

МІСЦЕ СЛОВОТВОРЕННЯ В РОЗВИТКУ СУЧАСНОГО УКРАЇНСЬКОГО ЛЕКСИКОНУ —

У статті відповідно до знаннеорієнтованої системи машинного перекладу розглядається можливість оптимізації обсягу перекладного словника засобами словотворення. Головну увагу зосереджено на методиці формування словника квазісуфіксів та формату його словникової статті, характеристиці побудови бази даних продуктивних моделей суфіксального словотворення.

Ключові слова: оптимізація обсягу перекладного словника, знаннеорієнтована система машинного перекладу, словотвірна модель, словотвірний тип, формат словникової статті, база даних моделей суфіксального словотворення.

Лексичний склад будь-якої мови — це динамічна система, функціонування і розвиток якої визначається механізмом утворення слів. Для практичних потреб людство навчилося систематизувати слова у словниках завдяки різним принципам подання матеріалу, різноманітним методикам добору та опису текстового слова. На сьогодні у світі наявна багата спадщина різноманітних словників і з мінімальним, і з інтегральним описом слова¹. Якщо від початку лексикографії слугували матеріали картотек, то сьогодні — мільйонні корпуси текстів з різношаровою лексикою². Не поступається фактичному матеріалу й вагома теоретична база, логічним продовженням якої стала теорія лексикографічних систем³.

Актуальною залишається проблема обсягу словника у двох її виявах — максимально повній фіксації наявних у мові слів (з урахуванням суб'єктивності у формуванні реєстру) та ідентифікації нових лексем. Якщо в середині минулого століття питання кількості та фіксації словникових одиниць активно дискутувалося в контексті теорії та практики перекладу, зокрема машинного⁴, то сьогодні йдеться про опис великого за обсягом словника для будь-якої галузі мовознавства. Потреба представлення об'єктивного світу в слові, викликана швидкими темпами цивілізаційного розвитку людства, зумовлює інтенсивне поновлення

¹ *Апресян Ю. Д.* Лексическая семантика. Синонимические средства языка. — М., 1995. — 473 с.; *Караулов Ю. Н.* Лингвистическое конструирование и тезаурус литературного языка. — М., 1981. — 357 с.; *Баранов А. Н.* Введение в прикладную лингвистику : Учебник. — М., 2001. — 360 с.

² Британський національний корпус містить понад 100 млн слововживань [<http://www.patcorp.ox.ac.uk>], стільки ж — Корпус іспанської мови [<http://www.corpusdelespanol.org/>], Український національний лінгвістичний корпус — понад 42 млн слововживань, Національний корпус російської мови — понад 35 млн слововживань [http://ruscorp.org/corpora_stat.html], Корпус польської мови Наукового видавництва PWN — близько 40 млн слововживань [<http://www.korpus.pwn.pl>].

³ *Широков В. А.* Феноменологія лексикографічних систем. — К., 2004. — 327 с.

⁴ *Ревзин И. И., Розенцвейг В. Ю.* Основы общего и машинного перевода. — М., 1964. — С. 180–185.

лексичного складу будь-якої мови. Свідченням такої тенденції є словники нових слів і значень, лінгвістичні дослідження темпів та специфіки розвитку словникового складу⁵.

Для розуміння нового слова, зокрема в системах автоматичного аналізу, одну із провідних ролей відіграє словотворення, оскільки відображає процес породження та сприйняття не тільки наявних у мові лексичних одиниць, а й слів, що позначають нові реалії. Моделювання механізмів словотворення як одного з компонентів морфологічної моделі мови передбачає застосування сучасних як теоретичних, так і практичних здобутків у царині лінгвістичних знань та інформаційних технологій. Теоретичні та практичні розвідки, присвячені вивченню словотвору та побудові модуля автоматичного аналізу та синтезу словотвірних явищ, спеціальні лінгвістичні бази даних, у тому числі й словотвірні словники та словники морфем, за своїм характером є аналітичними, відображають позицію лінгвіста-дослідника⁶. Зважаючи на це, у контексті автоматичного опрацювання текстів природної мови існує потреба у створенні відповідних методик, які б забезпечували словотвірний аналіз без урахування знань про конкретну лексичну одиницю, одночасно подаючи її як елемент лексико-граматичної системи мови.

Відповідно до завдань розроблення знаннеорієнтованої системи машинного перекладу подамо принципи побудови словників афіксів для української, російської та англійської мов на основі запропонованої методики виділення словотворчих афіксів⁷. Ці словники розглядаємо як інформаційний ресурс — частину словотвірної моделі мови, що може доповнюватися та бути інтегрованою до системи автоматичного морфологічного опрацювання тексту природної мови. Основою формування баз даних стали граматичні та орфографічні словники обраних мов та корпус різножанрових текстів.

Потужний арсенал теоретичних розвідок та теоретико-прикладних праць зі словотвору містить необхідний матеріал для моделювання словопороджувального механізму мови. Зокрема, колектив співробітників Інституту мовознавства ім. О. О. Потебні НАН України разом зі співробітниками Інституту української мови НАН України розробили проект спеціальної автоматизованої інформаційної системи — морфемно-словотвірний фонд української мови⁸. Цей проект, крім фактичної бази морфемно-словотвірного фонду української мови, має дослідницьку частину, головне завдання якої — побудова морфемних словників з показниками продуктивності використання окремих морфем при словотворенні.

⁵ Див. бібліографію: *Клименко Н. Ф., Карпіловська Є. А. та ін.* Динамічні процеси в сучасному українському лексиконі : Монографія.— К., 2008.— 336 с.; *Струганець Л. В.* Динаміка лексичних норм української літературної мови ХХ століття.— Тернопіль, 2002.— 348 с.

⁶ *Клименко Н. Ф., Карпіловська Є. А. та ін.* Словотвірна морфеміка сучасної української літературної мови.— К., 1998.— 162 с.; *Клименко Н. Ф., Карпіловська Є. А.* Морфемно-словотвірний фонд української мови як дослідницька та інформаційно-довідкова система // *Мовознавство*.— 1990.— № 6.— С. 41–50; *Клименко Н. Ф., Карпіловська Є. А. та ін.* Словник афіксальних морфем української мови.— К., 1998.— 434 с.; *Карпіловська Є. А.* Кореневий гніздовий словник української мови : Гнізда слів з вершинами-омографічними коренями.— К., 2002.— 912 с.; *Степанова М. Д.* Вопросы моделирования в словообразовании и условия реализации моделей // *Вопр. языкознания*.— 1975.— № 4.— С. 24–37; *Тагабилева М. Г., Березуцкая Ю. Н.* Словообразовательная разметка Национального корпуса русского языка: задачи и методы.— <http://www.dialog-21.ru/dialog2010/materials/pdf/73.pdf>; *Образование употребительных слов русского языка / Под ред. Л. Н. Засориной*.— М., 1979.— 279 с.

⁷ *Замаруева І. В.* Комп'ютерна модель розуміння природно-мовної текстової інформації // *Проблеми програмування*.— 1999.— № 2.— С. 96–102.

⁸ *Клименко Н. Ф., Карпіловська Є. А.* Морфемно-словотвірний фонд української мови...— С. 41–50.

Зважаючи на розробку знаннеорієнтованої системи машинного перекладу, що спирається на вилучення закодованих у текстовій інформації знань з предметної галузі, запропонована нами словотвірна модель повинна передусім забезпечувати оптимізацію побудови електронних перекладних словників. У своїй структурі модель містить словник квазісуфіксів та квазіпрефіксів. Для виділення одиниць базового лексичного складу мови ми пропонуємо спеціальну методичку, суть якої полягає у виділенні словотворчих квазіформантів з початку та з кінця слова — за методом аналогії.

Мовознавці неодноразово залучали метод аналогії до своїх досліджень зі словотвору. Так, Ф. О. Нікітіна обґрунтувала саме явище аналогії та його значення у словотвірних процесах⁹. Став у нагоді метод аналогії і для Д. С. Уорта під час розроблення словотвірного словника російської мови¹⁰. Цей метод успішно використовували російські вчені, зокрема Г. Г. Белоногов, для граматичної ідентифікації нових слів¹¹. У нашому дослідженні принцип аналогії може бути сформульований так: аналогічним формально-семантичним схемам побудови слів відповідають певні словотвірні значення. Для застосування цього підходу необхідно виявити всі або найбільш імовірні формально-семантичні схеми та визначити відповідні їм словотвірні значення.

З-поміж ключових понять словотвору, розроблених лінгвістами для об'єктивації процесів породження та сприйняття словотворчих процесів, актуальними для нашого дослідження є словотворчий афікс, словотвірне значення, словотвірний тип, словотвірна модель, словотвірна категорія¹². Базовою одиницею запропонованої студії є словотвірна модель — структурно-семантична словотвірна схема, що об'єднує певну кількість слів, для яких можна встановити однозначне співвідношення словотворчого форманта (словотворчого афікса, за допомогою якого утворюється лексична одиниця) та твірної основи. Словотворчий формант — це структурна одиниця, яка виконує дериваційну функцію. За своєю формою словотворчий формант може співвідноситися з одиницею, що виділяється під час традиційного словотвірного аналізу, а також може бути виділеним за необхідності суто формально. Твірна основа в межах нашого дослідження може мати також формальний характер і не відповідати вимогам традиційного словотвірного аналізу, наприклад, щодо напрямку мотивації.

Лексико-граматичне співвідношення між операційно виявленими словотворчим формантом та твірною основою — це категорійне словотвірне значення, притаманне визначеній словотвірній моделі, а отже, і лексичній одиниці, яка відповідає заздалегідь встановленим правилам.

⁹ Нікітіна Ф. А. Влияние аналогии на словообразование (На материале родственных языков).— К., 1973.— 198 с.

¹⁰ Уорт Д. С. Русский словообразовательный словарь. Введение // Новое в зарубежной лингвистике.— М., 1983.— С. 227–260.

¹¹ Белоногов Г. Г., Зеленков Ю. Г. Еще раз о принципе аналогии в морфологии // НТИ.— Сер. 2. Информ. процессы и системы.— 1995.— № 3.— С. 29–32.; Белоногов Г. Г., Зеленков Ю. Г. и др. Метод аналогии в компьютерной лингвистике // Там же.— 2000.— № 1.— С. 21–30.

¹² Клименко Н. Ф. Словотвірний тип // Українська мова : Енциклопедія / Ред. кол.: Русанівський В. М., Тараненко О. О. (співголови) та ін.— К., 2000.— С. 645–646; Клименко Н. Ф. Словотвірна категорія // Там же.— С. 641; Клименко Н. Ф. Словотвірна модель // Там же.— С. 642. Клименко Н. Ф. Формант // Там же.— С. 799; Словотвір сучасної української літературної мови / За ред. М. А. Жовтобрюха.— К., 1979.— 404 с.; Быкова Л. А. Современный русский литературный язык : Морфемика и словообразование.— К., 1974.— 94 с.; Кубрякова Е. С. Типы языковых значений : Семантика производного слова.— М., 2008.— 208 с.

Визначені під час дослідження словотвірні моделі об'єднані словотвірними типами, що підпорядковані словотвірним категоріям. Словотвірний тип розуміємо як формально-семантичну схему побудови похідних слів, мотивованих твірними словами певної частини мови, за допомогою того самого словотвірного форманта з тим самим значенням. Словотвірну категорію розглядаємо як сукупність похідних слів, що мають спільне словотвірне значення і використовують однаковий спосіб словотворення.

У своєму дослідженні ми розрізняємо й поняття продуктивності та уживаності словотвірних категорій¹³. Продуктивність визначаємо як властивість, притаманну мові й фіксовану в словниках та граматиках. Цю категорію розглядаємо у площині як кількісних даних, так і доцільності, можливості виділення тієї чи іншої моделі. Уживаність — це спроможність словотвірних одиниць до використання їх у словах лексичного складу мови.

Розглянемо процедуру виділення словотворчого афікса, визначення продуктивної словотвірної моделі, словотвірного типу та словотвірної категорії на прикладі українського словотворчого форманта *-ка*¹⁴. Так, у граматиках української мови суфікс *-ка* кваліфікується як неоднозначний: його вживають на позначення іменників жіночого роду (*архітекторка*), опрідметненої дії (*мандрівка*), як зменшено-пестливий суфікс (*рибка*). За інверсійним словником, укладеним на основі аналізованих нами текстів, та з урахуванням даних орфографічних словників української мови лексичні одиниці з формально виділеним квазісуфіксом *-ка* охоплюють значний масив лексики. Проте з огляду на обраний нами операційний підхід виділення словотворчого форманта та словотвірної моделі цей корпус лексичних одиниць є доволі строкатим. Наприклад, до нього потрапили слова, які за розроблюваною нами методикою визначення словотвірної моделі, не можуть бути проаналізованими: *русалка, вівсянка, шибка, лелека, ластівка, ложка*. Незважаючи на спільність форманта, для цих слів на синхронному етапі аналізу не визначаємо однозначне співвідношення похідного та твірного слів. Не потрапляють до нашого аналізу слова, в яких за одним із значень лексеми можна поділити на словотворчий афікс та твірну основу, а за іншим — ні. Наприклад, слова *лебідка, губка, скибка, помадка, колодка*, один із лексико-семантичних варіантів яких — це «зменшено-пестливе від іменника», а другий — означає поняття безвідносно до твірного. Не стали об'єктом нашого аналізу стилістично марковані слова з маркером «розмовне» типу *комуналка, стіннівка, читалка* та ін.

Розглянемо вісім словотвірних моделей із словотворчим формантом *-ка*¹⁵ (див. табл. 1). До алгоритму аналізу залучено інформацію про лексико-граматичне значення похідного слова (поле 4), про граматичне значення твірного слова (поле 5), вказано квазісуфікс, який необхідно додати для одержання пошукового слова (поле 3). Інформацію про словотвірну категорію, словотвірний тип, модель словотворення містять відповідно поля 6, 7 та 8. Застосований операційний підхід накладає обмеження й на одиниці, у яких виділений словотворчий

¹³ Клименко Н. Ф. Основи морфеміки сучасної української мови. — К, 1998. — С. 99–103.

¹⁴ Цей продуктивний формант виділяємо, зважаючи на дослідження словотвору сучасної української мови.

¹⁵ Застосовані позначення: іж — іменник жіночого роду; іжі — іменник жіночого роду, істота; іч — іменник чоловічого роду; ічі — іменник чоловічого роду, істота; дн — дієслово недоконаного виду (поле 4, поле 5); о — особа; одрд — опрідметнена дія, результат дії (поле 6); змп — зменшено-пестливе; ож — особа жіночої статі (поле 7); нм — особа за національністю, місцем проживання; вдп — особа за видом діяльності, ідейними переконаннями; іпфм — особа за ідейними переконаннями, формою мислення.

формант та отримана в такий спосіб твірна основа не дозволяють однозначно сформулювати категорійне словотвірне значення та ідентифікувати модель аналізу, а отже, й синтезу. На такі обмеження вказують поле 10 та з урахуванням стилістичної характеристики — поле 9. Якщо стилістичні обмеження поля 9 вказують на реальну мовленнєву картину, наприклад, у таблиці — це стилістично маркована розмовним стилем модель «зменшено-пестливе до іменника жіночого роду», то в полі 10 містяться винятки, які не відповідають нашій системі аналізу. Так, слово *голубка* можна ідентифікувати за моделлю: «особа жіночої статі за національністю, місцем проживання», оскільки граматичні значення і твірної основи, і твірного слова відповідають встановленому правилу, і словотворчий формант дозволяє встановити твірне слово — *голуб* — іменник чоловічого роду, істота. З огляду ж на знання про довкілля, слово *голубка* не має значення цієї моделі, його розпізнавання здійснюватиметься на етапі семантичного аналізу.

Подальший аналіз дистрибутивних можливостей словотворчого форманта **-ка** обраного корпусу лексики продемонстрував необхідність розширення контексту його реалізації для ефективності алгоритму аналізу. Наприклад, словотвірний тип «опредметнена дія, результат дії», де похідне слово — іменник жіночого роду, а твірне — дієслово недоконаного виду, містить близько 10 моделей, становлячи 3,74 % матеріалу (у табл. 2 подано моделі з найбільшою кількістю одиниць). Загалом виділено близько 65-ти моделей із формантом **-ка**, які представляють словотвірні типи: «особа жіночої статі», «опредметнена дія, результат дії», «зменшено-пестливе» та ін. Обраний підхід дозволяє уникнути неоднозначності завдяки залученню різної інформації (див. табл. 2).

Таблиця 1

Словотвірні моделі зі словотворчим формантом -ка (фрагмент)										
Поле 2	Поле 3	Поле 4	Поле 5	Поле 6	Поле 7	Поле 8	Поле 9	Поле 10	Поле 11	Поле 12
Квазі-суфікс 1	Квазі-суфікс 2	Гр. зн. 1	Гр. зн. 2	Сл. кат.	Сл. тип.	Сл. мод.	Ст.	Винятки	Приклади	Кількість ¹⁶
-ка	-а	іж	іж	іж	змп			р	садибка, шубка	40
-ка	0	іжі	ічі	о	ож	нм		голубка	капубка, швабка	1,2
-ка	-ець	іжі	ічі	о	ож	нм		кубанка	американка	13,6
-ка	0	іжі	ічі	о	ож	вдіп			оригіналка, меломанка	8,6
-ка	-ин	іжі	ічі	о	ож	нм			парижанка, марсіанка	3,05
-ка	-ець	іжі	ічі	о	ож	іпфм	р		початківка	0,8
-ка	-чати	іж	дн	д	одрд				відзнака, ознака	0,14
-ка	-кувати	іж	дн	д	одрд				атака, пака	0,08

¹⁶ Відносну частоту моделей встановлювали щодо загальної кількості слів із цим формантом в аналізованому списку слів.

Словотвірні моделі типу «опредметнена дія, результат дії» із словотворчим формантом -ка (фрагмент)								
Поле 2	Поле 3	Поле 4	Поле 5	Поле 6	Поле 7	Поле 8	Поле 10	Поле 11
Квазі-суфікс 1	Квазі-суфікс 2	Гр. зн. 1	Гр. зн. 2	Сл. тип	Сл. мод.	Сл. мод. 1	Виятки	Приклади
-вка	-вляти	іж	дн	одрд				надбавка, заставка
-дка	-джувати	іж	дн	одрд				висадка, розсадка
-ска	-шувати	іж	дн	одрд				навіска
-вка	-вати	іж	дн	одрд				поливка
-івка	-увати	іж	дн	одрд			гвинтівка, готівка	мандрівка готівка
-ірка	-іряти	іж	дн	одрд				перевірка

Аналіз текстів, зокрема в системах машинного перекладу, показав, що відсутність перекладних еквівалентів у машинному словнику пов'язана як з випадками опису нових понять (нових фрагментів знань), так і з продуктивним афіксальним словотворенням, на частку якого припадає близько половини так званих нових слів (наприклад: *near* — *nearly*; *вчитель* — *вчителька* тощо).

Словотвірна модель у розроблюваній системі машинного перекладу включає словник префіксів та словник суфіксів. Розглянемо процедуру аналізу на прикладі розроблюваних словників суфіксів.

Так, формат словникової статті продуктивних суфіксів має вигляд:

$\langle S_i \langle S_k \rangle$ [поле граматичних ознак] # [поле семантичних ознак]>

де S_i — послідовність літер, яку треба відкинути, S_k — послідовність літер, яку треба додати. Поле граматичних ознак містить інформацію про частину мови слова з відповідним суфіксом та відповідні граматичні категорії (рід, істота тощо), поле семантичних ознак — відповідну цьому суфіксу семантику (наприклад, розмір, рід діяльності тощо).

Процедура опрацювання не відомого системі слова полягає в тому, що у випадку, 1) якщо слово (у початковій формі) не знайдено в перекладному словнику, слід перейти на словник суфіксів; 2) якщо суфікс знайдено, відкинути S_i від слова і додати S_k ; 3) перейти на перекладний словник; 4) якщо слово знайдено, взяти перекладний еквівалент та синтезувати необхідну словоформу через синтетичний словотвірний словник вихідної мови; 5) якщо слово не знайдено, перейти на словник префіксів (опрацювання аналогічним способом); 6) якщо слово не знайдено, перейти на рівень семантичного розпізнавання.

Так, російський суфікс **-ниц(а)** налічує близько 1000 словникових одиниць¹⁷, які можна подати у словнику суфіксів двома словниковими статтями:

*ница (ник) * іменник жс. роду, істота#рід діяльності для жінки;*

*ьница (ь) * іменник жс. роду, істота#рід діяльності для жінки,*

що забезпечують опрацювання понад 600 слів, які доречно представляти в перекладному словнику лише для іменників чоловічого роду, істот. Решту слів

¹⁷ Зализняк А. А. Грамматический словарь русского языка: Словоизменение. Около 100000 слов.— М., 1980.— 880 с.

необхідно додати до перекладного словника. Наприклад, в українській мові словотворчий формант зі значенням «зменшено-пестливе», що вживається для творення іменників жіночого роду **-ка**, охоплює близько 2 000 лексем, які в перекладному словнику можна подати однією статтею:

*ка (а) *іменник жс. роду, неістота # зменшено-пестливе.*

Реєстр кожного із словників формували на основі інформації, взятої з граматик аналізованих мов, граматичних та орфографічних словників. Цей фактографічний ресурс сприяв визначенню списку продуктивних типів і моделей словотворення.

Для формування остаточного списку активних моделей було залучено морфологічно розмічений корпус текстів та розроблено систему АРМ-лінгвіста¹⁸. У застосованій нами системі автоматичного морфологічного аналізу використано флективний підхід за словником квазізакінчень. Словник словоформ складено за текстами різної жанрово-тематичної спрямованості. За необхідності залучалися дані з граматик української та російської мов. Основою пропонованого принципу є розроблений метод, який спирається на позиційно-цифрове кодування граматичної інформації у словниковій статті. У такий спосіб кожна аналізована словоформа отримує свій код, що містить інформацію про її частино-мовну належність та конкретне граматичне значення (див. рис. 1).

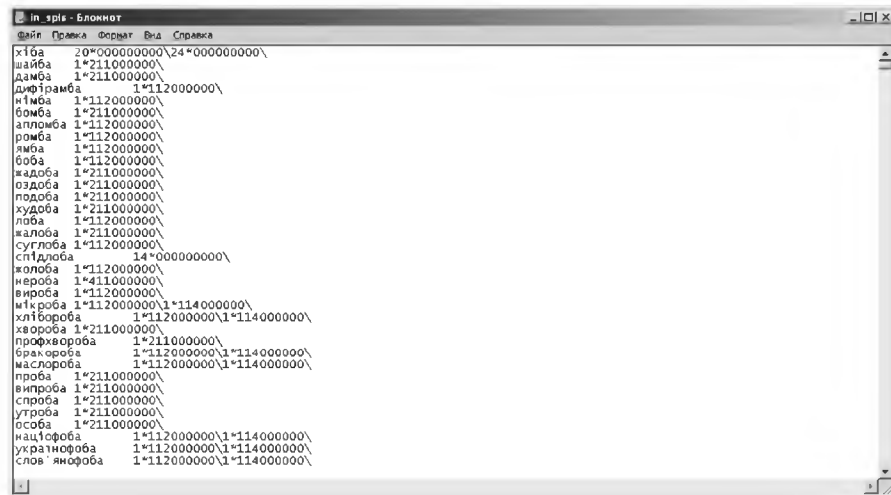


Рис. 1. Фрагмент інверсійного словника словоформ української мови.

Саме на основі інверсійного словника та завдяки розробленій пошуковій системі АРМ-лінгвіста визначалися вживані моделі суфіксального словотворення.

У результаті формування словників суфіксів для української, російської та англійської мов було розроблено базу квазісуфіксів (див. рис. 2), яку пропонуємо як необхідний універсальний інструментарій у вивченні та дослідженні аналізованого явища. Сформований лінгвістичний ресурс унаочнює результати визначення словотвірних моделей, містить усі необхідні квазісуфікси, встановлені за обраною методикою; подає продуктивні та активні моделі словопороджувального механізму; подає інформацію про похідне та твірне слово; репрезентує статистичну інформацію, достатню для характеристики квазісуфікса, моделі, типу.

¹⁸ Замаруєва І. В., Шипнівська О. О. Морфемна обробка текстів у системах машинного перекладу // Вісн. Київ. нац. ун-ту ім. Тараса Шевченка.— К., 2008.— № 20.— С. 61–63.

ID	suf_p	suf_1	suf_2	gram_p	gram_1	type	style_1	model	model1	model2	expert	example	freq
1	ка	хувати		ок	дн	дрд						атака, пака	5
2	ка	чати		ок	дн	рд						відзнака, ознака	4
3	ка	а		ок	ок	змл						свідка, шубка	2300
4	ка	0		ок	лч	ок					нм	голубка	85
5	ка	0		ок	лч	ок					р	оригіналка, професіоналка, мело	409
6	ка	ин		ок	лч	ок					нм	парисанка, марсанка	174
7	ка	ець		ок	лч	ок					нм	кубанка	776
8	ка	ець		ок	лч	ок					р	пфм	41
9	бка	оба		ок	ок	змл					р	зуділка	5
10	бка	бляти		ок	дн	од						обробка	4
11	бка	бати	бувати	ок	ад	дрд						розрубка, вирубка	17
12	ка	ити		ок	ад	дрд						насадка, висадка	16
13	вка	вляти		ок	дн	дрд						надбавка, заставка	14
14	дка	дити	дживати	ок	дн	дрд						висадка, розсадка	24
15	ска	сити	шувати	ок	дн	дрд						навіска	19
16	оска	осити		ок	дн	дрд						розноска, виноска	23
17	ка	увати		ок	дн	дрд					грамірка	затримка, підтримка, подка	56
18	ка	ати	увати	ок	ад	дрд						відгадка, розгадка, вигадка	37
19	вка	вати	ти	ок	дн	дрд						полівка	34
20	дка	дати	сти	ок	дн	дрд						закладка	21
21	дка	дувати	дати	ок	дн	дрд						розвідка	7
22	дка	джати		ок	дн	дрд						зарядка	5
23	ка	0		ок	лч	ок					пфм	сервілка	437
24	вка	ляць		ок	лч	ок					мп	чернілка	87
25	вка	ляць		ок	лч	ок					р	спартаківка, початківка, динамі	69
26	вка	ова		ок	ок	змл						головка, корівка	7
27	вка	увати		ок	дн	дрд						шнурівка, м'яздрівка	1
28	дка	ода		ок	ок	змл					р	подка	2
29	юка	га		ок	ок	змл					р	спазюка, смюка	8
30	ока	ога		ок	ок	змл					р	ночка	5
31	юка	га		іс	іс	зг					р	злудююка	2
32	вка	0		ок	лч	псдо						лінгвістика, софистика	41
33	ка	нк		ок	лч	псдо						ботаніка, техніка	38
34	йка	сь		ок	лч	ок					мп	плебейка	37

Рис. 2. База квазісуфіксів

ID	suf_p	suf_1	suf_2	gram_p	gram_1	type	style_1	model	model1	model2	expert	example	freq
72	ня	ти		іс	дн	дрд						остівування	2
73	ня	ти		іс	ад	дрд						комізування	3660
74	вня	ати	увати	іс	ад	дрд						обрубня	984
77	дування	дживати	дити	іс	дн	дрд						впровадження	87
82	єня	увати		іс	дн	дрд						заміщення, навішення, проголо	87
83	єня	ювати		іс	дн	дрд						освітлення	56
17	ка	увати		ок	дн	дрд					грамірка	затримка, підтримка, подка	56
75	єня	ювати		іс	дн	дрд						отруєня	48
18	ка	ати	увати	ок	ад	дрд						відгадка, розгадка, вигадка	37
19	вка	вати	ти	ок	дн	дрд						полівка	34
75	єня	іти	ювати	іс	ад	дрд						удостосня	24
14	дка	дити	дживати	ок	дн	дрд						висадка, розсадка	24
16	оска	осити		ок	дн	дрд						розноска, виноска	23
92	ня	тиса		іс	дн	дрд						ліствування	23
20	дка	дати	сти	ок	дн	дрд						закладка	21
15	ска	сити	шувати	ок	дн	дрд						навіска	19
11	бка	бати	бувати	ок	ад	дрд						розрубка, вирубка	17
12	ка	ити		ок	ад	дрд						насадка, висадка	16
38	нка	ти		ок	дн	дрд						оранка, читанка	14
85	щєня	стити		іс	дн	дрд						пещєня	14
13	вка	вляти		ок	дн	дрд						надбавка, заставка	14
94	єзня	осити		іс	дн	дрд						завезєня	12
84	єня	ити		іс	дн	дрд						тлумачєня, завезєня	12
88	чєня	чувати		іс	дн	дрд						перехрчувєня	8
21	дка	дувати	дати	ок	дн	дрд						розвідка	7
90	єня	ити		іс	дн	дрд						худєня	7
81	єзня	ізати		іс	дн	дрд						спостереєня	7
86	єзня	осити		іс	дн	дрд						внесєня	7
1	ка	хувати		ок	дн	дрд						атака, пака	5
22	дка	джати		ок	дн	дрд						зарядка	5
67	єчєня	ізати		іс	дн	дрд						залєчєня	5
42	пка	чувати		ок	дн	дрд						викрчувати	5
93	єзня	капи		іс	дн	дрд						замієня	4
91	єня	юватєся		іс	дн	дрд						спієня	4

Рис. 3. Модель типу «опредметнена дія, результат дії»

Подання лінгвістичної інформації в електронному форматі сприяє ефективності дослідження явища, дозволяє отримувати потрібну інформацію як у процесі дослідження, так і в разі залучення її до інших систем автоматичного аналізу тексту. Так, можна відібрати певну модель з необхідною інформацією про її продуктивність у мові, наприклад, модель типу «опредметнена дія, результат дії» словотвірної категорії «дія» з урахуванням їх частотного розподілу (за спадною) (див. рис. 3).

Отже, урахування продуктивного та активного афіксального словотворення дозволить розпізнавати «нові» слова, які утворюються завдяки цьому способу, навіть у разі відсутності їх у перекладному машинному словнику. Крім того, урахування словотвірної моделі дозволяє сформулювати вимоги до функціональної повноти перекладних словників у системах машинного перекладу. Застосована операційна методика виділення моделей словотворення на основі принципу аналогії є засобом економії в системі автоматичного опрацювання текстів. Залучення до аналізу нових текстів дасть можливість перевірити та в разі потреби доповнити наявну базу, з'ясувати не лише продуктивність, а й регулярність моделей у мовленні.

O. V. SHYPNIVSKA

THE PLACE OF THE DERIVATION IN THE DEVELOPMENT OF THE MODERN UKRAINIAN LEXICON

In the paper the possibility of optimization of the dictionary volume by the use of derivation according to the knowledge-based machine translation system is considered. The procedure of the creation of kwasisuffixes dictionary and the format of its entry are presented. The main principles of databases for the productive models of the suffixal derivation are described.

Key words: the optimization of the translated dictionary volume, the knowledge-based machine translation system, derivation model, derivation type, the format of the entry, the databases for the productive models of the suffixal derivation.