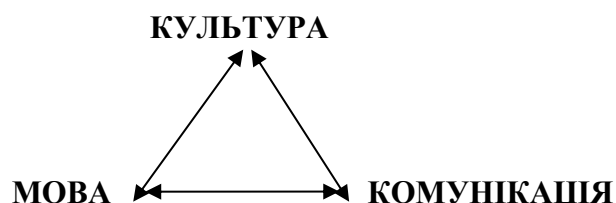


**КОНЦЕПЦІЯ**  
**ВСЕУКРАЇНСЬКОГО ЛІНГВІСТИЧНОГО ДІАЛОГУ**

# ОСНОВНІ ПРИНЦИПИ І АСПЕКТИ ФОРМУВАННЯ ВСЕУКРАЇНСЬКОГО ЛІНГВІСТИЧНОГО ДІАЛОГУ

## Наукові засади Всеукраїнського лінгвістичного діалогу

Мова є необхідним компонентом базового трикутника будь-якої світової культури:



Досі в історії невідомі прецеденти, коли б мова руйнувалася, а пов'язана з нею цивілізація залишалася.

Проте слід зазначити, що при переході до мережецентричної фази еволюції цивілізаційна роль природної людської мови докорінним чином змінюється. Адже Всесвітня Мережа вже перетворилася на середовище, де розігрується когнітивно-комунікативний сценарій розвитку цивілізації.

Економіка, заснована на знаннях, і так звана «цифрова економіка» вимагають колосальних обсягів нової науково-технічної, бізнесової, оперативної та іншої інформації, виробництво якої також включено в загальний когнітивно-комунікативний процес і яка у своїй найважливішій, концептуальній частині представлена саме структурами природної мови. За даними відомої фірми IDC, ще в 2011 році загальний світовий обсяг створених і реплікованих людством даних склав понад 1,8 зеттабайт (18 трильйонів Гб). За прогнозами цієї ж фірми, обсяг даних на планеті буде, як мінімум, подвоюватися кожні два роки до 2020 року, а надалі цей темп ще більше зростатиме.

Якщо зазначений тренд вважати справедливим – а на сьогодні немає серйозних підстав сумніватися в ньому, то вже зараз обсяг світових, адаптованих Мережею даних перевищив величину  $10^{14}$ , а можливо – і  $10^{15}$  Гб.

Однак виникають цілком обґрунтовані сумніви, що такі великі обсяги інформації можуть бути ефективно оброблені, а головне – адекватно сприйняті й усвідомлені їхніми реципієнтами-адресатами в реальному часі сучасного буття. Експоненційне зростання інформації не супроводжується адекватним зростанням людських можливостей щодо її засвоєння, а тим більш ефективного використання.

Виникає головне протиріччя сучасної епохи: закон еволюції вимагає зростання обсягів інформації, що виробляється і сприймається людством, а суспільство як сума індивідів і відносин між ними внаслідок особливостей людської природи, не в змозі належним чином скористатися цими обсягами і

змушене обмежувати їх виробництво, обробку і використання. Еволюція нібито сама «вмикає» механізми свого гальмування, які іноді виявляються в досить драматичних формах. Усе більшої популярності набувають катастрофічні сценарії розвитку цивілізації, гіпотези технологічної сингулярності та ін.

Зазначені особливості сучасного етапу розвитку Мережі поставили проблему створення інструментів, заснованих на принципах штучного інтелекту, здатних перебрати на себе значну частину функцій основного когнітивного тракту людини. Таким чином проблема інтелектуалізації Мережі впевнено пересунулася на передній план актуальності. **А серед багатьох аспектів інтелектуалізації особливо виділяється мова.**

Дійсно, на сьогодні вже цілком очевидно, що лінгвістичне забезпечення Мережі відіграє роль провідного чинника і основного інтерфейсу, що забезпечує взаємодію Людини з Мережею і Людини з Людиною через Мережу. Набуває все більшої актуальності проблема інтеграції ідей і технологій віртуалізації, великих даних та інтелектуалізації, головним чином, через механізми природної мови. **Проте особливість лінгвістичних засобів полягає в тому, що мови існують лише у формі окремих національних мов, носії яких, маючи фахову підготовку в галузі лінгвістичних технологій, власне, і здатні розробляти мовно-інформаційні компоненти програмного забезпечення на належному рівні та прийнятної якості.** Основою такого роду розробок є добре кодифіковані, анотовані та репрезентативні моделі і масиви лінгвістичних даних, що представляють усі (в ідеалі) аспекти функціонування тієї чи іншої мови – як у когнітивному, так і в комунікативному плані. Такими засобами служать лексикографічні середовища та лексикографічні агенти, а також лінгвістичні корпуси, які оперують великими цифровими масивами лінгвістично кваліфікованих і анотованих текстів.

**Отже, у мережецентричному світі природна людська мова стає не тільки (і навіть не стільки) гуманітарним, скільки (і насамперед) технологічним чинником, тому технологічний статус мови значною мірою визначатиме характер розвитку цифрової мережецентричної цивілізації у найближчий період, а разом із цим – і ієрархію мов, яка демонструватиме спроможність тієї чи іншої національної мови до успішної конкуренції серед усього кола світових мов.**

Слід наголосити на високій наукоємності зазначених інструментів, за допомогою яких тільки й можливе проведення сучасних лінгвістичних досліджень та розроблення високоефективних мовно-інформаційних застосувань (додатків). Отже, кожне національне мовознавство опиняється перед своєрідним викликом: або створити необхідну гаму інтелектуальних артефактів, орієнтованих на відповідну мову, або приректи її на деградацію у цифровому мережевому світі і як неминучий наслідок – на занепад нації.

З викладеного однозначно випливає, що без створення сучасної системи національних лінгвістичних ресурсів та національної лінгвістичної інфраструктури ефективний розвиток мережецентричного суспільства та економіки знань в Україні неможливий. Сформульовані в такий спосіб задачі зараз перетворилися на серйозні виклики національній та інформаційній безпеці держави.

**Таким чином, створення інтелектуалізованої, віртуалізованої, всеохопної, універсальної за функціоналом та сформованої на засадах Big Data системи національних лінгвістичних ресурсів та національної лінгвістичної інфраструктури є головною проблемою, на розв'язання якої спрямовується Концепція.**

### **Аналіз причин виникнення проблеми та обґрунтування необхідності її розв'язання**

Проблема забезпечення всебічного розвитку і функціонування української мови в усіх сферах суспільного життя, створення єдиного культурного простору України та збереження цілісності культури **виникла історично**, проте її розв'язанню досі не приділялося належної уваги.

Слід наголосити, що зміст і значення зазначеної проблеми для розвитку української держави останнім часом зазнали істотних змін у зв'язку із глобальними цивілізаційними зрушеннями й переходом світу до мережецентричної фази розвитку. Як результат, розв'язання даної проблеми набуло надзвичайно високої наукоємності й потребує запровадження спеціальних методів, засобів та організаційних форм.

Необхідно зазначити, що дослідження й розробки в зазначених напрямках в Україні ведуться, хоча й розрізнено, фрагментарно і без належної координації.

Створені в деяких осередках інформаційно-комп'ютерні лінгвістичні засоби (лінгвістичні корпуси, віртуальні системи професійної взаємодії, засоби концептуалізації та аналізу лінгвістичних даних тощо) не стали надбанням усього загалу користувачів й застосовуються лише локально. Сказане значно знижує не лише ефективність досліджень і розробок у цій у галузі, але й, що найголовніше – масштаби та якість застосування інтелектуальних мовно-інформаційних артефактів у державній, суспільній та приватній практиці.

Потребує нових і цілеспрямованих зусиль проблема підготовки сучасних кадрів найвищої кваліфікації з різних ділянок інформаційно-лінгвістичної технології, спроможних ставити й розв'язувати найскладніші й найактуальніші задачі у галузі сучасних засобів інтелектуального опрацювання українсько-мовної інформації на світовому рівні.

Зовсім недостатніми також є рівень та масштаби поширення українськомовного культурного продукту, створення висококультурного національно-мовного середовища.

Залишаються нерозвиненими механізми інтенсивного впровадження української мови в усі ділянки державно-громадської діяльності, засоби контролю за виконанням мовного законодавства та інших нормативних актів.

Значну проблему становить і недостатня розвиненість української термінологічної та термінографічної бази, що знижує можливості інтегрування до світової інформаційно-інноваційної системи.

Слід зауважити, що розв'язати ці проблеми традиційними методами неможливо внаслідок їхньої високої трансдисциплінарності.

Отже, на перший план зараз висувається проблема консолідації зусиль лінгвістів, культурологів, інформатиків, фахівців у галузі штучного інтелекту, спеціалістів з різних предметних галузей на ниві створення сучасної системи мовно-інформаційних ресурсів, спроможної забезпечити необхідну інформаційно-технологічну підтримку на шляху України до мережецентричного суспільства знань.

### **Конкретизація мети Концепції**

Конкретною метою Концепція визначає створення універсальної системи національних лінгвістичних ресурсів та національної лінгвістичної інфраструктури, орієнтованої на забезпечення всебічного розвитку і функціонування української мови в усіх сферах суспільного життя, єдиного культурного простору України, збереження цілісності національної культури та забезпечення конкурентоспроможності української мови у світовому інформаційно-лінгвістичному середовищі.

Оптимальний варіант досягнення поставленої мети мусить базуватися на цільових засадах, що вимагає переосмислення засадничих настанов у цій галузі, а саме, орієнтацію на створення двокомпонентної системи:

- а) національних лінгвістичних ресурсів;
- б) національної лінгвістичної інфраструктури,

що на державному рівні визначається як найважливіше науково-технологічне завдання, яке має безпосередній вплив на проблематику національної та інформаційної безпеки держави.

При цьому:

- 1) до системи національних лінгвістичних ресурсів включаємо:

– **національну словникову базу України**, а саме:

традиційні словники та цифрові системи, які подають, граматичний, семантичний, прагматичний, статистичний, лінгвогеографічний і т. ін. опис української мови та інших основних мов світу (насамперед тих, що мають міжнародний статус), Державний тезаурус України; відповідні цифрові

лінгвістичні корпуси, включаючи Державний інформаційно-лінгвістичний корпус нормативно-правової бази України; цифрову систему національної термінології та термінографії, національну цифрову енциклопедично-довідкову базу;

– **засоби цифрової гуманітаристики:** електронні системи розпізнавання та синтезу природних мов (фонетична та письмова форми), системи фонетично-графічних трансформацій; інтелектуальні людиномашинні інтерфейси, засновані на принципах природної мови; інтелектуальні цифрові системи семантичної інтерпретації та концептографічного аналізу, а також засоби лінгвістичної експертизи природномовних текстів; засоби і технології конвертації (парсингу) паперової лінгвістичної та літературної спадщини до цифрової форми, засновані на корпусних технологіях; системи, засоби і технології електронного навчання української мови та літератури; засоби і технології української жестової мови;

– **методи та цифрові лінгвістичні системи спеціального призначення;**

2) до системи національної лінгвістичної інфраструктури включаємо:

– **національну лінгвістичну мережу** як комплекс спеціалізованих програмно-апаратних засобів комунікації між локальними лінгвістичними осередками – суб'єктами системи національних лінгвістичних ресурсів;

– **головний репозиторій системи національних лінгвістичних ресурсів;**

– **віртуальні системи професійної взаємодії у лінгвістиці, літературі та культурі** (віртуальні лінгвістичні лабораторії, літературні й загально-культурні енциклопедичні системи та системи спеціального призначення);

– комплекс нормативно-програмних засобів для забезпечення функціонування та мережевої взаємодії учасників «Всеукраїнського лінгвістичного діалогу».

### **Шляхи та способи реалізації Концепції**

Зважаючи на обмежені фінансові, матеріально-технічні та кадрові можливості України та виходячи з необхідності прискореного розвитку системи національних лінгвістичних ресурсів та національної лінгвістичної інфраструктури шляхом консолідації наявного науково-технологічного потенціалу, **необхідно:**

- заснувати систему соціально-професійної взаємодії у мовній галузі – **«Всеукраїнський лінгвістичний діалог»** як розподілену мережеву систему **суб'єктів та об'єктів національного мовно-інформаційного процесу** із включенням до нього сучасних інтелектуалізованих, віртуалізованих, універсальних за функціоналом та сформованих на засадах Big Data інструментів лінгвістичного дослідження, лінгвістичних ресурсів та лінгвістичної інфраструктури, яка дозволить шляхом виділення відносно невеликих фінансових коштів забезпечити піднесення на якісно новий рівень

української лінгвістичної технології та прикладної лінгвістики, теорії і практики штучного інтелекту, їхньої практичної результативності та ефективної інтеграції до загальноєвропейського і світового інформаційно-лінгвістичного та модерного культурного простору;

- створити загальнонаціональний репозиторій національних інформаційно-лінгвістичних ресурсів «Всеукраїнського лінгвістичного діалогу» як ядра і центрального елемента національної лінгвістичної інфраструктури України. Зважаючи на масштаби та складність цієї системи, покласти функцію головної установи з проектування й забезпечення її функціонування та доступу до неї найширшого кола користувачів на НАН України.

З метою координації відповідних робіт МОНУ та НАНУ створюється Міжвідомча координаційна рада «Всеукраїнського лінгвістичного діалогу».

Міжвідомча Рада «Всеукраїнський лінгвістичний діалог» визначає порядок формування, використання та розвитку системи національних лінгвістичних ресурсів та національної лінгвістичної інфраструктури України, здійснює планування та контроль виконання заходів Державної цільової програми.

З метою підготовки кадрів найвищої кваліфікації з прикладної та математичної лінгвістики (докторів філософії та докторів наук з фізико-математичних, технічних та філологічних напрямів) необхідно:

- відкрити докторантуру та аспірантуру з відповідної спеціальності у провідних університетах та наукових установах України;
- створити мережу спільних (МОНУ та НАНУ) осередків «Всеукраїнського лінгвістичного діалогу» в університетах України.

### **Огляд першочергових заходів «Всеукраїнського лінгвістичного діалогу»**

Виходячи з викладеного пропонується скомпонувати комплекс заходів з двох розділів:

#### **I. Громадські, загальнокультурні, наукові та організаційні заходи.**

#### **II. Всеукраїнські проекти.**

До першого розділу доцільно ввести заходи, пов'язані із міжнародними та всеукраїнськими акціями на підтримку рідної мови, такими як Міжнародний день рідної мови, День слов'янської писемності і культури, Свято українського словника, День української писемності та мови тощо. Доцільно розробити орієнтовні сценарії проведення таких заходів. Проводити їх слід на різних рівнях – загальнодержавному, регіональному (обласному), районному, на рівні окремих установ (наприклад, університетів), територіальних громад тощо. Безпосередню організацію проведення цих заходів доцільно покласти на осередки Всеукраїнського лінгвістичного діалогу різних рівнів.

Також на осередки Всеукраїнського лінгвістичного діалогу спільно з освітніми, науковими, державними та муніципальними інституціями слід покласти проведення наукові та науково-організаційні заходи (семінари, конференції, круглі столи тощо). Слід вирішити питання про проведення в м. Києві Міжнародного з'їзду славістів та конгресів Міжнародної асоціації українців, що значно підвищить авторитет української мови і України взагалі на міжнародному рівні.

Другий розділ заходів присвячений виконанню Всеукраїнських проектів, які охоплюють весь обшир мовних проблем як на цивілізаційному, так і на національному рівнях. До числа Всеукраїнських проектів включено такі:

- 1. УКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ ЛІНГВІСТИЧНИЙ КОРПУС;**
- 2. НАЦІОНАЛЬНА СЛОВНИКОВА БАЗА;**
- 3. СИСТЕМА НАЦІОНАЛЬНОЇ ТЕРМІНОЛОГІЇ І ТЕРМІНОГРАФІЇ;**
- 4. ПАРСИНГ ЛЕКСИКОГРАФІЧНИХ СИСТЕМ;**
- 5. ФОРМУВАННЯ НАЦІОНАЛЬНОЇ ЛІНГВІСТИЧНОЇ ІНФРАСТРУКТУРИ**
- 6. ФОРМУВАННЯ СИСТЕМИ НАЦІОНАЛЬНИХ ЛІНГВІСТИЧНИХ РЕСУРСІВ УКРАЇНИ;**
- 7. ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ ОПРАЦЮВАННЯ ТЕКСТІВ. ЗАСТОСУВАННЯ КОРПУСНИХ ТЕХНОЛОГІЙ І КОНЦЕПТОГРАФІЧНИХ СИСТЕМ ТИПУ ULISS, UNILEX, POLYHEDRON В ОСВІТІ ТА ІНЖЕНЕРІЇ ЗНАТЬ. ІНТЕЛЕКТУАЛЬНІ ТРАНСДИСЦИПЛІНАРНІ ІННОВАЦІЙНО-ІНВЕСТИЦІЙНІ СИСТЕМИ.**
- 8. ЕЛЕКТРОННЕ НАВЧАННЯ;**
- 9. ЛІНГВІСТИЧНЕ ЗАБЕЗПЕЧЕННЯ МІЖНАРОДНОЇ ДІЯЛЬНОСТІ, ПОСОЛЬСТВ ТА КОНСУЛЬСТВ УКРАЇНИ ЗА КОРДОНОМ;**
- 10. ІНТЕЛЕКТУАЛЬНІ МОВНО-ІНФОРМАЦІЙНІ ІНТЕРФЕЙСИ;**
- 11. ЕЛЕКТРОННИЙ ПІДРУЧНИК;**
- 12. ВСЕУКРАЇНСЬКИЙ ПОРТАЛ «УКРАЇНСЬКА МОВА». СТВОРЕННЯ ЛІНГВІСТИЧНИХ ІГОР ТА ТРЕНАЖЕРІВ. ЗАПОЧАТКУВАННЯ СОЦІАЛЬНО-ПРОФЕСІЙНОЇ МЕРЕЖІ «ВСЕУКРАЇНСЬКИЙ ЛІНГВІСТИЧНИЙ ДІАЛОГ»;**
- 13. УКРАЇНСЬКИЙ ПРАВОПИС. ДОВІДКОВА СЛУЖБА З УКРАЇНСЬКОЇ МОВИ;**
- 14. ЛОГІКО-ЛІНГВІСТИЧНА ЕКСПЕРТИЗА ТА ЮРИДИЧНА ЛІНГВІСТИКА;**
- 15. ДЕРЖАВНИЙ ТЕЗАУРУС УКРАЇНИ;**
- 16. ДЕРЖАВНИЙ ЛІНГВІСТИЧНИЙ КОРПУС НОРМАТИВНО-ПРАВОВОЇ БАЗИ УКРАЇНИ;**
- 17. ЛІНГВІСТИЧНІ ТЕХНОЛОГІЇ СПЕЦІАЛЬНОГО ПРИЗНАЧЕННЯ.**



## Очікувані результати

З урахуванням викладеного передбачено досягнення таких результатів:

1. Мова державної системи та законодавства. Лінгвістична експертиза нормативно-правової бази України. Державний тезаурус України. Державний лінгвістичний корпус нормативно-правової бази України.

2. Наука та освіта. Наукові дослідження та розробки у галузі сучасних інтелектуальних лінгвістичних технологій та систем. Інтелектуальні трансдисциплінарні інноваційно-інвестиційні системи.

3. Онлайнкові загальнодоступні українськомовні лінгвістичні ресурси, освітні курси та системи. Електронне навчання української мови.

4. Всеукраїнська онлайндова довідкова служба з української мови та система державної сертифікації рівня володіння українською мовою.

5. Заходи, засоби та процеси інтеграції України до світового мовно-інформаційного простору.

6. Спеціальні аспекти застосування мов.

7. Системні відношення та фундаментальні аспекти еволюції в тріаді «Інформація–Мова–Інтелект».

8. Мовні аспекти українського мистецтва та модерна лінгвокультуро-логія;

9. Мова мас-медіа та соціальних комунікацій. Гармонізація мовного простору України.

10. Громадські акції, пов'язані із популяризацією української мови.

У результаті завдяки інтенсивному розвитку системи Всеукраїнського лінгвістичного діалогу буде забезпечено високий статус української мови як державної на всій території України та у всіх сферах її функціонування.

Особливого розвитку набуде сучасна цифрова мережева форма функціонування та застосування української мови.

Буде створено і впроваджено в доступ користувачам всю гаму національних лінгвістичних ресурсів (цифрові словники та інструментальні системи опрацювання українських текстів, система національної термінології та термінографії, засоби концептографічного аналізу та опрацювання знань, інтелектуальні трансдисциплінарні інноваційно-інвестиційні системи, національний лінгвістичний корпус обсягом не менше 5 мільярдів слововживань, цифрові бібліотеки української мови та літератури, збудовані за корпусними технологіями).

Буде розроблено і впроваджено електронні локальні та онлайнкові підручники нового покоління та інші електронні навчальні засоби української мови для користувачів різного рівня, починаючи з дошкільного і закінчуючи курсами для професіоналів та іноземців.

Буде створено:

- систему сертифікації рівня володіння українською мовою для державних службовців та осіб, які подають клопотання про прийняття до громадянства України;
- систему моніторингу функціонування державної мови, підтримки та розвитку національного правописного процесу;
- мовно-інформаційний інструментарій для перекладачів з та *на* українську мову (для основних світових мов).

Буде розроблено й впроваджено систему віртуальних лабораторій у галузі мовознавства, літературознавства, культури, музейної справи та логіко-лінгвістичної експертизи текстів.

Усі зазначені засоби та системи буде інкорпоровано до національної лінгвістичної інфраструктури України у вигляді **Українського лінгвокультурного порталу**, який являтиме собою вільно доступну систему модерних лінгвістично-культурних ресурсів, що перебуває у постійному розвитку та вдосконаленні.

### **Заключні положення**

Зважаючи на стрімку, майже непередбачувану еволюцію сучасного світу та необхідність на сучасному цивілізаційному зламі докорінного зміцнення основ державно-національного устрою України, її національної та інформаційної безпеки, «Всеукраїнський лінгвістичний діалог», який являтиме собою загальнодержавну систему, що забезпечує створення та впровадження в державну й суспільну практику комплексу національних лінгвістичних ресурсів та національної лінгвістичної інфраструктури, стане потужним чинником суспільної консолідації, утвердження гуманітарного і технологічного статусу української мови на рівні модерних викликів.