УДК 811.161.2'367.4:004

Наталія Дарчук, канд. філол. наук, доц.,
Маргарита Лангенбах, канд. філол. наук, асист.
КНУ імені Тараса Шевченка, Київ

ЕЛЕКТРОННИЙ СЛОВНИК МОВИ ТАРАСА ШЕВЧЕНКА: МЕТОДИКА І ТЕХНОЛОГІЇ УКЛАДАННЯ

Стаття присвячена теоретико-методичним аспектам Електронного словника мови Тараса Шевченка. Розглядається внутрішня структура та інтерфейс словника, лінгвістична параметризація текстового матеріалу. Оцінюються переваги електронних словників у цілому та даного проекту зокрема, перспективи його використання у лінгвістичних дослідженнях.

Ключові слова: *електронний словник, Тарас Шевченко, комп'ютерна лексикографія, лінгвістична база даних.*

Однією з характерних рис сучасної науки є активне використання комп'ютерних інструментів, які не лише оптимізують дослідницький процес, але й допомагають оформити напрацьовані результати в максимально зручному для користувачів вигляді. Лінгвістика у цьому сенсі не виняток. Зокрема, однією з галузей, тісно пов'язаних із інформаційними технологіями, є комп'ютерна лексикографія, до завдань якої належать:

– створення електронних картотек на базі розмічених текстових масивів, формування машинних фондів різної структури та призначення;

– розробка систем автоматичного конвертування паперових словників у комп'ютерний формат;

– поповнення словників в автоматичному режимі опрацювання інформації;

– укладання словопокажчиків, конкордансів, частотних словників;

– використання мультимедійних і гіпертекстових засобів представлення інформації (зображення, звукові або відео додатки, гіперпосилання);

– укладання перекладних словників на основі електронних корпусів паралельних текстів;

– укладання автоматичних словників;

– створення інтегрованих лексикографічних комп'ютерних систем тощо [Перебийніс : 9–10].

Серед здобутків українського електронного словникарства – комп'ютерний "Словник афіксальних морфем української мови", розроблений в Інституті мовознавства НАН України ім. О. О. Потебні [Карпіловська : 55]; електронні версії 11-томного тлумачного "Словника української мови" (онлайн-версію розміщено за адресою http://www.sum.in.ua) та академічного "Орфографічного словника української мови" (http://slovnyk.ua/), розроблені Українським мовно-інформаційним фондом НАН України; термінологічні словникові бази даних та спеціальна комп'ютерна система укладання словників PolyDic, створені колективом науковців університету "Львівська політехніка" [Мисак], онлайн-конкорданс творів Г. Сковороди (http://www.artsrn.ualberta.ca/skovoroda/) та ін. Колекція різноманітних словників розміщена на порталі http://mova.info [Дарчук 1996], що підтримується лабораторією комп'ютерної лінгвістики Інституту філології Київського національного університету імені Тараса Шевченка. Новітнім проектом цього колективу є Електронний словник мови Тараса Шевченка, приурочений до святкування 200-річного ювілею поета.

Як зазначає Є. Карпіловська, важлива перевага електронних словників – структурування інформації, що уможливлює роботу користувача з потрібними йому даними в зручному й гнучкому режимі доступу [Карпіловська : 73]. Л. Бєляєва також називає такі властивості електронного словника, як значне пришвидшення процесу пошуку інформації, її збалансованість, системність і ретельність добору [Беляева : 73]. Оскільки постать Тараса Шевченка винятково важлива для української філології і постійно привертає увагу науковців, цілком закономірною є потреба створення сучасного зручного інструменту для роботи з його творчою спадщиною. У цьому полягає **актуальність** і **новизна** проекту.

Метою роботи було укладення електронного словника з урахуванням сучасних лексикографічних методик і технологій для створення розгорнутого опису мови поета та оптимізації роботи дослідників із текстами його творчої спадщини.

Така мета передбачала виконання низки **завдань**, з-поміж яких:

1) лінгвістичний аналіз текстів Т. Шевченка;

2) укладання бази даних зафіксованих мовних одиниць із їхніми граматичними та кількісними характеристиками;

3) створення зручного користувацького інтерфейсу, за яким можна було б здійснювати пошук, сортувати та статистично опрацьовувати 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.

Укладання словника відбувалося в декілька етапів, на яких матеріал було піддано лінгвістичному і статистичному опрацюванню. За результатами створювалися словникові статті з багатоаспектною інформацією про лексичні одиниці, морфеми або словосполучення та частотою вживання їх у досліджуваних текстах.

Лінгвістична підготовка матеріалу для словника здійснювалася двома способами. Перший етап було проведено шляхом автоматичного морфологічного і синтаксичного аналізу. Автоматичний морфологічний аналіз був застосований для визначення частин мови та їх граматичних форм. На етапі автоматичного синтаксичного аналізу відбувалося виокремлення словосполучень і приписування їм необхідної інформації (роль словоформи у сполуці, тип синтаксичного зв'язку). На другому етапі опрацювання матеріалу здійснювалося автоматизовано, за підтримки лінгвістів, які контролювали правильність результатів автоматичного аналізу, редагували й усували виявлені помилки.

Оскільки синтаксичні і лексичні характеристики проявляються у взаємодії різної кількості одиниць, але переважно в межах речення, граничним контекстом був обраний відрізок від крапки до крапки, тобто речення. Мінімальним відрізком був контекст у три позиції 1- X +1 (де X – аналізована словоформа), тобто словоформа у препозиції та постпозиції до аналізованої словоформи.

Синтаксична інформація у словнику представлена моделями словосполучень. Опис синтаксису через словосполучення мотивується тим, що це міжрівнева лексико-морфолого-синтаксична одиниця [Головин : 190–193]. Словосполучення водночас є і **номінативною одиницею,** і елементарною **синтаксичною конструкцією.** На основі словосполучень формуються члени речення і синтаксичні конструкції, розкривається сутність міжслівних зв'язків.

Словосполученнями були визнані підрядні, сурядні та предикативні сполуки, оскільки всі вони демонструють важливу властивість: кожен із компонентів цих конструкцій – носій морфологічних ознак із релятивною функцією, тобто властивістю приєднувати до себе словоформи і приєднуватися до них, утворюючи певний синтаксичний комплекс.

Сполучуваність слів описувалася за такими параметрами:

1) тип сполуки, до якої входить аналізована словоформа (за частиномовною належністю головного члена): іменникова, прикметникова, дієслівна, прислівникова, числівникова, займенникова.

2) роль словоформи у сполуці (тільки для підрядних конструкцій, оскільки в сурядних та у випадку координації обидва члени є рівноправними):

а) ядрова (словосполучення, в яких аналізована словоформа є головним);

б) ад'юнктна (словосполучення з аналізованою одиницею в ролі залежного члена).

3) тип синтаксичного зв'язку: підрядний, сурядний, координація.

Статистичні характеристики одиниць словника подані у вигляді двох параметрів – абсолютної та середньої частоти. Самої лише абсолютної кількості прикладів може бути недостатньо для відтворення об'єктивної картини, оскільки слова не завжди рівномірно розподілені по текстах. Для врахування такого явища вводиться поняття середньої частоти, а також середнє квадратичне відхилення абсолютної частоти – параметр, який характеризує рівномірність розподілу слів у текстах.

Уся інформація заносилася до бази даних (рис. 1), яка має таку структуру: колонки з номерами слова, речення й тексту; словосполучення; кожна зі словоформ, що його складають;

їх початкові форми (леми), граматичні класи словоформ-компонентів словосполучення; тип синтаксичного зв'язку. Укладання лінгвістичної бази даних сприяє систематизації і чіткому структуруванню інформації, а також суттєво спрощує процес поповнення та редагування матеріалу [Лангенбах]. До того ж, виконання різноманітних пошукових завдань вимагає втручання лінгвіста-дослідника до представлення інформації, унаочнення відношень між різними її типами (виведення, включення, виключення, заміна, суміщення тощо), що уможливлюється застосуванням лінгвістичних баз даних [Карпіловська : 36]. Отже, бази даних є формалізованими моделями мовних об'єктів і створюють фактографічне підґрунтя для розв'язання багатьох лінгвістичних задач [Дарчук 2008 : 205].

Інтерфейс словника складається з чотирьох сторінок. Перша сторінка містить інформацію про проект, решта подає різноманітну лінгвістичну інформацію. Структура сторінок організована за єдиним принципом (рис. 2): основне поле пропонує лінгвістичні параметри добору інформації, пошук контекстів за словоформами або лексемами, морфемною моделлю слова, частиномовною належністю, категорійними характеристиками (рід, число, відмінок, час, особа тощо), синтаксичною моделлю словосполучення). У результатах виводяться знайдені одиниці з кількісними даними і контекстами вживання. Ліва бічна панель сторінки містить меню зі статистичними параметрами, а також посилання на інші сторінки словника.

Параметри опису одиниць словника відкривають широке поле для лінгвістичного аналізу творчості Т. Шевченка. Так, наприклад, дослідження моделей сполучуваності дає відповідь на такі питання:

– чи можливі випадки, коли словоформа певної частини мови у реченні не має залежних від неї елементів, або, навпаки, нічому не підпорядкована;

– словами яких частин мови може керувати слово (частина мови);

– до яких комплексів може входити і в якій ролі;

– якими комплексами може керувати;

– чи може дане слово утворювати предикативні сполуки, і якщо так, то з якими класами слів тощо.

id	textID	Sent	inwrd	incls	slvospoluch	wrd2	cls2	model	predl	predli	sz	insubcl	inwform	subcls2	wform2
374	10301	10	запроданий	А	Запродана жидам	жид	Й	А И			АС	Ж	Запродана	О	жидам
3472	10377	24	критий	А	криту Китайкою	китайка	К	А И			АС	Л	криту	Т	Китайкою
6459	10533	4	слiпий	А	слiпих Невольникiв	невольник	Й	А И			ІС	Е	слiпих	У	Невольникiв
6455	10533	3	святий	А	Святая сило	сила	К	А И			ІС	Ж	Святая	К	сило
6299	10523	18	розп'ятий	А	розп'ятий Син	син	Й	А И			ІС	И	розп'ятий	И	Син
8263	14794	23	вкритий	А	вкрита Свитиною	свитина	К	А И			АС	Ж	вкрита	Т	Свитиною
5847	10503	24	одринутий	А	одринутий людьми	люди	И	А И			АС	И	одринутий	Ю	людьми
2752	10351	3	повитий	А	Повит красою	краса	К	А И			АС	А	Повитi	Т	красою
240	10295	20	винний	А	винна голубка	голубка	К	А И			ІСЗ2	Ж	винна	И	голубка
1989	10326	4	малий	А	Малих дiток	дiтки	И	А И			ІСЗ4	Е	Малих	У	дiток
1997	10326	10	запроданий	А	запродані Жидові	жид	Й	А И			АС	А	запродані	Д	Жидові
2746	10351	2	розбитий	А	не розбите Серце	серце	Л	А И			КЗ	С	розбите	И	Серце
2698	10348	37	чужий	А	чужому краю	край	Й	А И			ПП1	Ш	чужому	П	краю
7747	10571	3	творчий	А	Творче неба	небо	Л	А И			ІСЗ3	С	Творче	Р	неба
7807	10578	2	политий	А	политая кровiю	кров	К	А И			АС	Ж	политая	Т	кровiю
2241	10333	86	темний	А	темним гаї	гай	Й	А И			ПП1	Ш	темним	П	гаї
1091	10311	44	пiдбитий	А	пiдбите Брехнею	брехня	К	А И			АС	С	пiдбите	Т	Брехнею
8374	17965	293	битий	А	Битий горем	горе	Л	А И			АС	И	Битий	Т	горем
1694	10325	31	завзятий	А	завзятих слав'ян	слав'янин	Й	А И			ІСЗ4	Е	завзятих	У	слав'ян
2337	10339	1	давнiй	А	давнi лiта	лiта	И	А И			ІСЗ4	У	давнi	А	лiта
7746	10571	3	добросердий-м	А	Добросердим-малим Тихолюбц	тихолюбець-св	Й	А И			СУ	О	Добросердим-г	О	Тихолюбцям-сі
2367	10340	4	неситий	А	неситiї Гріхами	гріх	Й	А И			ІСЗ4	А	неситiї	Ю	Гріхами
2380	10342	1	земний	А	земних владик	владика	Й	А И			ІСЗ4	Е	земних	У	владик
8043	14791	9	писанi	А	писані Катами	кат	Й	А И			АС	А	писані	Ю	Катами
8052	14791	26	бiдний	А	бiдних покритка	покритка	К	А И			ІСЗ4	Е	бiдних	У	покриток
963	10310	16	жонатий	А	жонатими парубками	парубок	Й	А И			СУ	Ю	жонатими	Ю	парубками
953	10310	11	високий	А	висока станом	стан	Й	А И			АС	Ж	висока	Т	станом
2433	10345	5	покинутий	А	покинутий Рабом	раб	Й	А И			АС	И	покинутий	Т	Рабом
2449	10346	3	славних	А	славних окова	окова	К	А И			ІСЗ4	Е	славних	Ю	оковами
870	10309	72	молодий	А	молодих дівчат	дівчина	К	А И			ІСЗ4	Е	молодих	У	дівчат
2566	10348	16	кradений	А	крадене Дідами	дід	Й	А И			АС	Х	крадене	Ю	Дідами
8039	14791	7	просвiщенний	А	просвiщенних християн	християнин	Й	А И			ПП1	Е	просвiщенних	У	християн
5556	10492	28	мережанний	А	мережанi сльозами	сльоза	К	А И			ІСЗ4	А	мережанi	Ю	сльозами
2320	10337	2	вкритий	А	вкрите Плодом	плід	Й	А И			АС	С	вкрите	Т	Плодом
2633	10348	69	напоєний	А	напоєна кров'ю	кров	К	А И			ДС	Ж	напоєна	Т	кров'ю
5633	10498	7	безталанний	А	безталанне байстря	байстря	Л	А И			ІС	С	безталанне	И	байстря
2792	10354	9	нехрещений	А	Нехрещений сину	син	Й	А И			ІС	И	Нехрещений	К	сину
7655	10568	5	убогий	А	Убогих серцем	серце	Л	А И			АС	Е	Убогих	Т	серцем
2754	10351	3	умитий	А	Умиті сльозою	сльоза	К	А И			АС	А	Умиті	Т	сльозою
1837	10324	2	повний	А	повен води	вода	К	А И			АС	И	повен	Р	води

Рис. 1. Структура бази даних

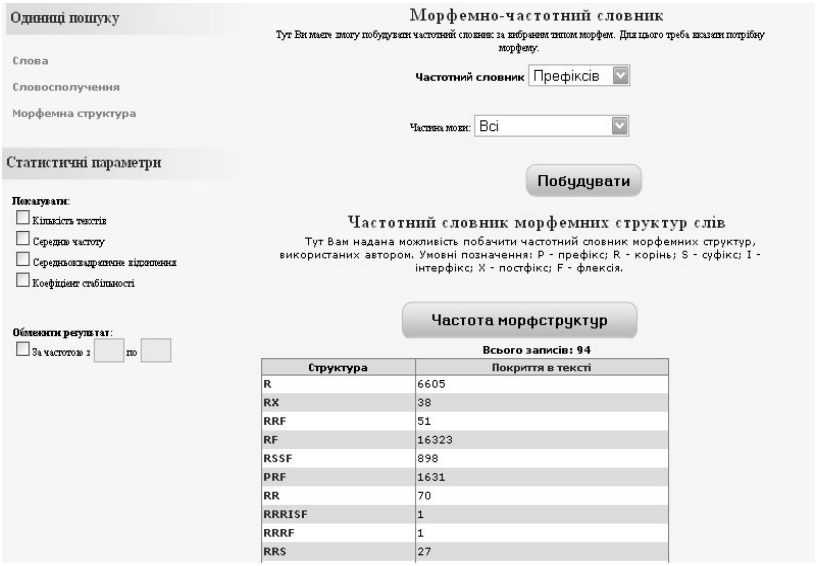


Рис. 2. Сторінка пошуку

В якості ілюстрації наведемо відомості про отриманий у результаті дослідження синтаксису творів Т. Шевченка перелік частиномовних моделей сполучуваності:

I. Моделі дієслова

1. Ядрові моделі дієслова

– дієслово + іменник: ***благословить*** *дітей;*

– дієслово + прикметник: ***ставати*** *зеленим;*

– дієслово + прийменник + іменник: ***дивлюсь*** *на тебе;*

– дієслово + дієслово: *жити* ***хочу****.*

2. Ад'юнктні моделі дієслова

– дієслово + дієслово: *ліг* ***одпочить;***

3. Предикативні моделі дієслова

У цьому виді зв'язку дієслово є присудком або підметом:

– іменник + дієслово: *Вітер віє.*

4. Сурядні моделі дієслова

– дієслово + дієслово: *гралися, хвалили.*

II. Моделі іменників

1. Ядрові зв'язки іменників

– прикметник + іменник: вольнії ***села***;

– займенниковий прикметник + іменник: *мій* ***квіте***;

– іменник + іменник: ***день*** радості;

2. Предикативні зв'язки іменника

– іменник + дієслово: *верба похилилась.*

III. Моделі прикметників

1. Ядрові моделі прикметників

– прислівник + прикметник: *дуже* ***цікаве***;

– прикметник + прийменник + іменник: ***великая*** *в женах*;

– прикметник + сполучник + іменник: ***червоних*** *як калина.*

2. Ад'юнктні моделі прикметників

– дієслово + прикметник: ***був*** *дужий*;

– прикметник + іменник: *вольнії* ***села***.

3. Предикативні моделі прикметників

– іменник + прикметник: ***ангелом*** *святим.*

4. Сурядні зв'язки прикметників

– прикметник + сполучник + прикметник*: зелений і синій*.

IV. Моделі прислівників

1. Ядрові зв'язки прислівника. Ядровим вважається прислівник, якщо він не може бути опущений без порушення змісту, тобто є лексично зв'язаним

– іменник + прислівник: *гріха* ***менше***;

– прислівник + прислівник: *досі* ***нудно***;

– прислівник + прийменник + займенниковий іменник: ***тихо*** *в мене.*

2. Ад'юнктні зв'язки прислівників

– прислівник + прислівник: ***досі*** *нудно;*

– прислівник + прикметник: *надто* ***молодую***.

3. Предикативні зв'язки прислівників

– іменник + прислівник: *билини кругом.*

4. Сурядні зв'язки прислівників

– прислівник + сполучник + прислівник: *любенько та тихо.*

Коло потенційних досліджень із застосуванням словника охоплює морфологічні, лексичні, синтаксичні й стилістичні розвідки. Цьому сприяє різноманітність лінгвістичної розмітки текстів та гнучкість інтерфейсу користувача, що свідчить про високу ефективність електронних словників і, зокрема, словника мови Тараса Шевченка.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. *Беляева Л.* Потенциал автоматизированной лексикографии и прикладная лингвистика / Л. Беляева // Известия РГПУ им. А. И. Герцена. – № 134. – СПб, 2010. – С. 70–79.
2. *Головин Б.* Введение в языкознание / Б. Головин. – М.: Высшая школа, 1966. – 328 с.
3. *Дарчук Н.* Комп'ютерна лінгвістика (автоматичне опрацювання тексту): підручник / Н. Дарчук. – К.: Видавничо-поліграфічний центр "Київський університет". – 351 с.
4. *Дарчук Н., Грязнухіна Т.* Частотний словник сучасної української публіцистики / Н. Дарчук, Т. Грязнухіна // Мовознавство. – 1996. – № 4–5. – С. 15–18.
5. *Карпіловська Є.* Вступ до прикладної лінгвістики**:** комп'ютерна лінгвістика: Підручник / Є. Карпіловська – Донецьк: ТОВ "Юго-Восток, Лтд", 2006. –188 с.
6. *Лангенбах М.* Електронна база моделей семантико-синтаксичної сполучуваності іменників української мови / М. Лангенбах. – Науковий вісник Волинського національного університету імені Лесі Українки. – Луцьк: Вежа, 2008. – с. 249–252.
7. *Мисак Р.* Комп'ютерні словники: класифікація та укладання / Р. Мисак // Збірник наукових праць: "Проблеми української термінології". – Львів, 2008. – С. 52–55.
8. *Перебийніс В., Сорокін В.* Традиційна та комп'ютерна лексикографія / В. Перебийніс, В. Сорокін. – К.: Видавничий центр КНЛУ, 2009. – 218 с.

Стаття надійшла до редакції 10.03.13

Дарчук Наталия, канд. филол. наук, доц.,
Лангенбах Маргарита, канд. филол. наук, ассист.
КНУ имени Тараса Шевченко, Киев

Электронный словарь языка Тараса Шевченко: методика и технологии создания

Статья посвящена теоретико-методическим аспектам Электронного словаря языка Тараса Шевченко. Рассматривается внутренняя структура и интерфейс словаря, лингвистическая параметризация текстового материала. Оцениваются преимущества электронных словарей в целом и данного проекта в частности, перспективы его использования в лингвистических исследованиях.

Ключевые слова: *электронный словарь, Тарас Шевченко, компьютерная лексикография, лингвистическая база данных.*

Darchuk Nataliia, Ph D, Doc.,
Langenbakh Margaryta, Ph D, Assistant Professor
Taras Shevchenko National University of Kyiv, Kyiv

The Electronic Taras Shevchenko Language Dictionary: methodics and technologies of creation

The article is dedicated to the theoretic and methodic aspects of the Electronic Taras Shevchenko Language Dictionary. The internal structure, interface and

linguistic parameters of dictionary is described. The article shows the advantages of electronic dictionaries, especially the Electronic Taras Shevchenko dictionary, and prospects of their usage in linguistic researches.

Keywords: *electronic dictionary, Taras Shevchenko, computer lexicography, linguistic database.*

УДК 811.161.2'3774

Анатолій Мойсієнко, д-р філол. наук, проф.
КНУ імені Тараса Шевченка, Київ

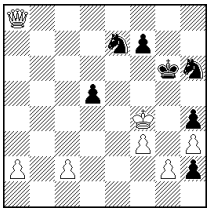
МАТЕРІАЛИ ДО ШАХОВОГО СЛОВНИКА УКРАЇНСЬКОЇ МОВИ
(задачні та етюдні теми українських авторів)

Вперше друкується українською мовою фрагмент укладеного автором шахового словника.

АКТИВІЗА́ЦІЇ ТЕ́МА. В етюдній композиції: біла фігура, пасивна в початковій позиції, чи така, що в процесі гри потрапляє під зв'язку, виявляється замкненою тощо, за певних умов набуває активності, визначає результат гри. Термін запропоновано на початку сорокових років XX ст. дніпропетровським шах. композитором П. Бондаренком (1905–1993).

Бондаренко П., Кузнецов О. "Шахматы в СССР", 1964 (почесна відзнака).

+ (7+7)



1. Фа6+ f6 2. Фf1 Кеf5 3. Фh1 Кg3 4. Ф:h2 Крf7, і білі король і ферзь практично виявляються замкненими, поля e3, g1 небезпечні через коневу вилку. Однак після 5. a4 Кре7 6. a5 Крd7 7. a6 Крс6 8. c3 у чорних цуґцванґ, вони змушені віддати пішака **d** (8…d4 9. cd), і хоч обидва білі пішаки на лівому фланзі стають легкою здобиччю чорного короля, білі король і ферзь отримують простір для активної переможної гри.