

файлів пакету MSOffice і 146 файлів текстових повідомлень динамічної бібліотеки дійшов висновку, що результати локалізації не відповідають вимогам ДСТУ 3966-2000. У процесі локалізації було виявлено низку проблем, пов'язану з перекладом термінів: загалом в процесі локалізації нових ІТ продуктів було запроваджено 350 нових термінів [4, с. 123].

Цей приклад локалізації підтверджує, що переклад термінів розглядається як невід'ємна частина локалізації і характеризується специфічними труднощами, пов'язаними з цим новим жанром технічного тексту, що зумовлено постійним оновленням та поповненням терміносистеми новими одиницями та високим рівнем міжмовної інтерференції. Тому професійні перекладачі повинні мати ґрунтовні знання мови-перекладу та володіти навичками лінгвістичного аналізу термінології, тобто мати мовну компетенцію. Крім того, постійно слід відстежувати зміни в термінологічному складі мов оригіналу та перекладу, що прискорить процес локалізації програмного продукту.

### **Список використаних джерел**

1. Бондарчук Л. Особливості локалізації програмного забезпечення українською мовою / Л. Бондарчук // Лексикографічний бюлетень: 36. наук. пр. – К.: Ін-т української мови НАН України, 2006. – Вип. 14. – С. 184–187.
2. Булаховський К. Стан і проблеми розвитку україномовної локалізації комп'ютерних програм» / К. Булаховський // Українська мова, 2010, № 4. – С. 93–97.
3. Єлісеєва С. В. Переклад і локалізація у сфері інформаційних технологій / С. В. Єлісеєва // Наукові праці. Філологія. Мовознавство. Випуск 243. Том 255, 2015. – С. 32–37.
4. Міщенко А. Л. Мультилінгвальна текстова комунікація у світлі сучасних лінгвістичних технологій (на матеріалі технічної документації): дис.... д-ра філол. наук: спец. 10.02.21 / А. Л. Міщенко. – Київ, 2015. – 503 с.
5. Чрділелі Т. В. Лінгвістичні особливості локалізації програмного забезпечення / Т. В. Чрділелі, Н. В. Кожемяченко // Наукові записки Національного університету «Острозька академія». Серія: Філологічна. – 2016. – Вип. 62. – С. 345–348.

*Михайлик Р. П., канд. філол. наук, Ph. D.  
Гарвардський університет, м. Кембридж*

## **РОЗВИТОК І ПОПУЛЯРИЗАЦІЯ УКРАЇНСЬКОМОВНИХ ТЕХНОЛОГІЙ В УКРАЇНІ ТА СВІТІ**

Інформаційні технології розпізнавання, розуміння та синтезу мовлення швидко розвиваються і використовуються в різних галузях суспільного життя: індустрії, освіті, медицині тощо [1]. Найрізноманітніші програми, сервіси й прилади допомагають нам знайти інформацію, не заблукати у місті або надиктувати текст на телефон. Незабаром такі розумні помічники, як Amazon Echo або Google Home увійдуть у дім кожного. На жаль, українськомовні технології не завжди доступні (наприклад, важко знайти GPS українською), незважаючи на величезну

кількість високопрофесійних ІТ спеціалістів – вихідців з України. У цьому інформаційному огляді будуть вказані деякі причини відсутності або непоширеності мовленнєвих технологій, розроблених або локалізованих для України.

Під мовленнєвими технологіями маємо на увазі, наприклад, розпізнавання та розуміння мовлення (ASR – automatic speech recognition) та синтез мовлення за текстом (TTS – text-to-speech). Оскільки одним із найприродніших засобів спілкування між людьми є усне мовлення, спілкування між людиною й комп'ютером також поступово переходить від текстових команд до звукових. Завдання, які вирішують у сфері розпізнавання та розуміння мовлення, такі: автоматичне перетворення мовленнєвого сигналу на текст; введення інформації голосом; усний переклад з однієї мови на іншу, смислова інтерпретація голосових повідомлень; розпізнавання мовлення по телефону для замовлення різноманітних послуг тощо. Спостерігаємо значний прогрес для англійської та інших «значних» мов у всіх цих напрямках. Для української мови теж існують деякі розробки. Наприклад, Nuance Dragon дає можливість диктувати текст українською (з 2012 року), Google Search має звукове забезпечення багатьма мовами, включно з українською (з 2013 року), також з жовтня 2016 року Google Translate додав український голос до списку охоплених мов, а Cloud Speech API перетворює звуковий файл на текст 80 мовами, в тому числі українською (2017). Чи легко було розробити ці продукти, і чи знають про їхнє існування в Україні? На жаль, хоча ASR і TTS стають все розвинутіші загалом, поки що перспективи щодо розвитку української не дуже оптимістичні.

Google Cloud Platform дає змогу Google обробляти величезну кількість даних, отриманих від мільярдів пошукових запитів (надрукованих або усних). Google був у змозі знайти відповідники «текст-звук» з величезною кількістю прикладів людської мови. Зокрема, TTS «читає» текст певною мовою, використовуючи різні моделі: мовні моделі нормалізують текст, фонематичні моделі автоматизують транскрипції слів, а акустичні моделі відповідають за вибір «чистих» звукових файлів природної мови [3]. Наприклад, речення, яке має цифри, скорочення та аббревіації, таке як *«ВВП України в 1-му кварталі 2016 р. становить 453,2 млрд грн, зміна дефлятора – 20,5 %»* має бути спочатку нормалізованим, а потім затранскрибованим і вимовленим згідно з правилами стандартної української граматики і вимови. І тут виникають кілька проблем для розробників українськомовного TTS чи ASR продукту.

Де знайти велику кількість речень і текстів якісною природною українською мовою (не суржилом і не книжним стилем)? Безперечно, вони є на сайтах уряду, різних управлінь та освітніх закладів, але це все зразки офіційно-ділового стилю. Інший стиль можна було б знайти на сайтах про спорт, розваги, подорожі, технологію, моду, музику, але більшість з них – російськомовні. Дописи на Facebook та в блогах мали би дати зразки розмовної мови, але в них переважає або суміш російської з українською, або відвертий суржик. В інших країнах на допомогу ІТ сектору приходять мовознавці, які збирають мовленнєві зразки в різноманітні корпуси, але в Україні немає якісного національного корпусу. Зразки усного мовлення – це також необхідна частина матеріалу для аналізу даних. Пошук Google

дає достатньо багато таких зразків для інших мов, але фрази, які надходять з України – важкі для обробки, адже у них 4 «мови» в одному реченні! Наприклад, у запиті «Хочу скачати гру енгри бьордс» вжито українську (*хочу*), російську (*скачать*), суржик (*ігру*) й англійську (*енгри бьордс*). Порівняйте: «Хочу скачати гру Злі пташки» (українська), «Хочу скачати игру Злые птицы» (російська), та «I want to download the game Angry birds» (англійська). Очевидно, що алгоритми, розроблені для однієї мови, можуть не впоратися із сумішню чотирьох.

Інша проблема – внутрішня, мовознавча. З українською мовою важко працювати, оскільки є дуже мало доступних і надійних лінгвістичних матеріалів. Є велика кількість письмових публікацій про українську мову українською: словники, граматики, описи мови письменників, підручники і навіть кілька граматик англійською [3, 4]. Однак деяких онлайн ресурсів досі не вистачає: наприклад, доступного національного корпусу, повного лексикону всіх морфологічних форм, баз даних зразків усного живого мовлення. Мовознавчі дослідження та інформаційна технологія в основному не пов'язані між собою. Нові терміни та запозичення не згадані в підручниках граматики й повільно стандартизуються. Локалізація продуктів чи назв компаній часто відбувається через російську мову і вибір одного варіанта є проблематичним. Наприклад, як вимовити і написати Google або Windows українською? Що «правильніше»: гугл або гугл?, віндовс, виндовс, віндовз або віндоуз? Як правильно прочитати е-мейл-адресу (email@google.com), кількість грошей (3,99 €), або номер кредитної картки (9838-5010-0722-2548)? Українські філологи напевно зможуть відповісти на ці питання і, може, навіть вказати сторінку підручника з правилами, але у розробників немає можливості працювати в українських бібліотеках, й онлайн пошук був би набагато ефективнішим.

Загалом, експерти з великих ІТ компаній часто недооцінюють перспективність української мови і, незважаючи на досить велику цифру потенційних покупців (40-мільйонне населення України), вважають за краще зосередитися на розробці нових продуктів російською мовою. Двомовність більшості українців – добре відомий факт. Вони справді можуть використовувати технологічні новинки російською мовою. Ця безпосередня перевага, однак, перетворюється в недолік упродовж тривалого часу: українська мова залишається знехтуваною у світовому ІТ-секторі, та її використання в Україні продовжує звужуватися, поступаючись російській у ще одній сфері – галузі мовних технологій.

Як збільшити українську присутність в ІТ? Необхідні зацікавлені професіонали, якісні дані й лінгвістичні дослідження міжнародного рівня. Почати можна із простого: налаштувати операційну систему (комп'ютер або телефон) на українську або переконатись, що мова Google пошук відповідає мові країни, тобто українська для України. Соціальні медіа (FB, блоги, музика, спорт, розваги) повинні мати український контент. Мовознавцям потрібно рухатися вперед разом з технологіями: більше довідкових матеріалів має бути адаптовано для прикладних лінгвістичних досліджень. Стандартні словники мають давати повну інформацію про слово – транскрипцію, морфологію й орфографію. Інститут української мови НАН України, Український мовно-інформаційний фонд НАН Украї-

ни та інші інституції могли б докласти більше зусиль для збирання й аналізу *живої* української мови та для локалізації запозичень.

На завершення потрібно наголосити, що українськомовні технології (наприклад, синтезовані голоси від UkrVox 4.2 (1994) до Google TTS (2016)) існують, але їхнє вдосконалення, використання та популяризація залежить від самих українців.

### Список використаних джерел

1. Moskvitch, K. The machines that learned to listen. – <http://www.bbc.com/>, 15 February 2017.

2. McLellan, Ch. How we learned to talk to computers, and how they learned to answer back. – <http://www.techrepublic.com/>

3. Словник української мови: в 11 тт. / АН УРСР. Інститут мовознавства; за ред. І. К. Білодіда. – К.: Наукова думка, 1970–1980.

4. Pugh, S. and Press, I. Ukrainian: A Comprehensive Grammar. – London: Routledge, 2005.

*Сєркова Г. М.  
Донецький національний університет імені Василя Стуса,  
м. Вінниця*

## ВИБІР ІНТЕГРОВАНОЇ ІНФОРМАЦІЙНОЇ СИСТЕМИ ЯК ОПТИМАЛЬНИЙ НАПРЯМ ЕКОНОМІЧНОГО РОЗВИТКУ ПІДПРИЄМСТВА

На кожному підприємстві є різні рівні управління, на яких циркулюють специфічні інформаційні потоки. Для обробки інформації використовують різні інформаційні технології, які реалізуються за допомогою відповідних інформаційних систем.

Корпоративні (інтегровані) інформаційні системи управління на кожному підприємстві можна описувати за рівнями та процесами обробки інформації (табл. 1).

*Таблиця 1*

**Стратифікація інформаційних технологій за операціями**

Рівень 1	Вхід 2	Обробка 3	Вихід 4	Користувачі 5
ESS	Сукупні дані	Аналіз і прийняття рішень, моделювання	Рішення, стратегії, плани	Вище керівництво
KWS	Технологічні дані, база знань	Моделювання, аналіз, прогнозування	Моделі, результати аналізу, графіки, таблиці, звіти	Аналітики, ІТ-професіонали