

по радіо, чи з телевізора фразу «На саміті було досягнуто консенсусу щодо надання Україні чергового траншу програми стенд-бай». Літом у трамваї знайомий кандидат наук виразився так: «Іду до офісу, залишу там кейс, а звідти – на фазенду». Коментувати – зайве. Не дай Бог, щоби прижилася в Україні така мова...

І ще одне питання, яке при складанні Словників треба би було постійно мати на увазі. Це – хибні поняття. Ще в 1954 р. викладач з Москви М.Калінушкін, працюючи в Китаї, видав книжку, де (тепер важко здогадатись, чи це був недогляд, чи автор мав якість інші думки), помилково об'єднав різні за призначенням і дією типи відцентрових машин одним поняттям – «нагнетатели» (по-українському – «нагнітачі»). Нам вперше вдалось його «виловити» і не вписати у словник. Але, чи не пропустили ми інші «шедеври»?

Весь час, складаючи словники, нас гризли сумніви, але «за спиною» ми відчували колег – доброзичливців, які не один раз на підправляли, підкидали свіжі думки. Спасибі їм за це.

Хоч сумніви так і залишились. Ось два приклади, найближчі до нашої тематики: труба – рура, хемія – хімія.

Більшість наукових і технічних термінів, якщо дивитися глибоко, мають, скоріше всього, греко-латинське походження. Одні з них приходили до нас через англійську, інші – німецьку, а ще інші – французьку та інші мови. Питання тільки, як глибоко вони вкорінилися, на яких територіях розповсюдились, з яких можна легко будувати похідні терміни тощо. Якщо, наприклад, на основі слів труба і рура вдається утворити окремі рівнозначні ряди (труба – трубопровід, трубка і т. п., рура – рурогін, рурка і т. п.), то з хемією і хімією цього не можна зробити. Слово «хемія» вживалось в українській мові до 50-х років; «хемія» – пишеться в більшості країн світу, від «хемія» побудовані такі терміни як «хемосорбція», «хемотропізм», «хемолімінесценція» тощо. На жаль, така побудова терміна на базі «хімія» – неможлива. І хоча слово «хімія» частіше вживається і навіть силоміць введене в сьогоденний нормативний словник, ми перевагу віддали терміну «хемія», поставивши на другому місці його синонім «хімія».

У даний час місячником "Ринок Інсталяційний" редагуються і видаються два словники: інсталяційних термінів та з механіки рідин і суміжних наук.

ПРО АВТОМАТИЗАЦІЮ ПОШУКУ ТЕРМІНІВ У НАУКОВО-ТЕХНІЧНИХ ТЕКСТАХ

© Надія Міщенко, Олександр Міщенко*, 2000

Інститут кібернетики НАН України, Київ

* Портлендський державний університет, США

Проблему створення термінологічних словників різного виду (тлумачних, дво- чи багатомовних) можна розглядати у трьох аспектах:

- збір термінів, вжитих у вітчизняній та іноземній науковій літературі, їх тлумачення та переклад іншомовних термінів українською мовою;
- розробка архітектури електронних версій таких словників для певного фаху та її наповнення;
- автоматизація розробки паперових версій на основі електронних.

У науковій літературі переважно висвітлюються два останні аспекти, проте сьогодні перший аспект також є актуальним у зв'язку з розповсюдженням мережами Інтернету науково-технічної текстової інформації різними мовами, що забезпечує оперативне введення в обіг нових наукових досягнень та відповідних термінів у наукових та освітніх закладах України.

Цілком очевидно, що пошуком, тлумаченням та перекладом наукових термінів українською мовою найефективніше займатися фахівцям відповідних областей, які найчастіше знають іноземну мову наукових текстів свого фаху, зокрема англійську. У даній роботі розглядається автоматизація пошуку термінів у таких областях інформатики як програмування та проектування цифрових пристроїв, які є областями наукової та практичної діяльності авторів.

Загальноприйнятою є думка про те, що англійська науково-технічна література є чи не єдиним джерелом нових термінів у інформатиці. Однак цей імператив може зашкодити вітчизняному термінознавству, адже у науковому світі українські вчені користуються заслуженим авторитетом, і більшість тих, хто їде за кордон продовжувати наукові дослідження, користуються вже набутою і творять нову наукову термінологію.

Варто пригадати також, що перша у світі Енциклопедія кібернетики [1] була створена в Україні і вийшла друком 1973 року українською мовою, а 1974 — російською завдяки подвижницькій праці академіка В.М.Глушкова, засновника Інституту кібернетики в Україні. Цей Інститут нині реорганізований у Кібернетичний центр імені В.М.Глушкова НАН України. У створенні Енциклопедії кібернетики брали участь провідні вчені України та інших республік тодішнього СРСР.

Перш ніж перейти до викладу основної теми доповіді, зауважимо, що надалі поняттям "термін" може бути позначена не лише окрема лексична одиниця, але й усталене словосполучення, складниками якого можуть бути загальноживані слова чи добре відомі терміни. Однак вважаємо, що ядром словосполучення неодмінно є іменник. Приклад терміну — усталеного словосполучення з відомих слів: binary decision diagram (BDD), у перекладі українською — діаграма бінарного вибору (ДБВ). Через високу частоту вживання цього терміну він замінений аббревіатурою, що взагалі є властивим для термінів, виражених усталеними словосполученнями.

Надалі новими термінами будемо називати терміни, відсутні в наявних україномовних термінологічних словниках. Такі терміни можуть бути створені фахівцем, або знайдені ним у процесі реферування фахової літератури. Не виключається цілеспрямований пошук термінів на основі обробки текстів свого фаху. Обробка фахових текстів включає пошук ключових слів текстів, визначення частоти їх вживання, пошук синонімів, наявність омонімії тощо. Вивчається контекст вживання термінів іноземною мовою з метою якомога точнішого їх перекладу.

Система FEST. Для автоматизації пошуку термінів у науково-технічній літературі, розміщеній на електронних носіях, пропонується мобільна програмна система послівної обробки текстів FEST [2]. Мобільність системи забезпечується можливістю приєднання до неї за допомогою вхідного параметра словників лексем різних мов для обробки відповідних вхідних текстів. Конкретно система може працювати над текстами українською, російською або англійською мовами. Можлива також обробка змішаних текстів (із цитуванням у межах тексту іншими мовами).

У системі передбачається, що лексика науково-технічних текстів певною мовою поділяється на кілька модулів:

- модуль загальноживаної лексики Мз, до якої належать також усталені словосполучення (складні прийменники та сполучники, вставні словосполучення);
- модуль галузевої лексики Мг, спільної для текстів, що належать до кількох фахових областей даної галузі;
- модуль лексики конкретного фаху Мф.

Ми не зачіпаємо семантичних проблем розподілу лексики на модулі, а використовуємо прагматичний підхід: оскільки фахові модулі лексики формуються на основі науково-технічних текстів, лексема, що зустрічається в текстах певного фаху частіше, ніж інші

лексеми (за винятком службових), здебільшого є терміном даного фаху. Термін, зміст якого є спільним для кількох фахових областей, може належати словнику галузевої лексики. До якого словника віднести той чи інший термін, остаточно вирішує фахівець.

Модулі формуються людиною (розробниками системи чи фахівцем) у вигляді текстових описів (специфікації) лексики спеціальною формальною мовою Dual [2], яка містить засоби подання всіх даних, необхідних для морфологічного аналізу словоформ та перевірки узгодження їх із сусідніми словоформами в текстах. Кожний з модулів може бути представлений в комп'ютері кількома файлами. Для роботи з літературою конкретною мовою автоматично (системою FEST) генеруються кілька машинних словників цієї мови на основі різних конфігурацій відповідних текстових модулів: повний словник С на основі всіх модулів, окремі словники Сз, Сг, Сф (відповідно за модулями Мз, Мг, Мф), Сзг (за модулями Мз та Мг), Ст (за модулями Мг та Мф).

Машинний словник Сз створюється заздалегідь, а решту можна генерувати безпосередньо перед використанням, оскільки відповідні модулі постійно поповнюються в міру обробки текстів. Оскільки мова специфікацій має засоби для вказівки фаху, в якому вживається лексема, то відповідні словники являють собою тезауруси відповідного фаху.

Генерація галузевих та фахових словників за текстами певної галузі чи фаху, по суті, створює персональну систему обробки фахових текстів [3], яка характеризується порівняно невеликим обсягом словників і майже повною відсутністю омонімії лексем. У разі потреби галузевий словник може бути поповнений лексикою вже існуючих словників.

Розмаїття словників лексики різних мов використовується в системі FEST для виконання різних завдань, одним із яких є пошук термінів у науково-технічних текстах, що здійснюється на основі статистичної обробки текстів із використанням машинних словників, побудованих за різними конфігураціями модулів. У процесі створення словників та пошуку термінів у текстах система FEST виконує лише рутинну частину роботи (перевірку правильності специфікацій лексики, генерацію словників за специфікаціями, морфологічний аналіз текстів, у процесі якого перевіряється узгодження іменників із сусідніми словоформами та складання списків відомих чи невідомих словоформ, а по завершенню аналізу тексту — формування частотних списків слів та словосполучень). Творча ж робота (складання специфікацій, аналіз частотних списків, відбір нових термінів для специфікації, їх тлумачення чи переклад) виконується фахівцем.

Зазначимо також, що запит до системи щодо виконання тієї чи іншої функції подається їй у вигляді спеціального параметра, за допомогою якого можна замовляти різні типи частотних списків:

- списки невідомих (відсутніх у словнику) словоформ або фрагментів тексту, що складаються з таких словоформ;
- списки відомих лексем або словосполучень із них.

Функції системи FEST. Розглянемо функції системи залежно від типу словника.

За відсутності словника (тобто, з порожнім словником) систему FEST доцільно використовувати лише для побудови частотного списку невідомих лексем (словоформ), який є джерелом для першочергового розширення словника. Зауважимо, що кожний сеанс роботи системи над новим текстом може виявити слова, відсутні в словнику, які, проте, можуть не заноситися до нього, якщо цього не вимагає методика виконання системою поставленого завдання.

Використовуючи словник Сз, знову-таки можна одержати частотний список невідомих слів, у якому, однак, серед найуживаніших слів можуть траплятися терміни. Фахівець визначає, які саме слова можна вважати термінами, описує їх мовою Dual, що дозволяє автоматично ввести їх до словника Ст.

Використовуючи розширений словник С (Сз, доповнений новими термінами Ст), можна одержати частотний список словосполучень із відомих (наявних в словнику С) слів, до яких входять знайдені терміни. Це робиться з метою пошуку стійких словосполучень за участю нових термінів. Якщо словосполучення стійке, тобто зустрічається часто, то з певною ймовірністю його можна вважати новим терміном. Додатково аналізуються всі входження цього словосполучення до тексту, які фіксуються в списку словосполучень, після чого приймається остаточне рішення про його статус.

Введення нового терміну в словник Ст дозволяє простежити його вживання в інших текстах, що належать до одного або кількох фахових напрямків даної галузі. Щоб зробити цей аналіз цілеспрямованішим і ефективнішим, для подальшого аналізу вживання терміна в текстах можна використати лише словник Ст. У цьому випадку можна одержати частотний список слів або словосполучень, складений лише з термінів. Усі інші словоформи будуть пропущені, як невідомі.

Розробляється алгоритм автоматичної побудови опису лексем, частота входження яких до тексту з різними закінченнями дозволяє ототожнити ці закінчення з однією з заданих у системі послідовностей закінчень і тим самим визначити парадигму лексеми.

Приклад використання системи FEST. Наведемо результат аналізу термінів у статті одного з авторів "Implicit Algorithms for Multi-Valued Input Support Minimization" (у перекладі українською мовою — "Неявні алгоритми мінімізації числа вхідних змінних у неповністю визначених багатозначних функціях"). Стаття знаходиться у файлі bemdd.txt обсягом біля 34Кб.

Для аналізу тексту використано словник Сз загальноновживаних слів англійської мови, сформований за науковими текстами. В результаті аналізу одержано частотний список невідомих слів, переважна більшість яких є термінами. Наводимо початковий фрагмент цього списку, переклад яких без контексту не є складним (за винятком абревіатур та терміну cofactor).

Таблиця 1

Файл bemdd.txt, enmorf.tbl, eng.spc

НОВІ СЛОВА

№ з/п	Частота вживання	Відсоток	Рядок	Основа - закінчення
1.	420	8.291	23	the
2.	147	2.902	24	variable (-s, -0)
3.	71	1.402	1	support (-0, -s)
4.	55	1.086	12	function (-s, -0)
5.	43	0.849	27	set (-0, -s)
6.	42	0.829	1	input (-0, -s)
7.	39	0.770	18	value (-s, -0)
8.	35	0.691	16	BEMDD (-s, -0)
9.	35	0.691	90	MISF (-s, -0)
10.	32	0.632	12	multi-valued
11.	30	0.592	18	output (-0, -s)
12.	24	0.474	1	minimization
13.	24	0.474	254	inessertial
14.	23	0.454	58	procedure (-s, -0)
15.	23	0.454	401	cofactor (-s, -0)
16.	22	0.434	71	binary
17.	20	0.395	12	decomposition

Всього слів у файлі - 5066, Нерозпізнаних - 2442 (48%)

На першому місці за частотою вживання традиційно виявляється артикль.

За одержаним списком слів складено їх специфікацію і згенеровано системою FEST словник термінів Ст. Після цього проведено аналіз тексту за словником Ст без словника Сз і сформовано частотний список словосполучень, з якого вибрані ті, що містять найуживаніші терміни з наведеного вище фрагмента. Нижче подаємо кілька таких словосполучень та їх переклад чи тлумачення українською мовою:

Binary-Encoded Multi-valued Decision Diagrams (BEMDDs) — Багатозначні Діаграми

Вибору, закодовані Двійковими Змінними (БДВДЗ);

BEMDD traversal routine — процедура обходу БДВДЗ;

Multi-valued-input multi-valued-output Incompletely Specified Functions (MISFs) —

Неповністю Визначені Багатозначні Функції (НВБФ);

decomposition of MISFs — декомпозиція НВБФ;

multi-valued input support minimization — мінімізація числа вхідних змінних у неповністю визначених багатозначних функціях;

implicit support minimization — мінімізація числа вхідних змінних, що базується на т.зв. неявних методах;

the support of MISF — множина вхідних змінних НВБФ;

input support variables — вхідні змінні деякої функції;

encode the support — кодувати множину вхідних змінних;

compatibility of cofactors — сумісність кофакторів (кофактор функції $F(x,y,z)$ відносно змінної x є функцією, яку одержуємо після заміни змінної x її значенням; наприклад, для $x=1$ кофактором даної функції буде функція $F(1,y,z)$);

cofactor domain — область визначення кофактора;

set of multi-valued benchmarks — набір тестів для визначення ефективності реалізації тих чи інших алгоритмів;

inessential variables — несуттєві змінні;

the benchmark functions — функції, що входять до набору тестів;

multi-valued decomposition — декомпозиція багатозначних функцій.

Ці терміни набувають значного поширення в англійській науковій літературі останніх років. Проводиться також аналіз літератури з програмування.

Насамкінець зробимо кілька зауважень, що стосуються проблем термінології не лише в кібернетичі.

Щодо калькування. Якщо термін вдало перекладений російською мовою, або в перекладі українською мовою з англійської надто громіздкий (зауважимо лаконічність англійської мови), то калькування відповідно з російської мови чи англійської може бути виправдане. Одним із небагатьох, на нашу думку, вдалих кальок з англійської є термін "софтвр", що вживається на позначення програмних продуктів. Невдале калькування термінів російською мовою неминуче з огляду на недостатню оперативність у перекладі та виданні наукової літератури українською мовою. Те ж саме стосується і фахового сленгу, який існує і в англійському середовищі, незважаючи на лаконічність англійської мови. Вихід із такої ситуації відомий — це оперативне створення нових словників або розширення існуючих за рахунок нових термінів українською мовою.

Українська наукова продукція з інформатики ще не має широкого представлення на електронних носіях, зокрема, в Інтернеті, що унеможливило автоматизований пошук наукових термінів в текстах українською мовою. Натомість одним із засобів для налагодження оперативного збору нових термінів було б унормування цього процесу, наприклад, шляхом запровадження обов'язкового супроводу статті чи монографії додатком,

де б наводилися вжиті автором терміни та їх коротке тлумачення. Крім того, щоб зацікавити фахівців цією роботою, доцільно вважати науковою статтею і дозволити до друку у відповідних наукових часописах набір відсутніх у словниках термінів, зібраних та пояснених автором на основі нових наукових праць із його фахової області. При цьому фахівцеві доцільно мати у своєму розпорядженні поточну версію електронного словника термінів відповідного фаху. Осередком збору такої інформації може бути відповідна термінологічна комісія при провідному з даного фаху науковому чи освітньому закладі.

1. *Енциклопедія кібернетики. В 2х т. - Київ: Наук. думка, 1973.* 2. Міщенко Н.М., Федюрко В.В., Щоголева Н.М. *Мобильная система статистической обработки профессиональных текстов в целях определения их тематики // Проблемы программирования. - 1999. - № 2. - С.103-110.* 3. Міщенко Н.М., Федюрко В.В., Щоголева Н.М. *Про персональні мовні системи для фахівців-нелінгвістів // Матеріали 8-ї Міжнар. конф. "KDS-99: Знання-Діалог-Рішення" (Київські, 13-18 вересня 1999р.). - Донецьк: журн. Искусственный интеллект. - 1999. - Вып. 2 - С.300-306.*

ПРОБЛЕМА ІНШОМОВНИХ ЗАПОЗИЧЕНЬ В СУЧАСНІЙ УКРАЇНСЬКІЙ ФІНАНСОВО-ЕКОНОМІЧНІЙ ТЕРМІНОЛОГІЇ (на матеріалі третього видання "Фінансового словника")

© Геннадій Вознюк, Анатолій Загородній, 2000
ДУ "Львівська політехніка"

Формування в Україні ринкових відносин, організація нових виробничих, комерційних і маркетингових структур, вихід підприємств на зовнішній ринок зумовлюють дедалі активніше оперування поряд із тими поняттями й термінами, що були властиві плановій економіці і частина з яких не втратила актуальності й у наші дні, новими категоріями і термінами, властивими економіці вільного підприємництва.

Сьогодні українська фінансово-економічна думка, дедалі природніше вписуючись у світові інтеграційні процеси, потрапляє й під англомовні термінологічні впливи, запозичуючи (інколи досить механічно й недоцільно) разом із відповідними поняттями та категоріями їхню словесну форму і входячи, як раніше в російськомовне, так тепер в англомовне поле залежності. Дехто з науковців, виробничників, економістів і фінансистів-практиків вважає, що, інтернаціоналізуючи таким чином нашу термінологію, ми збагачуємо її, полегшуючи собі шлях до економічного прогресу. На жаль, губимо через такий підхід часом більше, ніж віднаходимо.

Гадасмо, цей спрощений шлях некритичного й інколи навіть бездумного копіювання, який збіднює і власну термінологію, і власну економічну думку, і національну ментальність загалом, часто, навпаки, ускладнює розуміння і відтворення понять, окреслюючи їх механічно запозиченими термінами. Історія нашої мови переконливо свідчить, що українська мова, яка завжди активно контактувала і зі спорідненим слов'янським, і з греко-латинським, і з германо-романським, і з тюркським мовним матеріалом, творчо використовуючи його для своїх актуальних потреб, цілком спроможна гнучко й точно перекласти, у разі потреби скалькувати, по-своєму осмислити і власними фонетико-морфологічними засобами відтворити те чи інше запозичене поняття. Причому в жодному разі не йдеться про вузьконаціональне спрямування цього процесу, ані про архаїзацію сучасної української термінології. Звернення до джерел повинно живити термінознавчу думку, але не спотворювати чи засмічувати її. Отже, мовиться не про лінгвістичну "хірургію" (як іноді уявляє собі цей процес дехто навіть із академічних мовознавців), а про реабілітаційну "лінгвотерапію", якої, фігурально кажучи, потребує засмічений чужинними