

# ПЕРСПЕКТИВНІ НАПРЯМКИ РОЗВИТКУ НАУКОВОЇ ДУМКИ В УКРАЇНІ

*В.А. Широков, д.т.н., О.Г. Рабулець*

## **ФОРМАЛІЗАЦІЯ В ГАЛУЗІ ЛІНГВІСТИКИ**

*Розглядаються сучасні аспекти формалізованих моделей представлення національної мови.*

*Considered are modern aspects of formalized models of representation of a national language.*

*Я вірю, що будь-яка наукова думка,  
яка досягла рівня включення її  
в наукову теорію, підпадає під  
вплив міцї аксіоматичного методу  
й тим самим математики.*

**Д. ГІЛЬБЕРТ**

### **1. ІНФОРМАЦІЯ І МОВА**

Уже давно не прийнято розглядати мову як абстрактну «в себе і для себе» сутність. Природна історія мови, її буття як універсальної інформаційної субстанції все наполегливіше поляризує дослідження в напрямі вивчення, абстрагування, фіксації і побудови комплексних інформаційних моделей даних, елементами котрих є як власне структури мови, так і структури мислення. Метою відзначених побудов *par excellence* є те, що пов'язано з пізнавальною діяльністю людини: одержанням, передаванням, трансформацією знань тощо. Таким чином, когнітивний аспект лінгвістики в наш час перетворився на провідний у прикладних галузях мовознавства.

Масштаби когнітивного освоєння дійсності є такими, що, за нашими уявленнями, вони наполегливо потребують моделювання мовних явищ не у фрагментарному, але репрезентативному, цілісному обсязі, який у праці<sup>1</sup> було кваліфіковано як *основний когнітивний тракт*, що включає такі фази (етапи, блоки ...):

СПРИЙНЯТТЯ → ВІДЧУТТЯ → ПЕРЕЖИВАННЯ →  
УСВІДОМЛЕННЯ → РОЗУМІННЯ →  
РЕФЛЕКСІЯ → РЕАКЦІЯ.

У той же час прикладна спрямованість когнітивно-лінгвістичних досліджень визначає методи і навіть сам стиль робіт у цій галузі, які диктують необхідність залучення максимально формалізованого апарату. У цьому немає нічого виняткового. Історія мовознавства значною мірою є історією формалізації уявлень про природу мови. Свідомо або ні, але зусилля лінгвістів протягом століть були спрямовані на досягнення саме цієї мети.

Наука в цій галузі ще у давнині мала вражаючі досягнення – згадаємо хоча б створення санскриту. Поза сумнівом, що це потребувало високого рівня рефлексії у його винахідників. Набагато пізніший приклад – спроба Лейбніца створення досконалої мови уже може слугувати прообразом сучасних підходів до формалізації мовних феноменів. Ще пізніші спроби створення штучних мов (есперанто, волапук, ...) свідчать про те, що такі спроби, по суті, не припинялися ніколи.

Із початком комп'ютерної цивілізації формально-мовна проблематика одержала потужний імпульс розвитку, правда, на наш погляд, дещо однобічний, який спрощено трактує мовні явища.

Спробуємо розібратися, у чім тут справа. Наведемо думку Д.С. Лихачова<sup>2</sup>, котру можна було б поставити епіграфом до даного розділу:

*“У літературознавстві існує своєрідний комплекс власної неповноцінності, викликаний тим, що вона не належить до кола точних наук. Вважається, що висока ступінь точності в будь-якому випадку служить ознакою науковості. Звідси різні спроби підпорядкувати літературознавство точній методиці дослідження і пов'язані з цим обмеження діапазону літературознавства, що надають йому більш або менш камерного характеру.*

*Як відомо, для того, щоби наукова теорія могла вважатися точною, її узагальнення, висновки, дані повинні спиратися на певні*

однорідні елементи, з якими можна було б виконувати різні операції (комбінаторні, математичні в тому числі). Для цього досліджуваний матеріал повинен бути формалізований.

Оскільки для точності потрібна формалізація обсягу дослідження й самого дослідження, усі спроби створити точну методiku проведення досліджень у літературознавстві так чи інакше пов'язані із прагненням формалізувати матеріал літератури. І в цьому прагненні, хочу це підкреслити з самого початку, немає нічого одіозного. Формалізується будь-яке знання, і будь-яке знання саме формалізує матеріал. [підкреслення наше (В.Ш.)] Формалізація стає неприпустимою тільки тоді, коли вона насильно приписує матеріалу той ступінь точності, котрої він не має і по суті своїй не може мати."

Незважаючи на те, що наведене висловлювання стосується літературознавства, його з повною підставою можна адресувати і до мовознавства. А у ньому справа виглядає так.

Історія лінгвістики фіксує виділення з мовного континууму цілого ряду дискретних одиниць, які цей континуум складають, формують, утворюють. Цей процес супроводжується створенням й цілої низки моделей. Протягом багатьох століть було розвинено розгалужений концептуальний апарат лінгвістики. Тим не менше формальну теорію природної мови досі не побудовано. Можна поставити питання так: чому досі не винайдено математичну теорію природної мови, котра по відношенню до нього відіграла б роль, аналогічну тій, яку відіграло, приміром, диференційне й інтегральне числення по відношенню до фізичної науки?

Зрозуміло, спроби створення формального апарату лінгвістики неодноразово вживалися в історії мовознавства і з ними навіть пов'язаний відповідний термін – структурно-математична лінгвістика. Але, зізнаємося, у філології апарат структурно-математичної лінгвістики, теорія формальних мов, теорія формальних граматик і різноманітні їх узагальнення не відіграли ролі, подібній до ролі математичного аналізу й теорії диференційних рівнянь у фізиці.

Це сталося, мабуть, тому, що конструкції структурно-математичної лінгвістики не торкаються системотвірних властивостей природної мови і, по суті, являють ніщо інше як

своєрідні способи подання поняття алгоритму. І незважаючи на те, що природна мова чимось нагадує алгоритм (хоча б тим, що вона, так само як і алгоритм, розгортається лінійно), інтуїтивне почуття опирається проти визначення мови як алгоритму.

У наш час потреби у формалізації мови надзвичайно зросли. Це пов'язано із бурхливим розвитком комп'ютерних технологій та комунікацій, де природна мова все більш очевидно набуває рис основного інформаційного інструменту. Виявилося, що навчити комп'ютери розуміти природну мову практичніше, ніж перековувати на програмістів десятки й сотні мільйонів простих людей.

Таким чином, у результаті бурхливої інформатизації суспільства постала проблема побудови такої формальної теорії мови, котра була б спроможна адекватно відобразити її інтелектуальні, когнітивні функції, з урахуванням феноменів розуміння, значення, смислу, комунікації, навчання, передання знань і т. ін., на основі чого стала б можливою розробка засобів лінгвотехнології, спроможних виконувати точний автоматичний переклад з одної мови на іншу, екстракцію і локалізацію знань у текстах, виконувати автоматичний логічний аналіз природномовних масивів, забезпечувати інтелектуальний природномовний діалог людини з комп'ютером та інші інтелектуальні операції з об'єктами природної мови. Більш того, зараз необхідні комп'ютери з розвиненим сенсорно-перцептивним апаратом, можливостями поєднання логічного і образного аспектів мислення. Потреби сучасної технології диктують необхідність вищого ступеня інтеграції інтелектуальних функцій в людиномашинних системах, а це породжує цілу низку проблем – як науково-технічного так і гуманітарного порядку.

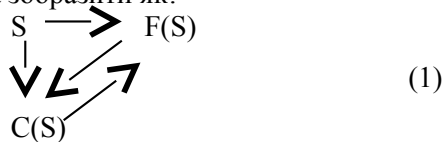
Не можна сказати, що ця проблематика є зовсім новою. Починаючи з 50-х років ХХ століття, у даній галузі відбувається справжній бум. Праці Хомського ініціювали цілий ряд напрямів у галузі формалізації синтаксичної структури мови. Виникло і сформулося поняття формальних граматик і його узагальнення, котрі, як уже відзначалося, лягли в основу цілого наукового напрямку – структурно-математичної лінгвістики. Були проведені значні дослідження статистичних властивостей мови. Роботи в галузі

машинного перекладу привели до створення теорії мови-посередника. Значні зусилля були спрямовані на формалізацію семантичної теорії. Серед помітних результатів даного напрямку відзначимо моделі «Смисл–Текст» Мельчука, модифікації яких використовуються і в наш час. Різний час ставить різні проблеми. У 70-і роки сформувалася когнітивна лінгвістика – напрям на межі когнітології і філології, який пов’язує теорію мови і теорію знань. Все більшого значення надається дослідженню взаємозв’язку механізмів розуміння, значення, смислу. На початку 80-х років все більшу увагу спеціалістів привертає сформульована Шенком гіпотеза інтегральної обробки, згідно з якою розуміння природної мови відбувається за одночасного опрацювання її семантичних, синтаксичних і когнітивних структур. У наш час ця доктрина визнається домінуючою в когнітивній і комп’ютерній лінгвістиці. Вона інтенсивно використовується і при розробці методів штучного інтелекту в застосуванні до лінгвістичної проблематики.

Незважаючи на таке велике різноманіття підходів (а може і завдяки йому), достатньо формальної теорії мови, котра могла б претендувати на роль канонічної, поки що не побудовано. Це стимулює дослідників раз за разом повертатися до витоків мовних явищ і пропонувати все нові підходи.

Існує велика кількість визначень мови. Більшість із них є варіаціями на тему, запропоновану Чейфом: мова – це система, яка в досить складний спосіб здійснює зв’язок між світом звуків і світом значень. Але для формалізації визначення подібного типу не годяться.

Якщо придивитися (“прислухатися”) до мови з позиції того, “що вона робить”, “як вона працює”, можна зробити висновок: *мова представляє собою певний інструмент – своєрідну “машину”, яка забезпечує перетворення “форми” на “зміст” і навпаки.* Але не довільної “форми”, а “форми” лінійних послідовностей певних дискретних об’єктів (звуків і звукових комплексів, знаків і знакових комплексів). Символічно це можна зобразити як:



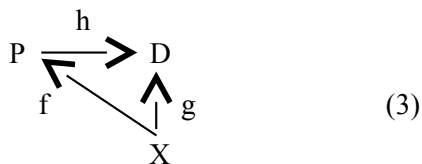
де  $F(S)$  – “формальна” частина опису (представлення) об’єкта  $S$ , котра представляє собою множину лінійних послідовностей дискретних елементів, які належать до певним чином визначеної скінченної множини (“алфавіту”);  $C(S)$  – “змістова” частина опису (представлення) об’єкта  $S$ , яка відповідає множині “форм”  $F(S)$ . Якщо стрілки в діаграмі (1) інтерпретувати як певні відображення, то, вважаючи, що існують механізми, які роблять діаграму (1) комутативною, можна забезпечити “самоузгодженість” моделі – по суті, забезпечити можливість моделювання феноменів розуміння.

Діаграма (1) співзвучна з визначенням складності та інформації за Колмогоровим<sup>3</sup>. Там також відбувається перетворення «змісту» у «форму» лінійних послідовностей.

Відповідна математична конструкція формулюється так. Нехай  $X$  – деяка скінченна множина. Зокрема цю множину можна інтерпретувати як множину вхідних значень для застосування деякого алгоритму  $h$ . Нехай  $Y$  – множина можливих результатів застосування алгоритму  $h$  до множини  $X$ . Нехай  $X' \subset X$  – область визначення алгоритму  $h$ :

$$h: X \rightarrow Y. \quad (2)$$

Нехай  $h(x)$  – обчислювана функція, яка збігається із застосуванням алгоритму  $h$  до елемента  $x \in X'$ . Нехай, врешті-решт, встановлено взаємно-однозначну відповідність між елементами з  $X$  та множиною двійкових слів скінченної довжини, тобто послідовностей вигляду  $d = d_1 d_2 \dots d_n$ , де  $d_i = 0$  чи  $1$ . Позначимо через  $|d|$  довжину слова  $d$ , тобто число нулів і одиниць в його двійковому представленні. Назвемо описом двійкового слова  $d$  певне двійкове слово  $p$ , яке є аргументом деякої однозначної обчислюваної функції  $h$ , такої, що  $h(p) = d$ . Отже, маємо діаграму:



де  $g$  – взаємно однозначне, а  $h$  і  $f$  – однозначні відображення, причому  $h \circ f = g$ , символом « $\circ$ » позначена композиція відображень. Власне, з описаної конструкції вже можна зробити висновок, що

відображення  $g$  можна інтерпретувати як відображення «зміст»  $\rightarrow$  «форма».

Різні слова мають різні описи, але одне слово може мати необмежену кількість описів. Для деякого  $x$  розглянемо множину  $\{p\} = P_h(x)$ , яка складається з усіх  $p$  таких, що  $h(p) = x$ . Нехай:

$$K_h(x) = \begin{cases} \min |p|, \text{ якщо } P_h(x) \text{ – скінченна множина} \\ \infty, \text{ якщо } P_h(x) \text{ – нескінченна множина.} \end{cases} \quad (4)$$

Величина  $K_h(x)$  називається складністю слова  $x$  і являє собою довжину найкоротшого слова  $p$ , в якому міститься повний опис слова  $x$  при фіксованому способі  $h$  його реконструкції. Якщо такого способу не існує, то вважається, що складність слова  $x$  є нескінченною. Недолік вказаного методу – довільність у виборі  $h$ . Але існує теорема [3], яка стверджує: для довільної обчислюваної функції  $h$  і довільного слова  $x$  існує обчислювана функція  $T$ , що виконується така нерівність:

$$K_T(x) \geq K_h(x) + C(h) \quad (5)$$

де  $C$  – константа, яка залежить тільки від  $h$ .

Формули (3)–(5) задають міру складності конструктивного об'єкта  $x$  або, що є тим самим, міру складності двійкового слова, яке перебуває з  $x$  у взаємно-однозначній відповідності. У свою чергу, об'єкт  $x$  може бути інтерпретований як число, алгоритм, узагальнене слово в деякому скінченному алфавіті або деяка теоретична конструкція, в тому числі й модель певного об'єкта. Зазначена обставина дає підставу для інтерпретації формул (3)–(5) як таких, що задають міру кількості інформації, яка міститься в об'єкті  $x$  [3]. Отже, можна зробити висновок, що і мовно-розумові конструкції, моделі та об'єкти також мають таку інформаційну природу, до кількісного опису якої можна застосувати поданий вище формалізм.

Очевидна “схожість”, спорідненість формули (1) з визначенням складності (інформації) за Колмогоровим наводить на думку про те, що в основі і того й іншого лежать ті самі закономірності. Сама форма представлення міри складності свідчить про певний процес, результатом котрого стала генерація “алфавіту” – знакової системи представлення об'єкта. Ця ситуація є настільки загальною, що

дозволяє зробити висновок про те, що в основі будь-якого процесу, явища, об'єкта, системи і т. под. лежить певна дискретна сукупність – назовемо її сукупністю *елементарних інформаційних одиниць*, у визначенні якої ключову роль відіграє поняття *лексикографічного ефекту в інформаційних системах*. Відзначений ефект має семіотичне походження і вперше описаний у роботі<sup>4</sup>. Його сутність зводиться до такого.

Загальна риса всіх процесів інформаційного обміну та взаємодії – численні перетворення інформації з однієї форми на іншу, причому сучасні природничі теорії однозначно підтверджують принципову дискретизацію процесу взаємодії та обміну, і отже, – принципову дискретизацію процесу опису дійсності. Зазначена дискретизація має принаймні одну спільну для всіх відомих процесів рису, яка, на наш погляд, має фундаментальний характер: спостерігаючи та узагальнюючи поведінку різних систем, доходимо висновку, що *у процесі еволюції (динаміки, саморозвитку...) системи будь-якої природи в її структурі індукується певна підсистема відносно сталих дискретних сутностей* ("підсистема порядку"), які відіграють роль її *елементарних інформаційних одиниць*, так що всі інші феномени системи становлять у певний спосіб організовані комбінації цих елементарних інформаційних одиниць. Як ілюстрацію викладеному наведемо деякі приклади.

Незважаючи на те, що за сучасними уявленнями світобудова вміщена в чотиривимірний просторово-часовий континуум, всі спостережувані величини є принципово квантованими. Вони взагалі залежать від невеликого числа так званих "світових констант" (сталой Планка, швидкості світла, заряду та маси електрона тощо), що разом з числами  $e$  та  $\pi$  і дискретною множиною цілих чисел становить своєрідний "алфавіт" фізики, через посередництво якого виражаються всі змістові твердження про поведінку фізичних систем (значення спостережуваних величин).

Аналогічну ситуацію маємо і при науковому описі інших систем: всі речовини є деякими "словами" в "алфавіті" хімічних елементів, а їх взаємоперетворення – "реченнями" в цій "мові"; білки в основному складаються з 20 залишків амінокислот, молекули ДНК будуються з чотирьох типів нуклеотидів і т. д.



Показово, що зазначена поведінка є характерною не тільки для теорій, що описують конкретні об'єкти реального світу (природні та технічні), але й для концепцій, які оперують ідеальними об'єктами, абстракціями й розумовими конструкціями. Справді, навіть опис процесів дискретизації, квантування сигналів, твердження теореми Котельникова можна інтерпретувати не тільки як принципову можливість представлення певного континуального універсуму через посередництво дискретних множин, але і як певний закон природи, що стверджує обов'язковість існування для будь-якої системи такої дискретної підсистеми відносно сталих сутностей (за нашою термінологією – підсистеми елементарних інформаційних одиниць), котра реалізує можливість представлення “через своє посередництво” будь-якого факту з розглядуваного універсуму.

Зазначена підсистема має властивості, споріднені з властивостями лексичної системи природної мови: вона “генерує” в своїй структурі щось на кшталт тезауруса і граматики з притаманними цим конструкціям властивостями знаковості, значення, змісту, поліморфності, омонімії (ізоморфізму), синонімії (гомоморфізму), полісемії (багатозначності), еліпсиса, метонімії, комбінованості і т. д; вона є носієм як “плану вираження”, так і “плану змісту”. Розгортання, взаємодія, взаємні впливи та взаємозумовленість обох зазначених планів у системі елементарних інформаційних одиниць підкоряється певним закономірностям, причому багатовікову контраверзу натуралістів і конвенціоналістів ми схильні розв'язувати в позитивному сенсі: існують приклади систем обох типів, тобто таких, де план виразу детермінований планом змісту, так і таких, де зв'язок між ними має конвенціональний характер.

Сукупностям елементарних інформаційних одиниць притаманна властивість “субстанціальності”, як і будь-яким агрегаціям, причиною яких є певні об'єктивні процеси (у цьому випадку – лексикографічний ефект), внаслідок чого зазначені сукупності, як правило, відзначаються певною стабільністю своїх характеристик, які забезпечують їх локалізацію у відповідних областях системних параметрів.

Тут ми не торкатимемося питань походження, типології та класифікації лексикографічних ефектів, оскільки в цій галузі ще

багато невивченого. Інтуїтивно зрозуміло, що їх походження пов'язане з процесами упорядкування і розупорядкування матерії, які у своїй основі мають дотичність до явищ структурованості, складності, неоднорідності матерії, тобто тих її характеристик, що мають відношення до інформаційних властивостей. Типологія лексикографічних ефектів, якби вона була відома, надала шлях і до побудови класифікаційних схем елементарних інформаційних одиниць.

Вказана сукупність явищ у їх всезагальності і становить зміст лексикографічного ефекту. Можна стверджувати, що при вивченні будь-яких предметних галузей фахівці фактично займаються дослідженням лексикографічних ефектів, які відбуваються в них або є для них характерними. Таким чином, лексикографічний ефект має не тільки феноменальну сторону, але й методологічну, оскільки він несе певний “потенціал операціональності”, стимулюючи в процесі моделювання тих чи інших систем встановлювати і визначати відповідні комплекси елементарних інформаційних одиниць, враховуючи, репрезентуючи та конкретизуючи їх властивості.

У подальшому комплекс елементарних інформаційних одиниць системи  $S$ , що розвинувся внаслідок лексикографічного ефекту (або ефектів)  $Q$ , позначатимемо через  $I^Q(S)$  або просто  $I(S)$ , коли посилання на тип лексикографічного ефекту не суттєве. Система елементарних інформаційних одиниць, будучи носієм цілого ряду властивостей, має певну структуру, що є їх проявом. Зокрема, в будь-якій системі елементарних інформаційних одиниць завжди виділяється ядро – певна власна підсистема  $I_0^Q(S) \subseteq I^Q(S)$ , і визначається породжувальна процедура  $\pi$ , такі що:

$$\pi: I_0^Q(L) \rightarrow I^Q(L). \quad (6)$$

Трійку  $(I^Q(S), I_0^Q(S), \pi)$  також будемо ототожнювати з системою елементарних інформаційних одиниць і використовувати це позначення поряд з  $I^Q(S)$ ,  $I(S)$ ,  $I_0^Q(S)$  як еквіваленти, за умовчанням вважаючи, що породжувальна процедура  $\pi$  визначена, відома і зрозуміла з контексту. Наведемо два приклади, що ілюструють визначення (6).

Для множини хімічних елементів  $I_0^{\text{CHEM}}(S)$  система всіх ізотопів (яка, зрозуміло, містить й саму  $I_0^{\text{CHEM}}(S)$ ) складає  $I^{\text{CHEM}}(S)$ , а породжувальною процедурою  $\pi$  в цьому випадку буде операція

приєднання до ядра відповідного атома допустимого для даного елемента числа нейтронів (0, 1, 2, ...).

Для множини лексем української мови  $I_0^{\text{WORD}}(U)$  в канонічній (вихідній формі) як  $I^{\text{WORD}}(U)$  може виступати повна сукупність текстових словоформ, а  $\pi$  в цьому випадку інтерпретується як оператор парадигматизації (оператор побудови повної парадигми до відповідних лексем, тобто функція, що подає до кожної лексеми української мови в канонічній формі її повну парадигму).

## 2. СТРУКТУРНА ТЕОРІЯ ЛЕКСИКОГРАФІЧНИХ СИСТЕМ

За прототип (а точніше – базовий конструктив) побудови формальної теорії мови ми обираємо лексикографічну модель даних, покладену в основу структурної теорії лексикографічних систем (надалі паралельно вживатимемо скорочення: Л-система), яка ґрунтується на викладеному понятті лексикографічного ефекту.

Відзначимо, що з частинними випадками (або реалізаціями) лексикографічних систем в науці і техніці оперують дуже давно: ними виступає велика кількість різноманітних інформаційних систем, баз даних та знань, у коло яких потрапляють і всі традиційні словники та комп'ютерні словникові системи. Зрозуміло, що машинний словник спроможний ефективно виконувати свої функції лише тоді, коли його структура достатньо повно відображає форму і зміст одиниць мови, що є об'єктами лексикографування – і тенденція до відтворення цієї повноти насправді спостерігається.

На нашу думку, проектування інформаційно-лінгвістичних систем повинно базуватися на глибокому вивченні насамперед феноменології мови, яка сама “підказує” вибір адекватного апарату, а також конструкції відповідних моделей. Незважаючи на те, що метою інформаційної науки є витлумачення предметної галузі (у нашому випадку – лінгвістичних фактів) мовою моделей даних, самі типи та конструкції цих моделей повинні впливати з предметної галузі і якомога повніше враховувати специфіку мовних явищ. Саме виходячи з цієї умови, була сформульована структурна теорія лексикографічних систем, заснована на феноменології лексикографічного ефекту, послідовне застосування якої веде до певних узагальнень і створення методології побудови лексикографічних моделей.

Вихідним пунктом аналізу, результатом якого стало формулювання теорії лексикографічних систем, виявилось дослідження значної кількості структур реально існуючих традиційних словників, їх узагальнення і побудова відповідних моделей. З метою деталізації та інформаційно-лексикографічної конкретизації було проведено дослідження загальних структуротвірних ефектів і елементів лексикографічних систем, які абстрагуються з традиційних словників, перетворюючись на елементи інфологічних моделей лексикографічних систем "загального положення". Цей шлях веде до поняття *структури лексикографічних систем*.

Очевидно, що структура традиційних словників не випадкова, оскільки в ній сконцентровано багатовіковий досвід цілих поколінь лексикографів – тому вона, як правило, вільна від суб'єктивних смаків та уподобань розробників інформаційних систем. Досвід лексикографування, як різновиду інтелектуальної діяльності, в міру його накопичення, від систематизації власне філологічних фактів (і навіть ще вужче – від систематизації даних про лексичні одиниці) поступово поширюється й на систематизацію даних про світ, знання про який, в свою чергу, зосереджені у природній мові як цілісній інформаційній системі.

Універсальність явища лексикографічного ефекту спричиняє тенденцію до лексикографування будь-якого мовного феномена. Тому в лексикографічній практиці існують приклади створення словників, в яких лексикографуються навіть і такі одиниці мови, які не мають безпосереднього вербального представлення. Так, спроба лексикографування синтаксичних структур зроблена, наприклад, у Г. Золотової<sup>5</sup>. У згаданій праці відзначається: "Як фізичний світ навколо нас складається з елементарних частинок, найдрібніших з відомих частинок матерії, так і синтаксична будова нашої мови організується різноманітними, але регулярними комбінаціями елементарних, або мінімальних, одиниць, далі неподільних на синтаксичному рівні. У лінгвістиці на сучасному етапі її розвитку визріла потреба в осмисленні поняття елементарних синтаксичних одиниць, з яких, як стає все очевидніше, будуються й усі інші, більш складні конструкції". І далі: "Синтаксею названо мінімальну, далі неподільну семантико-синтаксичну одиницю російської мови, котра

виступає одночасно як носій елементарного смислу і як конструктивний компонент більш складних синтаксичних побудов, яка, отже, характеризується певним набором синтаксичних функцій".

Аналогічні спроби лексикографування семантичних структур не тільки відображають загальну тенденцію до лексикографічного опису всіх мовних явищ, але й представляють відповіді на запити практики щодо розробки все більш досконалих систем лінгвістичного забезпечення.

Із викладеного випливає методологічна коректність включення одиниць будь-якого мовного рівня до складу елементарних інформаційно-лексикографічних одиниць тієї чи іншої лексикографічної системи. У такий спосіб лексикографуються семантичні, синтаксичні, когнітивні та інші структури, які, як правило, не мають безпосереднього вербального представлення в системі природної мови. До такого ж типу робіт близькі й праці зі створення словників: ідеографічних, дієслівного керування, еквівалентів слів, фразеологізмів тощо.

Вивчення багатовиддя структур існуючих традиційних словників веде до певних узагальнень, які можна не тільки покласти в основу теоретичної лексикографічної схеми, а й використати при проектуванні конкретних інформаційних систем лінгвістичного напрямку й створенні відповідного програмного забезпечення. Оскільки в лексикографії вже давно відбулося розмежування між поняттями "словник" та "список слів", "перелік", "індекс", "інвентар", то словник, як абстрактна лексикографічна система, обов'язково наділений структурою, що містить мінімум дві необхідні частини: реєстрову (ліву) та інтерпретаційну (праву). Саме наявність інтерпретаційної частини відрізняє словник від звичайного списку слів. Але словник має й більш глибоку структуру, яка відображається у будові реєстрової та інтерпретаційної частин словника в цілому та його окремих словникових статей, а також у структурі міжстатейних та міжсловникових відображень. Відтак, словник становить спеціальний вид тексту, в якому в систематизованому і структурованому вигляді подається опис лексики певної мови (або сукупності мов). Однак словник природно розглядати і як специфічний об'єкт техніки, а саме – інформаційну систему, де через посередництво поліграфічного виконання означено певні лінгвістичні

ефекти за допомогою шрифтових виділень, позиційного розміщення, спеціальних позначок тощо, які відіграють роль ідентифікаторів відповідних інформаційних змінних. Складність будови словника полягає ще й у тому, що не всі елементи його структури явно означені вказаним вище способом. У структурі реальних словників, як правило, велика кількість неявних структуротвірних елементів, виявлення яких часто є досить складним завданням. Процес абстрагування словникової (лексикографічної) структури становить своєрідну розшифровку, реконструкцію того лексикографічного ефекту, котрий привів до утворення цієї структури, і розгортається з використанням декількох положень, сформульованих спочатку в мовознавстві, але по суті таких, що мають загальносистемний характер.

Вихідним пунктом побудови структурної моделі лексикографічних (словникових) систем стала орієнтація на *багатоаспектне представлення* знакової природи лексичних одиниць як найбільш компактних та інформативних у природній мові. З позицій теорії лексикографічного ефекту це означає виділення в досліджуваній мовній системі підсистеми її елементарних інформаційних одиниць і встановлення їх системних характеристик.

Наступний момент полягає в урахуванні *дихотомічної структури* кожної елементарної інформаційної одиниці (і повної їх сукупності), котра відображається в *багатовимірному співвідношенні форми та змісту*, носієм якого є визначений клас елементарних інформаційних одиниць.

Багатоаспектність представлення знакової природи одиниць природної мови в традиційних словниках (або елементарних інформаційних одиниць у загальних лексикографічних системах) забезпечується врахуванням семіологічних, лінгвістичних (фонетичних, морфематичних, граматичних, семантичних, стилістичних та інших) і когнітивних особливостей об'єктів лексикографування – залежно від типу словника й глибини характеристики лексикографічного ефекту, який виступає предметом дослідження в кожному конкретному випадку. В інформаційно-лексикографічній моделі зазначеним особливостям протиставляється певна множина комплексів даних та/або знань.

Зауважимо, що у мовному (мовленнєвому) потоці онтологічна природа мови виступає неподільною на окремі складові, які фігурують у концептуальних представленнях. З цього випливає вимога щодо необхідності використання комплексних ("інтегрованих") моделей мовних явищ, і отже, при розробці комп'ютерних систем опрацювання мови постає завдання розробки таких формалізованих моделей, які були б налаштовані на ефективне представлення інтеграційних процесів, урахуваючи водночас специфіку лінгвістичних об'єктів. У такий спосіб критерій багатогранності представлення знакової природи одиниць природної мови веде до необхідності побудови комплексних, інтегрованих моделей даних, придатних до поєднання концептуальних представлень різних за своєю природою мовних явищ.

Дихотомічність структури елементарних інформаційних одиниць в інформаційно-лексикографічній моделі за аналогією з тим, як це робиться в більшості традиційних словників, виявляється у структурній організації лексикографічної системи й впливає з фундаментальних положень сучасної лінгвістики, яка оперує поняттями форми і змісту лексичних одиниць та поняттями внутрішньої та зовнішньої форми слова, феноменологія яких глибоко простежена на мовному матеріалі.

Як відзначається в дослідженні В.М. Русанівського<sup>6</sup>, мова має дуалістичну функцію: з одного боку – це матеріальна основа, на яку мислення спирається у процесі свого функціонування, а з іншої – матеріал, у якому воно фіксується, стаючи доконаним фактом. Об'єктами дослідження складників мовнорозумового потоку є як фізична ("матеріальна"), так і змістова ("ідеальна") сторони. Звукову субстанцію мовлення, отже, можна вважати його формою, а інформаційні властивості – змістом.

З огляду на сказане, звукова реалізація мовлення може поділятися на елементи різного ступеня агрегованості – інтонаційно цілісні одиниці (інтонемі), комбінації голосних і приголосних (склади), власне голосні і приголосні (звуки) і т. д. Цей процес необмежений, оскільки виділення і класифікація складників звуків мовлення залежать від цілого ряду чинників, у тому числі й від прогресу акустики, фонології і т. ін. Фізичний процес мовлення

належить до незворотних і (як і багато інших акустичних явищ) дисипативних процесів. Зазначені властивості фізичної субстанції мовлення разом із властивостями мовленнєвого апарату визначають її зовнішні інформаційні характеристики.

У свою чергу, письмова форма мови моделює її усну форму, так що взагалі справедлива послідовність: <модель дійсності – мислення> → <модель мислення – усна мова> → <модель усної мови – письмова мова>. Оскільки наведені моделі фізично реалізуються в єдиній системі (пов'язаній з індивідуумами, соціальними спільнотами, системами культури...), цілком можливими і реальними є їх взаємодія та взаємовпливи. Таким чином, писемний варіант мови так само виступає у функції і моделі мислення, і моделі дійсності.

Безпосереднє буття мови у вигляді мовленнєвої діяльності, а також писемності та інших способів фіксації мовних актів на фізичних носіях, відмінних від природномовних, репрезентує властивість мови “мати зовнішню форму”. Зовнішня форма взагалі можлива завдяки властивості мови служити позначкою (“представником”) феноменальної сторони дійсності, і оскільки мова сама становить певну дійсність, то в ній закладені потенції для позначення (“представлення”) самої себе.

Система, яка виступає репрезентантом феноменальної сторони дійсності, повинна мати й відповідну організацію. Оскільки відмінність між явищем і суттю є відносною, а між феноменальною і сутнісною дійсністю немає різкої межі, то її не мусить бути і в мові як моделі дійсності. Цей факт і реалізується у властивості слова мати внутрішню форму, яка має зв'язок із представленням ноуменальної частини його власного буття в системі мови. Зовнішня і внутрішня форми таким чином є взаємопов'язаними і разом складають форму слова на противагу його змісту – сумі конкретних значень.

Такого роду міркування дають підстави для твердження, що відношення “форма-зміст” та поняття внутрішньої і зовнішньої форми носять всезагальний характер, представляють універсальну властивість елементарних інформаційних одиниць, що індукуються в процесі розвинення будь-якого лексикографічного ефекту і, будучи формалізованими у вигляді відповідних моделей даних, спроможні скласти субстрат моделей інформаційних систем довільної природи та походження, а для мовноорієнтованих моделей даних вони є



просто обов'язковими. Зазначені поняття, на наш погляд, мають потенціал конструктивності, оскільки будь-який „зміст” існує лише у певній формальній оболонці, тобто у вигляді деякої “форми”, що дозволяє застосовувати уніфікований підхід до побудови їх репрезентантів у науковій теорії.

Розглянемо деяку систему  $S$ , концептуальний опис якої ми прагнутимо представити у вигляді певної лексикографічної системи. Оскільки нас цікавлять насамперед лінгвістичні факти, то системою  $S$  у нас виступатиме саме природна мова, або сукупність природних мов, або певна підсистема (певний аспект) природної мови.

Системі  $S$  притаманна ієрархія лексикографічних ефектів. Так, для природномовної системи можна навести приклади цілої низки лексикографічних ефектів, результатом яких постало виділення з мовленнєвого потоку окремих фонем, складів, морфем, слів (словоформ), лексем, словосполучень, еквівалентів слів, речень і т. д. Усі названі структурні одиниці виступають як складники відповідних класів елементарних інформаційних одиниць стосовно тих чи інших типів природномовних лексикографічних ефектів.

У подальшому будемо розуміти під *лексикографічною системою (Л-системою)* спеціальне інформаційне середовище, в якому розвивається (реалізується) певний лексикографічний ефект (або лексикографічні ефекти).

Для побудови практично корисної схеми моделювання відзначених явищ необхідно мати набір інформаційних конструктивів, які специфікують структурні елементи Л-систем, дозволяючи розробку конкретних застосувань. У такий спосіб приходимо до необхідності конструктивної теорії Л-систем, в основу якої покладено лексикографічну модель даних і яка вводиться у такий спосіб.

Визначимо певну множину  $V(I^Q(S))$ , як множину описів класу  $I^Q(S)$ , іншими словами, визначимо відображення:

$$D : I^Q(S) \longrightarrow V(I^Q(S)) \quad \text{і} \quad \Delta : V(I^Q(S)) \longrightarrow I^Q(S) \quad (7)$$

так що для кожного елемента  $x \in I^Q(S)$  однозначно визначено його опис  $V(x)$  як елемент множини  $V(I^Q(S))$  ( $V(x) \in V(I^Q(S))$ ):  $D x = V(x)$ . Отже,  $V(I^Q(S))$  має будову:

$$V(I^Q(S)) = \cup_{x \in I^Q(S)} V(x), \quad (8)$$

де  $V(x)$  представляється у вигляді слова (тексту) в деякому скінченному алфавіті  $A = \{a_1, a_2, \dots, a_n\}$ :

$$V(x) = v_1(x) v_2(x) \dots v_{k(x)}(x), \quad v_i(x) \in A, \quad i=1, 2, \dots, k(x). \quad (9)$$

Відзначимо, що серед символів  $v_i(x) \in A$  можуть бути й такі, що повторюються.

На  $V(x)$  визначимо певну структуру  $\beta(x)$  у такий спосіб. Припустимо, що для всіх  $V(x) \in V(I^Q(S))$  існує єдине правило, за яким будь-який  $V(x)$  – як слово вигляду (9) – розбивається на множину підслів  $\beta_i(x)$  наступним чином:

$$\{\beta_i(x), i = 1, 2, \dots, q\} = \beta[V(x)]; \beta[V(x)] \subseteq B[V(x)], \quad (10)$$

де  $B[V(x)] = \{v_{i_1} v_{i_2} \dots v_{i_p}, 1 \leq i_1 < i_2 < \dots < i_p \leq k(x), k = 1, 2, \dots, n, v_{ij} \in \{v_i(x)\}\}$ .

Назвемо  $\beta[V(x)]$  структурою  $V(x)$  за умови, що:

$$x \in \beta[V(x)]; V(x) \in \beta[V(x)], \beta_i(x) \neq \beta_j(x) \text{ при } i \neq j. \quad (11)$$

Покладемо за визначенням:

$$\beta[V(I^Q(S))] = \cup_{x \in I^Q(S)} \beta[V(x)] \quad (12)$$

Очевидно, що  $V(I^Q(S)) \in \beta[V(I^Q(S))]$ .

Позначимо через:

$$\beta_i = \cup_{x \in I^Q(S)} \beta_i(x), \quad i = 1, 2, \dots, q, \text{ а також } \beta = \cup_i^q \beta_i. \quad (13)$$

Зрозуміло, що  $\beta \equiv \beta[V(I^Q(S))]$ .

Відзначимо, що деякі з елементів  $\beta_i(x)$ ,  $i = 1, 2, \dots, q$ , можуть бути порожніми елементами для певних  $x \in I^Q(S)$ ; у цьому випадку вони опускаються в (10 – 13).

Через  $\sigma[\beta]$  позначимо певну структуру, визначену на  $\beta$ , і отже, на  $V(I^Q(S))$  – надалі називатимемо  $\sigma[\beta]$  макроструктурою  $V(I^Q(S))$  (обмеження  $\sigma[\beta]$  на  $V(x)$ :  $\sigma[\beta] \big|_{V(x)} \equiv \sigma(x)$  породжує мікроструктуру  $V(x)$ ). Активне формулювання цього факту полягає у встановленні процедури (оператора, процесу...)  $\sigma$ , який породжує на  $\beta$  структуру  $\sigma[\beta]$ :

$$\sigma: \beta \longrightarrow \sigma[\beta]. \quad (14)$$

Зрозуміло, що на  $V(I^Q(S))$  можлива генерація цілого ряду неізоморфних структур  $\sigma[\beta]$ . У їх ролі можуть виступати будь-які моделі даних (у тому числі й реляційна), логіко-математичні моделі (зокрема, логічні числення типу логіки предикатів), конструкції формальних граматик тощо.

Завдяки можливості інтерпретації  $I^Q(S)$  як класів, елементами яких є об'єкти будь-якого походження, природною є об'єктно-орієнтована інтерпретація моделі. Відношення між елементами класу  $I^Q(S)$  індуються системою унарних відношень  $r[\beta_1]$  на  $\beta_1 = \{x\}$  та відображенням:

$$\Delta: V(I^Q(S)) \longrightarrow I^Q(S); \quad \Delta r[\beta_1]. \quad (15)$$

У такий спосіб комплекс  $I^Q(S)$  постає представником онтологічної природи реальності, що підлягає моделюванню, у той час як  $V(I^Q(S))$ ,  $\beta$ ,  $\sigma[\beta]$  представляють її концептуальну сторону.

Подальший розгляд концентруватиметься навколо відношень "форма-зміст", носієм яких є комплекс  $I^Q(S)$  і які розвиваються й реалізуються в середовищі  $\{I^Q(S), D, V(I^Q(S)), \beta, \sigma[\beta]\}$ . При цьому сукупність властивостей і якостей комплексу  $I^Q(S)$ , внаслідок скінченності й обмеженості перцептивно-сенсорних і пізнавальних потенцій "суб'єкта D", розподіляється на дві – не цілком чіткі і не цілком чітко відокремлювані частини. У концептуальній же схемі, реалізованій в описі  $V(I^Q(S))$ , вказані частини повинні бути відокремлені; іншими словами, необхідною умовою коректності її побудови є існування процедури, яка здійснює таке відокремлювання. Зі сказаного випливає існування комутативної діаграми:

$$\begin{array}{ccc}
 & V(I^Q(S)) & \\
 & \swarrow \quad \searrow & \\
 F & & C \\
 \swarrow \quad \searrow & & \swarrow \quad \searrow \\
 \Lambda(I^Q(S)) & & P(I^Q(S)) \\
 & H &
 \end{array} \quad (16)$$

$$F(V(I^Q(S))) = \Lambda(I^Q(S)); \quad C(V(I^Q(S))) = P(I^Q(S)); \quad \Lambda(I^Q(S)) \cap P(I^Q(S)) = \emptyset, \quad (1.15)$$

причому  $H \circ F = C$ , де символом „ $\circ$ ” позначено композицію відображень;

$$\Lambda(I^Q(S)) = \cup_{x \in I^Q(S)} \Lambda(x) \quad P(I^Q(S)) = \cup_{x \in I^Q(S)} P(x) \quad (17)$$

До оператора  $H$  (зазвичай і там, де це можливо і доцільно) визначається й обернений оператор:  $H^{-1}$ . Для прикладу візьмемо як  $V(I^Q(S))$  – парадигматичну  $L$ -систему для української мови. Тоді  $\Lambda(I^Q(S))$  набуває інтерпретації у вигляді множини вихідних слівформ українських лексем, а  $P(I^Q(S))$  – множина відповідних повних парадигм. Позначивши вихідну форму лексеми  $x$  через  $x_0$ , її повну парадигму через  $[x]$ , будь-який елемент цієї парадигми через  $\chi(x)$ ;  $H \equiv H^{PAR}$ ;  $H^{-1} \equiv H^{LEM}$ , одержуємо  $H^{PAR}x_0 = [x]$ ;  $H^{LEM}\chi(x) = x_0$ . Якщо  $H^{PAR}$  є однозначним відображенням, то  $H^{LEM}$ , взагалі кажучи, відображення неоднозначне, що створює чималі проблеми при розв’язанні ряду завдань лінгвотехнології, зокрема при автоматичній природномовній індексації українських текстів.

На  $\Lambda(I^Q(S))$  і  $P(I^Q(S))$  індукуються макроструктури:

$$F(\sigma[\beta]) = \lambda[\beta] \text{ і } C(\sigma[\beta]) = \rho[\beta] \quad (18)$$

і відповідні їм мікроструктури:

$$\lambda[\beta] \big|_{V(x)} \equiv \lambda(x); \quad \rho[\beta] \big|_{V(x)} \equiv \rho(x) \quad (19)$$

як обмеження  $\lambda[\beta]$  і  $\rho[\beta]$  на  $V(x)$ .

Відзначимо, що діаграма (16) – тобто об’єкти  $V(I^Q(S))$ ,  $\Lambda(I^Q(S))$ ,  $P(I^Q(S))$  і відображення  $F$ ,  $G$ ,  $H$ , – фактично будується незалежно від структури  $\sigma[\beta]$ . Походження і смисл її складових елементів зовсім інші. А саме:  $\Lambda(I^Q(S))$  співвідноситься з тією частиною опису  $V(I^Q(S))$ , яке – в певному розумінні – представляє форму  $I^Q(S)$ , а, відповідно,  $P(I^Q(S))$  зіставляється з тією частиною опису  $V(I^Q(S))$ , що відповідає за зміст  $I^Q(S)$ . Наведена специфікація практично реалізує наше переконання в тому, що відношення "форма-зміст" є універсальними, притаманними об’єктам будь-якої природи і походження, оскільки вони залучаються до процесу взаємодії з суб’єктом, котрий їх досліджує.

**Визначення.** Вісімка об’єктів  $\{I^Q(S), D, V(I^Q(S)), \beta, \sigma[\beta], F, C, H\}$  визначає *елементарну лексикографічну модель даних*, а її конкретна реалізація – *елементарну лексикографічну систему*.

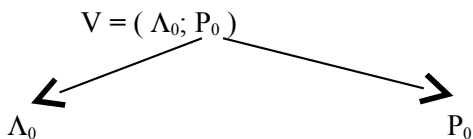
Зауважимо, що будь-який елемент  $\beta_i$  (або будь-яка їх сукупність), що належить до структур,  $\beta$ ,  $\sigma[\beta]$ ,  $\lambda[\beta]$ ,  $\rho[\beta]$  може набути інтерпретації як елементарної Л-системи. Таким чином одержуємо можливість виокремити в структурі початкової елементарної Л-системи низку інформаційно-лінгвістичних підструктур, які тлумачимо як окремі Л-системи. Перевизначивши у такий спосіб структуру вихідної Л-системи, одержуємо Л-модель та Л-систему загального положення (неелементарну); вона представляється у вигляді об'єднання певної кількості елементарних Л-систем із можливими відображеннями та зв'язками між ними. Таким чином, Л-система загального положення представляється у вигляді графа  $G = \{V = \{V_i\}; R = \{R_{kl}\}\}$ , де  $V = \{V_i\}$  – множина вершин, якими виступають елементарні Л-системи  $V_i$ , що входять до  $G$ , а  $R = \{R_{kl}\}$  – множина ребер графа  $G$ , так що  $e R_{kl}$  поєднує  $V_k$  та  $V_l$ .

Зокрема, ніщо не заважає дивитися на  $\Lambda(I^Q(S))$  і  $P(I^Q(S))$  як на окремі, автономні елементарні Л-системи, а це надає можливість побудови такого розвинення:

$$\begin{array}{ccc}
 & & H_0 \\
 V(I^Q(S)) = (\Lambda(I^Q(S)) \equiv \Lambda_0(I^Q(S))) & \xrightarrow{\quad} & P_0(I^Q(S)) \equiv P(I^Q(S)) \\
 \begin{array}{c} \swarrow \\ F_{01}^{\Lambda} \end{array} & & \begin{array}{c} \swarrow \\ F_{01}^P \end{array} \\
 \begin{array}{c} \searrow \\ C_{01}^{\Lambda} \end{array} & & \begin{array}{c} \searrow \\ C_{01}^P \end{array} \quad (20)
 \end{array}$$

$$\begin{array}{ccccccc}
 \Lambda_{01}^{\Lambda}(I^{Q1}(S)) & & P_{01}^{\Lambda}(I^{Q1}(S)) & \xrightarrow{\quad} & \Lambda_{01}^P(I^{Q2}(S)) & & P_{01}^P(I^{Q2}(S)) \\
 & & H_{01}^{\Lambda} & & H_{01}^P & & 
 \end{array}$$

Звернімо увагу на зміну типу лексикографічного ефекту на другому поверсі – замість  $Q$  тепер маємо  $Q_1$  й  $Q_2$ , відповідно. Отже, приходимо до комплексів об'єктів  $I^{Q1}(S)$  і  $I^{Q2}(S)$ . Продовжуючи цей процес, одержуємо рекурсивне розвинення лексикографічної системи  $V(I^Q(S))$ :





За нашою роботою<sup>7</sup> назвемо цей процес *рекурсивною редукцією лексикографічної системи*. Він нагадує своєрідний інформаційний “мікроскоп”, що “виявляє” з кожним кроком усе більш тонкі деталі структури лексикографічної системи.

Надалі позначатимемо процес рекурсивної редукції Л-системи  $V(I^Q(S))$  через  $RR\downarrow[V(I^Q(S))]$ . У визначення цього процесу входять характеристики всіх операторів F, G, H на всіх наявних рівнях рекурсивної редукції разом із результатами їх дії, а також усі макро- і мікроструктури  $\sigma, \lambda, \rho$ .

Таким чином, приходимо до загального визначення лексикографічної системи у вигляді сімки:

$$\{I^Q(S), D, V(I^Q(S)), \beta, \sigma [\beta], RR\downarrow[V(I^Q(S))], \Sigma\}, \quad (22)$$

де символом  $\Sigma$  позначено архітектуру, в якій реалізовано *лексикографічну модель даних*:

$$\{I^Q(S), D, V(I^Q(S)), \beta, \sigma [\beta], RR\downarrow[V(I^Q(S))]\}. \quad (23)$$

Визначення окремих елементів моделі у формулі (23) представлене формулами (3–21); архітектура  $\Sigma$ , зазвичай, обирається трирівневою, такою, що відповідає ANSI/SPARK<sup>8</sup>.

Основні складові архітектури ANSI/SPARK використовуватимемо в такій інтерпретації:

$$ARCH\_LS = \{KM, EXM, INM; \Phi, \Psi, \Xi\}, \quad (24)$$

де, в останній формулі, KM є концептуальною моделлю лексикографічної системи LS. Символом  $EXM = \{ex M\}$  позначено множину її зовнішніх моделей, які відповідають цій концептуальній моделі KM, а  $INM = \{in M\}$  – відповідна множина її внутрішніх моделей. Через  $\Phi = \{\phi\}$  позначено множину відображень KM в EXM:

$$\phi : KM \rightarrow ex M, \text{ де } ex M \in EXM; \quad (25)$$

відповідно,  $\Psi = \{\psi\}$  – множина відображень KM в INM :

$$\psi : KM \rightarrow in M, \text{ де } in M \in INM; \quad (26)$$

$\Xi = \{\xi\}$  – множина відображень INM в EXM.

При цьому ми зупиняємося на такій інтерпретації елементів архітектури.

Концептуальна модель (*концептуальний рівень представлення*) предметної галузі – це знакова, семантична модель, у котрій в однозначному, скінченному та несуперечливому вигляді інтегруються уявлення різних фахівців про предметну галузь.

У внутрішній моделі (*внутрішньому рівні представлення*) визначаються типи, структури та формати представлення, зберігання та маніпулювання даними, алгоритмічна база та операційно-програмне середовище, в яке „занурюється” концептуальна модель при її реалізації.

Зовнішня модель (*зовнішній рівень представлення*) відображає погляди кінцевих користувачів і, отже, прикладних програмістів на інформаційну систему. В ній реалізується система засобів, які дозволяють користувачеві здійснювати дозволені контакти і маніпулювання даними, представленими на внутрішньому рівні.

Одній концептуальній моделі може відповідати декілька внутрішніх та зовнішніх моделей, тому відображення  $\Phi: KM \rightarrow EXM$  та  $\Psi: KM \rightarrow INM$ , взагалі кажучи, не ін’єктивні. Але відображення  $\phi: KM \rightarrow exM$ , та  $\psi: KM \rightarrow inM$  будуються ін’єктивними (але, зазвичай, не бієктивними). Визначимо множину  $\Xi$  відображень  $INM$  в  $EXM$ :

для  $\forall inM \in INM$  та  $\forall exM \in EXM \exists \xi \in \Xi$  таке, що:

$$\xi(inM) = exM. \quad (27)$$

При цьому відображення  $\phi$ ,  $\psi$ ,  $\xi$  будуються в такий спосіб так, що діаграма:



є комутативною:  $\xi \circ \psi = \phi$ . Вимога комутативності цієї діаграми є суттєвою, оскільки гарантує узгодженість між усіма рівнями архітектури системи.

### 3. ПОНЯТТЯ ЛЕКСИКОГРАФІЧНОГО СЕРЕДОВИЩА

У реальній дійсності об'єкти мови функціонують в їх цілісності, не поділеній на окремі складові їх концептуального представлення. У лексикографічній системі це виявляється в тому, що представлення мовних об'єктів передбачає поєднання різних типів лексикографічних ефектів, а також поєднання та узгодження різнорідних (гетерогенних) лексикографічних структур. У свою чергу, це потребує узгодження між усіма елементами архітектури Л-систем, які підлягають процесу інтеграції.

Проведені нами численні експерименти зі створення конкретних комп'ютерних реалізацій інтегрованих лінгвістичних об'єктів дозволили зробити висновок про необхідність побудови спеціального мовно-інформаційного середовища, яке б із самого початку було адаптованим до процесів інтегрування різних лексикографічних систем і містило необхідні засоби та конструктиви для здійснення зазначених процесів, а також і фіксації їх результатів у вигляді інтегрованих лексикографічних систем, що мають різні лексикографічні структури.

У такий спосіб нами було запропоновано поняття *лексикографічного середовища*, яке вводиться в наступний спосіб.

Визначення. Вважається, що задано *лексикографічне середовище (Л-середовище)  $ML$* , якщо:

1. Задано клас Об  $ML$  елементів, кожен з яких є діаграмою вигляду (28) і представляє певну Л-систему (не обов'язково елементарну). Зазначені елементи називаються об'єктами Л-середовища  $ML$  – позначатимемо їх великими латинськими літерами:  $A, B, C, \dots$ .

2. Для кожної пари пари об'єктів  $A, B$  з  $ML$  задано множину  $\text{Hom}_{ML}(A, B)$ , яка називається множиною морфізмів  $A$  в  $B$ ; замість  $f \in \text{Hom}_{ML}(A, B)$  часто пишуть:

$$f$$

$f: A \rightarrow B$  або  $A \rightarrow B$ . При цьому  $f: \text{KM}_A \rightarrow \text{KM}_B$ ;  $f: \text{INM}_A \rightarrow \text{INM}_B$ ;

$f: \text{EXM}_A \rightarrow \text{EXM}_B$ ;  $f(\varphi_A) = \varphi_B$ ;  $f(\psi_A) = \psi_B$ ;  $f(\xi_A) = \xi_B$  та  $f(\xi_A) \circ f(\psi_A) = f(\varphi_A)$ .

3. Для кожної трійки об'єктів  $(A, B, C)$  з  $ML$  задано відображення

$$\mu: \text{Hom}_{ML}(A, B) \times \text{Hom}_{ML}(B, C) \rightarrow \text{Hom}_{ML}(A, C)$$



(образ  $\mu(f, g)$  пари  $(f, g)$ , де  $f \in \text{Hom}_{ML}(A, B)$ ,  $g \in \text{Hom}_{ML}(B, C)$ ), буде позначатися  $f \circ g$  або  $fg$  і називатися композицією морфізмів  $f$  і  $g$ ).

4. Множини  $\text{Hom}_{ML}(A, B)$  і композиція морфізмів задовольняють такі аксіоми:

(а) Асоціативність: для кожної трійки морфізмів  $f, g, h$

$$\begin{array}{c} f \quad g \quad h \\ A \rightarrow B \rightarrow C \rightarrow D \quad f(g \circ h) = (f \circ g) \circ h. \end{array}$$

(б) Існування одиниці: для кожного  $A \in \text{Ob } ML$  існує морфізм  $1_A: A \rightarrow A$ , ( $1_A \in \text{Hom}_{ML}(A, A)$ ), такий що  $1_A \circ f = f$  і  $g \circ 1_A = g$  для довільних морфізмів  $f \in \text{Hom}_{ML}(B, A)$  і  $g \in \text{Hom}_{ML}(A, B)$ .

(с) Якщо пари  $(A, B)$  і  $(A', B')$  різні, перетин  $\text{Hom}_{ML}(A, B)$  і  $\text{Hom}_{ML}(A', B')$  – порожній. Серед морфізмів з  $\text{Hom}_{ML}$ , відповідно до їх властивостей, можуть виділятися мономорфізми, епіморфізми та ізоморфізми.

Нехай дано два лексикографічні середовища  $ML_1$  і  $ML_2$ . Коваріантний (відповідно контрваріантний) функтор  $F$  з  $ML_1$  в  $ML_2$  складається з:

(а) відображення  $A \rightarrow F(A)$ , яке протиставляє кожен об'єкт  $A$  з  $ML_1$  об'єкт  $F(A)$  з  $ML_2$ ;

(б) відображеннь  $F(A, B): \text{Hom}_{ML_1}(A, B) \rightarrow \text{Hom}_{ML_2}(F(A), F(B))$  – для коваріантного й  $F(A, B): \text{Hom}_{ML_1}(A, B) \rightarrow \text{Hom}_{ML_2}(F(B), F(A))$  для контрваріантного функтора, визначених для всіх пар  $(A, B)$  об'єктів з  $ML_1$  і таких, що (якщо замість  $F(A, B)(u)$  писати просто  $F(u)$ )  $F(1_A) = 1_{F(A)}$  і  $F(vu) = F(v)F(u)$  (відп.  $F(vu) = F(u)F(v)$ ).

Отже, елементами специфікації Л-середовища  $ML$  виступають: клас систем  $\{S\}$ ; клас лексикографічних ефектів  $\{\{Q\}(S)\}$ , що розвиваються в кожній  $S$  з  $\{S\}$ ; сукупність класів елементарних інформаційних одиниць  $\{I^Q(S)\}$ , що генерується в кожній  $S$  з  $\{S\}$  відносно кожного лексикографічного ефекту  $Q$  з  $\{\{Q\}(S)\}$ ; клас відображень  $\{d\} \{I^Q(S)\}$  в  $\{V(I^Q(S))\}$ ; клас структур  $\{\beta, \sigma[\beta]\}(V(I^Q(S)))$ ; клас процесів рекурсивної редукції  $\{RR \downarrow [V(I^Q(S))]\}$ ; множини  $\{INM\}$ ,  $\{EXM\}$  внутрішніх і зовнішніх моделей та відображень  $\{\varphi\}$ ,  $\{\psi\}$ ,  $\{\xi\}$ , визначених формулою (24), відповідно; до специфікації Л-середовища  $ML$  входить і клас морфізмів  $\text{Hom}_{ML}$ .

#### 4. ЗАКЛЮЧНІ ЗАУВАЖЕННЯ

На теперішній момент автори можуть стверджувати, що в єдиній моделі даних у принципі можливе поєднання різних формальних конструкцій, які використовуються при моделюванні мовних явищ. Це стосується як окремих лінгвістичних алгоритмів, так і більш агрегованих об'єктів, представлених синтаксичними структурами і взагалі формальними граматиками різних типів, моделями лінгвістичних знань, заснованих на фреймах, продукціях тощо.

Проте, на думку авторів, проблема є насправді більшою: її найістотніші аспекти сконцентровано в галузі типології лексикографічних ефектів, яка тісно пов'язана з іншою фундаментальною проблемою – проблемою суб'єктивного у структурі світобудови. Адже основне системотвірне відношення конструкції Л-систем – відношення “форма – зміст” – є по суті нічим іншим, як констатацією існування суб'єкта, який сприймає дійсність, персоніфіковану в комплексі  $I^Q(S)$ . Дійсно, відображення  $D$ , що фігурує у формулі (7), природно інтерпретувати як дію суб'єкта, який за допомогою своїх перцептивно-сенсорних органів і мовнорозумового апарату, або через посередництво створених ним приладів, або за допомогою і того, і другого здійснює сприйняття, пізнання й опис комплексу  $I^Q(S)$ . Відношення відзначеного типу є системотвірними не тільки для природної мови, а й взагалі для систем, в яких відбуваються процеси, символічно зображені у формулах (7). Мовні та мовнорозумові процеси у межах сформульованого підходу вкладаються в більш широкий і загальний контекст, зберігаючи при цьому свою специфіку.

<sup>1</sup>Палагин А.В., Широков В.А. Принципы когнитивной лексикографии;

<sup>2</sup>Лихачев Д.С. Еще о точности литературоведения // О филологии. – М., 1989. – С. 27-30; <sup>3</sup>Колмогоров А.Н. Три подхода к определению понятия "количество информации" // Теория информации и теория алгоритмов. – М., 1987. – С. 213-223; <sup>4</sup>Широков В.А. Інформаційна теорія лексикографічних систем. – К., 1998; <sup>5</sup>Золотова Г.А. Синтаксический словарь. Репертуар элементарных единиц русского синтаксиса. – М., 1988; <sup>6</sup>Русанівський В.М. Структура лексичної та граматичної семантики. – К., 1988; <sup>7</sup>Широков В.А. Інформаційна теорія та системотехнічні засади комп'ютерної лексикографії.

Автореф. дис. ... д.т.н. – К., 1999; <sup>8</sup>ANSI/X3/SPARK Study group on data base management systems: interim report, FDT, 7:2, ACM, – New York, 1975.

*Л.І. Шевченко, канд. філол. наук*

## **МЕНТАЛЬНІСТЬ У ДИСКУРСІ ІНТЕЛЕКТУАЛІЗАЦІЇ ЛІТЕРАТУРНОЇ МОВИ**

*У концептах теорії інтелектуалізації літературної мови аналізується ментальність як мовний код нації.*

*The category of mentality as the linguistic code of a nation is analyzed in terms of the theory of intellectualization of a literary language.*

Гуманітарні дослідницькі парадокси значною мірою відбиті у розумінні категорії ментального: численні теоретичні роботи і методики аналізу залишають відкритим питання дефінітивної окресленості терміна. Зауважимо, точність визначення швидше нівелювала б проблему, ніж наближала до істини, тому що немає однозначного розуміння обсягу поняття: ментальне співвіднесене з первісними універсальними уявленнями про світ? диференційованими характеристиками світорозуміння і світовідчуття етносу / етносів? ментальне є специфікаційною константою чи еволюційно (культурно) змінною атрибутивністю національного світобачення? Лінгвістичний парадокс виявляється і в критеріях наукового опису: ментальність корелюється з «картиною світу» (тоді як «картина світу» відбиває насамперед первісний, язичницький світогляд), і ментальне ж, як правило, розглядають у контексті національних характеристик, тобто нівелюється логіка історичного, понятійного й ретроспективного аналізу мовних явищ. Ментальне включають у контекст культури, в той час як культура передусім синонімізована з інтегральною цивілізаційною характеристикою А. Боннера: «цивілізований», тобто «прищеплений» до культури. В епістемології присутнє і розуміння поняття «ментального» як відповідника до «інтелектуального», зокрема в концептах «школи інтелектуальної історії», де принципово декларується увага до «розвитку і зміни ментальних досягнень людства», «історії ідей»,