ӘОЖ 681.327.12
ҒТАХР 65.011.56

https://doi.org/10.55452/1998-6688-2025-22-4-155-167

[1,3,4]Амангелді Н.,
PhD, ORCID ID: 0000-0002-4669-9254,
e-mail: amangeldi_n_3@enu.kz
[1,2]*Еримбетова А.,
т.ғ.к., қауымдастырылған профессор, PhD,
ORCID ID: 0000-0002-2013-1513,
*e-mail: aigerian8888@gmail.com
[1,4]Газизова Н.Е.,
магистр, ORCID: 0009-0007-7278-4202,
e-mail: g.nazerke2755@gmail.com
[1,3]Турсынова Н.А.,
магистр, ORCID ID: 0000-0002-1857-9028,
e-mail: tursynova_na@enu.kz
[4]Болатбекқызы К.,
бакалавр, ORCID ID: 0009-0000-7659-4974,
e-mail: kerbezbolatbekkyzy03@gmail.com

[1]Ақпараттық және есептеуіш технологиялар институты, Алматы қ., Қазақстан
[2]Еуразия технологиялық университеті, Алматы қ., Қазақстан
[3]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана қ., Қазақстан
[4]«SignBridge» ЖШС, Астана қ., Қазақстан

ГЛОСАРЛЫҚ ҚАБАТТЫ ҚОЛДАНУ АРҚЫЛЫ ҚАЗАҚ ТІЛІНДЕГІ МӘТІНДІ ФОРМАЛИЗАЦИЯЛАУ

Аңдатпа

Ымдау тілін автоматты өңдеу технологиялары цифрлық трансформация кезеңінде теңсіздікке тап болған есту және сөйлеу қабілетінде шектеуі бар қоғам мүшелерінің өзекті қажеттілігіне айналды. Соңғы жылдары ымдау тілін табиғи тілмен бір деңгейдегі формалды құрылым ретінде қарастырып, оны автоматты жүйелерге бейімдеу мәселесі зерттеушілер назарын ерекше аударып келеді. Табиғи тілдегі ақпаратты ымдау тіліне автоматты аудару міндетін орындау үшін аралық қабат ретінде ымдау тілінің мәтіндік нысаны глостар пайдаланылады. Осы мақсатта аталған зерттеуде табиғи тілді өңдеу әдістері арқылы қазақ тілінің морфологиялық ерекшеліктерін қамтитын қазақ тіліндегі мәтінді ымдау тілінің глостарына түрлендірудің жаңа тәсілі ұсынылады. Атап айтқанда, ByT5-кіші үлгісі негізіндегі Seq2Seq архитектурасы қолданылады. Алынған нәтижелер ымдау тілінің ішкі құрылымын сақтайтын ықшам әрі мағыналық тұрғыдан бай глостар тізбектерін қалыптастыратынын көрсетеді. Глостар тізбегі ымдау тіліндегі қимылдарды жазбаша тілге ұқсас логикалық бірліктер ретінде бейнелейтін интерпретацияланатын аралық қабаттың жұмысын автоматтандыруға мүмкіндік береді. Түрлендірілген глостар тізбегі ымдау тілінің құрылымын сақтай отырып, артық ақпаратты азайтады және сөйлемнің үйлесімділігін арттырады. Демек, ымдау тілінің аватарларын басқаруда тек семантикалық тұрғыдан мәнді бірліктердің қолданылуы есептеу жүктемесін төмендетеді. Қысқа әрі мағыналық тұрғыдан бай глостар – ымдау тілінде қол қимылдарын синтездеудің тиімді ресурсы.

Тірек сөздер: ымдау тілі, глостар тізбегі, табиғи тілді өңдеу, формалдау, Seq2Seq, ByT5.

Кіріспе

Дүниежүзілік денсаулық сақтау ұйымының мәліметтері бойынша, әлемде есту қабілетінде шектеу бар 430 миллионға жуық адам бар, олардың 70 миллионнан астамы ымдау тілін

қолданатын қауымдастықтарға жатады [1]. ЫМ тілдері есту немесе сөйлеу мүмкіндігі шектеулі адамдар үшін ана тілі болып саналады және олардың мәдениеті мен жеке тұлғасының негізін құрайтын толыққанды, дербес тілдік жүйелерді құрайды [2]. Халықаралық ұйымдар ым тіліне қол жетімділік пен есту қабілеті шектеулі адамдардың құқықтарының сақталуы арасында тығыз байланыс бар екенін ерекше атап көрсетеді [2]. Ана тілін толық меңгермеген жағдайда, есту қабілеті шектеулі адамдар қоғамда тең мүмкіндіктерге ие бола алмайды. ЫМ тілін еркін меңгеру және қолдану оларға толыққанды әлеуметтік қатысу мен өмірдің барлық салаларында тең құқықтарды қамтамасыз етеді.

Цифрлық дәуір есту қабілеті шектеулі қауымдастықтар үшін жаңа теңсіздік формаларын тудырып отыр. Көптеген онлайн-контент және цифрлық сервистер еститін пайдаланушыларға бағытталған және негізінен дыбыстық немесе жазбаша түрде ұсынылған, бұл оларды есту қабілеті шектеулі адамдар үшін қол жетімсіз етеді [3]. Тіпті субтитрлер мен мәтіндік транскрипциялар да мәселені толық шеше алмайды: көптеген есту қабілеті шектеулі адамдар үшін жазбаша тіл – ана тілімен тікелей байланысты емес екінші тіл.[3].

Ақпаратқа қол жеткізудің шектеулі болуының негізгі себептерінің бірі – ым тілдерінің лексикалық ресурстарының жеткіліксіздігі. Мысалы, Қытай ым тілінде 902 белгіден тұратын базалық тізім қолданулардың шамамен 77%-ын қамтиды, бұл жиілік сөздігінің ықшам «ядросын» көрсетеді [4]. Ағылшын ым тілінің лексикалық деректер жинағы ASL-LEX шамамен 1000 ASL белгісін қамтиды, ал ауызша тілдерде ондаған мың лексемалар бар [5]. ЫМ тілдерінде иконикалық бірліктердің үлесі ағылшын және испан тілдеріне қарағанда жоғары [6]. Шектеулі сөздік қоры мәтінді түсінуге тікелей әсер етеді: мектеп оқушылары арасында ол жазбаша материалды түсіну деңгейімен корреляцияланады [7], ал есту қабілеті шектеулі студенттердің сөздік қоры мен «дүниені тануы» естуші құрдастарына қарағанда статистикалық тұрғыдан төмен [8]. Бала кезіндегі тілдік депривация ұзақ мерзімді когнитивтік және академиялық салдарға әкелуі мүмкін, бұл ым тіліне ерте қол жеткізудің маңызды екенін көрсетеді [9].

Цифрлық теңсіздік, әсіресе COVID-19 пандемиясы кезінде айқын көрінді: көптеген есту қабілеті шектеулі адамдар қашықтан медициналық кеңестерге және денсаулыққа қатысты сенімді ақпаратқа қол жеткізу кезінде қиындықтарға тап болды, себебі материалдар сирек ым тілінде ұсынылды [10].

Бұл теңсіздіктің себептері лексикографиялық және әлеуметтік-мәдени факторлармен [6–10] қатар техникалық аспектілерге де байланысты. Қазіргі технологиялар жиі есту қабілеті шектеулі пайдаланушылардың қажеттіліктерін ескермей жасалады. Көптеген цифрлық сервистер естуші пайдаланушыларға арналып әзірленеді және ым тілін қолданушы қауымдастықтардың қатысуынсыз жобаланады, бұл ым тілінде сөйлейтін пайдаланушылар үшін олардың практикалық құндылығын төмендетеді [11].

Бұл мәселе әсіресе Text-to-Gloss (TtG) тапсырмаларында айқын көрінеді, себебі бұл тапсырмада мәтінді автоматты түрде глостар тізбегіне аудару үшін үлкен және сәйкестендірілген корпустар (бейне + глостар) қажет болады. Қолданыстағы бастамаларға қарамастан, мұндай ресурстар әлі де шектеулі болып қалуда және негізінен ASL [12–15], азиялық [16–17] және кейбір еуропалық [18–21] ым тілдерінде негізделген, ал көптеген аз ресурсты ым тілдері үшін деректер жиынтықтары жоқ немесе көлемі мен қол жетімділігі бойынша шектеулі.

Ғылыми зерттеулерге шолу

Ымдау тілдерін автоматты түрде өңдеу саласындағы заманауи зерттеулер негізгі екі үлкен бағытта жүргізілуде: глосқа негізделген (gloss-based / TtG) және глоссыз (end-to-end/ gloss-free). Глосқа негізделген әдісте бастапқы мәтін алдымен глос тізбегіне аударылады, содан кейін глостар ымға/мультимедиаға трансляцияланады. Ал глоссыз әдіс аралық аннотациясыз, мәтін мен видео немесе негізгі нүктелерден тұратын кеңістік-уақыттық белгілерді тікелей сәйкестендіруге ұмтылады.

Глосқа негізделген әдістер

TtG-модельдері қолданыстағы NMT (Neural Machine Translation) тәсілдеріне сүйену мүмкіндігінің арқасында мәтінді ым тіліне аудару міндетін жүйелеудің алғашқы сәтті талпыныстарының бірі болды. Глосқа негізделген әдістер тапсырманы екі кезеңге бөлуге мүмкіндік береді: text→gloss және gloss→sign. Мұндай модульдік сипат машиналық аудармадағы тексерілген технологияларды қолдануды жеңілдетеді. Мәселен, Ye et al. (2023) [22] мақсатты домендерде мәтін генерациясы арқылы кері аударманы масштабтауға негізделген бірегей тәсілді ұсынды. Авторлар GPT-2 тілдік моделіне сүйенетін Prompt-based Domain Text Generation әдісін әзірлеп, жасанды домендік мәтіндерді қалыптастырды. Бұл мәтіндер кейіннен «глосс – мәтін» синтетикалық жұптарының көзі ретінде back-translation процедурасында пайдаланылды. Мұндай амал глосқа негізделген аудармада деректердің елеулі тапшылығын еңсеруге мүмкіндік беріп, әртүрлі деректер жиынтықтарында сапаны едәуір жақсартқанын көрсетті. Дегенмен, бұл әдіс модельдің генерациясының дұрыстығына тәуелді болып қала береді және мағынасы жағынан сәйкес келмейтін мәтіннің пайда болу ықтималдылығын жоққа шығармайды, бұл алынған шешімдердің әртүрлі жағдайларға бейімделуін шектеуі мүмкін.

Zhu et al. (2023) [23] тағы бір маңызды үлес қосып, мәтінді глосқа нейрондық машиналық аударудың әдістерін қарастырып, көптілді және аз дерекпен оқытылған жағдайларға баса назар аударды. Зерттеушілер төрт негізгі тәсілді талдады:

1. Деректерді аугментациялау – back және forward-translation қолдану, алдын ала өңдеулерді біріктіру және синтетикалық мысалдарға тегтер енгізу.
2. Ішінара бақыланатын NMT – үлкен моно-корпустардан сөйлемдерді «көшіру» арқылы модельді оқыту.
3. Трансферлік оқыту – үлкен корпустарда алдын ала оқыту жүргізіп, кейін мақсатты параллель деректерде fine-tuning жасау.

Көптілді NMT – бір модельді бір уақытта неміс тілі→глосс және неміс тілі→ағылшын тілі бағыттарында оқыту, мақсатты тілге арналған токендерді пайдалану.

Модельдер қарапайым Transformer архитектурасына негізделген (1 қабатты энкодер, 2 қабатты декодер, сөздік көлемі шамамен 32 000 болатын BPE-токенизация) және неміс тіліндегі DGS корпустарында (PHOENIX-Weather 2014T және Hamburg DGS Corpus) оқытылып, американдық ым тілі бойынша (ASL, NCSLGR, шамамен 1500 сөйлем) тексерілді. Сапаны бағалау BLEU, ChrF метрикалары және ым тілін меңгерген адамдардың бағалауы (Direct Assessment, шкала 0–6) арқылы жүргізілді. Ең жоғары нәтижелер көптілді модельде аугментация қолданылған кезде тіркелді: PHOENIX-та BLEU = 28,5 (+6 базалық модельмен салыстырғанда), DGS-та BLEU = 13,4 (+2). Сондай-ақ, ым тілін пайдаланушылардың орташа бағасы базалық модельге қарағанда айтарлықтай жоғары болды. Зерттеу нәтижелері көрсеткендей, аталған әдістердің кешенді қолданылуы мәтінді глосқа аударудың сапасын елеулі түрде арттырады.

Xu et al. (2022) [24] зерттеуінде басымдық практикалық құралдық аспектілерге берілген: олар глосс сөздігіне арналған автоматтандырылған GlossFinder жүйесін ұсынды. Жүйе 2000-ға жуық белгілерді қамтитын корпусқа сүйеніп, пайдаланушы енгізген глостарды сәйкестендіріп, тиісті түсіндірмесін береді. GlossFinder аударма мәселесін шешпегенімен, мәтін мен глостар арасында көпір орнататын қосымша құрал ретінде маңызды және болашақта машиналық аударма жүйелерінде көмекші компонент ретінде қолданылуы мүмкін. Жүйенің шектеулері ретінде сөздіктің салыстырмалы шағын көлемі мен контекстік өңдеудің болмауы атап өтіледі, бұл жүйені тек ішінара шешім ретінде сипаттайды.

Mohamed et al. (2022) [25] трансформерлерге негізделген тәсілді іске асырып, ASLG-PC12 корпусында мәтін мен глостар арасындағы екі бағытты аударма үшін seq2seq моделін оқытты. Қос бағытта да кіріс деректері алдын ала өңделіп, Transformer энкодеріне беріледі, ал декодер шығыс тізбекті генерациялайды. Энкодер мен декодердің әрбір қабаты self-attention

механизмін және позицияға тәуелсіз толық байланысқан желіні (feed-forward) қамтып, қалдық байланыстар мен нормализацияны қолданады. Декодерде encoder–decoder attention қосылып, self-attention қабаттары болашақ позицияларды бүркемелеу арқылы автоагрессивті шығысты қамтамасыз етеді. Осы тәсіл модельге тізбектегі әртүрлі қашықтықтағы байланыстарды ескеріп, ым тілінің грамматикасы мен контекстін дұрыс өңдеуге мүмкіндік берді. Модельдер BLEU және ROUGE метрикалары бойынша бағаланды. Мәтіннен глоска аударуда ең жоғары көрсеткішті 2-қабатты модельге тиесілі: BLEU-4 ≈ 96,89%, ROUGE ≈ 98,78%. Глостан мәтінге аударуда да 2-қабатты модель ең жақсы нәтиже көрсетті: BLEU-4 ≈ 84,82%, ROUGE ≈ 96,90%. Бұл нәтиже глоска негізделген аудармада терең модельдердің әлеуетін айғақтайды.

Глосарлық көрініс көптеген жылдар бойы мәтінді ым тіліне аудару саласында негізгі тәсіл ретінде қарастырылып келгенімен, оның қолданылуы бірқатар шектеулермен қатар жүреді. Дәл осы мәселені Müller et al. (2023) “Considerations for meaningful sign language machine translation based on glosses” атты еңбегінде жан-жақты талдады [26]. Авторлар глосс деңгейіне негізделген қолданыстағы зерттеулерді жүйелеп, оның артықшылықтары мен әлсіз жақтарын айқын көрсетті.

Зерттеуде глостарды аударма үдерісіндегі аралық буын ретінде пайдаланған ғылыми еңбектер корпусы жүйелі түрде сарапталды. Негізгі назар глостардың машиналық оқыту міндетін жеңілдететініне аударылды, себебі олар ым тіліндегі пайымдауларды жазба тіл құрылымына ұқсас токендер тізбегі түрінде бейнелейді. Бұл тәсіл дәстүрлі нейрондық машиналық аударма (NMT) архитектураларын қолдануға және жақын салалардағы әдістерді оңай бейімдеуге мүмкіндік береді. Алайда Müller және әріптестері осы әдістің бірқатар маңызды шектеулерін ерекше атап өтеді:

- Лингвистикалық толық еместік: глостар тек қимылдардың сызықтық ретін белгілейді және ым тілінің кеңістіктік компоненттерін, вербалды емес құралдарын және көпөлшемділігін жеткізе алмайды.
- Деректердің шектеулілігі: негізгі қолданылатын корпус RWTH-PHOENIX-Weather-2014T көлемі жағынан шағын әрі тақырыптық тұрғыда шектеулі, бұл модельдердің жалпылау қабілетін төмендетеді.
- Зерттеулердің жеткіліксіз ашықтығы: жұмыстардың аз бөлігі ғана глосарлық тәсілдің шектеулерін нақты мойындайды.
- Метрикалардың әртүрлілігі: BLEU, METEOR, COMET және басқа көрсеткіштердің әртүрлі нұсқаларын пайдалану нәтижелерді салыстыруды қиындатады, бұл саладағы прогресті объективті бағалауға кедергі келтіреді.

Müller және әріптестерінің (2023) еңбегі глосарлық әдістердің жеңілдетілген аралық деңгей ретінде белгілі бір практикалық құндылығы бар екенін көрсеткенімен, олардың қазіргі күйінде ым тілінің барлық аспектілерін толық бейнелей алмайтынын көрсетеді. Мұндай тәсілдердің одан әрі дамуы деректердің сапасын арттырумен қатар, бағалау сапасын жетілдіруді де талап етеді.

Глоссыз әдістер

Глосс аннотацияларын қолданбайтын ым тілі аудармасының әдістері саласында соңғы жылдары end-to-end әдістерін дамытуға, деректерді байытуға және бейнеден релевантты визуалды көріністерді алуға бағытталған арнайы алдын ала өңдеу амалдарына негізделген бірқатар зерттеулер пайда болды. Lin et al. (2023) мақаласында Gloss-Free End-to-End Sign Language Translation (GloFE) жүйесі ұсынылған – бұл аралық глосс деңгейінсіз, ым тілін тікелей мәтінге аудару жүйесі [27]. Модель бейнежазбаларды визуалды энкодер (CTR-GCN + Multi-Scale TCN) арқылы өңдеп, скелеттік белгілерді анықтап, оларды трансформерлік энкодерге береді, ал мәтіндік декодер мақсатты тілдегі сөйлемді генерациялайды. Глосс орнын толықтыру үшін conceptual anchors механизмі қолданылады: мақсатты мәтіндегі негізгі сөздер (зат есімдер мен етістіктер) «тірек» қызметін атқарады, ал cross-attention және

контрастивті триплеттік оқыту визуалды белгілерді осы концепттермен сәйкестендіреді. Модель OpenASL және How2Sign деректер жиынтықтарында глоссыз оқытылды, ал WLASL датасетінде бірқалыпты белгілерді тану үшін алдын ала үйретіліп, GloFE озық деңгейлі нәтижелер көрсетті. GloFE әдісі глоссыз ым тілін аударуда тиімді, бірақ бірқатар шектеулер бар. Әдіс деректер көлеміне және таңдалған концепттердің сапасына тәуелді, негізгі сөздерді қолмен таңдау қажет, есептеу қиындығы жоғары, маңызды емес сөздер жіберілуі мүмкін және қимылдардың барлық ерекшеліктері толық есепке алынбайды.

Ye et al. (2023) еңбектерінде end-to-end ым тілін аудару үшін кросс-модальды деректерді аугментациялау әдісін (XmDA) ұсынды, оның мақсаты бейне және мәтін белгілері арасындағы алшақтықты азайту және белгіленген деректердің жетіспеушілігінің орнын толтыру [28]. XmDA екі техниканы біріктіреді: Cross-modality Mix-up (бейне белгілері мен глосс эмбеддингісін араластыра отырып, репрезентацияларды теңестіру) және Cross-modality Knowledge Distillation («глосс→мәтін» модельін мұғалім ретінде пайдаланып мәтін генерациясын үйрету). Негізгі архитектура – трансформерлік SLT (CTC-классификаторымен) – глосс эмбеддингтерін енгізу арқылы кеңейтіліп, аталған техникаларды қолдану арқылы оқытылды. Модель неміс тіліндегі PHOENIX-2014T және қытай тіліндегі CSL-Daily стандартты деректер жиынтықтарында оқытылып, бағаланды. Эксперименттер сапаның айтарлықтай өскенін көрсетті: мысалы, PHOENIX-2014T-де BLEU 22.79-дан 25.36-ға дейін, ал CSL-Daily-де 11.61-ден 21.58-ге дейін өсті; сонымен қатар ROUGE/ChrF метрикалары бойынша да, әсіресе сирек кездесетін сөздер мен ұзақ сөйлемдер үшін жақсарулар байқалды. Аталған әдіс кросс-модальды аугментация арқылы ым тілін аударудың сапасын арттыруда тиімді болғанмен, есептеу жүктемесінің жоғары болуы, алдын ала үйретілген мұғалім-модельдерге тәуелділігі және бастапқы деректердің сапасына сезімталдығы сияқты шектеулер бар.

Shi et al. (2022) доменді кеңейту және «табиғи ортадағы» бейнематериалдарда оқыту мәселесіне арналған зерттеуде OpenASL атты ауқымды корпусты ұсынды [29]. Ол онлайн-платформалардан (YouTube) жиналған және жүздеген сағаттық, түрлі орындаушылар мен тақырыптарды қамтитын материалдардан тұрады. Авторлар мұндай алуан деректер жиынтығында оқыту үшін қолдың, дененің және бет артикуляциясының тұрақты белгілерін шығаратын арнайы архитектуралық шешімдер, сондай-ақ бейне сапасының әркелкілігі мен домендік ығысуларға назар аудару қажеттігін көрсетті. Нәтижесінде тар шеңберлі корпустардан тыс жұмыс істей алатын «ашық» SLT жүйелерін құру мүмкін болды.

Бірқатар еңбектер табысты глосссыз жүйелердің маңызды компоненті ретінде бейне ағынын алдын ала өңдеу мен тығыз нормализациялау тәсілдеріне басым назар аударады. Әсіресе keypoint-based тәсілдер бейнені скелеттік нүктелер векторына ауыстыруды және арнайы дене бөліктеріне бағытталған нормализацияны, кадрларды стохастикалық таңдауды және аугментация процедураларын дамытуды ұсынады. Мұндай әдістер [30] дұрыс нормализация мен негізгі нүктелерді агрегациялау аударма модульдері үшін тиімді көріністер құруға мүмкіндік беретінін, әрі трансформерлік архитектуралармен және генеративті декодерлермен үйлесімділікті сақтайтынын көрсетеді.

Өз кезегінде, Avramidis және Möller (2022) ым тілін аудару тапсырмасы аясында, глосстарды мәтінге аудару үшін нейрондық машиналық аударма (NMT) әдістерін зерттеді [31]. Авторлар неміс ым тілінің екі параллель корпусы – PHOENIX14T және Public DGS Corpus бойынша екі NMT архитектурасы (RNN және Transformer) арқылы тәжірибелер жүргізген. Жұмыс барысында модельдердің гиперпараметрлері, токенизация әдістері (BPE, unigram және арнайы токенизация) оңтайландырылып, деректерді аугментациялаудың екі техникасы қолданылған: кері аударма және парафразалау. Нәтижелер айтарлықтай жақсы нәтижелер көрсетті: сәйкес корпустарда оқытылған модельдер үшін BLEU көрсеткіші 5.0 және 2.2-ге өсті. Сонымен қатар, RNN модельдері Transformer модельдерінен жоғары нәтиже берді, ал сегментация үшін BPE қолдану ең тиімді болып шықты. Дегенмен, әдістің негізгі шектеулері – аннотацияланған деректердің сапасына тәуелділік және қолданыстағы NMT әдістерін ым тілінің ерекшеліктеріне бейімдеудің қиындығы.

Cao et al. (2022) Task-aware Instruction Network (TIN-SLT) тұжырымдамасын және көпдеңгейлі аугментация схемаларын ұсынды [32]. Мұнда модельге арналған «нұсқаулар» мен белгілерді біріктіру тетіктері алдын ала үйретілген трансформерлердің тілдік қабілеттерін барынша пайдалануға бағытталған. Негізгі идея – оқыту барысында модельді міндетке сәйкес динамикалық бағыттап, аугментация деңгейлері арқылы оқыту мысалдарының таралуын өзгерту. Бұл тәсіл шектеулі video→text жұптары жағдайында жалпылау қабілетін арттыруға мүмкіндік береді. Әдістің негізгі шектеуі – модельдің күрделілігінің өсуі және оқыту уақытының ұлғаюы.

Аз ресурты тілдерде жүргізілген практикалық жұмыстар, мысалы, Sousa et al. (2023) еуропалық португал тілінен португал ым тіліне аударуға арналған жұмысы, мақсатты тілдің және жергілікті корпустың ерекшеліктерін ескере отырып, архитектураларды және аугментация әдістерін бейімдеудің қажеттілігін көрсетеді [33]. Оқыту үшін эксперттер тарапынан аннотацияланған үлгілердің алтын стандартына негізделген көпдоменді LGP-5-Domain корпус жасалды. Эксперименттер көрсеткендей, ұсынылған әдіс жалғыз аналог жүйе PE2LGP-дан асып түсіп, BLEU және басқа аударма метрикалары бойынша жоғары көрсеткіштерге жетті. Негізгі шектеу – аннотациялардың сапасына тәуелділік және NMT моделін ым тілінің ерекшеліктеріне бейімдеудің күрделілігі.

Жүргізілген шолу ым тілдерінің машиналық аудармасы саласындағы зерттеулердің екі бағытта дамып отырғанын көрсетеді: біріншісі – глостарды аралық деңгей ретінде қолдану, екіншісі – толық глоссыз әдіс. Екінші бағыт терең оқыту саласындағы прогрес пен ірі мультимодальды модельдердің пайда болуына байланысты танымалдылығы күннен-күнге артып жатса да, дәл қазіргі уақытта глосарлық әдістер тұрақты нәтижелерді көрсетуде. Глостар сөйлем құрылымын формализациялауға, қимылдарды өңдеу кезіндегі көпмағыналықты азайтуға және машиналық аударма архитектуралары үшін модельдеуді жеңілдетуге мүмкіндік береді. Сонымен қатар, глосарлық аннотациялардың болуы модельдерді интерпретациялауға және ашық көрсетуге мүмкіндік береді, бұл жүйелерді білім беру немесе инклюзивті цифрлық шешімдерде қолдану кезінде ерекше маңызды.

Қолжетімді ым тілі корпустарының шектеулі және әркелкі жағдайында глосарлық әдістер деректерді стандарттауға және салыстырмалы эксперименттер жүргізуге мүмкіндік бере отырып, негізгі рөл атқара бермек. Бұл зерттеушілер қауымдастығына әрі қарайғы инновациялар мен салыстырмалы талдаулар үшін берік негіз қалыптастырады. Сондықтан, алдағы перспективада глостар академиялық зерттеулерде де, қолданбалы міндеттерде де маңызды элемент болып қала береді деп қорытындылауға болады, ал одан әрі прогресс, ең алдымен, глосарлық және глоссыз әдістерді интеграциялауға байланысты болады, мұнда глостар сенімді негіз ретінде қызмет атқаратын болады.

Мұндай тұжырымдар глосарлық әдістің қазіргі таңда өзінің маңыздылығын жоғалтпағанын және практикалық қолдануда тұрақты нәтижелер беретінін айқын көрсетеді. Осы контекстте аталған жұмыста глостарды аралық деңгей ретінде пайдалануға негізделген тәсілді таңдалдық. Тек модельдің құрылымын формализациялау ғана емес, сонымен қатар қазақ тілінің морфологиялық ерекшеліктерін ескеріп, ым тіліндегі мағыналық сәйкестікті нақты беруге бағытталған әдіс. Төменде біздің жүйеде қолданылған мәліметтерді алдын ала өңдеу және модельді оқыту кезеңдері егжей-тегжейлі сипатталады.

Материалдар мен әдістер

Модельді оқытуға арналған корпусты дайындау барысында оның лингвистикалық тазартылуына ерекше назар аударылды, өйткені қазақ тілінің құрылымында ым тіліне тікелей баламасы жоқ және модельді оқытуда қиындық тудыруы мүмкін элементтер бар. Мұндай ерекшеліктердің бірі – көмекші етістік формалары, мысалы, жатыр, тұр, отыр және басқа да формалар. Олар жазбаша тілде әрекеттің күйі немесе түрін білдіру үшін қолданылады, бірақ ым тілінде түсіп қалады, өйткені негізгі мағына олардың алдында тұрған негізгі етістікпен

беріледі. Деректерді өңдеу барысында мұндай көмекші сөздер алып тасталып, сөйлемде тек негізгі етістік қалдырылды.

Сондай-ақ ым тілінде қолданылмайтын жалғаулықтар мен шылау сөздер, мысалы, мен, және («и»), сондай-ақ ым тілдік контексте жеке мағыналық жүк артпайтын басқа да ұқсас элементтер де алынып тасталды. Мұндай сүзгілеу кіріс деректерін жеңілдетіп, модельдің негізгі мағыналық бірліктерге назар аударуына мүмкіндік берді, бұл мәтін мен глосс арасындағы сәйкестікті дәлірек меңгеруге ықпал етті.

Деректерді лингвистикалық тазарту бойынша негізгі ережелер 1-кестеде көрсетілген. Онда бастапқы сөйлемдердің және алдын ала өңдеу кезеңдерінен кейінгі түрлендірілген нұсқалардың мысалдары ұсынылған.

Кесте 1 – Корпусты лингвистикалық алдын ала өңдеу ережелерінің мысалдары

№	Алдын ала өңдеу ережесі	Бастапқы мәтін үзіндісі	Нормаланған нұсқасы	Ескерту
1	Көмекші етістіктерді алып тастау (отыр, жатыр, тұр, жүр)	Мен кітап оқып отырмын	Мен кітап оқимын	Ым тілінде шақ пен түр негізгі етістікпен беріледі
2	Шартты бөлшектерді алып тастау (-ақ, -ау, -ай, -да/-де/-та/-те)	Осы-ақ менің досым-ай келді	Осы менің досым келді	Бұл бөлшектер эмоция/күшейтуді білдіреді, ым тілінде мимикамен беріледі
3	-мыс, -міс жалғаулы кіріспелерді алып тастау	Ол айтыпты-мыс	Ол айтыпты	Ым тілінде бұл болжам мағыналары қолданылмайды
4	Постпозициялық формаларды алып тастау (-ды/-ді, -ты/-ті)	Айтқан-ды кеше	Айтқан кеше	Бұл формалар стильдік рөл атқарады, мағыналық негізге әсер етпейді
5	Күшейткіш сын есімдерді бастапқы түрге келтіру	әп-әдемі, жып-жылы, аппақ	әдемі, жылы, ақ	Ым тілінде күшейткіш мағына мимикамен не қимылмен беріледі

Модельді оқыту

Мәтінді ым тіліне аудару жүйесіндегі модельді оқыту кезеңі Seq2Seq архитектурасы және трансформер негізіндегі назар (attention) механизмі арқылы жүзеге асырылды. Бұл тәсіл машиналық аударма мен тізбектерді түрлендіру мәселелерін шешуде жоғары тиімділікті қамтамасыз етеді.

Негізгі модель ретінде ByT5-small қолданылды – бұл T5 моделінің байттық модификациясы, ол көптілді мәтіндермен жұмыс істей алады және сирек кездесетін таңбаларды жеке токен сөздігісіз-ақ тікелей өңдеуге қабілетті. Энкодер-декодер құрылымының арқасында бұл модель табиғи тілдегі сөйлемдерді формализацияланған глоссаларға түрлендіруге қолайлы, ал бұл – ымдарды генерациялауға дейінгі негізгі қадам.

Оқыту және инференс параметрлерінің толық конфигурациясы 2-кестеде ұсынылған.

Кіріс және мақсатты тізбектердің токенизациясы ByT5 ішкі токенизаторы арқылы жүзеге асырылды: ұзындығы 128 токенге дейін қысқартылып, әрі қарай бекітілген өлшемге дейін толықтыру (padding) жүргізілді. Мұндай тәсіл бірдей ұзындықтағы batch-терді қалыптастыруға мүмкіндік беріп, есептеуді тиімдірек етіп, оқыту үдерісін жеделдетті.

Мақсатты глосалар токен идентификаторлары түрінде кодталып, модельге labels параметрі арқылы берілді – бұл жүйеге кіріс және шығыс деректерді тікелей салыстыру арқылы үйренуге мүмкіндік берді.

Оқыту Hugging Face Transformers кітапханасының көмегімен және Seq2SeqTrainer класы арқылы жүргізілді, бұл процесті оңай баптауға және бақылауға жол ашты.

Негізгі гиперпараметрлер:

- Learning rate: 5×10^{-4};
- Batch size: 8 (оқыту және валидация үшін);
- Эпох саны: 10;
- Weight decay: 0.01 (модельдің жалпылау қабілетін арттыру және артық үйренудің алдын алу үшін).

Batch-терді қалыптастыру үшін DataCollatorForSeq2Seq пайдаланылды, ол тізбектердің ұзындығын динамикалық толықтыру арқылы артық паддингті азайтты.

Әр эпох аяқталған соң модельдің сапасы бағаланды. Бұл ретте тек соңғы бақылау нүктесі (checkpoint) сақталды – бұл диск кеңістігін үнемдеуге және аралық нұсқалардың жинақталуын болдырмауға мүмкіндік берді.

Инференс кезеңінде ең ықтимал глосс тізбегін табу үшін beam search әдісі қолданылды, 5 сәуле (beam) параметрімен. Нәтиже сапасын арттыру үшін қосымша шектеулер енгізілді:

n-грамманың қайталануына тыйым (ұзындығы 3);

Қайталануға айыппұл – мәні 1.2;

Ерте тоқтау – толық аудармаға жеткен кезде.

Бұл тәсіл шығыс тізбектеріндегі артықтықты азайтып, байланыстылықты күшейтіп, ым тіліне тән құрылымды сақтауға көмектесті.

Нәтижелер

Оқыту, 1-суретте көрсетілгендей, 10 эпохаға дейін жүргізілді. Негізгі мәтінде тек 1–5-эпохалардың нәтижелері көрсетілді, себебі 6–10-эпохалардың көрсеткіштері қосымша материалға шығарылды. Валидациялық шығын 1–3-эпохаларда тұрақты төмендеп, минималды мәнге 3-эпохада жетті де (≈0.2565), 4–5-эпохаларда қосымша айтарлықтай жақсару байқалмады (≈0.260–0.261).

[1040/1040 16:06, Epoch 10/10]

Epoch	Training Loss	Validation Loss
1	No log	0.303398
2	No log	0.270347
3	No log	0.256496
4	No log	0.261146
5	0.898300	0.260486
6	0.898300	0.268210
7	0.898300	0.266074
8	0.898300	0.280682
9	0.898300	0.287509
10	0.155400	0.294966

Сурет 1 – Оқыту нәтижелері

Осылайша, модельдің ең тиімді күйі валидациялық шығынның ең төменгі мәні тіркелген 3-эпохада анықталды. Тренинг шығынының 1–4-эпохалар аралығында тіркелмеуі, яғни «no log» ретінде белгіленуі себепті, модельді бағалау және таңдау тек валидациялық көрсеткіштерге негізделіп жүзеге асырылды.

Сапалық мысалдар 2-суретте ұсынылғандай, (мәтін→глосс) көмекші етістіктер мен эмоциялық бөлшектердің түсіріліп, мағыналық өзектің сақталатынын көрсетті.

```
текст: Сәлеметсіздер ме, құрметті оқушылар!
глосс: Сәлем, құрмет оқушы!

текст: Бүгінгі сабаққа қош келдіңіздер.
глосс: бүгін сабақ қош келу

текст: Бүгінгі сабағымыздың тақырыбы – Жер ғаламшарының пайда болуы.
глосс: Бүгін сабақ тақырып –Жер ғаламшар пайда болу.

текст: Бүгінгі сабағымыздың мақсаты – Жер ғаламшарының пайда болуын түсіну.
глосс: Бүгін сабақ мақсат – Жер ғалам пайда болу түсіну.

текст: Адамдар ежелден өзін қоршаған әлемді, оның ішінде ең бірінші, барлығымыздың ортақ үйіміз – Жер шарын танып-білуге ұмтылған.
глосс: Адам ежелден өз қоршау әлем – Жер шар тану білу ұмтылу.

текст: Бізге Жердің қалай пайда болғаны туралы әртүрлі ұлттардан көптеген аңгімелер мен мифтер жеткен.
глосс: Біз Жер қалай пайда болды туралы әр түрі ұлт көп аңгіме мифтер болды

текст: Ежелгі адамдар Жер, Күннен және басқа жұлдыздардан да үлкен деп ойлаған.
глосс: Ежелгі адам Жер, Күн басқа жұлдыз үлкен ойлау болу

текст: Адамдарда ғибадатхана мен пирамида салу тәжірибесі болған.
глосс: Адам ғибадатхана пирами салу тәжірибе болды

текст: Алайда Жердің қалай пайда болғаны туралы білім жоққа тең еді.
...
глосс: Құдай Жер адам көңіл шығу бар
```

Сурет 2 – Сапалық мысалдар

Талқылау

Әдеби шолу ым тілін машиналық аудару зерттеулерінің глосты аралық деңгей ретінде қолданатын тәсілдер және глоссыз end-to-end әдістер бағытын көрсетеді. Соңғы жылдары ірі мультимодальды модельдерге сүйенетін глоссыз бағыт танымал болғанымен, қазірше глоссқа негізделген әдістер тұрақты нәтижелер көрсетіп келеді. Бұл зерттеуде таңдаған интерпретацияланатын глосаралық қабат бағытымен үйлеседі.

Глостар сөйлем құрылымын формализациялап, қимылды өңдеу кезіндегі көпмағыналылықты азайтады және классикалық нейрондық машиналық аударма және тізбектен-тізбекке негізделген архитектураларына бейімдеуді жеңілдетеді. Бұған қоса, глосарлық аннотация интерпретациялану мен ашықтықты қамтамасыз етеді, бұл білім беру және инклюзивті цифрлық шешімдер үшін маңызды артықшылық болып саналады.

Lin және әріптестері ұсынған GloFE жүйесі тұжырымдық тірек нүктелер арқылы ымнан тікелей мәтін генерациялайды [27]. Жақсы нәтижелеріне қарамастан, әдіс дерек көлеміне, «тірек» сөздерді таңдаудың сапасына және есептеу ресурстарына тәуелді, яғни кейбір маңызсыз сөздер жоғалуы мүмкін, ал қимылдың барлық ерекшеліктері түгел ескерілмейді.

End-to-end тәсілдерде деректердің жетіспеушілігін азайту мақсатында Ye және әріптестері кросс-модальды аугментация әдісін ұсынды. Бұл тәсіл бейне мен мәтіннің репрезентацияларын бір кеңістікте сәйкестендіріп үйретеді, сондай-ақ алдын ала үйретілген глостан мәтінге түрлендіретін моделді мұғалім ретінде қолданып, негізгі end-to-end модельге қосымша білім береді. Осылайша, модель визуалды және тілдік ақпарат арасындағы байланысты тиімдірек үйрене алады. Бұл тәсіл гибридті бағыт құруға тиімді [28].

Sousa және әріптестерінің өз зерттеуінде мақсатты тіл мен жергілікті корпустың ерекшеліктерін ескеріп, модель құрылымын және деректерді толықтыру тәсілдерін бейімдеу бейімдеу қажет екенін көрсетті [33]. Бұл нәтиже KSL секілді тілдер үшін домендік корпусты өсіру мен сапалы аннотацияны бірінші орынға қою керектігін көрсетеді.

Осы зерттеуді ұсынылған глоссқа негізделген құбыр аз ресурсты жағдайда интерпретацияланатын әрі тұрақты негіз береді. Кеңістіктік және вербалды емес (non-manual) белгілерді ескеретін толыққанды модель қалыптастыру үшін, шолуда ұсынылған end-to-end тәсілдердің жекелеген механизмдерін глосс негізіндегі архитектурамен гибридтік үйлестіру келешегі зор бағыт болады. Осылайша, глосс сенімді өзек ретінде сақталып, end-to-end компоненттер күрделі көрнекі белгілерді меңгеруді күшейтеді.

Қорытынды

Бұл зерттеу seq2seq/byt5-small үлгісінің қазақ тіліндегі мәтінді ымдау тілі глостарына аударудың тиімділігін көрсетеді. Seq2Seq архитектурасы және byt5-кіші үлгісі қазақша сөйлемдерді байт деңгейінде өңдейді, нәтижесінде формальданған глостарға тұрақты түрлендірулер жасалады. Глостау тәсілі ымдау тілінің қозғалыстарын жазбаша тілге ұқсас логикалық бірліктер ретінде көрсететін интерпретацияланатын аралық бағдарлама қабаты ретінде қызмет етеді. Бұл құрылым дәстүрлі нейромашиналық аударма архитектурасын пайдалануға мүмкіндік береді және модель әрекетін талдауды жеңілдетеді.

Модель арқылы жасалған қысқартылған глостар реті ымдау тілін көрнекі көрсететін қимылдарды синтездеу үшін аватар жүйелеріне біріктірілуі мүмкін. Жоба мәтінді глостарға және мультимедиялық форматтағы ымға түрлендіру үдерісін қарастырды. Аударылған глостар тізбегі ымдау тілінің құрылымын сақтай отырып, артықшылықты азайтады және сөйлемнің үйлесімділігін жақсартады. Демек, ымдау тілінің аватарларын басқару үшін тек семантикалық мағыналы бірліктер ғана пайдаланылады, бұл есептеу талаптарын азайтады. Қысқа және мағыналық жағынан бай глостар ым тілінде қол қимылдарын синтездеудің тиімді ресурсы қызметін атқарады.

Глоска негізделген тәсілдің бірнеше шектеулері бар. Глостар ең алдымен ымдар тізбегін көрсетеді және ымдау тілінің кеңістіктік компоненттерін, мимикаларын немесе вербалды емес белгілерді толық түсірмейді. Бұл экспрессивті құрылымды сипаттаудың дәлдігін шектейді. Болашақ зерттеулер глостарға негізделген және глостарға негізделмеген әдістерді біріктіретін гибридті үлгілерді әзірлеуге назар аударуы керек. Мысалы, қазіргі зерттеулер кросс-модальды күшейту және мұғалім мен оқушыны оқыту әдістері арқылы глосс пен бейне сәйкестігін жақсарта алады. Бұл тәсілдер ымдау тілінің кеңістіктік және экспрессивтік ерекшеліктерін дәлірек көрсету үшін екі жүйенің де артықшылықтарын біріктіру арқылы машиналық аударманың сапасын жақсартуға бағытталған.

Қаржыландыру туралы ақпарат. Бұл зерттеу Қазақстан Республикасының Ғылым және жоғары білім министрлігінің Ғылым комитеті тарапынан қаржыландырылды (Грант № BR24992875).

ӘДЕБИЕТТЕР

1 FAQ. WFD – World Federation of the Deaf. URL: https://wfdeaf.org/contact/faqs/ (date of access: 20.09.2025).

2 Human rights. Deaf History Europe. URL: https://deafhistory.eu/index.php/component/zoo/item/human-rights#:~:text=,from%20deaf%20people%E2%80%99s%20human%20rights (date of access: 20.09.2025).

3 Six difficulties: Sign language avatar. DW Innovation. URL: https://innovation.dw.com/articles/six-difficulties-sign-language-avatar (date of access: 20.09.2025).

4 Zou, Y., Lin, H. A basic General Service List for Chinese Sign Language. Journal of Deaf Studies and Deaf Education, 30(3), 405–418 (2025). https://doi.org/10.1093/jdsade/enaf012.

5 Caselli, N.K. et al. ASL-LEX: A lexical database of American Sign Language. Behavior research methods, 49(2), 784–801 (2017). https://doi.org/10.3758/s13428-016-0742-0.

6 Perlman, M. et al. Iconicity in signed and spoken vocabulary: a comparison between American Sign Language, British Sign Language, English, and Spanish. Frontiers in psychology, 9, 1433 (2018). https://doi.org/10.3389/fpsyg.2018.01433.

7 Zhao, Y. et al. Relationship between vocabulary knowledge and reading comprehension in deaf and hard of hearing students. The Journal of Deaf Studies and Deaf Education, 26(4), 546–555 (2021). https://doi.org/10.1093/deafed/enab023.

8 Convertino, C. et al. Word and world knowledge among deaf learners with and without cochlear implants. Journal of Deaf Studies and Deaf Education, 19(4), 471–483 (2014). https://doi.org/10.1093/deafed/enu024.

9 Hall, W.C. What you don't know can hurt you: The risk of language deprivation by impairing sign language development in deaf children. Maternal and child health journal, 21(5), 961–965 (2017). https://doi.org/10.1007/s10995-017-2287-y.

10 Deaf community: Sign language equals rights. Human Rights Watch URL: https://www.hrw.org/news/2022/09/23/deaf-community-sign-language-equals-rights (date of access: 20.09.2025).

11 AI and machine translation: A threat to the deaf community. Deaf Journalism URL: https://www.deafjournalism.eu/ai-and-machine-translation-a-threat-to-the-deaf-community/ (date of access: 20.09.2025).

12 Duarte, A. et al. How2sign: a large-scale multimodal dataset for continuous american sign language. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2735–2744 (2021). https://doi.org/10.1109/CVPR46437.2021.00276 Dataset: http://how2sign.github.io/ openaccess.thecvf.comhow2sign.github.io.

13 Li, D. et al. Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. Proceedings of the IEEE/CVF winter conference on applications of computer vision, 1459–1469 (2020). URL: https://openaccess.thecvf.com/content_WACV_2020/papers/Li_Word-level_Deep_Sign_Language_Recognition_from_Video_A_New_Large-scale_WACV_2020_paper.pdf. Dataset: https://dxli94.github.io/WLASL/ openaccess.thecvf.comdxli94.github.io.

14 Athitsos, V. et al. The american sign language lexicon video dataset. 2008 IEEE computer society conference on computer vision and pattern recognition workshops. IEEE, 2008, pp. 1–8. Dataset: https://www.bu.edu/asllrp/av/dai-asllvd.html.

15 Albanie, S. et al. Bbc-oxford british sign language dataset. 2021. arXiv preprint arXiv:2111.03635 (2024). URL: https://www.robots.ox.ac.uk/~vgg/data/bobsl/ (see arXiv:2111.03635). robots.ox.ac.ukarXiv.

16 Zhou, H. et al. Improving sign language translation with monolingual data by sign back-translation. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (2021), pp. 1316–1325. Dataset: CSL-Daily, https://ustc-slr.github.io/datasets/2021_csl_daily/ openaccess.thecvf.comVisual Sign Language Research Group.

17 Ham, S. et al. Ksl-guide: A large-scale korean sign language dataset including interrogative sentences for guiding the deaf and hard-of-hearing. 2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021). IEEE, 2021, pp. 1–8. Dataset: https://github.com/ChelseaGH/KSL-Guide (also listed at sign-lang@LREC). GitHubsign-lang.uni-hamburg.de.

18 Cormier, K., & Schembri, A. British Sign Language Corpus [Data collection]. UK Data Archive, 2018. https://doi.org/10.5255/UKDA-SN-851521 reshare.ukdataservice.ac.uk.

19 Belissen, V., Braffort, A., Gouiffès, M. Dicta-Sign-LSF-v2: remake of a continuous French sign language dialogue corpus and a first baseline for automatic sign language processing. LREC 2020, 12th Conference on Language Resources and Evaluation (2020). URL: https://aclanthology.org/2020.lrec-1.740/. Dataset page: https://www.sign-lang.uni-hamburg.de/lr/compendium/corpus/dictasignlsfv2.html aclanthology.orgsign-lang.uni-hamburg.de.

20 Camgoz, N. C. et al. Neural sign language translation. Proceedings of the IEEE conference on computer vision and pattern recognition, 7784–7793 (2018). URL: https://www-i6.informatik.rwth-aachen.de/~koller/RWTH-PHOENIX-2014-T/ www-i6.informatik.rwth-aachen.de.

21 Konrad, R., Hanke, T., Langer, G., Blanck, D., Bleicken, J., Hofmann, I., … Schulder, M. MEINE DGS – annotiert. Public Corpus of German Sign Language, 3rd release [Dataset]. Universität Hamburg, 2020. https://doi.org/10.25592/dgs.corpus-3.0 sign-lang.uni-hamburg.de

22 Ye, J. et al. Scaling back-translation with domain text generation for sign language gloss translation. arXiv preprint arXiv:2210.07054 (2022) https://doi.org/ 10.18653/v1/2023.eacl-main.34.

23 Zhu, D., Czehmann, V., Avramidis, E. Neural machine translation methods for translating text to sign language glosses. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 12523–12541 (2023). https://doi.org/10.18653/v1/2023.acl-long.700.

24 Xu, C. et al. Automatic gloss dictionary for sign language learners. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 83–92 (2022). https://doi.org/10.18653/v1/2022.acl-demo.8.

25 Mohamed, A. et al. A deep learning approach for gloss sign language translation using transformer. Journal of Computing and Communication, 1(2), 1–8 (2022).

26 Müller, M. et al. Considerations for meaningful sign language machine translation based on glosses. arXiv preprint arXiv:2211.15464 (2022).

27 Lin, K. et al. Gloss-free end-to-end sign language translation. arXiv preprint arXiv:2305.12876 (2023). https://doi.org/10.18653/v1/2023.acl-long.722.

28 Ye, J. et al. Cross-modality data augmentation for end-to-end sign language translation. arXiv preprint arXiv:2305.11096 (2023). https://doi.org/10.18653/v1/2023.findings-emnlp.904.

29 Shi, B. et al. Open-domain sign language translation learned from online video. arXiv preprint arXiv:2205.12870 (2022).

30 Kim, Y., Baek, H. Preprocessing for keypoint-based sign language translation without glosses. Sensors, 23(6), 3231 (2023).

31 Angelova, G., Avramidis, E., Möller, S. Using neural machine translation methods for sign language translation. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, 273–284 (2022).

32 Cao, Y. et al. Explore more guidance: A task-aware instruction network for sign language translation enhanced with data augmentation. arXiv preprint arXiv:2204.05953 (2022).

33 Sousa, C., Coheur, L., Moita, M. Enhancing Accessible Communication: from European Portuguese to Portuguese Sign Language. Findings of the Association for Computational Linguistics: EMNLP 2023, 11452–11460 (2023).

[1,3,4]Амангелді Н.,
PhD, ORCID ID0000-0002-4669-9254,
e-mail: amangeldi_n_3@enu.kz
[1,2]*Еримбетова А.,
к.т.н., ассоциированный профессор, PhD,
ORCID ID: 0000-0002-2013-1513,
*e-mail: aigerian8888@gmail.com
[1,4]Газизова Н.Е.,
магистр, ORCID ID: 0009-0007-7278-4202,
e-mail: g.nazerke2755@gmail.com
[1,3]Турсынова Н.А.,
магистр, ORCID ID: 0000-0002-1857-9028,
e-mail: tursynova_na@enu.kz
[4]Болатбекқызы К.,
бакалавр, ORCID ID: 0009-0000-7659-4974
e-mail: kerbezbolatbekkyzy03@gmail.com

[1]Институт информационных и вычислительных технологий, г. Алматы, Казахстан
[2]Евразийский технологический университет, г. Алматы, Казахстан
[3]Евразийский национальный университет им. Л.Н. Гумилева, г. Астана, Казахстан
[4]ТОО «SignBridge», г. Астана, Казахстан

ФОРМАЛИЗАЦИЯ ТЕКСТА НА КАЗАХСКОМ ЯЗЫКЕ С ИСПОЛЬЗОВАНИЕМ ГЛОССАРНОГО СЛОЯ

Аннотация

Технологии автоматической обработки жестового языка стали актуальной потребностью членов общества с нарушениями слуха и речи, которые сталкиваются с неравенством в эпоху цифровой трансформации. В последние годы вопрос рассмотрения жестового языка как формальной структуры, равной естественному языку, и его адаптации к автоматическим системам привлекает повышенное внимание исследователей. Для выполнения задачи автоматического перевода информации с естественного языка на жестовый язык в качестве промежуточного слоя используются глоссы, то есть текстовая форма жестового языка. В данном исследовании предлагается новый метод преобразования текста на казахском языке, учитывающий морфологические особенности казахского языка, в глоссы жестового языка с использованием методов обработки естественного языка. В частности, применяется архитектура Seq2Seq на основе модели ByT5 small. Полученные результаты показывают, что сформированные последовательности глосс являются компактными и семантически насыщенными, сохраняя внутреннюю структуру жестового языка. Последовательность глосс позволяет автоматизировать работу интерпретируемого промежуточного слоя, представляющего жестовые движения как логические единицы, аналогичные письменному языку. Преобразованная последовательность

глосс сохраняет структуру жестового языка, уменьшает избыточность и повышает связность предложения. Таким образом, использование только семантически значимых единиц при управлении аватарами жестового языка снижает вычислительные затраты. Короткие и семантически насыщенные глоссы являются эффективным ресурсом для синтеза движений рук в жестовом языке.

Ключевые слова: жестовый язык, последовательность глосс, обработка естественного языка, формализация, Seq2Seq, ByT5.

[1,3,4]Amangeldy N.,
PhD, ORCID ID: 0000-0002-4669-9254,
e-mail: amangeldi_n_3@enu.kz
[1,2]*Yerimbetova A.,
Cand. Tech. Sc., PhD, Associate Professor, ORCID ID: 0000-0002-2013-1513,
*e-mail: aigerian8888@gmail.com
[1,4]Gazizova N.,
MSc, ORCID ID: 0009-0007-7278-4202,
e-mail: g.nazerke2755@gmail.com
[1,3]Tursynova N.,
MSc, ORCID ID: 0000-0002-1857-9028,
e-mail: tursynova_na@enu.kz
[4]Bolatbekkyzy K.,
Bachelor, ORCID ID: 0009-0000-7659-4974,
e-mail: kerbezbolatbekkyzy03@gmail.com

[1]Institute of Information and Computational Technologies, Almaty, Kazakhstan
[2]Eurasian Technological University, Almaty, Kazakhstan
[3]L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
[4]«SignBridge» LLP, Astana, Kazakhstan

DEVELOPMENT OF A MODEL FOR REAL-TIME RECOGNITION OF KAZAKH SIGN LANGUAGE USING MEDIAPIPE AND DEEP LEARNING METHODS

Abstract

Technologies for automatic processing of sign language have become an urgent need for members of society with hearing and speech impairments who face inequality in the era of digital transformation. In recent years, the issue of considering sign language as a formal structure equal to natural language and adapting it to automatic systems has attracted increasing attention from researchers. To perform the task of automatically translating information from natural language into sign language, glosses, which are the textual representation of sign language, are used as an intermediate layer. For this purpose, this study proposes a new method for converting Kazakh language text, which reflects the morphological features of the Kazakh language, into sign language glosses using natural language processing techniques. In particular, a Seq2Seq architecture based on the ByT5 small model is applied. The obtained results demonstrate that the generated gloss sequences are compact and semantically rich while preserving the internal structure of sign language. The gloss sequence makes it possible to automate the work of an interpretable intermediate layer that represents sign language movements as logical units similar to written language. The transformed gloss sequence preserves the structure of sign language, reduces redundancy, and improves sentence coherence. Thus, the use of only semantically meaningful units to control sign language avatars reduces computational requirements. Short and semantically rich glosses serve as an effective resource for synthesizing hand movements in sign language.

Keywords: sign language, gloss sequence, natural language processing, formalization, Seq2Seq, ByT5.

Мақаланың редакцияға түскен күні: 01.10.2025