DOI 10.18524/2307–4558.2024.42.325844

УДК [811.161.2+811.111]'322.4'25

ЯРОЦЬКА Галина Сергіївна,
доктор філологічних наук, професор кафедри прикладної лінгвістики Одеського національного університету імені І. І. Мечникова; Французький бульвар, 24/26, м. Одеса, 65058, Україна; тел.: +380503916600;
e-mail: g.yarotskaya@gmail.com; ORCID ID: https://orcid.org/0000–0003–4456–3852

БУШУЄВ Даніл Володимирович,
магістрант кафедри прикладної лінгвістики Одеського національного університету імені І. І. Мечникова;
Французький бульвар, 24/26, м. Одеса, 65058, Україна; тел.: +380982442627;
e-mail: danilbgm20@gmail.com; ORCID ID: https://orcid.org/0009–0000–0373–5414

ОСОБЛИВОСТІ РОЗПІЗНАВАННЯ СПЕЦІАЛЬНИХ СИМВОЛІВ І ТЕРМІНІВ СИСТЕМАМИ МАШИННОГО ПЕРЕКЛАДУ

Анотація. ***Метою*** статті є дослідити здатність систем машинного перекладу і штучного інтелекту розпізнавати спеціальні символи й терміни. Аналіз здійснено з опертям на результати машинного перекладу з англійської мови українською і навпаки. ***Об'єкт*** дослідження — лінгвістична терміносистема, автоматичний пошук перекладацьких відповідників трьома системами машинного перекладу та штучного інтелекту; ***предмет*** — розпізнавання системами машинного перекладу та штучного інтелекту спеціальних символів й термінів у перекладі з англійської мови українською і, навпаки, виявлення проблем машинного перекладу, оцінка доречності й недоречності варіантів. У студії в межах підходу теорії перекладу та комп'ютерної лінгвістики використано як загальнонаукові ***методи***, так і суто лінгвістичні: аналіз, синтез, порівняльний і зіставний методи, методи соціолінгвістичного та перекладацького аналізу, описовий метод. ***Результатами*** роботи є визначення проблемних випадків розпізнавання, що трапляються за умови машинного перекладу наукових текстів, та їх аналіз. Наукові тексти можуть містити спеціальні символи для передання особливостей фонетики або правопису іншомовного слова, в тому числі терміна; загальні терміни можуть мати відмінності у правописі англійської або української мови. У деяких наукових текстах лінгвістичні терміни можуть братися в лапки, які також спричинюють труднощі їх розпізнавання системами. ***Галузь застосування*** проведеного дослідження полягає у можливостях використання його результатів у подальших наукових розвідках у галузі комп'ютерної лінгвістики, лінгвістичному моделюванні та корпусній лінгвістиці, теорії і практики перекладу, практичного курсу перекладу. ***Висновки*** дослідження вбачаються у результатах порівняльного аналізу машинного перекладу від різних систем: якість перекладів є досить високою; доречність певних варіантів перекладу. Проте існу-

ють ***проблеми*** розпізнавання спеціальних символів і термінів, що потребує подальшого вдосконалення й оптимізації машинного перекладу.

Ключові слова: машинний переклад, штучний інтелект, розпізнавання, комп'ютерна лінгвістика, теорія перекладу, спеціальний символ, термін.

Постановка проблеми. Машинний переклад (*далі* – МП), який є молодою галуззю науки на перетині перекладознавства й інформатики, належить до мовознавчої галузі комп'ютерної лінгвістики. МП своїм виникненням мотивований практичними потребами, що постали у 50-ті роки XX ст., коли зросли обсяги і ємність науково-технічної інформації. Спочатку прогнозувалося, що системи МП можуть легко забезпечувати переклад будь-яких науково-технічних текстів, але перші спроби засвідчили, що виникли суттєві потреби у нетривіальних даних різного типу про функціювання мовних систем [2, т. 3, с. 35]. Ця галузь у наші часи зазнала потужної цифровізації, вдосконалення систем МП. Розробка нових систем штучного інтелекту вплинула на МП появою нових методів використання інформаційних технологій у перекладацькій діяльності. Це й визначає актуальність досліджень щодо оптимізації машинного перекладу, зокрема нашої наукової розвідки.

Зв'язок із попередніми та суміжними дослідженнями. У 1949 р. американський вчений Уоррен Уівер видав меморандум під назвою «Translation», у якому одним із перших розглянув можливості використання комп'ютерів для мовного перекладу. З того часу проблеми машинного та класичного перекладу стали одними із найактуальніших у прикладній лінгвістиці. У нашій роботі ми спиралися на концепції кількох шкіл машинного перекладу, зокрема на праці Т'єрі Пуабо (Thierry Poibeau) [15], виконані у Лабораторії комп'ютерних наук Париж-Норд (LIPN) Сорбонни Північного Парижу (USPN), у Лабораторії Латтіс Нової Сорбонни (Laboratoire Lattice) і в лабораторії Департаменту теоретичної та прикладної лінгвістики Кембриджського університету; на праці українських дослідників В. Волошина [1] і Є. Карпіловської [3]; на деякі інші роботи.

Формулювання завдань. Мету дослідження можна схарактеризувати як виявлення проблем розпізнавання системами машинного перекладу спеціальних символів і термінів у наукових текстах. Це зумовило вирішення низки завдань, що стосуються, по-перше, визначення ролі систем МП та штучного інтелекту у перекладацькій діяльності на матеріалі аналізу спеціальних символів і термінів, виявлених у текстах кількох статей із соціолінгвістики (мовні проблеми білінгвізму); по-друге, — встановлення причин виникнення проблем розпізнавання спеціальних символів і недоречного перекладу термінів. У процесі роботи використано як загальнонаукові методи аналізу й синтезу, так і власне лінгвістичні: порівняльний, зіставний, соціолінгвістичного та перекладацького аналізу, описовий метод. Новизна дослідження зумовлена тим, що перекладацький аналіз було здійснено вперше на матеріалі наукових текстів із соціолінгвістики.

Виклад основного матеріалу дослідження. Основними напрямами прикладної лінгвістики є такі: автоматичні методи переробки текстової інформації, реферування і перекладу; лексичне забезпечення автоматизованих систем; стандартизація науково-технічної термінології [1, с. 12]. Зазначимо, що «машинний переклад — це процес перетворення комп'ютером тексту, створеного засобами однієї природної мови, в текст, оформлений засобами іншої природної мови» [3, с. 164].

Теоретичне підґрунтя дослідження базується на розумінні перекладацьких норм, які поділяються на: ***норми очікування*** цільової аудиторії, стосовно того, яким має бути переклад, та ***професійні***. Доречність перекладу визначаємо розумінням доречності для певної комунікативної ситуації. «При цьому якість не означає, що можливий тільки один вірний переклад, бо норм можна дотримуватися різними способами. Норми очікування дозволять робити оцінні твердження щодо перекладу» [5, с. 60]. Отже, якісний переклад наукового лінгвістичного тексту має відповідати терміносистемі, що спирається на відомості з наявних двомовних словників лінгвістичної термінології.

Машина почерговими процедурами порівнює вхідні одиниці одного рівня з одиницями інших рівнів, записаних в її пам'яті. Якщо такого порівняння немає, машина зреалізовує передбачувану модель перекладу фрагмента тексту [1, с. 314].

Системи МП використовують різні підходи та засоби щодо виконання перекладу. Результат може бути прямим перекладом з однієї мови іншою без проміжного унаявлення або непрямим, коли система спочатку намагається подати більш узагальнений контекст. Це проміжне унаявлення також може бути незалежним від мови, щоб зробити можливим прямий переклад одного оригінального тексту різними цільовими мовами. «Кожна система є унікальною і використовує більш-менш оригінальний підхід до проблеми» [15, с. 25].

Далі розглянуто і проаналізовано приклади розпізнавання спеціальних символів і термінів за допомогою трьох систем МП: Google Translate (*далі* – GT) [11], DeepL Translate (*далі* – DL) [8], Glosbe Translate (*далі* – GL) [10]. Для встановлення доречностей розпізнавання термінів використано такі електронні словники, з яких два перекладні: Glosbe [9], Lingea [13] — і три тлумачні: СЛОВНИК.ua [4], Oxford Learner's Dictionaries [14], Cambridge Dictionary [7].

Перший текст — анотація статті «Деколонізація суржику: тяглість і зміни в ідеології використання змішання мов» української дослідниці Наталі Кудрявцевої.

Текст мовою оригіналу	«*The Ukrainian-Russian mixed variety suržyk has been a contested issue ever since it arose during the processes of Ukraine's colonisation and Russification by the Russian Empire and the USSR. <...> This ideology elevating a "pure" (Russian) language portrayed the languages of other Soviet republics as less legitimate while non-standard varieties were stigmatised. The dismantling of the USSR and the change in the status of the Ukrainian language saw a similar heightened concern for linguistic correctness. The necessity to legitimize Ukrainian as the (only) language of the Ukrainian nation encouraged the ideology which sought to distinguish a "high-status language" from the Ukrainian language varieties used by rural people who would often speak in a mix. In the 1990s, suržyk became iconic of cultural lowness and was seen as "contamination" caused by Russification. In this talk, I aim to explore the attitudes towards suržyk among people who decided to convert from Russian to speaking Ukrainian after the 2014 Russian aggression and the 2022 full-scale Russian invasion. I draw on interviews with 65 participants of volunteer Ukrainian language initiatives who reside primarily in Ukraine's east. These speakers may be associated with the so called "neo-suržyk" — a Russian-based variety different from the "old" Ukrainian-based suržyk developed in Ukraine since the late 19th century. Through a micro-level discourse analysis, I aim to trace how new, more positive attitudes towards suržyk emerged in the processes of "linguistic conversion" and de-Russification. I conclude by interpreting such views of suržyk as "a stage in the learning of/transition to Ukrainian", "a weapon", "a dialect or 'colour' of the Ukrainian language", as well as the overall appreciation of suržyk as being "better than Russian", as a sign of linguistic decolonisation*» [12, с. 17].
Результати машинного перекладу	
Система	Переклад
GT	«*Змішаний українсько-російський різновид суржику є дискусійним питанням ще відтоді, як він виник під час процесів колонізації та русифікації України Російською імперією та СРСР. <...> Ця ідеологія, що підносить «чисту» (російську) мову, зображувала мови інших радянських республік як менш легітимні, а нестандартні різновиди таврували. Розпад СРСР і зміна статусу української мови показали таку ж підвищену турботу про лінгвістичну правильність. Необхідність узаконити українську як (єдину) мову української нації заохочувала ідеологію, яка прагнула відрізнити «мову високого статусу» від українських мовних різновидів, якими користуються сільські жителі, які часто розмовляли змішано. У 1990-х роках суржик став іконою культурної низькості й розглядався як «забруднення», спричинене русифікацією. У цій доповіді я маю на меті дослідити ставлення до суржику серед людей, які вирішили перейти з російської на українську після російської агресії 2014 року та повномасштабного російського вторгнення 2022 року. Я спираюся на інтерв'ю з 65 учасниками волонтерських україномовних ініціатив, які мешкають переважно на сході України. Ці мовці можуть бути пов'язані з так званим «неосуржиком» — російською різновидом суржику, який відрізняється від «старого» українського суржику, який розвивався в Україні з кінця 19 століття. Через аналіз дискурсу на мікрорівні я прагну простежити, як у процесах «лінгвістичної конверсії» і дерусифікації з'явилося нове, більш позитивне ставлення до суржику. На завершення я тлумачу такі погляди на суржик як «етап вивчення/переходу до української», «зброю», «діалект чи «кольор» української мови», а також загальне оцінювання суржику як «краще російської», як ознака мовної деколонізації*».
DL	«*Українсько-російський мішаний суржик є дискусійним питанням відтоді, як він виник під час процесів колонізації та русифікації України Російською імперією та СРСР. <...> Ця ідеологія, що вивищувала "чисту" (російську) мову, зображувала мови інших радянських республік як менш легітимні, а нестандартні різновиди стигматизувала. Демонтаж СРСР і зміна статусу української мови призвели до такого ж посилення уваги до мовної коректності. Необхідність легітимізації української мови як (єдиної) мови української нації стимулювала ідеологію, яка прагнула відрізнити «мову високого статусу» від різновидів української мови, якими користуються сільські жителі, що часто розмовляють мішаниною. У 1990-х роках суржик став символом культурної низькості і розглядався як «забруднення», спричинене русифікацією. У цій доповіді я маю на меті дослідити ставлення до суржику серед людей, які вирішили перейти з російської на українську мову після російської агресії 2014 року та повномасштабного російського вторгнення 2022 року. Я спираюся на інтерв'ю з 65 учасниками волонтерських україномовних ініціатив, які проживають переважно на сході України. Ці мовці можуть бути пов'язані з так званим «неосуржиком» — російськомовним різновидом суржику, відмінним від «старого» українського суржику, що розвивався в Україні з кінця 19 століття. За допомогою аналізу дискурсу на мікрорівні я прагну простежити, як у процесі «мовної конверсії» та дерусифікації з'явилося нове, більш позитивне ставлення до суржику. На завершення я інтерпретую такі погляди на суржик як «етап у вивченні/переході на українську мову», «зброю», «діалект або «колір» української мови», а також загальне сприйняття суржику як «кращого за російську», як ознаку мовної деколонізації*».
GL	«*Українсько-російський змішаний сорт Суржик був спірним питанням з тих пір, як він виник під час процесів колонізації та русифікації України Російською імперією та СРСР. <...> Ця ідеологія, що підвищувала "чисту" (російську) мову, зображувала мови інших радянських республік як менш легітимні, в той час як нестандартні різновиди були стигматизовані. Розпад СРСР та зміна статусу української мови призвели до подібної підвищеної турботи про мовну коректність. Необхідність легітимізувати українську мову як (єдину) мову української нації заохочувала ідеологію, яка прагнула відрізнити «мову високого статусу» від різновидів української мови, що використовуються сільськими людьми, які часто говорять змішано. У 1990-х рр. сюржик став іконою культурної низості і розглядався як ?? зараження ??, викликане русифікацією. У цій доповіді я прагну дослідити ставлення до сержика серед людей, які вирішили перейти з російської мови на українську після російської агресії 2014 року та повномасштабного російського вторгнення 2022 року. Я опираюся на інтерв'ю з 65 учасниками волонтерських ініціатив української мови, які проживають переважно на сході України. Ці носії мови можуть бути пов'язані з так званим ?? нео-суржиком ??, який відрізняється від ?? старого ?? українського суржика, що розвивався в Україні з кінця 19 століття. Завдяки аналізу дискурсу на мікрорівні, я прагну відстежити, як з'явилися нові, більш позитивні ставлення до хірургії в процесі «лінгвістичного переходу» та «дерусифікації». На завершення, я інтерпретую такі погляди на suržyk як «ступінку в навчанні/переходу до української мови», «зброю», «диалектом або «кольором» української мови», а також загальну оцінку suržyk як «кращого за російський», як ознаку мовної деколонізації*».

Слід зазначити, що в тексті, термін *суржик* англійською передано як *suržyk*. Тут використано нетипову транслітерацію: український звук [*ж*] треба передавати диграфом *zh*.

У тексті вжито такі терміни: **1.** *contested issue* — ***дискусійним питанням*** (GT, DL); ***спірним питанням*** (GL) — обидва варіанти перекладу слова *contested* є правильними [9], але в контексті науки доречніше використати варіант *дискусійне* питання, тобто варіанти від систем МП GT, DL; **2.** *linguistic correctness* — ***лінгвістичну правильність*** (GT); ***мовної коректності*** (DL); ***мовну коректність*** (GL) — щодо слова *concern* доречним є варіант від DL; *linguistic* — доречніші варіанти від систем МП DL, GL, бо в українській мові *лінгвістичний* це більше те, що стосується науки [4]; *correctness* — доречнішими є варіанти від систем МП DL та GL [9]; **3.** *legitimize* — ***узаконити*** (GT); ***легітимізації*** (DL); ***легітимізувати*** (GL) — *легітимізація* це *процес обґрунтування законності* та *прагнення представити свої дії як правомірні* [4], що доречно за контекстом, тому доречними є варіанти від систем МП DL, GL; **4.** *high-status language* — ***мову високого статусу*** (усі системи) — отриманий варіант є доречним [4]; **5.** *contamination* — ***забруднення*** (GT, DL); ***зараження*** (GL) — *contamination* перекладається як *контамінація, забруднення, зараження* [13]; за контекстом правильним варіантом є *забруднення*, тому варіант від системи МП GL *зараження* є недоречним; **6.** *Speakers* — ***мовці*** (GT, DL); ***носії мови*** (GL) — за контекстом, *speakers* — мається на увазі *респонденти, які взяли участь у соцопитуванні*, тому можна вважати варіант від системи МП GL *носії мови* недоречним, бо тоді в оригіналі було б *native speakers* [9]; **7.** *neo-suržyk* — ***неосуржиком*** (GT, DL); ***нео-суржиком*** (GL) — слова з іншомовним префіксом *нео-* пишуться разом [4], тому доречними є варіанти від систем МП GT, DL; **8.** *linguistic conversion* — ***лінгвістичної конверсії*** (GT); ***мовної конверсії*** (DL); ***лінгвістичного переходу*** (GL) — доречним є варіант від системи МП DL [7]; **9.** *Ukrainian-based suržyk* — ***українського суржику*** (GT, DL); ***українського суржика*** (GL) — за електронним словником української мови, форма родового відмінку слова *суржик — суржику* [4]; жоден з варіантів перекладу *Ukrainian-based* не є доречним, бо йдеться про *заснованого на базі української мови суржику*; **10.** *De-Russification* — ***дерусифікації*** (усі системи) — отриманий варіант є доречним [4].

Слід зауважити, що використання спеціальних символів, не характерних для певної мови, дійсно може викликати складнощі у розпізнаванні тексту машиною, в даному випадку в системі МП GL: у першому реченні тексту розпізнано спеціальний символ *ž* і було отримано правильний переклад терміна *suržyk — Суржик*, хоча з невеликою помилкою — слово скальковано з англійської, бо написано з великої літери: в англійській назви мов, діалектів, говірок тощо, як правило, пишуть з великої літери; нарешті, отримано проблемні варіанти — *сюржик / сержик*. Словосполучення *towards suržyk* перекладено як *до хірургії*, що зовсім не відповідає контексту, бо про хірургію там взагалі не йдеться; термін *хірургія* англійською буде *surgery*, за транскрипцією *[ˈsɜːdʒərɪ]* [13], що за правописом і вимовою нагадує термін *suržyk;* це може бути другою причиною виникнення проблеми. Також це може пояснити дві інших перекладацьких недоречності в тексті — *сюржик / сержик*. Далі система зовсім не розпізнала цей термін і залишила його написання латиницею.

Усі системи по-різному розпізнали лапки, які в оригіналі “ ”, а в МП — « » (усі системи); а в окремих випадках система GL зовсім не розпізнала лапки, поставивши на їхнє місце знаки питання (*?? ??*). У системі МП GL виникли проблеми розпізнавання літери *і* в слові *dialect*, де в перекладі на українською замість *і* отримано *и*; *variety* перекладено як *сорт*, що є недоречним в контексті лінгвістики [4].

Наступним матеріалом для перекладу послугував текст наукової статті «Мовна ситуація у Великобританії» української дослідниці Ганни Штомпель.

Текст мовою оригіналу	«<...> *Але помилково було би зараз завершити розмову про мови британців, оскільки актуальним стало питання повсталих майже з небуття ще двох кельтських мов, які вважалися мертвими ще років десять тому. Вкрай мало людей знають або чули хоча б щось про корнську чи менську (менкську)!* <...> *Але на величезний подив, навіть пересічного англійця, завдяки титанічній праці ентузіастів-лінгвістів, що на основі двомовних — англо-корнських та англо- менських словників, та зібраним по крихтам літературним джерелам цими мовами, почали стрімке динамічне відродження літературних варіантів цих мов. На даний момент на півострові Корнуолл та острові Мен діють дошкільні заклади, школи, факультети університетів де навчання провадиться корнською чи відповідно менською. Ожилі мови відвойовують все нові й нові позиції у суспільстві. Іноді трапляються, навіть, комічні ситуації коли, наприклад, водій з Англії (до якої територіально належить Корнуолл) не може дістатися потрібного пункту, оскільки таблички на дорогах написані тільки по-корнськи*» [6, с. 1].
Результати машинного перекладу	
Система	Переклад
GT	«<...> *But it would be a mistake to end the conversation about the languages of the British now, since the issue of two more Celtic languages, which were considered dead ten years ago, arose almost from nothingness, has become relevant. Very few people know or have heard at least something about Cornish or Menk (Menk)!* <...> *These are the languages of the local population of the British peninsula of Cornwall and the Isle of Man, located in the Irish Sea. Until the 2000s, there was a lot of different, often contradictory information about them.* <...> *But to the great surprise of even an average Englishman, thanks to the titanic work of linguist enthusiasts, on the basis of bilingual — English-Cornish and Anglo-Mannish dictionaries, and collected bits and pieces of literary sources in these languages, a rapid dynamic revival of literary variants of these languages*

Система	Переклад
	began. At the moment, there are pre-schools, schools, and university faculties operating on the Cornwall peninsula and the Isle of Man, where education is conducted in Cornish or, respectively, in Manx. Living languages are winning new and new positions in society. Sometimes there are even comical situations when, for example, a driver from England (to which Cornwall territorially belongs) cannot reach the desired point, because the signs on the roads are written only in Cornish».
DL	(британська) «<...> *But it would be a mistake to end the conversation about the languages of the British now, as the issue of two more Celtic languages that were considered dead ten years ago has become relevant. Very few people know or have heard at least something about Cornish or Menoese!* <...> *These are the languages of the local population of the British Cornish peninsula and the Isle of Man, located in the Irish Sea. Until the 2000s, there was a lot of diverse, often contradictory information about them.* <...> *But to the great surprise of even the average Englishman, thanks to the titanic work of enthusiastic linguists, who based their bilingual dictionaries on English-Cornish and English-Meno, and collected bits and pieces of literary sources in these languages, a rapid and dynamic revival of literary variants of these languages began. At the moment, there are pre-schools, schools, and university faculties on the Cornish peninsula and the Isle of Man where education is provided in Cornish or Manx, respectively. Revived languages are gaining more and more ground in society. Sometimes there are even comical situations when, for example, a driver from England (to which Cornwall belongs geographically) cannot get to the desired destination because the signs on the roads are written only in Cornish*». (американська) «<...> *But it would be a mistake to end the conversation about the languages of the British now, as the issue of two more Celtic languages that were considered dead ten years ago has become relevant. Very few people know or have heard at least something about Cornish or Menoese!* <...> *These are the languages of the local population of the British Cornish peninsula and the Isle of Man, which is located in the Irish Sea. Until the 2000s, there was a lot of diverse, often contradictory information about them.* <...> *But to the great surprise of even the average Englishman, thanks to the titanic work of enthusiastic linguists, who based their bilingual dictionaries on English-Cornish and English-Meno, and collected bits and pieces of literary sources in these languages, a rapid dynamic revival of literary variants of these languages began. At the moment, on the Cornish peninsula and the Isle of Man, there are pre-schools, schools, and university departments where education is provided in Cornish or Manx, respectively. Revived languages are gaining more and more ground in society. Sometimes there are even comical situations when, for example, a driver from England (to which Cornwall belongs geographically) cannot get to the desired destination because the signs on the roads are written only in Cornish*».
GL	«<...> *But it would be a mistake to end the discussion of the languages of the Britons at this point, since the question of the near-extinction of two other Celtic languages that were considered dead a decade ago has been raised. Very few people know or have ever heard of Cornish or Minskoff!* <...> *They are spoken by the indigenous people of the British peninsula of Cornwall and the Isle of Man in the Irish Sea. Until the 2000s, there was a great variety of, and often conflicting, information about them.* <...> *But to the amazement of even the average Englishman, thanks to the titanic work of enthusiastic linguists, based on bilingual ?? Anglo-Cornish and Anglophone dictionaries, and collected from the tiny literary sources of these languages, a rapidly dynamic revival of literary variants of these languages has begun. Preschools, schools, and university departments in Cornwall and the Isle of Man currently teach Cornish or Mannish, respectively. Language centers are taking on new and new positions in society. There are even comical situations, such as when a driver from England (to which Cornwall belongs territorially) cannot reach the desired point because the road signs are written only in Cornish*».

Зафіксовані терміни та власні назви: **1.** *повсталих <мов>* — ***arose*** (GT); - (DL); ***the near-extinction*** (GL) — доречним є варіант від системи МП GL, який перекладається як *які майже вимерли*, бо маються на увазі мови, які вважали мертвими, англійською можна сказати *extinct language* [9]; **2.** *мертвими <мовами>* — ***dead*** (усі системи) — отриманий варіант недоречний, бо англійською термін *мертва <мова>* — *extinct language* [9]; **3.** *ентузіастів-лінгвістів* — ***linguist enthusiasts*** (GT); ***enthusiastic linguists*** (DL, GL)– доречним є варіант від систем МП DL та GL [14]; **4.** *двомовних* — ***bilingual*** (усі системи) — отриманий варіант є доречним [13]; **5.** *літературним джерелам* — ***literary sources*** (усі системи) — отриманий варіант є доречним [14]; **6.** *відродження* — ***revival*** (усі системи) — отриманий варіант є доречним [7]; **7.** *літературних варіантів <мов>* — ***literary variants*** (усі системи) — отриманий варіант за контекстом є цілком доречним**, хоча англійською мовою доречніше *literary language*** [9]; **8.** *ожилі <мови>* — ***living languages*** (GT); ***revived languages*** (DL); ***language centers*** (GL) — стосовно мов, доречним є варіант від системи МП DL [13; 14]; у системи МП GL виникли труднощі з розпізнаванням цього терміну; **9.** *позиції* — ***positions*** (GT, GL); ***ground*** (DL) — за контекстом *позиції* перекладється як *ground* [7]; отже, доречним є варіант від системи МП DL; **10.** *територіально* — ***territorially*** (GT, GL); ***geographically*** (DL) — *territorially* це *стосовно країни, групи людей, місцезнаходження*, а *geographically — місце у фізичному просторі* [7]; тому доречними є варіанти від систем МП GT, GL; **11.** *пункту* — ***point*** (GT, GL); ***destination*** (DL) — доречним є варіант від системи МП DL [7].

Варто зауважити, що в усіх системах зафіксовані проблеми розпізнавання власної назви *острову Мен* і спільнокореневих слів: *Menk* (GT); *Menoese* (DL); *Minskoff* (GL), тоді як англійською — *Isle of Man*, *Manx* [7].

Висновки. Отже, системи машинного перекладу і штучного інтелекту мають деякі проблеми у розпізнаванні спеціальних графічних символів: тире, лапок, деяких інших. У тексті, написаному англійською мовою, жодна система не розпізнала лапки такими, якими їх подано в оригіналі. Зафік-

совано також випадки неврахування орфографії у перекладі деяких термінів і випадки перекладу недоречними за контекстом еквівалентами. У тексті, написаному українською мовою, немає спеціальних символів, але наявна велика кількість термінів і власних назв. Було встановлено недоречні за контекстом еквіваленти перекладу та проблеми розпізнавання деяких власних назв і похідних присвійних прикметників. Проте було отримано також і значну кількість доречних результатів від усіх систем, що є свідченням постійного вдосконалення систем машинного перекладу. Це дає можливість адресатові інтерпретувати текст перекладу як адекватний та еквівалентний тексту оригіналу в цілому. Отже, виникає можливість визначити тексти перекладу як тотожні за змістом тексту-джерелу. Засобом оптимізації вважаємо вдосконалення розпізнавання спеціальних символів і збільшення словникового складу термінів і власних назв у системах машинного перекладу. **Перспективи** дослідження вбачаємо в організації збору й аналізу експертних оцінок фахівців із певних наукових галузей і професійних перекладачів щодо визначення якості машинного перекладу.

Література

1. Волошин В. Г. Комп'ютерна лінгвістика : навч. посібник. Суми : Університетська книга, 2004. 382 с.
2. Загнітко А. П. Словник сучасної лінгвістики : поняття і терміни : у 4 т. Донецьк : ДонНУ, 2012. Т. 3. 424 с.
3. Карпіловська Є. А. Вступ до прикладної лінгвістики : комп'ютерна лінгвістика : посібник. Донецьк : Юго-Восток, 2006. 188 с.
4. СЛОВНИК.ua. URL : https://slovnyk.ua/index.php/ (дата звернення : 5.10.2024).
5. Тащенко Г. В. Актуальні проблеми теорії та практики перекладу : конспект лекцій. Харків : ХНУ ім. В. Н. Каразіна, 2021. 165 с.
6. Штомпель Г. В. Мовна ситуація у Великобританії. *View metadata, citation and similar papers at core.ac.ukprovided by Electronic archive of Kirovograd National Technical University.* 2010. URL : https://core.ac.uk/reader/81588500/ (дата звернення : 5.10.2024).
7. Cambridge Dictionary. URL : https://dictionary.cambridge.org/uk/dictionary/ (дата звернення : 5.10.2024).
8. DeepL Translate. URL : https://www.deepl.com/uk/translator/ (дата звернення : 3.10.2024).
9. Glosbe. URL : https://uk.glosbe.com/ (дата звернення : 2.10.2024).
10. Glosbe Translate. URL : https://translate.glosbe.com/ (дата звернення : 2.10.2024).
11. Google Translate. URL : https://translate.google.com/?hl=uk&sl=auto&tl=en&op=translate/ (дата звернення : 4.10.2024).
12. Kudriavtseva N. Decolonising Suržyk : Continuity and Change in the Ideologies of Mixed Language Use. *Workshop June 13–14, 2024 Justus Liebig University Giessen and online. Abstracts* : book of abstracts / organisers prof. dr. M. Wingender, dr. L. Pidkuimukha. Giessen : Justus Liebig University Giessen, Giessen Center for Eastern European Studies (GiZo), German Research Foundation (DFG), 2024. P. 17–18.
13. Lingea. URL : https://www.dict.com/ (дата звернення : 4.10.2024).
14. Oxford Learner's Dictionaries. URL : https://www.oxfordlearnersdictionaries.com/definition/english/ (дата звернення : 5.10.2024).
15. Poibeau T. Machine Translation. London : The MIT Press, 2017. 214 p.

References

1. Voloshyn, V. H. (2004), *Computational Linguistics* [*Komp"yuterna linhvistyka*], University Press, Sumy, 382 p.
2. Zahnitko, A. P. (2012), *Dictionary of Modern Linguistics : concepts and terms* [*Slovnyk suchasnoyi linhvistyky : ponyattya i terminy*], DonNU Press, Donetsk, Vol. 3. 424 p.
3. Karpilovska, Ye. A. (2006), *Introduction to Applied Linguistics : Computational Linguistics* [*Vstup do prykladnoyi linhvistyky : komp"yuterna linhvistyka*], Yuho-Vostok, Donetsk, 188 p.
4. SLOVNYK.ua. Available at : https://slovnyk.ua/index.php/ (Accessed : 5.10.2024).
5. Tashchenko, H. V. (2021), *Current Problems of the Translation Theory and Practice* : *lecture notes* [*Aktual'ni problemy teoriyi ta praktyky perekladu : konspekt lektsiy*], V. N. Karazin Kharkiv National University, 165 p.
6. Shtompel, H. V. (2010), "Language Situation in Great Britain", *View metadata, citation and similar papers at core.ac.ukprovided by Electronic archive of Kirovograd National Technical University* ["Movna sytuatsiya u Velykobrytaniyi", *View metadata, citation and similar papers at core.ac.ukprovided by Electronic archive of Kirovograd National Technical University*], Available at : https://core.ac.uk/reader/81588500/ (Accessed : 5.10.2024).
7. *Cambridge Dictionary,* available at : Available at : https://dictionary.cambridge.org/uk/dictionary/ (Accessed : 5.10.2024).
8. *DeepL Translate,* available at : https://www.deepl.com/uk/translator (Accessed : 3.10.2024).
9. *Glosbe,* available at : https://uk.glosbe.com/ (Accessed : 2.10.2024)
10. *Glosbe Translate,* available at : https://translate.glosbe.com/ (Accessed : 2.10.2024)
11. *Google Translate,* available at : https://translate.google.com/?hl=uk&sl=auto&tl=en&op=translate/ (Accessed : 4.10.2024).
12. Kudriavtseva, N. (2024), "Decolonising Suržyk", *Continuity and Change in the Ideologies of Mixed Language Use,* Justus Liebig University Giessen, GiZo, DFG, pp. 17–18.
13. *Lingea,* available at : https://www.dict.com/ (Accessed : 4.10.2024)
14. *Oxford Learner's Dictionaries,* available at : https://www.oxfordlearnersdictionaries.com/definition/english/ (Accessed : 5.10.2024).
15. Poibeau, T. (2017), *Machine Translation,* The MIT Press, London, 214 p.

Galyna S. IAROTSKA,
Doctor of Philology (Grand PhD), Professor of the Applied Linguistic Department,
Odesa I. I. Mechnikov National University, 24/26 Frantsuzkyi Blvd., Odesa, 65058, Ukraine, tel.: +380503916600;
e-mail: g.yarotskaya@gmail.com; ORCID ID: https://orcid.org/0000–0003–4456–3852

Danil V. BUSHUIEV,
Student of master's degree of the Department of Applied Linguistics, Odesa I. I. Mechnikov National University;
24/26 Frantsuzkyi Blvd., Odesa, 65058, Ukraine, tel.: +380982442627; e-mail: danilbgm20@gmail.com;
ORCID ID: https://orcid.org/0009–0000–0373–5414

SPECIFIC FEATURES OF IDENTIFYING SPECIAL CHARACTERS AND TERMS BY MACHINE TRANSLATION SYSTEMS

Summary. The **purpose** of the article is to investigate the ability of machine translation systems to recognize special characters and terms based on the results of translating from English to Ukrainian and vice versa to identify problems and establish the causes of their occurrence. The **object** of the research is the results of identifying spelling and terminology features in machine translation of scientific texts in the field of sociolinguistics on such topics as bilingualism and language conflicts translated from English to Ukrainian and vice versa, using three machine translation systems; the **subject** is special characters and terms identified by machine translation systems as the result of translating from English to Ukrainian and vice versa, to identify problems, relevant and irrelevant variants. The research in the scope of translation theory and computational linguistics uses both general scientific **methods** and purely linguistic methods such as: analysis, synthesis, comparative and correlative methods. The **results** of the work are to establish identification problems that occur in machine translation of scientific texts and their analysis. Scientific texts may contain special characters to represent specific features of phonetics or spelling of a foreign language word or term and general terms may have spelling differences in English or Ukrainian. Special characters such as quotation marks, dashes and transliteration symbols may cause difficulties identifying them by systems. The obtained results of the study allow to conclude. The **practical value** is in the possibility of using the results in research in the field of computational linguistics, linguistic modelling and corpus linguistics, translation theory and practice, practical course of translation. The **conclusions** of the research can be seen in the further detailed explanation of the identified problems of special characters and terms, to propose, on the basis of that, ways to optimize machine translation.

Key words: machine translation, artificial intelligence, identifying, computational linguistics, translation theory, special character, term.

Статтю отримано 17.10.2024 р.