

*До 100-річчя
Національної академії наук України*

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
УКРАЇНСЬКИЙ МОВНО-ІНФОРМАЦІЙНИЙ ФОНД

ЛІНГВІСТИЧНО-ІНФОРМАЦІЙНІ СТУДІЇ

ПРАЦІ
УКРАЇНСЬКОГО МОВНО-ІНФОРМАЦІЙНОГО ФОНДУ
НАН УКРАЇНИ

У П'ЯТИ ТОМАХ

Дослідження 1991–2018 років

За загальною редакцією *В. А. Широкова*

КИЇВ 2018

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ
УКРАЇНСЬКИЙ МОВНО-ІНФОРМАЦІЙНИЙ ФОНД

ТОМ II

**ГРАМАТИЧНІ
СИСТЕМИ**

УКРАЇНСЬКИЙ МОВНО-ІНФОРМАЦІЙНИЙ ФОНД

*Рекомендовано до друку вченою радою
Українського мовно-інформаційного фонду НАН України
(протокол № 9 від 23 жовтня 2018 р.)*

Рецензенти:
академік НАН України *О. В. Палагін*,
член-кореспондент НАН України *А. О. Загнітко*

Автори другого тому:
*В. А. Широков, Т. П. Любченко, Є. В. Купріянов,
І. В. Шевченко, К. В. Широков*

Лінгвістично-інформаційні студії: праці Українського мовно-інформаційного фонду НАН України : у 5 т. / В. А. Широков та ін. Т. 2 : Граматичні системи. Київ. Український мовно-інформаційний фонд НАН України. 2018. 300 с.

Другий том монографії «Граматичні системи» присвячено викладу граматичних властивостей певного кола мов, який ґрунтується на розвинутому в попередньому томі формалізмі семантичних станів та теорії лексикографічних систем. Запропоновано факторизацію семантичного стану мовної одиниці на граматичну та лексичну компоненти. Аналіз граматичних станів здійснюється в межах, які залишають незмінною лексичну частину відповідного повного семантичного стану. В такий спосіб визначається коло граматичних аспектів, що складають предмет розгляду. Загалом дослідження концентрується навколо словозмінних властивостей мовних систем.

Формально визначено поняття словозмінної парадигми для флективних мов (із синтетичним компонентом у словозміні) – української та російської, а також поняття парадигматичного відношення та оператора парадигматизації. Побудовано лексикографічні словозмінні системи для вказаних мов. Реалізація названих систем на великих репрезентативних масивах лексичних даних уможливила побудови практично вичерпних словозмінних класифікацій – української та російської, а також дозволила створити відповідні віртуальні лексикографічні лабораторії. Аналогічні результати отримано для мов із аналітико-синтетичним компонентом у словозміні – німецької та іспанської.

Проведено дослідження іменної словозміни сучасної турецької мови, яка належить до класу аглютинативних. Побудовано загальну морфологічну структуру турецьких іменникових лексем. Із застосуванням формалізму семантичних станів створено теорію граматичної ідентифікації іменникових форм як таких, що перебувають у стані часткової суперпозиції (іменниково-дієслівної) граматичних станів. Такий підхід дозволив збудувати достатньо повну словозмінну класифікацію турецького іменника та відповідну граматичну Л-систему. На основі здобутих результатів створено віртуальну лексикографічну лабораторію турецької іменної словозміни, яку верифіковано на масиві понад 30 тис. лексем (близько 60 млн словоформ).

© Український мовно-інформаційний фонд
НАН України, 2018

© В. А. Широков, Т. П. Любченко,
Є. В. Купріянов, І. В. Шевченко,
К. В. Широков, 2018

ЗМІСТ

ВСТУП.....	7
Розділ 7. ЗАГАЛЬНІ ПИТАННЯ МОДЕЛЮВАННЯ ГРАМАТИЧНИХ СИСТЕМ ФЛЕКТИВНИХ МОВ.....	9
7.1. Граматична параметризація лексики і феномен словозміни.....	9
7.1.1 Феномен словозміни у формалізмі семантичних станів.....	9
7.1.2. Принципи побудови морфологічної (парадигматичної) лексикографічної системи для мови з флективним компонентом	10
Розділ 8. ГРАМАТИЧНІ СИСТЕМИ ДЛЯ МОВ ІЗ СИНТЕТИЧНИМ КОМПОНЕНТОМ У СЛОВОЗМІНІ.....	19
8.1. Українська та російська словозмінна класифікації та електронний граматичний словник	19
8.1.1. Парадигматичні типи та словозмінні параметри	20
8.1.2. Парадигматичні класи та відношення парадигматизації.....	30
8.1.3. Парадигматичні класи з урахуванням акцентуації	31
8.1.4. Дефектність та варіативність словозмінної парадигми	35
8.1.5. Оператор парадигматизації	40
8.1.6. Побудова лексикографічної системи ЕГС російської мови.....	41
8.1.7. Побудова лексикографічної системи ЕГС української мови	46
8.1.7.1. Загальна структура словозміни української лексики.....	47
8.1.7.2. Іменник	50
8.1.7.2.1. Система словозмінних параметрів іменника.....	50
8.1.7.2.2. Процедура визначення словозмінних параметрів іменника	59
8.1.7.3. Дієслово	62
8.1.7.3.1. Система словозмінних параметрів дієслова.....	62
8.1.7.3.2. Процедура визначення словозмінних параметрів дієслова	67
8.1.7.4. Прикметник	70
8.1.7.4.1. Система словозмінних параметрів прикметника	70
8.1.7.4. 2. Процедура визначення словозмінних параметрів прикметника.....	71
8.1.7.5. Алгоритмічна специфіка парадигматизації інших частин мови.....	72
8.1.8. Технологічні аспекти створення ЕГС.....	73
8.1.8.1. Структура бази даних (внутрішній рівень представлення).....	74
8.1.8.2. Програмні засоби для створення ЛБД російського ЕГС	78
Розділ 9. МОДЕЛЮВАННЯ СЛОВОЗМІННИХ СИСТЕМ ІЗ СИНТЕТИЧНИМ І АНАЛІТИЧНИМ СЛОВОЗМІННИМИ КОМПОНЕНТАМИ	83
9.1. Концептуальне моделювання словозмінної системи німецької мови.....	83
9.2. Л-система ЕГС німецької мови.....	92
9.2.1. Структура бази даних ЕГС німецької мови	92

9.2.2. Програмний інтерфейс для підготовки та редагування граматичної ЛБД	95
9.2.3. Інструментальні засоби для супроводу граматичної ЛБД німецької мови.....	98
Розділ 10. Л-СИСТЕМА СЛОВОЗМІННОГО СЛОВНИКА ІСПАНСЬКОЇ МОВИ	100
10.1. Загальнотеоретичні питання моделювання іспанської словозміни.....	102
10.1.1. Проблема визначення частин мови	103
10.1.2. Формальне визначення поняття словозміни.....	106
10.1.3. Квазіоснови та квазіфлексії	109
10.2. Парадигматична класифікація іспанської лексики	112
10.2.1. Парадигматичні типи та їхні характеристики	112
10.2.2. Номінативний парадигматичний тип	120
10.2.3. Дієслівний парадигматичний тип	125
10.2.4. Займенниковий парадигматичний тип	139
10.2.5. Детермінантний парадигматичний тип	143
10.3. Побудова Л-системи словозмінного словника іспанської мови	145
10.3.1. Граматичні стани як об'єкт граматичного словника	145
10.3.2. Концептуальна модель Л-системи граматичного словника іспанської мови.....	150
10.3.3. Сигнатура словозмінної парадигми.....	161
10.3.4. Лексикографічна параметризація словозміни	163
Розділ 11. ГРАМАТИЧНІ СИСТЕМИ ДЛЯ МОВ З АГЛЮТИНАТИВНИМ КОМПОНЕНТОМ У СЛОВОЗМІНІ	167
11.1. Проблема граматичної кваліфікації частин сучасної турецької мови	167
11.2. Морфологічна структура іменника сучасної турецької мови	178
11.3. Застосування теорії семантичних станів до граматичної ідентифікації турецької лексики.....	185
11.4. Формальні засади побудови словозмінних класифікацій	195
11.5. Морфонологічні та морфотактичні принципи словозмінної класифікації турецької лексики	199
11.5.1. Формалізація словозміни турецького іменника та виділення системи його парадигматичних класів.....	203
11.5.2. Виділення підкласів у словозмінних класах турецького іменника на базі процесів в основі	206
11.5.3. Інші процеси в системі відмінювання турецького іменника.....	213
11.6. Концептуальна модель, структура бази даних та зовнішній інтерфейс електронного граматичного словника турецької мови	214
ДОДАТКИ	222
ДОДАТОК А. Таблиці прикладів української словозміни.....	222
ДОДАТОК Б. Приклади парадигматичних класів української словозміни	235
ДОДАТОК В. Таблиці прикладів російської словозміни.....	236
ДОДАТОК Г. Відношення між елементами множин Ω , T , P , C , F , Π	245
ДОДАТОК Д. Таблиця переходів від класифікації А. Залізняка до класифікації, прийнятої в Л-системі ЕГС російської мови (для іменників).....	251
ДОДАТОК Е. Парадигми у німецькій мові	276
ЛІТЕРАТУРА.....	291

ВСТУП

Грамматична проблематика останнім часом цікавить не лише теоретиків мовознавства, які зобов'язані нею займатися за службовими, так би мовити, обов'язками, але все більше – прикладників, які створюють системи комп'ютерного опрацювання мови. В обшир цього опрацювання входить невизначено велике коло завдань, пов'язаних з інформаційним пошуком, перекладом з однієї мови на іншу («з усіх на всі»), інженерією знань та лінгвістичною когнітологією, дослідженням системних зв'язків у тріаді «інформація – мова – інтелект», розпізнаванням і синтезом мовних і текстових образів, інтерпретацією змісту текстів, проведенням різнопланових лінгвістичних експертиз, навчанням мови, створенням різного роду віртуальних систем професійної лінгвістичної взаємодії, розробленням корпусних лінгвістичних технологій і багато іншого.

Грамматичний опис для окресленого кола завдань становить інтерес, переважно, в тому сенсі, що він забезпечує інтерпретацію мовних об'єктів з боку форми, що принципово важливо з погляду формалізації цього опису, оскільки формалізація (і якомога глибша) є необхідною умовою створення ефективних, наділених інтелектуальними властивостями артефактів, які здійснюють мовне опрацювання. Зі сказаного випливає необхідність розроблення універсальної формалізованої граматичної концепції, яка б охоплювала максимально можливе коло мовних феноменів і ефективно працювала на повному обсязі мовних одиниць, а не на окремих, хоча й показових прикладах.

Автори усвідомлюють, що висловлюватися на тему фундаментальних початків граматики – науки, історія якої налічує не одну тисячу років, – не зовсім безпечно. Проте автори мають свій власний, у чомусь унікальний досвід граматичного опрацювання лексики різних мов і ними створені дуже великі, справді репрезентативні граматично анотовані комп'ютерні масиви лексики, що надає матеріал до певних граматичних узагальнень концептуального характеру. Зазначені узагальнення будуються на ґрунті феноменологічного підходу до вивчення природних систем, до яких ми відносимо і природну мову, наукові засади якого викладено в першому томі нашої монографії. З іншого боку, вони, як було докладно розібрано у цій праці, мають своїх попередників у особі видатних російських учених А. М. Колмогорова, В. А. Успенського та А. А. Залізняка, в працях яких, написаних ще в далекі 50–60-ті роки минулого століття, сформульовано початки нового наукового підходу до граматичного опису

лінгвістичних феноменів, засновані на понятті стану мовної одиниці¹. На жаль, ініціатива згаданих учених свого часу не одержала розвитку і була незалежно перевідкрита у вигляді теорії семантичних станів мовних одиниць одним із авторів цієї книги (*В.Ш.*²) в інший час і з інших міркувань. Мета нашої праці – продемонструвати феноменологічність граматики природної мови і, отже, підтвердити близькість науки про мову до інших дисциплін природничо-наукового, а отже і технологічного циклу. Актуальність цього завдання ми відчуваємо у зв'язку з тим навантаженням, що лягає нині на лінгвістичну технологію і нерозривно пов'язану з нею інженерію знань, яка все виразніше набуває рис глобальної індустрії й вимагає для свого розвитку все більш витончених і потужних інтелектуальних мовних інструментів.

На закінчення додамо, що наші спостереження над феноменологічними аспектами граматики переконують нас у тому, що мова і наука про мову мають набагато більше спільного з природними явищами і природничими науками, ніж це прийнято вважати в традиційній лінгвістиці. Хочеться висловити переконання, що надалі ця спільність набуде набагато більш певних рис.

¹ Успенский В. А. К определению падежа по А. Н. Колмогорову. В сб. «Бюллетень Объединения по проблемам машинного перевода». № 5. Москва: [И МГПИИЯ], 1957. С. 11–18; Зализняк А. А. Русское именное словоизменение. Москва : Наука, 1967.

² Широков В. А. Лексикографічне представлення семантичних станів. Математические машины и системы. Київ, 1999; Широков В. А. Гуманітарна традиція і технологічний статус мови. Мовознавство. 2001. № 3. С. 128–132; Широков В. А. Феноменологія лексикографічних систем: монографія. НАН України. Укр. мов.-інформ. фонд. Київ : Наук. думка, 2004. 327 с.