

ВИКОРИСТАННЯ СЕРЕДОВИЩА «WHISPER» ДЛЯ ТРАНСКРИБУВАННЯ ДІАЛЕКТНОГО МОВЛЕННЯ (на матеріалі наддністрянського говору)

У роботі проаналізовано використання ASR-моделі «Whisper» на матеріалі діалектного мовлення. Проведено оцінку ефективності роботи та проблем, які постають перед програмою в ході опрацювання запису. Акцент зроблено на фонетичних, морфологічних, а також лексичних особливостях транскрибованого тексту. З'ясовано перспективи використання ASR-інструментів для подальших діалектологічних досліджень.

Ключові слова: «Whisper», ASR-інструменти, діалектне мовлення, діалектизми, наддністрянський говір, штучний інтелект (ШІ), транскрибування.

The research justifies using the ASR model “Whisper” on dialect speech material. An assessment of the efficiency of the work and the problems that the program faces during the processing of the recording was carried out. The emphasis is on the phonetic, morphological and lexical features of the transcribed text. Trends in the development of using ASR tools to future dialectology research.

Key words: “Whisper”, ASR tools, dialect speech, dialectisms, naddnistri-anskyi dialect, artificial intelligence (AI), transcription.

З розвитком технологій штучного інтелекту (AI) перед дослідниками в різних сферах відкрилося нове вікно можливостей, винятком не стала і діалектологія. Одним із векторів засто-

сування AI у цій сфері стало використання автоматичного розпізнавання мовлення (ASR) для транскрибування діалектного матеріалу з метою швидкого та ефективного аналізу даних і в перспективі — створення корпусу українських діалектів.

Автоматичне розпізнавання мовлення — це сфера, де поєднуються досягнення комп'ютерного зору (CV), оброблення природної мови (NLP) та специфічні аспекти аудіооброблення, як-от зниження рівня шуму, покращення якості звуку, виокремлення мовних сигналів у складних акустичних середовищах [1, с. 39]. В останнє десятиліття з поширенням глибоких нейронних мереж (DNN) ASR-інструменти почали «моделювати складні зв'язки між аудіосигналами та текстом, вивчаючи важливі характеристики на різних рівнях» [1, с. 40]. Це дозволило вивести якість транскрибування на цілком новий рівень, як наслідок, ми маємо на ринку десятки інструментів, доступних для різних типів задач. Одним із найефективніших на сьогодні залишається «Whisper».

«Whisper» — це «модель для розпізнавання мовлення, розроблена компанією OpenAI і випущена у 2022 році. Модель вирізняється високою точністю розпізнавання мовлення навіть у складних акустичних умовах, таких як наявність шуму або низька якість запису [1, с. 41]. Модель «Whisper» була навчена на «680 000 годин аудіо, що становить одну з найбільших колекцій даних для навчання моделі розпізнавання мовлення» [1, с. 42].

Однак використання навіть провідних моделей породжує низку труднощів у їхньому застосуванні. Ключовими проблемами у роботі з діалектним мовленням є:

- відсутність механізму code-switching, тобто програма не може перемикатися між кількома мовами і транскрибує текст у рамках однієї лексичної системи;
- нівелювання фонетичних особливостей певного говору, програма коригує їх відповідно до літературних норм;
- насамкінець — складність у розпізнанні власне діалектизмів.

Для детальнішого ознайомлення з особливостями використання ASR-інструментів було проведено аналіз транскрипції діалектологічного запису, зробленого студенткою НаУКМА Новицькою Лідією у 2011 р. на території с. Гайок, Бережанського району Тернопільської області, що належить до наддністрянського говору. Вибір запису мотивовано тим, що цей говір «об'єднує багато говірок з вузьколокал. особливостями на різних структур. рівнях, що створює його внутр. діал. розмаїтість; водночас усім цим говіркам притаманний ряд важливих спільних рис» [ЕУМ].

Найхарактерніше роботу середовища «Whisper» демонструють фонетичні особливості говору (перше слово у парі – вимовлене мовицею, друге – транскрибоване «Whisper'ом», тут і далі):

- 1) сильне «укання», що проявляється у зближенні ненаголошеного [о] з [у] (кúрова – корова, курові – корови);
- 2) ненаголошене [а] після м'яких приголосних часто переходить в [е]:
 - а. подівлю́се – подивлю́ся, вді́алосе – вдало́ся, господа́їне – господині́;
 - б. мані́вце́ма – мані́вце́ма, теле́їта – теле́та, свє́та – свєта́;
 - с. в'є́за́ла – в'ї́зала;
- 3) наголошене [а] в [о]:
 - а. пома́го́в – пома́гав, сказа́́в – сказа́в, ї́хов – ї́хав;
 - б. вті́ко́в – вті́ков;
- 4) звук [р'] звичайно переходить у твердий [р] (госпо́дара – господа́ру, буракі́ – бураки́, ліхтаро́х – ліхтара́х);
- 5) приголосні [ц], [с] у суфіксах -ск-, -цк- зазвичай тверді:
 - а. ні́мец – ні́мець;
 - б. че́ска – че́ски;
- 6) палатальні [н'], [д'], [т'] у кінці складу перед наступним приголосним часто переходять в [й] (ші́сна́йцєть – 16);
- 7) широке вживання вставного [н] після губних та в інших категоріях слів (памне́таю – пам'ята́ю, запамне́тало – запам'ята́ло);

- 8) відсутність заміни етимологічного [о] на [а] (горі́чов водóв – горячу воду);
- 9) наявність палатального [л] (ля́мпу – лямпу);
- 10) функціональне посилення фонем [ф]:
 - a. фі́рман – фірмен, фі́ру – фіру;
 - b. фóсах – хосах, фóсі – хосі;
- 11) перехід [а] в [е], [і], [и] (почиткú – початку);
- 12) тверда вимова звука [і] (тýї – ті, ре́крутьв – рекрутив);
- 13) наявність протетичних приголосних (довуко́ла – до Укола).

Вагомою частиною також є морфологічні та синтаксичні риси. Найпоширенішими з них є вживання препозитивної частки *ся*, у якій фонема [а], відповідно до особливостей мови, перейшла в [е]: *ше се боя́ла – ще сі боя́лася; всьо се лиши́ло – все сі лиши́ло; тай се подиві́в – та сі подивив*. А також закінчення -ов, -ев орудного відмінка однини жіночого роду: *ба́бов – бабом; з сестро́в – з сестрою; з дру́гов дівчинов – з другою дівчиною*. Збереження програмою автентичного закінчення відбулося лише в одному випадку: *мало́в дити́нов – з малов дитину*. Останньою характерною рисою мови, що послідовно трапляється в матеріалі є збереження кінцевого твердого [т] у формах 3-ї особи одн. теперішнього часу дієслів II дієвідміни, коли наголос падає на закінчення: *стаю́т – стаю́ть; перу́т – перу́ть*.

Не можна також обійти увагою роботу «Whisper а» з афіксами, яка також є доволі мінливою:

- 1) префіксальні морфеми (зачі́ли – почали, відобра́ти – відобрати);
- 2) суфіксальні морфеми (вті́кнути – втікнути);
- 3) флексійні морфеми (пішла́ зі сві́та – пішла зі світу, доро́гів – дорogів).

Слів-речень, себто вигуків у матеріалі не так і багато, однак варто зазначити, що всі вони передані доволі коректно: *футь – футь; трас, трас, трас, трас, трас, трас*. Числівники, натомість уніфіковано відповідно до стандартів літературної

мови: *три-штíри* – *три-чотири*; *едén* – *один*. Інші – записані числами.

Що ж до синтаксичних ознак, то вони здебільшого опрацьовані непослідовно, зокрема архаїчне вираження майбутнього часу в І ос. мн.: як *будém варíли* – як *будемо варити*.

Особливу увагу варто звернути на діалектизми та іншу просторічну лексику. Їхнє транскрибування, як правило, відбувається за фонетичним принципом, ото ж їх можна поділити на дві великі групи:

- 1) відтворені достоту: *кубéльку* – *кубельку*; *зла́пали* – *злапали*; *німа́к* – *німак*; *німака́ забíли* – *німака забили*; *суматосі* – *суматосі*; *віділи* – *виділи*; *відів* – *видів*; *відите* – *видите*; *пра́ла* – *прала*; *праникі* – *праники*; *шмарка́та* – *шмарката*; *сутерíнах* – *сутеринах*;
- 2) відтворені неточно: *та́мкай барабо́ль* – *там кайбарабо́ль*; *ту́ткай* – *тут, кай*; *зачну́* – *почну*, *німакі* – *німецькі*, *гармідер* – *гармеди*; *ту́ю* – *цю*; *жие́* – *вже є*; *бамбéтель* – *бомбетель*; *тро́щі* – *трошки*; *це́бер* – *цибер*.

Найчастотніші помилки в роботі з цими типами лексики: нерозрізнення голосних фонем (*п'еду* – *п'яду*; *сатíна* – *сатіна*), намагання підібрати фонетично близькі відповідники з літературної мови (*тро́щі* – *трошки*), неврахування пауз у мовленні (*та́мкай барабо́ль* – *там кайбарабо́ль*), або ж навпаки – хаотичне розділення (*ту́ткай* – *тут, кай*; *та́мкай* – *там, кай*).

Детальний аналіз роботи ASR-інструментів, зокрема на формах мовлення відмінних від «чистого» літературного дозволить у майбутньому вдосконалити їх і, відповідно, розширити поле використання. Такі середовища, як «Whisper», уже допомагають у створенні анотованих корпусів діалектного мовлення, тож розуміння механізмів та принципів їхньої роботи не лише на технічному, а й на мовознавчому рівні (зокрема на прикладі різних говірок) допоможе прискорити процес опрацювання записів і зробить їх доступними для широкого кола дослідників.

Список використаних джерел

1. Мрозек Є. Р. Аналіз сучасних підходів до розв'язання задач розпізнавання мовлення. *Системи керування та комп'ютери*. 2024. № 4. С. 39–49.
2. Українська мова : енциклопедія / НАН України, Ін-т мовознав. ім. О. О. Потебні, Ін-т укр. мови ; редкол.: В. М. Русанівський [та ін.]. Вид. 2-ге, випр. і допов. Київ : Вид-во «Українська енциклопедія» ім. М. П. Бажана, 2004. 820 с.
3. Introducing Whisper. *OpenAI*. URL: <https://openai.com/research/whisper> (дата звернення: 28.05.2025).