

**А. А. ЛУЧИК**

Національний університет «Києво-Могилянська академія»  
Київ, Україна  
Електронна пошта: allal@meta.ua  
<https://orcid.org/0000-0002-4218-355X>

**О. С. ПАЛЬЧЕВСЬКА**

Львівський державний університет безпеки життєдіяльності  
Електронна пошта: palch56@ukr.net  
<https://orcid.org/0000-0002-2090-4870>

## **МОВНА РІЗНОМАНІТНІСТЬ В УКРАЇНСЬКОМУ НАЦІОНАЛЬНОМУ ЛІНГВІСТИЧНОМУ КОРПУСІ**

У статті описані властивості побудованого на текстових матеріалах Українського національного лінгвістичного корпусу (далі — УНЛК) обсягом близько 20 млн слововживань мікрополя «мови світу». Репрезентативність, збалансованість за жанрами і функційними стилями текстів корпусу, а також можливість використання повнотекстового пошукового інструментарію дозволили отримати максимально достовірні дані про функціонування номінацій мов світу в українській мові. За допомогою кількісних підрахунків референтів мов світу в УНЛК утворено ядро, навколоядерна, напівпериферійна і периферійна зони мікрополя «мови світу». Ядро мікрополя сформоване найчастотнішими у корпусі одиницями, навколоядерна зона — менш уживаними, периферійна — маловживаними, віддалена периферія — поодинокими одиницями. Встановлено, що всі складники ядра мікрополя «мови світу» є референтами індоєвропейських мов. Навколоядерну зону формують номінації на позначення мов індоєвропейської, тюркської, сино-тибетської, афро-азійської, фіно-угорської та японсько-рюкюської сімей. До периферійної зони увійшли одиниці, які репрезентують мови індоєвропейської, тюркської, фіно-угорської, картвельської, корейської, афро-азійської, нігеро-конголезької та тупі сімей. Кількісні підрахунки номінацій УНЛК, які увійшли у досліджуване мікрополе «мови світу», висвітлили вплив екстралінгвальних факторів на особливості моделювання мовної інформації, серед яких: генетичні витоки мов світу, близькість культур, ментальності, вірувань, територіальне контактування народів, особливості їхньої міжмовної взаємодії, економічна й політична вага країн, мешканці яких є носіями тієї чи іншої мови. Побудоване мікрополе відкриває шляхи для нових спостережень за мовою крізь призму самої мови та актуалізує її властивості щодо цього на сучасному етапі функціонування.

Ключові слова: мови світу, українська мова, лексико-семантичне поле, ядерна зона, навколоядерна зона, напівпериферійна зона, периферійна зона, Український національний лінгвістичний корпус.

З метою дослідження української мови та укладання сучасних словників в Українському мовно-інформаційному фонді НАН України протягом десятиліть створювався Український національний лінгвістичний корпус (далі — УНЛК). Сьогодні він уже має обсяг близько 20 млн слововживань [ЛІС 4 : 17–18] у текстах класичної і сучасної художньої літератури (проза і поезія), словників (близько 240), мемуарно-бібліографічної літератури, журнальної публіцистики і літературної критики, газетної публіцистики і новин, а також літератури наукової, науково-популярної й навчальної, релігійної й філософсько-релігійної, документів офіційно-ділової, виробничо-технічної та юридичної діяльності. Обсяг ресурсів необмежений, корпус постійно поповнюється новими даними.

У системі УНЛК сформовані дві підсистеми: бібліографічна та повнотекстова. Бібліографічна підсистема є інструментом для збирання, збереження, моделювання й використання інформації про природну мову у цифровому форматі. Друга підсистема дає можливість здійснювати повнотекстовий пошук тієї чи тієї мовної одиниці або фрагмента тексту з максимально бажаною кількістю слів. Повнотекстовий пошук супроводжується додатковими параметрами, ними можуть бути: урахування порядку слів, пошук у певній підмножині об'єктів, використання процедури лематизації, лексикографічної бази даних, синонімічних рядів, вибір граматичних параметрів для кожного слова обраного фрагмента тексту.

Формування підмножини об'єктів, тобто так званого «кошика» (тимчасового сховища) дає можливість користувачеві зберігати певну інформацію для подальшої роботи в інших сеансах. Користуючись таким інструментом, дослідник може відібрати джерела певного автора, стилю чи жанру і працювати лише з цією частиною корпусу. Така процедура дає можливість виділити із загального корпусу свій власний підкорпус, зорієнтований на розв'язання особистих завдань. Зазначений підхід дозволив відмовитися від занадто суб'єктивної ідеї створення «еталонного корпусу» і надав можливість досліджувати засобами корпусу «нееталонні», аномальні мовні явища [Лучик, Остапова : 33–37].

Хоча в Українському мовно-інформаційному фонді корпус використовується передусім як джерельна база лінгвістичної інформації для створення фундаментальної багатотомної лексикографічної системи «Словники української мови», сфера застосування УНЛК значно ширша. Активне використання у сьогоденній науці корпусної методології, як уважають Дж. Егберт і П. Бейкер, пов'язане з розумінням того, що корпуси надають об'єктивні свідчення про «реальну мову» [Egbert, Baker, 1]. Зважаючи на те, що колекція текстів тут характеризується репрезентативністю, збалансованістю за жанрами і стилями, отримані за допомогою повнотекстового пошукового інструментарію результати надають максимально достовірні дані про мову, а за її допомогою і про факти навколишньої дійсності. Безсумнівно, це сприяє розв'язанню нагальних питань лінгвістики, зокрема побудові і мовної та концептуальної картин світу, лексико-семантичного поля тощо.

Природне прагнення людини отримати цілісну картину світобудови зумовлює обстеження мови в усіх можливих її виявах як об'єктивної джерельної бази для реалізації поставленої мети, оскільки саме мова як одна з найголовніших ознак людини є тією рушійною силою, яка «керує нашим розумом» та «запалює наші чуття і думки» [Гердер : 96]. Застосування лінгвістичних корпусів, зокрема й УНЛК з його сучасним інструментарієм, ви-

водить рівень пізнання картини світобудови крізь призму мови на вищій щабель наукових пошуків, які здійснюються нині у багатовекторній парадигмі [Широков, Лучик : 3–16].

Як відомо, однією з можливостей лінгвістичного корпусу є надання матеріалу для вивчення частотності вживання мовних одиниць як загалом, так і їхнього безпосереднього функціонування в текстах різних стилів. Здійснення кількісних підрахунків тих чи тих мовних фактів за допомогою лінгвістичних корпусів, звичайно, підноситься на інший рівень наукових пошуків, надаючи більшої чіткості та, зрештою, максимальної однозначності для інтерпретації. Саме за допомогою лінгвістичних корпусів із зазначеним інструментарієм певною мірою усуваються ті труднощі, коли, як зазначає В. В. Левицький, «фантазувати та рефлексувати, спираючись на формантні частоти чи частоту основного тону, на закон Вернера або Грімма, на оцінки семантичного диференціалу або, нарешті, на частоту вживання слова або словосполучення, набагато складніше і незручніше, ніж із “легкістю надзвичайною” фантазувати і рефлексувати з приводу концептів — примарних, фантомних категорій, що утворюються уявою дослідника, позбавлених більш або менш чітко окреслених меж» [Левицький : 111–133].

У пропонованому дослідженні скористаємося цим інструментарієм для побудови мікрополя «мови світу» як одного зі складників лексико-семантичного поля «мова», що, з одного боку, сприятиме пізнанню відповідного провідного концепту у когнітивній картині світу, а з другого, — за допомогою пізнання мовних форм, зокрема номінацій мов світу, привідкриє знання про факти об’єктивної дійсності, актуальні для української нації. Ядро такого поля формується найчастотнішими у корпусі одиницями, навколяядерна зона — менш уживаними, периферійна — маловживаними, віддалена периферія — поодинокими одиницями.

Слово *мова* на позначення природної людської мови як опозиції до штучних мов і мов тварин має досить високу частотність в УНЛК. Показник уживань зазначеної лексеми — 241931. Цей факт зумовлений насамперед провідною роллю мови для суспільства, у тому числі сучасного, що впевнено оволодіває новітніми інформаційними технологіями, завдяки чому підтверджуються уявлення про природну мову як найзручнішу і найгнучкішу інформаційну систему, з якою не здатна конкурувати жодна інша, відома людству чи створена ним [Широков : 5].

Однією з унікальних властивостей людської мови є те, що вона, з одного боку, позначає факти об’єктивної дійсності, а з другого, — передає їхню різноманітність, індивідуальність у досить жорсткій формі позначень та відношень різних конкретних виявів. Отже, мова як абстрактне уявлення про загальну людську мову реалізується насамперед у своїх численних конкретних етнічних репрезентантах. Так, лексема *мова* вступає в гіперо-гіпонімічні відношення з одиницями на позначення конкретних, так званих етнічних, мов, яких, за різними оцінками, у світі налічується від 4500 до 6000. Саме наявність опозиційних відношень між лексемою *мова* й одиницями на позначення її конкретних представників дозволяє твердити про можливість побудови лексико-семантичного поля з ядреною одиницею *мова*, що вступає в опозиційні відношення з назвами конкретних мов світу. Останні ж утворюють між собою еквіполентні опозиції з інтегральною семою ‘мова’ і диференційними семами, що відбивають уявлення про конкретний етнос — носія певної мови. Для побудови мікрополя «мови світу» були обрані номінації,

які репрезентують офіційні та найпоширеніші мови світу, а також мови тих народів, які найтісніше взаємодіють з носіями української мови, тобто всі слов'янські та мови національних меншин України. Загальна кількість цих номінацій становить 160 одиниць, назви 96 з них репрезентовані в УНЛК.

Аналіз частотності позначень етнічних мов дозволяє стверджувати, що ядро мікрополя формують номінації, які мають показники уживання в УНЛК понад 1 тис. Таких мов, за даними корпусу, налічується 7. Оскільки корпус сформовано на основі українського дискурсу, цілком логічно, що сюди входить номінація *українська мова* з кількістю слововживань 28162, виявлених у 1866 текстах, які належать до різних стилів. За кількістю слововживань до української найбільше наблизений елемент мікрополя *російська мова*, показник якого — 9328. Представлене це позначення у 1244 текстах. Номінація *англійська мова* має показник 3922, на четвертому місці за кількістю слововживань перебуває одиниця *польська мова* з показником 2896. До ядра мікрополя також входять одиниці: *німецька мова* — показник 2804, *французька мова* — 1434, *грецька мова* — 1085. При цьому, якщо перші шість номінацій референціюють поняття приблизно рівною мірою у текстах усіх функційних стилів, то одиниця *грецька мова* фіксується переважно в науковому стилі.

Зі спостережень над ядром мікрополя «мови світу» випливає, що всі його складники є позначеннями індоєвропейських мов. Це передусім ті з них, що належать до слов'янських — східно- і західнослов'янських мов: *українська мова*, *російська мова* і *польська мова*. Потрапляння до ядра мікрополя «мови світу» позначень слов'янських мов є абсолютно зрозумілим явищем, оскільки, на думку В. В. Лучика, «незважаючи на те, що вже майже півтори тисячі років як перестала існувати праслов'янська етномовна спільність і слов'яни розселилися на різні території, почали користуватися різними мовами, жили здебільшого в різних державах, досі збереглися спорідненість їхніх мов і типові антропологічні дані, близькість культур, вірувань і ментальності, відчуття етнічної солідарності та генетичної спільності» [Лучик 2013 : 173]. Вагоме місце у складі ядра мікрополя «мови світу» займають референти таких мов: германських — *англійська мова*, *німецька мова*, романських — *французька мова*, а також грецьких мов — *грецька мова*. Очевидно, така частотність уживання зазначених елементів в УНЛК пов'язана як з економічно-культурною взаємодією носіїв цих мов, так і з генетичними зв'язками одиниць мовних систем. Підтвердженням останньому слугують спостереження за особливостями українських соматизмів як одного з найдавніших прошарків лексики. Так, за даними «Етимологічного словника української мови» у семи томах, основні номінації людського тіла в українській мові сформувалися у праслов'янську добу, а близько 32 % соматизмів належать до одиниць індоєвропейського походження [Лучик 2012a : 409]. Найчастіше українські соматизми виявляють зв'язки з відповідниками мов, які входять до таких груп: 1) слов'янська, 2) германська, 3) балтійська, 4) італійська, 5) іранська, 6) грецька [Лучик 2012b : 207]. Саме репрезентанти цих мов світу переважно і представлені у ядрі аналізованого мікрополя.

Навколоядерну зону формують позначення, показники яких — від 100 до 1000 слововживань. Таких репрезентантів тут налічується 22 одиниці, вони є референтами мов шести сімей — індоєвропейської, тюркської, сино-тибетської, афро-азійської, фіно-угорської, японсько-рюкюської. Найуживанішими, а відтак максимально наблизеними до ядра мікрополя «мови світу», виступають позначення слов'янських мов — *чеська мова* з показником 945

та *білоруська мова* з показником 563. За кількістю слововживань одиниць навколоядерної зони на третьому місці перебуває референт афро-азійських мов — *мова іврит* з показником 422. Іспанській мові і санскриту належить тут четверта і п'ята сходинки з показниками, відповідно, 363 і 343 слововживань. Найуживанішим референтом іншої сім'ї, власне, тюркських мов, виявляється одиниця *турецька мова*, показник слововживань якої становить 333 у 109 текстах корпусу. Ці слововживання рівною мірою представляють художній, публіцистичний, науковий і офіційно-діловий стилі. Ще 11 одиниць навколоядерної зони є репрезентантами мов індоєвропейської сім'ї: романської, слов'янської, балтійської та германської груп мов. Так, романські представлені словосполученнями *італійська мова* і *румунська мова*, відповідно, з показниками слововживання 306 і 196.

Крім номінацій *чеська мова* і *білоруська мова*, до навколоядерної зони увійшли референти п'яти слов'янських мов: *болгарська мова* з показником слововживань 263, *словацька мова* — 215, *словенська мова* — 209, *македонська мова* — 186, *сербська мова* — 127. Балтійська група представлена у навколоядерній зоні номінацією *литовська мова*, що у корпусі має 118 слововживань, а германська група — мовою *ідиш* — 129 слововживань.

Досить високий показник у корпусі має номінація *угорська мова*, рідна мова угорців та одна з мов національних меншин України, що належить до фіно-угорської сім'ї. Кількість слововживань *угорська мова* дорівнює 259. Меншою частотністю характеризується вживання вербалізатора ще однієї мови цієї сім'ї — *фінська мова*, що має показник 106.

Якщо висока частотність в УНЛК наведених вище позначень мов світу значною мірою пояснюється тісними генетичними зв'язками її одиниць з українською мовою, то активна вживаність позначень інших мов зумовлена насамперед культурно-економічними, демографічними, релігійними тощо чинниками. Так, номінації двох робочих мов Генеральної асамблеї та деяких інших органів Організації Об'єднаних Націй (ООН) теж, за даними УНЛК, виступають складниками навколоядерної зони мікрополя. Референт китайської мови, яка належить до сино-тибетської сім'ї, а її носіями є понад 1,2 млрд осіб, має у корпусі показник 188. Найпоширеніша семітська мова афро-азійської сім'ї, *арабська*, носіями якої є близько 240 млн мовців, обмежено з релігійною метою використовують понад 1,5 млрд мусульман усього світу, вербалізується у корпусі 221 раз. Японо-рюкюська група мов представлена у навколоядерній зоні номінацією *японська мова*, яка, хоч юридично і не має статусу офіційної, при цьому, посідаючи 9 місце за кількістю носіїв, фактично такою є. В УНЛК референт *японська мова* має показник 170.

Компонентами навколоядерної зони мікрополя є референти двох тюркських мов кипчаської групи — кримськотатарської, однієї з мов корінних народів України, і татарської. Отже, хоча ЮНЕСКО зарахувало кримськотатарську до числа мов, що вимирають, за даними УНЛК, її вербалізатори досить активно функціонують, про що свідчить позиція референта *кримськотатарська мова* у навколоядерній зоні мікрополя «мови світу» з показником 141. Ресурси, в яких виявлено цю номінацію, належать передусім до офіційно-ділового стилю, зокрема урядових видань, публіцистичного та наукового стилів. Слововживань на позначення *татарська мова* у корпусних текстах, серед яких переважають ті, що належать до художнього стилю, налічується 106.



До периферійної зони мікрополя «мови світу» зараховано номінації з показниками від 10 до 100 слововживань. Усього таких одиниць нараховується 31, що представляють мови індоєвропейської, тюркської, фіно-угорської, картвельської, корейської, нігеро-конголезької та тупі сімей. Найуживанішими периферійної, як і ядерної та навколоядерної, зони виявляються референти індоєвропейських мов: *шведська мова* з показником 84, *хінді* — 81, *хорватська мова* — 81, *голландська мова* — 72 (інша назва — *нідерландська мова* — 13), *латиська мова* — 59, *португальська мова* — 54, *верхньолужицька мова* — 52, *нижньолужицька мова* — 50, *вірменська мова* — 41, *перська мова* — 36, *циганська мова* — 32 (інша назва — *ромська мова* — 10), *ірландська мова* — 31, *данська мова* — 26, *албанська мова* — 23, *чорногорська мова* — 19, *боснійська мова* — 18, *каталонська мова* — 13, *ісландська мова* — 13.

В УНЛК фіксується три репрезентанти тюркської сім'ї, одна монгольської алтайської макросім'ї, що входять до периферійної зони мікрополя «мови світу»: *казахська мова* з показником 21, *узбецька мова* — 10, *караїмська мова* — 10, а також *монгольська мова* з кількістю слововживань 25. Потрапляння номінації караїмської мови, носіями якої на сьогодні є кількасот людей, до семантичного поля «мови світу» пояснюється тим, що вона є мовою національних меншин України.

Так і мова комі, яка одночасно є однією з державних мов Республіки Комі. З показником слововживань 25 номінація *мова комі* (інша назва — *комі-зирянська мова* — 1), як і номінація *естонська мова* з показником 54, репрезентують в УНЛК представників фіно-угорської сім'ї.

Двома референтами на периферії поля представлена нігеро-конголезька сім'я: *мова суахілі* з показником 34 і *мова йоруба* — 10. Інші мовні сім'ї, за даними корпусу, мають по одній номінації, що виступає складником периферії побудованого нами мікрополя. П'ять мовних сімей мають по одному репрезентанту на периферії поля. Картвельську сім'ю представляє номінація *грузинська мова* з показником 72. Лінгвістичну сім'ю тупі, що містить близько 70 аборигенських діалектів Південної Америки, репрезентує номінація *мова гуарані* з 52 слововживаннями в УНЛК. Корейську мовну сім'ю, сформовану власне корейською мовою, яку іноді вважають і мовою острова Чеджу, вербалізує у корпусі номінація *корейська мова*, уживана тут 36 разів у публіцистичних і наукових текстах. Номінація *арамейська мова* з показником 28 представляє на периферії поля афро-азійську сім'ю. Референт найпоширенішої автохтонної мови Південної Америки — *мова кечуа*, що належить до сім'ї кечуа, фіксується у 9 текстах корпусу 14 разів.

Найрозгалуженішою є віддалена периферія мікрополя «мови світу». Значною мірою тут представлені номінації мов, носії яких мають спільне з українським народом радянське минуле. Вони репрезентують національні меншини на теренах України, і є значними за кількісним складом націями світу. Загалом ця зона формується 37 номінаціями мов, що належать до 9 сімей. Показники уживання номінацій зазначеного фрагмента мікрополя становлять від 1 до 9. Як у ядерній, навколоядерній і периферійній зонах, так і тут домінують вербалізатори індоєвропейських мов, слововживань яких налічується у корпусі від одного до восьми: *таджицька мова* (8), *урду мова* (7), *африкаанс мова* (6), *бенгальська мова* (4), *пенджабі мова* (3), *курдська мова* (3), *орія мова* (2), *люксембурзька мова* (2), *маратхі мова* (1), *гуджараті мова* (1), *пушту мова* (1), *сингальська мова* (1).

Представлені тут номінації мов тюркської сім'ї з показниками уживаності від 1 до 8: *кримчацька мова* (8), *киргизька мова* (7), *чуваська мова* (7), *туркменська мова* (7), *азербайджанська мова* (5), *гагаузька мова* (3), *башкирська мова* (1). Мови фіно-угорської сім'ї референціюються трьома номінаціями: *мордовська-ерзя мова* (4), *удмуртська мова* (3), *марійська мова* (1). Зафіксована лише одна номінація мов афро-азійської сім'ї з показником 1: *оромо мова*.

Тільки віддалена периферія мікрополя «мови світу» формується за участю номінацій мов австронезійської, дравідійської, північнокавказької, австро-азійської і тай-кадської сімей. Представниками австронезійської сім'ї є номінації 5 мов, уживання яких має показники від одного до трьох: *індонезійська мова* (3), *яванська мова* (2), *малайська мова* (2), *філіппінська мова* (2), *сунданська мова* (1). Номінації референтів дравідійської сім'ї поповнюють віддалену периферію чотирма одиницями, серед яких тільки *тамільська мова* має показник уживання 2, а решта — по одному: *телугу мова*, *малайлам мова*, *каннада мова*.

Північнокавказькі мови, представлені в УНЛК, мають дещо вищі показники уживання. Так, номінація *чеченська мова* має показник 4, а *аварська мова* — 3. Як північнокавказька, так і австро-азійська сім'я репрезентовані у корпусі номінаціями двох мов: *в'єтнамська мова* з показником 3 і *кхмерська мова* з показником 2. Номінація *лаоська мова*, що є референтом тай-кадської сім'ї, фіксується в корпусі лише один раз.

Як зазначає О. Б. Ткаченко, лише з середини ХХ ст. відбувається зростання «економічної і політичної ваги країн Азії та Африки, природно, піднесення впливовості їхньої культури і поширених там мов» [Ткаченко : 361].

Можливо, саме таким пізнім періодом міжмовної взаємодії і пояснюється те, що деякі офіційні мови світу ще не мають жодного вживання в УНЛК, наприклад: *дорі мова*, *біслама мова*, *тигринья мова*, *амхарська мова*, *кірибати мова*, *сесото мова*, *маврикійська мова*, *бходжупурі мова*, *малагасійська мова*, *мальдівська мова*, *маршальська мова*, *непальська мова* та ін.

Проте назви мовних сімей, референти яких наявні у мікрополі «мови світу», переважно представлені у корпусі: *індоєвропейські мови* з показником 315, *тюркські мови* — 200, *фіно-угорські мови* — 13, *дравідійські мови* — 3, *картвельські мови* — 1, *австро-азійські мови* — 1, *тупі мови* — 1, *тай-кадські мови* — 1.

Отже, мікрополе «мови світу» лексико-семантичного поля «мова», утвореного на основі матеріалів УНЛК, упорядковується 96 позначеннями офіційних, а також мов народів, які найтісніше взаємодіють із носіями української мови. Застосування методу кількісних підрахунків топологічно увиразнює ядрну, навколяядерну, напівпериферійну та периферійну зони мікрополя «мови світу». Розподіл мовних одиниць висвітлює факти об'єктивної дійсності, найактуальніші і менш актуальні для носіїв української мови. Так, ядро поля «мови світу», сформоване сімома найуживанішими номінаціями — *українська мова*, *російська мова*, *англійська мова*, *польська мова*, *німецька мова*, *французька мова*, *грецька мова*, — засвідчує актуальність цих позначень для мовної картини світу українців ХХ — початку ХХІ ст.

З аналізу навколяядерної зони, утвореної 21 одиницею, випливає, що до активних для українців номінацій належать як позначення індоєвропейських, що вже встановлюється зі спостережень над ядром мікрополя, так і тюрк-

ських, афро-азійських, фіно-угорських, сино-тибетських, японо-рюкюських мов. Фрагменти мовної картини світу українців розширюються за рахунок уявлень про мови тупі, кечуа, нігеро-конголезькі, монгольські, корейську, австро-азійські, австронезійські, північнокавказькі, дравідійські, тай-кадські, номінації яких формують периферію і віддалену периферію аналізованого мікрополя.

Кількісні підрахунки номінацій мов світу в УНЛК засвідчують, що на їхню активність впливають генетичні витоки, близькість культур і ментальності, територіальне контактування народів, особливості їхньої міжмовної взаємодії, економічна і політична вага країн, мешканці яких є носіями тієї чи іншої мови. Побудоване мікрополе «мови світу» розширює уявлення про концептуальне поле «мова», необхідність структурування якого вже визріла у лінгвістичних колах. Виявлені активні номінації мов світу, що ввійшли до ядерної і навколоядерної зон, здатні зорієнтувати фахівців на ті не охоплені увагою сфери перекладацької лексикографічної діяльності, які вже стали актуальними для носіїв української мови. Одиниці напівпериферійної і периферійної зон мікрополя «мови світу» з низькими показниками слововживання увиразнюють шляхи розширення ментальних змін носіїв української мови. Пізнання досі не обстежуваних мов забезпечуватиме розвиток вітчизняної перекладної лексикографії, коли вихідною точкою її науково-дослідного досягнення буде українська мова у зіставленні з мовою, номінація якої поки що входить до напівпериферійної або периферійної зони мікрополя «мови світу».

## ЛІТЕРАТУРА

- Гердер И. Г. Идеи к философии истории человечества. Москва : Наука, 1977. 703 с.
- Левицький В. В. Опыт использования квантитативных методов на кафедре германского, общего и сравнительного языкознания Черновицкого университета. *Проблеми квантитативної лінгвістики* : зб. наук. пр. Чернівці : Рута, 2005. С. 111–133.
- Лінгвістично-інформаційні студії: праці Українського мовно-інформаційного фонду НАН України : у 5 т. / В. А. Широков та ін. Т. 4 : Корпусна та когнітивна лінгвістика. Київ : Український мовно-інформаційний фонд НАН України, 2018. 246 с.
- Лучик А. А. Вербальна модель зовнішнього портрета слов'янина (на матеріалі «Етимологічного словника української мови»). *Акцентологія. Етимологія. Семантика: До 75-річчя академіка НАН України В. Г. Скляренка*. Київ : Наук. думка, 2012а. С. 407–417.
- Лучик А. А. Генетичні зв'язки українських соматизмів з їхніми відповідниками у неслов'янських мовах (на матеріалі «Етимологічного словника української мови»). *Сучасні дослідження з іноземної мови* : зб. наук. пр. Ужгород : ФОП Бреза, 2012б. Вип. 10. С. 207.
- Лучик А. А., Остапова І. В. Синтагматична параметризація еквівалентів слів у парадигмі корпусної лінгвістики. *Людина. Комп'ютер. Комунікація* : зб. наук. пр. Львів : Вид-во Львів. політехніки, 2017. С. 33–37.
- Лучик В. В. Вступ до слов'янської філології. Київ : Академія, 2013. С. 173.
- Ткаченко О. Б. Міжмовні контакти української мови. *Українська мова* : енциклопедія / редкол.: Русанівський В. М., Тараненко О. О. (співголови) та ін. Київ : Вид-во «Укр. енциклопедія» ім. М. П. Бажана, 2007. С. 361.
- Широков В. А., Лучик А. А. Парадигмальні засади лінгвістики першої половини ХХІ століття. *Мовознавство*. 2021. №5. С. 3–16.



Широков В. А. Информационная теория лексикографических систем. Київ : Довіра, 1998. С. 5.

Egbert J., Baker P. Introduction. *Triangulating Methodological Approaches in Corpus-Linguistic Research*. New York ; London, 2016. P. 1–19.

## REFERENCES

- Gerder I. G. (1977). *Idey k filosofii istorii chelovechestva*. Moscow: Nauka, 96. [In Russian].
- Levickij V. V. (2005). *Opyt ispol'zovaniya kvantitativnykh metodov na kafedre germanskogo, obshchego i sravnitel'nogo yazykoznaniiya Chernovickogo universiteta. Problemy kvantitativnoi linhvistyky: Zbirnyk naukovykh prats*. Chernivtsi: Ruta, 111–133. [In Russian].
- Linhvistychno-informatsiini studii: pratsi Ukrainskoho movno-informatsiinoho fondu NAN Ukrainy : in 5 vols. (2018). V. A. Shyrovok etc. (Eds.). Vol. 4: *Korpusna ta kohnityvna linhvistyka*. Kyiv: Ukrainskyi movno-informatsiinyi fond NAN Ukrainy, 17–18. [In Ukrainian].
- Luchyk A. A. (2012a). Verbalna model zovnishnioho portreta slov'iany (na materialy «Etymolohichnoho slovnyka ukrainskoi movy»). *Aktsentolohiia. Etymolohiia. Semantyka: Do 75-richchia akademika NAN Ukrainy V. H. Skliarenka*. Kyiv: Naukova dumka, 407–417. [In Ukrainian].
- Luchyk A. A. (2012b). Henetychni zv'iazky ukrainskykh somatyzmiv z yikhnyimi vidpovidnykamy u neslov'ianskykh movakh (na materialy «Etymolohichnoho slovnyka ukrainskoi movy»). *Suchasni doslidzhennia z inozemnoi movy*. Zbirnyk naukovykh prats. Uzhhorod: FOP Breza. Iss. 10, 207. [In Ukrainian].
- Luchyk A. A., Ostapova I. V. (2017). Syntahmatychna parametryzatsiia ekvivalentiv slov u paradyhmi korpusnoi linhvistyky. *Liudyna. Komp'uter. Komunikatsiia: Zbirnyk naukovykh prats*. Lviv: Vydavnytstvo Lvivskoi politekhniki, 33–37. [In Ukrainian].
- Luchyk V. V. (2013). *Vstup do slov'ianskoi filolohii*. Kyiv: Akademiia, 173. [In Ukrainian].
- Tkachenko O. B. (2007). Mizhmovni kontakty ukrainskoi movy. *Ukrainska mova. Entsyklopediia*. V. M. Rusanivskiy, O. O. Taranenko, M. P. Ziabliuk etc. (Eds.). Kyiv: Vydavnytstvo «Ukrainska entsyklopediia» im. M. P. Bazhana, 361. [In Ukrainian].
- Shyrovok V. A., Luchyk A. A. (2021). Paradyhmalni zasady linhvistyky pershoi polovyny XXI stolittia. *Movoznavstvo*, (5), 3–16. [In Ukrainian].
- Shyrovok V. A. (1998). *Informatsiina teoriia leksykoorafichnykh system*. Kyiv: Dovira, 5. [In Ukrainian].
- Egbert J., Baker P. (2016). Introduction. *Triangulating Methodological Approaches in Corpus-Linguistic Research*. New York ; London, 1–19.

### A. A. LUCHYK

National University «Kyiv-Mohyla Academy», Kyiv, Ukraine

E-mail: allal@meta.ua

<https://orcid.org/0000-0002-4218-355X>

### O. S. PALCHEVSKA

Lviv State University of Life Safety, Lviv, Ukraine

E-mail: palch56@ukr.net

<https://orcid.org/0000-0002-2090-4870>

## LANGUAGE DIVERSITY IN THE UKRAINIAN NATIONAL LINGUISTIC CORPUS

The article describes the features of the microfield «Languages of the World» based on the Ukrainian National Linguistic Corpus (hereinafter referred to as the UNLC) text materials with a volume of about 20 million word usages. The representativeness, the corpus texts' genres and functional styles balance as well as the full-text search toolkit usage ability, made it possible to obtain the most reliable data on the nominations denoting world languages functioning in the Ukrainian language. With the help of the world languages

referents quantitative calculations, the core, nearcore, semi-peripheral and peripheral zones of the “Languages of the World” microfield were formed in the UNLC. The core of the microfield is formed by the most frequent units in the corpus, the nearcore zone – by less frequent units, the peripheral zone – by the units that are not frequently used, and the remote periphery – by the units that are rarely used. It was found that all the “Languages of the World” microfield core components are referents of the Indo-European languages. The nearcore zone is formed by nominations designating the languages of the Indo-European, Turkic, Sino-Tibetan, Afrasian, Finno-Hungarian and Japanese-Ryukyu families. The peripheral zone includes units representing the languages of the Indo-European, Turkic, Finno-Hungarian, Kartvelian, Korean, Afro-Asiatic, Niger-Congolese and Tupi families. Quantitative calculations of UNLC nominations, which were included in the «Languages of the World» microfield under study or analysis, highlighted the influence of extralingual factors on the features of language information modeling, including: the world languages genetic origins, cultures proximity, mentality, beliefs, peoples territorial contact, their interlinguistic interaction features, the economic and political weight of countries whose inhabitants are speakers of one or another language. Therefore, the microfield opens the way for new observations on the language through the prism of the language itself and actualizes its properties in relation to the above issue at the current stage of functioning.

Keywords: world languages, Ukrainian language, lexical-semantic field, core zone, nearcore zone, semi-peripheral zone, peripheral zone, Ukrainian National Linguistic Corpus.

Дата надходження до редакції — 21.09.2022

Дата затвердження редакцією — 03.10.2022