

**Широков В.А.**

академік Національної академії наук України, доктор технічних наук, професор
директор Українського мовно-інформаційного фонду НАНУ, Київ, Україна,
vshirokov48@gmail.com

МОВА В МЕРЕЖЕЦЕНТРИЧНОМУ СВІТІ

У роботі аналізуються процеси зміни цивілізаційної ролі природної людської мови при переході до мережецентричної фази еволюції. Відзначається, що зростання інформації, що функціонує в Мережі, вже вийшло на експонентний режим і її обсяги вже не можуть бути ефективно оброблені, сприйняті і усвідомлені їхніми реципієнтами-адресатами. Таким чином, проблеми створення комп'ютерних інструментів, заснованих на принципах штучного інтелекту (проблеми інтелектуалізації Мережі), впевнено пересунулися на передній план актуальності, а серед багатьох аспектів інтелектуалізації особливо виділяється мова. Технологічний статус мови визначатиме характер розвитку цифрової мережецентричної цивілізації в найближчий період, а разом з цим - ієрархію мов, яка демонструватиме здатність мови тієї чи іншої національної або етнічної спільноти до успішної конкуренції в колі світових мов.

З метою збереження мовного розмаїття та відповідних мовних картин світу в статті пропонується організація «Всесвітнього Лінгвістичного Діалогу» як розподіленої мережевої системи соціально-професійної взаємодії в галузі лінгвістики у вигляді виртуалізованої науково-дослідницької та соціальної інфраструктури, лінгвістичних ресурсів та інструментів лінгвістичного дослідження і проектування. Обговорюються концептуальні характеристики і системні параметри «Всесвітнього Лінгвістичного Діалогу», а також науково-організаційні заходи, спрямовані на його створення.

МЕРЕЖЕЦЕНТРИЧНИЙ СВІТ, ТЕХНОЛОГІЧНИЙ СТАТУС МОВИ, ВСЕСВІТНІЙ ЛІНГВІСТИЧНИЙ ДІАЛОГ, СИСТЕМА ЛІНГВІСТИЧНИХ РЕСУРСІВ ТА ЛІНГВІСТИЧНОЇ ІНФРАСТРУКТУРИ

Широков В. А. Языки в сетцентрическом мире. В работе анализируются процессы изменения цивилизационной роли естественного человеческого языка при переходе к сетцентрической фазе эволюции. Отмечается, что рост информации, функционирующей в Сети, уже вышел на экспоненциальный режим и её объёмы уже не могут быть эффективно обработаны, восприняты и осознаны их реципиентами-адресатами. Таким образом, проблемы создания компьютерных инструментов, основанных на принципах искусственного интеллекта (интеллектуализации Сети), уверенно передвинулись на передний план актуальности, а среди множества аспектов интеллектуализации особенно выделяется язык. Технологический статус языка будет определять характер развития цифровой сетцентрической цивилизации в ближайший период, а вместе с этим – и иерархию языков, демонстрирующую способность языка той или иной национальной либо этнической общности к успешной конкуренции в кругу мировых языков.

С целью сохранения языкового разнообразия и соответствующих языковых картин мира в статье предлагается организация «Всемирного Лингвистического Диалога» как распределенной сетевой системы социально-профессионального взаимодействия в области лингвистики в виде виртуализированной научно-исследовательской и социальной инфраструктуры, лингвистических ресурсов и инструментов лингвистического исследования и проектирования. Обсуждаются концептуальные характеристики и системные параметры «Всемирного Лингвистического Диалога», а также научно-организационные меры, направленные на его создание.

СЕТЦЕНТРИЧЕСКИЙ МИР, ТЕХНОЛОГИЧЕСКИЙ СТАТУС ЯЗЫКА, ВСЕМИРНЫЙ ЛИНГВИСТИЧЕСКИЙ ДИАЛОГ, СИСТЕМА ЛИНГВИСТИЧЕСКИХ РЕСУРСОВ И ЛИНГВИСТИЧЕСКОЙ ИНФРАСТРУКТУРЫ

Shirokov V.A. Languages in a network-centric world. The paper analyzes the processes of changing the civilizational role of the natural human language during the transition to the network-centric phase of evolution. It is noted that the growth of information functioning on the Web has already reached an exponential mode and its volumes can no longer be effectively processed, perceived and realized by their recipients. Thus, the problems of creating computer tools based on the principles of artificial intelligence (Network intellectualization) have confidently moved to the forefront of relevance, and among many aspects of intellectualization, language is especially distinguished. The technological status of the language will determine the development modus of digital network-centric civilization in the near future, and with it the hierarchy of languages, demonstrating the ability of a language of one or another national or ethnic community to compete successfully in the world's circle.

In order to preserve linguistic diversity and the corresponding linguistic pictures of the world, the article proposes the organization of the World Linguistic Dialogue as a distributed network system of social and professional interaction in the field of linguistics in the form of a virtualized research and social infrastructure, linguistic resources and tools of linguistic research and design. The conceptual characteristics and system parameters of the World Linguistic Dialogue, as well as scientific and organizational measures aimed at its creation are discussed.

NETWORK-CENTRIC WORLD, TECHNOLOGICAL STATUS OF THE LANGUAGE, WORLD LINGUISTIC DIALOGUE, A SYSTEM OF LINGUISTIC RESOURCES AND LINGUISTIC INFRASTRUCTURE

Вступ

При переході до мережецентричної фази еволюції цивілізаційна роль природної людської мови докорінно змінюється. Всесвітня Мережа вже перетворилася в середу, 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 до загального когнітивно-комунікативного процесу. В своїй найважливішій, концептуальній частині ця інформація представлена переважно структурами природної мови.

Є всі підстави вважати, що зростання інформації, що функціонує в Мережі, вже вийшло на експонентний режим і досягло колосальних обсягів, які збільшуються в постійно зростаючій прогресії. Цілковито обґрунтованими вбачаються сумніви, що такі великі обсяги інформації можуть бути ефективно опрацьовані, а головне – адекватно сприйняті й усвідомлені їхніми реципієнтами-адресатами в реальному часі сучасного буття. Адже експоненційне зростання інформації не супроводжується адекватним зростанням людських можливостей щодо її засвоєння, а тим паче ефективного використання.

Виникає головне протиріччя сучасної епохи [1]: закон еволюції вимагає зростання обсягів інформації, виробленої і сприйнятої 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 ніби сама «вмикає» механізми свого гальмування, що іноді проявляються в досить драматичних формах. Все більшої популярності набувають катастрофічні сценарії розвитку цивілізації, гіпотези технологічної сингулярності, трансгуманістичні концепції тощо.

1. Статус мови в мережецентричному світі

Зазначені особливості сучасного етапу розвитку Мережі висунули проблему створення інструментів, заснованих на принципах штучного інтелекту і здатних взяти на себе значну частину функцій основного когнітивного тракту людини [2]. Таким чином проблема інтелектуалізації Мережі впевнено пересунулася на передній план актуальності, а серед безлічі аспектів інтелектуалізації особливо виділяється мова.

Справді, 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.

Однак особливість лінгвістичних засобів полягає в тому, що мови існують лише у формі окремих національних мов, носії яких, маючи професійну підготовку в області лінгвістичних технологій, власне, й здатні розробляти лінгвоінформаційні компоненти програмного забезпечення на належному рівні та прийнятної якості.

Основою такого роду розробок служать кодифіковані, анотовані й репрезентативні моделі та масиви лінгвістичних даних, що представляють всі (в ідеалі) аспекти функціонування тієї чи іншої мови – як в когнітивному, так і в комунікативному плані. Такими засобами служать лексикографічні середовища [3] та інтелектуальні лексикографічні агенти [4], а також лінгвістичні корпуси [5], які оперують великими цифровими масивами лінгвістично кваліфікованих і анотованих текстів, концептографічні та онтографічні системи обробки та інтелектуальної інтерпретації, різного роду перетворювачі гетерогенних лінгвістичних даних і т. ін.

У цьому зв'язку слід особливо відзначити проблему багатомовності або, як її ще називають, Вавилонський синдром. Як відомо, зараз у світі налічується понад 6000 різних мов, які дуже варіюють за своїм інформаційним та офіційним статусом. Щорічно в світі зникає кілька десятків мов, а число найбільш активно використовуваних мов (у тому числі – в Мережі) можна оцінити величиною приблизно в сто одиниць. При цьому залишається невирішеною і навіть загострюється проблема зникнення малих та міноритарних мов, що становить серйозну загрозу деградації мовного розмаїття і безповоротної втрати цілих мовних картин світу. Створення високоякісних інтерлінгвальних технологій та інструментів ефективного оперування таким складним, динамічним й недостатньо сформованим мовним середовищем наштовхується на цілий ряд труднощів, пов'язаних з проблематикою взаємовідносин між мовною, розумовою та образною підсистемами свідомості та підсвідомості, причому як на індивідуальному, так і на соціальному рівнях. На наше переконання, універсальна мова (інтерлінгва, лінгва франко ...) не може бути створена на субстраті якої-небудь одної або декількох національних мов, а повинна формуватися на основі якихось інших, поки що не дуже зрозумілих принципів.

У той же час очевидно, що в сучасному мережецентричному світі природна людська мова все більш виразно набуває рис не тільки (і навіть не стільки) гуманітарного, скільки (і перш за все) технологічного чинника. Завдяки цьому технологічний статус мови значною мірою визначатиме характер розвитку цифрової мережецентричної цивілізації в найближчий період, а разом з цим – й ієрархію мов, яка буде демонструвати здатність мови тієї чи іншої

національної або етнічної спільноти до успішної конкуренції в колі світових мов.

Варто зробити особливий акцент на високій науковості та трансдисциплінарності створення відповідних інформаційно-лінгвістичних технологій та інструментів.

2. Проєкт «Всесвітній Лінгвістичний Діалог»

Розв'язання цієї проблеми, на нашу думку, можливо лише в середовищі спеціального інтегрованого мультилінгвального середовища. Ми пропонуємо назвати його «Всесвітнім Лінгвістичним Діалогом», що мислиться як далекосяжне узагальнення ідеї «Всеукраїнського Лінгвістичного Діалогу», сформульованої нами ще в 2003 році [6]. «Всесвітній Лінгвістичний Діалог» ми уявляємо собі як розподілену мережеву систему соціально-професійної взаємодії в галузі лінгвістики у вигляді віртуалізованої науково-дослідницької та соціальної інфраструктури, лінгвістичних ресурсів та інструментів лінгвістичного дослідження і проектування.

Це дозволить шляхом виділення відносно невеликих фінансових коштів забезпечити піднесення на якісно новий рівень кадровий склад і фактичні результати в галузі мультилінгвальної інноваційної лінгвістичної технології, теорії і практики штучного інтелекту, їхньої практичної результативності та ефективної інтеграції загальносвітового інформаційно-лінгвістичного, наукового і культурного простору.

Виходячи з викладеного, метою даного проєкту ми вважаємо створення загальносвітової соціально-професійної Мережі «Всесвітній Лінгвістичний Діалог» для збереження мов народів світу, розвитку мовного розмаїття і мультилінгвальної взаємодії в кіберпросторі у вигляді розподіленої, відкритої, загальнодоступної й безперервно еволюціонуючої системи лінгвістичних ресурсів та лінгвістичної інфраструктури.

Таким чином, *визначальними концептами проєкту стають лінгвістичні ресурси та лінгвістична інфраструктура.*

При цьому:

1) до системи *лінгвістичних ресурсів* включаємо:

– Всесвітню словникову базу, а саме:

традиційні словники й цифрові лексикографічні системи, які надають лексикографічний опис (граматичний, семантичний, прагматичний, статистичний, етимологічний, лінгвогеографічний та ін.) мов загальносвітового лінгвістичного простору, а також відповідні лінгвістичні корпуси; цифрову систему термінології і термінографії, цифрову енциклопедичну та довідкову базу;

– засоби цифрової гуманітаристики: електронні системи розпізнавання та синтезу природних мов (фонетична і письмова форми); системи фонетико-графічних трансформацій; інтелектуальні

людинамашинні інтерфейси, засновані на принципах природної мови; інтелектуальні цифрові системи семантичної інтерпретації і 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 мовам; засоби і технології жестових мов; – дослідницькі цифрові лінгвістичні системи;

2) система *лінгвістичної інфраструктури* міститиме:

– Всесвітню лінгвістичну мережу як комплекс спеціалізованих програмно-апаратних засобів доступу і комунікації між локальними лінгвістичними центрами – суб'єктами системи Всесвітніх лінгвістичних ресурсів (ВЛР);

– головний репозиторій системи Всесвітніх лінгвістичних ресурсів;

– віртуальні системи професійної взаємодії в лінгвістиці (віртуальні лінгвістичні лабораторії, мережеві літературні та загальнокультурні енциклопедичні системи і системи спеціального призначення, в тому числі інструментальні);

– комплекс навчальних і нормативно-програмних засобів для забезпечення функціонування і мережевої взаємодії учасників і членів «Всесвітнього лінгвістичного діалогу».

3. Основні напрями «Всесвітнього Лінгвістичного Діалогу»

Зрозуміло, що така універсальна система загальносвітових лінгвістичних ресурсів і відповідної інфраструктури вимагає спеціального і достатньо кваліфікованого проектування. Досвід автора участі в міжнародному проєкті MONDILEX [7] і проєкті створення Національної словникової бази України [8] дозволяє зробити висновок про достатню реалістичність формування глобальних соціально-професійних лінгвістичних систем, які в силу їх відкритості надади б можливість усім мовам і всім фахівцям і аматорам не тільки активно представити свої зацікавленості і практичні напрацювання в загальносвітовому лінгвістичному процесі, а й скористатися найбільш просунутими і ефективними мовними технологіями та інструментами лінгвістичного дослідження, розробленими для найбільш розвинених у цьому відношенні світових мов. Таким чином, проблема збереження мовного розмаїття отримала б ефективний інструментарій для свого вирішення.

Серед завдань проєкту «Всесвітній Лінгвістичний Діалог» вважаємо за необхідне відзначити наступні напрями.

1. Корпусна лінгвістика. Створення лінгвістичних корпусів та корпусних технологій опрацювання текстів.

2. Комплекс віртуальних лексикографічних лабораторій для всіх мов проекту. Розробка і створення інструментальної системи, призначеної для проектування учасниками проекту власних, системно узгоджених віртуальних лексикографічних лабораторій.

3. Головний репозиторій системи Всесвітніх лінгвістичних ресурсів. Засоби адміністрування і система доступу до Головного сховища ВЛР.

4. Засоби міжмовної адаптації та інтеграції різних лінгвістичних ресурсів і систем.

5. Когнітивно-концептографічний комплекс обробки та аналізу природномовних текстів.

6. Інтегровані інструментальні платформи інтелектуальної обробки мовних даних і лінгвістичних структур.

7. Моделювання основного когнітивного тракту людини і розробка відповідних технологічних застосунків.

8. Розробка інтелектуальних лінгвістичних інтерфейсів.

9. Дослідження в галузі ДНК-лінгвістики в контексті біоінформатики і розробка відповідних застосунків.

10. Розробка засад еволюційної лінгвістики.

Зрозуміло, що всі відзначені напрямки повинні розроблятися в обсязі всіх мов - учасників проекту, а в ідеалі охоплювати всі мови народів Світу.

4. Заключні зауваження

Для проектування такої глобальної системи необхідно забезпечити консолідацію зусиль і формування інтернаціонального дослідницького колективу фахівців, що має досвід відповідної роботи. З огляду на складність і масштаб завдання, вбачається доцільним і необхідним прийняття необхідних політичних рішень на міжнародному рівні. Автор, на жаль, не дуже ясно уявляє собі, хто б міг стати суб'єктом даної міжнародної ініціативи. У той же час, її зміст, як ми уявляємо, в основних рисах може бути зведений до двох положень:

– по-перше, констатація доцільності створення загальносвітової соціально-професійної мережі «Всесвітній Лінгвістичний Діалог» для збереження мов народів світу, розвитку мовного розмаїття і мультилінгвальної взаємодії в кіберпросторі у вигляді розподіленої, відкритої, загальнодоступної і безперервно еволюціонуючої системи лінгвістичних ресурсів і лінгвістичної інфраструктури.

– по-друге, організація міжнародної робочої групи фахівців для розробки концепції «Всесвітнього Лінгвістичного Діалогу» і підготовки пропозицій для урядів країн Світу, громадських і приватних організацій зі створення Міжнародного Фонду «Всесвітній Лінгвістичний Діалог».

На основі викладеного ми робимо висновок, що доля світового мовного розмаїття вирішальною мірою залежить від створення глобального мережевого лінгвістичного процесу, який був би здатний надати всім мовам світу можливість рівноправного і ефективного в ньому участі. На нашу думку, ідея «Всесвітнього Лінгвістичного Діалогу» найбільш повно відповідає цьому прагненню.

Список літератури:

- [1] Широков В. А. Эволюция как универсальный естественный закон (Пролегомены к будущей общей теории эволюции). Бионика интеллекта. – Часть I. 2017. № 1 (88). Сс. 3–14. (https://www.ulif.org.ua/system/files/shirokov_bionica_2017.pdf) Часть II., 2017. № 2 (89), Сс. 3-18. <https://www.ulif.org.ua/system/files/bionica-2017-part2.pdf>. Часть III., 2018. – № 1 (90). – С. 3–18. <https://www.ulif.org.ua/system/files/bionica-2018-part3.pdf>
- [2] Palagin A. V., Широков V. A. Principles of cognitive lexicography // International journal «Informational theories & application». – 2000. – Vol. 9. – № 2. – P. 43–51.
- [3] В. А. Широков та ін. Лінгвістично-інформаційні студії: Праці Українського мовно-інформаційного фонду НАН України: у 5 т. Том 1. Наукова парадигма та основні мовно-інформаційні структури. Київ, Український мовно-інформаційний фонд НАН України. 2018. 271 с. ISBN 978-966-02-8683-2. ISBN 978-966-02-8684-9 (Т. 1). DOI 10.33190/978-966-02-8683-2/8684-9
- [4] В. А. Широков та ін. Лінгвістично-інформаційні студії: Праці Українського мовно-інформаційного фонду НАН України: у 5 т. Том 5. Віртуалізація лінгвістичних технологій. Київ, Український мовно-інформаційний фонд НАН України. 2018. 286 с. ISBN 978-966-02-8683-2. ISBN 978-966-02-8690-0 (Т. 5). DOI 10.33190/978-966-02-8683-2/8690-0
- [5] В. А. Широков та ін. Лінгвістично-інформаційні студії: Праці Українського мовно-інформаційного фонду НАН України: у 5 т. Том 4. Корпусна та когнітивна лінгвістика. Київ, Український мовно-інформаційний фонд НАН України. 2018. 300 с. ISBN 978-966-02-8683-2. ISBN 978-966-02-8689-4 (Т. 4). DOI 10.33190/978-966-02-8683-2/8689-4
- [6] Широков В. А. Всеукраїнський лінгвістичний діалог у контексті теорії лексикографічних систем // Мовознавство. – 2003. – № 6. – С. 3-7.
- [7] Main results of MONDILEX project. Ludmila Dimitrova, Violetta Koseska-Toszewa, Radovan Garabik, Tomaz Erjavec, Leonid Iomdin, Volodymyr Shyrokov. <https://ispan.waw.pl/journals/index.php/cs-ec/article/view/cs.2011.017> DOI: <https://doi.org/10.11649/cs.2011.017> Copyright (c) 2015 Ludmila Dimitrova, Violetta Koseska-Toszewa, Radovan Garabik, Tomaz Erjavec, Leonid Iomdin, Volodymyr Shyrokov. License URL: <http://creativecommons.org/licenses/by/3.0/pl/>
- [8] Указ Президента України від 07. 08. 1999 р. № 967 «Про розвиток національної словникової бази». Зокрема, в ході реалізації згаданих проектів було створено близько 50 віртуальних лексикографічних лабораторій, частина з яких експлуатується в промисловому режимі.

Надійшла до редакції 15.01.2020