

ВИКОРИСТАННЯ ЛІНГВО-ІНФОРМАЦІЙНОГО ПІДХОДУ В ДОСЛІДЖЕННІ СКЛАДНИХ НОМІНАЦІЙ НА ФОНЕТИЧНОМУ РІВНІ

Лариса Азарова (Вінниця, Україна)

У статті розглянуто фонетичний аналіз виявів «золотої» пропорції в композитах і юкстапозитах на акустичному рівні шляхом оцінювання співвідношень часових інтервалів, вимірювання яких здійснено за допомогою спеціальної апаратури цифрового опрацювання звукових сигналів на основі лінгво-інформаційного підходу.

Ключові слова: *фонетика, акустичний, «золота» пропорція, композити, юкстапозити, лінгво-інформаційний підхід.*

In the article the phonetic analysis of «golden» proportion in composite and juxtaposite words at the acoustic level by the time interval correlation valuation has been done. The duration of the phonation have been measured with the special apparatus for digital signal processing on the lingvo-information's base.

Key-words: *phonetics, acoustic, «golden» proportion, composite words, juxtaposite words, lingvo-information's base.*

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 в теперішній час і в перспективі.

У нашу епоху – епоху комп'ютеризації та інформатизації змінюється методологія наукового пізнання: вдосконалюються його методи і використовуються саме ті, які витримали випробування часом. З'являється нова методика, яка дозволяє глибше охопити дослідження об'єктивного світу, явищ природи, пояснити їхню сутність.

Ми пропонуємо як базову форму дослідження словотвору використовувати лінгво-інформаційний підхід. Суть його полягає в тому, що під час досліджень і аналізу складних слів використовуються не тільки

традиційні лінгвістичні методи і прийоми, а й теорії штучного інтелекту, теорії інформації, комп'ютеризованого оброблення лінгвістичної інформації тощо. Лінгво-інформаційний підхід ґрунтується на поєднанні як традиційних лінгвістичних методів і прийомів словотвору, так і запозичених методів і прийомів інформаційних технологій. У цьому й полягає актуальність цього дослідження. Цій проблемі присвячено велику кількість наукових праць [6-13]. У той же час слід зауважити, що лінгво-інформаційний підхід до вивчення словотвору використано ще недостатньо. Особливо це стосується дослідження закономірностей складних слів на фонетичному рівні.

Метою даної статті є аналіз закономірностей побудови складних слів на фонетичному рівні за допомогою лінгво-інформаційного підходу.

Особливо треба відзначити важливість використання лінгво-інформаційного підходу під час досліджень словотвору на рівні писемної й усної мови. Усне мовлення, як таке, не є відтворенням писемного тексту. Це потік звуків, про які не можна сказати, що вони чітко розподілені на слова або мають чітку граматичну структуру. Усне мовлення є здебільшого суто особистим, воно слугує для розмови й складається у процесі самої розмови. Водночас написаний текст у більшості випадків попередньо обмірковується і може бути прочитаний за відсутності автора. Усне мовлення пов'язане з часовим континуумом. Слухачі змушені сприймати його, як воно надходило, фразу за фразою. З метою наукового аналізу мовлення записують і в пошуках структури досліджують окремі уривки. Подібні записи передбачають апостеріорне вивчення мовлення, яке досліджують поза природним часовим континуумом. Як тільки мовлення виявляється записаним, наприклад, у вигляді інтонограми (синусоїди або імпульси на екрані комп'ютера чи осцилографа), воно зображується у формі, аналогічної до письма. Його можна вивчати в цілому або розбити на відрізки й елементи, інтерпретувати як послідовність складів або ще дрібніших одиниць. Важливо пам'ятати, що подібне зображення у вигляді дискретних елементів роблять лише з метою вивчення мови. Саме ж мовлення являє собою безперервний потік; між

окремими його 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 особливого завдання – реєстрації тих особливостей голосу мовця, які істотні саме для нього [9; 10].

На противагу усному мовленню писемний текст значно легше аналізувати, хоч тут також потрібно виконати великий обсяг роботи. Писемний текст призначений для громадського використання. Його заздалегідь обмірковують, і він меншою мірою зазнає змін залежно від інтересів мовців. Різні умови, у яких використовують писемне й усне мовлення, а також і різна мета, якій вони служать, позначаються на їхній структурі.

Розмовне мовлення рідко буває «правильним» щодо граматики й синтаксису. Фрази можуть бути незакінченими, деякі слова або сполучення слів повторюють декілька разів і в різних варіантах. Під час писання становище інше. Той, хто пише, не може спостерігати за своїми читачами, може лише уявити їхні можливі вагання. Тому він заздалегідь обмірковує текст і здебільшого суворіше дотримується правил. Для відбору потрібного слова та оцінювання різних варіантів граматичної побудови людина може витратити стільки часу, скільки їй знадобиться [7; 5].

На противагу цьому, усне мовлення будується на основі порівняно обмеженого словника (статистичні дослідження засвідчили, що англійці й американці в 96% телефонних розмов використовують не більше 737 різних слів [8: 290-324]; але при цьому слова можуть вільніше сполучуватися в різні структури з повтореннями, акцентуванням, з великою різноманітністю форм та відповідною надлишковістю. Під час писання нам доводиться чимось компенсувати відсутність жестів й інтонацій, якщо ми хочемо запобігти неоднозначності. Для цього треба вводити мовну надлишковість шляхом значного розширення словника і суворішого дотримання правил граматики [8; 15].

Слід зауважити, що лінгво-інформаційний підхід досить широко використовується в прикладній лінгвістиці, особливо під час комп'ютерного оброблення текстів, зокрема складання частотних словників, відбору лексики з частотного словника, складання статистичних моделей текстів на рівні словосполучень, алгоритмічного відбору стійких словосполучень, розподілу лінгвістичних одиниць, ймовірно-статичного виявлення в тексті доміантних одиниць. Паралельно з елементами лінгвістичної теорії широко використовуються математичний апарат теорії ймовірностей і закон Ципфа-Мандельброта [15]. Під час опису системи мови в рамках квантитативної й алгебраїчної лінгвістики застосовують також теорію нечітких множин [14; 4].

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 утворюють; комп'ютерного наближення слововживань тексту до їхньої канонічної форми; комп'ютерного реферування тексту шляхом його компресії [13]. Найскладнішою проблемою використання інформаційних технологій для автоматизованого оброблення текстів є комп'ютерний переклад, що охоплює лінгвістику тексту, а також лексичний, семантичний та граматичний переклади. Цим питанням присвячено велику кількість наукових праць у кінці ХХ ст. [1; 2; 3; 5; 12]. Під час автоматизованого оброблення текстів аналізують 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 перебирають можливі значення слів із метою вивчення сенсу речень (фраз). Номінативну одиницю розглядають як елементарну складову в системній ієрархії оброблюваного тексту.

У свою чергу такі номінативні одиниці як композити та юкстапозити також являють собою лінгвістичні системи, тільки нижчого рівня відносно речень (фраз) або текстів. Порівняно з простими лексемами вони є більш інформативними з погляду виявлення своїх значень. Це пов'язано з тим, що

кількість можливих значень у складних словах буде меншою, ніж у простих номінаціях, значення яких набагато більше залежить від сенсу речення. З іншого боку, і сенс самого речення істотно залежить від значень його складників. Особливо наочно це можна пояснити, аналізуючи якість комп'ютерного перекладу текстів, написаних іноземними мовами за допомогою спеціальних прикладних програм-автоперекладачів. Якість такого перекладу залежить насамперед від повноти наборів значень слів, які представлені в словнику (словниках), а також від чіткості розмежування самих тематичних словників згідно з потрібною галуззю знань. Практичний досвід засвідчує, що недотримання цих правил, навіть у рамках близькоспоріднених слов'янських мов (українська-російська), породжує неякісний переклад. Якщо ж мови належать до різних сімей (українська-англійська), то перекладені речення можуть являти собою повну нісенітницю.

Поєднання у складній номінативній одиниці декількох основ чи слів конкретизує значення першої з акцентом на опорний компонент. Отже, замість існування багатьох значень в окремих компонентах складна (об'єднана) номінація має здебільшого менше сенсових значень, а іноді тільки єдине. Тому використання складних номінацій спрощує розпізнавання сенсу речення як людиною, так і автоматизованою системою оброблення текстів.

Відкритими залишаються питання: які використовують способи об'єднання компонентів складних слів в єдину конструкцію і наскільки вони є гармонійними; які простежуються закономірності побудови складних номінацій і якими є сполучувальні можливості основ у двокомпонентних складних словах.

Лінгво-інформаційний підхід на рівні структурного аналізу ґрунтується, з одного боку, на засадах лінгвістичної теорії словотвору, а з іншого – на використанні деяких положень теорії інформації та штучного інтелекту. На фонетичному рівні вказаний підхід передбачає, крім згаданого вище, використання спеціалізованої інформаційно-вимірювальної системи мовних сигналів на основі персонального комп'ютера. Загальною метою досліджень є

виявлення за допомогою лінгво-інформаційного підходу закономірностей гармонійної побудови складних слів. Тому методологія цих досліджень використовує також філософію гармонії і як напрям цієї філософії – концепцію «золотої пропорції».

Цифрове оброблення аналогових (звукових) сигналів почало широко використовуватися в останні десятиліття у зв'язку з появою цифрових обчислювальних машин, зокрема високопродуктивних персональних комп'ютерів. Крім персонального комп'ютера, до каналу цифрового оброблення звукових сигналів здебільшого входять: мікрофон (мікрофони), підсилювачі, фільтри, аналого-цифрові та цифро-аналогові перетворювачі, цифровий сигнальний процесор тощо. Певні досягнення в галузі цифрового оброблення звукових сигналів відділу аналого-цифрових систем має Вінницький національний технічний університет. Згадана апаратура дає змогу досить точно (похибка не перевищує 1.0%) дослідити форму звукового сигналу (інтонограму), який виникає під час вимовляння того чи іншого складного двокомпонентного слова, визначити центр поділу номінативної одиниці на компоненти, виміряти тривалість самого слова і його компонентів, зафіксувати дані в пам'яті комп'ютера. Для обчислення пропорцій побудови досліджуваних складних одиниць на фонетичному (акустичному) рівні треба передусім чітко визначити центр їх поділу на компоненти у вигляді часової відмітки. Далі за цією відміткою вимірюємо довжину (тривалість) великого компонента. Виміряні значення тривалості звучання цілого слова й великого компонента фіксуємо. Кількісне значення пропорції на фонетичному рівні обчислюємо з відношення:

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

у якому індекс ϕ означає пропорцію на фонетичному рівні. 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 в оцінюванні цього показника. Водночас слід пам'ятати, що такі індивідуальні особливості окремої особи, як дещо неоднакове почуття «пропорції гармонійності», «зсув логічного наголосу в слові», можуть призвести до певного розмивання значень Π_{ϕ} для слова, яке вимовляють різні особи. Тому для зменшення похибки в оцінюванні пропорції звучання композитів і юкстапозитів під час проведення практичних дослідів треба залучати фахівців-фонетистів, професійних дикторів. Саме в них правильно «поставлено» дикцію, вони чітко вимовляють слова, передусім їхні закінчення, мають розвинене почуття «часової пропорційності». Тому немає нічого дивного в тому, що красива, правильна вимова, у якій реалізоване гармонійне часове співвідношення звуків і складів, – це справжнє мистецтво.

Набір кількісних показників для оцінювання Π_{ϕ} , під час дослідження складних слів на фонетичному рівні повинен бути аналогічним до того, що використовується у процесі структурного аналізу. Він має охоплювати ті співвідношення, які входять до складу сімей «золотих» p -пропорцій. Якщо в межах структурного підходу (вимірювання в складах) зручно було використовувати цілочислові відношення типу: $3/2$; $5/3$; $8/5$ і т.д., то в межах фонетичного для вимірювання часових інтервалів потрібно послуговуватися переважно дробовими показниками. Це пов'язано з тим, що відношення виміряних часових інтервалів у двокомпонентних словах є здебільшого дробовим числом. Систематизована таблиця значень дробових показників Π_{ϕ} для сімей «золотих» p -пропорцій, що враховує p -числа Фібоначчі та числа Люка, спираючись на раніше наведені дані, має таку форму:

P	П_ф	П_ф[*]
0	2.00	2.00
1	1.50; 1.67; 1.75; 1.60; 1.57; 1.625	1.62
2	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

У цій систематизації **П_ф** відповідає значенням коренів узагальненого рівняння «золотих» *p*-пропорцій. Як дробові показники слід також використовувати корені узагальненого рівняння «золотих» *S*-пропорцій, зокрема такі:

	S	3	4	5
П_ф	1.84	1.93	1.97	

З огляду на все це ми пропонуємо таку послідовність дослідження пропорцій побудови композитів і юкстапозитів у процесах мовлення.

На першому етапі проводимо систематизацію відібраних складних слів за різними класифікаційними ознаками: за характером походження, за належністю до тієї чи іншої частини мови і т.д.

Другий етап передбачає процедуру «озвучування» відібраних і систематизованих двокомпонентних складних слів. Під час практичної

реалізації другого етапу нам допомагали диктори-фонетисти, які послідовно вимовляли кожне слово, відібране для аналізу, а звуковий сигнал, що його «генерує», перетворювався в електричні коливання (інтонограми) і фіксувався в пам'яті комп'ютера за допомогою спеціальної апаратури цифрового звукозапису. Для проведення досліджень на цьому етапі було використано можливості студії звукозапису Вінницької телерадіокомпанії, а також науково-технічного відділу аналого-цифрових систем Вінницького національного технічного університету.

На третьому етапі ми провели фонетично-часовий аналіз досліджуваних номінативних одиниць на матеріалі інтонограм. Вимірювали загальну тривалість звучання композита або юкстапозита, а також тривалість звучання великого компонента. Особливо важливим під час виконання цієї процедури було максимально точне визначення центру поділу (часової «відмітки») слова на компоненти, тобто момент переходу від першого компонента до другого. Розташування цієї часової відмітки повинно припадати на ту грань, якою закінчується останній склад першого компонента й починається перший склад другого компонента. Згадані вище вимірювання уже не вимагали (як на другому етапі) спеціальної студійної апаратури. Для їх реалізації достатньо сучасного персонального комп'ютера, обладнаного засобами мультимедіа, зокрема навушниками або звуковими колонками. Для керування функціонуванням комп'ютера в цьому режимі потрібне лише спеціальне прикладне програмне забезпечення. Уже створено прикладні комп'ютерні програми, які можна використовувати з цією метою. Ми використали пакет прикладних програм «**Wave for Windows**», який виявився практично досить зручним, бо він розрахований не лише на фахівця з комп'ютерної техніки чи звукотехніки, а й на звичайного користувача комп'ютерних технологій, зокрема на фахівця-лінгвіста.

Підсумком третього етапу було отримання кількісних значень тривалості звучання всього двокомпонентного слова й великого компонента. Приклади

вимірювання вказаних характеристик у словах *сон-трава* і *хлібозакупівля* за допомогою комп'ютерних інтонограм подані на рис. 1; 2.

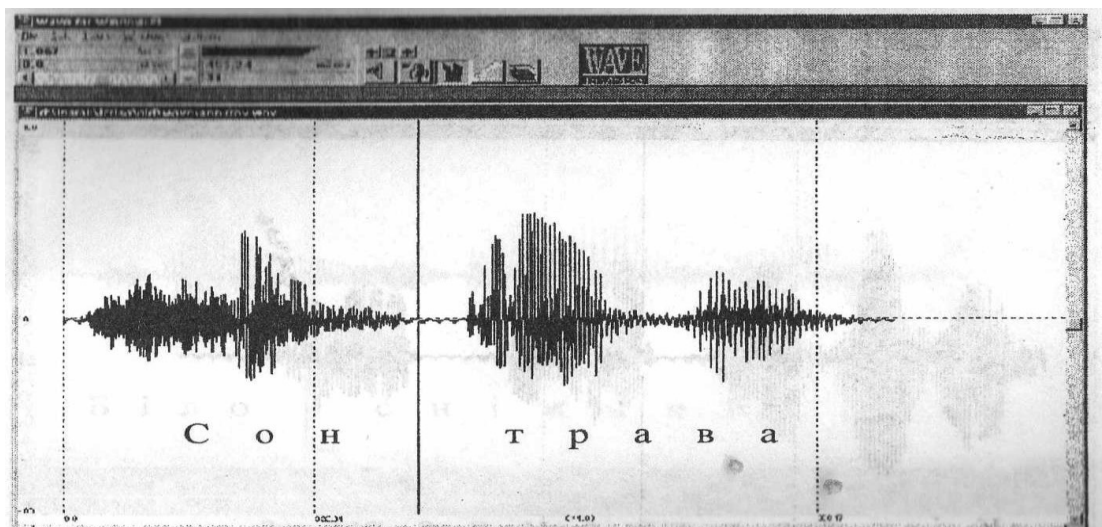


Рис. 1

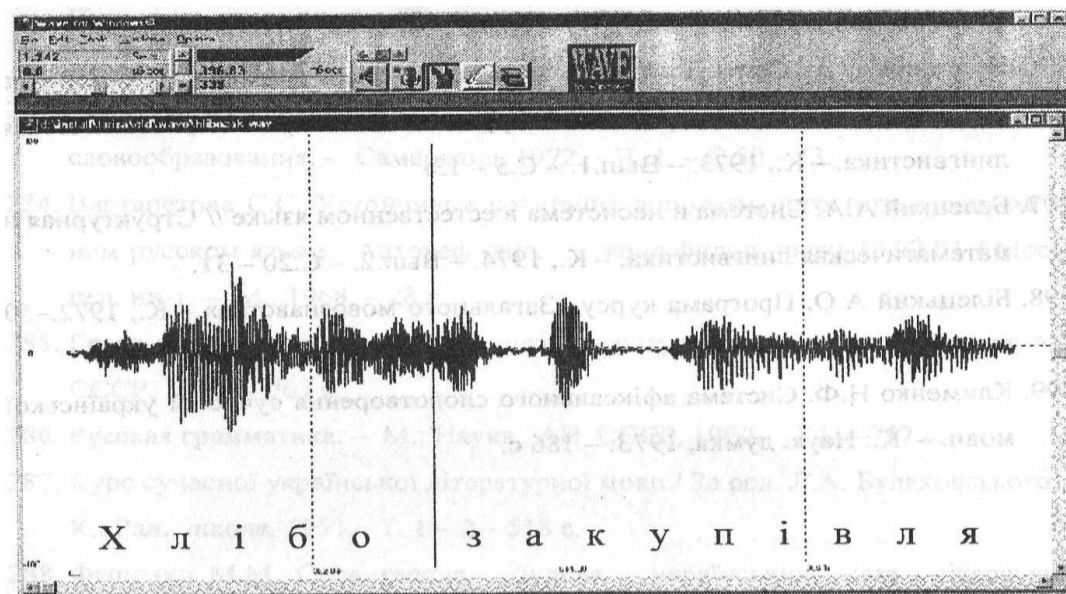


Рис. 2

В інтонограмі слова *сон-трава* (рис. 1) центр його поділу (часова відмітка – вертикальна пряма лінія) припадає на значення 455 мс, загальна ж тривалість (довжина) становить 1067 мс. Отже, тривалість (довжина) великого компонента *трава* дорівнює $1067 \text{ мс} - 455 \text{ мс} = 612 \text{ мс}$. У слові *хлібозакупівля* (рис. 2) центр поділу припадає на значення 396 мс, а загальна тривалість дорівнює 1242 мс.

Відповідно тривалість великого компонента *закупівля* становить 1242 мс - 396 мс = 846 мс.

На завершальному, четвертому, етапі ми визначили пропорції побудови двокомпонентних складних слів, обчисливши відношення відповідно до наведеної основної формули для Π_{Φ} . Далі отримане значення Π_{Φ} зіставили з числовими показниками таблиць для сімей «золотих» p - і S - пропорцій. За результатами цього кваліфікаційного аналізу досліджувані двокомпонентні слова занесено до складу відповідних сімей.

Так, наприклад, пропорція побудови юкстапозита належить до сім'ї «золотої» 1-пропорції

455 мс 612 мс

Сон | трава $\Pi_{\Phi} = 1067 \text{ мс} : 612 \text{ мс} \sim 1,74$

Пропорція побудови композита *хлібозакупівля* належить до сім'ї «золотої» 2- пропорції:

396 мс 846 мс

Хлібо | закупівля $\Pi_{\Phi} = 1242 \text{ мс} : 846 \text{ мс} \sim 1,47$

Інтонограми досліджуваних складних слів засвідчують, що переважна більшість (47 композитів і юкстапозитів) належать до сім'ї «золотої» 1-пропорції, 17 слів – до сім'ї «золотої» 2-пропорції, 5 слів – до сім'ї «золотої» 3-пропорції. Крім того, 3 слова нараховує сім'я «золотої» $S=3$ -пропорції та по одному слову сім'ї «золотої» $S=3$ - і 5-пропорцій.

Наявність у побудові композитів та юкстапозитів на фонетичному (акустичному рівні) рівні чітко визначених гармонійних пропорцій і часових співвідношень не є випадковим. З точки зору лінгво-інформаційного підходу це є закономірним і пов'язано з тим, що в процесі словотворення людський мозок гнучко розв'язує проблему «конструювання» складних номінацій, які були б гармонійними та зручними, а на фонетичному рівні – ще й завадостійкими.

Отже, фонетичний аналіз має свої методики визначення пропорцій побудови композитів і юкстапозитів. Вони використовують систему кількісних

показників гармонії у вигляді сім'ї дихотомії, сімей «золотих» пропорцій, а також р-чисел Фібоначчі. Механізм роботи головного мозку людини пов'язаний з гармонійністю словотворення. Це засвідчили проаналізовані моделі розумової діяльності мовця в рамках лінгво-інформаційного підходу, що пояснюють появу наявних простих і складних слів. Оскільки сама мова і її елементи є продуктом розумової діяльності мовця, то сприйняття людиною гармонії на рівні абстрактного мислення, її почуттів відбивається на побудові складних номінацій у вигляді гармонійних пропорцій між довжиною слова та його компонентами.

БІБЛІОГРАФІЯ

1. Вертель В. А. Автоматические словари в системе бинарного вероятностного МП / В. А. Вертель, Е. А. Вертель, Кричевич и др. // Инженерная лингвистика: Уч.зап. ЛГПИ им. Герцена. – Л., 1971. – С.135 – 176.
2. Вертель В. А. Алгоритм получения частотного словаря с учетом длины словоформ / В. А. Вертель, Е. А. Вертель – М.: Наука, 1970. – 189 с.
3. Гончаренко В. В. Лексикографические, лингвистические и инженерно-лингвистические вопросы построения автоматического словаря (англо-русский автоматический словарь по полупроводникам): Автореф. дис. ... канд. фшол. наук: 10.02.04 / В. В. Гончаренко / Кишинев, гос. пед. ин-т. – Кишинев, 1972. – 28 с.
4. Заде Л.А. Основы нового подхода к анализу сложных систем и процессов принятия решений / Заде Л.А. // Математика сегодня. –1974. – № 7. – С.65 – 93.
5. Мельничук И. А. Автоматический перевод / И. А. Мельничук, Р. Д.Равич. – М.: Наука, 1967. –293 с.
6. Bloch B. Outline of Linguistic Analysis, Linguistic Society of America / B. Bloch, G. L Trager – Baltimore: Waverley Press, 1942. – 227 p.
7. Bodmer F. The Loom of Language / Bodmer F. – New York: W.W. Norton & Co., 1944. – 289 p.
8. French N. R. The Words and Sounds of Telephone Conversations / N. R. French, C. N. Carter, W. Koenig. – Baltimore: Bell System Tech. J.9., 1930. – 298 p.
9. Jones D. An outline of English Phonetics / Jones D. – Cambridge, England: W. Heffer & Sons. – 1950. – 211 p.
10. Jones D. The Phoneme: Its Nature and Use / Jones D. – Cambridge, England: W. Heffer & Sons. – 1950. – 175 p.

11. Miller G. A. Language and Communication / Miller G. A. – New York: McGraw-Hill Book Co., 1951. – 255 p.
12. Pershke S. Possibilities of Further Development of Automatic Language Translation // *Pensiero e linguaggio in operazioni* / Pershke S. – Milano. – 1970. – J.1. – P.76 – 109.
13. Piotrowski R. Automatic Pattern Recognition applied to semantic problems. Computational and mathematical linguistics R. Piotrowski, L. Pabilina // *Proceedings of the 1973 International Conference on Computational Linguistics*. – Pisa, 1974. – P. 64 – 87.
14. Zadeh L.A. Outline of a New Approach to the Analysis of Complex Systems and Decision Processes / Zadeh L.A.// *IEEE TS MC*. – 1973. – Vol. SMC-3. Jan. – P. 5 – 98.
15. Zipf G. K. Human Behavior and the Principle of Least Effort / Zipf G. K. – Cambridge, Mass: Addison Wesley Publishing Co., 1949. – 356 p.

ВІДОМОСТІ ПРО АВТОРА

Лариса Азарова – доктор філологічних наук, професор, завідувач кафедри мовознавства Вінницького національного технічного університету.

Наукові інтереси: словотвір на діяхронічному та синхронічному рівні, науково-технічна термінологія.