

слово – за аналогією до вже наявних утворень чи на базі словосполучень.

<sup>1</sup>Русанівський В.М. Староукраїнська і слов'яноруська мови на тлі розвитку інших слов'янських літературних мов // Мовознавство. – 1997. – №1. – С. 3; <sup>2</sup>Огієнко І. Історія української літературної мови. – К., 1995. – С. 109; <sup>3</sup>Русанівський В.М. Історія української літературної мови. – К., 2001. – С. 101 – 102; <sup>4</sup>Житецкий П.Г. Очерк литературной истории малорусского наречия в XVII веке с приложением словаря книжной малорусской речи по рукописи XVII века. – К., 1889. – С. 3; <sup>5</sup>Русанівський В.М. Історія укр. літ. мови. – С. 103; <sup>6</sup>Огієнко І. Знач. праця. – С. 107; <sup>7</sup>Чепіга І.П. Концепція літературної мови староукраїнських книжників XVI – перш. пол. XVII ст. // Жанри і стилі в історії української літературної мови. – К., 1989. – С. 82; <sup>8</sup>Німчук В.В. Мовознавство на Україні в XIV–XVII ст. – К., 1985. – С. 198; <sup>9</sup>Русанівський В.М. Історія укр. літ. мови. – С. 70; <sup>10</sup>Огієнко І. Знач. праця. – С. 107; <sup>11</sup>Там само. – С. 108; <sup>12</sup>Там само. – С. 110; <sup>13</sup>Русанівський В.М. Історія укр. літ. мови. – С. 76; <sup>14</sup>Василевская Е.А. Словообразование в русском языке. – М., 1962. – С. 27; Кубрякова Е.С. Типы языковых значений: семантика производного слова. – М., 1981. – С. 64; <sup>15</sup>Троицкий В.П. Основные принципы словообразования // Уч. зап. І Ленингр. госуд. пед-та иностр. яз. – Л., 1940. – Т. 1. – С. 304.

*О.О. Шипнівська, м.н.с.*

## **КІЛЬКІСНА ТА ЯКІСНА ХАРАКТЕРИСТИКА ТИПІВ МОРФОЛОГІЧНИХ ОМОНІМІВ СУЧАСНОЇ УКРАЇНСЬКОЇ МОВИ**

*У статті розглядається якісна та кількісна характеристика типів морфологічних омонімів сучасної української мови.*

*The article deals with the qualitative and quantitative description of types of the morphological homonyms in the contemporary Ukrainian language.*

Одним із типів мовної неоднозначності є омонімія – явище яке притаманне будь-якому мовному рівню і є перешкодою для комунікативного процесу. З появою систем автоматичної обробки мовної інформації (писемних текстів) проблема омонімії як одного з типів мовної багатозначності набуває особливого значення.

Дослідження, що ставлять завдання ідентифікації словоформ одного слова – машинний переклад, автоматична корекція тексту, лінгвостатистика, автоматичне індексування – не можуть обійтися без автоматичного усунення омонімії текстових одиниць. Автоматичний синтаксичний та морфологічні аналізи вимагають однозначного визначення граматичного значення текстових словоформ. А тому проблемам автоматичного усунення граматичної омонімії як для української, так і для російської мов, присвячено багато робіт (роботи Т.О. Грязнухіної<sup>1</sup>, С.С. Білокриницької<sup>2</sup>, Ю. В Орехова<sup>3</sup>, Т. М. Молошної<sup>4</sup>).

Успішне розв'язання поставленого завдання можна здійснити на основі всебічного аналізу графічних омонімів та опису їх функціонування в текстах. Об'єктом даного дослідження є граматичні омоніми, тобто слова з однаковим графемним складом та різними граматичними значеннями.

Джерельною базою для виявлення морфологічних омонімів став словник словоформ, сформований на основі розробленого в Українському мовно-інформаційному фонді НАН України граматичного словника української мови. Реєстр словника обсягом 171 620 лексем (без урахування власних назв) за допомогою програми синтезу парадигм<sup>6</sup> було розгорнуто в список словоформ обсягом 3 094 174, який вважається достатньо репрезентативним для вирішення поставленої задачі.

Автоматичне визначення повного списку омографів проводилося шляхом використання розробленої в УМІФі програми автоматичного морфологічного аналізу. У розробленій системі морфологічного аналізу для експлікованого представлення граматичної інформації про належність неомонімічної словоформи до певного лексико-граматичного класу та конкретизації її граматичного значення в тексті використовуються двосимвольні коди. Перший символ такого коду вказує на частиномовну характеристику словоформи, другий символ коду вказує на конкретне граматичне значення. Для омонімічних одиниць тексту система автоматичного морфологічного аналізу визначає ланцюжок двосимвольних кодів, що відповідають усім теоретично можливим граматичним значенням словоформи, які можуть реалізуватися нею в мовленні. Наприклад, словоформа *вчений* має такий ланцюжок <АИЙИ>, що означає: АИ – ад'єктив, ЙИ – іменник чоловічого роду.

У системі для позначення лексико-граматичного класу використовуються такі символи: І – іменник, Д – дієслово, А – прикметник, Ч – числівник, М – займенник, О – займенник-прикметник, П – прийменник, С – сполучник, Н – вигук, Н – прислівник, Б – частка.

Окрім того, іменники диференціюються за родовою приналежністю, а тому мають свої символи: Й – іменник чоловічого роду, К – іменник жіночого роду, Л – іменник середнього роду та окремо розглядаються множинні іменники – И. Аналогічно має своє позначення дієслово згідно відповідно до видової приналежності: г – дієслова доконаного і Г – недоконаного виду.

Після морфологічної розмітки з 3 094 174 словоформ словника 1 554 372 словоформ одержали омонімічні коди. За допомогою програмних засобів з отриманого списку було виділено 610 612 омографів, опис їх здійснюється в термінах морфологічних типу та моделі. Омонімічні типи фіксують міжчастиномовну омонімію без врахування конкретного граматичного значення. Виразником конкретних граматичних значень, які можуть реалізуватися відповідними типами є моделі (див. вище наведений приклад омонімічного ланцюжка для словоформи *вчений*).

Як показали результати аналізу, міжчастиномовні омографи становлять 2 % (12 014) серед загальної кількості виділених за словником словоформ омонімів. Решта 98 % (588 598) – це омоніми в середині парадигми окремих граматичних класів. Опис їх здійснюватиметься лише на рівні моделей і не є предметом даного дослідження.

Виділена група міжчастиномовних омонімів розподіляється між 82 омонімічними типами, які репрезентують існуючі в українській мові омонімічні відношення між різними граматичними класами. Шість з них охоплюють 95 % усіх міжчастиномовних омографів. До них належать: АІ (*вартіві, наречені*), ДІ (*коти, лови*), АД (*бадьорим, білим*), ДН (*значимо, мислимо*), ІН (*візаві, ажур*), АН (*балакуче, благословенно*). Саме вони і можуть вважатися ядром морфологічних омонімів української мови. Всі вони представлені повнозначними частинами мови: змінними – іменником, ад'єктивом, дієсловом та незмінною – прислівником. Заслугує на увагу той факт, що при цьому мова використовує усі теоретично можливі комбінації цих класів.

Решта 76 типів знаходяться на периферії і становлять лише 5 % морфологічної омонімії. У цій групі 31 тип представлені лише одним словом. Наприклад, 17 трикомпонентних типів мають лише по одному омографу: ННЬ ‘геть’, НІЬ ‘ага’, НПЬ ‘на’, НІН ‘цабе’, АІС ‘тале’, ДПІ ‘край’, ДМО ‘тих’, ДСЬ ‘хоч’, ІНС ‘доки’, ІНЬ ‘гаразд’, МНО ‘всього’, МОС ‘та’, МСЬ ‘що’, НОБ ‘саме’, НПЬ ‘за’, АНД ‘ревно’, ‘гулі’ NAI.

У частини двокомпонентних та трикомпонентних типів з одним омографом (типи МН, НС, АО, МОЧ, ДІН) один з компонентів належить до діалектної лексики або є рідковживаним, застарілим, а тому вони можуть бути віднесені до непродуктивних омонімічних типів сучасної української мови. В системах АОТ, орієнтованих на обробку текстів, що не належать до художнього стилю або не є науковими розвідками з лінгвістики зазначеними типами можна взагалі знехтувати.

Усі виділені нами типи морфологічної омонімії з їх абсолютною частотою представлено в таблиці 1:

Таблиця 1.

тип	кількість омографів
АІ	5682
ДІ	2406
АД	2388
ДН	389
ІН	342
АН	272
МО	68
АДІ	68
НІ	67
НП	32
СЬ	30
ІО	23
ІП	20
АО	15
НС	14
МОЧ	14

ІЧ	12
МН	10
НЬ	9
ІМ	7
ІЬ	7
НД	7
ІНП	6
АІН	6
МЧ	6
НСЬ	5
НЬ	5
ІМО	5
НДІ	5
ІС	4
ОЬ	4
НО	4
ОЧ	4
НСЬ	4

НН	4
ДП	4
НЧ	3
ДН	3
ДНП	3
ДО	3
ДЬ	3
АЧ	2
ДС	2
НП	2
НСЬ	2
ОС	2
ДНО	2
МНЧ	2
МНО	2
ИСЬ	2
МОЬ	2
САЬ	1
МНОЬ	1
МОС	1
НПЬ	1
ННЬ	1
МНОСЬ	1
НЬ	1
НСЬ	1
НН	1
МСЬ	1
МЬ	1
НОЬ	1
НПЬ	1
НАИ	1
МС	1
ДП	1
ДНС	1
ДС	1
ДСЬ	1
ДЧ	1

АНД	1
АИС	1
МНОСЬ	1
ННЬ	1
ПСЬ	1
МНО	1
МНОС	1
ХА	1
ДМ	1
ДМО	1
НС	1

У відношення морфологічної омонімії вступають всі одинадцять лексико-граматичних класів: іменник, дієслово, ад'єктив, прислівник, сполучник, числівник, займенник, займенник-прикметник, прийменник, вигук, частка. Іменник, дієслово, прислівник, частка корелюють з 10 лексико-граматичними класами. Сполучник не вступає в омонімічні відношення з числівником, вигук – з займенником, займенник-прикметник – з прийменником та вигуком, ад'єктив – зі сполучником, займенником, прийменником, займенник – з ад'єктивом, прийменником, вигуком, прийменник – з ад'єктивом, числівником, займенником, займенником-прикметником, прийменником, числівник – із сполучником, прийменником, вигуком, часткою.

Аналіз омонімічних типів з погляду їхнього компонентного складу показав, що довжина омонімічних ланцюжків кодів варіюється в діапазоні від двох до шести. Дані про розподіл міжчастиномовних омографів за визначеною характеристикою подані у таблиці 2.

Таблиця 2.

Довжина омонімічного ланцюжка кодів	Абсолютна частота
двосимвольні	11904
трисимвольні	146
чотирисимвольні	9
п'ятисимвольні	1
шестисимвольні	1

Як видно з наведеної таблиці, більшість омографів представлені двокомпонентними омонімічними ланцюжками. Визначене нами ядро морфологічної омонімії також представлене саме двосимвольними кодами. Малочисельними є чотири-, п'яти-, шестикомпонентні типи. Чотирікомпонентні типи: 'як, так' ІНСЬ, 'тому, того' ІМНО, 'усього' МНОЬ, 'тим' МНОС, 'помість' ІПСЬ, 'коли' ДІНС, 'но' НІСЬ. П'ятикомпонентний МНОСЬ – 'все' та шестикомпонентний ІМНОСЬ – 'усе'.

Серед типів морфологічних омонімів, представлених у виділеному нами ядрі, найбільш продуктивним є тип ад'єктив / іменник з 5 682 омографами. Регулярність такого типу пояснюється поширеним в сучасній українській мові явищем субстантивації, адже саме такі словоформи і становлять основну частку цього типу: *безробітних, ведучого, вченого, шампанському, твердокрилі*'.

Для порівняння тип дієслово / іменник майже у 2 з половиною рази менший за описаний вище тип: *вигреби, видзвони, відділи, забіли винагороди, нагороди, погрози, бурю, мережу, неволю, тверджу*.

Двокомпонентні типи морфологічних омонімів, де у відношення морфологічної омонімії вступають змінні лексико-граматичні класи та прислівник, у виділеному нами ядрі не є такими продуктивними, як описані вище типи. Тип дієслово / прислівник має 388 омографів: *'загущено, благословенно, докупи, тривало, колись'*. Тип іменник / прислівник має 342 омографи: *'авансом, порою, світло'*. 272 омографи має тип ад'єктив / прислівник: *'гвалтовне, гостріше, мальовниче, найдешевше'*.

До периферії типів морфологічних омонімів належать як двокомпонентні, так і три-, чотири-, п'яти-, шестикомпонентні типи. 38 двокомпонентних типів охоплюють 428 омографів, 44 трикомпонентних – 156 омографи, 7 чотирикомпонентних – 9 омографів. П'яти та шестикомпонентні типи мають по одному омографу.

Таким чином, одержані результати буде прийнято за основу подальшого опису типів морфологічних омонімів на рівні представлення їх омонімічними моделями. Окрім того, вони слугуватимуть вихідними даними на етапі дослідження функціонування виділених типів омонімів в українських текстах для визначення контекстних диференційних ознак, що буде застосовуватися в алгоритмах автоматичного зняття граматичної неоднозначності в системах АОТ.

<sup>1</sup>Грязнухіна Т.О., Братищенко Л.Г., Дарчук Н.Б., Критська В.І., Пуздирева Т.К., Орлова Л.В. Шляхи уникнення омонімії в системі автоматичного морфологічного аналізу // Мовознавство. – 1989. – №5. – С. 3–12; <sup>2</sup>Белокриницкая С.С. Различные типы омонимии и способы их различия при машинном переводе // Вопросы языкознания. – 1960. – № 2. – С. 97–102; <sup>3</sup>Орехов Ю. В. Автоматическое различение омонимии словоформ // НТИ. – Сер. 2. Информационные процессы и системы. – М., 1977. – № 4. – С. 34–36; <sup>4</sup>Молошная Т. Н. Вопросы различения омонимов при машинном

переводе с английского языка на русский // Проблемы кибернетики. – Вып. 1. – М., 1958. – С. 213–221; <sup>5</sup>Малаховский Л.В. Теория лексической и грамматической омонимии. – Л., 1990. – 240 С; <sup>6</sup>Шевченко І. Алгоритмічна словозмінна класифікація української лексики // Мовознавство. – 1996. – С. 40–44.

*О.В. Бугаков, асп.*

## **АВТОМАТИЧНА ІДЕНТИФІКАЦІЯ СКЛАДЕНИХ ПРИЙМЕННИКІВ В УКРАЇНСЬКОМУ ТЕКСТІ**

*Подано формальні правила розмежування складених прийменників від синтаксичних конструкцій типу “однослівний прийменник + повнозначне слово” в тексті (модуль автоматичного контекстного аналізу).*

*Formal rules that distinguish complex prepositions from syntactic structures “one-word preposition + full meaning word” in text (module of automatic context analysis) are given.*

Розробка систем автоматичної обробки текстової інформації (АОТ), зокрема, автоматичного синтаксичного аналізу (АСА), вимагає розв’язання проблеми ідентифікації аналітичних конструкцій – сполучень повнозначного і службового слів, в яких службове слово виражає граматичне значення повнозначного і, відповідно, усієї конструкції в цілому<sup>1</sup>. Кількість компонентів може коливатися від 2 до 5: згідно з; з огляду на те що. До таких аналітичних конструкцій (“складені об’єкти” – Л.В.Ричкова<sup>2</sup>, “еквіваленти слова” – Р.П.Рогожнікова<sup>3</sup>, А.А.Лучик<sup>4</sup>) відносимо: аналітичні форми дієслова майбутнього часу (*буду писати*), минулого часу (*зробив був*), умовного способу (*пішов би*), описові форми ступенів порівняння прикметників і прислівників (*найбільш ерудований*), аналітичні сполучники (*у разі коли*), еквіваленти прислівників (*не в міру*), еквіваленти часток (*лише б*), складені прийменники (*відповідно до*) тощо. Останні і є об’єктом нашого дослідження.

За даними попереднього дослідження, проведеного на текстах обсягом 100 тис. слововживань, частота вживання складених прийменників (СП) серед усіх прийменникових вживань становить 3,2%.

У прийменникових сполученнях, що одержують статус прийменника, повнозначні слова в поєднанні з конкретними