

ТЕОРЕТИЧНІ ТА МЕТОДОЛОГІЧНІ ЗАСАДИ ПРОФЕСІЙНОЇ ОСВІТИ

УДК 378.1:004

С.М. АМЕЛІНА,

*доктор педагогічних наук, професор, завідувач кафедри іноземної філології і перекладу
Національного університету біоресурсів і природокористування України*

Р.О. ТАРАСЕНКО,

*кандидат технічних наук, доцент кафедри соціальної педагогіки
та інформаційних технологій в освіті
Національного університету біоресурсів і природокористування України*

ФОРМУВАННЯ ІНФОРМАЦІЙНОЇ КОМПЕТЕНТНОСТІ МАЙБУТНІХ ПЕРЕКЛАДАЧІВ НА ОСНОВІ ВИКОРИСТАННЯ МІЖНАРОДНИХ ІНФОРМАЦІЙНИХ РЕСУРСІВ

У статті розглянуто можливості використання міжнародних інформаційних ресурсів у процесі формування інформаційної компетентності майбутніх перекладачів. Представлено характеристики корпусів і баз даних директоратів Єврокомісії та інших міжнародних організацій. Акцентовано увагу на важливості розглянутих інформаційних ресурсів у професійній діяльності перекладачів на сучасному етапі, який характеризується широким застосуванням новітніх технологій перекладу, що базуються на використанні систем автоматизованого перекладу.

Ключові слова: інформаційні ресурси, бази даних, інформаційна компетентність, перекладач.

Постановка проблеми. Процес професійної підготовки конкурентоспроможних перекладачів у вищих навчальних закладах невід’ємно пов’язаний із формуванням їхньої інформаційної компетентності. Одним із важливих аспектів цього процесу є пошук шляхів ефективного використання джерел якісних термінологічних інформаційних ресурсів з галузевим спрямуванням у навчальному процесі. При цьому орієнтиром мають бути не окремі видання, а бази даних в електронному вигляді, які забезпечують структурування матеріалів, визначення умов та порядку доступу до них, наявність мовного різноманіття.

Аналіз останніх досліджень. Предметом досліджень вітчизняних і зарубіжних науковців була низка питань, пов’язаних з інформаційним забезпеченням професійної підготовки майбутніх перекладачів, а саме: створення медіаінформаційних центрів в університетських кампусах (М. Пучков), підготовка студентів до застосування інформаційних технологій у перекладацькій діяльності (А. Кутузов, С. Попов), класифікація інформаційних ресурсів (С. Талья, П. Ваккарі, Дж. Фрай, П. Вутерс).

Формулювання мети. Метою статті є визначення шляхів використання міжнародних інформаційних ресурсів у процесі формування інформаційної компетентності майбутніх перекладачів.

Виклад основного матеріалу. Нині існує велика кількість інформаційних баз даних, які включають бібліографічні записи, реферати документів, повнотекстові матеріали щодо

сільського господарства і суміжних галузей [2]. Використання цього виду інформаційних ресурсів створює умови для формування складових інформаційної компетентності майбутніх перекладачів аграрної галузі, які базуються на можливості опрацювання сучасної, науково вивіреної, фахово досконалої та стандартизованої термінології. Найбільш відомими і потужними серед таких баз даних є:

– AGRIS – International Information System for the Agricultural Sciences and Technology – міжнародна інформаційна система із сільськогосподарських наук і технологій. Забезпечується національними центрами країн-членів FAO і Координаційним центром AGRIS у Римі з 1975 р. Обсяг записів становить 2,7 млн. Матеріали подані англійською мовою.

– AGORA – Access to Global Online Research in Agriculture – глобальні дослідження в сільському господарстві. Забезпечується Продовольчою сільськогосподарською організацією ООН (FAO). Матеріали подані англійською, арабською, іспанською, французькою мовами.

– AGRICOLA – Agricultural On-Line Access – інформаційна система США із сільського господарства. Забезпечується Національною сільськогосподарською бібліотекою США з 1970 р. Обсяг записів становить 4,8 млн. Матеріали подані англійською мовою.

– CAB Abstracts – The World's Leading Agriculture Database – Міжнародна інформаційна система із сільськогосподарських наук і технологій. Забезпечується Сільськогосподарським бюро Британської Співдружності CAB International з 1972 р. Обсяг записів становить 6,3 млн. Матеріали подані англійською мовою.

– FSTA – Food Science and Technology Abstracts – Міжнародна реферативна система з продовольства і технологій. Забезпечується Міжнародною інформаційною службою з продовольства IFIS з 1969 р. Обсяг записів становить 890 тисяч. Матеріали подані англійською мовою.

– DOAJ – Directory of Open Access Journals – доступ до повних текстів наукових журналів. Забезпечується Lund University Libraries. Матеріали подані англійською мовою.

– FAO Catalogue On-line – Інтерактивний каталог FAO. Забезпечується Меморіальною бібліотекою ім. Девіда Любліна. Матеріали подані англійською, арабською, іспанською, французькою мовами [1].

Важливе значення у процесі підготовки перекладачів для аграрної галузі мають ресурси всесвітньої організації із сільського господарства FAO (Food and Agricultural Organization) при ООН, які включають найновішу інформацію з різних галузей сільського господарства та суміжних наук. Доступ до цих ресурсів став можливий через депозитарні бібліотеки FAO, статус яких вона надає найкращим науковим бібліотекам аграрного спрямування. Зокрема в Україні статус депозитарної бібліотеки FAO надано тільки науковій бібліотеці Національного університету біоресурсів та природокористування України.

На особливу увагу заслуговують бази інформаційних ресурсів, які за способом, формою та форматами подання і зберігання текстових матеріалів максимально адаптовані для безпосереднього їх використання у системах автоматизованого перекладу. Останнім часом на створення комплексу таких баз спрямовує свої зусилля низка структур Єврокомісії (ЄК), зокрема: Європейський парламент (European Parliament), Об'єднаний дослідний центр Генерального директорату Європейської комісії (Joint Research Centre of Directorate General of the European Commission – JRC), Генеральний директорат з перекладу (Directorate-General for Translation – DGT), агенція «Європейський центр з профілактики і контролю захворювань» (agency «European Centre for Disease Prevention and Control» (ECDC)), Генеральний директорат з питань освіти та культури (Directorate General for Education and Culture – DG EAC) [3]. Кожна із цих структур започаткувала власну базу перекладених офіційних матеріалів майже на усіх мовах країн – членів Євросоюзу (ЄС), які позиціонуються як паралельні корпуси або бази паралельних текстів у відповідних файлових форматах. Такі ресурси, як правило, використовуються професійними перекладачами у поєднанні із системами автоматизованого перекладу (CAT-системи) для підвищення швидкості та якості перекладів [4]. При цьому важливо відзначити багатоаспектність їх застосування, що може охоплювати створення баз пам'яті перекладів, екстрагування термінів до термінологічних баз, створення спеціальних словників для реалізації функцій автоматичного введення термінів у процесі перекладу. Незважаючи на різні підходи до побудови таких баз, перекладачі ма-

ють можливість, залежно від потреби, компіювати на їх основі власні двомовні чи багатомовні ресурси, пов'язуючи різні мовні версії одного і того ж документа. У контексті нашого дослідження цінність цих баз полягає ще й у тому, що переважна більшість із них містить великий обсяг матеріалів з аграрної та продовольчої тематики. Підтвердженням значущості зазначених баз для підтримки процесу перекладу буде слугувати їх характеристика, зокрема:

– JRC-Acquis – багатомовний паралельний корпус 22 мовами ЄС, що містить в цілому більше 1 мільярда слів. Цей корпус був першим із започаткованих Єврокомісією у 2006 році. У ньому сконцентровані повнотекстові документи ЄС (директиви, адміністративні документи, декларації, резолюції, угоди, акти тощо) в TEI форматі, які після певних перетворень можуть бути використані у системах автоматизованого перекладу. JRC-Acquis є єдиним з корпусів ЄС, де вирівнювання текстів було здійснено з використанням двох різних типів програмного забезпечення. Це дозволило провести порівняння ефективності вирівнювання і надавати користувачам можливість вибору результатів, які краще відповідають їхнім потребам. Завдяки використанню спеціалізованих програмних засобів користувачі мають змогу створювати двомовні вирівняні корпуси для будь-якої мовної пари.

– DGT-Acquis – багатомовний паралельний корпус 23 мовами ЄС. Він створений у 2012 році і включає в себе матеріали усіх серій «Офіційного журналу Європейського Союзу» (Official Journal), починаючи з 2004 року, які охоплюють публікації на теми, пов'язані із суспільним життям, політикою, економікою, охороною здоров'я, інформаційними технологіями, юриспруденцією, сільським господарством, продовольчою політикою, соціальними питаннями. Записи бази подані у форматі XML Formex-4, і їх кількість тільки англійською мовою становить 98 мільйонів слів.

– DCEP – багатомовний паралельний корпус 23 мовами ЄС і турецькою, що складається з документів Європарламенту, створених у період з 2001 по 2012 роки і опублікованих на офіційному сайті Європарламенту. Корпус створений у 2015 році, є найновішим корпусом Євросоюзу і містить понад 1,3 мільярда слів. DCEP не містить стенографічних звітів пленарних засідань Європейського парламенту, щоб уникнути дублювання з корпусом EuroParl. Повнотекстові документи корпусу можуть бути подані у структурованому або неструктурованому вигляді у форматах XML або SGML. Структуровані документи дозволяють користувачам використовувати для їх обробки такий набір інструментів, що забезпечує найбільшу його ефективність при застосуванні за певних умов. Використання варіантів неструктурованих текстів забезпечує більшу швидкість їх обробки.

– DGT-TM – база пам'яті перекладів, яка містить законодавчі матеріали на 24 мовах ЄС, у тому числі всі договори, правила і директиви, прийняті Європейським Союзом, а також постанови Європейського Суду. Ця база започаткована у 2011 році, але включає зазначені матеріали з 2004 року. Обсяг записів бази перевищує 1 мільярд слів, які представлені у форматі TMX. Такий формат представлення даних забезпечує максимальну ефективність його використання у системах автоматизованого перекладу. При вирівнюванні паралельних тестів, призначених для розміщення у базі DGT-TM, здійснюються спеціальні заходи, що передбачають ігнорування як коротких несуттєвих речень, так і занадто довгих, а також усунення явних невідповідностей. Метою таких заходів є створення TMX файлів, які будуть максимально зручними для перекладачів. При цьому здійснюється багаторівнева лінгвістична перевірка всіх документів висококваліфікованими перекладачами, що спеціалізуються в конкретних предметних галузях. Крім того, якість перекладів контролюється юридичними службами та видавничим управлінням ЄК. Багато уваги приділяється забезпеченню узгодженості термінології, яка здійснюється із залученням представників адміністрації держав – членів ЄС. При цьому перекладачі ЄС працюють у сучасному IT-середовищі, яке надає їм значну кількість спеціалізованих інструментів для оптимізації роботи і забезпечення якості та узгодженості перекладів. Незважаючи на існування певної суб'єктивності в перекладі в цілому, якість перекладених матеріалів у базі DGT-TM відповідає високому рівню. Результатом такого підходу до супроводу бази DGT-TM є найвищий показник завантажень ресурсів цієї бази порівняно з іншими корпусами та базами Євросоюзу.

– ECDC-TM – база пам'яті перекладів, яка містить матеріали 23 мовами ЄС, а також норвезькою та ісландською, основна частина яких стосується тем, що пов'язані з питання-

ми охорони здоров'я, профілактики та контролю захворювань. База створена у 2012 році і містить 320 тисяч записів, що набагато менше ніж обсяг вищезгаданих ресурсів. Але перевага бази ECDC-TM полягає у тому, що вона охоплює такі предметні сфери, матеріали з яких відсутні в інших базах. Записи бази подано у форматі TMX і супроводжуються програмним забезпеченням, що дозволяє формувати двомовні бази пам'яті перекладів для будь-якої мовної пари.

– EAC-TM – база пам'яті перекладів, яка створена з файлів, що містять переклади електронних форм проектів або заявок на фінансування програм, а також форми звітності за двома програмами ЄС: «Life-long Learning» та «Youth in Action Programme». База створена у 2013 році і містить понад 540 тисяч записів, які охоплюють кілька предметних сфер, а саме: освіта, професійна підготовка, культура, молодь та спорт. Особливістю цієї бази є те, що вона містить матеріали більшою кількістю мов, ніж будь-який інший корпус чи база Єврокомісії. Крім прийнятих 22 мов, вона включає в себе документи хорватською, ісландською, норвезькою і турецькою мовами, оскільки ці країни мають право на участь у програмах EAC. Як і у попередній базі, записи подано у форматі TMX, а наявність спеціального програмного забезпечення дозволяє формувати двомовні бази пам'яті перекладів для будь-якої мовної пари.

Враховуючи важливість розглянутих інформаційних ресурсів у професійній діяльності перекладачів на сучасному етапі, який характеризується широким застосуванням новітніх технологій перекладу, що базуються на використанні систем автоматизованого перекладу, варто акцентувати на їх цінності у формуванні інформаційної компетентності майбутніх перекладачів. Зазначені паралельні корпуси та бази пам'яті перекладів, особливо ті, ресурси яких наведено у форматах TMX, доцільно використовувати у навчальному процесі для формування знань та умінь щодо здійснення комплексу операцій з термінологічними масивами, їх обробки та подання у формах, які відповідають сучасним умовам перекладацької діяльності. Для цього доцільно запроваджувати у структуру навчальних планів підготовки майбутніх перекладачів окремі спецкурси, наприклад «Інформаційні технології в перекладацьких проектах», зміст яких має бути спрямований на вивчення зазначених аспектів.

Оскільки одним із важливих видів діяльності перекладача при застосуванні систем автоматизованого перекладу є формування галузевих термінологічних баз шляхом екстрагування термінів із спеціалізованих текстів, то ефективність вивчення цієї теми у процесі викладання спецкурсу може бути пов'язана саме із застосуванням ресурсів паралельних корпусів та баз пам'яті перекладів Євросоюзу. Тому відпрацювання принципово нового підходу до отримання вихідного термінологічного матеріалу засобами SDL MultiTerm Extract доцільно здійснювати з використанням отриманих із баз вирівняних паралельних текстів у форматах TMX. Сутність такого процесу зводиться до автоматичного опрацювання спеціалізованих текстів, з яких за певними алгоритмами вилучається масив термінів, характерних для галузі, до якої належав текст. При цьому терміни безпосередньо додаються до термінологічних баз спеціалізованих форматів. Не маючи прецеденту такої діяльності на попередніх етапах навчання, студенти отримують можливість оцінити ефективність продемонстрованого способу отримання термінологічних ресурсів та запропонованого інструментарію. Для досягнення максимального ефекту при опануванні процесу формування термінологічних баз шляхом вилучення термінології засобами SDL MultiTerm Extract необхідно дотримуватись певних критеріїв відбору текстів, які призначені для застосування їх як індивідуальних завдань, а саме: подання двомовних матеріалів у вигляді вирівняних паралельних текстів у форматі TMX, вузькогалузева спрямованість тематики зазначених текстів, насиченість термінами.

На наступному етапі вивчення спецкурсу може здійснюватися формування умінь з використання функції SDL Trados Studio AutoSuggest. Окреме вивчення цієї функції зумовлене тим, що вона створює умови для суттєвого підвищення ефективності формування контенту цільового тексту у процесі перекладу і полягає в автоматичному пропонуванні системою варіантів записів термінів при їх введенні. Однак для відпрацювання такої функції необхідно створити спеціалізований словник на основі TMX файлів, сумарний обсяг текстових сегментів яких має становити не менше 25 тисяч записів. Досягти такого обсягу бази тільки за

рахунок власноруч створених записів досить складно. Тому з метою налаштування студентів на вивчення та застосування цієї функції SDL Trados Studio їм спочатку варто запропонувати відпрацювати навички отримання уже сформованих TMX файлів. Як уже зазначалося, такі файли є у вільному доступі на порталі Єврокомісії. На основі отриманих TMX файлів із баз DGT-TM, ECDC-TM, EAC-TM з достатньою кількістю записів студенти можуть створити спеціальні словники у форматі BPM та відпрацювати використання функції AutoSuggest у реальних умовах здійснення перекладу і оцінити її ефективність.

Особливе значення у структурі спецкурсу має розділ, у якому безпосередньо опановується процес перекладу засобами CAT-систем з використанням пакета SDL Trados Studio. При відборі вихідного матеріалу для практичних завдань з перекладу необхідно дотримуватися певних критеріїв, зокрема: підбирати тексти, які належать до аграрної галузі (тексти доцільно брати з міжнародних інформаційних баз із сільськогосподарських наук і технологій AGRIS, AGORA тощо); надавати перевагу текстам насиченим таблицями, схемами, діаграмами; відбирати тексти (готувати їх фрагменти), сегменти яких збігаються із сегментами спеціально підготовлених баз пам'яті перекладів на рівні 10% – 15% за чіткими збігом (exact match) та 20% – 25% за нечіткими збігами (fuzzy match). Вибір таких показників чітких та нечітких збігів із базою пам'яті перекладів дозволяє оцінити майбутнім перекладачам дієвість застосування таких баз у процесі перекладу, відпрацювати навички редагування сегментів, переклад яких запропоновано системою SDL Trados Studio частково, а також сконцентрувати свою увагу на самостійному здійсненні перекладу значної частини тексту. Важливим джерелом термінологічного ресурсу для підготовки таких баз пам'яті перекладів можуть слугувати паралельні корпуси або бази паралельних текстів Єврокомісії.

Висновок. Враховуючи важливість матеріалів, акумульованих у міжнародних корпусах і базах даних, та їх придатність для використання у системах автоматизованого перекладу, доцільно застосовувати такі інформаційні ресурси у процесі формування інформаційної компетентності майбутніх перекладачів у вищих навчальних закладах України. Цінними ресурсами для підготовки майбутніх перекладачів для аграрної галузі є, зокрема, бази Єврокомісії (JRC-Acquis, DGT-Acquis, DCEP, DGT-TM, ECDC-TM, EAC-TM) і бази даних із сільськогосподарських наук і технологій (AGRIS, AGORA, AGRICOLA, CAB, FSTA, DOAJ, FAO Catalogue Online). Перспективи подальших наукових розвідок можуть бути пов'язані із пошуком шляхів обробки інформаційних ресурсів зазначених джерел у навчальному процесі.

Список використаних джерел

1. Белорусская сельскохозяйственная библиотека [Электронный ресурс]. – Режим доступа: <http://belal.by/basesr.html>
2. Тарасенко Р.О. Фахово-орієнтовані ресурси сучасного інформаційного простору як джерело інформаційної підтримки професійної діяльності перекладача / Р.О. Тарасенко // Педагогіка формування творчої особистості у вищій і загальноосвітній школах: зб. наук. пр. / [редкол.: Т.І. Сущенко (голов. ред.) та ін.]. – Запоріжжя: КПУ, 2014. – Вип. 37 (90). – С. 399–406.
3. Steinberger R. An overview of the European Union's highly multilingual parallel corpora / R. Steinberger, M. Ebrahim, A. Poulis, M. Carrasco-Benitez, P. Schlüter, M. Przybyszewski, S. Gilbro / Language Resources and Evaluation Journal (LRE). 2014, Volume 48, Issue 4, pp. 679–707.
4. Steinberger R. DGT-TM: A freely Available Translation Memory in 22 Languages / R. Steinberger, A. Eisele, S. Klocek, S. Pilos, P. Schlüter/ Proceedings of the 8th international conference on Language Resources and Evaluation (LREC'2012), Istanbul, 21-27 May 2012. – P. 454–459.

References

1. *Belorusskaia selskokhoziaistvennaia biblioteka* [Belarus Agricultural Library]. Available at: <http://belal.by/basesr.html> (In Russian).
2. Tarasenko, R.O. (2014). *Fakhovo-orientovani resursy suchasnokho informatsiinoho prostoru yak dzherelo informatsiinoi pidtrymky profesiinoi diialnosti perekladacha* [Professionally-oriented resources of modern information space as a source of information support of pro-

fessional translators]. *Pedahohika formuvannia tvorchoi osobystosti u vyshchii i zahalnoosvitnii shkolakh* [Pedagogy formation of creative personality in higher and secondary schools]. Zaporizhzhia, KPU, issue 37 (90), pp. 399-406 (In Ukrainian).

3. Steinberger, R., Ebrahim, M., Poulis, A., Carrasco-Benitez, M., Schlüter, P., Przybyszewski, M., Gilbro, S. (2014). An overview of the European Union's highly multilingual parallel corpora. *Language Resources and Evaluation Journal (LRE)*, vol. 48, issue 4, pp. 679-707.

4. Steinberger, R., Eisele, A., Klocek, S., Pilos, S., Schlüter, P. (2012). DGT-TM: A freely Available Translation Memory in 22 Languages. *Proceedings of the 8th international conference on Language Resources and Evaluation (LREC'2012)*, Istanbul, pp. 454-459.

В статье рассмотрены возможности использования международных информационных ресурсов в процессе формирования информационной компетентности будущих переводчиков. Представлены характеристики корпусов и баз данных директоратов Еврокомиссии и других международных организаций. Акцентируется на важности рассмотренных информационных ресурсов в профессиональной деятельности переводчиков на современном этапе, который характеризуется широким применением новейших технологий перевода, основанных на использовании систем автоматизированного перевода.

Ключевые слова: информационные ресурсы, базы данных, информационная компетентность, переводчик.

The article discusses the possibility of using international information resource in the process of formation of information competence of future translators. The characteristics of corpora and databases of the European Commission Directorates and other international organizations are presented. The attention is paid to the importance of the reviewed information resources in professional activities of translators at the present stage, which is characterized by the extensive use of the latest translation technologies based on computer-assisted translation systems.

Key words: information resources, databases, information competence, translator.

Одержано 25.10.2016.