



Lublin, May 19, 2026

REVIEW OF THE FOREIGN SCIENTIFIC CONSULTANT

on the dissertation work of Abduali Balzhan Abdirasilkyzy on the topic:
“Development and research of a model and tools for an automated system for generating
meeting transcripts for the Kyrgyz–Kazakh language pair”
submitted for the degree of Doctor of Philosophy (PhD) in the educational program:
8D07109 – Automation and Internet of Things

The dissertation of Abduali Balzhan Abdirasilkyzy addresses a timely and scientifically significant problem: the development of intelligent tools for automatic generation of meeting transcripts for the Kyrgyz–Kazakh language pair. Despite the linguistic proximity of these two Turkic languages, speech processing and automatic translation in this direction remain insufficiently studied. The agglutinative nature of Turkic languages, their morphological complexity, and the large variety of word forms complicate the building of robust language models, making this research both necessary and relevant.

Balzhan demonstrated a systematic and comprehensive approach to the research, conducting an in-depth analysis of modern Automatic Speech Recognition (ASR) models — including Whisper, MMS, and Turkic ASR — as well as neural machine translation models such as mT5, Gemma, and NLLB. Established evaluation metrics including BLEU, chrF, WER, and TER were applied to objectively assess the quality of the developed systems.

The main scientific contributions achieved in the course of this work are as follows:

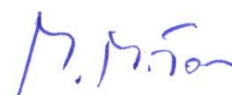
1. A large-scale parallel corpus consisting of 782 293 parallel sentences was developed for the Kyrgyz–Kazakh language pair, representing a significant contribution to low-resource NLP infrastructure for Turkic languages.
2. A Complete Set of Endings (CSE)-based Morphology-Aware Post-Processing method was proposed for correcting Kyrgyz speech recognition outputs. This method directly addresses the morphological complexity of agglutinative Turkic languages and improves the accuracy of ASR results in real-world conditions.
3. A prototype of an integrated ASR → MT system for automatic Meeting Transcript generation for the Kyrgyz–Kazakh language pair was developed. The system covers the full pipeline: speech-to-text conversion, Morphology-Aware Post-Processing, neural machine translation from Kyrgyz into Kazakh, and transcript generation.
4. Experimental evaluation of the developed models was conducted. The quality of ASR models was assessed, the effectiveness of the Morphology-Aware Post-Processing method was analyzed, and the results of Kyrgyz-to-Kazakh translation were presented and compared with existing approaches.



The practical significance of this work is considerable. The proposed system enables automation of meeting transcript generation in government institutions, educational organizations, and business environments. It contributes to the development of speech technologies for the Kyrgyz and Kazakh languages, provides a universal solution applicable to multilingual intelligent systems, and serves as a foundation for developing effective software tools for other low-resource languages.

The dissertation is well-structured and logically organized. It consists of an introduction, three chapters, a conclusion, and a list of references. The first chapter covers the theoretical foundations of ASR and machine translation systems; the second presents the methodology of the developed system, including its client-server architecture and data processing stages; and the third chapter presents experimental results, comparisons with existing approaches, practical application features, system limitations, and future development directions.

In conclusion, I confirm that the dissertation by Abduali Balzhan Abdirasilkyzy is a completed, independent scientific study conducted at a high academic level. In main opinion, the work fully satisfies the requirements for PhD dissertations in the field of 8D07109 – Automation and Internet of Things with respect to scientific novelty, theoretical contribution, methodological rigor, and practical significance. I strongly recommend the dissertation for defence.



Dr. Marek Milosz
Professor of university
Lublin University of Technology
Faculty of Electrical Engineering and
Computer Science
Department of Computer Science
E-mail: m.milosz@pollub.pl

ЛЮБЛИН ТЕХНОЛОГИЯ УНИВЕРСИТЕТИ
Электротехника және компьютерлік ғылымдар факультеті
Компьютерлік ғылымдар кафедрасы

PL-20-618 Люблин, Надбыстшицка 36Б, Польша

Телефон/факс +48 81 538 43 49

Почта: inst-inf@pollub.pl

Люблин, 19 мамыр 2026

“Қырғыз-қазақ тілдік жұбы үшін кездесу стенограммаларын жасауға
арналған автоматтандырылған жүйенің моделі мен құралдарын әзірлеу
және зерттеу” тақырыбындағы 8D07109 – Автоматтандыру және

Internet of Things

білім беру бағдарламасы бойынша философия докторы (PhD) дәрежесін
алу үшін ұсынылған Абдуали Балжан Абдирасилқызының

диссертациялық жұмысына

ШЕТЕЛДІК ҒЫЛЫМИ КОНСУЛЬТАНТТЫҢ ПІКІРІ

Абдуали Балжан Абдирасилқызының диссертациялық жұмысы қырғыз–қазақ тілдік жұбы үшін кездесу стенограммаларын автоматты түрде құруға арналған интеллектуалды құралдарды әзірлеу мәселесіне арналған өзекті әрі ғылыми тұрғыдан маңызды зерттеу болып табылады. Бұл екі түркі тілінің лингвистикалық жақындығына қарамастан, осы бағыттағы сөйлеуді өңдеу және автоматты аударма мәселелері әлі де жеткілікті деңгейде зерттелмеген. Түркі тілдерінің агглютинативті табиғаты, морфологиялық күрделілігі және сөз формаларының көптүрлілігі сенімді тілдік модельдерді құруды қиындатады, сондықтан бұл зерттеу қажетті әрі өзекті болып табылады.

Балжан зерттеу барысында жүйелі әрі кешенді тәсіл қолдана отырып, қазіргі заманғы автоматты сөйлеуді тану (ASR) модельдеріне — Whisper, MMS және Turkic ASR — сондай-ақ mT5, Gemma және NLLB секілді нейрондық машиналық аударма модельдеріне терең талдау жүргізді. Әзірленген жүйелердің сапасын объективті бағалау үшін BLEU, chrF, WER және TER сияқты кеңінен қолданылатын бағалау метрикалары пайдаланылды.

Осы жұмыс аясында қол жеткізілген негізгі ғылыми нәтижелер төмендегідей:

1. Қырғыз–қазақ тілдік жұбы үшін 782 293 параллель сөйлемнен тұратын ауқымды параллель корпус әзірленді. Бұл түркі тілдеріне арналған төмен ресурсты табиғи тілдерді өңдеу инфрақұрылымын дамытуға елеулі үлес болып табылады.

2. Қырғыз тіліндегі сөйлеуді тану нәтижелерін түзетуге арналған Complete Set of Endings (CSE) негізіндегі Morphology-Aware Post-Processing әдісі ұсынылды. Бұл әдіс агглютинативті түркі тілдерінің морфологиялық күрделілігін тікелей ескеріп, нақты жағдайларда ASR нәтижелерінің дәлдігін арттырады.

3. Қырғыз–қазақ тілдік жұбы үшін кездесу стенограммаларын автоматты түрде құруға арналған біріктірілген ASR → MT жүйесінің прототипі әзірленді. Жүйе толық өңдеу тізбегін қамтиды: сөйлеуді мәтінге түрлендіру, Morphology-Aware Post-Processing, қырғыз тілінен қазақ тіліне нейрондық машиналық аударма және стенограмма генерациясы.

4. Әзірленген модельдерге эксперименттік бағалау жүргізілді. ASR модельдерінің сапасы бағаланып, Morphology-Aware Post-Processing әдісінің тиімділігі талданды, сондай-ақ қырғыз тілінен қазақ тіліне аударма нәтижелері ұсынылып, қолданыстағы тәсілдермен салыстырылды.

Жұмыстың практикалық маңыздылығы зор. Ұсынылған жүйе мемлекеттік мекемелерде, білім беру ұйымдарында және бизнес орталарында кездесу стенограммаларын автоматты түрде құру үдерісін автоматтандыруға мүмкіндік береді. Сонымен қатар, ол қырғыз және қазақ тілдері үшін сөйлеу технологияларын дамытуға үлес

қосады, көптілді интеллектуалды жүйелерде қолдануға болатын әмбебап шешім ұсынады және басқа да төмен ресурсты тілдерге арналған тиімді бағдарламалық құралдарды әзірлеуге негіз болады.

Диссертациялық жұмыс құрылымдық жағынан жүйелі әрі логикалық тұрғыдан дұрыс ұйымдастырылған. Жұмыс кіріспеден, үш бөлімнен, қорытындыдан және пайдаланылған әдебиеттер тізімінен тұрады. Бірінші бөлімде ASR және машиналық аударма жүйелерінің теориялық негіздері қарастырылған; екінші бөлімде әзірленген жүйенің әдістемесі, оның ішінде клиент–сервер архитектурасы мен деректерді өңдеу кезеңдері сипатталған; үшінші бөлімде эксперименттік нәтижелер, қолданыстағы тәсілдермен салыстыру, практикалық қолдану ерекшеліктері, жүйенің шектеулері және болашақ даму бағыттары ұсынылған.

Қорытындылай келе, Абдуали Балжан Абдирасилқызының диссертациялық жұмысы жоғары академиялық деңгейде орындалған, аяқталған әрі дербес ғылыми зерттеу екенін растаймын. Жұмыс ғылыми жаңашылдығы, теориялық маңыздылығы, әдіснамалық негізділігі және практикалық құндылығы тұрғысынан 8D07109 – Автоматтандыру және Internet of Things білім беру бағдарламасы бойынша PhD диссертацияларына қойылатын талаптарға толық сәйкес келеді. Диссертациялық жұмысты қорғауға ұсынуға толық негіз бар деп есептеймін.

/Қолтаңба/

Доктор Марек Милош

ЛЮБЛИН ТЕХНОЛОГИЯ
УНИВЕРСИТЕТИ

Электротехника және

Компьютерлік ғылымдар факультеті
Компьютерлік ғылымдар

кафедрасы профессоры

email: m.milosz@pollub.pl