

найхарактернішими представниками ад'єктивної лексики з високим семантичним потенціалом. Ці прикметники можна об'єднати в декілька груп. Більшість прикметників із високим потенціалом полісемії є конкретними, основні значення яких характеризують ті чи інші сторони базових понять природознавства, що у процесі когнітивної діяльності були поширені й на інші предмети. Найпомітнішу частину прикметників із високим семантичним потенціалом складають ті з них, що позначають просторово-часові відношення, наприклад: *high, short, quick, running*, ознаки з точки зору сили, міцності, гостроти: *strong, weak, tight, loose, set*, слова на позначення відчуттів, наприклад: *cold, dry, light, soft, hard, heavy* та інші. Досить помітну частину високополісемічних прикметників складають оцінні слова, наприклад: *nice, fine, fair, good, foul* та інші.

Очевидним є той безперечний факт, що основна маса прикметників із високим потенціалом – прості (90 %). І тільки незначна частина слів цього діапазону є афіксально маркованими, точніше, похідними, що цілком узгоджується з рисунком 3. Складних ад'єктивних одиниць із високим семантичним потенціалом не виявлено. Наші дослідження також показали, що переважна більшість цих слів – англосаксонського походження.

На основі проведених досліджень ми робимо висновок про те, що всі полісемічні прикметники англійської мови можна умовно поділити на три групи залежно від величини притаманного їм потенціалу творення нових значень. Першу групу складають прикметники зі звичайним потенціалом полісемії, другу групу утворюють слова з підвищеним потенціалом, третю групу становлять прикметники з високим потенціалом полісемії, що зумовлено словотвірними та семантичними чинниками. Як бачимо, нескладний математичний аналіз статистичних даних дозволив нам зробити цілком конкретні лінгвістичні висновки. Впадає у вічі той факт, що семантичний потенціал наростає не плавно, а стрибкоподібно, тобто величина потенціалу квантується (змінюється дискретно) – звичайний → підвищений → високий, що дає змогу стверджувати про наявність своєрідного „квантового ефекту в лінгвістиці”.

У перспективі доцільно провести аналогічні дослідження семантичного потенціалу різних лексико-граматичних класів слів у різних мовах, а також здійснити порівняльний аналіз семантичного потенціалу певних груп слів.

Література

1. Биофизика / Костюк П. Г., Гродзинский Д. М., Зима В. Л., Магура И. С., Сидорик Е. П., Шуба М. Ф. – К.: Выща школа, 1988. – 504 с.
2. Папп Ф. О некоторых количественных характеристиках словарного состава языка // Slavica VII. – Debrecen, 1967. – С. 51–58.
3. Поликарпов А. А., Каримова Г. О. Словарь «Новые слова и значения» (НСЗ–1984) как источник данных об эволюции языка // Прикладная лингвистика и автоматический анализ текста. Тез. докл. научн. конф. – Тарту, 1988.
4. Тулдава Ю. А. О некоторых квантитативно-системных характеристиках полисемии // Учен. зап. Тартуского ун-та. – Вып. 502. Linguistica XI. – Тарту, 1979. – С. 107–141.

С. Курстак*

Київський національний лінгвістичний університет (Київ)
УДК 81'322

АЛГОРИТМ ЗНЯТТЯ ОМОНІМІЙ ТИПУ «ІМЕННИК-ДІЄСЛОВО» В АНГЛІЙСЬКІЙ МОВІ ТА ІМПЛЕМЕНТАЦІЯ ЙОГО В ПРОГРАМНОМУ ПРОДУКТІ

Актуальність статті зумовлена тим, що жодне точне дослідження словникового складу мови не може обійтися без глибоко розгляду питань омонімії. Не менш важливою є й розробка проблем омонімії для вирішення задач інженерної, структурної, комп'ютерної лінгвістики в цілому.

Омонімія (від гр. *homos* – однаковий і *опута* – ім'я) – це прояв особливої якості мовного знаку: різним означуваням можуть відповідати тотожні означення. Через це омонімія створює труднощі в процесі комунікації. Часто слухач опиняється в складній ситуації: яке з декількох значень, які виражаються певною мовною формою, треба обрати для правильного розуміння повідомлення. Труднощі можуть виникати не лише в слухача (читача), а й в мовця (письменника), який задля оптимізації процесу спілкування

* © С.Курстак, 2006

намагається будувати повідомлення так, щоб його сприйняли однозначно. Окрім іншого, омонімія додає складності у вивченні іноземної мови, коли студент стикається з такою проблемою: одна й та сама мовна форма може мати абсолютно різні значення – факт, на який в рідній мові він зазвичай не звертає уваги. Аналіз таких форм суттєво погіршує сприйняття іншомовного тексту або повідомлення.

Значні труднощі додає омонімія при створенні систем автоматичного аналізу тексту, а також в лексикографічній роботі: при виборі одиниць для частотного словника, при визначенні меж між окремими словами в загальному словнику мови, при розробці способів представлення в ньому омонімічних слів і т.п.

Повне вирішення проблем омонімії наразі є недосяжною мрією, але воно розпочне революцію в найрізноманітніших галузях науки і повсякденного життя: від машинного перекладу і штучного інтелекту до побутових роботів і автоматичних систем забезпечення життя та надання послуг. Всі ці системи оперують з мовним матеріалом і нерідко – з звуковим. Такі системи зазвичай не мають права на помилку, тому повідомлення які вони сприймають/відсилають повинні бути стовідсотково однозначними. І хоча вивчення омонімії ведеться вже давно, нажаль в цій області досягнення невеликі. А час вимагає поглибленого дослідження проблем омонімії, вияву її внутрішніх закономірностей, визначення місця, яке вона займає в загальній системі мови.

Основним структурним елементом омоніміки в мові є **омонімічний ряд**. Омонімічним рядом називається сукупність слів (лексем) або словоформ однієї й тієї ж мови в один і той самий період його існування, взаємно тотожних в плані вираження, але суттєво різних в плані змісту.

Омонімічний ряд – це свого роду макросистема, цілісність якої базується на загальності форми її елементів. Вивчення омонімічного ряду може йти як по лінії дослідження його системних характеристик, так і по лінії аналізу його структури. В першому випадку омонімічний ряд розглядається як складна єдність, і тоді нас цікавлять його якості як цілого, в другому ж він розглядається як розчленована єдність, і тоді нас цікавить його склад і внутрішня організація. В курсовій роботі переважно використовується другий підхід.

Омоніміка мови представляє собою сукупність всіх омонімічних рядів мови, одні з яких знаходяться між собою в визначених системних відносинах (наприклад, ряди вихідних і похідних омонімів або ряди основних і похідних омоформ), інші – без жодного зв'язку між собою. Таким чином, омоніміка в цілому не утворює єдиної системи. Інтенсивний розвиток омонімії в англійській мові очевидно спричинений більш ніж одним фактором. Насправді, існує кілька непов'язаних між собою причин, таких як односкладовий характер англійської мови і її аналітична структура.

Велика кількість омонімів також тісно пов'язана з такою характерною рисою англійської мови як фонетичною ідентичністю слова і основи, чи, іншими словами, предомінантністю вільних форм серед найчастотніших коренів. Факт, що частотність слів знаходиться в оберненій залежності від їх довжини, вказує на те, що односкладні слова будуть найчастотнішими. Крім того, так як найчастотніші слова полісемантичні, природним є те, що вони можуть набути значень, які суттєво відрізнятимуться від вихідних. Коли посередницькі відношення руйнуються, деякі з цих нових значень можуть втратити будь-які зв'язки з рештою структури і розпочати незалежне існування. Цей феномен відомий під назвою *дезінтеграція* або *розділення полісемії*.

Причини виникнення омонімії можна розділити на дві групи:

- омонімія, що викликана конвергентним розвитком звуку, коли два чи три слова різного походження випадково співпадають в звуковому складі;
- омонімія як розвинута полісемія, через дивергентний розвиток значення (включаючи конверсію).

Обидві причини можуть поєднуватися з втратами закінчень і іншими морфологічними процесами. Як приклад можна навести два латинських дієслова: *cadere* 'падати' і *sarere* 'тримати(ся)', в першому з яких, з плином часу 'd' змінилося на 'r', частково під впливом аналогічної форми з старо-французької мови. Такий тип омонімії неважко впізнати. Існування другого типу – питання до дискусії. В. І. Абаєв вважає омонімічними лише ті словоформи, які є різними з точки зору етимології. Всі інші – просто полісемічні. Багато інших вчених не згодні з В. І. Абаєвим і наполягають на необхідності семантичних і структурних критеріїв для розрізнення омонімії і полісемії.

На відміну від 'sarere' члени омонімічного ряду 'box', які утворилися внаслідок дезінтеграції, є етимологічно поєднаними. Ідентичність форми – не випадкова; вона базується на генетичних зв'язках. Всі слова цього ряду (5 за даними Оксфордського словника) походять одне від одного і простежуються до латинського 'bixus'.

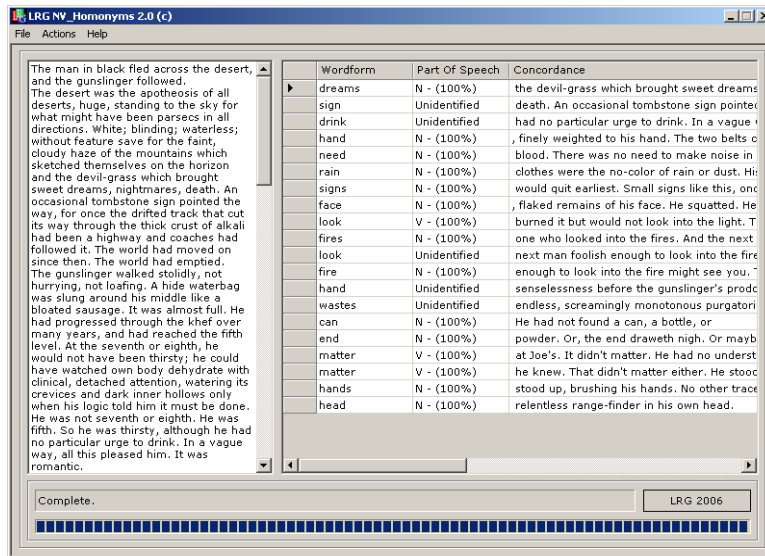
Такі омоніми можуть частково або повністю походити один від одного, але спільна точка їх зародження знаходиться за межами англійської мови. В цих словах, з появою нового значення, яке суттєво відрізняється від інших, семантична структура вихідного слова розділяється. Нове значення починає незалежне існування і отримує власну семантичну структуру. Звідси й термін *дезінтеграція*.

Підбиваючи підсумок, омонімія типу «іменник-дієслово», яка викликана з одного боку процесом втрати закінчень і т.п. і з іншого – конверсією (див. Таб. 1 з аналізом походження та конверсії омонімічних форм), є однією з найвизначніших характеристик сучасної англійської мови.

У роботі проводилося дослідження омонімічного складу англо-українського навчального словника другого рівня. Омоніми досліджувалися як групи, виділені за типом омонімії. Предмет дослідження: 266 омонімічних пар типу «іменник-дієслово» англійської мови. На основі проведеного дослідження запропонувати спосіб вирішення омонімії типу «іменник-дієслово» на прикладі 266 омонімічних пар.

Для проведення дослідження з англо-українського навчального словника другого рівня вибрано всі омонімічні пари, трійки, четвірки та п'ятірки. Омоніми розділено на групи за типом і подано їх частоту. На основі найчастотнішої групи «іменник-дієслово» укладено механічну вибірку з 266-ти омонімічних пар. Засобами власноруч створеної програми на мові програмування С# запропоновано спосіб автоматичного зняття омонімії в її частинному варіанті «іменник-дієслово» для 266-и омонімічних пар. Омонімічні словоформи закладено в масив. Окрім 266 вихідних форм, в масиві присутні їх омонімічні похідні (наприклад: "advance", "advances" і т.п.);

Отримавши текст, програма розділяє його на словоформи і, з ціллю полегшення доступу до нього, записує в базу даних. В базі даних кожне слово отримує своє унікальне місце і номер. З текстом в такому вигляді, наступною операцією є порівняння словоформ введеного тексту з омонімічними формами в масиві. При відповідному збігові починає роботу модуль "Determine", в якому зібрані диференційні, діагностичні ознаки та властивості відповідних частин мови, за якими однозначно (або з високою ймовірністю правильною відповіді) можна зняти омонімію. За допомогою низки простих правил в алгоритм "Determine" включена також система визначення вірогідності правильності результату. На виході користувач отримує впорядковану в вигляді списку інформацію про омонімічні форми, які зустрілись в введеному тексті і визначену частину мови для кожної з них з відповідною вірогідністю.



Модуль “Determine” розділений на 2 частини:

набір правил для іменника (NDefining)

- Якщо слово подане з великої літери і стоїть не на початку речення;
- Якщо слово закінчується на “s”, “s”’.
- Якщо попереднє слово – неомонімічний прикметник (вибірка з 760 найчастотніших прикметників з англо-українського навчального словника 2-ого рівня);
- Якщо попереднє слово – артикль;
- Якщо попереднє слово – присвійний займенник;
- Якщо попереднє слово – числівник (кількісний або порядковий);
- Якщо попереднє та наступне слова в різних варіаціях – займенники, прислівники, прийменники...

набір правил для дієслова (VDefining)

- Якщо попереднє слово – особовий займенник;
- Якщо попереднє слово – допоміжне дієслово;
- Якщо попереднє слово – прислівник, а перед ним – особовий займенник.
- Різноманітні комбінації часток, прийменників і т. д.

В кожному наборі правила також поділяються на *графічні (однозначні)*, які визначають частину мови безпосередньо за графічними показниками або іншими, які можуть вказати на частину мови без контекстного аналізу, та *синтаксичні (контекстні)*, які визначають частину мови за контекстом, сполучуваністю і т. д.

Згадані вище ознаки майже стовідсотково визначають частину мови, в той час як інші надають лише часткову вірогідність і спрацьовують лише в взаємному поєднанні. Кількість комбінацій суттєво велика. Ця робота обмежена розміром вибірки і масштабами дослідження. І все ж запропонований спосіб зняття омонімії типу «іменник-дієслово» спрацьовує для даного масиву омонімічних форм з високою вірогідністю правильного результату.

*А. Прыстаўка**

Беларускі дзяржаўны ўніверсітэт (Мінск)
УДК 81'322

ТЫПАЛОГІЯ СЕМАНТЫЧНЫХ ПЕРАЎТВАРЭННЯЎ ТЭКСТАЎ

Зварот да вызначэння тыпаў семантычных пераўтварэнняў тэкстаў абумоўлены важнасцю гэтага пытання для патрэб камп'ютэрнай лінгвістыкі. Даследаванні асаблівасцей будовы чалавечага мыслення можа растлумачыць далейшыя кірункі працы на ўдасканаленні перакладных праграм, праграм рэферыравання тэкстаў і г. д.

Прац, прысвечаных даследаванню семантыкі, дастаткова шмат і ў лінгвістыцы, і ў логіцы, але, на нашу думку, яны носяць больш філасофскі, чым практычны характар.

Мы разглядаем пераўтварэнні семантыкі, грунтуючыся на законах мовы як сістэмы, даказаных Агульнай тэорыяй сістэм (АТС) Ю. А. Урманцава і пацверджаных на розных узроўнях у межах сістэмнай лінгвістыкі У. А. Карпавым [гл. 1; 2].

Разгледзім з'яву тэкставых рэмінісцэнцый (ТР) у газетных загаловах, пабудаваных на падставе прэцэдэнтных тэкстаў (ПТ). Прэцэдэнтны тэкст – гэта інварыянт сітуацыі, які ў выглядзе закончанага выказвання захоўваецца ў свядомасці носьбіта мовы ў гатовым выглядзе. Самымі распаўсюджанымі прыкладамі прэцэдэнтных тэкстаў з'яўляюцца прыказкі, прымаўкі, выслоўі, крылатыя выразы, афарызмы і г. д. Загаловкі публіцыстычных артыкулаў, створаныя на падставе прэцэдэнтных тэкстаў, прыцягваюць увагу, выклікаюць адпаведныя асацыяцыі, правакуюць цікавасць да артыкула, а таму даволі папулярны ў сучасным перыядычным друку. Разгледзім газетныя загаловкі сучаснай беларускамоўнай прэсы, пабудаваныя на аснове прэцэдэнтных тэкстаў літаратурнага ці фразеалагічнага характару (“Ці загляне сонца ў наша аконца”, “А хто там стаіць?”, “Зямля па-надзейнымі крыламі” і г. д.).

Змены ці пераўтварэнні, якія адбыліся з прэцэдэнтнымі тэкстамі ў загаловах, падпарадкаваны сістэмным законам. Згодна АТС “аб'ект-сістэма можа або заставацца ў

* © А. Прыстаўка, 2006