) \rightarrow (p^Д \sim (x_1^Д \vee x_1^Т)); \\ \Phi_{44}^{ДЗ} &= p_U^Б \sim x_1^Д p^Д q^M; \\ \Phi_{45}^{ДЗ} &= (p_U^Б \vee p_U^Д \vee p_U^З \vee p_U^\Gamma) q_U^M \sim p^З q^T (x_1^З \vee x_1^C); \\ \Phi_{46}^{ДЗ} &= (p_U^Б q_U^M \vee p_U^Д q_U^Т \vee p_U^Т q_U^M) \sim p^З q^M (x_1^З \vee x_1^C); \\ \Phi_{47}^{ДЗ} &= G_{12}^{ДЗ} \rightarrow p^Г q^T (x_1^\Gamma \vee x_1^X); \\ \Phi_{48}^{ДЗ} &= p^Б \sim x_1^Б. \end{aligned}$$

Наведена вище система рівнянь є формальним еквівалентом зв'язків, що існують між звуками сонорних, парних і позапарних дзвінких та глухих приголосних і графічними еквівалентами мови, що їм відповідають.

За допомогою запропонованого опису буквофонемних зв'язків у вигляді системи рівнянь алгебри предикатів можна здійснити формальний перехід від фонетичного запису слова до його графічного подання (задача аналізу), за заданим орфографічним текстом отримувати його фонетичну транскрипцію (задача синтезу).

Отримані формальні моделі наглядно показують, що математичний апарат алгебри скінченних предикатів й вибрана методика формального опису дозволяє отримати їх компактний формальний опис, а також опис акцентних характеристик відмінюваних і дієвідмінюваних частин української мови. Такі моделі відповідають двом основним вимогам формалізації, а саме: достатньо точно описують вихідний мовний об'єкт та можуть бути реалізовані засобами обчислювальної техніки. Розроблені формальні моделі акцентних залежностей мови можна використати для економного запису словарних структур природної мови в лексикографічних базах даних.

6. Морфологічний аналіз

У цьому розділі висвітлено шляхи математичного моделювання морфологічного рівня української мови. Об'єктом формального опису є письмова, тобто орфографічна, а не фонематична реалізація словоформ української мови. При цьому окремою словоформою є відтинок тексту між двома проміжками.

Особливу роль під час словозміни відіграють закінчення словоформ (флексії), що передають, як правило, граматичні значення в українських словах. У морфології української мови існує певна залежність (відношення) між флексією та оточуючим її текстом. Завдання полягає у математичному описі існуючої залежності, тобто у формалізації поняття флексії. Текст, що охоплює закінчення, є неоднорідним щодо нього. Розрізняємо найближчий текст, що межує безпосередньо з закінченням у словоформі, і подальший текст, що межує зі словоформою, у якій перебуває закінчення, а також зосередимо подальший розгляд на методиці формального опису процесів зміни слів за допомогою флексій, що можна застосувати й до інших лінгвістичних одиниць тексту.

Дію тексту, що оточує закінчення (як ближчого, так і подальшого), замінимо введенням набору ознак, який проінтерпретуємо як смисл закінчення. Під смислом розумітимемо набір проміжних змінних, що поєднані рівняннями, які описують закінчення, з рівняннями, що описують решту частини тексту. Набір значень цих змінних інтерпретуваємо як значення смислу. Смисл розглядається як змінний вектор $X = (x_1, x_2, \dots, x_m)$, де m – число компонентів вектора смислу, а змінні x_1, x_2, \dots, x_m – компоненти смислу.

Переважна більшість закінчень містить одно-, дво-, трибуквені закінчення та нульову флексію. Це дало підставу для формального подання закінчення у вигляді трикомпонентного вектора $Y = (y_1, y_2, y_3)$, де y_1, y_2, y_3 – літери, що стоять зліва направо на першому, другому та третьому місцях закінчення відповідно.

Для того щоб отримати компактний та повний опис фрагментного морфологічного відношення $L(X, Y)$, необхідно вибрати компоненти вектора X з урахуванням внутрішньої структури мови. Через те, що об'єкт, який моделюється, є достатньо складною для математичного опису системою, зручно розбити його на ряд підсистем, які надалі називатимемо «мовними блоками», що пізніше об'єднуються теоремою про розкладання в алгебрі скінченних предикатів.

Для кожного мовного блоку фіксують значення деяких із компонентів смислу, які ми назвемо обмеженими. Другу групу ознак, що служать безпосередньо для вибору закінчення всередині деякого мовного блока, назвемо змінними. Обмежувальним ознакам поставимо у відповідність змінну ξ ,

а змінним – X . Таким чином, маємо можливість уточнити структуру вектора смислу $X = (\xi, X) = (\xi_1, \xi_2, \dots, \xi_k, x_1, x_2, \dots, x_m)$, де k, m – число обмежувальних та змінних компонентів вектора смислу відповідно.

Змінні ознаки в свою чергу неоднорідні щодо закінчення. Впливові ближнього тексту на закінчення відповідатимуть лексичні ознаки, а подальшого – синтаксичні.

Отже, завдання полягає у формальному описі засобами АСП фрагментарного морфологічного відношення $L = (\xi_1, \xi_2, \dots, \xi_k, x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n)$, що буде моделлю флексійного оброблення українських слів. Морфологічне відношення визначається своєю характеристичною функцією $L_\varphi(X, Y)$.

Звуження обсягу математичного опису здійснено введенням шести граматичних ознак. Спочатку введено ознаку ξ_1 виду морфеми зі значеннями: закінчення (з), не закінчення (н) (префікс, корінь, суфікс, тощо).

Ознака змінюваності слова, експлікована змінною ξ_2 , може приймати одне з двох значень – змінюване (z_M), незмінюване (n_{3M}).

Введення змінної ξ_3 , що відповідає ознаці типу словозміни, дозволило окремо описувати відмінювання іменних частин мови ($\xi_3 = i$) та відмінювання дієслів ($\xi_3 = d$).

Іменні частини мови розрізняються за ознакою типу відмінювання (змінна ξ_4). Змінна ξ_4 визначена на множині значень: субстантивна (с), ад'єктивна (а), займенникова (з) та числівникова (ч).

Введемо граматичну ознаку змінної іменної частини мови (ξ_5), що може приймати одне з таких значень: іменник (і), прикметник (п), займенник (з), числівник (ч), дієприкметник (дп).

Шоста граматична ознака, що забезпечує звуження об'єкта математичного моделювання, ξ_6 – вид словоформи зі значеннями: регулярна (р), нерегулярна (н).

Проведені лінгвістичні дослідження показали, що необхідно ввести дев'ятнадцять граматичних ознак: x_1 – відмінок зі значеннями: називний (н), родовий (р), давальний (д), знахідний (з), орудний (о), місцевий (м), кличний (кл); x_2 – рід зі значеннями: чоловічий (ч), жіночий (ж), середній (с); x_3 – число зі значеннями: одиниця (о), множина (м); x_4 – ознака одухотвореності зі значеннями: одухотворена (о), не одухотворена (н); x_5 – варіант закінчення зі значеннями: архаїчне (а), сучасне (с); x_6 – ознака наголошеності закінчення зі значеннями: наголошене (н), ненаголошене (н'); x_7 – остання буква основи зі значеннями, заданими на множині букв та знаків українського алфавіту; x_8 – вид основи зі значеннями: тверда (т), м'яка (м); x_9 – вимога встановлення знака «над буквою» $\tilde{}$ зі значеннями: ставити (с), не ставити (н); x_{10} – повнота словоформи зі значеннями: повна (п), коротка (к); x_{11} – ознака репрезентації зі значен-

нями: інфінітив (і), особова форма (о); x_{12} – спосіб зі значеннями: дійсна (д), наказова (н); x_{13} – часу зі значеннями: минуле (м), не минуле (m^{-3}); x_{14} – особи зі значеннями: перша (1), друга (2), третя (3); x_{15} – тип впливу дієслівної основи зі значеннями: перший (1), другий (2); x_{16} – відмінювання зі значеннями: перше (1), друге (2); x_{17} – ознака наявності частки «ви» зі значеннями: так (т), ні (н); x_{18} – передостання буква основи зі значеннями, заданими на множині більшості букв українського алфавіту; x_{19} – наявність нормативних обмежень зі значеннями: так (т), ні (н).

Лінгвістичні дослідження формальних моделей словозміни української мови виявили специфічний характер індивідуального морфологічного предиката $L_i(X, Y)$, який можна зобразити у вигляді кон'юнкції двох предикатів, що залежать від меншого числа змінних. Лінгвістичною основою такого зображення предиката $L_i(X, Y)$ є різна природа змінних ознак X , що відповідають двом механізмам: синтаксичного впливу тексту та впливу закінчень основ.

Предикати L_{i1} та L_{i2} змістовно інтерпретуються як таблиці впливу основ та парадигматичні таблиці синтаксичного впливу відповідно. Функція $y_j = F(X)$, ($j = \overline{1,3}$), що відповідає предикатові $L_i(X, Y)$, задана на декартовому добутку $A \times B$ значеннями в M у вигляді:

$$Y_j = \varphi(t, s), \text{ де } t = \eta(X_I), s = \mu(X_{II});$$

$$A = X_I \in \{x_1, x_2, \dots, x_5, x_{10}, x_{11}, \dots, x_{14}\};$$

$$B = X_{II} \in \{x_6, x_7, x_8, x_{15}, x_{16}, \dots, x_{19}\}.$$

Для отримання функцій $s = \mu(X_{II})$ та $t = \eta(X_I)$ розроблено відповідні алгоритми. Множина цих номерів прийнята як області значення функції $s' = \mu(X_{II})$:

$$s^1 = \text{ПРГ}(x_7) x_8^t \vee \text{ЗЯЗ}(x_7) \vee \text{ГОЛ}(x_7) \vee x_7^H;$$

$$s^2 = \text{ПРГ}(x_7) x_8^M; \quad s^3 = \text{ГОЛ}(x_7) \vee x_7^H.$$

Отримано два класи суміжності, перенумеровані довільно та прийняті як значення функції $t' = \eta(X_I)$:

$$t^1 = (x_1^H \vee x_1^O \vee x_4^H) x_2^M x_3^O;$$

$$t^2 = (x_1^P \vee x_1^3 \vee x_4^O) x_2^M x_3^O.$$

Предикат $L_{14}(X, Y)$ відповідає функції $y_j = F_i(X)$, що відображає залежність закінчень регулярних слів іменних частин мови ад'єктивного відмінювання від смислу. Подамо цю функцію у вигляді:

$$y_j = \varphi_1(s_1, t_1), \text{ де } s_1 = \mu_1(X_{II}), t_1 = \eta_1(X_I);$$

$$X_I = \{x_1, x_2, x_3, x_4, x_5, x_{10}\}, X_{II} = \{x_6, x_7, x_8\}.$$

Парадигматична таблиця ад'єктивного відмінювання є квадратною матрицею, а таблиця впливу основ – прямокутною. У парадигматичній таблиці класи суміжності результуючого розбиття Q_j' перенумеровані довільним чином.

Способи впливу подальшого тексту на закінчення регулярних словоформ за числом класів суміжності:

$$\begin{aligned} t_1^1 &= (x_1^h \vee x_1^3 x_4^h) x_2^u x_3^o x_{10}^p; \\ t_1^2 &= ((x_1^p \vee x_1^3 x_4^o) \wedge x_2^u x_3^o \vee x_1^p x_2^c x_3^o) x_{10}^p; \\ t_1^3 &= x_1^d (x_2^u \vee x_2^c) x_3^o x_{10}^p; \\ t_1^4 &= (x_1^o (x_2^u \vee x_2^c) x_3^o) x_{10}^p; \\ t_1^5 &= x_1^m (x_2^u \vee x_2^c) x_3^o x_{10}^p; \\ t_1^6 &= x_1^h x_2^j x_3^o x_{10}^p; \\ t_1^7 &= (x_1^p \vee x_1^d \vee x_1^o x_5^a \vee x_1^m) x_2^j x_3^o x_{10}^p; \\ t_1^8 &= x_1^3 x_2^j x_3^o x_{10}^p; \\ t_1^9 &= x_1^o x_2^j x_3^o x_5^c x_{10}^p; \\ t_1^{10} &= (x_1^h \vee x_1^3) x_2^c x_3^o x_{10}^p; \\ t_1^{11} &= (x_1^h \vee x_1^3 x_4^h) x_3^m x_{10}^p; \\ t_1^{12} &= (x_1^p \vee x_1^a x_4^o \vee x_1^m) x_3^m x_{10}^p; \\ t_1^{13} &= x_1^o x_3^m x_{10}^k; \\ t_1^{14} &= x_2^u x_3^o x_{10}^k; \\ t_1^{15} &= x_2^c x_3^o x_{10}^k; \\ t_1^{16} &= x_2^c x_3^o x_{10}^k; \\ t_1^{17} &= x_3^m x_{10}^k; \end{aligned}$$

$$\begin{aligned} t_1^1 &\rightarrow (y_1^h \vee y_1^i \vee y_1^o) y_2^j y_3^u; \quad t_1^2 \rightarrow (y_1^a \vee y_1^j) y_2^m y_3^h; \\ t_1^3 &\rightarrow (y_1^o \vee y_1^e) y_2^b y_3^i; \quad t_1^4 \rightarrow (y_1^h \vee y_1^i) y_2^m y_3^u; \\ t_1^5 &\rightarrow (y_1^o \vee y_1^e) y_2^m y_3^u; \quad t_1^6 \rightarrow (y_1^b \vee y_1^j) y_2^u y_3^m; \\ t_1^7 &\rightarrow (y_1^h \vee y_1^i) y_2^j y_3^u; \quad t_1^8 \rightarrow (y_1^o \vee y_1^e) y_2^u y_3^{j^o}; \\ t_1^9 &\rightarrow (y_1^o \vee y_1^b) y_2^u y_3^{j^o}; \quad t_1^{10} \rightarrow (y_1^a \vee y_1^j) y_2^u y_3^x; \\ t_1^{11} &\rightarrow (y_1^a \vee y_1^e) y_2^u y_3^p; \quad t_1^{12} \rightarrow (y_1^a \vee y_1^j) y_2^t y_3^u; \\ t_1^{13} &\rightarrow (y_1^h \vee y_1^i) y_2^m y_3^h; \\ t_1^{14} &\rightarrow (y_1^u (y_2^u (y_3^u \vee y_3^b) \vee y_2^u y_3^{j^o}); \\ t_1^{15} &\rightarrow (y_1^a \vee y_1^j) y_2^u y_3^u; \\ t_1^{16} &\rightarrow (y_1^o \vee y_1^e) y_2^u y_3^u; \quad t_1^{17} \rightarrow (y_1^i \vee y_1^j) y_2^u y_3^u. \end{aligned}$$

Способи впливу ближчого тексту на закінчення регулярних словоформ за числом класів суміжності:

$$\begin{aligned} s_1^1 &= \text{ПРГ}(x_7) x_6^h x_8^t; \quad s_1^2 = \text{ПРГ}(x_7) x_6^h x_8^m; \\ s_1^3 &= \text{ЗЯЗ}(x_7) x_6^h; \quad s_1^4 = \text{ШИП}(x_7) x_6^h; \quad s_1^5 = x_6^h x_8^h; \end{aligned}$$

$$s_1^6 = x_6^h (\text{ПРГ}(x_7) \vee x_7^h \vee \text{ГОЛ}(x_7) \vee x_7^h);$$

$$s_1^7 = x_6^h (\text{ЗЯЗ}(x_7) \vee \text{ШИП}(x_7));$$

$$s_1^8 = x_6^h (\text{ГОЛ}(x_7) \vee x_7^h).$$

Функція $y_j = \varphi_1(s_1, t_1)$ відповідності структури закінчення (предикативі L_{14}):

— для першої букви закінчення:

$$y_1^a = (t_1^6 \vee t_1^{15})(s_1^1 \vee s_1^3 \vee s_1^4 \vee s_1^5 \vee s_1^6 \vee s_1^7);$$

$$y_1^j = (t_1^6 \vee t_1^{15})(s_1^2 \vee s_1^8);$$

$$y_1^h = (t_1^4 \vee t_1^{11} \vee t_1^{12} \vee t_1^{13} \vee t_1^{17})(s_1^1 \vee s_1^5 \vee s_1^6) \vee t_1^1 (s_1^1 \vee s_1^5);$$

$$y_1^i = (t_1^4 \vee t_1^{11} \vee t_1^{12} \vee t_1^{13} \vee t_1^{17})(s_1^2 \vee s_1^3 \vee s_1^4 \vee s_1^7 \vee s_1^8) \vee t_1^1 (s_1^2 \vee s_1^3 \vee s_1^4 \vee s_1^8);$$

$$y_1^o = (t_1^2 \vee t_1^3 \vee t_1^5 \vee t_1^7 \vee t_1^9 \vee t_1^{10} \vee t_1^{16})(s_1^1 \vee s_1^3 \vee s_1^6 \vee s_1^7) \vee t_1^1 (s_1^6 \vee s_1^7);$$

$$y_1^e = (t_1^2 \vee t_1^3 \vee t_1^5 \vee t_1^7 \vee t_1^9 \vee t_1^{10} \vee t_1^{16})(s_1^2 \vee s_1^4 \vee s_1^5 \vee s_1^8);$$

$$y_1^y = t_1^8 (s_1^1 \vee s_1^3 \vee s_1^4 \vee s_1^5 \vee s_1^6 \vee s_1^7);$$

$$y_1^{j^o} = t_1^8 (s_1^2 \vee s_1^8);$$

$$y_1^u = t_1^{14} (s_1^1 \vee s_1^4 \vee s_1^5);$$

— для другої букви закінчення:

$$y_2^m = t_1^3 \vee t_1^4 \vee t_1^5 \vee t_1^{13};$$

$$y_2^h = t_1^1 \vee t_1^7 \vee t_1^{14} \& s_1^8;$$

$$y_2^x = t_1^{12}; \quad y_2^r = t_1^2;$$

$$y_2^u = t_1^6 t_1^8 \vee t_1^9 \vee t_1^{10} \vee t_1^{11} \vee t_1^{14} (s_1^1 \vee s_1^2 \vee \dots \vee s_1^7) \vee t_1^{15} \vee t_1^{16} \vee t_1^{17};$$

— для третьої букви закінчення:

$$y_3^h = t_1^{13}; \quad y_3^{j^o} = t_1^8 \vee t_1^9;$$

$$y_3^j = t_1^6; \quad y_3^e = t_1^{10} \vee t_1^{11};$$

$$y_3^b = s_1^2 \& t_1^{14};$$

$$y_3^o = t_1^2; \quad y_3^y = t_1^3;$$

$$y_3^u = t_1^1 \vee t_1^4 \vee t_1^5 \vee t_1^7 \vee t_1^{12} \vee t_1^{14} (s_1^1 \vee s_1^3 \vee s_1^4 \vee \dots \vee s_1^8) \vee t_1^{15} \vee t_1^{16} \vee t_1^{17}.$$

Формалізація нерегулярних словоформ з нестандартними закінченнями

$$\bigvee_{\langle \beta_1, \beta_2, \dots, \beta_{20} \rangle_{i=1}}^n Z_i^{\beta_i} (y_1 = Z_{n+1}^{\beta_{n+1}}) (y_2 = Z_{n+2}^{\beta_{n+2}}) (y_3 = Z_{n+3}^{\beta_{n+3}}),$$

де $\langle \beta_1, \beta_2, \dots, \beta_{20} \rangle$ — деяке слово з словника системи, $\langle Z_1, Z_2, \dots, Z_{20} \rangle$ — зображення слова зі словника, $\langle y_1, y_2, y_3 \rangle$ — закінчення слова, $\bigvee_{\langle \beta_1, \beta_2, \dots, \beta_{20} \rangle}$ — означає

логічну суму всіх елементів словника, найдовше слово якого містить не більше двадцяти букв.

7. Формалізація принципу уніфікації багатовходових k -значних структур

В узагальненому вигляді універсальні просторові k -значні структури, згідно з узагальненим способом завдання функцій k -значної логіки за допомогою таблиць істинності та принципу симбіозу дво- та багаторівневого кодування і засобів, включають до свого складу такі компоненти: паралельний аналого-цифровий перетворювач (елемент розпізнавання k -значної змінної); дешифратор (ДШ), селектор, комутатор, паралельний цифро-аналоговий перетворювач (ключовий комутатор, або підсумовувач струмів) [5,8].

Висновки

Вирішення задач формалізації принципів організації багатовходових k -значних структур забезпечує побудову новітньої концепції синтезу структур для високошвидкісних обчислювальних систем; застосування просторового та часового паралелізму на структурному й алгоритмічному рівнях та k -значних методів кодування; створення процедурних і функціональних мов, паралельних машин баз знань і логічного виводу [2–5].

Єдино відомим нам об'єктивним носієм знання та інтелекту є людина, а виразником, засобом до зовнішнього спілкування та носієм інтелекту є людська мова, що й складатиме об'єкт і напрямок наступних досліджень.

На підставі вивчення даних фонетики і орфографії української мови розроблено:

— систему фонетичних ознак, а також систему предикатів для формального опису системи фоне-

тичних ознак, що дозволяє здійснювати їх автоматичне оброблення;

– методика формального опису букво-фонемних відношень, побудовані математичні відношень голосних та приголосних звуків української мови, на основі яких формалізуються закономірності поєднання голосних і приголосних букв та звуків, що утворюють пари за дзвінкістю-глухістю й твердістю-м'якістю, а також поза парних приголосних;

– математичну модель акцентних характеристик української мови. Усі ці результати дозволяють отримати компактну модель для економного запису словників природної мови в пам'яті ЕОМ.

Із допомогою опису букво-фонемних зв'язків у вигляді системи рівнянь АСП можна:

– в інтелектуальному інтерфейсі здійснити формальний перехід від фонетичного запису слова до його графічного зображення (задача фонетичного аналізу);

– за даним орфографічним текстом отримати його фонетичну транскрипцію (задача синтезу мови);

– перейти до системного та структурного синтезу процесорів функціонально-зорієнтованих на аналіз та синтез мови, які є необхідною складовою частиною систем штучного інтелекту.

Наведений перелік характеризує різні рівні роботи з інформацією, виділяє якісно різні групи інформаційних технологій, особливо підкреслюючи можливості роботи із знаннями на рівні контекстно-залежного опису предметної галузі, тобто на рівні семантики і контекстної залежності трактування кожної інформаційної послідовності.

Між першим і подальшими рівнями подання знань проходить стіна, що відокремлює інтелектуальні системи від не інтелектуальних, а саме поняття знань не є винятковою приналежністю інтелектуальних систем і кардинально змінюється на кожному рівні його подання.

Список літератури: 1. *Лачинов, В.М.* Информодинамика или Путь к Миру открытых систем [Текст] / В.М. Лачинов, А.О. Поляков. – Санкт-Петербург: Издательство СПбГТУ. – 1999. – 432 с. 2. *Четвериков, Г.Г.* Формалізація принципів побудови універсальних k -значних структур мовних систем штучного інтелекту [Текст] / Г.Г. Четвериков // Доповіді НАН України. – 2001. – №1 (41). – С. 76–79. 3. *Бондаренко, М.Ф.* Основи теорії синтезу надшвидкодійних структур мовних систем штучного інтелекту [Текст] / М.Ф. Бондаренко, З.Д. Коноплянко, Г.Г. Четвериков. – К.: ІЗМН, 1997. – 264 с. 4. *Бондаренко, М.Ф.* Основи теорії багатозначних структур і кодування в системах штучного інтелекту [Текст] / М.Ф. Бондаренко, З.Д. Коноплянко, Г.Г. Четвериков. – Х.: Фактор-Друк, 2003. – 336 с. 5. *Пат. 2147789 РФ, МКВ Н 03 К 19/02,*

Н 03 М 1/00. Функциональный преобразователь с многозначным кодированием / М.Ф.Бондаренко, З.Д. Коноплянко, Г.Г.Четвериков (Україна). – №97101717/09; Заявл. 04.02.97; Опубл. 24.04.2000, Бюл.№11. – 6 с. 6. *Четвериков, Г.Г.* Алгебро-логічні засоби моделювання природної мови. Повідомлення 1 [Текст] / Г.Г. Четвериков // Бионика интеллекта. – 2004. – №1(61). – С.42–45. 7. *Четвериков, Г.Г.* Алгебро-логічні засоби моделювання природної мови. Повідомлення 2 [Текст] / Г.Г. Четвериков // Бионика интеллекта. – 2005. – №2(62). – С.12–19. 8. *Четвериков, Г.Г.* Алгебро-логічні засоби моделювання природної мови. Повідомлення 3 [Текст] / Г.Г. Четвериков // Бионика интеллекта. – 2007. – №1(66). – С. 48–52. 9. *Широков, В.А.* Очерк основных принципов квантового лингвистики [Текст] / В.А. Широков // Бионика интеллекта. – 2007. – №1(66). – С. 25–32. 10. *Шабанов-Кушнарченко, Ю.П.* Теория интеллекта. Математические средства [Текст] / Ю.П. Шабанов-Кушнарченко. – Харьков: Выща шк., 1984. – 144 с. 11. *Шрейдер, Ю.А.* О понятии “математической модели языка” [Текст] / Ю.А. Шрейдер. – М.: Знание, 1971. – 63 с.

Надійшла до редколегії 12.02.2010 р.

УДК 007. 519.714:519.766

Концепция унификации методов и средств построения параллельных многозначных структур языковых систем / Г.Г. Четвериков // Бионика интеллекта: науч.-техн. журнал. – 2010. – № 1 (72). – С. 3–11.

Статья посвящена анализу проблемы создания систем искусственного интеллекта, которые позволяют моделировать на логическом и аппаратном уровнях различные подходы математического описания естественного языка, и которые являются элементами k -значной структурной организации информационно-интеллектуальных технологий. Показана необходимость и возможность разработки общей теории построения интеллектуального интерфейса и систем искусственного интеллекта, которая могла бы стать методологической основой целеустремленного создания новых интеллектуальных и информационных технологий.

Библиогр.: 11 найм.

UDK 007. 519.714:519.766

The concept of unification methods and construction tools parallel multiple-valued structures language systems / G.G. Chetverikov // Bionics of Intelligence: Sci. Mag. – 2010. – № 1 (72). – P. 3–11.

Article is devoted to analysis problem of creation the artificial intellect systems which allow to model at logic and hardware levels the various approaches of the mathematical description natural language and which are elements of the k -unit structural organization of information-intellectual technologies. Necessity and possibility of working out the general theory of construction the intellectual interface and an artificial intellect systems which could become methodological basis of purposeful creation new intellectual and information technology is demonstrated.

Ref.: 11 items.