

Удосконалення програми можливо через урахування всіх ДО: 1) варіантів форм допоміжного дієслова ('ave, av, a, 've, hev, haf; 'as, haith); 2) варіантів частки not (ain't, hain't ent); 3) способів утворення перфектного часу за допомогою have (has) got.

Таким чином, програма створювалася на основі статистичних розрахунків про вживання того чи іншого варіанта форми і може стати у пригоді тому, хто має на меті вивчення англійської мови.

Література

1. Дворжецкая М. П., Перебийнос В. И. Методические рекомендации по изучению системных и функциональных характеристик словоизменительных форм английского глагола (Perfect Tenses Active). Выпуск 4. — Львов, 1995. — С. 13–28.
2. Міліх Н. Г., Перебийніс В. С., Рукіна Е. П. Морфологія англійського дієслова. — К.: Либідь, 1995. — С. 13.

Я. Чаварга*

(Ужгород)

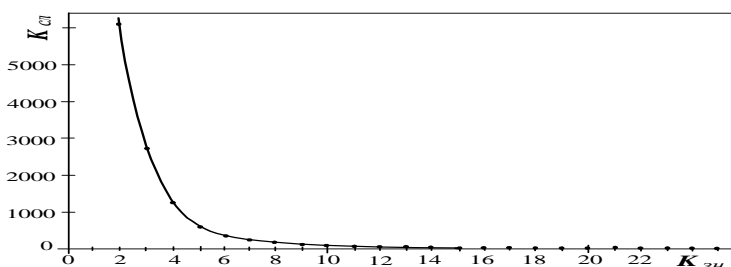
УДК 81'322.344

СЕМАНТИЧНИЙ ПОТЕНЦІАЛ ПРИКМЕТНИКІВ АНГЛІЙСЬКОЇ МОВИ

The article deals with the exponential dependence of the quantity of polysemantic adjectives in Modern English on the quantity of their meanings. The concrete form of this exponent is found and its linguistic interpretation is given. Three semantic potentials of English polysemantic adjectives are distinguished on the basis of our investigations: the usual, the increased and the high potential of creating new meanings.

Загальновідомо, що семантичний обсяг слова та кількість слів із певним числом значень, зафіксованих у словнику статистично пов'язані: найбільшу частку в ньому складають однозначні слова, далі йдуть (у порядку спадання частот) слова з двома, трьома і т. ін. значеннями. Такий розподіл є універсальною квантитативно-системною характеристикою полісемії природних мов. Вищезгадану статистичну залежність підтверджують і результати нашого дослідження, графічне зображення яких показано на рис. 1:

Рисунок 1. Залежність кількості прикметників від кількості значень.



За зовнішнім виглядом крива на рис. 1 нагадує відомий у математиці тип кривих, які називаються експонентами й описуються формулою:

$$y = a \cdot e^{-bx} \quad (1)$$

де для нашого випадку y – кількість слів ($K_{сл}$ на графіку),
 x – ступінь полісемічності ($K_{зн}$ на графіку),
 a – константа, що визначає масштаб по осі y , точку перетину кривої з віссю y ,
 b – константа, що задає крутизну експоненти,
 e – основа натурального логарифму ($e \approx 2,718$).

* © Я. Чаварга, 2006

Мета нашого дослідження – з'ясувати, наскільки строго залежність кількості слів від кількості значень узгоджується з експоненціальною залежністю, і знайти конкретний вигляд цієї експоненти та дати її лінгвістичну інтерпретацію.

Відзначимо, що експоненціальна залежність інтенсивності реакції (відчуття) від величини сили подразнення, в тому числі й у психофізичних експериментах, добре відома в науці, що вивчає поведінку живих організмів, – у біофізиці. Це так званий закон Вебера – Фехнера, який стверджує, що інтенсивність відчуття пропорційна натуральному логарифму сили подразнення [1:439]. У нашому випадку роль подразника відіграє поява нового предмета, явища чи властивості і, відповідно, потреба в номінації певного предмета чи явища засобами мови, а роль реакції – саме найменування (ім'я). Подібність нашої кривої до експоненти, а також низький рівень розкиду експериментальних точок вказує на те, що процес когнітивної діяльності людини в галузі творення нових значень слів на основі вже наявних не є хаотичним, безсистемним, а підпорядковується строгим закономірностям, характерним для живих організмів узагалі, і діяльності людини в напрямку творення полісемії зокрема.

З'ясуванню питання аналітичного виду залежності кількості слів від кількості значень (встановленню математичної формули) в літературі, в тому числі і в лінгвістичній, приділяють значну увагу, оскільки правильно визначена математична формула може сприяти з'ясуванню механізму творення полісемії, а відтак і впливу на неї когнітивної діяльності, що її визначає, оскільки при цій діяльності вона, власне, і твориться. Ф. Папп, 1967, [2] у своїх дослідженнях полісемії в угорській мові на базі тлумачного словника спробував представити результати досліджень як

$$y = V 2^{-x} \quad (2)$$

де y – частка у словнику слів, що мають x значень,
 V – обсяг словника.

Ю. А. Тулдава, досліджуючи полісемію в естонській мові [4], запропонував формулу, яка краще, ніж формула Ф. Паппа, узгоджується з експериментальними даними:

$$p(m) = a \cdot e^{-b \cdot m^c} \quad (3)$$

де $p(m)$ – частка слів із даною кількістю значень,
 m – кількість значень слова,
 a, b, c – параметри, що залежать від того, яка мова, яка частина мови та який словник чи текстовий масив досліджується.

У дослідженнях Ю. А. Тулдави $c \approx 0,5$, тобто формула має вигляд

$$p(m) = a \cdot e^{-b \sqrt{m}} \quad (4)$$

У всіх згаданих випадках робилася спроба представити результати експерименту через одну експоненту. Однак, якщо експериментальна або теоретична крива є результатом складання декількох простих експонент, то встановити це за зовнішнім виглядом графіка типу того, що зображений на рис. 1, важко, а в більшості випадків просто неможливо, оскільки при накладанні декількох експонент одна ділянка кривої плавно переходить у наступну. При цьому створюється враження наявності однієї кривої. Отже, при аналізі графіка на рис. 1 нам треба відповісти на запитання: чи є насправді дана експериментальна крива експонентою, чи тільки подібна на неї за формою (наприклад, гіпербола), або чи не є ця крива сумою декількох простих експонент типу

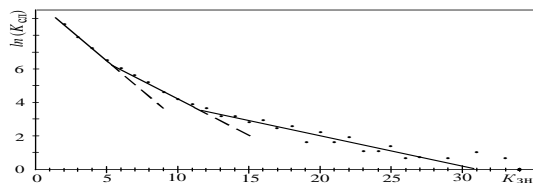
$$y = a \cdot e^{-b \cdot m}.$$

Для того, щоб відповісти на це запитання, криву треба прологарифмувати. У разі, якщо отриманий графік буде мати вигляд прямої лінії, це означатиме, по-перше, що вихідна (експериментальна) крива насправді є експонентою, і, по-друге, що процес творення багатозначності ймовірно є однаковим для всього масиву прикметників.

Якщо ж крива є комбінацією двох або більше експонент типу $y = a \cdot e^{-b \cdot x}$, то це означатиме, що творення полісемії відбувається за одним і тим самим законом, але матеріал (масив прикметників) по-різному робить свій вплив на дію закону. З курсу математики відомо, що в цьому випадку при логарифмуванні замість однієї прямої лінії отримаємо ламану лінію, при цьому прямолінійні ділянки ламаної відповідають тим

значенням аргументу, при яких дана експонента робить домінуючий внесок у загальну експоненту. Отже, щоб визначити, скільки різних експонент репрезентує крива на рисунку 1, представимо її у логарифмічній шкалі. На рис. 2 показано графік залежності кількості слів від кількості значень у логарифмічній шкалі.

Рисунок 2. Залежність кількості прикметників від кількості значень у логарифмічній шкалі.



Як бачимо, логарифмічний графік досить впевнено можна інтерпретувати як комбінацію трьох прямих ліній, а це свідчить про те, що на кожній із ділянок домінує одна з трьох експонент, тобто в кожному випадку утворення нових значень слова відбувається за законом Вебера – Фехнера. Як видно з графіка, вплив першої експоненти є домінуючим до кількості значень 5. Від кількості значень 5 і до 12 переважає друга експонента, і, нарешті, від 12 й вище домінує третя експонента.

Графік на рисунку 2 не тільки вказує нам на наявність трьох різних причин, що впливають на форму кривої, але й уможливує встановлення загального аналітичного виду залежності кількості слів від кількості значень. Оскільки на логарифмічному графіку наявні три лінійні ділянки, це дає нам підстави зробити припущення про те, що вираз, яким можна аналітично представити експериментальну криву, повинен мати вигляд:

$$y = y_0 + y_1 e^{-b_1 m} + a_2 e^{-b_2 m} + a_3 e^{-b_3 m} \quad (5)$$

За допомогою програми ORIGIN 7 ми отримали формулу для інтерпретації залежності кількості прикметників від кількості значень в англійській мові:

$$y = 35499 \cdot e^{-1,007 m} + 3049 \cdot e^{-0,395 m} + 36 \cdot e^{-0,099 m} \quad (6)$$

Результати розрахунків за цією формулою разом з експериментальними даними представлені в таблиці:

Таблиця. Розрахунки внеску кожної експоненти формули (6) в загальну криву і порівняння з експериментальними даними

К-сть значень	exp1	exp2	exp3	exp1+ exp2+ exp3=	екс- пери- мент	теорія/ експ.
2	4737	1384	30	6151	6124	1,004
3	1731	932	27	2690	2740	0,982
4	632	628	24	1285	1275	1,008
5	231	423	22	677	638	1,061
6	84	285	20	390	389	1,003
7	31	192	18	242	261	0,927
8	11	129	16	158	174	0,908
9	4	87	15	107	103	1,039
10	2	59	13	75	68	1,103
11	1	40	12	53	47	1,127
12	-	27	11	39	42	0,929
13	-	18	10	29	24	1,208
14	-	12	9	22	23	0,957
15	-	8	8	17	12	0,895
16	-	5	7	14	19	0,737

К-сть значень	exp1	exp2	exp3	exp1+exp2+exp3=	екс-перимент	теорія/експ.
17		4	7	11	12	0,917
18		2	6	10	13	0,769
19		2	6	7	5	1,400
20		1	5	6	9	0,667
21		1	5	5	5	1,000
22		-	4	5	7	0,667
23			4	5	3	1,667
24			3	4	3	1,333
25			3	4	4	1,000
26			3	4	2	2,000
27			3	3	2	1,500
28			2	3	0	-----
29			2	3	2	1,500
30			2	2	0	-----
31			2	2	3	0,667
32			2	2	0	-----
33			1	2	2	1,000
34			1	2	1	2,000
35			1	2	1	2,000
36			1	1	1	1,000

Незважаючи на те, що формула (6) має вигляд складного виразу, насправді це простіша залежність порівняно з формулою Ю. А. Тулдави (4), оскільки являє собою суму простих експонент, у той час як у формулі Ю. А. Тулдави наявна нелінійна залежність показника степеня від кількості значень, тобто \sqrt{m} , лінгвістичну інтерпретацію якій знайти важко, якщо взагалі можливо.

У випадку, коли математична залежність між двома величинами має вигляд декількох доданків, як у (6), використання математики дозволяє встановити внесок кожної складової в загальну криву – для формули (6) можна встановити величину внеску кожної з трьох експонент у сумарну експоненту. На рисунку 3 графічно представлено результати таких обчислень. Так, графік 1 побудований як:

$$y_1 = \frac{a_1 e^{-b_1 m}}{a_1 e^{-b_1 m} + a_2 e^{-b_2 m} + a_3 e^{-b_3 m}} \quad (7)$$

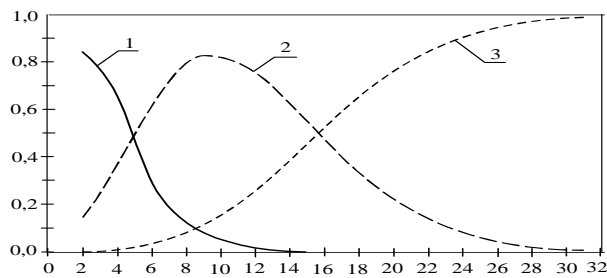
графік 2 як:

$$y_2 = \frac{a_2 e^{-b_2 m}}{a_1 e^{-b_1 m} + a_2 e^{-b_2 m} + a_3 e^{-b_3 m}} \quad (8)$$

а графік 3 як:

$$y_3 = \frac{a_3 e^{-b_3 m}}{a_1 e^{-b_1 m} + a_2 e^{-b_2 m} + a_3 e^{-b_3 m}} \quad (9)$$

Рисунком 3. Графіки внеску окремих експонент формули (5) у загальну криву, програма ORIGIN 7: 1 – внесок першої експоненти; 2 – внесок другої експоненти; 3 – внесок третьої експоненти.



Той факт, що наша експериментальна крива залежності кількості прикметників від ступеня полісемії виявилася сумою трьох експонент, свідчить на користь гіпотези про те, що на утворення нових значень, крім інших факторів, впливає семантичний потенціал слова. Як правило, слова при утворенні є однозначними, але в ході подальшого вживання вони можуть послідовно реалізувати свій асоціативно-семантичний потенціал у вигляді нарощування числа своїх значень. На матеріалі російської мови при аналізі даних словника „Новые слова и значения” [3] висуваються припущення про вичерпання потенціалу семантичного розвитку слів, сповільнення темпу утворення нових значень слів при просуванні до вищих семантичних зон аж до повної зупинки на певному етапі. Ми припускаємо, що наявність трьох експонент у згаданому законі пов'язана з тим, що слова не просто мають внутрішній потенціал творення багатозначності, але ще й діляться на декілька типів за величиною цього потенціалу, за різними закладеними в них потенційними можливостями творення полісемії.

Аналізуючи графіки на рисунку 3, бачимо, що перша крива описує внесок у полісемію прикметників англійської мови з порівняно невисоким (назвемо його звичайним) потенціалом творення багатозначності. Друга крива показує прикметники з дещо вищим потенціалом творення полісемії (назвемо його підвищеним потенціалом), які складають порівняно менш частотну групу. Третю групу прикметників представляють слова з високим семантичним потенціалом, (див. третю криву).

Ми скористаємося отриманими графічними результатами для того, щоб дати відповідь на запитання, що насправді являють собою ці потенціали, і в чому полягає їхня лінгвістична суть. Згідно з графіком 1 на рисунку 3, для з'ясування лінгвістичної суті першого потенціалу потрібно проаналізувати слова з полісемією 2–5. Переважна більшість цих слів повинні мати спільну ознаку, яка і складе характеристику для першого семантичного потенціалу прикметників. Це основна кількість досліджуваних нами слів (10777 із 12019, або 89, 67 %). Визначення когнітивних моделей цих слів через вивчення та зіставлення словникових дефініцій показали, що більшість із них, які входять до діапазону 2–5 значень, тлумачаться за таким словниковим шаблоном: 1) of or pertaining to ...; 2) related (relating) to ...; 3) caused by ...; 4) having (composed of, made of) ...; 5) being in ... тощо. Водночас наші спостереження показали, що основний масив прикметників цього діапазону складають багатоморфемні (похідні та складні) слова – 10116 із 10777, ~94%. Простих прикметників у цьому діапазоні відносно небагато – усього 661, ~6%, що за графіком 3 на рис. 3 майже відповідає внеску слів із високим потенціалом. Отже, кількість багатоморфемних слів приблизно у 15, 3 раз більша, ніж кількість одноморфемних. Цей факт дає підставу для висновку про те, що афіксально марковані прикметники мають звичайний семантичний потенціал. У словах із першим (звичайним) потенціалом чітко проявляється вплив словотвірного фактора на полісемію.

Другий семантичний потенціал досягає максимуму (~82% внеску) при кількості значень 9–12. Отже, відповідно до графіка 2 на рис. 3, суть лінгвістичного змісту другого (підвищеного) потенціалу полісемії треба шукати, аналізуючи слова з кількістю значень 9–12. Наші дослідження показали, що переважна більшість цих адективних одиниць є абстрактними, тобто такими, що позначають абстрактні ознаки. Наприклад: *Christian, classical, drawn, exclusive, imperfect, legal, personal, primitive, privy, progressive, sovereign, spiritual*. Відношення між простими й афіксально маркованими прикметниками змінюється: перепад між їхньою кількістю суттєво переміщується у бік зростання кількості простих слів. Кількість багатоморфемних одиниць перевищує кількість одноморфемних у 4, 65 разів проти 15, 3 разів у першій групі, що добре узгоджується з ходом графіків на рис. 3.

Принагідно зауважимо, що в англійській мові прикметників із підвищеним потенціалом творення багатозначності приблизно в десять разів менше, ніж прикметників зі звичайним потенціалом. Це дає вагому підставу для висновку про те, що однією з головних причин появи у прикметників більш високого (підвищеного) потенціалу творення полісемії є узагальнення поняття до абстрактного рівня.

Графік 3 на рисунку 3 вказує на те, що лінгвістичну природу високого семантичного потенціалу потрібно шукати у списку слів у діапазоні від 16 значень і більше. Ця група прикметників є порівняно нечисленною. У досліджуваному нами матеріалі нараховується 99 слів із багатозначністю 16 і вище, що складає всього 0, 82 % від загальної кількості прикметників. Ці 99 слів дають 2154 значення, або 5, 4 % від сумарної кількості значень. Ми проаналізували слова з кількістю значень 20 і більше як такі, що є

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 означуванням 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