

*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 I

**НАУКОВА ПАРАДИГМА
ТА ОСНОВНІ
МОВНО-ІНФОРМАЦІЙНІ
СТРУКТУРИ**

УКРАЇНСЬКИЙ МОВНО-ІНФОРМАЦІЙНИЙ ФОНД

УДК 061.1:81(477)(082)

DOI 10.33190/978-966-02-8683-2/8684-9

*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 першого тому:

*В. А. Широков, Т. О. Грязнухіна, Т. П. Любченко, І. В. Остапова,
І. В. Шевченко, Л. Л. Шевченко, К. В. Широков*

Лінгвістично-інформаційні студії: праці Українського мовно-інформаційного фонду НАН України: у 5 т. / В. А. Широков та ін. Т. 1 : Наукова парадигма та основні мовно-інформаційні структури. Київ. Український мовно-інформаційний фонд НАН України. 2018. 271 с.

У першому томі «Лінгвістично-інформаційних студій» викладено засади наукової парадигми Українського мовно-інформаційного фонду, а саме: послідовне втілення погляду на мову як інформаційну систему певного типу, феноменологічний підхід до вивчення мовних явищ та еволюціонізм, який спонукає розглядати мовні феномени не в їх окремішності, а як елементи Всесвіту, що постійно змінюється, підкоряючись певним всезагальним закономірностям.

Сформульовано сім лінгвістичних презумпцій – загальних тверджень про природу мови, якими автори керуються, будуючи свої теоретичні концепції та їх застосування. Викладено авторську інтерпретацію системного підходу, яка ґрунтується на експлікації та взаємодії структурних, субстанційних та суб'єктних властивостей досліджуваних мовних систем.

На основі зазначених фундаментальних настанов сформульовано універсальний феноменологічний принцип – лексикографічний ефект в інформаційних системах. Цей ефект висвітлює механізм породження з континууму дискретних комплексів відносно стійких семіотично і семантично визначених сутностей, котрі, комбінуючись, породжують феноменальні прояви досліджуваних об'єктів. Інформаційно-феноменологічний підхід та формалізація проявів лексикографічного ефекту в застосуванні до мовних систем приводять авторів до формулювання теорії лексикографічних систем та теорії семантичних станів, які надають потужні концептуальні інструменти до пояснення великого масиву мовних явищ та побудови високоефективних лінгвістичних технологій.

Для фахівців у галузі теоретичної, математичної та прикладної лінгвістики, інформаційних та лінгвістичних технологій, аспірантів і студентів вишів. Книга може бути цікавою для фахівців у галузі антропології культури, когнітології, прикладної семіотики та системного аналізу.

© Український мовно-інформаційний фонд
НАН України, 2018

© В. А. Широков, Т. О. Грязнухіна,
Т. П. Любченко, І. В. Остапова,
І. В. Шевченко, Л. Л. Шевченко,
К. В. Широков, 2018

ISBN 978-966-02-8683-2

ISBN 978-966-02-8684-9 (Т. 1)

ЗМІСТ

ВСТУП.....	7
ЧАСТИНА 1. НАУКОВА ПАРАДИГМА УКРАЇНСЬКОГО МОВНО-ІНФОРМАЦІЙНОГО ФОНДУ НАЦІОНАЛЬНОЇ АКАДЕМІЇ НАУК УКРАЇНИ	21
Розділ 1. ПРО ІНФОРМАЦІЙНУ ДОКТРИНУ, ЯКУ АВТОРИ ВВАЖАЮТЬ БЛИЗЬКОЮ ДО СВОЇХ УЯВЛЕНЬ ПРО СВІТ І ПРО МОВУ. АБО, ІНШИМИ СЛОВАМИ, ПРО ЇХ ІНФОРМАЦІЙНИЙ СВІТОГЛЯД	22
1.1. Наукове поняття інформації та інформаційно-енергетичні взаємодії	22
1.2. Алгоритмічна складність А. М. Колмогорова і пов'язана з нею інформаційна міра	33
Розділ 2. ЛЕКСИКОГРАФІЧНИЙ ЕФЕКТ В ІНФОРМАЦІЙНИХ СИСТЕМАХ	42
Розділ 3. ФЕНОМЕНОЛОГІЧНИЙ ПІДХІД ДО ВИВЧЕННЯ МОВНОЇ СИСТЕМИ	54
3.1. Логіко-феноменологічні основи опису мови	54
3.2. Феноменологічна картина мовного світу та її концептуальне представлення	59
Розділ 4. СИСТЕМНІСТЬ У МОВІ ТА ЛІНГВІСТИЦІ	65
4.1. Структура, субстанція і суб'єкт у визначенні поняття системи в лінгвістиці	65
4.2. Системні прояви в мовній феноменології	68
ЧАСТИНА 2. ОСНОВНІ МОВНО-ІНФОРМАЦІЙНІ СТРУКТУРИ ТА ЇХ ТЕОРЕТИЧНИЙ ОПИС	71
Розділ 5. ТЕОРІЯ СЕМАНТИЧНИХ СТАНІВ МОВНИХ ОДИНИЦЬ	72
5.1. Перше уявлення про лінгвістичні стани і пов'язане з ними поняття лінгвістичних спостережуваних.....	72
5.2. Подальша формалізація лінгвістичних станів	92
5.3. Лінгвістичні стани у формалізмі нечітких множин	97
5.4. Формалізований підхід до визначення семантичних станів слів	99
5.5. Зауваження щодо концепції станів та феноменологічної складності устрою світу	108
Розділ 6. ТЕОРІЯ ЛЕКСИКОГРАФІЧНИХ СИСТЕМ	115
6.1. Структура і архітектура лексикографічних систем	115
6.2. Лексикографічні середовища	133
6.3. Інтегровані Л-системи та методика їх побудови	135

6.4. Індекси лексикографічних систем.....	140
6.5. Лексикографічні системи і словники.....	141
6.6. Семантичні стани та лексикографічні числення.....	149
6.6.1. Основні положення теорії числень.....	149
6.6.2. Визначення лексикографічного числення.....	155
6.6.3. Синтез семантичних станів та лексикографічних числень.....	159
6.6.4. Узагальнена лінгвістична змінна.....	160
ДОДАТОК	
Еволюція як універсальний природний закон.....	168
СПИСОК ЛІТЕРАТУРИ.....	264

ВСТУП

1.

Публікація багатотомного наукового видання завжди потребує певного роз'яснення намірів автора або авторського колективу: з якою метою була виконана велика за обсягом робота, які принципи покладено в її основу, яку аудиторію мали на меті автори і, зрештою, якої реакції вони чекають на свою публікацію.

Автори свідомі того, що зараз, на початку викладу, їм навряд чи вдасться сформулювати вичерпні концептуально-логічні відповіді на ці питання. Тим паче, що поява нашої праці мала й певне зовнішнє стимулювання (про це ще йтиметься далі), аж ніяк не пов'язане із логікою розгортання наукової парадигми, що становить зміст роботи. Внаслідок сказаного ми обираємо історико-логічний спосіб викладу і почнемо з історії заснування і становлення Українського мовно-інформаційного фонду Національної академії наук України, маючи сподівання, що в ході цього викладу тією чи іншою мірою будуть прояснені питання, які ми поставили вище.

Український мовно-інформаційний фонд НАН України було засновано в березні 1991 року на останній хвилі наукової програми Державного комітету СРСР з науки і техніки «Створення машинного фонду російської мови та машинних фондів мов народів Радянського Союзу». Зараз, коли пройшло понад 30 років (а згадана програма була заснована ще у 1986 році), ця проблематика, як це не дивно, зовсім не здається несучасною або застарілою. Проблема створення «машинних фондів мов» – а нині їх заступили лінгвістичні корпуси та відповідні системи аналізу природномовної інформації – досі залишається актуальною. Вона потребує для свого розв'язання значних обчислювальних ресурсів, а головне – адекватних і ефективних формалізованих моделей природної мови.

Питання про масштаби обчислювальних ресурсів є більш-менш зрозумілим, адже зазначені масштаби визначаються, насамперед, обсягами репрезентативних текстів, які підлягатимуть опрацюванню, а вони є надзвичайно великими.

Стосовно другого питання, а саме – адекватних і ефективних формалізованих моделей природної мови, то воно виявляється набагато складнішим і потребує певних додаткових міркувань для розуміння і раціональної постановки відповідних дослідницьких завдань. На наше переконання ці аспекти зосереджені в комплексі системно-концептуальних відношень тріади «Мова – Інформація – Знання». Саме в цьому комплексі містяться і парадигмальні інваріанти наукової діяльності Фонду. Розглянемо це питання докладніше.

Науково-дослідний підхід Українського мовно-інформаційного фонду Національної академії наук України за своїм фундаментальним спрямуванням відрізняється від більшості традиційних мовознавчих установ. В основу його наукової парадигми покладено *три базових принципи*:

Послідовне втілення підходу до мови як до інформаційної системи певного типу, що диктує необхідність оперування фундаментальними визначеннями наукових понять *інформації, знання, системи* та їхніми деталізаціями, які визначаються специфікою об'єкта «природна мова».

Феноменологічний підхід до вивчення мовних явищ, який забезпечує епістемологічний баланс між емпіризмом та метафізикою і, в такий спосіб, вводить методологію лінгвістики до кола, близького методологічним засадам природничих наук.

Еволюціонізм, який спонукає розглядати мовні феномени не в їх окремішності, а як елементи Всесвіту, що постійно змінюється, підкоряючись певним всезагальним закономірностям.

Викладові перших двох принципів (інформаційному і феноменологічному) будуть присвячені окремі розділи нашої праці. Останній – еволюційний принцип, унаслідок великого обсягу матеріалу та з метою забезпечення цілісності й неперервності викладу, ми окреслимо в основному тексті лише контурно, а