

Старишко Ю. В.

## ПРАКТИЧНІ ЗАСАДИ ЛЕКСИКОГРАФІЧНОГО МОДЕЛЮВАННЯ ДЕРИВАЦІЙНИХ ОДИНИЦЬ (на матеріалі «Українсько-російського словотворчого словника» З. С. Сікорської)

*У статті проаналізовано способи відображення словотвірної структури в «Українсько-російському словотворчому словнику» (З. С. Сікорська), на основі яких створено дериваційні моделі, що забезпечать функціонування Віртуальної інструментальної дериваційної лабораторії та Інтегрованої віртуальної лексикографічної лабораторії з автоматизованим словотвірним аналізом.*

**Ключові слова:** моделювання дериваційних одиниць, Віртуальна інструментальна дериваційна лабораторія, Інтегрована віртуальна лексикографічна лабораторія з автоматизованим словотвірним аналізом, квазісемантичні класи.

Серед переліку наявних лексикографічних систем України на сьогодні не створено інструментальне програмне забезпечення, що здійснювало б маніпулювання дериваційними одиницями. Ґрунтовний словотвірний аналіз слів української мови розкрито у працях Є. А. Карпіловської «Кореневий гніздовий словник української мови» [1], колективному виданні «Шкільний словотвірний словник сучасної української мови» (Н. Ф. Клименко, Є. А. Карпіловська, Л. П. Кислюк) [2]; проте зазначені системи ізольовано функціонують у цифровому форматі. Аналіз праць та повна відсутність відповідного програмного забезпечення свідчать, що створення Віртуальної інструментальної дериваційної лабораторії є актуальним завданням.

На основі аналізу зазначених праць вирішено будувати комп'ютерну систему на матеріалі «Українсько-російського словотворчого словника» (З. С. Сікорська), оскільки він

- має найбільш повну кількість словотвірних моделей, структури яких є достатньо простими та несуперечливими (головний критерій);
- містить достатньо репрезентативний мовний матеріал (близько 42 тисяч слів);
- становить культурно-історичну цінність (перший у слов'янському мовознавстві перекладний словотвірний словник).

На матеріалах цього словника зусиллями співробітників Українського мовно-інформаційного фонду Національної академії наук України (УМІФ НАН України) створюється Віртуальна інструментальна дериваційна лабораторія. Завдяки на-

явним на сьогодні електронним лексикографічним системам УМІФ (Електронний граматичний словник (<http://lcorp.ulif.org.ua/dictua/>); система російської словозміни на основі словника А. А. Залізняка; Віртуальна лексикографічна лабораторія «Словник української мови» (<http://lcorp.ulif.org.ua/vll/>), система природномовного аналізу корпусного типу (<http://idx.ulif.org.ua:8888/idx/>)) стає можливим інтеграція цієї системи в єдину Інтегровану віртуальну лексикографічну лабораторію з автоматизованим словотворчим аналізом (інтегрована система – сукупність двох або кількох взаємопов'язаних систем, у якій функціонування однієї з них залежить від результатів функціонування іншої (інших) так, що цю сукупність можна розглядати як єдину систему [4, с. 65]).

У такому випадку необхідно врахувати, що створювана нами програма має містити не тільки набір функцій для автоматизованого словотвірного аналізу, а мати ще й такий функціонал:

- відображення:
  - реєстру слів;
  - повної словозміни реєстрової одиниці;
  - частиномовної належності;
  - тлумачення реєстрованої лексеми;
  - твірного слова;
  - типу словотворення;
- розклад реєстрових слів на морфеми, їхня класифікація за квазісемантичними параметрами;
- побудова (репрезентація) дериваційної моделі.

Зазначена система повністю ідентична структурі паперового словника. Морфологічна будова, словотворчі засоби та правила похідності (див. у передмові до словника [3, с. 6–14] не змінюватимуться і не редагуватимуться авторами системи).

Словникова стаття словника складається зі словотвірної пари слів української мови (похідне у початковій формі + твірне) і відповідної словотвірної пари російською мовою. Якщо дериват в українській мові (або російській) співвідноситься з непохідним у російській мові (або українській) словом чи не має адекватного відповідника (перекладається описово), словотвірна пара вказується лише в українській мові (або лише російській), наприклад:

*молоч+áрств-о* ← *молóч/н/ий* (-н-) молочная промышленность

Роздільником між твірним і похідним словом слугує стрілка (←), яка вказує на напрям похідності, проте необхідно врахувати випадки варіативності похідного слова: *крас+ень, крас+унь* ← *крас/ів/ий*. Морфемна будова твірних слів (у паперовому словнику позначається через похилу риску (/)) не враховується системою, оскільки, згідно з принципом замкнутості словника, твірне слово має розкладатися на морфемі в іншій словниковій статті:

*механіз+áтор* ← *механ/із/áці[ü/a]*, (-áцій-), *механ/ізм* (-м);  
*механіз+áці[ü-a]* ← *механ/із/ува/ти* (-ува-);  
*механіз+ува-ти* ← *механ/зм* (-м).

Словотвірна структура похідних слів відображена в словниковій статті у вигляді конструкції: твірна основа (твірні основи, якщо слово складне) + словотворчий засіб (словотворчі засоби), що бере безпосередню участь в утворенні цього слова; закінчення (матеріально виражене) відділяється рисою. Це дає можливість побудувати універсальну модель простих похідних слів (одноосновних):

$$P + O_{(x)} + (-*)S + F,$$

де P – префіксальна частина:

$$P = \left\{ \begin{array}{l} \emptyset; \\ \{p_k\}, k = 1, \dots, n, \text{ де } p_k \in \{p_1, \dots, p_n\} \end{array} \right.$$

– множина всіх префіксів (префіксоїдів) української мови;  $O_{(x)}$  – основа твірного слова як компонент семантичної бази похідного:

$O_{(x)} = \{o_k\}, k = 1, \dots, m, \text{ де } p_k \in \{p_1, \dots, p_m\}$  – множина усіх твірних основ української мови; S – суфіксальні форманти (пуста множина, позначена знаком  $\emptyset$ ) (*кул+e+мет+ $\emptyset$* ):

\* Параметр вводиться для безафіксного способу словотворення.

$$S = \left\{ \begin{array}{l} \emptyset; \\ \{s_k\}, k = 1, \dots, l, \text{ де } s_k \in \{s_1, \dots, s_l\}; \end{array} \right.$$

F – словозмінна флексія або закінчення:

$$F = \left\{ \begin{array}{l} \emptyset; \\ \{f_k\}, k = 1, \dots, r, \text{ де } s_k \in \{s_1, \dots, s_r\}. \end{array} \right.$$

Певні труднощі для побудови системи створюють внутрішні закінчення, після яких слідує постфікси *-ся, -сь, -небудь*: *зі/ü/тү/ся, котрій/сь, як/ий/-небудь*. Будемо вважати їх квазіфлексією.

Морфонологічні явища (чергування, нарощення, накладання), що відбуваються при словотворенні, в паперовому словнику подаються в дужках (пор. *дорóж+н-ій* ← *дорóг/а* (г-ж); *вигляд+á-ти* ← *ві/гля/ну/ти* (-ну-, ≠-д); *четв+ер-о* ← *чотір/и* (-ир-, о-е, ≠-е)). Інформацію такого типу ми не враховуємо, оскільки інтеграція з Граматичним словником містить ці дані у відповідних парадигматичних класах.

Для багатоосновних слів матимемо моделі:

$$P_1 + O_{(x)1} + (-)S_1 + P_2 + O_{(x)2} + (-)S_2 + \dots + P_n + O_{(x)n} + (-)S_n + F \text{ (пор. } \textit{мал+о+у+дій+н-ий, вод+о+о+чис+н+ий, мед+сан+бáт),}$$

де  $P_y$  – префіксальна частина відповідної основи ( $y = \{1, 2, \dots, n\}$ ):

$P_y = \left\{ \begin{array}{l} \emptyset; \\ \{p_k\}, k = 1, \dots, n, \text{ де } p_k \in \{p_1, \dots, p_n\} \end{array} \right.$  – множина усіх префіксів (префіксоїдів) української мови;

$O_{(x)y}$  – основи твірного слова ( $y = \{1, 2, \dots, n\}$ ):

$O_{(x)y} = \{o_k\}, k = 1, \dots, m, \text{ де } p_k \in \{p_1, \dots, p_m\}$  – множина усіх твірних основ української мови;

S – суфіксальні форманти (у тому числі інтерфікси) ( $y = \{1, 2, \dots, n\}$ ):

$$S_y = \left\{ \begin{array}{l} \emptyset; \\ \{s_k\}, k = 1, \dots, l, \text{ де } s_k \in \{s_1, \dots, s_l\}; \end{array} \right.$$

F – словозмінна флексія або закінчення:

$$F = \left\{ \begin{array}{l} \emptyset; \\ \{f_k\}, k = 1, \dots, r, \text{ де } s_k \in \{s_1, \dots, s_r\}. \end{array} \right.$$

Складні слова, написані через дефіс, репрезентує така модель:

$$P_1 + O_{(x)1} + (-)S_1 /-/+ P_2 + O_{(x)2} + (-)S_2 /-/+ \dots + P_n + O_{(x)n} + (-)S_n + F: \textit{капітан/-/+лейтенант, наукóв+о/-/+популярний.}$$

Аналіз статей засвідчує, що кількість плюсів (плюс (+) – поєднує словотворчі частини слова; матеріально виражене закінчення в похідних виділяється дефісом (-)) нараховує не більше 4-5: *культур+н+о/-/+побут+óв-ий, вод+о+о+чис+н+ий*.

Залучення до системи 20-томного Словника сучасної української мови (СУМ 20) відкриває нові можливості для Інтегрованої віртуальної

лексикографічної лабораторії з автоматизованим словотвірним аналізом: на основі формул квазісемантики, визначених у СУМ 20, виділяємо квазісемантичні класи дериваційних морфем (пор. клас префіксів зі значенням відсутності чого-небудь: *а-* (*абіогенез, абіогенний, алогічний, аперіодичний, аперіодично, аритмічний, аритмічно, аритмія, асиметричний, асиметрично, асиметрія, асимільований, асимільовати, асимільоватися, асиміляція*); *дез-* (*дезактивізація, дезінтегру-*

*вати, дезінфекція, дезорієнтований*); *диз-* (*дизадаптація, дизаритмічний*); *дис-* (*дисбаланс, диспалаталізація*)). Загальний список зазначених лексичних одиниць нараховує 739 морфем (список ще поповнюється).

Отже, пропонується Інтегрована віртуальна лексикографічна лабораторія з автоматизованим словотвірним аналізом забезпечує вирішення широкого кола питань з українського й частково з російського словотворення.

#### Список літератури

1. Карпіловська Є. А. Кореневий гніздовий словник української мови / Є. А. Карпіловська. – К. : Українська енциклопедія, 2002. – 912 с.
2. Клименко Н. Ф. Шкільний словотвірний словник сучасної української мови / Н. Ф. Клименко, Є. А. Карпіловська, Л. П. Кислюк. – К. : Наукова думка, 2005. – 264 с.
3. Сікорська З. С. Українсько-російський словотвірний словник / З. С. Сікорська. – 2-ге вид., перероб. і доп. – К. : Освіта, 1995. – 256 с.
4. Широков В. А. Елементи лексикографії / В. А. Широков. – К. : Довіра, 2005. – 304 с.

*Yu. Staryshko*

### PRACTICAL PRINCIPLES OF LEXICOGRAPHICAL MODELLING OF DERIVATIVE UNITS (a case study of Ukrainian-Russian Word-Building Dictionary edited by Z. S. Sikorska)

*This article analyses the models of dictionary entries and the ways to display the word-building structure in Ukrainian-Russian Word-Building Dictionary (Z. S. Sikorska). Basing on this analysis, the author develops general models of derivatives (derivative models for simplexes and complex words are different). These developed derivative models will ensure functioning of the Virtual Instrumental Derivative Laboratory. Integration of the mentioned above system with the United Integrated Virtual Lexicographic Laboratory with automated analysis of word-building is possible because of already existing on-line lexicographical systems developed by the Ukrainian Lingua-Information Fund of NAS of Ukraine (Online Grammar Dictionary; System of Russian Inflection, based on the dictionary edited by Zalizniak A. A.; Virtual Lexicographical Laboratory "Explanatory Dictionary of the Ukrainian Language"; corpus system of natural language analysis). Integration with the system of Dictionary of the Ukrainian Language in 20 volumes is extremely important, because this makes it possible to determine a semantic meaning of the word stem and also to extract quasi-semantic classes of prefixes and prefixoids, suffixes and suffixoids. Thus, the proposed system of automated analysis of word-building provides a solution for a wide range of problems related to Ukrainian and, partly, Russian word building.*

**Keywords:** modelling of derivatives, Virtual Instrumental Derivative Laboratory, Integrated Virtual Lexicographic Laboratory with automated analysis of word-building, quasi-semantic classes.

*Матеріал надійшов 26.09.2016*