

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ У ПОЛЬСЬКІЙ ЛЕКСИКОГРАФІЇ: СУЧАСНИЙ СТАН ТА ПЕРСПЕКТИВИ

У статті аналізується сучасний стан польської комп'ютерної лексикографії, вказуються основні періоди її розвитку, визначаються перспективи застосування інформаційних технологій у створенні електронних словників нового покоління.

Ключові слова: електронний словник, інформаційні технології, польська комп'ютерна лексикографія, тлумачний словник.

Усебічне й масштабне проникнення комп'ютерної техніки в різні сфери суспільної діяльності зумовлює активізацію роботи лексикографів-славістів щодо впровадження інформаційних технологій та створення нових словникових ресурсів із використанням комп'ютерних систем, які надають нові інструментальні можливості для фіксації й дослідження одиниць мови. Сьогодні розвиток комп'ютерної лексикографії характеризується застосуванням принципів автоматизованої побудови й укладання реєстрів традиційних паперових словників, розробленням електронних словників як елементів інформаційно-пошукових підсистем та створенням лексикографічних систем як мовно-інформаційних об'єктів принципово нового типу¹.

У сучасній слов'янській лексикографії пріоритетними є наукові дослідження, пов'язані з автоматичним одержанням текстів різних словників, комп'ютерними програмами різноманітного застосування в системах штучного інтелекту, машинного перекладу та в інформаційно-пошукових системах (М. Банько, А. М. Баранов, П. Жмігородський, М. М. Пещак, В. А. Широков, З. Салоні та ін.)².

Крім значної економії часу та інтелектуальних ресурсів, комп'ютерне опрацювання лексикографічних даних дає можливість підтримувати точність і ретельність контролю лексикографічного процесу, оперативно виправляти помилки, створювати нові лексикографічні та термінографічні комплекси³. За-

¹ Широков В. А. Елементи лексикографії.— К., 2005.— С. 18.

² Баранов А. Н. Введение в прикладную лингвистику.— М., 2003.— 360 с.; Пещак М. М. Нариси з комп'ютерної лінгвістики.— Ужгород, 1999.— 199 с.; Широков В. А. Лінгвістичні та технологічні основи тлумачної лексикографії.— К., 2010.— 295 с.; Широков В. А. Комп'ютерна лексикографія.— К., 2011.— 351 с.; Bańko M. Z pogranicza leksykografii i językoznawstwa.— Warszawa, 2001.— 336 s.; Saloni Z. Perspektywy polskiej leksykografii jednojęzycznej // Poradnik językowy.— Warszawa, 1996.— N 7.— S. 1–18; Żmigrodzki P. Wprowadzenie do leksykografii polskiej.— Katowice, 2009.— 310 s.

³ Дубичинский В. В. Теоретическая и практическая лексикография.— Вена ; Х., 1998.— С. 55.

стосування інформаційних технологій у традиційному словникарстві дозволяє також проводити різні лексикологічні та лексикографічні роботи в автоматизованому режимі й надає можливість створення лінгвістичних, зокрема лексикографічних, баз даних і знань з метою використання їх у дослідницькому та технологічному режимах.

Як показує вивчення наукових джерел, у лінгвістичній славістиці немає однозначного трактування багатьох понять, пов'язаних з комп'ютерною лексикографією. Так, польські лексикографи поняття «електронний словник» застосовують до різного типу об'єктів: 1) словникових модулів, убудованих у програми для різноманітних (не тільки лексикографічних) використань; 2) окремих комп'ютерних програм, записаних на електронних носіях, що містять лексикографічну базу даних, а також спеціальних пошукових інструментів; 3) словникових систем, доступних з інтернету; 4) перекладних словників у вигляді окремих мініатюрних комп'ютерних пристроїв; 5) систем, які полегшують уведення текстів⁴. Однак найчастіше термін «електронний словник» у теорії та практиці лексикографування застосовується на позначення перекладних словників у вигляді окремих комп'ютерних засобів. Водночас поняття «комп'ютерний словник» уживається як щодо об'єктів, перелічених у другому пункті, так і в загальному використанні поряд із терміном «електронний словник»⁵.

Таку ж неоднозначність трактування спостерігаємо й у східнослов'янських дослідників. Зокрема, російський учений А. М. Баранов автоматичним (електронним) словником називає словник у спеціальному машинному форматі, призначений для використання користувачем або комп'ютерною програмою оброблення тексту⁶. Український дослідник В. В. Дубічинський поняття «електронний словник» визначає як частину автоматизованої системи інтелектуального призначення із зібраними й певним чином упорядкованими найменуваннями понять у тій галузі знань, з якими працює або на які спирається ця автоматизована система⁷.

Перші польські електронні словники з'явилися в 90-х роках ХХ ст. У цей період виходять у світ «Komputerowy słownik języka polskiego» (1996)⁸ та електронна версія 11-томного словника польської мови В. Дорошевського (1997)⁹.

За нашими спостереженнями, електронний варіант великого тлумачного словника В. Дорошевського є комп'ютерною програмою, за допомогою якої можна проводити перегляд окремих словникових статей у графічному режимі, а також пошук словникового матеріалу за певними лексикографічними параметрами. Оснащена такою пошуковою системою, комп'ютерна версія в алфавітному порядку дає повний доступ до реєстрових одиниць одинадцятитомного тлумачного словника В. Дорошевського (1958–1961 рр.). Натомість «Komputerowy słownik języka polskiego»¹⁰, в основі якого лежить лексичний матеріал трьохтомного «Słownika języka polskiego» М. Шимчака, а також «Słownika wyrazów obcych» Е. Соболя, надає швидкий доступ до лексикографічної інформації кожного з перелічених джерел з можливістю пошуку та систематизації словникового матеріалу за відповідними ремарками.

⁴ Żmigrodzki P. Op. cit.— S. 251–252.

⁵ Ibid.— S. 252.

⁶ Баранов А. Н. Знач. праця.— С. 85.

⁷ Дубічинський В. В. Українська лексикографія: історія, сучасність та комп'ютерні технології.— Х., 2004.— С. 72.

⁸ Komputerowy słownik języka polskiego.— Warszawa, 1996.

⁹ Słownik języka polskiego [Wersja elektroniczna] / Red. W. Doroszewski.— Warszawa, 1958–1969, 1997.— Т. 1–11.

¹⁰ Назва продукту з погляду металексикографії є неточною, оскільки комп'ютерним словником може називатися будь-яка інша електронна лексикографічна праця даного типу.

Саме після появи цього словника польські лексикографи П. Жмігродський, З. Салоні, К. Томчак та ін. відзначають переваги електронного носія над паперовими виданнями. Серед основних позитивних характеристик комп'ютерних лексиконів учені називають: гнучкість інформації, зручність оперування лексикографічними даними, швидкий доступ до реєстрових слів та їх пошук за різноманітними критеріями, можливість інтеграції кількох словників в одній програмі¹¹. Ретельно опрацьований комп'ютерний словник, на думку П. Жмігродського, знімає будь-які обмеження, притаманні традиційному словникарству¹².

Як відомо, на початку ХХІ ст. у польській комп'ютерній лексикографії спостерігалася певна стагнація у застосуванні цифрових технологій. Електронні словники розглядалися здебільшого як додаткова форма паперової версії (через доступ в Інтернеті або додавання оптичного диска до традиційного словника). Бар'єром на шляху розвитку інформаційних технологій став невисокий попит на електронні словники з боку масового користувача¹³. На думку М. Банька, комп'ютеризація інформації, з одного боку, збільшує можливості паперових словників (дозволяє здійснювати вибірковий пошук реєстрових одиниць, фрагментів слів тощо), а з другого, — виявляє непослідовності й прогалини друкованого тексту. Можливості комп'ютерних словників значно б зросли, якби вони із самого початку створювалися в електронній формі¹⁴.

Якісно інший етап розвитку польської комп'ютерної лексикографії розпочався тоді, коли електронний словник перестав сприйматися дослідниками лише як зведення вказівок щодо вживання або значення того чи іншого слова і почав розглядатися як джерело певним чином структурованої лінгвістичної інформації, на основі якої виявляються додаткові механізми функціонування мови як системи. Це підтверджує висновки В. А. Широкова про загальну тенденцію зміщення акценту комп'ютерно-лексикографічних праць із технології укладання словника до постановки й розв'язання питань опрацювання тексту, реалізованого словником. Для сучасної комп'ютерної лексикографії одним із важливих завдань стає дослідження природномовного тексту словника з метою формалізації відображених у ньому мовних явищ¹⁵.

Польська комп'ютерна лексикографія серед інших завдань передбачає створення комплексних лексикографічних систем, що інтегрують репрезентативну лінгвістичну та когнітивну інформацію певної природної мови або сукупності мов. Ідеалом теорії і практики польської лексикографії залишається універсальний словник, який з максимальною повнотою міг би відобразити мовну стихію, усі розряди одиниць лексикону зі всіма властивими їм ознаками. Вимогою до такого словника є подання всебічної інформації про специфіку реєстрових одиниць: їх значення, основні граматичні характеристики, стилістичні властивості, контексти вживання, фразеологізми та інші семантико-синтаксичні структури, до складу яких входить реєстрове слово.

Інноваційним продуктом польської лексикографії є комп'ютерна версія «Uniwersalnego słownika języka polskiego» С. Дубіша (2004 р.)¹⁶. Цей словник,

¹¹ Saloni Z. Op. cit.— S. 16; Tomczak K. O wyższości słownika komputerowego nad tradycyjnym // Poradnik językowy.— Warszawa, 1997.— N 7.— S. 59–64.

¹² Zmigrodzki P. Teraźniejszość i przyszłość polskiej leksykografii językoznawczej // Poradnik językowy.— Warszawa, 2003.— N 1.— S. 25.

¹³ Ibid.— S. 25.

¹⁴ Bańko M. Stan i perspektywy współczesnej polskiej leksykografii jednojęzycznej // Językoznawstwo w Polsce.— Opole, 2003.— S. 132–133.

¹⁵ Широков В. А. Інформаційна теорія лексикографічних систем.— К., 1998.— С. 70.

¹⁶ Uniwersalny słownik języka polskiego [Wersja elektroniczna] / Red. S. Dubisz.— Warszawa, 2004.

у якому використано досвід укладання різних типів словників (тлумачних, орфографічних, фразеологічних, словників іншомовних слів, словників синонімів, антонімів тощо), репрезентує різноаспектну словникову інформацію. В електронній версії словника, крім значного лексикографічного матеріалу з великими можливостями пошуку (окремо винесено індекси «Hasła słownika», «Przysłowia», «Terminy i frazeologizmy», «Indeks a tergo»), доступна інтерактивна мережа відсилань між реєстровими одиницями, які пов'язані певними значеннєвими відношеннями (синонімічними, антонімічними, гіперо-гіпонімічними), що надає широкі можливості для проведення різного роду лінгвістичних досліджень. Лексикографічна праця С. Дубіша є першим польським тлумачним словником універсального типу, лінгвістичний матеріал якого ґрунтується на основі Корпусу текстів польської мови, що дозволяє мінімізувати суб'єктивність в інтерпретації значень реєстрових одиниць.

Окремо варто спинитися на Граматичному словнику польської мови, який укладено в електронній формі 2007 року (автори: З. Салоні, В. Грущинський, М. Волинський, Р. Волош)¹⁷. У лінгвістичній полоністиці відзначається, що цей словник створено за допомогою спеціальних лінгвістичних алгоритмів, у результаті чого подано граматичну характеристику близько 254 тис. польських мовних одиниць. Формат такого словника дозволяє фіксувати всі флексійні форми описаних реєстрових слів (понад 4 млн словоформ), а також здійснювати пошук за окремими граматичними класами: іменниками, прикметниками, числівниками, дієсловами та іншими частинами мови. Окрім того, реєстрові одиниці словника розміщено двома способами: а fronte — звичайний алфавітний порядок слів, а також а tergo (зворотний порядок), що дозволяє автоматично виявляти лексеми, які мають подібну словозмінну парадигму.

На сьогодні значна частина польських словникових систем функціонує в онлайн-режимі. Вони надають користувачам можливість прямого доступу до фундаментальних словникових ресурсів через Інтернет. Найважливішими серед них є: «Słownik języka polskiego» (www.sjp.pwn.pl)¹⁸, який є електронною версією словника «100 tysięcy potrzebnych słów» Є. Бральчика (2005 р.), «Uniwersalny słownik języka polskiego» (www.usjp.pwn.pl) — доступ платний, «Wielki słownik ortograficzny PWN» (www.so.pwn.pl) та «Słownik wyrazów obcych PWN» (swo.pwn.pl).

Однак польські лексикографічні праці, що виходять протягом останніх п'ятнадцяти років, усе ж не можуть претендувати на роль повного джерела опису лексики сучасної польської мови. У зв'язку із значним поширенням наукової та технічної термінології різних галузей знань, процесу термінологізації мовних одиниць, упровадженням у польську мову величезного масиву нової лексики, насамперед іншомовного походження, зростає потреба створення нового великого академічного словника польської мови, який міг би прийти на зміну словнику В. Дорошевського, від часу створення якого пройшло понад п'ятдесят років.

З одного боку, зазначені мовні процеси є ознакою нормального природного розвитку мови, але з другого, — приводять до певних проблем при лексикографуванні польської лексики загальномовними та термінологічними словниками, оскільки лексикографічні джерела не завжди вчасно фіксують зміни, що відбулися в лексичній семантиці сучасної літературної мови, часто не встигаючи врахувати темпи розвитку загальномовної і термінологічної лексики. Розв'язання цієї проблеми полягає у створенні нових словникових праць тлу-

¹⁷ Słownik gramatyczny języka polskiego / Z. Saloni, W. Gruszczyński i in. — Warszawa, 2007.

¹⁸ Усі вказані електронні адреси актуальні станом на квітень 2012 року.

мачного та термінологічного типу із застосуванням інформаційно-комп'ютерних технологій, що дозволить ефективніше вирішувати поставлені лексикографічні завдання.

У перспективі польських лексикографів — «Wielki słownik języka polskiego»¹⁹, робота над яким здійснюється авторськими колективами академічних наукових осередків Варшави, Кракова, Катовіц, Торуня. В основі проекту лежить загальна концепція нового великого тлумачного словника дескриптивного характеру у виключно електронному варіанті, що виконуватиме функцію основного інформативного джерела польської мови, охоплюючи польську лексику, починаючи від 1945 року. Укладачі передбачили необхідність вичерпного опису якнайбільшої кількості польських лексичних одиниць згідно із сучасним станом мовознавчої науки та новітніми стандартами світової лексикографії.

Основними принципами лексикографування польської лексики у майбутньому словнику є:

1. Доступ до мовних одиниць в електронній формі, що сприятиме постійному збагаченню ресурсів словника, даватиме можливість коректування, удосконалення лексикографічного опису реєстрових слів.

2. Синхронний опис — словник охоплюватиме новітню польську лексику останнього десятиліття ХХ ст. — поч. ХХІ ст.

3. Автономність — незалежність словника від лексикографічних концепцій попередніх тлумачних лексиконів.

4. Дескриптивність — характеристика мовних одиниць із урахуванням їх функціонування та уживаності в сучасній польській мові.

5. Урахування надбань сучасної мовознавчої науки, зокрема в галузі семантики, граматики і прагматики²⁰.

Така постановка завдання є для польського словникарства доволі новою. Вона вимагає вироблення нових концептуальних засад, що враховували б як суто лінгвістичні аспекти, так і технологічні моменти лексикографічного процесу. Це підтверджує висновки про необхідність і перспективи активізації впровадження у традиційне слов'янське словникарство новітніх комп'ютерних методів опрацювання лексикографічних даних, розроблення нової методики та засобів комп'ютеризації процесу лексикографування аж до створення лексикографічних систем нового типу.

V. V. CHUMAK, R. P. TYMOSHUK

INFORMATION TECHNOLOGIES IN POLISH LEXICOGRAPHY: CURRENT STATE AND PROSPECTS

This article analyzes the current state of Polish computer lexicography, indicates major periods of its development. The prospects of applying information technologies of composing new generation dictionaries are determined.

Key words: electronic dictionary, information technologies, Polish computer lexicography, explanatory dictionary.

¹⁹ Докладну інформацію про проект «Wielkiego słownika języka polskiego» можна знайти за адресою: <http://www.wsjp.pl/>

²⁰ *Zmigrodzki P.* Jak powstaje «Wielki słownik języka polskiego» // *Język polski wczoraj — dziś — jutro* / Red. B. Czopek-Kopcuch. — Kraków, 2010. — S. 151–152.