

УДК 519.62



Т. М. Федорова

ХНУРЕ м. Харків, Україна, tanja_fedorova@mail.ru

ПОБУДОВА ЛОГІЧНОЇ МЕРЕЖІ ДЛЯ ОБРОБКИ ФЛЕКСІЙ ІМЕННИКІВ ІТАЛІЙСЬКОЇ МОВИ

Стаття присвячена побудові математичної моделі обробки флексій логічною мережею для іменників італійської мови. Модель характеризується системою бінарних відношень, що задаються формулами відповідних предикатів, що дає можливість реалізовувати її на програмно – алгоритмічному рівні та використовувати для автоматичного вирішення задач аналізу, синтезу та нормалізації словоформи.

МОВА, ІМЕННИК, ГРАМАТИЧНІ ОЗНАКИ, ФЛЕКСІЯ, БІНАРІЗАЦІЯ, ВІДНОШЕННЯ, ДВОДІЛЬНИЙ ГРАФ, ПРЕДИКАТ, ЛОГІЧНА МЕРЕЖА

Вступ

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.

Для сучасного етапу розвитку науково-технічного прогресу характерне переважає неформалізованих чи недостатньо формалізованих проблем, алгоритмічне розв'язання яких чи то не існує, чи то не може бути отримане на наявних сьогодні машинах.

Оскільки всі розумові здібності, які необхідно передати машині, вже наявні у людини на достатньо високому рівні розвитку, ніякий інший інтелект, крім людського, науці недоступний, то міркувати машини будуть за тими ж законами, що й людина.

Лінгвістичні праці містять неформалізовані описи морфологічних явищ, які складно перекласти на математичну мову. Людина здатна розуміти їх «з півслова». Крім того, людина, що не приступала до вивчення морфології, фактично уже володіє мовою, морфологічні явища їй знайомі з власної мовної практики. Математичні описи морфології мови, які адресовані обчислювальній машині, мають бути зрозумілі їй. Запити інтелектуалізації потребують від комп'ютерів здатності розуміти людську мову, 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 рішення. 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 розуміти побачене

й почуте, спрямоване на вирішення аналогічних питань [18, 19].

Застосування комп'ютерних технологій в обробці текстової інформації ставить нові завдання та відкриває великі перспективи у сучасній лінгвістиці. 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 та лінгвістичних алгоритмів, адаптованих до використання в сучасних мовно-інформаційних системах, пошукових засобах, а також інших комп'ютерних системах, пов'язаних з інтелектуальним опрацюванням мовної інформації.

Актуальною постає проблема створення морфологічних класифікацій слів та засобів граматичної ідентифікації лінгвістичних структур у природномовних текстах. У цьому зв'язку необхідними виявляються комп'ютерні програми автоматичної граматичної ідентифікації лексичних одиниць, спроможні самостійно, без участі людини приписувати параметри граматичного значення текстовим словоформам [3, 4].

Побудова моделі флексійної обробки та логічної мережі для іменників італійської мови

Італійська мова є прямим нащадком латинської мови, носієм найдавнішої і багатой культури. Незважаючи на це, італійська є сучасною живою мовою, яка постійно розвивається і передбачає безліч нюансів і особливостей. Основними характеристиками італійської мови є широке лексичне багатство, величезна різноманітність стилістичних і виразних прийомів, складна граматична структура.

Серед мов романської групи італійська звертає на себе увагу як найбільш діалектно роздроблена. Діалекти можна умовно розділити на три групи: північні, центральні і південні. Серед північних у свою чергу виділяють галло-італійські (ломбардський, пьемонтський, генуезький та емільянський) і венеціанські. Центральні включають в себе наріччя Тоскани, Умбрії, Північного Лаціо, області Марке. До південної групи відносять

неаполітанський, абруцський, калабрійський і апулійський діалекти, а також групи сицилійських.

Італійська мова — одна з романських мов (іта-ло-романська підгрупа), в якій явище словозміни є притаманним для основних значущих частин мови. Певна кількість словозмінних форм утворюється синтетично, тобто шляхом приєднання до основи слова флексій, які містять в собі інформацію про конкретні значення однієї або декількох граматичних категорій; також є частина граматичних форм, які утворюються аналітично (граматичні форми складаються більше ніж з одного слова). Як і всі західнороманські мови, італійська втратила іменне відмінювання, а спосіб утворення множини в італійській відрізняється, наприклад, від іберороманської [5].

Грамматична будова італійської мови типова для романських мов. Відмінювання іменників здійснюється за допомогою прийменників (di, a, da, in, su та ін.). Середній рід відсутній. Іменники зазвичай супроводжує артикль — визначений (il, la, lo, l', i, gli, le), що розвинувся з латинського займенника ille, або невизначений (un, una, uno, un), що розвинувся з латинського числівника unus. Багато прийменників зливаються з визначенням артиклем в одне ціле. Система дієслівних часів складається з простих часів, які безпосередньо розвинулися з відповідних латинських, і складних, глибокі коріння яких йдуть також до латині.

Найсерйозніші зміни в порівнянні з народною латиною відбулися в італійській фонетиці. При цьому розвитку багатьох звуків та їх поєднань в італійській є унікальним і утворює особливий «італійський» тип фонетики.

Характерною рисою італійської мови є велика кількість голосних. Більше того, в італійській мові немає слів, що закінчуються на приголосний звук, крім визначеного артикля il і прийменників con, in, per. В італійській мові є одна приголосна літера, яка не виконує фонетичної функції (не вимовляється), але вона присутня у написанні слів — це буква Hh (акка). Подібно російському твердому знаку, вона виконує розділювальну функцію. Займаючи місце між приголосною і голосною літерами, вона надає твердість голосній. В італійській мові порядок слів вільний.

Вихідним матеріалом для моделювання режимів опрацювання слів на фонетичному та морфологічному рівнях природної мови є лінгвістичні закономірності. Математичним апаратом моделювання є алгебра скінченних предикатів [4].

В італійській мові іменники мають два роди: чоловічий (maschile) і жіночий (femminile) і два числа: однину (singolare) і множину (plurale). Італійські іменники чоловічого роду на -o/-e у множині в більшості випадків мають закінчення -i, а іменники жіночого роду на -a — закінчення -e.

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 Y . Із морфем слова у якості фрагменту вибрано закінчення. Нехай $X = \{x_1, x_2\}$ — набір граматичних ознак, які характеризують вплив дальнього тексту на закінчення (x_1 — число зі значеннями одина (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^{ion} — іменники з суфіксом -ion, y^0 — всі інші іменники.

В табл. 1, яка називається парадигматичною, представлені закінчення іменників італійської мови.

Таблиця 1

Парадигматична таблиця

x_1^o	$x_2^ч$	- o - e - a
	$x_2^ж$	- a - e
x_1^m	$x_2^ч$	- i
	$x_2^ж$	- e - i

Контекст, в який включено іменник, впливає на форму слова граматичними ознаками x_1, x_2 . Нумеруємо осередки парадигматичної таблиці. Для кожного осередку парадигматичної таблиці вводимо свій номер q (табл. 2).

Таблиця 2

x_1^o	$x_2^ч$	1
	$x_2^ж$	2
x_1^m	$x_2^ч$	3
	$x_2^ж$	4

Виражаємо номери q осередків парадигматичної таблиці через ознаки x_1, x_2 :

$$\begin{aligned} x_1^o x_2^ч &= q^1; & x_1^m x_2^ч &= q^3; \\ x_1^o x_2^ж &= q^2; & x_1^m x_2^ж &= q^4. \end{aligned}$$

Проводимо бінаризацію щойно записаного відношення, яке пов'язує змінну q із змінними x_1, x_2 . Для цього знаходимо відношення P_1 , що зв'язує змінні x_1 і q :

$$P_1(x_1, q) = x_1^o (q^1 \vee q^2) \vee x_1^m (q^3 \vee q^4).$$

Далі відношення, що отримуємо, зображуються також у вигляді дводольних графів.

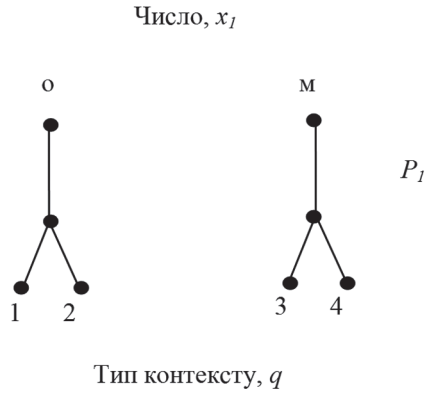


Рис. 1. Граф бінарного відношення $P_1(x_1, q)$

Знаходимо відношення P_2 , що зв'яже змінні x_2 і q :

$$P_2(x_2, q) = x_2^1 (q^1 \vee q^3) \vee x_2^2 (q^2 \vee q^4).$$

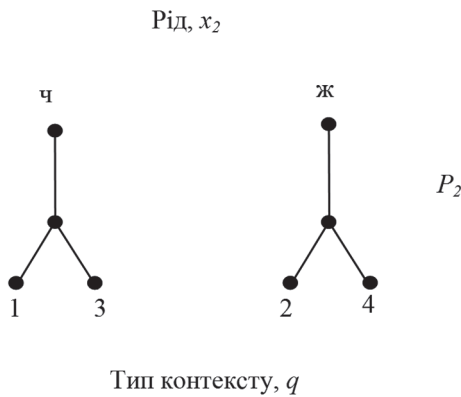


Рис. 2. Граф бінарного відношення $P_2(x_2, q)$

Розглянемо відношення P_3 , яке зв'яже номер впливу контексту закінчення z зі змінною q :

$$P_3(z, q) = (z^o \vee z^e \vee z^a) q^1 \vee (z^a \vee z^e) q^2 \vee z^i q^3 \vee (z^e \vee z^i) q^4.$$

Розглянемо відношення P_4 , яке зв'яже змінну основи слова y зі змінною закінчення слова z :

$$P_4(y, z) = (z^a \vee z^i \vee z^e) y^{ist} \vee (z^e \vee z^i) y^{ion} \vee (z^o \vee z^a \vee z^i \vee z^e) y^0.$$

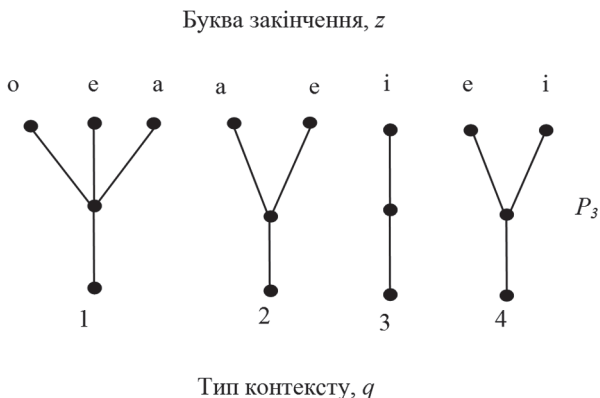


Рис. 3. Граф бінарного відношення $P_3(z, q)$

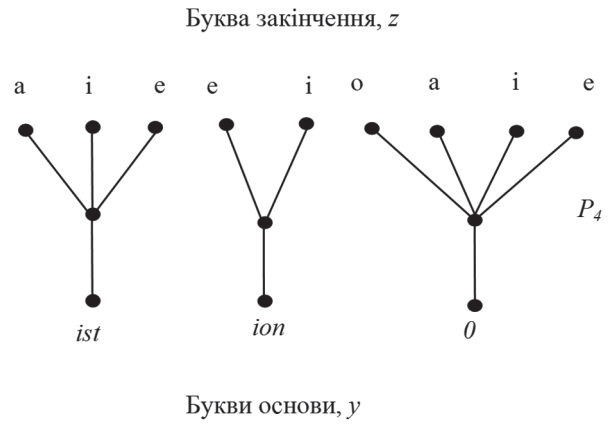


Рис. 4. Граф бінарного відношення $P_4(y, z)$

Побудовано математичну модель для іменників італійської мови. Вона характеризується системою бінарних відношень P_1-P_4 , що задаються формулами відповідних предикатів. Утворюючи кон'юнкцію всіх цих предикатів, отримуємо предикат моделі:

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

Предикату моделі відповідає відношення моделі P , що зв'яже між собою предметні змінні x_1, x_2, q, z, y . На рис. 5 зображена логічна мережа для відношення моделі P .

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, що відповідає даній гілці [6].

У побудованій мережі 5 полюсів і 4 гілки. Мережа має зовнішні і внутрішні 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 {о, м}, x_2 з значеннями {ч, ж}; 2) y з значеннями {ist, ion, o}; 3) z з значеннями {о, е, а, і}. 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.

Висновки

В статті була побудована модель флексійної обробки та логічна мережа для іменників італійської мови. Модель характеризується системою бінарних відношень та формулами відповідних предикатів. Мережа призначена для рішення системи рівнянь, що задаються відповідною моделлю.

Електронну схему, яка відповідає мережі, що зображена на рис. 2, можна використовувати для

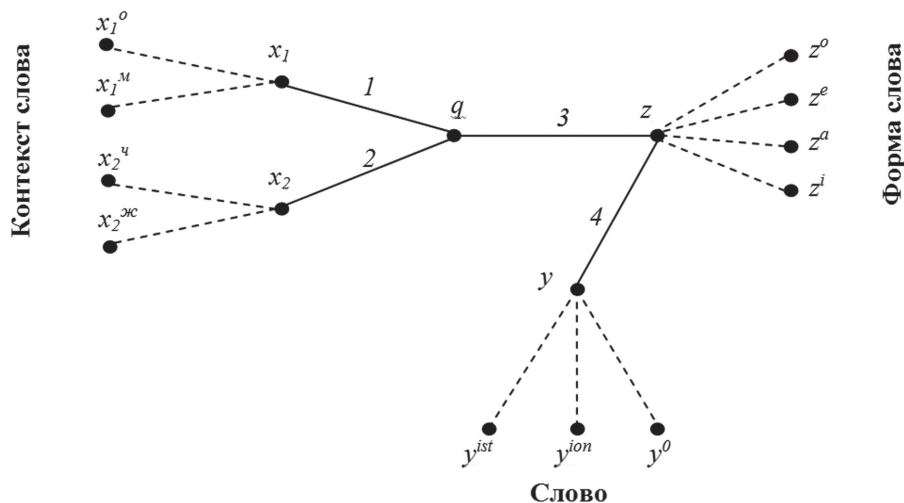


Рис. 5. Логічна мережа словозміни іменників італійської мови

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.

Список літератури: 1. *Бондаренко, М.Ф.* Основи території синтезу надшвидкодуючих структур мовних систем штучного інтелекту / М.Ф. Бондаренко, З.Д. Коноплянко, Г.Г. Четвериков. – К., 1997. – 263 с. 2. *Шабанов-Кушнарєнко, Ю.П.* Теорія інтелекту. Математическіє средства : моногр. / Ю.П. Шабанов-Кушнарєнко. – Х. : Вища шк., 1984. – 142 с. 3. *Широков, В.А.* Комп'ютерна лексикографія / В.А. Широков. – К. : Наукова думка, 2011. – 351 с. 4. *Широков, К.В.* Іменна словозміна у сучасній турецькій мові / К.В. Широков. – К. : Довіра, 2009. – 318 с. 5. *Чельшева, И.И.* Італьянський язык / И.И. Чельшева, Т.З. Черданцева // Языки мира. Романские языки. – М. : Academia, 2001. – С. 60. 6. *Логическая сеть как технология моделирования естественного языка / Ю.П. Шабанов-Кушнарєнко, В.И. Хаханов, Н.Т. Процай и др. //*

Информационные технологии – в науку и образование : материалы науч.-практ. конф. – Х. : ХНУРЭ, 2005. – С. 30–33.

Надійшла до редколегії 26.08.2013

УДК 519.62

Построение логической сети для обработки флексий существительных итальянского языка / Т.Н. Федорова // Бионика интеллекта: науч.-техн. журнал. – 2013. – № 2 (81). – С. 43–46.

Статья посвящена построению математической модели обработки флексий логической сетью для существительных итальянского языка. Модель характеризуется системой бинарных отношений и формулами соответствующих предикатов. Это позволяет реализовать ее на программно – алгоритмическом уровне и использовать для автоматического решения задач анализа, синтеза и нормализации словоформы.

Ил. 5. Библиогр.: 6 назв.

UDK 519.62

Building the logical network to handling the flexions of Italian nouns / T.N. Fyodorova // Bionica Intellecta: Sci. Mag. – 2013. – № 2 (81). – С. 43–46.

The article is devoted to constructing mathematical model of inflectional processing of logical network for the Italian nouns. The model is characterized by a system of binary relations and formulas related predicates. This allows to implement it on the program - algorithmic level and used to automatically solve the problems of analysis, synthesis and the normalization of the word form.

Fig. 5. Ref.: 6 items.