

мовах, а з вищенаведених прикладів стає зрозумілим, що така, на перший погляд, лексикографічна робота, як укладання словника дуже тісно переплітається з основами та правилами перекладу.

Література

1. Зорько Г. Ф., Майзель Б. Н., Скворцова Н. А. Большой русско-итальянский словарь. 3-е издание стереотипное. – М.: Издательство Русский Язык, 1999. – 1018 с.
2. Ковалев В. Ф. Итальянско-русский и Русско-итальянский словарь. 2-е издание, стереотипное. – М.: Русский Язык Медиа, 2005. – X, 1077 с.
3. Корунець І. В. Теорія і практика перекладу (аспектний переклад). Підручник. – Вінниця: Нова Книга, 2003. – 448 с.
4. Перебийніс В. І. Лексикографічне забезпечення навчального процесу з іноземної мови // Вісник Київського лінгвістичного університету. Київ, 1999. – Серія Філологія, Том 2, №1. – 233 с.

О. Лазарева*
УДК 81'374.33

ПРО МОЖЛИВІСТЬ АВТОМАТИЧНОГО УКЛАДАННЯ ТЕРМІНОЛОГІЧНИХ СЛОВНИКІВ ЗА ФАХОВИМИ ТЕКСТАМИ

The paper suggests an approach to creation of the vocabulary of terminological dictionaries. Word combinations of definite grammar structures being automatically extracted from the texts of some domain and arranged in a frequency list may constitute the basis for a terminological dictionary. The obtained vocabulary will be excessive and should be edited by an expert.

Термінологічний вибух як одна з проблем інформаційного сьогодення викликає необхідність пошуку шляхів інтенсифікації та інтелектуалізації процесів укладання та оновлення словників різного призначення. Це й паперові фахові та тлумачні словники, й словники, що використовуються в автоматизованих інформаційних системах (класифікатори, рубрикатори, інформаційно-пошукові тезауруси). Традиційно такі словники склалися фахівцями з предметної галузі, можливо у співпраці з лінгвістами, на основі аналізу цілої колекції документів з заданої тематики. І цілком очевидно, що сам процес накопичення масиву лексичних одиниць був і лишається дуже трудомісткою операцією. Крім того, зважаючи на стрімкий розвиток практично всіх галузей науки й техніки, з одного боку, і доволі довгий процес підготовки чи модернізації словників, з іншого, такі словники завжди відстають за складом лексики від поточного стану галузі, тобто не відповідають вимогам повноти.

Всі ці проблеми склали підставу для пошуку методики автоматизації первинного накопичення бази термінологічних одиниць окремої предметної галузі. Основоположним принципом такої методики є припущення (і наша глибока впевненість), що необхідним і достатнім джерелом термінологічної лексики є фахові тексти, тобто достатньо великий масив таких текстів може практично повністю покривати лексичний і термінологічний склад певної тематичної підмови. Другим важливим принципом методики є припущення, що потенційно будь-яке іменне словосполучення може бути терміном або може містити термін. Отже, для вирішення поставленої задачі формування лексичного складу термінологічного словника необхідно розробити методи, які б дозволяли автоматизувати процес виділення з текстів таких іменних словосполучень, структура яких відповідає структурі термінологічних одиниць. Сформований таким чином словник, безперечно, буде надмірним і буде потребувати постредагування. Однак, трудовитрати, пов'язані з постредагуванням, будуть несумірно меншими за трудомісткість формування такого словника традиційними методами. Точне визначення обсягу корпусу текстів, необхідних для складання словників, представляє окрему задачу і не розглядається в цій роботі. Однак приблизну оцінку можна зробити емпірично, а саме – в процесі створення словника спостерігати за швидкістю його зростання. Якщо надходження нових слів або словосполучень практично припиняється, тобто відсоток нових надходжень відносно, наприклад, загальної кількості слів у тексті складає менше деякого досить малого значення, то поповнення словника можна припинити.

Перш за все визначимось, які словосполучення ми будемо шукати в текстах. Дослідники термінології [4] виділяють досить багато синтаксичних структур термінів, які

* © О. Лазарева, 2006

включають, насамперед, найбільш розповсюджену модель «одиничний іменник», а також дво-, три- або багатокomпонентні словосполучення, які, в тому числі можуть містити прийменники, прислівники і дієслова. Однак найбільш продуктивними [2] є субстантивні словосполучення з прикметником або дієприкметником (ад'єктивний тип): *програмне забезпечення, багатозначна суфіксальна морфема*; субстантивні словосполучення з іменником (або їх послідовністю) в родовому відмінку (атрибутивний тип): *моделювання дискретних схем, неповне опрощення основи слова*. Оцінки, надані в роботі [3], показують, що терміни іншої структури, наприклад, з прийменниками або сполучниками складають менш одного відсотка від загального числа термінів. Часто дослідники, що працюють над аналогічними задачами, зокрема [1], вводять обмеження на довжину словосполучення, що буде включатися до словника – не більше 3 слів. З одного боку вони мають рацію – довжина більшості термінів не перевищує трьох. Але існують і набагато довші терміни, наприклад: *багатокomпонентна дрібнодисперсна суха суміш, область керування програмними засобами*, й ігнорувати їх вважаємо недоцільним.

Схематично типи найбільш продуктивних термінологічних конструкцій можна представити наступним чином:

$$S_1 := \langle \text{ад'єктив} \rangle \dots \langle \text{іменник} \rangle \text{ або } [A_i] \dots N;$$

$$S_2 := S_1 \{S_1\}^p \dots,$$

де квадратні дужки містять необов'язковий елемент, а фігурні – обов'язковий.

Цілком очевидно, що не всі словосполучення такої структури є термінами: це й слова загального вжитку, й вільні словосполучення. Але що є критерієм для включення чи не включення виділеної лексичної одиниці до словника? Це, перш за все, її галузева приналежність, і ціла низка інших критеріїв, що визначають властивості терміна як одиниці терміносистеми. А це вже екстралінгвістична інформація, формалізувати яку на сьогодні не представляється можливим. Тому ми ставимо задачу виділення всіх слів і словосполучень певної структури за формальними ознаками, а потім пропонуємо надати фахівцеві можливість редагування цього списку.

Якщо проаналізувати розташування термінів у тексті з урахуванням їхньої структури, то виявляється, що термін (зокрема, структури типу S_1 або S_2) може бути як самостійною синтаксичною одиницею речення, так і входити до складу словосполучення такої ж структури. Тому пропонується поділення виявлених у тексті початкових словосполучень на компоненти і формування з них усіх можливих комбінацій – **похідних дескрипторів**, синтаксична структура яких також відповідає структурі терміна.

Похідні дескриптори для конструкції S_1 формуються шляхом генерації усіх можливих комбінацій ад'єктивів з іменником. При цьому свідомо виключаються інвертовані ланцюжки типу $A_i + A_i N$. Це пояснюється тим, що порядок слідування ад'єктивів носить не випадковий характер, а зумовлений семантичними відношеннями актантів. Отже, наприклад, з ланцюжка $A_1 A_2 N$ буде сформоване 3 похідних дескриптори: $N; A_1 N; A_2 N$, які разом з початковим будуть занесені до словника.

Для атрибутивних словосполучень (тип S_2) процес генерації дескрипторів має дещо інший характер, виходячи з того, що відношення між компонентами цієї конструкції є лінійними, тобто напрямок синтаксичного зв'язку спрямований від попереднього компонента до наступного без стрибків: $S_1 \rightarrow S_1 \rightarrow \dots \rightarrow S_1$. Генерація полягає у формуванні всіх допустимих лінійних ланцюжків довжиною від 1 до n , де n – число ад'єктивних компонентів атрибутивної конструкції. Отже з початкового словосполучення цього типу, наприклад $A_1 N_1 + A_2 N_2 + A_3 N_3$, буде сформоване 5 похідних одиниць: $A_1 N_1; A_2 N_2; A_3 N_3; A_1 N_1 + A_2 N_2; A_2 N_2 + A_3 N_3$, і вони теж будуть занесені до словника. Крім того, за правилами обробки ад'єктивних словосполучень до словника додатково будуть включені елементи $N_1; N_2; N_3; N_1 + A_2 N_2; N_2 + A_3 N_3$.

Отримані в результаті описаних перетворень словосполучення додатково піддаються процедурі інвертування, тобто розташування елементів словосполучення таким чином, щоб головне слово, тобто перший іменник, опинилось на першій позиції. Це дає можливість при сортуванні кінцевого списку словникових одиниць за алфавітом отримати цікаву картину лексико-семантичних гнізд, об'єднаних загальним головним словом.

Крім того, для запобігання повторення одних і тих же слів і словосполучень, які в текстах зустрічаються у різних відмінково-числових формах, дескриптори потрібно нормалізувати, тобто привести до початкової форми головний іменник і узгоджені з ним ад'єктиви.

Важливим показником при відборі дескрипторів у тематичні термінологічні словники є показник частоти вживання даного слова чи словосполучення. Цілком очевидно, що ймовірність приналежності деякого дескриптора до складу термінів певної предметної галузі прямо пропорційна його частоті й тим вища, чим більшою є кількість його слів. Отже показник

$$K = (F-1)*Q,$$

де K – імовірнісний коефіцієнт, F – частота дескриптора, Q – кількість слів у словосполученні, можна використовувати як оціночний для прийняття рішення щодо включення певного слова чи словосполучення до словника. Сортування складеного словника за частотою, або за отриманим показником, незважаючи на уявну простоту цього методу, дає непогану вершину списку на достатньо великих обсягах текстів. Звичайно, найбільш частотними будуть однослівні дескриптори, попри все – загальноживані слова і загальнонаукові слова типу *система, метод* тощо. Але ж ми вже зазначали, що дана методика не враховує ані семантичні, ані прагматичні властивості слів, а використовує лише формальні методи для відбору претендентів на включення до словника, а вирішальне слово завжди надається експертові.

Література

1. Антонов А.В. Информационно-поисковая система Galaktika-ZOOM с элементами анализа на гипермассивах информации // НТИ. Сер. 1. – 2001. – № 8. – С.12–21.
2. Лазарва О.Я. Про деякі продуктивні моделі науково-технічних термінів. // Українська термінологія і сучасність: Зб.наук.праць. Вип.4. – К: КНЕУ, 2001. – С.185–188.
3. Лукашевич Н.В. Автоматизированное формирование информационно-поискового тезауруса по общественно-политической жизни России // НТИ. Сер.2. – 1995. – №3. – С. 21–24.
4. Суперанская А.В., Подольская Н.В., Васильева Н.В. Общая терминология. Вопросы теории. – М.: УРСС, 2003.

В. Шульгіна, к. філол. н. *
УДК 81'322.373

ПРОБЛЕМИ КОМП'ЮТЕРНОЇ ЛЕКСИКОГРАФІЇ У ГАЛУЗІ ІНФОРМАЦІЙНОЇ ДІЯЛЬНОСТІ

Метою статті є аналіз можливостей створення електронного забезпечення лексикографічними виданнями такої сфери, як документознавство. Проблема забезпечення навчального процесу на факультетах документознавства необхідною літературою включає в себе і підготовку термінологічних словників цієї потрібної сфери. Термінологічний словник документознавчої галузі є першою спробою узагальнення термінологічних понять, необхідних в роботі документознавців.

Прикладна лінгвістика як один з новітніх сучасних напрямків розвитку мовознавства виокремилася з традиційної лінгвістики не так давно. Серед проблем, якими займаються дослідники в галузі прикладної лінгвістики, проблема комп'ютерної лексикографії посідає одне з чільних місць [1]. Це пов'язано, насамперед, з активним розвитком комп'ютерної техніки, з її використанням у різних сферах науки та народного господарства. Комп'ютерна лексикографія є складовою частиною лексикографії як розділу традиційного мовознавства. Тому досвід лінгвістів-лексикографів, силами яких створювалося словникове багатство будь-якої мови, є надійним фундаментом для подальшого його розвитку у межах комп'ютерної лексикографії. Важливу роль у розвитку сучасної комп'ютерної лексикографії відіграють і новітні лексикографічні розробки, які є безпосередньою реакцією на вимоги сьогодення.

Останнім часом українська лексикографія поповнилася цілою низкою нових термінологічних словників. Поява цих видань зумовлена у певній мірі оновленням змісту багатьох наукових сфер українського суспільства, яке переживає часи набуття і зміцнення незалежності нашої держави, часи переходу від соціалістичної економіки до ринкової. Економіко-соціальні зміни, що відбуваються в Україні, не могли не позначитися і на розвиткові такої наукової галузі, якою є документознавство. Адже сама підготовка фахівців-документознавців для України стала справою зовсім новою – до розпаду СРСР таких спеціалістів готували лише в Московському історико-архівному інституті. Зараз, з урахуванням потреби в грамотних документознавцях, яку відчувають як державні установи, так і кожна приватна фірма, факультети документознавства та

* © В. Шульгіна, 2006