

17. Тищенко О. М., Кульбіда С. В. Українська жестова мова глухих як об'єкт лексикографічної параметризації / Лексикографічний бюлетень. – Київ, 2005. – Вип. 12. – С. 37–43.
18. Українські говірки Донеччини. Фонотека. – Донецький національний університет. – Донецьк., 2000. – CD, Вип. 1.
19. Фрадкіна Р. Н. Говорящие руки. Тематический словарь жестового языка глухих России. – М., 2001. – 597с.
20. Цыпин Р. Жестовый язык и письменность / Наше життя, 2002. – 16 серпня.

*Л. Тищенко\**

Інститут української мови НАН України (Київ)  
УДК 161.2.81'373.46.322

### НАУКОВА ТЕРМІНОЛОГІЯ В НАЦІОНАЛЬНОМУ КОРПУСІ УКРАЇНСЬКОЇ МОВИ

*Питання, порушені в статті, пов'язані з проблемою створення національного корпусу української мови, яку опрацьовують у відділі лексикології, лексикографії та національного корпусу Інституту української мови НАН України. У статті висвітлені засади формування підкорпусу наукової термінології, обґрунтовано принципи добору текстів, а також хронологічні межі.*

Для української мови завдання створення різноманітних корпусів текстів є дуже актуальним. Без таких корпусів і баз, побудованих за різними параметрами класифікації текстів та вміщеної в них інформації, неможливе успішне розв'язання багатьох важливих теоретичних і практичних проблем сучасної україністики, зокрема вироблення критеріїв унормування та кодифікації лексики й граматичного ладу сучасної української мови, вивчення властивих їм динамічних процесів, проблем мовного планування й прогнозування, створення надійного підґрунтя для розв'язання проблем варіантності та конкурентоздатності тих чи інших одиниць, питомої ваги в мові нових тенденцій, вивчення співвідношення інновацій і сталих явищ у різних підсистемах мови, вивчення тенденцій розбудови текстових структур і вироблення надійних рекомендацій для формування структури текстів нових типів. Особливо це стосується таких функціонально-стильових різновидів української мови, як ділова мова, законодавчі та інші державні документи (наприклад, мова дипломатії), реклама, мова засобів масової інформації, тексти з тих фахових галузей, які ще донедавна обслуговувалися російською мовою.

Однією із особливостей сучасної української мови є динаміка її лексико-семантичної системи, частиною якої є термінологічна лексика. Тому введення наукової термінології до національного корпусу української мови сприятиме глибшому й детальнішому опису історії української мови, точнішій хронологічній атрибуції лексичних і семантичних новацій, деталізованій демонстрації семантичного, стилістичного й фразеологічного порядку, адже, на думку В. Виноградова, «... історія слова повинна проводити весь зміст і ланцюжок змістових «приростів», усі «метаморфози». Вона намагається розкрити певні умови вживання слова в різні періоди його мовного життя» [2: 19].

У лінгвоукраїністиці від 60-х років минулого століття і до сьогодні існують значні досягнення в суміжних з корпусною лінгвістикою напрямках – прикладній (комп'ютерній) лінгвістиці, а саме праці В. С. Перебийніс, Т. О. Грязнухіної, Н. Ф. Клименко, М. М. Пешак, С. А. Карпівської, Л. І. Комарової, О. М. Демської-Кульчицької та створені комп'ютерні текстові ресурси, зокрема Комп'ютерна лабораторія КНУ ім. Т. Шевченка та лексикографічний процесор «СЛОВО», який розробили львівські дослідники О. М. Коссака, О. Коруд, Л. Хвишун, який допомагає створювати одно- та двомовні термінологічні словники. На сьогодні теоретичні засади одномільйонного корпусу української мови розроблено, а практичне його створення тільки починається [3]. Цю роботу проводить група національного корпусу Інституту української мови НАН України. Під час визначення аспектів створення термінологічного підкорпусу української мови виникають проблеми, які ми спробуємо вирішити. Це питання про жанри й джерела підкорпусу наукової термінології, про хронологічні межі, про метатекстову інформацію

---

\* © Л. Тищенко, 2006

цього підкорпусу, про розроблення системи кодування на основі стандартних мов кодування.

Основою нашого підкорпусу буде репрезентативна вибірка текстів 60-х років XIX століття – до сучасного періоду, зокрема: а) роботи видатних науковців України; б) публіцистика; в) науково-популярна література; г) ділові, офіційні документи. Ці джерела відповідатимуть наступним вимогам:

- видання українською мовою (в Україні чи за її межами);
- надання переваги сучаснішим перевиданням, які максимально зберегли орфографію й пунктуацію джерела.

При складанні корпусу, як ми вже говорили, важливим є питання про хронологічні межі, тобто тексти/уривки текстів якого періоду потрібно включати до науково-термінологічного підкорпусу. Системний підхід до творення нової, збирання та перегляду наявної народної термінології починається з другої половини XIX століття переважно в Галичині. Оскільки саме з II половини XIX ст. було закладено підґрунтя, на якому розвивалася, збагачувалася й удосконалювалася українська термінологія. Вперше проблема терміна в українській мові постала в 60-ті рр. XIX ст. на сторінках журналу «Основа», а також вченими-членами Наукового товариства імені Шевченка. Будемо вважати 60-ті роки XIX століття нижньою хронологічною межею опису термінології. Не залишаться поза увагою роботи науковців 20 – поч. 30-х років XX ст. – періоду так званого «золотого десятиріччя», тобто до нашого підкорпусу будуть внесені тексти/уривки текстів, починаючи з другої половини XIX ст. і закінчуючи XXI ст.

Не менш важливим є питання про жанри й джерела корпусу. Як зазначав А. М. Баранов, описуючи проблеми корпусу текстів сучасної російської публіцистики (90-і рр. XX ст.), «... текстів повинно було бути достатньо багато, щоб відобразити всі релевантні властивості проблемної галузі» [1: 131–132], у нашому разі – властивості термінології як окремого функціонально-стильового різновиду сучасної української літературної мови. Українська мова, зокрема наукова термінологія – це великий масив різних за походженням і семантичною/стилістичною віднесеністю слів і словосполучень: це вся без обмежень термінологія (як власне українська, так і запозичена, особливо терміни, які трапляються в декількох термінологічних системах, або терміни, які набули широкого (переносного) значення). Постає питання, скільки місця в загальному корпусі має займати термінологія? На нашу думку, це приблизно 10% всієї лексики.

Наступна проблема – це метатекстова інформація, завданням якої є забезпечення багатофункціонального використання корпусу різними дослідниками. Тексти будуть уводитися до корпусу й перетворюватимуться у корпусний формат TEI, де буде створено електронний заголовок до тексту, а полотно тексту розмічене на III рівні CSS. Модель електронного заголовка повинна містити дані про:

- 1) повну й коротку назву тексту;
- 2) графіку тексту (книжно-слов'янська; орфографія певного хронологічного періоду; використання латинської та інших видів графіки), наявність іноземних елементів у тексті, наявність формул, малюнків і т. ін.;
- 3) атрибуцію текстів його автором;
- 4) прізвище, ім'я, по-батькові автора/авторів текстів;
- 5) науковий/літературний псевдонім автора;
- 6) дата народження/дата смерті автора;
- 7) датування створення/датування першої публікації;
- 8) вказівку на першу публікацію (зібрання творів, окреме видання, наукове видання; журнал, газета, архівний матеріал і т. д.);
- 9) вказівку на наявність авторських (редакторських) коментарів, приміток, виправлень, спеціальних зауважень щодо підготування тексту до видання;
- 10) вказівку на спосіб уведення тексту в корпус (ручне введення, СД-версія, сканування, Інтернет, оригінал-макет видання, авторська копія);
- 11) вказівку на осіб, які вводили текст до корпусу, робили розмітку й філологічну перевірку тексту; осіб, які підготували формуляр тексту;
- 12) вказівку на місце зберігання тексту (вказати на заклад, куди текст відправлено після обробки – адреса тексту).

Загалом, дослідницькі можливості сучасних електронних текстових корпусів/підкорпусів досить різноаспектні. Але найважливішим аспектом аналізованого дослідного підходу є те, що він «змушує подивитися на багато проблем з погляду,

відмінного від традиційного опису мови» [5]. Як справедливо стверджує В. Риков, корпус повинен «мати власну концепцію, бути чітко детермінованою щодо призначення, а фактичний матеріал повинен репрезентувати реальні тексти національної мови, враховуючи весь історичний період її існування, варіантність мовної реалізації, стилістичну диференціацію і теорію поширення» [6]. Отже, упровадження корпусних методик в українське термінознавство повинно обов'язково враховувати, що незалежно від типу здійснюваних корпусних досліджень, їхня якість залежатиме передовсім від коректності побудови самого термінокорпусу, яка має відповідати національній філологічній традиції.

#### Література

1. Баранов А. Н. Введение в прикладную лингвистику. – М., 2001.
2. Виноградов. В. «Чтение древнерусского текста и историко-этимологические каламбуры». // Вопросы языкознания. – 1968. – №1 – С. 19.
3. Демська-Кульчицька О. М. Основи національного корпусу української мови. – К., 2005.
4. Коссак О., Коруд О., Хвищун Л. Система комп'ютерної підтримки словникових баз даних СЛОВО // Проблеми українізації комп'ютерів. – К., 1993.
5. Мустайоки А., Копотов М. К вопросу о статусе эквивалентов слова типа потому что, в зависимости от, к сожалению // Вопросы языка. – 2004. – №3. – С. 88–107.
6. Рыков В. В. Корпус текстов как отражение состояния русского языка. – <http://gukov-cl.narod.ru>. – 2002.

*А. Кузнцова\**

Інститут філології КНУ ім. Т. Шевченка (Київ)  
УДК 161.2.81'373.322

### ПАРАМЕТРИЗАЦІЯ ПУБЛІЦИСТИЧНИХ НЕОЛЕКСЕМ У КОРПУСНІЙ ЛІНГВІСТИЦІ

*У статті аналізуються публіцистичні неолексеми з погляду їхньої параметризації. Так, проаналізовано корпус Частотного словника українського публіцистичного стилю, запропоновано етапи роботи зі створення словника неологізмів, визначено перспективи дослідження.*

Характерною особливістю словникового складу української мови, як і будь-якої іншої, є його безперервний розвиток. Актуалізація в мові застарілих слів або значень, поява нових слів або розширення значень наявних слів спричиняється потребою людини називати нові предмети, явища, поняття, що з'являються у суспільно-економічному, політичному й культурному житті народу.

Пропонована робота присвячена дослідженню неолексем у публіцистичних текстах. Орієнтовані на пересічного читача, такі тексти є одним з найбільших джерел нових лексичних одиниць і саме в них ми маємо нагоду зафіксувати нові слова, які постійно виникають у мовленні. Але на сьогодні закономірності їх виникнення й функціонування ще не стали предметом фронтальних лінгвістичних досліджень.

Матеріалом нашого дослідження є корпус Частотного словника українського публіцистичного стилю – 2004, укладений студентами спеціалізації «Комп'ютерна лінгвістика». Лексичний склад словника групувався за рубриками: Політика (67 тис. слововживань), Суспільство (56 тис.), «Помаранчева революція» (54 тис.), Культура (46 тис.), Закон (29 тис.), Економіка (20 тис.), Спорт (12 тис.), Здоров'я (6 тис.), Релігія (6 тис.). Загальний обсяг вибірки становить 300 тис. слововживань. Словник укладений за матеріалами центральних та обласних газет «Українська Правда», «Україна молода», «Голос України», «Дзеркало тижня», «Наша Газета+», «Львівська газета», «Рівне вечірне», «ПРАВДА», «Критика», «Контекст», «Коментар», «Церква і молодь», виданих протягом кінця 2003 та всього 2004 року. Дослідження матеріалу словника має на меті створення параметризованого словника неологізмів, здійснення опису їх словотвірних особливостей і розподіл нової лексики за функціональними групами. Досягнути цієї мети можна завдяки вирішенню наступних завдань:

1. Створення корпусу текстів публіцистики, об'єднаних за хронологічним і тематичним принципом. Форма подачі – електронна. На цьому етапі ми отримуємо вже готовий матеріал, поділений у словнику за вищезгаданими рубриками і обмежений, відповідно, виданнями кінця 2003 та 2004 року.

\* © А. Кузнцова, 2006