

УДК 811.161.2''73.611

Л. Є. Азарова

Вінницький національний технічний університет, м. Вінниця

**ЛІНГВО-ІНФОРМАЦІЙНИЙ ПІДХІД ЯК БАЗИСНА ФОРМА  
ДОСЛІДЖЕННЯ СКЛАДНИХ НОМІНАЦІЙ НА СТРУКТУРНОМУ  
РІВНІ**

*Азарова Л. Є. Лінгво-інформаційний підхід як базисна форма дослідження складних номінацій на структурному рівні*

*Анотація – У статті розглянуто використання лінгво-інформаційного підходу для дослідження складних слів на структурному рівні. Розроблено основні положення теорії гармонійної побудови складних номінацій на засадах концепції «золотої» пропорції; запропоновано методологію дослідження гармонійної побудови складних слів у концепції «золотої» пропорції, що спирається на лінгво-інформаційний підхід та сучасні інформаційні технології*

*Ключові слова – композит, юкстапозит, лінгво-інформаційний підхід, «золота» пропорція, числа Фібоначчі, закон гармонійного поділу.*

*Азарова Л. Е. Лингво-информационный подход как базовая форма исследования сложных номинаций на структурном уровне*

*Аннотация – В статье рассмотрено использование лингво-информационного подхода для исследования сложных слов на структурном уровне. Разработаны основные положения теории гармонического построения сложных слов на основе концепции «золотой» пропорции; предложена методология исследования гармонического построения сложных слов в концепции «золотой» пропорции, которая основывается на лингво-информационном подходе и современных информационных технологиях.*

*Ключевые слова – композит, юкстапозит, лингво-информационный подход, «золотая» пропорция, числа Фибоначчи, закон гармонического деления.*

*Azarova L. E. Linguistic and informative approach as a basic form of nominations for the study of complex structural level*

*The development of the theory of linguistics, including derivation, requires a choice of methods and approaches, borrowing the achievements of other sciences. Experimental studies in linguistics must in turn also involve the use of methodological tools and hardware information technologies. The choice of basic shapes depending on the specific direction of linguistic research, the nature of the problems and challenges faced by linguists at present and in the future.*

*In our era – the era of computerization and informatization changing methodology of scientific knowledge, improved his methods and used the ones that have stood the test of time. There is a new technique that allows you to reach a deeper study of the objective world of natural phenomena to explain their essence. In this regard, developed the basic tenets of the theory of building a harmonious compound words based on the concept of «golden» proportions; proposed research methodology harmonious construction of complex words in the concept of «golden» proportions, based on linguistic-information approach and modern information technology; analyzed the expression «golden» proportions as a quantitative indicator of harmony in human mental activity and word formation; the basic principles of difficult words on the structural level; isolated and characterized noun Composites for their generators of parts belonging bases; investigated harmonious construction of two-component composite nouns, adjectives, participles, adverbs and verbs in fact Ukrainian and foreign origin; systematized on the analysis of new theoretical basis; The mechanism of harmonious construction of complex nominations for the structural level; established principles of specificity and yukstapozyt at the structural level.*

*Keywords – composite, yukstapozyt, linguistic-information approach, «golden» proportion, Fibonacci numbers, the law of harmonious division.*

У нашу епоху комп'ютеризації та інформатизації змінюється методологія наукового пізнання: вдосконалюються його методи і використовуються саме ті, які витримали випробовування часом; з'являється нова методика, яка дозволяє глибше охопити дослідження об'єктивного світу, явищ природи, пояснити їхню сутність. У теперішній час мова значно розширює свої комунікативно-суспільні функції. Продовжуючи залишатися найважливішим знаряддям спілкування у суспільстві, природна мова стає тією основою, на базі якої здійснюється взаємодія між людьми та комп'ютером. Вона не може безпосередньо функціонувати в комп'ютері або в комп'ютерних системах і мережах. Природна мова повинна перетворитися з асоціативно-евристичного ґрунту людського мислення на послідовно-алгоритмічну мову комп'ютера. Проблема такого перетворення вказаних мов у свою чергу породжує інженерно-лінгвістичну проблему, яка призводить до появи в мовознавстві низки методологічних наслідків.

Розвиток сучасного мовознавства характеризується внесенням у лінгвістичні дослідження таких технологічних обмежень й умов, які надавали б, з одного боку, логічну строгість теорії, а з іншого – забезпечували б об'єктивність і повноту описання лінгвістичного матеріалу. Прагнення логічної строгості теорії й об'єктивності характеризує загальну типологію, соціолінгвістику, квантитативне і структурне мовознавство.

Класична лінгвістика, як правило, відволікається від того матеріального середовища, у якому здійснюються інформаційні процеси, що нею вивчаються. Такий підхід відрізняється від стратегії й технології, які застосовують, наприклад, інженерні лінгвісти під час визначення особливостей функціонування природної мови в системах «людина-комп'ютер-людина» [1, с. 4]. У цьому плані слід зазначити, що комп'ютерні

системи мають порівняно з людиною набагато більшу завадостійкість і надійність. Якщо функціонування мозку під час обробки різноманітних текстів, перекладів, складанні словників залежить від фізіологічного й психічного стану людини, а також швидко ослаблюється внаслідок втомлюваності, то робота комп'ютерних систем характеризується значною стійкістю і тривалістю [1, с. 109].

Поява значної кількості нових слів породжує проблему сприйняття їх людиною і подальшого використання. Безумовно, що процес розширення обсягів тематичних словників у різноманітних галузях знань є об'єктивним. На нього впливають маса чинників, пов'язаних із досягненнями людини в науці, техніці, соціальній сфері та інших. Крім того, кількість галузей знань із часом постійно зростає. Це все викликає появу нових номінативних одиниць. Проте на цьому шляху простежуються дві тенденції. Перша поява «абсолютно» нових слів, як правило, простих, що вимагає додаткових комірок пам'яті (за аналогією з комп'ютером) у мозку людини. Друга тенденція – це утворення нових багатокомпонентних складних слів, переважно двоконпонентних. Причому компоненти складних слів можуть бути як новими, так і старими. Це дозволяє економити комірки пам'яті мозку, а пошук, ідентифікацію слів, а також розкриття їхнього нового значення здійснювати за асоціативною ознакою.

Автор пропонує як базисну форму дослідження словотвору використовувати лінгво-інформаційний підхід. Суть його полягає в тому, що під час досліджень й аналізу складних слів використовуються не тільки традиційні лінгвістичні методи і прийоми, а й теорії штучного інтелекту, теорії інформації, комп'ютеризованої обробки лінгвістичної інформації тощо.

Метою даної статті є визначення специфіки та принципів гармонійності побудови складних номінацій на структурному рівні за допомогою лінгво-інформаційного підходу.

Використання кількісного підходу, який доцільно застосовувати в сучасній лінгвістиці в дослідженні закономірностей гармонійної побудови складних номінацій, є конче актуальним. Лінгвісти досить часто, крім структурного принципу, вдаються й до кількісних характеристик, коли досліджують, описують певні мовні явища, їхні взаємні залежності. Наприклад, М. В. Арапов вказує, що лінгвіст, якщо він виступає як практик, обов'язково розглядає мову не тільки у якісному, але й в кількісному аспекті. Важко знайти лінгвістичну роботу, у якій взагалі не було б елементарних підрахунків. Підраховується кількість фонем у конкретній мові, складів різноманітної структури, способів вираження граматичних значень, певних словотвірних засобів; встановлюється кількісне співвідношення різноманітних класів слів та ін. У прикладній лінгвістиці на таких підрахунках ґрунтуються різноманітні евристичні процедури: виявлення ключових слів, встановлення парадигматичних і синтагматичних відношень [1, с. 15]. Варто згадати й О. О. Леонтьєва, який відзначив, що кількісні оцінки присутні практично в будь-якому дослідженні природничих мов. Описуючи конкретну мову, її дослідник обов'язково відзначить підсумки трудомістких підрахунків, оформлених у вигляді таблиць і графіків, емпіричних формул [5, с. 19]. Досить вагомими є висловлювання проф. М. О. Ковалевського ще в 1877 р. Він писав: «Вимірювальними способами боролася фізіологія з віталізмом у фізичних явищах життя організму і переборола ненаукову гіпотезу, що тільки прикривала людські пізнання. Вимірювальними способами їй судилося розсіяти пітьму, яка панувала в розумінні психічних процесів. Перші промені світла в цьому темному царстві вже виблиснули. Необхідна лише наполеглива праця – і мова чисел розкриє перед нами таємничу роботу головного мозку» [4, с. 51]. Аналізуючи ці слова М. О. Ковалевського, Бодуен де Куртене писав про створення «найвищої математики», котра «оволодіє також психічними й соціально-психологічними явищами», і далі, «як тільки така математика

з'явиться – настане час для справжніх законів мовного світу – законів, що гідно посідають місце поряд із законами точних наук» [3, с. 45].

Останніми десятиліттями, у зв'язку з поширеним розвитком засобів обчислювальної техніки та прикладної математики, кількісні методи все частіше застосовуються лінгвістами як за кордоном, так і в Україні: на філологічних факультетах вивчають математичні методи в лінгвістиці, публікується серія «Квантитативна лінгвістика» в Німеччині, Сполучених Штатах Америки, Англії. Із 60-х років кількісні методи активно використовують і в україністиці.

Актуальною є проблема дослідження закономірностей побудови двокомпонентних іменників-комполітів та юкстаполітів в українській мові. Конкретизація теми дослідження підказана тим, що іменники-комполіти – це одна з найбільш багаточисельних і граматично містких категорій у загальній системі частин мови. Саме ця граматична категорія, поєднуючи в собі узагальнене поняття предметності, позначена семантичною об'ємністю і структурною різноманітністю моделей. Для об'єктивного оцінювання закономірностей того чи іншого явища дуже важливим є вибір способів і засобів дослідження, визначення його мети. Наукове пізнання передбачає визначення явищ також і через їхню кількісну характеристику. Якісна сторона об'єкту нерозривно пов'язана з їхнім кількісним виявом.

Застосування кількісного підходу для дослідження двокомпонентних іменників-комполітів і юкстаполітів вимагає одержання числових оцінок, які найбільш повно та якісно відображають характер закономірностей побудови кожного окремого слова, дозволяючи поширення цих закономірностей на великі групи слів. У цьому зв'язку дуже важливим є питання вибору оцінювального показника, який би створював найбільш цілісне уявлення про об'єкт дослідження.

Оскільки об'єктом дослідження слугує двокомпонентне складне слово, то його доцільно розчленувати на компоненти, причому, у загальному плані –

на великий і малий компоненти (крім випадків дихотомії). Як і слово в цілому, так і його компоненти розчленовуються на елементарні одиниці – склади [2, с. 15]. Першу спробу використати склад під час вимірювання довжини слова належить німецькому вченому Фуксу [7, с. 229]. Таким чином довжина слова, а також компонентів вимірюється кількістю складів. Розглянуті чинники виступають передумовою для вибору оцінювального показника, який дає можливість порівнювати довжини всього слова й великого компонента ( за кількістю складів). У ролі такого показника виступає пропорція-відношення, що визначається виразом:

$$\Pi = \frac{\text{кількість складів у двокомпонентному слові}}{\text{кількість складів у великому компоненті}}$$

Аналіз, проведений на базі словникових джерел для двокомпонентних іменників-комполітів і юкстаполітів власне-українського та іншомовного походження, показав, що пропорція  $\Pi$  має чітко окреслені значення. При цьому кількість складів у всьому слові й у великому компоненті дорівнює так званім  $p$ -числам Фібоначчі або числам Люка [6]. Що стосується відношення  $\Pi$ , то воно має значення, які дорівнюють числам «золотої»  $p$ -пропорції. Слід зазначити, що  $p$ -числа Фібоначчі зображуються у вигляді рядів:

- $p = 0;$  1, 2, 4, 8.....;
- $p = 1;$  1, 1, 2, 3, 5, 8.....;
- $p = 2;$  1, 1, 1, 2, 3, 4, 6, 9.....;
- $p = 3;$  1, 1, 1, 1, 2, 3, 4, 5, 7, 10..... .
- $p = 4;$  1, 2, 3, 4, 5, 6, 8, 11, 15..... .
- $p = 5;$  1, 2, 3, 4, 5, 6, 7, 9, 12..... .
- $p = 6;$  1, 2, 3, 4, 5, 6, 7, 8, 10, 13..... .

Параметр  $p$  характеризує зв'язок наступного члена ряду з попереднім. При  $p = 0$  кожен наступний член підсумовується із самим собою або шляхом

подвоєння попереднього. При  $p = 1$  кожен наступний член, починаючи з третього, дорівнює сумі двох попередніх. У цьому випадку маємо 1-числа Фібоначчі або просто числа Фібоначчі. При  $p = 2$  (2-числа Фібоначчі) кожен наступний член, починаючи з четвертого, дорівнює сумі попереднього й того, що відстоїть за один. При  $p = 3$  (3-числа Фібоначчі) кожен наступний член, починаючи з п'ятого, дорівнює сумі попереднього й того, що відстоїть за 2 і тому подібне.

У ряді Люка 2, 1, 3, 4, 7, 11, 18, ... кожен наступний член, починаючи з третього, дорівнює сумі двох попередніх (як і в числах Фібоначчі). Відношення  $\alpha_p$  сусідніх  $p$ -чисел Фібоначчі з їхнім зростанням наближається до «золотих»  $p$ -пропорцій, значення яких наведені в таблиці [6, с. 18]:

$p$	0	1	2	3	4	5
$\alpha_p$	2.0	1.618	1.463	1.381	1.325	1.287

Систематизована таблиця значень пропорцій  $\Pi$  для сімей «золотих»  $p$ -пропорцій виглядає так:

$P$	$\Pi$					
0	2/1	4/2	6/3	8/4	10/5	...
1	3/2	5/3	7/4	8/5	10/6	...
2	4/3	6/4	9/6	13/9	...	
3	5/4	7/5	10/7	14/10	...	
4	6/5	8/6	11/8	...		
5	7/6	9/7	12/9	...		
6	8/7	10/8	13/10	...		

«Золота» пропорція позначає об'єкт, будова яких відповідає певному ідеалу рівноваги, гармонії, сумірності їхніх складників. Вона виникає внаслідок розв'язання геометричної задачі про ділення одиничного відрізка в крайньому та середньому відношеннях. При цьому відрізок  $a = 1$  поділяється



на дві нерівні частини  $b$  і  $a-b$  так, що відношення меншої частини  $a-b$  більшої  $b$  дорівнює відношенню частини  $b$  до всього відрізка  $a$ , тобто  $(a-b):b = b:a$ . За цих умов отримуємо квадратичне рівняння  $b^2 - b - 1 = 0$  з додатнім коренем  $b_1 = (1 + \sqrt{5}):2 \sim 1,618$ . У загальному випадку  $\alpha_p$  стає наслідком розв'язання рівняння  $(p+1)$ -го степеня:  $b^{p+1} - b^p - 1 = 0$  [6, с. 12].

Залежно від кількісних показників для чисельника й знаменника у формулі  $\Pi$ , а також самого значення  $\Pi$ , слова, що аналізуються, можуть умовно бути віднесені до сімей відповідної «золотої»  $p$ -пропорції. Наприклад, у сім'ї дихотомії мають місце таке значення  $\Pi$ :

держфонд ( $\Pi = 2/1 = 2,0$ ); біосфера ( $\Pi = 4/2 = 2,0$ );

1 | 1                                      2 | 2

ідолопоклонник ( $\Pi = 6/3 = 2,0$ ); мінералоутворення ( $\Pi = 8/4 = 2,0$ );

3 | 3                                      4 | 4

У сім'ї «золотої» 1-пропорції наявні утворення з такими показниками:

автограф ( $\Pi = 3/2 = 1,5$ ); льододробарка ( $\Pi = 5/3 = 1,67$ );

2 | 1                                      2 | 3

вагонобудування ( $\Pi = 7/4 = 1,75$ ); радіовимірювання ( $\Pi = 8/5 = 1,6$ );

3 | 4                                      3 | 5

У сім'ї «золотої» 2-пропорції мали місце такі показники:

авіапункт ( $\Pi = 4/3 = 1,33$ ); автообладнання ( $\Pi = 6/4 = 1,5$ );

3 | 1                                      2 | 4

радіолабораторія ( $\Pi = 9/6 = 1,5$ ).

3 | 6

Сім'ю «золотої» 3-пропорції репрезентують такі композити:

паразитолог ( $\Pi = 5/4 = 1,25$ ); бджолозапилювання ( $\Pi = 7/5 = 1,4$ ).

4 | 1                                      2 | 5

Аналогічні показники виявилися й серед юкстапозитів. Наприклад, у сім'ї дихотомії мали місце слова:

хліб-сіть ( $\Pi = 2/1 = 2,0$ ); буркун-зілля ( $\Pi=4/2=2,0$ );

1 | 1

2 | 2

кіловат-година ( $\Pi=6/3=2,0$ );

будівники-комсомольці( $\Pi=8/4=2,0$ ).

3 | 3

4 | 4

У сім'ї «золотої» 1-пропорції наявні такі юкстапозити:

Іван-чай ( $\Pi = 3/2 = 1,5$ );

камінь-самоцвіт ( $\Pi = 5/3 = 1,67$ ).

2 | 1

2 | 3

У сім'ї «золотої» 2-пропорції мали місце:

гриб-боровик ( $\Pi = 4/3 = 1,33$ );

батьківщина-мати ( $\Pi = 6/4 = 1,5$ );

1 | 3

4 | 2

У сім'ї «золотої» 3-пропорції:

член-кореспондент ( $\Pi = 5/4 = 1,25$ ); люмпен-пролетаріат ( $\Pi = 7/5 = 1,4$ ).

1 | 4

2 | 5

Нами було проаналізовано 5990 слів, серед яких двокомпонентні композити налічують 5631 одиницю, а юкстапозити – 359 одиниць. Серед двокомпонентних композитів сім'я дихотомії становить 1840 одиниць (32,7%), сім'я «золотої» 1-пропорції – 2679 одиниць (47,6%), «золотої» 2-пропорції – 864 одиниці (15,3%), «золотої» 3-пропорції – 204 одиниці (3,6%), «золотої» 4-пропорції – 42 одиниці (0,7%). Серед юкстапозитів сім'я дихотомії становить 125 одиниць (34,8%), «золотої» 1-пропорції – 166 одиниць (46,2%), «золотої» 2-пропорції – 56 одиниць (15,6%), «золотої» 3-пропорції – 8 одиниць (2,2%), «золотої» 4-пропорції – 4 одиниці (1,1%). Слід відзначити, що відсотковий розподіл між сім'ями «золотих»  $p$ -пропорцій серед двокомпонентних композитів і юкстапозитів загалом збігається – за винятком випадків з малими обсягами вибірки (наприклад, для юкстапозитів у сім'ї «золотої» 4-пропорції – всього 4 одиниці).

Аналіз досліджуваного фактичного матеріалу дає підстави вважати наведений вище розподіл композитів і юкстапозитів закономірним, оскільки побудова двокомпонентних іменників-композитів та юкстапозитів пов'язана

із закономірностями функціонування людського мозку і відповідає певним пропорціям у процесі їхнього формотворення. Пропорції побудови розглянутих складних номінацій виражаються низкою конкретних чисел і підлягають закономірному гармонійного поділу – «золотого» перерізу. Тому нові складні слова, які з'являтимуться в мові, доповнюватимуть вказані сім'ї за принципом формотворення звукової оболонки цих лексичних одиниць, оскільки слова й мова взагалі – це продукт розумової діяльності людини. Механізм роботи людського мозку суттєво впливає на процес становлення й розвиток мови як засобу людського спілкування взагалі й появу новоутворень у мові.

Отже, лінгво-інформаційний підхід достатньо широко використовується в інженерній лінгвістиці, особливо під час комп'ютерної обробки текстів, зокрема складання частотних словників, відбирання лексики з частотного словника, складання статистичних моделей текстів на рівні словосполучень, алгоритмічного вибирання стійких словосполучень, розподілу лінгвістичних одиниць, ймовірно-статичного виявлення в тексті домінантних одиниць. При цьому поруч з елементами лінгвістичної теорії широко використовуються математичний апарат теорії ймовірностей і закон Ципфа-Мандельброта, а також під час описання системи мови в разі квантитативної й алгебраїчної лінгвістики застосовується теорія нечітких множин. З іншого боку мовознавчі теорії поширені в рамках лінгвістичного забезпечення автоматизованих систем науково-технічної інформації, зокрема для автоматизованого розпізнавання сенсу тексту та одиниць, що його уворюють, комп'ютерного приведення слововживань тексту до їхньої канонічної форми; комп'ютерного реферування тексту шляхом його компресії. У той же час слід зауважити, що застосування лінгво-інформаційного підходу для дослідження словотвору є недостатньо розвинутим. Особливо це стосується дослідження композитів та юкстапозитів на структурному та фонетичному рівнях.

## Література

1. Арапов М. В. Квантитативная лингвистика. – М.: Наука, 1988. – 184 с.
2. Азарова Л.Є. Структура композитів у концепції «золотої» пропорції // Мова та історія: Періодичний збірник наукових праць. – К., 1999. – С.13 – 18.
3. Бодуэн де Куртенэ И. А. Избранные труды по общему языкознанию. – М.: Наука, 1963. – Т.1. – 191 с.
4. Ковалевский Н. О. Как смотрит физиология на жизнь вообще и «психическую» в особенности // Уч. зап. Императорского Казанского ун-та. – 1877. – Т.1. – 345 с.
5. Леонтьев А. А. Язык, речь, речевая деятельность. – М.: Просвещение, 1969. – 214 с.
6. Стахов А. П. Алгоритмическая теория измерения. – М.: Знание, 1979. – 437 с.
7. Фукс В. Математическая теория словообразования // Теория передачи сообщений. – М.: Иностран. лит-ра, 1957. – С. 221 – 249.