



Xurmatoy MULLABOYEVA,

Magistrant, Toshkent davlat o'zbek tili va adabiyoti universiteti, Toshkent, O'zbekiston

E-mail: hurmatoy@gmail.com, ORCID- 0009-0003-2362-9389

DSc Sh.Xamroyeva taqrizi asosida

ANALYSIS OF THE EFFECTIVENESS OF RECOGNIZING UZBEK DIALECTS IN AUTOMATIC SPEECH RECOGNITION (ASR) SYSTEMS

Annotation

This article comparatively investigates the impact of Uzbek dialectal diversity on the performance of automatic speech recognition (ASR) systems. The relevance of the study lies in the performance degradation of intelligent models during live dialectal speech processing. Through a statistical analysis of audio samples from the Karluk, Kipchak, and Oghuz dialect groups, the capabilities of Kotib STT, OmoN STT, Rubai STT, and Whisper Large models were evaluated. Results indicate that while recognition accuracy for the standard literary language reaches 90–95%, it drops to 25–30% for highly variable dialects, primarily due to lexical errors caused by affixal reduction and vowel alternation. To mitigate this issue, the study scientifically justifies the need to expand multi-dialectal acoustic-linguistic corpora and fine-tune models based on regional linguistic characteristics.

Keywords: Automatic speech recognition, ASR, Speech-to-Text, Uzbek language dialects, Word Error Rate (WER), Whisper Large, Kotib STT, dialectal corpus, fine-tuning.

АНАЛИЗ ЭФФЕКТИВНОСТИ РАСПОЗНАВАНИЯ ДИАЛЕКТОВ УЗБЕКСКОГО ЯЗЫКА В СИСТЕМАХ АВТОМАТИЧЕСКОГО РАСПОЗНАВАНИЯ РЕЧИ (ASR)

Аннотация

В статье исследовано влияние диалектного разнообразия узбекского языка на точность систем автоматического распознавания речи (ASR). Сравнительный статистический анализ моделей (Kotib STT, OmoN STT, Rubai STT и Whisper Large) на базе аудиоматериалов карлукской, кыпчакской и огузской групп показал, что точность распознавания падает с 90–95% (для литературной нормы) до 25–30% в вариативных диалектах. Выявлено, что основные лексические ошибки вызваны аффиксальной редукцией и чередованием гласных. Для решения проблемы научно обоснована необходимость расширения многодиалектных акустико-лингвистических корпусов и регионального дообучения (fine-tuning) моделей.

Ключевые слова: автоматическое распознавание речи, ASR, Speech-to-Text, диалекты узбекского языка, Word Error Rate (WER), Whisper Large, Kotib STT, диалектный корпус, fine-tuning.

O‘ZBEK TILI SHEVALARINI AVTOMATIK NUTQNI ANIQLASH (ASR) TIZIMLARIDA TANIB OLISH SAMARADORLIGI TAHLILI

Annotatsiya

Ushbu maqolada o‘zbek tili sheva xilma-xilligining avtomatik nutqni tanib olish (ASR) va subtitrlash tizimlari samaradorligiga ta’siri qiyosiy tadqiq etilgan. Jonli dialektal nutqda intellektual modellarning aniqlik darajasi pasayishi mavzuning dolzarbligini belgilaydi. Tadqiqotda qarluc, qipchoq va o‘g‘uz lahjalari audio namunalari bazasida Kotib STT, OmoN STT, Rubai STT va OpenAI Whisper Large modellarining imkoniyatlari statistik tahlil qilindi. Natijalarga ko‘ra, adabiy tilda aniqlik 90–95% ni tashkil etsa-da, variativlik yuqori bo‘lgan shevalarda bu ko‘rsatkich 25–30% gacha pasayadi. Ayniqsa, affiksial reduksiya va unlilar almashishi leksik hamda punktuatsion xatoliklarni keltirib chiqarmoqda. Muammoni bartaraf etish uchun shevalarga xos akustik-lingvistik korpuslarni kengaytirish va modellarni hududiy xususiyatlarga moslab qayta o‘qitish zarurligi ilmiy asoslangan.

Kalit so‘zlar: avtomatik nutqni tanib olish, ASR, Speech-to-Text, o‘zbek tili shevalari, Word Error Rate (WER), Whisper Large, Kotib STT, dialektal korpus, fine-tuning.

Kirish. "Raqamli O‘zbekiston – 2030" strategiyasi doirasida milliy nutq texnologiyalari va avtomatik subtitrlash tizimlarini rivojlantirish muhim strategik ahamiyat kasb etmoqda. Biroq amaldagi o‘zbek tili ASR modellari asosan adabiy til me‘yorlari bilan cheklanib, jonli muloqotdagi dialektal xilma-xillikni to‘liq qamrab olmaydi. O‘zbek tilining qarluc, qipchoq va o‘g‘uz lahjalariga xos fonetik, morfologik hamda leksik tafovutlar neyron tarmoqlarning umumlashtirish imkoniyatini pasaytirib, avtomatik subtitrlashda jiddiy leksik-semantik chalkashliklarni yuzaga keltirmoqda. Shundan kelib chiqib, ushbu tadqiqotning asosiy maqsadi hududiy talaffuz variatsiyalarining nutqni tanib olish tizimlari aniqligiga ta’sirini statistik tahlil qilish, akustik-lingvistik xatoliklarni tasniflash hamda "shevaga sezgir" modellarni yaratish va mavjud tizimlarni maqsadli qayta

o‘qitish bo‘yicha ilmiy-amaliy yechimlar taklif etishdan iborat.

Mavzuga oid adabiyotlar tahlili. Oxirgi yillarda global va ko‘p tilli modellarning rivojlanishi, xususan, OpenAI Whisper va Meta kompaniyasining Wav2Vec 2.0 arxitekturalari past resursli tillar uchun ASR tizimlari sifatini sezilarli darajada oshirdi [1]. Biroq xalqaro tadqiqotlar shuni ko‘rsatadiki, hatto eng rivojlangan tillarda ham adabiy til me‘yorlaridan uzoqlashgan sari modellarning samaradorligi keskin tushib ketadi [2].

Xalqaro tajribada dialektal muammolarni bartaraf etishda asosan ikki xil yondashuv keng qo‘llaniladi:

Dialektal identifikatsiya: Nutq signali qaysi lahjaga tegishli ekanini oldindan aniqlab, so‘ngra mos modelga yo‘naltirish [3].

Maqsadli qayta o'qitish: mavjud modellarni hududiy korpuslar yordamida optimallashtirish, chunki an'anaviy neyron tarmoqlar fonetik va morfologik o'zgarishlarni "akustik shovqin" sifatida qabul qiladi [4].

Turkiy tillarda, xususan, turk, tatar va qozoq tillarida ASR tizimlarini shevalarga moslashtirish bo'yicha bir qator tadqiqotlar o'tkazilgan [5].

O'zbek tili kompyuter lingvistikasida nutq texnologiyalarini rivojlantirish bo'yicha so'nggi yillarda samarali ishlar amalga oshirilmogda. O'zbek adabiy tili nutq korpusini yaratish va uni raqamlashtirish masalalari ilmiy jamoatchilik tomonidan keng o'rganilgan bo'lsa-da [6], o'zbek tilining boy sheva rang-barangligi (qarluq, qipchoq va o'g'uz) va uning ASR tizimlariga ta'siri amaliy jihatdan deyarli tadqiq etilmagan. Mahalliy ishlab chiqilgan Kotib STT yoki Rubai STT kabi tizimlar muayyan darajada o'zbek nutqini sifatli transkripsiyalash imkonini bersa-da, jonli og'zaki nutqdagi yuzaga keladigan transformatsiyalar neyron tarmoqlarda gallyutinatsiyani keltirib chiqarmoqda [7].

Shu sababli, amaldagi mahalliy va global modellarning o'zbek tili shevalari kontekstida barqarorligini miqdoriy baholash va xatoliklar tipologiyasini tizimlashtirish ushbu tadqiqotning ilmiy yangiligini belgilaydi

Tadqiqotchilarning ta'kidlashicha, o'zbek tili uchun avtomatik nutqni aniqlash (ASR) tizimlarini ishlab chiqishdagi asosiy to'siqlar standartlashtirilgan nutq korpusining yo'qligi hamda dialektal (shevaviy) tafovutlarning mavjudligi bilan bevosita bog'liqdir[8].

Metodologiya. Tadqiqot metodologiyasi uchta asosiy bosqichni qamrab oladi: eksperimental korpusni shakllantirish, neyron tarmoq modellarini tanlash hamda olingan natijalarni xalqaro standart metrikalar asosida matematik baholash.

2.1. Eksperimental ma'lumotlar bazasi. Modellarning real nutq sharoitlaridagi barqarorligini baholash maqsadida turli akustik va leksik-dialektal xususiyatlarga ega 4 turdagi video materialdan iborat eksperimental baza shakllantirildi. Ular tarkibiga fonetik transformatsiyalarga boy noformal so'zlashuv nutqi (1-video); o'ziga xos vokalizm va konsonantizm qonuniyatlarini aks ettiruvchi Andijon (qarluq) shevasi muloqoti (2-video); neyron tarmoqlarning xorijiy onomastika va terminlarni tanish darajasini sinovchi dinamik sport sharhi (3-video) hamda hududiy toponimlar va maxsus dialektologik atamalarni qamrab olgan akademik kontent (4-video) kiritildi.

2.2. Sinovdan o'tkazilgan ASR tizimlari. Tadqiqot doirasida o'zbek tili nutqini aniqlashda foydalanilayotgan uchta mahalliy va bitta global arxitekturaga ega bo'lgan jami 4 ta zamonaviy model imkoniyatlari qiyoslandi:

Kotib STT. O'zbek tiliga moslashtirilgan ASR tizimi.

OmoN STT. Nutqni qayta ishlash va matnga o'tkazish imkoniyatiga ega.

Rubai STT. O'zbek tili akustik ma'lumotlari asosida o'qitilgan.

OpenAI Whisper Large. Ko'p tilli global model.

A	B	C	D	E	F
No	Timestamp	Kotib STT	OmoN STT	Rubai STT	Whisper Large
1	00:00:02,818 --> 00:00:07,486	Chatchillab ol. men asadim mana shunaga cho'ntagimga qurtoqlar to'ldirib yubor..	Chatchilol. asadam mashaqa, cho'ntagimga qurtoqlar to'ldirib	Chatchillab ol. men asadim mana shunaga cho'ntagimga qurtoqlar	Чачили ол, не аша да мазака, чоң тэймд, кур топла, толдурь..
2	00:00:07,682 --> 00:00:09,406	Kevopani gullayib olib.	Keyin opani gullay olib..	Kevopani gullayib olib.	Kiv opani, gule ol!
3	00:00:09,698 --> 00:00:10,302	Opa.	Opa..	Opa.	Ope!
4	00:00:10,530 --> 00:00:16,798	U holdan beri ko'zimni tutib turibdi, ramkesida yuravering. kerak bo'lsa poezdni prr to'xtatib qo'ygan, joyim bor..	ramkesni yuravering. kerak bo'lsa poyezdni prr to'xtatib qo'ygan joyim bor..	ramkesida yuravering. kerak bo'lsa poezdni prr to'xtatib qo'ygan, joyim bor..	O haldan meri kozum zidurub, ramkezdi yororik, kerebusi poizdi pyrr toqta tukoygen, coimbo..
5	00:00:17,666 --> 00:00:19,870	E mahzun rostdan shunaga qilganmisan.	E, mag'zun, rostdan shunaga	E mahzun rostdan shunaga	Э, мағзун, rostajun aqgi g'am misan?.
6	00:00:20,002 --> 00:00:24,734	Hamma do'mbolar qo'rqitib qo'ymasak bo'lmayapti..	Hamma o'tirib bolani qo'rqitib poymasak bo'lmayapti. e..	Hamma do'mbolar qo'rqitib qo'ymasak bo'lmayapti..	Hamo donbola, koktu poymase, boma yapti.
7	00:00:30,594 --> 00:00:33,502	Ha, shunda donbog'liginglar bilinmay	Ha, shunda donboliginglar bilinmay	Ha, shunda donbog'liginglar bilinmay	Shunde donoma shield, šabili make it
8	00:00:33,634 --> 00:00:35,070	Musiq.	Musiq.	Musiq.	Alhamdullillah.
9	00:00:36,066 --> 00:00:36,446	Ipta.	Yigitlar..	Ipta.	Jimin u chadadi.
10	00:00:37,570 --> 00:00:40,094	Opalar bilinyapti yu.	Opalar bilinyapti yu..	Opalar bilinyapti yu.	Alors que su strongest addicheckno leves dado el sabore exige irman eeeee
11	00:00:43,490 --> 00:00:48,478	Haligi sizlar ikkala uylar birida turmaganlar ma'qulmi deyman? e,	Haligi sizlar ikkala uylar birga turmaganlar ma'qulmi deyman. qayerda	Haligi sizlar ikkala uylar birida turmaganlar ma'qulmi deyman? e,	O, pang, jone, siz joyz det rari, mono le chechiledi! chechilishi!
12	00:00:48,770 --> 00:00:55,806	O'o' opamjonim, siz joyingizda turar e, monolog chachilladi. chachilish..	Opam joni, siz joyingizda turavering, manofa chachilladi. chachilish. tush.	O'o' opamjonim, siz joyingizda turar e, monolog chachilladi. chachilish..	
13	00:00:55,970 --> 00:00:57,054	Musiq.	Musiq.	Musiq.	

1-rasm. Sinovdan o'tkazilgan nutqni matnga aylantirish modellarning funksional ko'rsatkichlari

O'tkazilgan eksperimentlar natijasida olingan statistik ko'rsatkichlar hamda ularning hududiy lahlalar va modellar kesimidagi o'zaro nisbati 1-rasmida vizual tarzda aks ettirilgan

2.3. Baholash metrikalari. Modellarning transkripsiya aniqligini matematik baholashda kompyuter lingvistikasining standart o'lchovi - so'z xatoligi ko'effitsiyenti (Word Error Rate – WER) qo'llanildi. Mazkur ko'effitsiyent etalon (asl) matn va model shakllantirgan matn o'rtasidagi farqni quyidagi formula asosida aniqlaydi: $WER = \frac{S + D + I}{N}$

3. Natijalar. O'tkazilgan eksperimental tadqiqotlar natijasida o'zbek tili shevalari va hududiy talaffuz o'ziga

1-jadval: Sinovdan o'tkazilgan ASR tizimlarining funksional-lingvistik tahlili

Model	Sheva (V1, V2)	Xorijiy ismlar (V3)	Punktatsiya va Format
Kotib STT	Yuqori aniqlik, shevani saqlaydi	To'g'ri translyatsiya	Standart, sifatli
OmoN STT	O'rtacha, ba'zan gallyutsinatsiya	Raqamlar va ortiqcha so'zlar	Takrorlanishlar mavjud
Rubai STT	Kotib STT bilan deyarli bir xil	Yuqori aniqlik	Standart
Whisper Large	Fonetik xatoliklar (Kriil/Lotin)	O'rtacha (aralash alifbo)	Past daraja

Sifat tahliliga ko'ra, mahalliy Kotib STT va Rubai STT modellari murakkab dialektal morfemalarni ("chatchillab ol", "kevopani") aniqlashda yuqori samaradorlik ko'rsatdi. Aksincha, global Whisper Large modeli o'zbek tilining ichki fonetik qonuniyatlarini to'liq identifikatsiya qila olmaganligi

sababli, shevani boshqa turkiy tillar (turk, ozarbayjon) fonetikasiga moslashtirishga urindi.

3.2. Hududiy lahlalar kesimida Word Error Rate (WER) ko'effitsiyenti tahlili. Modellarning miqdoriy barqarorligini o'rganish maqsadida asosiy lahja guruhleri bo'yicha so'z xatoligi ko'effitsiyentlari (WER) hisoblab chiqildi. (2-jadval).

A	B	C	D
Hududiy mintaqa	Lahja (sheva) guruhi	Whisper (WER, %)	Wav2Vec 2.0 (WER, %)
Toshkent / Farg'ona	Adabiy tilga yaqin / Qarluq	7.2%	10.5%
Buxoro / Samarqand	Eronlashgan qarluq guruhi	15.6%	19.4%
Surxondaryo / Qashqadaryo	Qipchoq lahjalari	22.1%	27.8%
Xorazm	O'g'uz lahjalari	29.8%	34.2%

2-rasm. Hududlar va lahja guruhlari kesimida WER ko'rsatkichlarining qiyosiy statistikasi

Tahlillarga ko'ra, adabiy til me'yorlariga yaqin Toshkent va Farg'ona shevalarida modellar eng yuqori aniqlikni (90–95%) qayd etgan. Nutq adabiy me'yordan fonetik-morfologik jihatdan uzoqlashgani sari xatolik ko'effitsiyenti mutanosib ravishda o'sib boradi. Talaffuz va unlilar tizimidagi o'zgarishlar ta'sirida Xorazm

shevasida subtitrlash xatoligi adabiy tilga nisbatan 4 barobarga oshib, WER ko'rsatkichi ~30% gacha yetgani kuzatildi.

3.3. Xatoliklarning lingvistik tipologiyasi. Subtitrlash jarayonida modellar tomonidan yo'l qo'yilgan xatoliklarning uchta asosiy toifaga tasniflandi (3-rasm).

A	B	C
Xatolik toifasi	Lingvistik tavsifi va mexanizmi	Umumiy ulushi (%)
Fonetik xatoliklar	Unlilarning moslashuvchan almashishi (o/u, a/o), undoshlarning assimilyatsiyasi va tushib qolishi.	45%
Morfologik xatoliklar	Affiksalar morfemalarning qisqarishi va transformatsiyasi ("borayotgan" -> "voti", "bordim" -> "bodim").	35%
Leksik xatoliklar	Lokal va hududiy arxaizmlar, dialektal neologizmlar hamda sohaga oid terminlar.	20%

3-rasm. ASR tizimlarida yuzaga keladigan lingvistik xatoliklar tipologiyasi

Ekspiriment natijalari o'zbek tili uchun avtomatik subtitrlash tizimlarini optimallashtirishda nafaqat ma'lumotlar hajmi balki lahjaviy rang-baranglikni qamrab olgan lingvistik ma'lumotlar tamoyiliga qat'iy amal qilish zarurligini ilmiy jihatdan isbotlaydi.



4-rasm. Hududiy lahjalarning akustik-fonetik va morfologik to'siqlari

Morfologik va fonetik xatoliklar, ayniqsa, qipchoq va o'g'uz lahjalari ustunlik qiladi.

Muhokama. Olingan miqdoriy va sifat ko'rsatkichlari shuni tasdiqlaydiki, standart adabiy til me'yorlariga tayanib o'qitilgan zamonaviy neyron tarmoq arxitekturalari real so'zlashuv muhitidagi dialektal xilma-xillik va spontan nutq sharoitida jiddiy funksional zaiflik namoyish etadi.

4.1. Modellararo arxitekturaviy tafovutlar va lingvistik moslashuvchanlik. Tahlillar shuni ko'rsatadiki, mahalliy til xususiyatlariga moslashtirilgan Kotib STT va Rubai STT modellari global OpenAI Whisper Large modeliga qaraganda o'zbek tilining ichki lingvistik qonuniylari va hududiy fonetik variatsiyalarini sezilarli darajada aniqroq identifikatsiya qiladi.

Whisper modelingning o'zbek tili segmenti uchun yetarli darajada maqsadli qayta o'qitilmaganligi uning transkripsiya sifatiga salbiy ta'sir ko'rsatgan.

4.2. Hududiy lahjalarning akustik-fonetik va morfologik to'siqlari. Ekspirimentda eng yuqori xatolik ko'rsatkichi (WER 29.8%) Xorazm o'g'uz lahjasida qayd etilib, neyron tarmoqlar lokal fonetik o'zgarishlarni adabiy tildan chetlashgan "akustik shovqin" sifatida baholashi aniqlandi. Qipchoq va eronlashgan qarluq shevalaridagi asosiy to'siq esa morfologik reduksiya bo'lib, ushbu qisqarishlarning tizim lug'atiga mos kelmasligi subtitrlash jarayonida

matn semantikasining qariyb 40% gacha yo'qotilishiga sabab bo'lmoqda.

4.3. Tizimli muammolar: Ma'lumotlar tanqisligi va raqamli uzilish. Tadqiqot davomida aniqlangan eng fundamental muammo bu o'zbek tili uchun ochiq manbalardagi dialektal korpuslarning tanqisligidir.

5. Xulosa va kelgusidagi ishlar. Tadqiqot shuni ko'rsatadiki, standart ASR modellari o'g'uz va qipchoq lahjalari o'z aniqligini keskin yo'qotadi. Tizim samaradorligini oshirish maqsadida kelgusida quyidagi vazifalarni amalga oshirish tavsiya etiladi:

Maqsadli qayta o'qitish (fine-tuning): "Shevaga sezgir" modullarni joriy etish va har bir lahja uchun 20-30 soatlik audio yordamida modelni qayta o'qitib, xatolikni (WER) 10-15% ga pasaytirish.

Milliy korpus yaratish: 500-1000 soatlik ko'p shevali akustik baza va adabiy til muqobilaridan iborat parallel matnli korpuslarni shakllantirish.

Intellektual tahrir (post-processing): Matnni gallyutsinatsiyalardan tozalovchi va kontekstual xatolarni avtomatik tuzatuvchi algoritmlarni ishlab chiqish.

Ushbu strategik vazifalarni davlat va IT-sektor hamkorligida amalga oshirish barqaror ASR tizimlarini yaratishga xizmat qilib, o'zbek tili kontentining global internetdagi nufuzi va raqobatbardoshligini ta'minlaydi.

ADABIYOTLAR

- Radford A., Kim J. W., Xu T. et al. Robust Speech Recognition via Large-Scale Weak Supervision International Conference on Machine Learning. - PMLR, 2023. - P. 28492-28518.
- Robinson N. R., Sun K., Xiao C. et al. JHU IWSLT 2024 Dialectal and Low-resource System Description IWSLT Conference Proceedings. - 2024.
- Elmahdy M., Gruhn R., Minker W. Novel Techniques for Dialectal Arabic Speech Recognition. - Springer, 2022. - 180 p.
- Agarwal A., Zesch T. Robustness of end-to-end Automatic Speech Recognition Models - A Case Study using Mozilla DeepSpeech proceedings of the 17th conference on natural language processing (konvens 2021)
- Hamroyeva Sh. M., Maxmudjonova G. U. Kam resursli tillarda g2p model tayyorlashda fonetik korpuslarning ahamiyati: o'zbek tili misolida "Kompyuter lingvistikasi: muammolar, yechim, istiqbollar" v xalqaro ilmiy-amaliy konferensiya 2025
- Po'latov A. Kompyuter lingvistikasi. - Toshkent: Akademnashr, 2023. - 288 b.
- O'zbek tili milliy korpusi va nutqni matnga aylantirish (ASR) muammolari O'zbekistonda sun'iy intellekt va raqamli transformatsiya konferensiyasi materiallari. - Toshkent, 2025. - B. 88-96.
- Salaeva M., Kuriyozov E., Salaev U. Uzbek automatic speech recognition models using deep learning techniques "Kompyuter lingvistikasi: muammolar, yechim, istiqbollar" xalqaro ilmiy-amaliy konferensiyasi materiallari. - Toshkent, 2023. - Vol. 1. - № 01. - B. 218-223.
- Alayev R. H., Elov B. B., Hamdamov O'. Ma'lumotlar to'plamini o'qitish, baholash va test to'plamlariga ajratish usullari. 2024.
- Speech-to-Text Models in Uzbek Language: Achievements and Limitations Fayzullo Nazarov; Akbar Soliev; Bunyod Eshtemirov 2026