

**СТВОРЕННЯ СЛОВНИКА МІЖЧАСТИНОМОВНИХ
МОРФОЛОГІЧНИХ ОМОНІМІВ СУЧАСНОЇ УКРАЇНСЬКОЇ МОВИ**

© Ольга Шипнівська, 2007

Український мовно-інформаційний фонд НАН України (Київ)
УДК 811.161.2

У статті пропонується спеціальна методика створення електронного словника міжчастиномовних морфологічних омонімів сучасної української мови на основі лінгвістичної бази даних. Розглядаються структура та головні функції такої лексикографічної системи.

Морфологічна омонімія як лексикографічна проблема є актуальною як для теоретичних досліджень, так і для практичних розробок¹. У теоретичному плані це стосується принципу інтегрального опису лексики будь-якої мови, за яким всі одиниці мови повинні мати належний опис у словнику, з можливістю створення загального вчення про функціонування мови в повному комплексі формальних та змістових засобів [6: 110; 3], у практичному – для побудови удосконалених спеціальних АОТ-орієнтованих словників та адекватних граматики.

Питання, пов'язані з лексикографічним описом омонімії, постали в зв'язку з необхідністю представлення цього явища в словниках і стосувались, передусім, проблем лексичної тотожності. Так, О. С. Ахманова зазначає: "Працюючи над створенням словників, я дійшла висновку, що загалом численні та складні лексикологічні питання можна звести до двох головних – проблеми "окремоті" та "тотожності слова" [2: 4].

Для розгляду дослідниками висувалися положення про можливість чи неможливість виділення омонімів у межах багатозначного слова, про необхідність встановлення критеріїв розмежування лексико-семантичних варіантів слова та омонімів. Розглядалися питання виділення та лексикографічного опису омонімів у тлумачних та спеціальних словниках [9: 34-36; 11; 8; 5; 11].

З розвитком комп'ютерної лексикографії з широким використанням електронних граматичних словників у системах автоматичної обробки текстової інформації задача формування електронних словників граматичних омографів стає надзвичайно актуальною. Для української мови такий словник розробляється вперше. Незначною мірою морфологічні омографи представлено в Словнику омонімів української мови О. М. Демської та І. М. Кульчицького як випадки лексико-граматичної омонімії [4].

Складність створення такого словника зумовлена відсутністю електронного граматичного словника, який представляє всю словозміну, та корпусу текстів.

У російській лексикографії створено два словники, в яких подано граматично неоднозначні слова: "Словарь омонимичных словоформ" Ж. Г. Аношкіної [1], "Словарь грамматических омонимов русского языка" О. М. Кім [7].

¹ У нашій роботі під морфологічними омонімами розуміємо графічно тотожні словоформи, які мають різні морфологічні значення.

Показовим у цьому випадку з огляду на спосіб представлення, обсяг та методику подання матеріалу є створений для російської мови Ж. Г. Аношкіною “Словарь омонимичных словоформ” – 3 400 омонімічних ряди [1]. Реєстр словника отримано внаслідок породження всіх словоформ із комп'ютерної версії Граматичного словника А. А. Залізняка.

У словнику подано омоформи-омографи, тобто форми різних (хоча часто близьких за змістом) слів, які мають однакове написання. Під час створення словника не враховувалася ознака ‘наголос’, яка у словнику А. А. Залізняка віднесена до словозмінних, оскільки словник зорієнтовано саме на писемний варіант мови.

Представляє як повну, так і часткову граматичну омонімію частин мови “Словарь грамматических омонимов русского языка” О. М. Кім [7]. Словник містить 5 000 омонімічних рядів.

Основою для автоматичного укладання електронного словника міжчастиномовних омографів української мови слугує лінгвістична база міжчастиномовних морфологічних омонімів української мови із залученням даних із лінгвістичної бази даних текстових морфологічних омонімів¹.

Список морфологічних омонімів української мови як таких, що задаються її граматичною системою, визначався на матеріалі словника словоформ, одержаного в результаті автоматичного синтезу парадигм, представлених у граматичному словнику української мови. Граматичний словник у повному обсязі представляє словозміну сучасної української мови, а тому подає репрезентативний лінгвістичний матеріал.

У результаті автоматичної морфологічної розмітки з 3.094.174 (171.620 лексем) словоформ словника 1.554.372 словоформам було поставлено у відповідність омонімічні коди за допомогою програми автоматичного морфологічного аналізу. На основі цього списку було виділено 610 612 омонімічних рядів.

У списку морфологічно неоднозначних одиниць словника (610 612 омографів) 98 % (600 409) становлять омографи в середині словозмінної парадигми одного лексико-граматичного класу, решта 2 % (10 203) – це морфологічні омографи між різними лексико-граматичними класами, які й стали об'єктом лексикографічного опрацювання. Міжчастиномовні морфологічні омоніми описуються 820 моделями (з 1486), які представляють 81 тип морфологічних омонімів².

¹ Про формування списку омонімів та лінгвістичної бази даних морфологічних омонімів докладно див. у роботах Шипнівська О. О. Формування реєстру омографів сучасної української мови // Мовні і концептуальні картини світу: Зб. Наук. пр. – Випуск 14. Книга 2. – К.: Вид. Дім Дмитра Бураго, 2004. – С. 255-259.; Шипнівська О. О., Стариков С. В. Формування лінгвістичної бази даних граматичних омонімів на матеріалі Українського національного лінгвістичного корпусу // Лексикографічний бюлетень: Зб. Наук. пр. / Відп. ред. д. філол. н., проф. В. Німчук. – К., 2006. – Вип. 14. – С. 188-194.

² Про характеристику типів міжчастиномовних морфологічних омонімів та специфіку їх функціонування в українських текстах докладно див. у роботах: Шипнівська О. О. Кількісна та якісна характеристика типів морфологічних омонімів сучасної української мови // Актуальні проблеми української лінгвістики: Зб. Наук. пр. – К.: Видавничо-поліграфічний центр „Київський університет”, 2005. – С. 36-42.;

Словник морфологічних омонімів сучасної української мови містить таку інформацію:

- реєстр міжчастиномовних морфологічних омонімів – близько 11 300 одиниць;
- кожному із представлених у реєстрі омографів поставлена у відповідність модель, що описує морфологічні характеристики його компонентів;
- для кожного компонента омонімічної моделі дається вихідна форма (вихідні форми виводилися автоматично за допомогою програми лематизації, розробленої в Українському мовно-інформаційному фонді НАНУ);
- про реалізацію реєстрових одиниць в Національному лінгвістичному корпусі обсягом 42 млн. слівформ (зірочкою позначені омоніми, які жодного разу не зустрілися в тексті).

Табличне представлення заданої інформації сформованого словника подано в табл. 1:

Таблиця 1. Табличне представлення електронного словника міжчастиномовних морфологічних омонімів

ММО	модель ММО	тип ММО	вихідні форми
багатію*	ГГЙДЙКЙП	ДІ	багатій багатіти
стелю	ГГКВ	ДІ	стелити стеля
ніж	ГЪЙВЙИСП	ДІС	ніж ніжити ніж

На рис. 1 подано електронну версію словника ММО. Вхід у словник здійснюється через реєстрову частину.

На рис. 2 подається формат словникової статті, доступ до якої здійснюється через одиниці реєстру міжчастиномовних морфологічних омонімів. Заголовковим словом статті є міжчастиномовний морфологічний омонім. Тлумаченням до омоніма є перелік тих граматичних значень, які він може мати. Цей перелік отримано автоматично за допомогою програми, яка трансформувала граматичні коди відповідного ланцюжка з омонімічного коду омографа. Стаття містить також інформацію про вихідні лексичні форми компонентів, представлених у коді.

Шипнівська О.О. Функціонування міжчастиномовних морфологічних омонімів в українських текстах // Мовознавство. – 2005. – №6. – С.70-80.

Реєстр	Частина мо...	Публ.	Наук.	Худ.
вдосконалений	А/Д		*	
вдосконалено	Д/Присл.			
вдягнений	А/Д	*		
веде	Д/І	*	*	*
ведений	А/Д	*		
веди	Д/І		*	*
ведмедеві	А/І			*
ведмедів	А/І	*		*
ведмежа	А/І	*		*
веду	Д/І	*	*	*
ведуча	А/І	*	*	
ведучий	А/І	*		
ведучим	А/І	*		
ведучими	А/І	*	*	
ведучих	А/І	*		
ведучі	А/І	*		
ведучій	А/І			
ведучім	А/І			
ведучого	А/І	*		
ведучої	А/І		*	
ведучому	А/І			
ведучою	А/І	*	*	
ведучу	А/І			
вельветові	А/І			
велюрові	А/І			
вензелеві	А/І			

Рисунок 1. Електронний словник міжчастиномовних морфологічних омонімів сучасної української мови

Омограф: **ведмедів**

Частина мови:

Ал'єктив
Іменник

Конкретне граматичне значення:

Ал'єктив, чоловічий рід, знахідний відмінок, однина
Ал'єктив, чоловічий рід, називний відмінок, однина
Іменник, чоловічий рід, родовий відмінок, множина
Іменник, чоловічий рід, знахідний відмінок, множина

Початкові форми:

ведмедів	0	Ал'єктив
ведмедь	0	іменник чоловічого роду, істота

Рисунок 2. Словникова стаття електронного словника міжчастиномовних морфологічних омонімів

Розроблена для словника пошукова система дає можливість формувати запити не лише за конкретним омонімом, а й за будь-якими видами інформації, представленої в словниковій статті: і значеннями граматичних класів, і конкретними граматичними значеннями компонентів омоніма.

Створений словник дає можливість формувати списки міжчастиномовних морфологічних омонімів за стилями (науковим, публіцистичним та художнім), списки тих омонімів, які реально зустрічаються в текстах. Наприклад, за словником можна виділити омоніми певного типу з характеристикою функціонування в текстах. На рис. 3 подано фрагмент словника, який показує омоніми типу дієслово/іменник та їхнє функціонування в текстах трьох стилів.

Реєстр	Частина мо...	Публ.	Наук.	Худ.
благодією	Д/І	*	*	
блюди	Д/І			
блюдодієм	Д/І			
блюдодій	Д/І			
блюдодію	Д/І			
блювала	Д/І			
блювали	Д/І			
бовванів	Д/І			*
бовкала	Д/І			
бовкало	Д/І			
болів	Д/І			*
бомби	Д/І	*	*	*
боре	Д/І			
бори	Д/І			*
борозни	Д/І	*		*
борони	Д/І			*
брати	Д/І	*	*	*
брежне	Д/І			*
бриджу	Д/І			
брила	Д/І	*	*	*
брили	Д/І	*		*
брило	Д/І			
брити	Д/І			

Рисунок 3.

Автоматичне генерування описаного словника за даними лінгвістичної бази даних є переконливим доказом доцільності використання спеціальних лінгвістичних баз даних як інструмента лінгвістичних досліджень.

Автоматично побудований на основі сформованих баз даних словник міжчастиномовних морфологічних омонімів обсягом близько 11 300 одиниць є інформаційною лексикографічною системою, розрахованою на широке коло запитів стосовно різних аспектів омонімії в граматичній системі української

мови та може слугувати для типологічних досліджень української та інших флективних мов. Створений словник є лексикографічним джерелом не тільки для фахівців, але й для тих, хто вивчає українську мову.

Література

1. Аношкина Ж. Г. Словарь омонимичных словоформ русского языка. – М., 2001 <http://www.irlras-cfrr.rema.ru/homofoms/>
2. Ахманова О. С. Очерки по общей и русской лексикологии. – М.: Гос. Учебно-педагог. изд-во Минист. просвещения РСФСР. – 1957. – 295 с.
3. Гак В. Г. Грамматика и тип словаря // Слово в грамматике и словаре. – М.: Наука, 1984. – С. 42-50.
4. Демська О. М., Кульчицький І. М. Словник омонімів української мови. – Львів: Фенікс, 1996. – 223 с.
5. Засорина Л. Н., Шашенкова Л. В. Много ли словарных грамматических омонимов в русском языке? (По материалам толкового словаря русского языка под ред. Д. Н. Ушакова) // Вопросы металингвистики. – М.: Изд-во Ленингр. ун-та, 1973. – С.103-110.
6. Караулов Ю. Н. Лингвистическое конструирование и тезаурус литературного языка. – М.: Наука, 1981. – 365 с.
7. Ким О. М., Островкина И. Е. Словарь грамматических омонимов русского языка. – М.: Астрель АСТ. Ермак, 2004. – 842 с.
8. Котельник Ф. С. Конверсионная омонимия и ее трактовка в словарях // Теоретические проблемы семантики и ее отражение в одноязычных словарях. – Кишинев: Штинца, 1982. – С. 144-148.
9. Палевская М. Ф. О лексико-грамматической омонимии и подаче лексико-грамматических омонимов в словаре // Тираспольский госуд. пед. ин-т. Ученые записки. – Вып. XI. – Кишинев: Штинца, 1960. – С. 107-115.
10. Тышлер И. С. Словарь лексических и лексико-грамматических омонимов современного английского языка. – Саратов: Изд-во Саратовского ун-та, 1975. – 378 с.
11. Шумарин С. В. Способы репрезентации функциональных омонимов в словарях: Автореф. дисс. на соиск. уч. ст. канд. филол. наук: 10.02.01 / Московский педагог. ун-т им. В. И. Ленина – М., 1992. – 16 с.