

Т. Юзепольська*

Київський національний лінгвістичний університет (Київ)
УДК 81'322**ПАРАДИГМАТИЧНІ КЛАСИ ІМЕННИКА АНГЛІЙСЬКОЇ МОВИ (КАТЕГОРІЇ ЧИСЛА: АНАЛІЗ ТА СИНТЕЗ)**

The article deals with the problem of classification of nouns: setting different types on the base of their plural forms. To differentiate the nouns morphological analysis has been performed. The data-base has been worked out that includes a great variety of noun characteristics. Applying this data-base, the application "Noun" has been created. It could help to find plural forms of the nouns and to transform the given Singular nouns into corresponding Plural forms.

Опрацювання текстової інформації незмінно залишається провідним завданням комп'ютерної лінгвістики. Наші знання про дійсність втілюються в певній вербальній формі. Навчити комп'ютер „розуміти” текст і означає наділити його здатністю видобувати з нього потрібну для виконання того чи іншого завдання інформацію. Таке „розуміння” тексту полягає у вмінні аналізувати його на різних рівнях представлення інформації: морфологічному, синтаксичному, логіко-семантичному і узагальнювати одержані внаслідок подібного аналізу результати у певній визначеній формі [2].

Об'єктом цього дослідження є морфологічні ознаки іменника англійської мови. Морфологічний аналіз слугує основою семантичної обробки інформації, що представляється природною мовою, і тому є необхідним компонентом процесу автоматичного індексування. В обчислювальній лінгвістиці під морфологічним аналізом розуміють аналіз окремих слів поза контекстом. У результаті морфологічного аналізу послідовність слів вихідного тексту замінюється на послідовність інформації про ці словформи.

Поняття „морфологічний аналіз” в обчислювальній лінгвістиці відрізняється від традиційного лінгвістичного уявлення про морфологічний аналіз. Відповідно до останнього морфологічний аналіз веде до визначення частини мови і морфологічної будови слова. Цілью аналізу є опис форми слова. В обчислювальній лінгвістиці морфологічний аналіз може слугувати засобом отримання різних даних, і навіть тих, що не стосуються морфології [1].

Предметом дослідження було обрано категорія числа іменників англійської мови, що в майбутньому може стати основою для автоматичного перекладу й розв'язання багатьох інших завдань комп'ютерної лінгвістики. Також результати морфологічного аналізу можуть бути вихідними даними етапів автоматичного аналізу науково-реферативного тексту — синтаксичного, лексичного, семантичного, аналізу текстових надфразних відношень, а також слугувати базою для виявлення особливостей морфологічної структури наукових рефератів.

Матеріалом дослідження є група іменників англійської мови, відібраних зі словника Macmillan Essential Dictionary (for learners of English). Саме на основі їх формального аналізу були визначені парадигматичні класи іменників і створена програма аналізу та синтезу форми множини.

Метою дослідження є створення програми, яка могла б утворювати форму множини заданого їй іменника, розпізнавати цю форму в тексті та визначати її початкову форму, тобто форму однини.

Визначаючи парадигматичні класи іменників, можна ще раз пересвідчитись у тому, що форма слова відіграє суттєву роль у визначенні певних граматичних категорій. Завдяки їй можна видобути з тексту багато граматичної інформації, яку потім можна використати у створенні певних систем автоматичного аналізу тексту. Однак, на жаль, не в усіх мовах форма говорить так ясно про свою граматичну приналежність. Це залежить від типу природної мови (аналітична чи флективна). Наприклад, у мови з багатоморфологією (українська чи латинська) аналіз окремо взятої словформи дає можливість отримати велику кількість різноманітної інформації про різні граматичні категорії. А в такій мові як англійська (аналітична) форма слова не дозволяє дістати всю ту інформацію, яку б нам хотілось. Більше того, виникає проблема лексико-граматичної омонімії

* © Т. Юзепольська, 2006

(наприклад, *answers* як форма множини іменника і як форма третьої особи однини дієслова; *(the) rich* як іменник, що позначає групу людей з певними особливими ознаками і як прикметник). Тому в таких випадках треба звертати увагу ще й на оточення того слова, що ми аналізуємо.

Створення бази даних «*mac noun*»

Щоб зробити класифікацію іменників за ознакою утворення множини, потрібно було опрацювати близько 11000 іменників, відібраних зі словника Macmillan. Для цього за допомогою Microsoft Access було створено базу даних "*mac noun*". Таблиця "*noun*" (рис. 1) складається з наступних полів: *id, wrd, c/u, rule_pl, plural, rule_pl2, plural2, signs, par_kl, chastotn, the, sphere, B/A, definition*.

В першу чергу було зафіксовано інформацію щодо сфери вживання (*formal (anomaly, refuse), informal (bike, flop), medical (diabetic, plasma), linguistics (gerund, participle)*) та частоти вживаності (частотності) деяких слів. Наступним етапом дослідження було визначення типу іменника (злічуваний (*balcony, ladder, yacht*), незлічуваний (*bodywork, marble, turbulence*)), що вживається лише в однині (*Singularia Tantum*) (*atmosphere, monotone, subconscious*), що вживається лише в множині (*Pluralia Tantum*) (*bedclothes, remains, acrobatics*), злічуваний або незлічуваний (*cake, exposure, size*) та ін.). Відомо, що не всі іменники англійської мови утворюють форму множини. Протягом цього етапу перевірка типу іменників відбувалася за допомогою словника Macmillan і граматики англійської мови. Таким чином було отримано ту групу іменників (7542 слів), над якою надалі буде проводитися аналіз. Потім були формально описані правила утворення форми множини простих та складених іменників (<e* (*basis, congressman, saleswoman*), >s (*anagram, house, formula*), <s-in-la >w (*brother-in-law, sister-in-law*), <v >es (*elf, werewolf, sheaf*) і т. д.). На завершальному етапі було визначено 98 парадигматичних класів на основі типів іменників, правил утворення форми множини та кількості слів в одному іменникові.

Створення програми «*Noun*»

Існує два підходи до створення програм такого роду: на основі певних правил та на основі словника. Програма, основана на першому підході виконує синтез та аналіз множини на основі заданих їй формальних правил. Зазвичай така програма займає зовсім небагато місця, але її недоліком є досить значна кількість помилок, адже неможливо задати абсолютно всі правила і врахувати усі тонкості мови. Програма, створена на основі словника, звичайно займає набагато більше місця, але в ній менше помилок. А так, як зараз комп'ютерні об'єми пам'яті майже необмежені, то, на мою думку, доцільніше використовувати саме такий підхід. Що й було зроблено в даному дослідженні.

Сам робочий словник, а саме форму множини усіх слів (адже ми мали лише список слів у формі однини), було отримано за допомогою допоміжної програми, написаної мовою C#. Програма "*Noun*" має два блоки (рис. 2): аналіз та синтез множини іменника.

У результаті дослідження парадигматичних класів іменників англійської мови було зроблено наступне:

- 1) Створена база даних з 11000 іменників, у якій міститься детальна характеристика кожного слова.
- 2) Сформульовані формальні правила утворення форми множини.
- 3) Визначений тип іменника: *countable or uncountable, countable, uncountable, plural, singular, singular or uncountable* та ін.
- 4) Призначений номер парадигматичного класу.
- 5) Визначена сфера вживання.

Рисунок 1. База даних "*mac noun*". Таблиця "*noun*"

id	wrd	c/u	rule_pl	plural	par_kl	signs	rule_pl2
6661	four		phrase		97		
6662	four-by-four	c	>s	four-by-fours	1		
6663	four-poster	c	>s	four-posters	1		
6664	foursome	c	>s	foursomes	1		
6668	Fourth of July		sing		95 vl		
6669	four-wheel drive	c	>s	four-wheel drives	3 2w		
6670	fowl	c	*	fowl	45 2p	>s	fc
6671	fox	c	>es	foxes	4		
6673	foyer	c	>s	foyers	1		
6674	fracas		sing		95		
6675	fraction	c	>s	fractions	1		
6677	fracture	c	>s	fractures	1		
6680	fragment	c	>s	fragments	1		
6682	fragrance	c/u	>s	fragrances	57		
6685	frailty	c/u	<i >es	frailties	62		
6686	frame	c	>s	frames	1 c/plural		
6688	framework	c	>s	frameworks	1		
6689	franchise	c	>s	franchises	1		
6691	frankfurter	c	>s	frankfurters	1		
6695	fraternity		u		94		
6697	fraud	c/u	>s	frauds	57		
6701	fray		sing		95		
6703	freak	c	>s	freaks	1		
6706	freckles		plural		92		

Цей етап дослідження вказує на те, що морфологічний аналіз слова дозволяє автоматизувати певні види роботи з мовними одиницями, що значно полегшує працю (роботу) лінгвістів. На основі цієї бази даних було створено програму «Noun», яка виконує аналіз та синтез форми множини англійського іменника. Вона утворює множину іменника від заданого, а також розпізнає у тексті усі іменники, що стоять у формі множини і потім визначає їх форму однини. Результати даного дослідження доводять, що проблема зняття омонімії залишається актуальною так само, як і розрізнення типу іменника, який одночасно може бути і злічуваним, і незлічуваним.

Рисунок 2. Програма «Noun»



Література

1. Василевский А. А., Марчук Ю. Н. Вычислительная лингвистика. – М., 1970.
2. Волошин В. М. Вступ до комп'ютерної лінгвістики (Донецький національний університет). — Донецьк: ТОВ „Юго-Восток, Лтд”, 2003. — 184 с.
3. Использование ЭВМ в лингвистических исследованиях. – К.: Наук. думка, 1999.
4. Каушанская В. Л. Грамматика английского языка. — М., 2000.
5. Качалова К. Н., Израилевич Е. Е. Практическая грамматика английского языка с упражнениями и ключами. — СПб.: БАЗИС, КАРО, 2004. — 608с.
6. Морфологический анализ научного текста на ЭВМ. – Киев: Наук. думка, 1989.
7. Digby Beaumont, Colin Granger. The Heinemann ELT English Grammar.
8. Macmillan English Dictionary (for learners of English).

В. Дембіцький*

Київський національний лінгвістичний університет (Київ)
УДК 81'33

СТУПЕНІ ПОРІВНЯННЯ ПРИКМЕТНИКІВ І ПРИСЛІВНИКІВ. СИНТЕЗ І РОЗПІЗНАВАННЯ В ТЕКСТІ ПАРАДИГМАТИЧНИХ КЛАСІВ АНГЛІЙСЬКИХ ПРИКМЕТНИКІВ ТА ПРИСЛІВНИКІВ

Стаття присвячена проблемі комп'ютерного оброблення письмових текстів, зокрема морфологічного аналізу. Автор пропонує програму, яка, керуючись правилами утворення ступенів порівняння, автоматично утворює їх для відібраних з англо-українського навчального словника прикметників та прислівників. Наступний модуль програми дозволяє виділити прикметники та прислівники у тексті, а також пояснює ступінь порівняння знайденого слова.

Здатність комп'ютера опрацювати мовний матеріал обмежена набором операцій і засобів, котрі доступні обчислювальній техніці на сучасному етапі її розвитку. Ми можемо використовувати комп'ютер як помічника при умові, що сформулюємо йому завдання так, щоб його можна було б викласти однією з мов програмування. Обов'язковою складовою частиною лінгвістичного забезпечення будь-якої системи

* © В.Дембіцький, 2006