

МЕТОДОЛОГІЯ ВИКОРИСТАННЯ ЛІНГВО-ІНФОРМАЦІЙНОГО ПІДХОДУ В ДОСЛІДЖЕННІ ГАРМОНІЙНОЇ ПОБУДОВИ СКЛАДНИХ НОМІНАЦІЙ

У статті обґрунтовано доцільність використання сучасних комп'ютерних технологій у рамках створення теорії гармонійної побудови складних слів в українській мові на засадах методології «золотої пропорції» та використання лінгво-інформаційного підходу; визначено основні принципи побудови складних слів на структурному й фонетичному рівнях.

Ключові слова: комп'ютерні технології, лінгво-інформаційний підхід, структурний та фонетичний рівні, «золота пропорція», композит, юкстапозит, гармонійні співвідношення.

В епоху комп'ютеризації та інформатизації змінюється методологія наукового пізнання: вдосконалюються його методи й використовуються саме ті, які витримали випробовування часом; з'являється нова методика, що дозволяє глибше охопити дослідження об'єктивного світу, явищ природи, пояснити їхню сутність.

Розвиток теорії мовознавства, зокрема словотвору, вимагає свого вибору методів і підходів, запозичення досягнень інших наук. Проведення експериментальних досліджень у мовознавстві повинно у свою чергу також передбачати використання методологічних прийомів і технічних засобів інформаційних технологій. Вибір базових форм залежить від конкретного напрямку лінгвістичних досліджень, характеру проблем і завдань, що стоять перед мовознавцями.

У теперішній час мова значно розширює свої комунікативно-суспільні функції. Продовжуючи залишатися найважливішим знаряддям комунікації людей, природна мова стає тією основою, на базі якої здійснюється спілкування між людьми з комп'ютером і його системами.

Бурхливий розвиток комп'ютерної техніки в останні десятиліття активно впливає на інформатизацію суспільства і сприяє поширенню використання комп'ютерних технологій в багатьох галузях людських знань, зокрема в мовознавстві. Проблемам комп'ютерної лінгвістики, одного з новітніх напрямів інформаційної технології сучасного мовознавства, присвячені публікації Агеєва В. М., Узілевського Г. Я., Андрющенка В. М., Баранова А. М., Волошина В. Г. та ін. Водночас ще не створено довершених універсальних теорій, які можна було б покласти в основу побудови практичних комп'ютерних програм. Тому розвиток нових комп'ютерних технологій у лінгвістиці є конче актуальним і необхідним.

Метою статті є обґрунтування доцільності використання комп'ютерних технологій у рамках створення теорії гармонійної побудови складних слів в українській мові на засадах методології «золотої пропорції» та використання лінгво-інформаційного підходу. Для досягнення цієї мети розв'язано такі завдання:

- розроблено основні положення теорії гармонійної побудови складних слів на засадах концепції «золотої пропорції»;
- запропоновано методологію дослідження гармонійної побудови складних слів, що спирається на лінгво-інформаційний підхід та сучасні комп'ютерні технології;
- визначено основні принципи побудови складних слів на структурному й фонетичному рівнях;
- висунуто гіпотезу про зв'язок виявлених гармонійних співвідношень із механізмом роботи мозку людини.

Предметом дослідження є складні слова в широкому розумінні, а саме: ті, що утворені способом складання – поєднанням двох або більше основ (композити) чи самостійних слів різного походження (юкстапозити); серед яких композитні іменники, прикметники, ті, що є

наслідком використання основоскладання та суфіксації: суфіксально-складні іменники та прикметники, а також суфіксальні деривати від складних твірних основ.

Для розв'язання вказаних завдань автор пропонує як базисну форму дослідження словотвору використовувати комплексний лінгво-інформаційний підхід, який ґрунтується на використанні інформаційних комп'ютерних технологій. Суть його полягає в тому, що під час досліджень й аналізу складних слів застосовуються не тільки традиційні лінгвістичні методи і прийоми, а й теорії штучного інтелекту, інформаційні технології опрацювання текстів.

Одним із таких комплексних підходів щодо розглянутої проблеми може слугувати методологія дослідження гармонії в лінгвістиці, зокрема у складних двокомпонентних композитах та юкстапозитах.

У методологічному аспекті для вимірювання довжини складного слова та його компонентів можна використовувати два підходи. Суть першого полягає в тому, що за одиницю вимірювання довжини беруть безпосередньо склад. При цьому довжина слова або його компонентів дорівнюватиме кількості складів. Цей підхід є зручним для аналізу номінативних одиниць, які писемно представлені в текстах.

Другий підхід полягає в тому, що за одиницю вимірювання довжини вибирають одиницю фізичної величини – часу, наприклад, мілісекунду. Це означає, що на фонетичному рівні номінативну одиницю розглядають у динаміці як відрізок потоку мовлення, яка акустично характеризується певною тривалістю. При цьому слід зазначити, що тривалість звучання композита (юкстапозита), як правило, не має сталого значення. Така тривалість залежить від низки чинників, зокрема, від характеру і темпу мовлення, який завжди певною мірою є індивідуальним у різних осіб, від інтонації мовця тощо. Уникнути вказаних незручностей можна, застосовуючи для аналізу відносний показник-пропорцію.

У випадку структурного рівня, коли номінативні одиниці зафіксовані у статичній формі й будь-які зміни їхньої довжини принципово неможливі, значення пропорції обчислюється за допомогою формули:

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

Для обчислення пропорцій побудови складних одиниць на фонетичному (акустичному) рівні треба чітко визначити центр їхнього поділу на компоненти у вигляді часової відмітки. Виміряні значення тривалості звучання цілого слова і великого компонента фіксуються. При цьому значення пропорції обчислюється з відношення:

$$P_{\phi} = \frac{\text{тривалість вимовляння двокомпонентного слова (мс)}}{\text{тривалість вимовляння великого компонента (мс).}}$$

Оскільки пропорція P_{ϕ} являє собою відносну величину, то такий суб'єктивний чинник як неоднаковий темп мовлення в різних осіб не може істотно впливати на точність оцінювання цього показника. Водночас, такі індивідуальні особливості окремої особи, як неоднакове почуття «гармонійності», «зсув логічного наголосу», можуть призвести до певного розмивання значень P_{ϕ} для слова, яке вимовляють різні особи. Тому для зменшення похибки в оцінюванні пропорцій звучання композитів і юкстапозитів під час проведення практичних дослідів було залучено фахівців-фонетистів, професійних дикторів. Саме в них правильно «поставлено» дикцію, вони чітко вимовляють слова, передусім їхні закінчення, мають розвинене почуття «часової пропорційності».

Систематизована таблиця значень пропорцій P , отримана в ході досліджень на структурному рівні, має вигляд:

Р	П
---	---

0	2,00	2,00
1	1,50; 1,67; 1,75; 1,60; 1,57; 1,625	1,618
2	1,33; 1,50; 1,44; 1,46	1,465
3	1,25; 1,40; 1,43	1,38
4	1,20; 1,38	1,32
5	1,17; 1,29	1,285
6	1,14; 1,30	1,26

Слід відзначити, що якщо в межах структурного підходу використовувалися цілочислові відношення типу: 3/2; 5/3; 8/5 і т.п., то в межах фонетичного для вимірювання часових інтервалів потрібно послуговуватися переважно дробовими показниками. Наведені в таблиці значення P_ϕ і P'_ϕ враховують p -числа Фібоначчі, числа Люка, а також значення «золотих» p -пропорцій [Соколов 1975: 38]. Як дробові показники слід також використовувати корені узагальненого рівняння «золотих» S -пропорцій [Стахов 2003: 38], зокрема, такі:

S	2	3	4	5
P_ϕ	1,618	1,84	1,93	1,97

Вимірювання тривалостей звучання композитів, юкстапозитів та їхніх компонентів, а також визначення центру поділу досліджуваної номінативної одиниці на компоненти здійснюється за допомогою спеціальної апаратури. До складу такої апаратури входять пристрої запису й відтворення звукових сигналів, принципи дії яких, у значній мірі, ґрунтуються на теорії цифрового оброблення мовних сигналів. Певні досягнення в цій галузі має Вінницький національний технічний університет, зокрема, відділ аналого-цифрових систем. Розроблена апаратура дає змогу досить точно (похибка не перевищує 1%) дослідити форму звукового сигналу (інтонограму), який виникає під час вимовлення тієї чи іншої номінативної одиниці, визначити центр поділу слова на компоненти, виміряти тривалість самого слова і його компонентів, зафіксувати дані в пам'яті комп'ютера. Озвучені номінативні одиниці, перетворені в цифрову форму і представлені на відповідному носії (компакт-диски, дискети), можна використовувати для подальших досліджень. Для їхньої реалізації достатньо наявності сучасного персонального комп'ютера, обладнаного засобами мультимедіа, зокрема, навушниками або звуковими колонками, а також спеціалізованого програмного забезпечення. У цьому плані досить зручним є пакет прикладних програм «Wave for Windows», який виявився досить зручним, оскільки він розрахований не лише на фахівців із комп'ютерної техніки чи звукотехніки, а й на пересічного користувача комп'ютерних технологій, зокрема, на фахівця-лінгвіста.

Використання комп'ютерних технологій на фонетичному рівні можна продемонструвати на прикладі дослідження слів *автотранспорт* і *хліб-сіль*:

437 мс 706 мс

авто | транспорт: $P_\phi = (437 \text{ мс} + 706 \text{ мс}) : 706 \text{ мс} \approx 1,618$

1143 мс

382 мс 554 мс

хліб | сіль: $P_\phi = 936 \text{ мс} : 554 \text{ мс} \approx 1,68$

936 мс

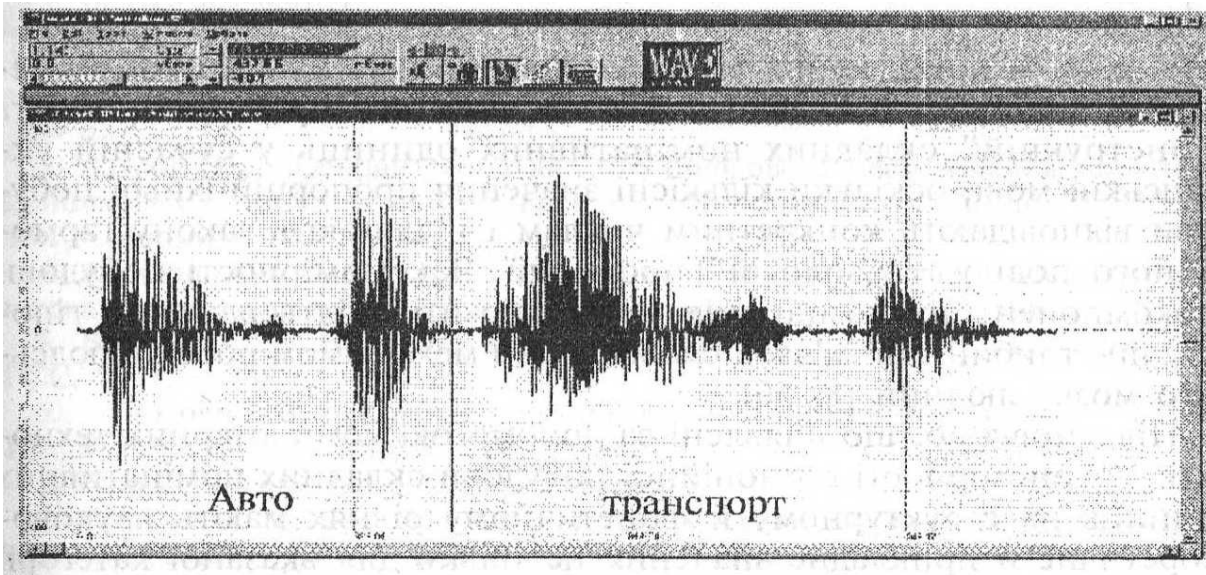


Рис.1. Комп'ютерна інтонограма слова «автотранспорт»

В інтонограмі слова *автотранспорт* (рис. 1) центр його поділу (часова позначка — вертикальна пряма лінія) припадає на значення 437 мс, загальна ж тривалість (довжина) становить 1143 мс. Отже, тривалість (довжина) великого компонента *транспорт* дорівнює $1143 \text{ мс} - 437 \text{ мс} = 706 \text{ мс}$. Отримане значення P_ϕ зіставили з числовими показниками таблиць для сімей «золотих» p - і S - пропорцій і зробили висновок, що *автотранспорт* належить до сім'ї класичної «золотої» 1-пропорції

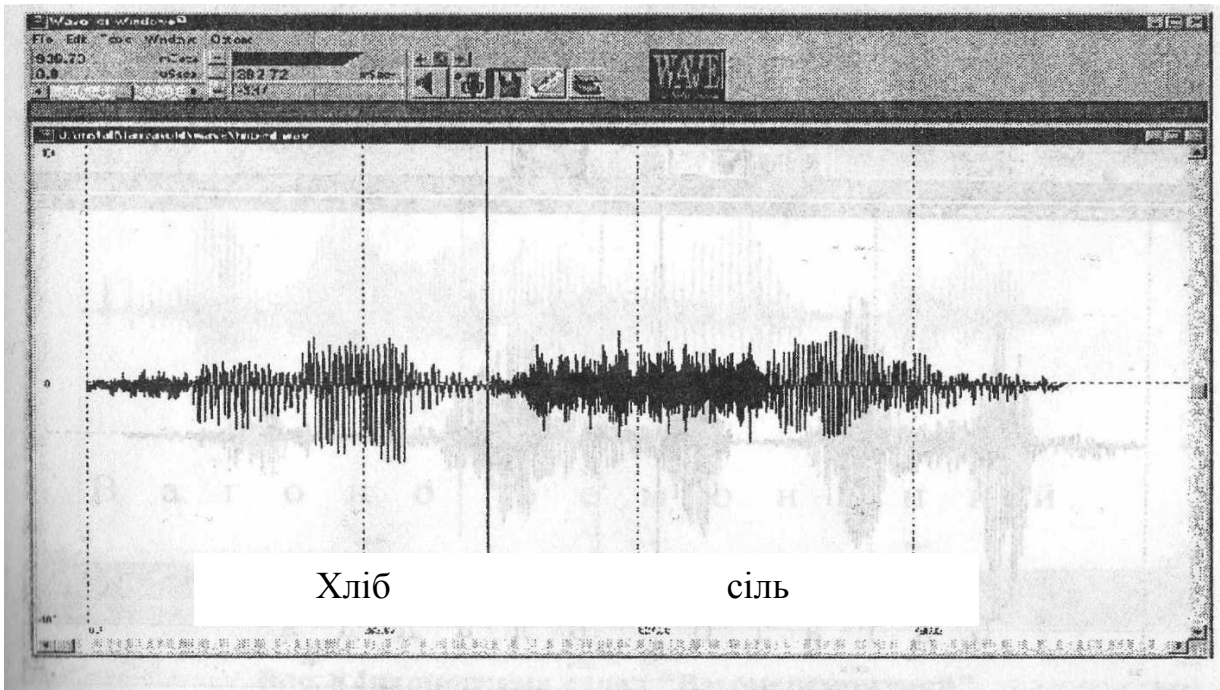


Рис.2. Комп'ютерна інтонограма слова «хліб-сіль»

В інтонограмі слова *хліб-сіль* (рис. 2) центр його поділу (припадає на значення 382 мс, загальна ж тривалість (довжина) становить 936 мс. Отже, тривалість (довжина) другого компонента *сіль* дорівнює $936 \text{ мс} - 382 \text{ мс} = 554 \text{ мс}$. Отримане значення P_ϕ зіставили з

числовими показниками таблиць для сімей «золотих» p - і S - пропорцій і з'ясували, що *хліб-сіль* теж належить до сім'ї класичної «золотої» 1-пропорції.

Композит

435мс 956 мс

авіа | промисловість

1391 мс

$P_{\phi} = (435 \text{ мс} + 956 \text{ мс}) : 956 \text{ мс} \approx 1,46$ належить до сім'ї «золотої» 2-пропорції, а юкстапозит

720 мс 620 мс

купівля-продаж

1340 мс

$P_{\phi} = (720 \text{ мс} + 620 \text{ мс}) : 620 \text{ мс} \approx 1,86$ – до сім'ї «золотої» 3-пропорції.

Варто зауважити також, що пропорції P_{ϕ} і P на структурному й фонетичному рівнях можуть відрізнятися. Так, наприклад, композит-іменник *водосховище* ($P_{\phi} = 1,33$) на фонетичному рівні входить до сім'ї «золотої» 2-пропорції, а на письмі він має пропорцію $P=5/3$ й належить до класичної «золотої» 1-пропорції. Юкстапозит *люкс-секунда* ($P_{\phi}=1,66$) на фонетичному рівні кваліфікуємо як одиницю сім'ї «золотої» 1-пропорції, а на структурному — ($P=4/3$) він належить до «золотої» 2-пропорції.

Проте пропорції більшості складних номінативних одиниць і під час вимови, й на письмі збігаються, утворюючи сім'ю класичної «золотої» 1-пропорції, наприклад:

3 : 2

дроворуб ($P=3/2$; $P_{\phi}=1,72$);

2 | 1

5 : 3

хата-читальня ($P=5/3$; $P_{\phi}=1,57$),

2 | 3

5 : 3

кінорежисер ($P = 5/3$; $P_{\phi} = 1,68$)

2 | 3

5 : 3

стежки-доріжки ($P=5/3$; $P_{\phi} = 1,75$)

2 | 3

Ми пропонуємо в межах лінгво-інформаційного підходу для глибшого і детальнішого аналізу гармонії звучання двокомпонентних складних слів у деяких випадках використовувати багатократний експеримент. Суть його полягає в тому, що те саме слово вимовляється декілька разів, але в різний спосіб, змінюючи темп вимовляння і логічний наголос, а також – різними дикторами. Наявні варіанти звучання досліджуваної номінації відповідним чином перетворюються за допомогою зазначеної вище апаратури оброблення звукових сигналів і фіксуються на цифрових носіях інформації. Після цього будь-який дослідник, фахівець із мовознавства, за допомогою персонального комп'ютера і засобів мультимедіа на основі інтонограм може наочно провести фонетичний (акустичний) аналіз відібраних композитів або юкстапозитів. Порівнявши отримані дослідним шляхом кількісні показники P_{ϕ} з табличними даними для сімей «золотих» p -пропорцій, можна зробити ґрунтовніший висновок про гармонійну побудову відповідної складної одиниці на фонетичному рівні.

Для прикладу можна навести результати гармонійної побудови двокомпонентного суфіксально-складного прикметника *білосніжний*. Під час дослідів було реалізовано шість варіантів вимови цього слова і отримано такі показники тривалості звучання та пропорцій P_{ϕ} .

I варіант:

357мс | 879мс

білосніжний $P_{\phi} = 1236\text{мс} : 877\text{мс} \approx 1,406$;

II варіант:

340мс | 744мс

білосніжний $\Pi_{\phi} = 1084\text{мс}:744\text{мс} \approx 1,457$;

III варіант:

278мс | 740мс

білосніжний $\Pi_{\phi} = 1018\text{мс}:740\text{мс} \approx 1,376$;

IV варіант:

415мс | 1085мс

білосніжний $\Pi_{\phi} = 1500\text{мс}:1085\text{мс} \approx 1,383$;

V варіант:

281мс | 681мс

білосніжний $\Pi_{\phi} = 962\text{мс}:686\text{мс} \approx 1,408$;

V варіант:

373мс | 863мс

білосніжний $\Pi_{\phi} = 1236\text{мс}:862\text{мс} \approx 1,43$.

Порівнявши отримані дані для Π_{ϕ} з табличними показниками для сімей «золотих» p -пропорцій, встановлюємо, що перший варіант належить до сім'ї «золотої» 2-пропорції, а інші п'ять – до сім'ї «золотої» 3-пропорції. Це підтверджує те, що у всіх випадках звучання слова *білосніжний* є гармонійним, хоч кількісні показники гармонії можуть відрізнятися. Слід також зауважити, що на структурному рівні цей композит належить до сім'ї дихотомії, оскільки пропорції побудови тут $\Pi = 4/2 = 2$.

Усі ці факти свідчать про те, що наявність на фонетичному рівні в досліджених композитах і юкстапозитах чітко визначених часових співвідношень типу сімей «золотих» p - і 5-пропорцій не є випадковою. Вона закономірна і зумовлена тим, що в процесі словотворення людський мозок гнучко розв'язує проблему побудови складних номінацій, які були б на фонетичному рівні «завадостійкими» і до того ж гармонійними. Цю завадостійкість і гармонійність забезпечує використання сімей «золотих» p - і S -пропорцій, які самі по собі вже є надлишковими і гармонійними.

Підсумки дослідження масивів слів, що належать до комп'ютеризованого словника, дають підставу твердити про гармонійність «конструкцій» складних номінативних одиниць у сучасній українській мові, оскільки кількісні значення пропорцій їхньої побудови відповідають конкретним числам і підлягають закону гармонійного поділу — «золотій» пропорції. Закономірності побудови двокомпонентних композитів і юкстанозитів підтверджують гіпотезу про глибинний зв'язок словотвору з механізмом роботи людського мозку [Соколов 1975: 56].

Застосування методів точних наук, зокрема формально-логічних та математичних методів, у лінгвістиці вже давно не викликає спротиву і в багатьох дослідженнях довело свою ефективність і перспективність. Використання логіки аналізу матеріалу, процедур, що вимагають іншого, відмінного від власне лінгвістичного, погляду на мовну систему, будову її одиниць, закономірності функціонування, відкриває можливості для одержання якісно нової, прихованої раніше від ока дослідника або інакше осмислюваної інформації про мову. Це зайвий раз доводить справедливність твердження Ф. де Соссюра про те, що погляд на об'єкт формує сам об'єкт. Це підтверджує також слова Луї Пастера: «Науки тільки виграють, якщо використовують запозичені одна в одній методи й факти» [Радо 1950: 180].

Особливо треба відзначити важливість використання лінгво-інформаційного підходу під час досліджень словотвору в рамках письмової й усної мови. Усна мова, як така, не являє собою відтворення письмового тексту. Це потік звуків, про які не можна сказати, що вони чітко розподілені на слова або мають чітку граматичну структуру. Усна мова, як правило, є суто особистою, вона слугує для розмови й складається в ході самої розмови. У той же час написаний текст у більшості випадків попередньо обмірковується і може читатися за відсутності автора. Усна мова пов'язана з часовим континуумом. Слухачі змушені сприймати її по мірі надходження, фразу за фразою. З метою наукового аналізу мови її записують і в пошуках структури досліджують окремі уривки. Подібні записи

передбачають апостеріорне вивчення мови, яка при цьому досліджується поза природним часовим континуумом. Як тільки мова виявляється записаною, наприклад, у вигляді інтонограми (синусоїди або імпульсу на екрані комп'ютера чи осцилографа), вона зображується у формі, аналогічній письму. Її можна вивчати в цілому або розбити на відрізки й елементи. Її можна інтерпретувати як послідовність складів або ще дрібніших одиниць. Важливо пам'ятати, що подібне зображення у вигляді дискретних елементів робиться лише з метою вивчення мови. Сама ж мова являє собою неперервний потік; між окремими її звуками немає чіткої межі. Фонетична транскрипція – це письмове символічне зображення, квантована послідовність дискретних символів, подібних друкованим літерам. Ці символи є квантованими одиницями для фонетистів, які застосовують їх для розв'язання особливої задачі – реєстрації тих особливостей голосу мовця, які істотні саме для нього [Азарова 2001].

На противагу усній мові письмовий текст значно легше аналізується, хоча при цьому також потрібно виконувати великий обсяг роботи. Письмовий текст призначається для громадського використання. Він заздалегідь обмірковується і в меншій мірі піддається зміненням під дією інтересів мовців. Різні умови, у яких використовуються письмова й усна мова, а також і різна мета, якій вони слугують, відображаються на їхній структурі. Слід зауважити, що 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, алгоритмічному вибиранні стійких словосполучень, розподілу лінгвістичних одиниць, ймовірно-статичному виявленні в тексті доміантних 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 [Агеев 1995].

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 композитів і юкстапозитів, зокрема двокомпонентних. Так під час автоматизованої обробки текстів аналізуються такі лінгво-інформаційні характеристики лексем, як довжина слова (у складах, звуках або літерах), статистична частота повторювань номінативних одиниць, типові елементи морфем (префікси, суфікси), а також перебираються можливі значення слів із метою визначення сенсу речень (фраз). У цьому випадку номінативна одиниця розглядається в кінцевому підсумку як елементарна складова в системній ієрархії оброблюваного тексту.

Отже, виявлені за допомогою комп'ютерних технологій закономірності гармонійної побудови складних номінативних одиниць на структурному і фонетичному рівнях мають науково-теоретичне і прикладне значення не тільки для вказаної категорії складних слів, а й словотвору в цілому.

Література

Агеев 1995: Агеев, В.Н. Человеко-компьютерное взаимодействие, процессы, модели [Текст] / В. Н. Агеев, Г. Я. Узилевский – М.: Мир книги, 1995. – 352 с.

Андрющенко 1980: Андрющенко В.М. Проблемы вычислительной лингвистики и автоматическая обработка текстов на естественном языке [Текст] / В. М. Андрющенко – М.: МГУ, 1980. – 184 с.

Баранов 2001: Баранов А.Н. Введение в прикладную лингвистику [Текст] / А. Н. Баранов – М.: Эдиторная УРСС, 2001. – 358 с.

- Волошин 2004: Волошин В.Г. Комп'ютерна лінгвістика: Навчальний посібник [Текст] / В. Г. Волошин – Суми: ВТД «Університетська книга», 2004. – 382 с.
- Радо В. Жизнь Пастера: Пер. с франц. [Текст] / В. Радо – М.: Наука, 1950. – 243 с.
- Соколов А.А. Математические закономерности электрических колебаний мозга [Текст] / А. А. Соколов, Я. А. Соколов. – М.: Наука, 1975. – 98 с.
- Стахов А.П. Новая математика для живой природы. Гиперболические функции Фибоначчи и Люка [Текст] / А. П. Стахов – Винница, 2003. – 264 с.

Азарова Лариса Евстахиевна

Методология лингво-информационного подхода в исследовании гармоничного построения сложных номинаций

В статье обоснована целесообразность использования современных компьютерных технологий в рамках создания теории гармоничного построения сложных слов в украинском языке на основе методологии «золотой пропорции» и использования лингво-информационного подхода; определены основные принципы построения сложных слов на структурном и фонетическом уровнях.

Ключевые слова: современные компьютерные технологии, лингво-информационный подход, структурный и фонетический уровень построения сложных слов, «золотая пропорция», композит, юкстапозит, гармоничные соотношения.

Azarova Larisa Evstahievna

Methodology of linguistic and informational approach in the study of harmonious building complex nominations

In the article the feasibility of using modern computer technology within the theory of building a harmonious complex words in Ukrainian methodology based on the «golden proportion» and use linguistic-information approach; the basic principles of the structural complex words and phonetic levels.

Keywords: modern computer technology, linguistic-informational approach, structural and phonetic level building compound words, «golden proportion» composite yukstapozyt, harmonious relationship.