

Дарчук Н. П., к. філол. н., доц.,
Лангенбах М. О., к. філол. н., асист.,
Інститут філології КНУ імені Тараса Шевченка, м. Київ

ЕЛЕКТРОННИЙ СЛОВНИК ЯК ДОСЛІДНИЦЬКА БАЗА ДАНИХ (на прикладі електронного словника мови Тараса Шевченка)

Стаття описує електронний словник мови Тараса Шевченка. Розглянуто структуру бази даних та інтерфейсу, визначено наукову вагу проекту як такого, що лежить у царині двох напрямів сучасного мовознавства: корпусної лінгвістики та комп'ютерної лексикографії. Дається перелік функціональних можливостей словника та сфер його застосування.

Ключові слова: електронний словник, Тарас Шевченко, комп'ютерна лексикографія, корпусна лінгвістика.

Вивчення творчості письменника, його індивідуального стилю, особливостей формулювання думок і творення образів неможливе без глибокого і всебічного аналізу його мови. Інформація з різних мовних рівнів (морфологічного, синтаксичного, семантичного), відомості про прагматичні, стилістичні та інші особливості авторських текстів були і є невичерпним джерелом для словників і граматик. Як показує лексикографічна практика, лінгвістично коректно сформований текстовий масив – запорука якісних мовознавчих досліджень.

Нескінченний інтерес до постаті Тараса Шевченка породжує все нові й нові дослідження його творчості. Оскільки ж сучасну наукову добу характеризує активне впровадження новітніх інформаційних технологій, закономірним кроком є створення відповідного інструментарію для роботи з текстами видатного поета. Це ще один ресурс для поглиблення наших уявлень про духовний світ митця, усвідомлення його місця у літературному процесі, зокрема у витворенні національної мови та культури загалом.

Водночас подібні розробки розглядаються в аспекті накопичення фактичної мовної інформації, створення філологічно компетентних текстових зібрань, чим займається окрема галузь мовознавчої науки – корпусна лінгвістика. В останні десятиліття минулого сторіччя зусилля багатьох країн були спрямовані на створення національних універсальних корпусів текстів із необхідними й достатніми кількісними та якісними параметрами для укладання на їх основі словників і граматик національних мов. Електронні корпуси текстів з їх технологією представлення дають змогу автоматизувати лінгвістичні дослідження і мають фундаментальне значення для лінгвістичної теорії і практики [1, с. 107]. На сьогодні створення корпусу розглядається як обов'язок стосовно національної мови. Роботи у цьому напрямку проводяться і в Україні, і першою важливою умовою розвитку української корпусної лінгвістики є накопичення текстів різних стилів і часових зрівів. Оскільки ж творчість Тараса Шевченка зіграла винятково значущу роль у становленні української мови, творчу спадщину поета не можна лишати поза увагою у площині корпусних розробок.

Зазначені вище аспекти зумовлюють **актуальність** і **новизну** проекту електронного словника мови Тараса Шевченка, розробленого в межах заходів святкування ювілею поета.

Метою нашої роботи було створення такого лінгвістичного і програмного продукту, який би, з одного боку, надавав широкі відомості про мову Тараса Шевченка, з іншого ж, був би зручним для використання у різноманітних філологічних дослідженнях, з усіма перевагами електронного словника – багаторазовим і багатоаспектним використанням, автоматичною переорганізацією, доповненням уміщеної в ньому інформації про ті чи інші мовні об'єкти тощо [4, с. 55]).

Сформульована мета передбачала виконання низки **завдань**, серед яких:

- лінгвістичне опрацювання текстів Т. Шевченка;
- укладання бази даних із граматичними та кількісними характеристиками мовних одиниць;
- створення зручного користувачького інтерфейсу, за яким можна було б здійснювати пошук, сортувати та статистично опрацьовувати всю зібрану в базі інформацію відповідно до потреб дослідників.

Об'єктом дослідження були оцифровані поетичні тексти Т. Шевченка, **предметом** – їх лексичні, морфологічні та синтаксичні характеристики.

Над проектом працював колектив лабораторії комп'ютерної лінгвістики, зокрема В. Сорокін, М. Лангенбах, Я. Ходаківська, О. Тютенко, а також група студентів магістратури у складі М. Білоконь, А. Дригіної, К. Ксьондзик, В. Мяснікової, М. Лисенко та А. Шкоди. Керівник проекту – Н. Дарчук.

Останні десятиліття лінгвістики є епохою розквіту лексикографії, оскільки цей період характеризує зростання кількості й різноманіття словників, удосконалення лексикографічних методик в плані системного та інтегрального опису мовних одиниць і фактів. У сучасній українській лексикографії на окрему увагу заслуговує питання створення та використання авторських словників [2, с. 44].

Електронний словник мови Тараса Шевченка є водночас частиною електронного Корпусу української мови і лексикографічним продуктом. Він об'єднує різні типи словникових відомостей – граматичні, частотні, конкордансні (інформацію про контекстуальне оточення одиниць) тощо. Інтерфейс словника складається з чотирьох сторінок. Перша сторінка містить загальний опис проекту, решта подає різноманітні лінгвістичні дані. Так, у розділі «Пошук слів» (рис. 1) передбачено можливість відбору контекстів за конкретними лексемами або словоформами. Окрім того, можна побудувати частотний словник лексем та словоформ. На основі цих відомостей можна робити висновки про лексичне ядро мови Тараса Шевченка, досліджувати тенденції вживання певних слів у їх конкретних граматичних та лексико-семантичних варіантах, зв'язках із контекстами тощо.

У розділі «Морфемно-частотний словник» (рис. 2) пошук організовано за такими параметрами: частинномовна належність слова, типи афіксів. Також у цьому розділі представлений частотний словник морфемних і морфних моделей, що зустрічаються в текстах Т. Г. Шевченка. Це дозволяє аналізувати мову поета з точки зору її морфологічних та морфемних особливостей.

Одиниці пошуку

Пошук слів

Слова Пошук за словом або першими буквами

Словосполучення

Морфемна структура

Частотний словник: Лексем Частота мови: Всі

Показати

Всього записів: 6889

| Слово | Частина мови | Абсолютна частота |
|--------|--------------|-------------------|
| а | сполучн. | 1224 |
| Аароня | ім. ч. р. | 1 |
| аби | сполучн. | 9 |
| аби-то | прислівн. | 1 |
| абичий | займ.-прикм. | 1 |
| або | сполучн. | 52 |
| абощо | част. | 2 |
| ав | част. | 1 |
| Авраам | ім. ч. р. | 1 |

Обмежити результат:
 За частотою з по

Рис. 1. Сторінка пошуку слів

Одиниці пошуку

Морфемно-частотний словник

Тут Ви маєте змогу побудувати частотний словник за вибраним типом морфем. Для цього треба вказати потрібну морфему.

Частотний словник: Префіксів

Частота мови: Всі

Побудувати

Частотний словник морфемних структур слів

Тут Вам надана можливість побачити частотний словник морфемних структур, використаних автором. Умовні позначення: Р - префікс; R - корінь; S - суфікс; I - інтерфікс; X - постфікс; F - флексія.

Частота морфструктур

Всього записів: 94

| Структура | Покриття в тексті |
|-----------|-------------------|
| R | 6605 |
| RX | 38 |
| RRF | 51 |
| RF | 16323 |
| RSSF | 898 |
| PRF | 1631 |
| RR | 70 |
| RRRISF | 1 |
| RRRF | 1 |
| RRS | 27 |

Обмежити результат:
 За частотою з по

Рис. 2. Сторінка морфемного пошуку

Одніці пошуку

Слова

Словосполучення

Морфемна структура

Пошук за словосполученнями

Головне слово (початкова форма)* **Залежне слово** (початкова форма)

* у цьому полі необхідно записати те слово, яке виступає у сполучі головним, незалежно від його позиції

Прійменник

Частина мови

Всі

Тип зв'язку

Всі

Тип словосполучення

Всі

Частина мови

прикметник

Граматична форма залежного слова

Всі

АБО

уведіть повністю словосполучення

Шукати

Всього записів: 2282

| Слово | Частота* | Словосполучення | Контекст |
|---------|----------|------------------|----------|
| алмаз | 1 | Алмазом добрим | Контекст |
| ангел | 2 | ангелом Святим | Контекст |
| ангел | 2 | ангелом святим | Контекст |
| апостол | 3 | апостола святого | Контекст |

Рис. 3. Сторінка пошуку за словосполученнями

Розділ «Пошук за словосполученнями» (рис. 3) дає можливість аналізувати синтаксис мови поета на рівні моделей словосполучень. Опис синтаксису через словосполучення мотивується тим, що це, по-перше, такі елементи мови, які зберігаються і відтворюються більшістю носіїв мови у цілісному вигляді [5, с. 92] і, по-друге, це міжрівневі лексико-морфолого-синтаксичні одиниці [3, с. 190-193]. Шукати словосполучення можна за їх конкретним лексичним складом, за якимось одним із слів-складників, за граматичними характеристиками (частиномовна належність компонентів, їх граматична форма, тип синтаксичного зв'язку). Модель словосполучення подається у словнику з кількісними даними та відповідними контекстами вживання, що відкриває перспективу для вивчення мови поета в аспекті її синтаксичних особливостей.

Вищезазначена інформація надала можливість укласти лінгвістично коректну базу даних текстів Тараса Шевченка як частину Корпусу української мови, яка поєднує у собі кілька типів словників і відкриває можливості для подальшого аналізу мови поета з точки зору її граматичної та лексичної структури, стилістичних і статистичних параметрів тощо. Методика, використана для створення електронного словника, може в подальшому бути застосована для опрацювання творчих набутоків інших авторів, що буде суттєвим внеском у розвиток української філології. Окрім того, подібні розробки є надзвичайно важливими в контексті такого перспективного напрямку, як корпусна лінгвістика. Для проведення теоретичних і прикладних досліджень у галузях когнітивної лінгвістики, стилістики, семасіології, функціональної граматики, словотвору, для створення автоматичних систем опрацювання тексту та систем штуч-

ного інтелекту бракує узагальнених даних про закономірності функціонування мовних одиниць. Необхідну лінгвістичну інформацію для опрацювання її у філологічних студіях можна одержати з корпусів текстів, розбудова яких є ознакою нашого часу.

Таким чином, можна стверджувати, що представлений нами словник забезпечує широку сферу застосування, яка охоплює як теоретичні розвідки, так і прикладні аспекти. Цьому сприяє достатня параметризація лінгвістичної розмітки текстів та гнучкість інтерфейсу користувача, що свідчить про високу ефективність цього словника та електронних словників у цілому.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бук С. Корпус текстів у лінгводидактиці (на матеріалі омонімії у корпусі великої прози Івана Франка) / С. Бук. – Вісник Львівського університету. Серія філологічна. Випуск 57. – Львів, 2012. – с. 106-116
2. Гливінська Л. Корпус української мови як джерело лексикографічної рецепції художнього тексту / Л. Гливінська. – OPERA SLAVICA, XXIII, 2013, 4. – с. 43-50
3. Головин Б. Н. Введение в языкознание / Б. Н. Головин. – М.: Высшая школа, 1966. – 332 с.
4. Карпіловська Є. А. Вступ до прикладної лінгвістики: комп'ютерна лінгвістика: Підручник / Є. А. Карпіловська. – Донецьк: ТОВ «Юго-Восток, Лтд», 2006. – 188 с.
5. Марчук Ю. Н. Компьютерная лингвистика / Ю. Н. Марчук. – М.: АСТ: Восток-Запад, 2007. – 317 с.

Стаття надійшла до редакції 17.04.2014 р.

Дарчук Наталья, канд. філол. н., доц.,

Лангенбах Маргарита, канд. філол. н.,

Институт филологии КНУ имени Тараса Шевченко, г. Киев

ЭЛЕКТРОННЫЙ СЛОВАРЬ КАК ИССЛЕДОВАТЕЛЬСКАЯ БАЗА ДАННЫХ (на примере электронного словаря речи Тараса Шевченко)

Статья описывает электронный словарь речи Тараса Шевченко. Рассматривается структура базы данных и интерфейса, определена научная ценность проекта, осуществляемого в русле двух направлений современного языкознания: корпусной лингвистики и компьютерной лексикографии. Перечислены функциональные возможности словаря и сферы его использования.

Ключевые слова: *электронный словарь, Тарас Шевченко, компьютерная лексикография, корпусная лингвистика.*

Darchuk Natalia, Cand.Phil.Sci, Associate Professor

Langenbakh Margaryta, Cand.Phil.Sci, Assistant Professor

Institute of Philology, Taras Shevchenko National University of Kyiv

ELECTRONIC DICTIONARY AS A RESEARCH DATABASE (by the example of Electronic Taras Shevchenko Language Dictionary)

The article presents the Electronic Taras Shevchenko Language Dictionary. The structure of database and user interface are described. The authors define the scientific value

of this project which belongs to the two branches of linguistics – the corpus linguistics and the electronic lexicography. The functionality and the possibilities of usage are represented.

Key words: *electronic dictionary, Taras Shevchenko, computer lexicography, corpus linguistics.*

УДК 81'255'221

Лисюк М.В., асп.,
Інститут філології КНУ імені Тараса Шевченка, м. Київ

ЕВРИСТИЧНИЙ ПІДХІД ДО ВІДТВОРЕННЯ НЕВЕРБАЛЬНОЇ ПОВЕДІНКИ В ПЕРЕКЛАДІ: ПОШУК АДЕКВАТНОСТІ

У статті розглядається евристичний підхід до перекладу художніх творів. Зокрема, аналізується вплив перекладацької евристики на особливості відтворення нюансів невербальної поведінки персонажів.

Ключові слова: *евристика, невербальна поведінка, креативність перекладу, окаціоналізми, девіації.*

Процес перекладу за своєю природою є творчою діяльністю, яка передбачає вибір можливого варіанту вирішення проблеми із задалегідь не визначеної кількості варіантів. Тому важко не помітити конгруентності понять переклад та евристика. «Неможливість алгоритмізації процесу перекладу, що спричинена об'єктивною відсутністю точних приписів, які б однозначно призводили до заданого результату, означає свободу вибору та необхідність творчого пошуку перекладачем. У зв'язку з цим оптимальним з точки зору встановлення творчої природи перекладацьких рішень вважаємо евристичний підхід» [13:71]. Евристика – психологічний термін, покликаний пояснити здатність людини обирати під час рішення задачі тільки найбільш осмислені варіанти. Фактично, евристика – це «такий метод пошуку, який зі значною вірогідністю дозволяє відбирати найвдаліші коридори в лабіринті рішення задач» [3: 212]. Евристика – наука про виникнення нового (думок, ідей, способів дії) у знанні і діяльності людини [16:676]. В обох цих випадках вибір може бути інтуїтивним і залежатиме від творчого потенціалу, компетентності, професіоналізму та всебізності суб'єкта, зокрема, перекладача. Головне завдання як для перекладу, так і для евристики – це створення якісно нових дієвих підходів до відповідної проблеми.

Проблему евристичного підходу до перекладу розглядали як вітчизняні, так і закордонні перекладознавці (Г. Казакова, Л. Коломієць, О. Швейцер, W.Wilss). Л. Коломієць розробила теорію різновекторності перекладацької інтерпретації, де детально описала особливості евристичного вектору [5: 133-478]. Всі дослідники сходяться на думці, що особливістю цього підходу є поєднання логічного шляху до пошуку рішень