

- Запропонована методика опрацювання лінгвістичних даних в електронній базі є узагальненням комплексу теоретичних і прикладних ідей сучасного мовознавства. Технологія конструювання бази робить її надзвичайно ефективним та раціональним інструментом (вона зберігає багато часу та людських ресурсів) для спеціалістів-філологів різного профілю.
 - Електронна база даних зі своєю методологією і технологією допомагає ефективно й оперативно здійснювати масштабні комплексні філологічні дослідження на рівні сучасної наукометрії.
 - Ця база даних використовується у навчальному процесі на філологічних факультетах університетів, забезпечуючи необхідний теоретичний і технологічний рівень підготовки фахівців.
 - Створена нами база даних починає використовуватися як готовий продукт і як модель для конструювання аналогічних баз університетами України.
- Література**
1. Апресян Ю. Д. О толковом словаре управления и сочетаемости русского глагола// Словарь. Грамматика. Текст. – М., – 1998. С.13.
 2. Караулов Ю. Н. Лингвистическое конструирование и тезаурус литературного языка. М., 1981. С.51.

О. Зубань*

Київський національний університет ім. Т. Шевченка (Київ)
УДК 81'33

ПАРАМЕТРИЗОВАНА БАЗА ДАНИХ ЯК ІНСТРУМЕНТ ДОСЛІДЖЕННЯ КОРПУСУ ТЕКСТІВ

Розвиток теорії і практики прикладної лінгвістики, зокрема методів комп'ютерного моделювання, дозволили по-новому сформулювати лексикографічне завдання, а саме – створення автоматизованої системи лінгвістичного аналізу тексту. Структура цієї системи базується на модульно-словниковій ідеології: системно-структурні відношення одиниць кожного рівня мовної системи представлені в окремому модулі: фонетичному, морфемному, словотвірному, морфологічному, синтаксичному, семантичному.

У традиційній лінгвістиці наукова значущість, вагомість результатів дослідження визначається перш за все репрезентабельністю матеріалу дослідження: чим більше мовних фактів, тим достовірніші спостережувані закономірності, але традиційна форма збирання і систематизації інформації (переважно паперова картотека та використання різноманітних паперових словників) сьогодні не задовольняє потреб дослідників-філологів. Потрібні нові інформаційні технології, які б оптимізували роботу дослідника. Тому в українському мовознавстві на сьогодні нагальною є проблема укладання електронних лінгвістичних словників, які мають формат параметризованих електронних баз даних, оснащених пошуково-класифікаційними програмними аналізаторами, що забезпечують:

- ефективне та оперативне проведення лінгвістичного аналізу;
- можливість аналізу великих лексичних масивів;
- отримання точних формальних характеристик мовних одиниць різних рівнів як у системі мови, так і реляційно-функціональних особливостей їх у тексті.

У сучасній комп'ютерній лінгвістиці досить плідно розвивається лексикографічний напрямок. Зокрема в українському мовознавстві сьогодні створено чимало різногалузевих електронних словників, з різноманітними системами навігації та інтерфейсами, частина з яких стала вже лінгвістичним комерційним продуктом і доступна для широкого кола користувачів. Такі бази даних покликані виконувати функцію своєрідних довідників для лінгвіста-дослідника і, без сумніву, є надзвичайно важливими для організації повномасштабного дослідження мови, але вони є **статичними**, їх не можна використовувати в режимі автоматизованого аналізу тексту. Тому особливої уваги сьогодні заслуговують ті електронні лінгвістичні продукти, які спрямовані на аналіз тексту і мають статус **динамічних** пошукових систем, які здатні в автоматичному або автоматизованому режимі вилучати інформацію про мовні одиниці з будь-якого параметризованого тексту.

Текст як основна форма збереження і передачі інформації, будучи результатом мовленнєвого акту, є інвентарем мовних одиниць, які комбінуються в ньому за законами

* © О. Зубань, 2006

мовної норми. А оскільки в мові немає нічого, чого б не було в мовленні, то текст водночас і реалізує, і формує систему мови. Тобто, правомірним і перспективним є опосередкований шлях вивчення мовних явищ через текст, оскільки аналізуються всі одиниці мовних рівнів як цілісна реальність у спостереженні.

В останні десятиріччя фахівцями різних сфер, дослідження яких дотичні з комп'ютерним аналізом текстів, гостро відчувається необхідність у точних функціональних характеристиках мовних одиниць у різних типах текстів. Таку лінгвостатистичну інформацію можна отримати за допомогою електронної бази даних, досліджувальні функціональні можливості якої забезпечуються взаємопов'язаною роботою різних модулів – фонетичного, морфемного, словотвірного, морфологічного, синтаксичного, семантичного.

Працюючи над проблемами конструювання параметризованих лінгвістичних баз даних, викладачі Інституту філології Київського національного університету імені Тараса Шевченка розробили методику формалізованого опису мовних одиниць різних рівнів, на основі якої створили комп'ютерні інструменти – пакети програм, що забезпечують вилучення інформації із тексту і створення різноманітних словників: частотних, граматичних, морфних структур, кореневих, афіксальних, словників кореневих гнізд, синтаксичних структур словосполучення та речення, різнотипних електронних карток, спрямованих у різні галузі лінгвістичних досліджень.

Розвиток теорії і практики прикладної лінгвістики, зокрема методів комп'ютерного моделювання, дозволили по-новому сформулювати лексикографічне завдання, а саме – *створення автоматизованої системи лінгвістичного аналізу тексту*.

Структура цієї системи базується на модульно-словниковій ідеології: системно-структурні відношення одиниць кожного рівня мовної системи представлені в окремому модулі: фонетичному, морфемному, словотвірному, морфологічному, синтаксичному, семантичному.

Робочою одиницею кожного модуля є слово (вихідна форма), що ґрунтується на лінгвістичному постулаті – слово – центральна одиниця мови, що на різних рівнях мовної системи характеризується різними типами структур, або ж як структурна одиниця входить до одиниць вищих мовних рівнів – словосполучення та речення; а також зумовлено текстовою спрямованістю створюваної автоматизованої системи аналізу, оскільки в графемному представленні будь-якого тексту визначається дискретна одиниця від пробілу до пробілу – слово. Отже, базовою одиницею кожного модуля виступає слово (вихідна форма) як формальна одиниця: задана послідовність графем, що через посередництво морфологічного модуля має постійний зв'язок із текстовою словоформою – послідовністю графем між пробілами тексту. У кожному модулі ця робоча одиниця формалізується через представлення її у термінах програмних процедур, що дозволяє описати різноманітні лінгвістичні структури слова (фонетичну, морфемну, словотвірну, морфологічну), словосполучення, речення. Формалізація опису цих структур здійснюється за допомогою єдиної метамови, що забезпечує зв'язок слова із структурами слів в усіх модулях, і зокрема з морфологічним модулем, який дозволяє ідентифікувати кожен словоформу тексту із її словниковою формою (вихідною) у кожному модулі.

Кожен модуль складається з двох блоків: 1) блок-словник: інвентар слів кожного модуля формується на основі 11-томного тлумачного словника української мови і в процесі роботи доповнюється новою лексикою (іншомовними словами та термінами різних галузей науки), загалом на сьогодні словникова база системи автоматизованого лінгвістичного аналізу складає ≈ 170 тис. слів української мови; 2) блок-аналізатор, що розглядається як надійний лінгвістичний інструмент мовознавчих досліджень, здатний працювати в режимах пошуку, класифікації, лінгвістичного аналізу на великих лексичних словникових та текстових масивах.

Інформаційний макет словника кожного модуля та його ідеологія визначалися специфікою електронного характеру та перспективою створення на їх основі автоматизованої системи лінгвістичного аналізу тексту, тому лінгвістичні аналізатори кожного модуля дозволяють в автоматизованому режимі приписувати слову формалізовану інформацію про його структуру залежно від рівня аналізу, а через ідентифікацію словоформи тексту із словом модуля отримати автоматичний аналіз словоформи.

Кожен модуль автоматизованої системи лінгвістичного аналізу спрямований на вирішення своїх завдань, що зумовлює структуру його словника, формалізацію запису структури слова та програмування аналізу.

Фонетичний модуль (на сьогодні розроблений лише проект модуля). У цьому модулі ставиться завдання представити графемний запис слів у двох типах структур: фонетичній та фонемній. Визначення цих структур забезпечується двома типами алгоритмів, оскільки процеси чергування та асиміляції чітко визначаються фонетичною позицією, що дозволяє їх алгоритмічно описати, наголос як необхідний знак транскрипції приписується лінгвістом. Одиницею словників (фонетичного та фонемного) у цьому модулі є слово у фонетичному та фонемному записі. Автоматизований фонетичний аналіз забезпечує аналізатор, який дозволяє отримувати інформацію про:

- фонемну будову слова/словоформи;
- кількість голосних: за характеристикою місця творення, ступеня підняття, лабіалізації, наголосу;
- кількість приголосних: за класифікаційними ознаками сонорна, шумна (дзвінка, глуха), губна, язикова (передньоязикова, середньоязикова, задньоязикова), глоткова, зімкнена, щілинна, африката, свистяча, шипляча, носова;
- фонемні моделі слів/словоформ, що дозволить розробити алгоритм складоподілу;
- складову структуру слова/словоформи: кількість складів (закритих, відкритих), їх моделі;
- коефіцієнт консонантності та вокалізму тексту.

Морфемний модуль. Словник цього модуля формують слова із визначеними морфемними структурами. Формалізація опису морфемних структур слів полягає у моделюванні структурно-функціональних зв'язків морфем у слові: представлення кожної морфемної структури слова у вигляді лінгвістичної моделі, яка визначає межі і тип кожного морфа і дозволяє автоматично описати кожну морфемну структуру через програмну процедуру, наприклад, *заледеніти* / P2R5I7S8F10. Автоматично сформована програмна процедура моделі морфемної структури, пов'язана з формою вираження кожного слова морфемної бази даних, дає повну лінгвістичну інформацію про морф, його структурні відношення з іншими морфемами і визначається як робоча одиниця морфемної бази даних.

Укладання морфемного словника та використання морфемного модуля в режимі лінгвістичного аналізу забезпечується зручним автоматизованим інструментом, який дозволяє доповнювати морфемну базу новою лексикою.

Формалізація морфемних структур слів через їх опис у термінах програмних процедур МБД дозволила створити на основі цієї бази даних автоматизовану систему лінгвістичного аналізу, здатну виконувати цілий ряд завдань в автоматичному режимі:

- групувати лексику у спільнокореневі та спільноафіксальні класи;
- класифікувати лексику за кількісно-морфемними моделями;
- створювати алфавітно-частотні кореневі та афіксальні словники;
- проводити морфемний аналіз вихідних словоформ;
- проводити морфемну сегментацію словоформ тексту.

Робота морфемного модуля із словоформами тексту забезпечується не тільки словником морфемного модуля, а й програмою автоматичного морфемного сегментування словоформ тексту, якою був доповнений модуль у процесі дослідження корпусу текстів поетичного стилю.

Формалізація опису морфемних структур у морфемному модулі дозволяє швидко і ефективно створювати нові лексикографічні лінгвістичні описи, що спрямовані на вирішення різноманітних проблем у галузі морфеміки. Сьогодні ведеться робота над створенням словника морфем, що репрезентує морфему як знакову парадигматичну одиницю з приписаним квантом значення, реалізовану в морфах у визначених дистрибуціях.

Словотвірний модуль. Укладання електронного словотвірного словника здійснюється за допомогою аналізатора морфемного модуля системи, що застосовується для виконання двох завдань: 1) автоматизованого групування лексики у спільнокореневі вибірки; 2) автоматизованої класифікації лексики у межах кожного спільнокореневого класу за кількісно-морфемними моделями.

Кожна вибірка спільнокореневої лексики формує у межах словотвірного модуля окреме поле, в якому морфемні структури слів класифікуються за кількісно-морфемними

моделями. За принципами словотвірної похідності було розроблено формалізовані методики опису словотвірних зв'язків між мотивуючим та мотивованим словами, що дозволяє автоматично побудувати робочу гіпотезу-модель словотвірного гнізда.

Класифікація слів спільнокореневої вибірки за словотвірними тактами є лише лінгвістичною гіпотезою, яка вимагає перевірки, що здійснюється в автоматизованому режимі: 1) перевірка правильності групування слів у спільнокореневої вибірки та розташування їх за словотвірними тактами; 2) визначення у кожному похідному слові словотвірної основи та словотвірного форманта; 3) додавання інформації про морфологічні процеси, що відбуваються на кожному такті словотвору.

Словотвірний модуль дозволяє аналізувати текстові слововживання за такими параметрами:

- спосіб творення слова та його словотвірна структура;
- місце слова у словотвірному гнізді;
- класифікація лексики за ознаками способу творення, словотвірними значеннями та словотвірними категоріями,

а також:

- моделювати потенційно можливі слова української мови;
- визначати неологізми та okazionalіzми в тексті;
- доповнювати електронний словотвірний словник новою лексикою.

Морфологічний модуль. Словник цього модуля представляє вихідні форми слів, яким приписано код граматичної інформації: частина мови, нерегулярні граматичні значення (наприклад, для іменника – рід, для дієслова – вид), код парадигматичного класу, за допомогою якого словоформи приписуються регулярні граматичні значення (наприклад, для іменника – відмінок, для дієслова – час, особа).

Морфологічний модуль у системі автоматизованого аналізу тексту є стрижневим, тому що всі інші модулі працюють із текстовою словоформою через ідентифікацію її з граматичним кодом морфологічного модуля. Аналізатор морфологічного модуля в автоматизованому режимі (із зняттям омонімії) здійснює параметризацію корпусу текстів.

Морфологічний аналізатор дозволяє:

- проводити автоматичний морфологічний аналіз словоформ тексту;
- групувати текстові слововживання за класифікаційними ознаками:
 - частина мови;
 - у межах іменних частин мови: рід, число, відмінок, лексико-граматичний розряд;
 - у межах дієслова: вид, спосіб, час, перехідність/неперехідність, стан, особа, число / рід, число;
 - у межах прислівника: лексико-граматичні розряди;
 - у межах прийменника: участь у керуванні відмінками іменників;
 - у межах сполучника: розряди за функцією та значенням;
 - у межах частки: розряди за функцією;
 - у межах вигука: розряди за значенням.
- створювати словник-конкорданс визначених словоформ або словоформ визначеного граматичного значення.
- отримувати статистичні дані за різними параметрами граматичної інформації словоформ.

Синтаксичний модуль. Цей модуль не має свого електронного словника, тому його аналізатор має тільки текстову ідеологію і дозволяє описувати синтаксичні структури речень, використовуючи морфологічну базу даних та автоматизований інструмент побудови дерева залежностей, що репрезентує формальну синтаксичну структуру речення через моделювання зв'язків між словоформами та визначення їхнього виду: підрядний (керування, прилягання, узгодження), сурядний та координація.

Застосування цього модуля в автоматизованому аналізі тексту дозволяє:

- визначати кількісні параметри та структуру надфразових єдностей: кількість абзаців, простих та складних речень в абзаці; словосполучень в абзаці та реченні, довжину речення у слововживаннях;
- визначати функціональні параметри: структуру речення через модель дерева залежностей, характер підрядного зв'язку словосполучення (керування,

прилягання, узгодження), морфологічну модель словосполучення у межах кожного типу підрядного зв'язку;

- на базі визначених моделей укладати словники структур словосполучень та речень з текстовою ілюстрацією цих структур.

Семантичний модуль. Цей модуль будується на словниках двох типів:

- Електронного словника синонімічних рядів;
- Словника-тезауруса.

Електронний словник синонімічних рядів, який укладено на базі лексики “Словника синонімів української мови” (Т. 1 – 1999, Т. 2 – 2000), систематизує слова за їх лексичними значеннями і групує лексику у 12 тис. синонімічних рядів, що покривають 50 тис. слів української мови;

Семантичний аналізатор цього словника дозволяє використовувати його в процесі створення словника-тезауруса.

Словник-тезаурус систематизує лексику не за лексичним значенням, а за ідеографічним принципом представлення поняттєвої спорідненості лексики – від концепту (поняття) до лексичних засобів його вираження, що дозволяє групувати семантично близькі лексичні одиниці, які з певними змістовими варіаціями передають одне поняття, у лексико-семантичні групи. Ідеографічна методика вибрана не випадково, тому що саме така класифікація дозволяє:

- формалізувати опис структурних відношень лексичної системи мови, а саме найтипівіші семантичні зв'язки у лексиці – синонімічні, антонімічні та полісемічні, через визначення місця кожної лексико-семантичної групи, а отже, кожної лексеми, у межах ідеографічної класифікації;
- отримати формалізований опис семантики слова через суму поняттєвих концептів;

Лінгвістичний аналізатор цього модуля дозволяє в автоматизованому режимі заповнювати синонімічну карту словника із врахуванням тлумачення слів за 11-томним тлумачним словником.

В укладанні обох типів словників семантичного модуля використовуються словники морфемного і словотвірного модулів, які дозволяють глибше зрозуміти і правильно описати семантику слова.

У результаті роботи семантичного модуля лінгвіст може отримувати інформацію про:

- семантику слова за логіко-концептуальною схемою синонімічної карти;
- класифікувати слова/словоформи за лексико-семантичними групами;
- визначати синонімічні, антонімічні та полісемічні слова в тексті;
- визначати та описувати різноманітні переносні значення слів, тропіку художніх текстів.

- укладати словники поетичних образів.

Створені різногалузеві лінгвістичні бази даних були використані у дослідженні корпусу текстів, до якого ввійшли тексти поетичного, публіцистичного та наукового стилів обсягом майже 2 млн. словоформ.

Корпус текстів був об'єднаний із модульною системою лінгвістичного аналізу, в результаті чого створюється параметризована база даних, що структурується на:

1. **Модуль-текст**, який репрезентує корпус текстів;

2. **Модуль-словник**, в якому систематизується різноманітна лінгвістична інформація в результаті аналізу текстів;

3. **Модуль-аналізатор** – інструмент лінгвістичних досліджень великих текстових масивів, що включає пакети програм, які можуть виконувати такі функції:

- забезпечують зв'язок корпусу текстів з лінгвістичними базами даних: морфологічною, морфемно-словотвірною, синтаксичною, семантичною;
- забезпечують роботу в автоматичному режимі пошуку і класифікації лексики за різними параметрами, а також морфологічного, морфемного, статистичного, синтаксичного та семантичного аналізів;
- формують словник-конкорданс мінімальних контекстових слововживань.

У логіці конструювання параметризованої бази даних корпус текстів є джерелом автоматизованого укладання різноманітних словників (модуль-словник) за допомогою модуля-аналізатора. У результаті лінгвістичних досліджень за допомогою параметризованої лінгвістичної бази були укладені такі словники:

1. **Інтегрований алфавітно-частотний словник поетичного мовлення** обсягом 31.000 слів і 69.000 словоформ з інформацією про слово: граматичною, лексико-граматичною і статистичною (середня частота словоформи (x_{ser}), середнє квадратичне відхилення середньої частоти словоформи (σ), середня частота лексики ($x_{serbases}$), середнє квадратичне відхилення середньої частоти лексики (σ_{bases}), абсолютна частота і коефіцієнт стабільності);

Рисунок 1. Фрагмент словника для дієслова “брати”

Лексема	Словоформа	Гр. іпф.	x_{ser}	σ	$x_{serbase}$	σ_{base}	abs	Abs-base	de	De-base
Брати	Брати	Інф	0,02	0,14	0,33	0,6385661	6	99	5	8
Брати	Беру	1 ос.т.	5,333333E-02	0,2654974	0,33	0,6385661	16	99	7	8
Брати	Береш	2 ос.т.	2,333333E-02	0,1716262	0,33	0,6385661	7	99	5	8
Брати	Бере	3 ос.т.	0,07	0,2800595	0,33	0,6385661	21	99	7	8
Брати	Берем	Іос.мн.т.	6,666667E-03	8,137704E-02	0,33	0,6385661	2	99	2	8
Брати	Беремо	Іос.мн.т.	3,333333E-03	5,763872E-02	0,33	0,6385661	1	99	0	8
Брати	Беруть	Зос.мн.т.	0,03	0,2215852	0,33	0,6385661	9	99	5	8
Брати	Брав	Чр.мін.	2,333333E-02	0,1716262	0,33	0,6385661	7	99	5	8
Брати	Брала	Жр.мін.	0,01	9,949874E-02	0,33	0,6385661	3	99	4	8
Брати	Брала	Мн.мін.	0,02	0,14	0,33	0,6385661	6	99	5	8
Брати	Бери	Нак.одн.	4,666667E-02	0,2404625	0,33	0,6385661	14	99	7	8
Брати	Беріть	Накмн	2,333333E-02	0,1509599	0,33	0,6385661	7	99	6	8

2. **Словник поета**, в якому лексика, параметризована і систематизована в таких словниках: *Алфавітно-частотний словник слів і слововживань; Морфемні словники різного лінгвістичного спрямування; Лексичні словники різного лінгвістичного спрямування; Словники тропів.*

Ці словники формують: окрему базу даних для кожного з 15 поетів поетичного стилю корпусу текстів (вибірка 20 тис. совоживань, загальний обсяг корпусу поетичних текстів 300.000 слововживань), а також зіставно-порівняльну базу спільної лексики поетів (105 списків для кожної пари поетів),

Укладання алфавітно-частотного словника слів і слововживань та морфемних словників із частотними характеристиками для кожного морфа (морфеми для коренів) забезпечується пакетом алгоритмів і програм автоматичного морфемного і морфологічного аналізу, а також пакетом програм автоматичної статистичної обробки лінгвістичних даних. Укладання алфавітно-частотного словника здійснюється у двох режимах: а) автоматичному – без зняття омонімії; б) автоматизованому – за умови зняття лексичної, а в деяких випадках лексико-граматичної омонімії.

Алфавітно-частотний словник представляє точні характеристики частоти, стабільності вживання слова/словоформи, його лексико-граматичні характеристики. Укладання за цими параметрами лексичних реєстрів, спільних для двох авторів, дозволяє здійснювати порівняльну характеристику лексики, а також досліджувати закономірності організації ідіостилу кожного поета, зокрема аналізувати в автоматизованому режимі такі групи слів, як неологізми, архаїзми, діалектизми – ця інформація систематизується у відповідних словниках.

Укладання морфемних словників здійснювалось після проведення автоматичного морфемного сегментування словоформ тексту.

У результаті роботи автоматичного морфемного сегментатора на базі корпусу текстів поетичного мовлення було укладено цілу низку морфемних словників, які характеризують ідіостиль поета на морфемному рівні і дозволяють провести зіставно-порівняльний аналіз морфемних структур різних поетів, а також різних мовних стилів (рис. 2).

Рисунок 2. Фрагмент результатів автоматичного морфемного сегментування словоформ поетичного тексту Л. Костенко

id	cls	morfem	morfema	comm
378422	ПТ	R	за	за RC RC
378423	ЙИ	R	щит	щитом RDFE RDFF
378424	ЙИ	F	ом	щитом RDFE RDFF
378425	АИ	R	смарагд	смарагдових RHSJFL RHSJFL
378426	АИ	S	ов	смарагдових RHSJFL RHSJFL
378427	АИ	F	их	смарагдових RHSJFL RHSJFL
378428	ЙИ	R	ліс	лісів RDFE RDFF
378429	ЙИ	F	ів	лісів RDFE RDFF
378430	X	N	.	

Словник лексики кожного поета через словник-конкорданс модуля-аналізатора має постійний зв'язок із корпусом текстів, що дозволяє автоматично ілюструвати текстове вживання кожної словоформи і будувати електронну картку (рис. 3).

Така картка дозволяє швидко та ефективно створювати різноманітні лексичні словники: *Словники синонімів, омонімів, неологізмів, діалектизмів, архаїзмів, фразеологізмів; Словники тропів: епітетів, метафор, метонімії, синекдох, порівнянь, оксюморонів, гіпербол.*

Ці словники укладаються в автоматизованому режимі, оскільки така інформація не може бути повністю формалізованою і вимагає участі лінгвіста. Наприклад, укладання словника неологізмів, словника синонімічних гнізд, ідеографічного словника супроводжується автоматичним морфемним та словотвірним аналізом, але вимагає тлумачення значення слова, введення його до відповідної лексико-семантичної групи синонімічної схеми або передбачає контроль за значенням слів при зіставленні їх із тією чи іншою тематичною групою електронного словника синонімічних гнізд.

Методика формалізованого лінгвістичного опису мовних одиниць у створенні системи автоматизованого лінгвістичного аналізу є узагальненням теоретичних і прикладних ідей сучасного мовознавства, що робить цю систему надзвичайно ефективним і раціональним інструментом лінгвістичних досліджень.

Рисунок 3. Фрагмент Частотного словника поетичного мовлення М. Вінграновського з текстовою ілюстрацією

Полем3	Полем1	Кодес.Полем2	kont	subre:	xser	sigma	xser
Батьківщина	батьківщини	кл.	ім.ж.р.в.	кК	0,1	0,43589	0,2
батько	батька	зн.	ім.ч.р.	ЙБ	0,05	0,217945	0,3

Запись: 76 из 7335

Побудувати картку

Примеры словоупотребления:

І так усе... задумливо лінією закреслилась ти в голубій івилі в тонкій руці з прив'яленою лілією на тлі очей моїх... на світу білім тлі... спогад про війну... сумні без батька двоє дітей цих...
 ...не люби свого батька - ту руку стару... не люби його садч вишневу кору... не люби свою матір в печалі жалі... не люби її кроки м'які і малі.

Ця методика може бути використана як еталонна при укладанні різноманітних електронних словників, картотек, автоматизованій класифікації лінгвістичного матеріалу, побудові навчальних комп'ютерних тренажерів, тестів тощо. З іншого боку, отримана параметризована інформація про організацію українських текстів на різних рівнях мовної системи дає можливість вивчити закономірності функціонування системи мови в різних стилях, комплексно досліджувати мовні особливості ідіостилів українських поетів та письменників.

*В. Старко, к. філол. н.**

Волинський державний університет імені Лесі Українки (Луцьк)
 УДК 161.2.81'373.374.322

ФРАЗЕОЛОГІЯ, ЛЕКСИКОГРАФІЯ І КОРПУС

The article deals with the problems of multiaspect investigation and lexicographic description of phraseology based on corpus data. The author discusses the frequency of idioms and proverbs and the demand it places on the size and content of corpora and compares three different corpora with respect to their suitability for phraseology research.

Корпусні дослідження фразеології є відносно новою галуззю лінгвістики. Чи не найзначніший внесок в її розвиток внесла Розамунд Мун, чия робота [3] була першою

* © В. Старко, 2006