

ІНТЕГРАЦІЯ СИСТЕМ АВТОМАТИЧНОГО РОЗПІЗНАВАННЯ МОВЛЕННЯ (ASR) ТА НЕЙРОННОГО ПЕРЕКЛАДУ В ОНЛАЙН-УСНОМУ ПЕРЕКЛАДІ: ВИКЛИКИ ТА ПЕРСПЕКТИВИ

Олександр Лисиченко

*аспірант кафедри теорії і практики перекладу з англійської мови
Київський національний лінгвістичний університет
м. Київ, Україна*

Стрімке зростання обсягів мультимедіального контенту та інтенсивний розвиток технологій штучного інтелекту зумовлюють підвищену потребу в автоматизованих системах онлайн-перекладу усного мовлення. Технології автоматичного розпізнавання мовлення (ASR) дозволяють здійснювати трансформацію аудіосигналу в текстову форму, яка в подальшому підлягає обробці за допомогою систем нейронного машинного перекладу (NMT). Синергія цих компонентів створює передумови для реалізації майже миттєвого багатомовного усного перекладу. Сьогодні системи ASR активно застосовуються у віртуальних помічниках, автоматизованих сервісах підтримки користувачів, а також для створення транскрипцій освітніх матеріалів [1]. У свою чергу, NMT-системи демонструють високі показники точності перекладу у різноманітних мовних парах. Інтеграція ASR і NMT у режимі реального часу відкриває широкі перспективи для розвитку онлайн-сервісів усного перекладу, водночас висуваючи низку складних технічних завдань. Проаналізовано теоретичні засади такої інтеграції, сучасний стан відповідних технологій і платформ, окреслено ключові проблеми та визначено перспективні напрями подальшого розвитку.

Архітектурно інтеграцію ASR і NMT зазвичай здійснюють у каскадному режимі: спочатку ASR-система розпізнає живий мовний сигнал і генерує текст, потім NMT-модель перекладає цей текст іноземною мовою. У такому «каскадному» підході підвищується складність системи та є ефект накопичення помилок (помилка ASR посилює помилку перекладу). Альтернативою є прямий (end-to-end) підхід, коли єдина нейронна модель без проміжної структури формує переклад на основі акустичних характеристик вхідного сигналу [2]. Пряма модель має меншу архітектурну складність і меншу втрату інформації, але потребує великих корпусів (large corpora) мовлення з

розміченим перекладом. Сучасні ASR-системи використовують глибині нейромережі (як приклад, Conformer, Wav2Vec 2.0 тощо), а NMT – трансформери з автоагресивним декодером. Інтегровані моделі, зокрема SeamusM4T та Whisper, здатні виконувати декілька задач одночасно — автоматичне розпізнавання мовлення, машинний переклад і генерацію усного мовлення (Github, 2025). Якість роботи нейронних систем машинного перекладу (NMT) істотно залежить від якості вхідного тексту: наявність орфографічних помилок, відсутність пунктуації або інші дефекти в транскриптах, сформованих системами ASR, можуть суттєво знижувати точність перекладу. У зв'язку з цим однією з актуальних проблем є підвищення стійкості моделей NMT до «шумних» вхідних даних. Незважаючи на це, каскадна архітектура, яка передбачає послідовну обробку мовлення через ASR і NMT, залишається домінантною у промислових застосуваннях. Водночас у наукових дослідженнях активно розробляються прямі (end-to-end) моделі та їх гібридні варіанти, орієнтовані на реалізацію синхронного перекладу.

Серед відкритих рішень асоціюються моделі OpenAI Whisper та її розвиток WhisperX – багатомовні нейронні ASR, навчені на великих корпусах розмовної мови [3]. Whisper є загального призначення: вона виконує як транскрипцію мовлення, так і автоматичний переклад, адже має внутрішній модуль перекладу. Це дозволяє використовувати Whisper у каскаді ASR+NMT чи навіть як єдину модель для деяких мовних задач. Ще одним перспективним інструментом є модель SeamlessM4T від Meta: мультимодальний транслятор, що об'єднує ASR, переклад і синтез мовлення для майже 100 мов [4]. SeamlessM4T підтримує розпізнавання мови, переклад мовлення в текст, переклад мовлення в мовлення і текстовий переклад, суттєво поліпшуючи узгодженість і знижуючи затримки у порівнянні з окремими етапами. На рівні дослідницьких розробок запропоновано моделі neural transducers для одночасного ASR і перекладу: наприклад, LAMASSU [7] демонструє єдиний потоковий трансдюсер, який одночасно виконує розпізнавання та переклад, стискаючи розмір моделі без втрати якості.

В онлайн-усному перекладі виникає кілька основних проблем. По-перше, фонетичні спотворення та варіанти вимови можуть призводити до хибних транскрипцій. Наприклад, помилкове розпізнавання подібно звучних слів (омофонів) чи втрати кінцевих

звуків типу редукції в українській або недодувних приголосних у англійській зменшують якість ASR. По-друге, шумове забруднення (фон, гул, перешкоди у записі) істотно погіршує розпізнавання. По-третє, акценти та діалекти: якщо ASR-модель навчена на стандартній вимові, сильний регіональний акцент знижує точність. Четвертою проблемою є затримки обробки. Забезпечення онлайн-перекладу в режимі реального часу вимагає мінімізації затримок у передачі результату: як системи розпізнавання мовлення, так і нейронного машинного перекладу повинні функціонувати з високою швидкістю обробки. Проте прагнення скоротити затримку шляхом передчасного запуску перекладу — за відсутності повного мовного контексту — часто супроводжується погіршенням якості результату [5]. Крім того, однією з ключових проблем залишається обмежена доступність релевантних даних: для багатьох мов, доменів і комунікативних ситуацій бракує великих корпусів пар «мовлення–переклад» у режимі реального часу. За відсутності таких спеціалізованих наборів даних моделі демонструють низьку здатність до адаптації у задачах прямої міжмовної інтерпретації.

Інтегровані ASR+NMT-системи мають значний потенціал у реальному часі. В освіті вони можуть стати важливими інструментами для мультимовної освіти: автоматично субтитрувати лекції іноземною мовою або допомагати студентам розуміти викладача їхньою рідною мовою. Розширення доступності навчання, особливо у великих університетах, потребує саме таких технологій. В науковій співпраці на міжнародних нарадах і симпозіумах система автоматичного перекладу може дозволити дослідникам дивитися доповіді інших мов без затримки, полегшуючи обмін знаннями. В дипломатії вже є спроби застосувати автоматичний переклад для багатомовних дебатів. Зокрема, потужні каскадні системи уже тестувались на реальних дебатах Європарламенту [6]. Це демонструє, що штучний інтелект може підсилити чи доповнити живих синхронних перекладачів у політичних чи ділових дискусіях. AI-переклад дозволяє багатомовним онлайн-заходам працювати без перекладачів-посередників — наприклад, платформи (KUDO, Interprefy тощо) вже використовують механізми ASR+NMT для створення субтитрів і голосових перекладів у реальному часі. Загалом, система SimulST спрямована на забезпечення «швидкого і

плавного перекладу» у реальному часі, що особливо корисно для розмов, відеоконференцій і швидкої комунікації.

Таким чином, інтегровані системи, що поєднують автоматичне розпізнавання мовлення (ASR) та нейронний машинний переклад (NMT), становлять перспективний напрям у подоланні мовних бар'єрів у сфері синхронного онлайн-перекладу усного мовлення. Результати сучасних досліджень свідчать про те, що каскадні архітектури ASR+NMT вже досягають рівня якості, порівнюваного з прямими (end-to-end) моделями мовлення-перекладу. Водночас зберігається низка критичних викликів, зокрема: необхідність підвищення робастності моделей до акустичних шумів, варіативних акцентів і спотворень сигналу; оптимізація архітектур із метою зменшення латентності; а також розширення багатомовних корпусів для ефективного навчання моделей. Подальший поступ у цій галузі пов'язаний з інтеграцією великих мовних моделей (LLM) для постредагування помилок ASR, застосуванням адаптивних стратегій інференсу та навчанням на мультизадачних наборах даних. Очікується, що такі системи матимуть усе ширше практичне застосування в освітніх, наукових, дипломатичних і бізнес-контекстах, забезпечуючи оперативну та якісну багатомовну комунікацію в режимі реального часу.

Список використаних джерел

1. *Benchmarking open source and paid services for speech to text: an analysis of quality and input variety* / A. Ferraro et al. *Frontiers in Big Data*. 2023. Vol. 6. DOI: 10.3389/fdata.2023.1210559
2. Bentivogli L., Cettolo M., Gaido M., Karakanta A., Martinelli A., Negri M., Turchi M. *Cascade versus direct speech translation: Do the differences still make a difference?* // *ACL-IJCNLP 2021: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, 2021. С. 2873–2887. DOI: 10.18653/v1/2021.acl-long.224
3. *GitHub – openai/whisper: Robust Speech Recognition via Large-Scale Weak Supervision*. GitHub. URL: <https://github.com/openai/whisper>
4. *Introducing SeamlessM4T, a Multimodal AI Model for Speech and Text Translations*. Meta Newsroom. URL: <https://about.fb.com/news/2023/08/seamlessm4t-ai-translation-model/#:~:text=Today,%20we're%20introducing%20SeamlessM4T,%20the,SeamlessM4T%20supports>
5. Liu X., Hu G., Du Y., He E., Luo Y. F., Xu C., Xiao T., Zhu J. *Recent Advances in End-to-End Simultaneous Speech Translation* // *Proceedings of the 33rd*

International Joint Conference on Artificial Intelligence (IJCAI-24) : Survey Track. 2024. C. 8136–8144. URL: <https://www.ijcai.org/proceedings/2024/0900.pdf>

6. Pérez-González-de-Martos A., Iranzo-Sánchez J., Giménez Pastor A., Jorge J., Silvestre-Cerdà J.-A., Civera J., Sanchis A., Juan A. *Towards Simultaneous Machine Interpretation // Interspeech 2021: Proceedings of the Annual Conference of the International Speech Communication Association. 2021. C. 2277–2281. DOI: 10.21437/Interspeech.2021-201*

7. Wang P., Sun E., Xue J., Wu Y., Zhou L., Gaur Y., Liu S., Li J. *LAMASSU: A Streaming Language-Agnostic Multilingual Speech Recognition and Translation Model Using Neural Transducers // Interspeech 2023: Proceedings of the Annual Conference of the International Speech Communication Association. 2023. C. 57–61. DOI: 10.21437/Interspeech.2023-2004*

LINGUISTYLISTIC DIFFICULTIES OF SIMULTANEOUS INTERPRETING ON THE EXAMPLE OF A PRESS CONFERENCE

Юрій Мельничук

асистент кафедри лінгвістики та перекладу

*Чернівецький національний університет ім. Юрія Федьковича
м. Чернівці, Україна*

Simultaneous interpretation is a form of oral translation that takes place nearly at the same time as the speaker's delivery. In this process, the interpreter listens to the speaker while concurrently conveying the message in another language for the audience. This method is commonly used at international conferences, meetings, press briefings, and other public events where participants speak different languages.

Simultaneous interpretation has several defining features. One of the key characteristics is simultaneity: the interpreter translates almost in real time, processing and analyzing the speaker's message while delivering it in another language. This process requires the use of specialized equipment, such as headphones, microphones, and interpretation booths, which help improve sound clarity and support the interpreter's performance [2].

Another important aspect is the interaction between the interpreter and the speaker. During the speech, interpreters may use gestures and non-verbal cues to maintain effective communication. The work also demands a high level of concentration, mental endurance, and advanced language skills. Because of the fast pace of the speaker's delivery, interpreters often