

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 опрацьовувати 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

автоматичної обробки письмових текстів є морфологічний аналіз, який, власне, і є одним з основних етапів роботи системи автоматичного аналізу тексту.

На основі опрацьованих (досліджених) прикметників та прислівників слід розробити алгоритми утворення парадигми англійських прикметників та прислівників (де це можливо), а також принципи розпізнавання їх (у будь-якому ступені порівняння) у тексті та утворення від вищого та найвищого ступенів початкової форми прикметника або прислівника, а також створити програму, яка б практично застосовувала отримані принципи (правила).

Для проведення дослідження з навчального англо-українського словника другого рівня відібрано усі реєстрові прикметники та прислівники і збережено у таблицях бази даних:

прикметників 721;
прислівників 260;

Згідно з граматиною англійської мови розроблено правила утворення і розпізнавання ступенів порівняння вищезгаданих частин мови.

На мові програмування C# складена програма, яка, керуючись правилами утворення ступенів порівняння, автоматично їх утворює для відібраних з англо-українського навчального словника прикметників та прислівників. Другий модуль програми дозволяє виділити прикметники та прислівники у тексті, а також пояснює ступінь порівняння знайденого слова.

Характеристика досліджуваних частин мови.

Залежно від значення прикметники поділяють на якісні (qualitative) та відносні (relative). Більшість якісних прикметників мають ступені порівняння, тоді як відносні взагалі не мають ступенів порівняння. Однак навіть деякі якісні прикметники не мають ступенів порівняння (greenish, darkish, incurable, unsuitable, chief, principal...). Більшість якісних прикметників можуть утворювати прислівники (adverbs) за допомогою додавання суфікса -ly.

Ступені порівняння прикметників.

Більшість якісних прикметників мають ступені порівняння (degree of comparison):

- **початкова форма (positive degree):**

початкова форма вказує на якість предмета і відповідає словниковій формі, тобто прикметники у початковій формі не мають ніяких закінчень: difficult, green. Однак часто коли говорять про однакову ступінь якості різних предметів використовують сполучник «as ... as» – «такий же..., як» або його заперечний варіант «not so ... as» – «не такий ..., як».

The line AB is as long as the line CD.

Конструкція «as (adjective) as» може бути доповнена у передпозиції прислівниками not, almost, twice, three times, half, one-third та ін.

Mabel is as clever an administrator as Robin.

New York is as distant from San Francisco as Boston is from London.

- **вищий (comparative degree) – позначає вищий рівень якості:**

наприклад, якщо потрібно вказати, що один предмет володіє більш вираженою ознакою у порівнянні з іншим предметом, то використовують прикметник у вищому ступені, який утворюється шляхом додавання суфікса «-er» до основи прикметника (який складається з одного або двох складів) наприклад:

short – shorter, dark – darker

- **найвищий (superlative degree) – позначає найвищий рівень якості,**

вказує на те, особа, предмет чи група володіє найбільш вираженою ознакою у порівнянні з усіма іншими (наприклад, biggest/(the) most beautiful)

Розбіжності у способах творення ступенів порівняння (одночасно і синтетичний, і аналітичний способи) спричинені історичним розвитком англійської мови.

Прикметники (як і прислівники) утворюють свої форми порівняння наступними способами:

1. За допомогою закінчення «-er»; «-est» (синтетичний спосіб).

Прикметники, які утворюють ступені порівняння шляхом додавання закінчень (синтетичним способом), включають односкладові прикметники та двоскладові з закінченням «-у». Наприклад: brave (1)¹, happy (2).

Ще кілька двоскладових прикметників інколи вживаються з закінченнями «-er» / «-est», наприклад: clever, quiet.

Варто відзначити, що односкладові дієслова у формі дієприслівників минулого часу (past participle), які у тексті вжиті як прикметники, зазвичай вживаються без закінчень.

2. За допомогою додавання прислівників more/most перед прикметником (аналітичний спосіб).

Як видно з наведеної вище інформації, односкладові прикметники зазвичай утворюють вищий і найвищий способи порівняння у перший спосіб (синтетичний), а багатоскладові – у другий спосіб (аналітичний).

Правила утворення ступенів порівняння прикметників:

Регулярні прикметники

Односкладові прикметники утворюють вищий ступінь порівняння шляхом додавання до форми прикметника у початковій формі суфікса «-er». Найвищий ступінь порівняння утворюється способом додавання суфікса «-est» (синтетичний спосіб): hard – harder – (the) hardest.

Односкладові прикметники, які закінчуються на «-e» утворюють ступені порівняння додаванням «-r», «-st»: Nice – nicer – (the) nicest. А якщо основа прикметника закінчується на букву «-у» з попереднім приголосним, то при додаванні суфікса «-er/-est» буква «-у» переходить в «-i»: dry – drier – (the) driest; easy – easier – (the) easiest.

В односкладових прикметників з комбінацією літер типу cvc (consonant-vowel-consonant/приголосний-голосний-приголосний), де кінцевим є будь-який єдиний приголосний, крім “w”, “x” або “y”, перед «-er»/ «-est» подвоюється кінцева приголосна у вищому і найвищому ступені порівняння, щоб зберегти закритий склад: big – bigger – (the) biggest.

Коли прикметник закінчується на “w”, “x” or “y”, перед яким знаходиться єдиний наголошений голосний, останній приголосний не подвоюється перед додаванням «-er»/ «-est», наприклад: slow – slower, gray – greayer.

Варто відзначити, що коли прикметник закінчується на єдиний приголосний, який стоїть після двох голосних, останній приголосний не подвоюється перед додаванням «-er»/ «-est», наприклад: loud – louder, neat – neater, soon – sooner.

Однак існують деякі багатоскладові прикметники, які утворюють вищий і найвищий ступені порівняння за допомогою закінчень (синтетичним способом). Деякі двоскладові прикметники також можуть утворювати ступені порівняння за допомогою «-er»/ «-est» (як правило прикметники, які закінчуються на -y; -ow; -er; -le), а також слова (polite, severe, sincere, common):

Прикметники, що мають 2 склади, але з наголосом на останньому складі:

complete – completer – (the) completest

concise – conciser – (the) concisest

Двоскладові прикметники, які закінчуються на «-у», змінюють «-у» на «-i»:

happy – happier – (the) happiest

nasty – nastier – (the) nastiest

Двоскладові прикметники, які закінчуються на ненаголошену голосну:

hollow – hollower – (the) hollowest (having a cavity, sunken)

Двоскладові прикметники, які закінчуються на [l], тобто «-el», «-le»

simple – simpler – (the) simplest

Двоскладові прикметники, які закінчуються на [r]

clever – cleverer – (the) cleverest

Багато двоскладових прикметників і всі прикметники, які мають більше двох складів утворюють ступені порівняння за допомогою відповідно more та most, less та least. Ці слова ставляться перед прикметниками у початковій формі (зазвичай цим способом утворюють ступені порівняння похідні та складні якісні прикметники, про які йшлося вище):

active – more active – (the) most active

¹ У дужках вказано кількість складів.

difficult – more difficult – (the) most difficult

Нерегулярні прикметники:

Деякі прикметники мають нерегулярні форми у вищому та найвищому ступенях:

good – better – (the) best

bad – worse – (the) worst

Деякі прикметники мають дві форми порівняння: друга має особливе значення,

яке, власне, є зовсім іншим словом:

many – more – most (with countable nouns)

much – more – most (with uncountable nouns)

far – farther – farthest (of distance)

far – further – furthest (of distance or time)

Further має ще одне значення, наприклад воно може мати значення «додатковий» (additional, extra).

Do you need any further information?

Near – nearer – nearest

near – nearer – next

Наприклад: «What is the name of the next village?»

(old – older – oldest) & (old – elder – eldest)

(late – later – latest) & (late – latter – last)

Прислівник – це частина мови, яка виражає деякі умови, що стосуються дії або стану, або відзначає певні характерні риси дії або якості. Прислівники вказують на те, коли, де, чому, за яких обставин чи як (у який спосіб) щось зроблено (у той час як прикметники показують ознаку предмета чи особи), або щось відбувалось чи відбувається. Основна синтаксична функція прислівника – обставинна. Прислівники часто закінчуються на «-ly», однак багато з них не мають такого закінчення, і до того ж воно не завжди означає, що слово з цим закінченням – прислівник. Наприклад, існує кілька прислівників на «-ly», які не мають відповідних собі прикметників: friendly, likely.

Деякі прислівники **мають дві форми**: в одній слово закінчується на «-ly», а у другій ні. У деяких випадках, ці дві форми мають різні значення:

He arrived late.

Lately, he couldn't seem to be on time for anything.

Прикметник прислівник(значення)

прислівник(значення)

good

well

difficult

with difficulty

public

publicly

deep

deep (place)

deeply (feeling)

direct

direct

directly (=soon)

Треба відзначити, що *well* крім прислівника, може бути ще й прикметником, що означає *healthy* (здоровий). Прикметник *well* частіше вживається як предикативний ад'єктив.

Well (предикативний прикметник, означаючий ім. you): *I hope you are well.*

Well (прислівник способу дії): *He did well on the examination.*

Подані нижче прикметники можуть також вживатись як прислівники (без змін): fast, hard, little, loud, hourly, enough, long.

Прислівники з і без «ly»

У наступній таблиці подано приклади пар прислівників, які тісно пов'язані, але мають різні значення.

Прислівник	Значення	Прислівник	Значення
hard	with effort	hardly	scarcely
high	opposite of low	highly	very; very well
late	opposite of early	lately	recently
near	opposite of far	nearly	almost
wide	opposite of narrow	widely	commonly

Ступені порівняння прислівників:

Більшість прислівників, головним чином прислівники способу дії (серед яких прості (simple) та похідні (derivative)), можуть мати ступені порівняння, що утворюються так само, як і ступені порівняння прикметників.

Прислівники, які мають ту ж початкову форму, що і відповідні прикметники, мають ту ж саму форму у вищому та найвищому ступенях порівняння. Наприклад, якщо слово має лише один склад, вищий ступінь утворюється за допомогою додавання до простої форми «-er» для утворення вищого ступеня, а щоб утворити найвищий «-est»: fast – faster – tastest, near – nearer – nearest.

Ми часто використовуємо more та most, less та least, цим самим виражаючи ступінь порівняння прислівника. Тут перша пара слів вживається для нарощення ознаки, а друга — для відображення спадання ознаки. Зазвичай прислівники, утворені від прикметника, інколи вживаються з закінченням «-ly», утворюють вищий ступінь саме таким способом: *clearly – more clearly – most clearly, softly – more softly – most softly*.

Варто відзначити, що прислівники, утворені додаванням «ly» до односкладового прикметника, інколи вживаються з закінченням «er» або «est». *We walked slower and slower.* Однак у сучасній англійській мові буде правильніше сказати: *We walked more and more slowly.*

Ступені порівняння прислівників often (часто), quickly (швидко), slowly (повільно) утворюються обома способами: often – more often/oftener – most often/oftenest; quickly – more quickly/quicker – most quickly/quickest; slowly – more slowly/slower – most slowly/slowest.

Деякі прислівники, як і відповідні їм прикметники, утворюють ступені порівняння не за правилами (нерегулярні форми). Вони мають ті ж форми порівняння, що і відповідні нерегулярні прикметники: well – better – best; far – farther/further – farthest/furthest.

Важливо, що такі прислівники, як sometimes, never, here, there, now, then, first, again, yesterday і daily, не мають ступенів порівняння.

Робота основного алгоритму визначення прикметників і прислівників, вжитих у вищому та найвищому ступенях (головним чином утворених синтетичним способом), а також і у початковій формі, побудована на аналізі кінцевих ланцюжків літер (квазіфлексій) слововживань, які відбивають особливості формотворчих елементів, властивих даному класу слів. Фактично, аналіз починається зі знаходженням основи слова у словнику і виділенням можливих змін у морфемному складі. Враховуючи зв'язок автоматичного морфологічного аналізу зі словником потрібно згадати про морфологічний аналіз зі словником основ.

Цей принцип використовується досить часто. Процес пошуку починається з того, що кожна словоформу, взяту з тексту, порівнюють на (повне) збігання з лексичними одиницями словника. Якщо словоформа не знайдена, то від слова відсікається потенційна формотворча морфема і знову відбувається пошук на збігання, але тепер з основою. Якщо вибрана і словникова форми основи збіглися, то першій надається інформація, яка міститься у словнику при початковій формі. Якщо шуканої словоформи в словнику немає, то вона вважається такою, що не належить до цієї частини мови або просто не внесена до бази даних.

У кожній мові існують особливості суфіксального словотворення і словозміни, які ускладнюють алгоритм розпізнавання основ. Так, при знаходженні англійської словоформи “bigger” при пошуку відкидається спочатку “er”. Залишок “big” шукається у словнику.

З вищесказаного випливає, що алгоритм власне морфологічного аналізу залежить від того, що розуміється під основою слова. Словник основ може містити одиниці двох видів: а) канонічні основи, тобто основи, які містяться у звичайному словнику (іменник у називному відмінку однини, дієслово в інфінітиві, прикметник і прислівник у початковій формі); б) основи, представлені графічно незмінною частиною словоформ (тобто основою).

Класифікація словоформ при використанні канонічних основ відбувається для прикметників та прислівників наступним чином: існують лише три форми (початкова, вищий та найвищий ступені порівняння).

Розробка принципів автоматичного аналізу і синтезу прикметників і прислівників англійської мови (початкова форма, вищий та найвищий ступінь).

Згідно з граматиною англійської мови прикметники та прислівники можуть утворювати ступені порівняння лише двома способами — аналітичним та синтетичним (або не утворювати їх взагалі).

Для того, щоб розробити систему автоматичного синтезу ступенів порівняння англійських прикметників, як і для утворення парадигми будь-якої іншої самостійної частини мови, перш за все потрібно укласти систему правил, які було б легко інтерпретувати під комп'ютерну програму. Тут обов'язково потрібно врахувати усі можливі нерегулярні випадки утворення парадигматичних форм, оскільки у іншому випадку програма просто буде невільно утворювати парадигму.

Так, якщо подивитись на прикметники, варто відзначити, що не всі з них можуть утворювати ступені порівняння, а деякі можуть утворювати тільки найвищий ступінь (utter – ...-utmost/uttermost). Суплетивні форми типу good – better – best потрібно просто задати окремо як винятки.

Основна проблема полягає у тому, чи має слово ступінь порівняння взагалі, навіть якщо воно і підпадає під певне правило словозміни, оскільки серед прикметників ступені порівняння мають лише більшість якісних, та й серед прислівників процент тих, які можуть утворювати ступені порівняння — досить невисокий.

Згідно з граматиною для прислівників та прикметників була розроблена система правил словозміни при утворенні ступенів порівняння. Таблиця з правилами подана нижче:

Таблиця . Уніфікована система правил утворення ступенів порівняння прикметників та прислівників англійської мови

правило	вищий	найвищий	приклад	пояснення
PR0	—	—	—	ступенів пор. нема
PR1	>er	>est	fast-faster-fastest	додавання частки до основи
PR2	<*>*er	<*>*er	big-bigger-biggest	подвоєння останньої приголосної плюс додавання частки до основи
PR3	>r	>st	wide-wider-widest	додавання частки коли основа закінчується на -e
PR4	<i>er	<i>est	sly-slier-sliest	додавання частки коли слово закінчується на -y
PR5	more <>	most <>	interesting – more interesting – most interesting	аналітичний спосіб
PR6	<***>	<***>	good-better-best	суплетивні форми

Розроблена система правил повністю задовольняє усі можливі випадки утворення парадигми англійських прикметників і прислівників, включаючи винятки, випадки, коли слово взагалі не має ступенів порівняння, а також коли утворює їх аналітичним або синтетичним способами.

Стосовно синтезу ніяких проблем не виникає через те, що кожне слово згаданих частин мови (за кількома винятками) має усталену форму утворення парадигми і повністю підпадає під відповідне правило словозміни. З іншого боку, потрібно відзначити, що при аналізі виникає проблема з омонімією (прикметника і прислівника, прислівника і прийменника, прикметника та іменника і т. д.). У зв'язку з цим без розроблення окремого алгоритму розпізнавання омонімії не можна стовідсотково автоматично опрацювати текст і розмежовувати ці частини мови, а потрібне втручання користувача у процес розпізнавання.

Суттєвою перевагою створеної програми є те, що вона може доповнюватись лексичними одиницями (оскільки зараз база даних містить лише прислівники та прикметники, які входять до навчального англо-українського словника II рівня).

Література

1. Использование ЭВМ в лингвистических исследованиях / Т. А. Грязнухина, Н. П. Дарчук, Н. Ф. Клименко и др.; Отв. ред. В. И. Перебийнос; АН УССР. Ин-т языковедения им. А. А. Потебни.— Киев: Наук. думка, 1990. — 288 с.

2. Крылова И. П., Гордон Е. М. Грамматика современного английского языка: Учебник для ин-тов и фак. иностр. яз.— 5-е изд.—М.: Книжный дом «Университет», 2000.— 488 с.— На англ. яз.
3. Рыбаков Ф. И. и др. Автоматическое индексирование на естественном языке/ Рыбаков Ф. И., Руднев Е. А., Петухов В. А. — М.: Энергия, 1980. — 160 с.
4. Карпіловська Є. А. Вступ до комп'ютерної лінгвістики / Донецький національний університет. — Донецьк: ТОВ «Ого-Восток, Лтд», 203. — 184 с.
5. Качалова К. Н., Израилевич Е. Е. Практическая грамматика английского языка: В 2-х т. К.: Методика, 2000 г. — 368 с.
6. Положення про навчально-наукові та кваліфікаційні роботи студентів Київського державного лінгвістичного університету / Укл.: Болдирєв Р. В., Соловей М. І., Сажко Л. А., Спіцин Є. С. — К.: Вид. центр КДУ, 1999. — 46с.
7. М. А. Ganshina, N. M. Vasilevskaya English Grammar: Ninth edition revised — М.: Higher school publishing house, 1964. — 548p.
8. Освітні лінгвістичні Інтернет-портали: www.webster.comnet.edu; www.ego4u.com; www.paulnoll.com; www.vionet.hit.bg; www.wordpower.ws; www.algebra.com; www.fortunecity.com; www.ielanguages.com

О. Бугаков*

Український мовно-інформаційний фонд НАН України (Київ)
УДК 81'322.5

ВСТАНОВЛЕННЯ ЗОН ПРИЙМЕННИКОВИХ ЗВ'ЯЗКІВ В УКРАЇНСЬКОМУ ТЕКСТІ (МОДУЛЬ АВТОМАТИЧНОГО СИНТАКСИЧНОГО АНАЛІЗУ)

The principles of constructing the algorithm for identification of prepositional connections zones in the Ukrainian text (module of parsing) are given. The algorithm has been constructed on the material of linguistic database as a subcorpus of the Ukrainian National Linguistic Corpus according to the analysis of prepositional connections zones in syntactic structure of the Ukrainian sentence.

Прийняте в Українському мовно-інформаційному фонді (УМІФ) поетапне здійснення синтаксичної розмітки лінгвістичного корпусу текстів базується на принципі часткової розмітки з використанням методу фракційного синтаксичного аналізу. Суть останнього полягає у побудові синтаксичної структури речення з окремих фракцій – синтаксичних конструкцій певного типу, а саме: ланцюжків слів, об'єднаних предикативним зв'язком, іменних безприйменникових конструкцій, прикметниково-іменникових конструкцій, прийменникових конструкцій (зон прийменникових зв'язків), на аналіз яких, власне, і спрямоване наше дослідження. Шляхом об'єднання синтаксичних конструкцій одержуємо повне представлення синтаксичної структури речення. Звідси й назва – фракційний синтаксичний аналіз.

Зона прийменникових зв'язків (ЗПЗ) містить прийменник, головне слово (тобто слово, яке керує прийменниково-іменниковою синтаксею) і залежне слово (тобто слово, яке підпорядковується головному за допомогою прийменника).

Метою дослідження є побудова алгоритму встановлення зон прийменникових зв'язків в українському тексті (як окремого модуля автоматичного синтаксичного аналізу (АСА)) з урахуванням якісних та кількісних характеристик організації прийменникових зв'язків в українському тексті.

Виведення алгоритмічних правил визначення ЗПЗ передбачає проведення аналізу на репрезентативному матеріалі з метою забезпечення належного рівня достовірності результатів. Таким матеріалом слугував Український національний лінгвістичний корпус (УНЛК) обсягом 36 млн. слововживань, створений в УМІФі.

Програмне забезпечення УНЛК, побудованого як інформаційна система, дозволяє створювати спеціалізовані субкорпуси, орієнтовані на розв'язання поставлених завдань. За допомогою спеціально розробленої в УМІФі програми зазначені субкорпуси переводяться у форматі баз даних з певною структурою, орієнтованою на проведення конкретних досліджень синтаксичної структури тексту. Лінгвістичні бази даних (ЛБД), які виконують функцію інструмента і матеріалу дослідження мовного явища, структуровані за таким принципом: текстовим сегментам (контекстам), що містять конкретну мовну одиницю (прийменник), ставляться у відповідність визначені наперед диференційні ознаки, за якими здійснюється аналіз. Структурування ЛБД за полями, що

* © О.Бугаков, 2006