

**ПРЕДСТАВЛЕННЯ СЛОВОВІРНОЇ МОДЕЛІ  
В ЗНАННООРІЄНТОВАНІЙ СИСТЕМІ МАШИННОГО ПЕРЕКЛАДУ**

*У статті розглянуто визначення словотвірної моделі на основі принципу аналогії відповідно до засад побудови знаннеорієнтованої системи машинного перекладу. Подано теоретичні передумови встановлення словотвірного значення моделі.*

*Ключові слова: словотвірна модель, словотвірне значення, принцип аналогії, знаннеорієнтована система машинного перекладу.*

До загальних проблем словотвірного модуля системи автоматичного розпізнавання лексем входить процес синтезу-аналізу нових та наявних в мові та мовленні слів [1; 7; 10]. Незважаючи на неодноразову практичну реалізацію задачі, розробни-

ки по-різному бачать методи її вирішення [1; 2; 7; 10]. Різноплановість лінгвістичних явищ та необхідність побудови ефективного лінгвістичного процесора передбачає пошук нових шляхів розв'язання проблеми.

У статті пропонуємо методику визначення словотвірної моделі на основі методу аналогії [1]. Як джерело формування повного списку моделей розглядаємо базу даних словотвірних афіксів. У контексті запропонованого підходу як один із способів ідентифікації невідомих для системи слів розглянемо можливість визначення словотвірного значення за словотвірною моделлю. Робота виконується в рамках проекту зі створення знаннеорієнтованої системи машинного перекладу.

Сучасні досягнення в царині як теоретичного, так і практичного словотвору, базуючись на методах морфемного та словотвірного аналізів, на солідному фактичному матеріалі конкретних мов, слугують фундаментом для розробки дієвого модуля розпізнавання слів [2; 3; 4; 5; 10]. Так, зокрема і для української мови поруч із традиційними морфемними, словотвірними, гніздовими словниками, розроблено морфемно-словотвірний фонд української мови як дослідницьку та інформаційно-довідкову систему [2]. Одиницею опису цієї бази стало слово, яке вручну було поділене на морфеми. Розробниками бази були створені словники морфем української мови з показниками продуктивності. На основі бази було сформовано список моделей морфемної структури слів та побудовано морфемну модель словотворення української мови [2].

Незважаючи на це, у разі вирішення практичних завдань щодо словотвірного аналізу важко напряму застосувати отримані результати для побудови адекватного модуля словотвірного аналізу та синтезу. По-перше, виникає проблема, що стосується обсягу словника, а, відповідно, і повного списку моделей. По-друге, існують певні неточності щодо продуктивності і регулярності моделей. Як правило, для визначення продуктивності дослідники залучають словниковий матеріал, який на момент аналізу не може бути достатньо повним стосовно синхронного періоду мови. Істотно не поліпшує ситуацію і визначення регулярності моделей та її частотних характеристик. Як демонструє дослідницький матеріал, поза увагою науковців залишаються лексеми, ймовірність вживання яких мала, хоча саме вони в тексті можуть нести головне семантичне навантаження. Зокрема, це стосується нових термінів в науковій, науково-популярній літературі та в публіцистичних жанрах. По-третє, всі наявні бази, як електронні, так і паперові, орієнтовані передусім на слово – лінгвістичну одиницю, що має конкретне лексичне значення, відповідно, аналізуються і синтезуються слова, що існують у мові, а сама система представляє пасивну граматику (за термінологією Л. В. Щерби). Натомість активна граMATика передбачає аналіз не відомих для системи слів.

Виходячи із практичних завдань розроблення знаннеорієнтованої системи машинного перекладу, що спрямована на вилучення закодованих в текстовій інформації знань про предметну галузь, запропонована модель словотвору повинна перш за все оптимізувати побудову електронних перекладних словників. Структура моделі нестить словник суфіксальних та префіксальних афіксів. Вирішення завдань щодо побудови таких словників і зумовило створення відповідної бази даних. У статті ми з'ясуємо деякі теоретичні передумови дослідження і, оскільки робота має практич-

не спрямування, розглянемо процедуру виділення моделей словотворення, можливість їх застосування у системі.

Опис та пояснення процесів створення, використання та сприйняття похідних слів в словотворі сприяли напрацюванню відповідної теоретичної бази та головних понять словотвору [3; 5; 6; 8]. Як перед мовцями, так і перед розробниками лінгвістичного процесора постає завдання визначення словотвірного значення слова завдяки певним механізмам.

Наша робота зорієнтована на встановлення закономірностей перетворень простих одиниць в складні та виявлення механізмів їх обмежень. Серед понять словотвору актуальними для нас є: словотвірний афікс, словотвірна модель, словотвірний тип та словотвірна категорія. Базова одиниця запропонованої праці – словотвірна модель – структурно-семантична словотворча схема, що поєднує визначену групу слів, для яких можливо встановити однозначне співвідношення словотвірного форманта і твірної основи. У нашому випадку словотвірний формант – структурна одиниця, що виконує дериваційну функцію. За своєю формою словотвірний формант може співвідноситися з одиницею, що виділяють при традиційному словотвірному аналізі, а також може бути виділена суто формально. Твірна основа може також мати суто формальний характер і не відповідати вимогам традиційного словотвірного аналізу.

Визначені під час дослідження словотвірні моделі підпорядковані словотвірним типам та словотвірним категоріям. Під словотвірним типом розуміємо формально-семантичну схему побудови похідних слів, мотивованих твірними певної частини мови, за допомогою форманта з одним і тим же значенням. Словотвірну категорію розглядаємо як сукупність похідних слів, які мають спільне словотвірне значення і використовують той самий спосіб словотворення. Лексико-граматичне співвідношення між формально виділеними словотвірним формантом і твірною основою – це категоріальне словотвірне значення, притаманне всій словотвірній моделі, і, відповідно, лексемі, що входить до цієї словотвірної моделі.

Для моделювання як аналізу, так і синтезу словотвірного процесу важливим є поняття про системний характер словотвірного значення як «результату взаємодії категоріального значення форманта та лексико-граматичного значення твірного слова» [6]. Так, Є. С. Кубрякова, відзначаючи регламентований та модельований характер семантики похідного слова, вказує на значущість для побудови значеннєвої структури похідного слова не тільки його мотивуючого слова й афікса, але й тієї моделі, за якою ця структура організована [4, 15-16]. Важливим щодо наших практичних завдань є висновок дослідниці щодо входження слова в певну формулу мотивації, що повторюється для всієї серії однотипних утворень», що має «власну, задану структуру» та думка про те, що «на певному рівні абстракції похідне слово і його дефініція можуть розглядатися як семантично тотожні одиниці [4].

У рамках розробленої нами системи найбільш прийнятним, на нашу думку, є моделювання словотвірного процесу за методом аналогії. Механізм словотвірного процесу такого типу полягає у відтворенні готового в мові зразка: «слово, створене за аналогією, семантично будується так само, як і його прототип», а тому його семантична структура обмежена визначеним значенням [4, 25-29]. У словотворенні за

аналогією лежить знання словотвірної моделі як низки однотипних структурно-семантичних утворень, об'єднаних одним афіксом, схемою побудови похідних одного типу, а також розумінням одного і того ж лексичного зразка.

Важливим для створення модуля автоматичного словотвірного аналізу є питання мотивованості. У контексті розробленої системи, найбільш прийнятним, на нашу думку, є підхід, що відображає саме інтуїцію лінгвіста, а, отже, і носія мови. За таким підходом, мотивоване слово сприймається як таке, що усвідомлюється як утворене від іншого, наявного в мові слова на певному синхронному відрізку, а мотивація – відношення між словами, одне із яких усвідомлюється як мотивуюче (твірне), а інше – мотивоване (похідне), причому питання про дійсну генетичну похідність і про хронологію в цьому випадку не є важливим» [8, с. 10].

Для нашої роботи актуальним є важливе питання закономірностей сполучуваності мотивуючого слова та словотвірного форманта. Обмеження полягають у тому, що мотивуючі слова мають властивості, які сприяють утворенню мотивованих слів тільки від саме цих мотивуючих і зашкоджують утворенню від інших. Обмеження, які можна пояснити властивостями мотивуючого, є вихідними. Існують обмеження, обумовлені можливостями ймовірного мотивованого, або обмеження, орієнтовані на результат: у лексиконі є слова, формально тотожні можливим новоутворенням [8, с. 116-130].

Розглянемо принципи визначення словотвірної моделі за методом аналогії на прикладі іменника української мови. Для формування списку продуктивних моделей був використаний морфологічно розмічений корпус текстів. На основі інверсійного словника та завдяки розробленій пошуковій системі АРМ-лінгвіста визначалися регулярні моделі суфіксального словотворення [8]. Для формування словника суфіксів ми пропонуємо розроблену базу даних.

Так, база суфіксів іменників, розроблена для сучасної української мови, містить 556 моделей, охоплюючи близько 50 % досліджуваного матеріалу. Крім того, 146 моделей є неоднозначними, тобто можуть мати більше, аніж одне словотвірне значення у діапазоні від 1 до 8 одиниць. Для виявлених моделей було встановлено 236 словотвірних формантів, за допомогою яких і функціонують визначені моделі. Як один із структурних компонентів були виділені і словотвірні типи – 134. У таблиці 1 подано формат бази даних суфіксів, розробленої для української мови.

Таблиця 1

**База даних суфіксів сучасної української мови (фрагмент)**

Поле2	Поле 3	Поле 4	Поле 5	Поле 6	Поле 7	Поле 8	Поле 9
Афікс_1	Афікс_2	Гр. зн_1	Гр. зн_2	Сл. тип	Сл. мод.	Обме- ження	Приклади
ння	ти	іс	дн	одрд			оспікування
ість	ий	іж	п	вс			радість, веселість
ння	ти	іс	дд	одрд			командування
ка	а	іж	іж	змп			садибка, шубка

ниця	ник	іжі	ічі	ож	різни-	вєслувальниця,
					ця	основоположниця
ація	увати	іж	дн	одрд	ситу-	апробація, моти-
					ація,	вація
					пунк-	
					туація	

База даних містить інформацію про план вираження та план змісту словотвірної моделі, а саме: словотвірний формант мотивованого слова (поле 2), афікс, який необхідно додати для одержання пошукового слова (поле 3), інформацію про лексико-граматичне значення похідного слова (поле 4), про граматичне значення твірного слова (поле 5), інформацію про словотвірний тип, модель словотворення (відповідно поля 6, 7), виявлені дані про обмеження (поле 8).

Розглянемо процедуру визначення словотвірної моделі, словотвірного значення. Так, наприклад, обрана словотвірна категорія «діяч» реалізується моделями словотвірного типу «діяч», де твірна основа – дієслово недоконаного виду (54 моделі). Найрегулярнішими тут є моделі із словотвірним формантами *-ач*, *-альник*, *-атор*, *-итель*, *-дник*, *-івник*, з твірними основами дієслів недоконаного виду на *-ати*, *-ати*, *-увати*, *-ити*, *-дувати*, *-увати* (відповідно). Словотвірне значення саме цих моделей сформоване так: «особа діяча, виконавець дії». Процедура визначення моделі включає: виділення словотвірного форманта, визначення твірної основи, визначення словотвірного форманта для твірної основи, що дає змогу однозначно утворити лексему. Для визначення словотвірного форманта необхідною передумовою є лексико-граматична інформація досліджуваної лексики, що як план змісту представлена у базі даних.

Застосований лінгвістом-аналітиком операційний підхід до визначення словотвірного форманта та твірної основи моделі накладає певні вихідні лінгвістичні обмеження. Так, для розглянутої вище моделі типу «діяч» із словотвірним формантом *-ач* важливою є інформація про лексико-граматичні властивості аналізованого слова – іменник, чоловічий рід, однина, істога, лексико-граматична інформація про твірну основу – дієслово, недоконаний вид. Якщо ж, наприклад, усунути один із компонентів плану змісту, скажімо, «істога», то ця модель не відповідатимуть словотвірному значенню «особа діяча, виконавець дії», оскільки до неї входитимуть слова, що можуть позначати предмети (*передавач*).

У моделях типу «діяч» зі словотвірним значенням «особа діяча, носія ознаки», утворених від прикметників, вихідним обмеженням стає саме інформація про твірну основу – прикметник. Наприклад, у моделі із словотвірним формантом *-ник* (*каверзник*, *підлабузник*). Таку ж обмежувальну силу має і довжина словотвірного форманта як мотивованого, так і мотивуючого слова. Наприклад, словотвірні моделі із формантом *-ець*: *-ець*, де формант твірної основи *-ний* (*правдомовець*, *співомовець*), *-вець*, з формантом твірної основи *-вий* (*сміливець*, *вродливець*). Значення твірної основи «дієслово недоконаного виду» та «дієслово доконаного виду» та відповідні форманти – *-ант*, *-унт* (для дієслів недоконаного виду), *-ун*, *-ант*, *-ент* (для дієслів доконаного виду) – уможливають встановлення іншого словотвірного значення

для лексем типу *‘інформант, конкурент, опонент, стрибун’* – «особа діяча, носія процесуальної ознаки за дієсловом». Вихідне обмеження моделей «особа діяча, виконавця дії», що визначає значення твірного слова – іменник чоловічого роду, неістота, у поєднанні із компонентами словотвірної моделі (словотвірним формантом твірного та мотивованого слова) дозволяє розширити словотвірне значення – «особа діяча, виконавця дії, що стосується поняття (з вказівкою на поняття)». Для аналізованого нами типу обмеження, зорієнтовані на результат не є поширеними. Так, лише у моделі зі значенням «діяч, носій ознаки» із словотвірним формантом *–овик* (*строковик, бойовик*) система не допускає лексем: *більшовик, меншовик*.

Отже, обраний формальний підхід визначення словотвірної моделі дозволяє об’єктивувати процес словотворення, оптимізувати обсяг перекладного словника, застосувати наявні в лінгвістиці знання щодо досліджуваних процесів та відкрити нові можливості у вивченні словотвору на синхронному рівні.

### Література

1. *Белогогов Г. Г.* Метод аналогии в компьютерной лингвистике / Белогогов Г. Г., Зеленков Ю. Г., Новоселов А. Б., Хорошилов Ал-др, А., Хорошилов Ал-сей. А. // НТИ. Сер. 2. Информ. процессы и системы. – 2000. – № 1 – С. 21–30.
2. *Клименко Н. Ф.* Морфемно-словотвірний фонд української мови як дослідницька та інформаційно-довідкова система / Клименко Н. Ф., Карпіловська Є. А. // Мовознавство. – 1990. – № 4. – С. 41–50.
3. *Ковалик І.* Вчення про словотвір / І. Ковалик. – Л. : Вид-во Львівського ун-ту, 1961. – 82 с.
4. *Кубрякова Е. С.* Типы языковых значений: Семантика производного слова / Кубрякова Е. С. – М., 2008. – 208 с.
5. Словотвір сучасної української літературної мови / За ред. М. А. Жовтобрюха. – К., 1979. – 404 с.
6. *Степанова М. Д.* Вопросы моделирования в словообразовании и условия реализации моделей / Степанова М. Д. // Вопросы языкознания. – 1975. – № 4. – С. 24–37.
7. Тагабилева М. Г. Словообразовательная разметка Национального корпуса русского языка: задачи и методы / Тагабилева М. Г., Березуцкая Ю. Н. [Електронний ресурс]. – Режим доступу : <http://www.dialog-21.ru/dialog2010/materials/pdf/73.pdf>.
8. *Улуханов И. С.* Мотивация в словообразовательной системе русского языка / Улуханов И. С. – М. : Узд-ий центр «Азбуковник», 2005. – 314 с.
9. *Шипнівська О. О.* Місце словотворення в розвитку сучасного українського лексикону / Шипнівська О. О. // Мовознавство. – 2012. – № 1. – С. 55-64.
10. *Erjavec T.* Machine learning of morphosyntactic structure: lemmatizing unknown Slovene words / Erjavec T., Džeroski S. // Applied artificial intelligence. – Vol. 18. – 2004. – P. 17-41.

*Shypnivska O. The presentation of the derivation pattern in the knowledge-based machine translation system*

*In the paper the identification of the derivation pattern by the use of the analogy principle according to the knowledge-based machine translation system is considered. The theoretical principles of the determination of the derivation meaning of patterns are described.*

*Keywords: derivation pattern, derivation meaning, the principle of analogy, the knowledge-based machine translation system.*

***Шипновская О. А. Представление словообразовательной модели в знаниеориентированной системемашинного перевода***

*В статье рассмотрено определения словообразовательной модели по принципу аналогии согласно разработке знаниеориентированной системы машинного перевода. Представлено теоретические предпосылки извлечения словообразовательного значения.*

*Ключевые слова: словообразовательная модель, словообразовательное значение, принцип аналогии, знаниеориентированная система машинного перевода.*