

«D070300 – Ақпараттық жүйелер» мамандығы бойынша философия докторы (PhD) дәрежесіне іздену үшін ұсынылған Шормакова Асем Ноябреннаның «Ағылшын тілінен қазақ тіліне машиналық аударманың пост-редакциялау үлгілерін, әдістерін және программалық құралдарын зерттеу және өңдеу» тақырыбындағы диссертациялық жұмысына ресми рецензенттің

СЫН-ПІКІРІ

р/н №	Критерийлер	Критерийлер сәйкестігі	Ресми рецензенттің ұстанымы
1.	Диссертация тақырыбының (бекіту күніне) ғылымның даму бағыттарына және/немесе мемлекеттік бағдарламаларға сәйкес болуы	1.1 Ғылымның даму бағыттарына және/немесе мемлекеттік бағдарламаларға сәйкестігі:	Ұсынылып отырған «Ағылшын тілінен қазақ тіліне машиналық аударманың пост-редакциялау үлгілерін, әдістерін және программалық құралдарын зерттеу және өңдеу» диссертациясы ғылымның даму бағытына сәйкес, себебі машиналық аударманы пост-редакциялау өзекті мәселелердің бірі болып саналады.
		1) Диссертация мемлекет бюджетінен қаржыландырылатын жобаның немесе нысаналы бағдарламаның аясында орындалған (жобаның немесе бағдарламаның атауы мен нөмірі); 2) Диссертация басқа мемлекеттік бағдарлама аясында орындалған (бағдарламаның атауы) 3) Диссертация Қазақстан Республикасының Үкіметі жанындағы Жоғары ғылыми-техникалық комиссия бекіткен ғылым дамуының басым бағытына сәйкес (бағытын көрсету)	Диссертациялық жұмыс PhD докторлық диссертациясының жоспарына сәйкес және мемлекет бюджетінен қаржыландырылатын «Қазақ тіліндегі ақпараттық-аналитикалық деректерді іздеу жүйесін дамыту» гранттық жобасының ғылыми-зерттеу жұмыстарының жоспарына сәйкес орындалды. (2018-2020 ж., мемлекеттік тіркеу нөмірі: No AP05132950). Аталған жобада қарастырылған тапсырманың біреуі кенінен қолданылған.
2.	Ғылымға маңыздылығы	<u>Жұмыс ғылымға елеулі үлесін қосады/қоспайды, ал оның маңыздылығы ашылған/ашылмаған.</u>	Бұл диссертациялық жұмыс ғылымға елеулі үлесін қосады. Зерттеу кезінде алынған нәтижелер ғылыми тұрғыда өте маңызды, себебі машиналық аударма қолданушылар саны қарқынды өсуде және күнделікті өмірде де машиналық аударманы қолдану жиілеп кетті. Сондықтан да машиналық аударманың сапасын жақсарту мақсатында пост-редакциялау әдісін қолдану өте маңызды болып тұр.
3.	Өзі жазу принципі	Өзі жазу деңгейі: 1) жоғары; 2) орташа; 3) төмен; 4) өзі жазбаған	Зерттеу жұмысын орындаушының диссертациялық жұмысты жазу барысында рәсімдеуі, түсіндіруі, сипаттауы жоғары деңгейде жазылған. Ғылыми жұмыстың жаңалығы жазу деңгейінің жоғарылығын көрсетеді.
4.	Ішкі бірлік принципі	4.1 Диссертация өзектілігінің негіздемесі: 1) негізделген;	4.1 Зерттелген ғылыми жұмыстың өзектілігінің негіздемесі негізделген. Ұсынылған жаңалығы, қолданылған, жақсартылған әдістердің тиімді

		2) жартылай негізделген; 3) негізделмеген.	колданылу реті өте жоғары деңгейде сипатталып дәлелделінген. Алғаш рет ұсынылып отырған технологияның айқындалған тапсырмалары ғылыми жұмыстың сапасы жоғары деңгейде екенін көрсетеді.
		4.2 Диссертация мазмұны диссертация тақырыбын айқындайды 1) <u>айқындайды</u> ; 2) жартылай айқындайды; 3) айқындамайды	4.2 Диссертация мазмұны диссертация тақырыбын айқындайды. Диссертация тақырыбы: «Ағылшын тілінен қазақ тіліне машиналық аударманың пост-редакциялау үлгілерін, әдістерін және программалық құралдарын зерттеу және өңдеу» болғандықтан ғылыми жұмыстың сипаттамасы толық мазмұнға сай.
		4.3. Мақсаты мен міндеттері диссертация тақырыбына сәйкес келеді: 1) <u>сәйкес келеді</u> ; 2) жартылай сәйкес келеді; 3) сәйкес келмейді	4.3 Зерттеу жұмысының мақсаты мен міндеттері диссертация тақырыбына сәйкес келеді. Диссертациялық жұмыстың нәтижесінде айқындалған мақсаттары мен міндеттері ағылшын тілінен қазақ тілінен аударылған машиналық аударманың пост-редакциялау нәтижелеріне сәйкес жүйелі түрде құрылған.
		4.4. Диссертацияның барлық бөлімдері мен құрылысы логикалық байланысқан: 1) <u>толық байланысқан</u> ; 2) жартылай байланысқан; 3) байланыс жоқ	4.4 Диссертацияның барлық бөлімдері мен құрылысы толық логикалық байланысқан. Зерттеу жұмысында жазылған барлық бөлімдер логикалық жүйеде жазылған. Машиналық аударма мен пост-редакциялау тақырыбы кеңінен талданып, қолданылған, жақсартылған әдістерді түсіндіре кетіп, ұсынылған жаңа PE-LC технологияның кезеңдері ретімен сипатталған.
		4.5 Автор ұсынған жаңа шешімдер (қағидаттар, әдістер) дәлелденіп, бұрыннан белгілі шешімдермен салыстырылып бағаланған: 1) <u>сыни талдау бар</u> ; 2) талдау жартылай жүргізілген; 3) талдау өз пікірін емес, басқа авторлардың сілтемелеріне негізделген	4.5 Автор ұсынған жаңа шешімдер (қағидаттар, әдістер) дәлелденіп, бұрыннан белгілі шешімдермен салыстырылып бағаланған. Ғылыми жұмысты автор жазған кезде жаңадан ұсынып отырған технологиясында қолданылған әдістерге кеңінен талдау жасаған. Бұрыннан белгілі әдістердің кемшілігі мен артықшылығын көрсете отырып, туралау әдісіне өз ғылыми үлесін қосып, бірнеше тапсырмаларды қамтитын тапсырмаларға талдау жасаған.
5.	Ғылыми жаңашылдық принципі	5.1 Ғылыми нәтижелер мен қағидаттар жаңа болып табыла ма? 1) <u>толығымен жаңа</u> ; 2) жартылай жаңа (25-75% жаңа болып табылады); 3) жаңа емес (25% кем жаңа болып табылады)	5.1 Зерттеу нәтижесінде алынған ғылыми нәтижелер мен қағидаттар жаңа болып табылады. Диссертациялық жұмыста алғаш рет ұсынылған PE-LC технологиясының алгоритмі толығымен жаңа болып табылады. Ғылыми жұмыста қарастырылған келесі кезеңдер жұмыстың жаңалығын айқындайды: 1 – қазақ тіліндегі сөйлемнің қате аударылған сөздерін анықтау; 2 – қате аударылған сөздердің синонимдер каталогын автоматты түрде қалыптастыру; 3 – қате аударылған сөзді мағынасы жағынан жақын синониммен ауыстырып түзетілген сөйлемнің машиналық аудармасын шығару. 4 – жоғарыдағы үш тапсырманы біріктіріп пост-редакциялау технологиясын құрастыру.

		<p>5.2 Диссертацияның қорытындылары жаңа болып табыла ма? 1) <u>толығымен жаңа</u>; 2) <u>жартылай жаңа</u> (25-75% жаңа болып табылады); 3) жаңа емес (25% кем жаңа болып табылады)</p>	<p>5.2 Диссертация нәтижесінде алынған қорытындылар толығымен жаңа болып табылады. Себебі зерттеу нәтижесінде алынған жақсы көпесткіштер мен жаңалықтар толық ашылып жазылған. Айтап айтсақ: алғаш рет пост-редакциялау технологиясы ұсынылған; автоматты түрде қазақ тілінде қате аударылған сөздерден каталог құрылған; синоним сөздер ішінен семантикалық таңдау негізінде енгізілген мәтінге қатысты мағынасы жағынан ең жақын, ықтималдығы жоғары сөзді шығарып толық пост-редакцияланған сөйлем шығарылды.</p>
		<p>5.3 Техникалық, технологиялық, экономикалық немесе басқару шешімдері жаңа және негізделген бе? 1) <u>толығымен жаңа</u>; 2) <u>жартылай жаңа</u> (25-75% жаңа болып табылады); 3) жаңа емес (25% кем жаңа болып табылады)</p>	<p>5.3 Диссертациялық жұмыста ағылшын тілінен қазақ тіліне аударылған мәтінді пост-редакциялау мәселесін шешу мақсатында қолданылған техникалық, технологиялық шешімдер жаңа және негізделген. Пост-редакциялау технологиясын іске асыру үшін ағылшын-қазақ мәтіндері қолданылып, Google, Apertium машиналық аудармалары кеңінен қолданылды.</p>
6.	Негізгі қорытындылардың негізділігі	<p>Барлық қорытындылар ғылыми тұрғыдан қарағанда ауқымды дәлелдемелерде <u>негізделген/негізделмеген</u> (qualitative research және өнертану және гуманитарлық бағыттары бойынша)</p>	<p>Зерттеу жұмысының нәтижесінде алынған ауқымды дәлелдемелер негізделген. Алғаш рет ағылшын – қазақ машиналық аударманың PE-LC пост-редакциялау технологиясы әзірленді. Ағылшын тілінен қазақ тіліне қате аударылған сөздерді табу әдісі кері аудармамен жетілдірілді. Қате аударылған қазақ тіліндегі сөздер синонимдерінің каталогын автоматты түрде қалыптастыру сұлбасы мен алгоритмі жасалды. Қате аударылған сөздің ықтималдығы жоғары синоним сөзді таңдау семантикалық текше әдісінің үлгісі мен алгоритмі бейімделді.</p>
7.	Қорғауға шығарылған негізгі қағидаттар	<p>Әр қағидат бойынша келесі сұрақтарға жауап беру қажет: 7.1 Қағидат дәлелденді ме? 1) <u>дәлелденді</u>; 2) шамамен дәлелденді; 3) шамамен дәлелденбеді; 4) дәлелденбеді 7.2 Тривиалды ма? 1) <u>ия</u>; 2) <u>жоқ</u> 7.3 Жаңа ма? 1) <u>ия</u>; 2) жоқ 7.4 Қолдану деңгейі: 1) <u>тар</u>; 2) орташа;</p>	<p>7.1 Диссертациялық жұмыста қорғауға шығарылған қағидаттар келесідей:</p> <ul style="list-style-type: none"> • Ағылшын–қазақ машиналық аударманың жаңа автоматты пост-редакциялау технологиясы. • Ағылшын тілінен қазақ тіліне қате аударылған сөздерді анықтаудың жетілдірілген әдісі. • Ағылшын тілінен қате аударылған қазақ сөздер синонимдерінің каталогын автоматты түрде қалыптастыру технологиясы. • Семантикалық текше негізінде ықтималдығы жоғары синонимді таңдау бейімделген әдісі. <p>Аталған қағидаттар толығымен дәлелденген.</p> <p>7.2 Зерттеу жұмысының айқын көрсетілген өзектілігі, ұсынылған жаңа технологиясы, зерттеу нәтижесінде алынған өте жақсы көрсеткіштер жұмысты тривиалды еместігін анық көрсетеді.</p>

		<p>3) <u>кең</u> 7.5 Мақалада дәлелденген бе? 1) <u>ия</u>; 2) <u>жоқ</u></p>	<p>7.3 Диссертациялық жұмыста қорғауға шығарылған қағидаттар толығымен жаңа. Себебі алғаш рет ұсынылып отырған пост-редакциялау технологиясы, автоматты түрде қазақ тіліндегі каталогты қалыптастыру мен семантикалық текше әдісін қолдана отырып шығарылған нәтиже алғаш рет ұсынылып отыр. 7.4 Зерттеу жұмысында ұсынылған технологияны қолдану деңгейі кең. Себебі тек ағылшын-қазақ мәтіндері үшін ғана емес, басқа тілдер жұбына да қолданылуға болады. 7.5 Ғылыми жұмыстың нәтижелері 14 ғылыми мақалаларда жарық көрген және бір авторлық куәлік алынған. Соның ішінде келесі мақалалар Скопус базасына енгізілген:</p> <ul style="list-style-type: none"> • Shormakova A., Zhumanov Z.H., Rakhimova D. "Post-editing of words in Kazakh sentences for information retrieval". <i>Journal of Theoretical and Applied Information Technology</i>, 2019, 97(6), p. 1896–1908. (Scopus 2021: Q4, CiteScore-1.3; Percentile- 30%) • Abeustanova (Shormakova) A., Tukeyev U. "Automatic Post-editing of Kazakh Sentences Machine Translated from English". <i>Studies in Computational Intelligence/Advanced Topics in Intelligent Information and Database Systems</i>, vol. 710– Springer International Publishing, 2017. – p. 283-295. (Scopus 2021: Q4, CiteScore-1.8; Percentile- 27%). • Rakhimova D., Assem S. "Problems of Semantics of Words of the Kazakh Language in the Information Retrieval". <i>Lecture Notes in Computer Science</i> (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019, 11684 LNAI, p. 70–81. (Scopus 2021: Q2, SJR=0.25, CS=2.1, Percentile-50%)
8.	<p>Дәйектілік принципі Дереккөздер мен ұсынылған ақпараттың дәйектілігі</p>	<p>8.1 Әдістеменің таңдауы - негізделген немесе әдіснама нақты жазылған 1) <u>ия</u>; 2) <u>жоқ</u></p> <p>8.2 Диссертация жұмысының нәтижелері компьютерлік технологияларды қолдану арқылы ғылыми зерттеулердің қазіргі заманғы әдістері мен деректерді өңдеу және интерпретациялау әдістемелерін пайдалана отырып алынған: 1) <u>ия</u>; 2) <u>жоқ</u></p>	<p>8.1 Диссертациялық жұмыста таңдалған әдістеме нақты жазылған. Туралау негізіндегі қате сөздерді табу алгоритмі, қазақ тіліндегі каталог қалыптастыру алгоритмі, семантикалық текше әдісі пайдаланылып, қолданылу реті нақты жазылған.</p> <p>8.2 Диссертация жұмысының нәтижелері компьютерлік технологияларды қолдану арқылы ғылыми зерттеулердің қазіргі заманғы әдістері мен деректерді өңдеу және интерпретациялау әдістемелерін пайдалана отырып алынған. Зерттеу жұмысы python бағдарламасының соңғы нұсқасында жазылып, зерттеу кезінде 250мың сөйлемнен тұратын ағылшын-қазақ параллельді мәліметтер қоры пайдаланылып, 3,5 млн синонимдер сөздік қорынан тұратын thesaurus.com онлайн сөздігі қолданылды.</p>

		<p>8.3 Теориялық қорытындылар, модельдер, анықталған өзара байланыстар және заңдылықтар эксперименттік зерттеулермен дәлелденген және расталған (педагогикалық ғылымдар бойынша даярлау бағыттары үшін нәтижелер педагогикалық эксперимент негізінде дәлелденеді):</p> <p>1) <u>ия</u>;</p> <p>2) жоқ</p>	<p>8.3 Теориялық қорытындылар, модельдер, анықталған өзара байланыстар және заңдылықтар эксперименттік зерттеулермен дәлелденген және расталған. Диссертациялық жұмыс нәтижесінде алынған эксперименттік нәтижелер өте жақсы көрсеткіштер көрсетті. Ұсынылған PE-LC технологиясының 2017 жылғы BLEU сапа көрсеткіші Google MA-дан +6%, Prompt MA-дан +11% жақсы нәтиже, 2019 жылы +3,88% және +8,37%, 2020 жылы Google MA-дан +13,1%, 2021 жылы +11,5% нәтижелер алынды. 2022 жылы эксперимент нәтижелері бойынша PE-LC пост-редакциялау технологиясының BLEU сапа көрсеткіші Google MA-дан +6,47% жақсы нәтиже көрсетті. PE-LC технологиясының сапа көрсеткіштері Google MA-дан TER -4,66% және WER -4,28% төмен көрсеткіш көрсетті, яғни бұл ұсынылып отырған технологияның салыстырмалы түрде жақсырақ нәтиже алынғанын көрсетеді.</p>
		<p>8.4 Маңызды мәлімдемелер нақты және сенімді ғылыми әдебиеттерге сілтемелермен <u>расталған</u> / ішінара расталған / расталмаған</p>	<p>8.4 Диссертациялық жұмыстағы маңызды мәлімдемелер нақты және сенімді ғылыми әдебиеттерге сілтемелермен расталған. Пайдаланылған ғылыми әдебиеттер зерттеу саласына сәйкес.</p>
		<p>8.5 Пайдаланылған әдебиеттер тізімі әдеби шолуға жеткілікті/жеткіліксіз</p>	<p>8.5 Пайдаланылған әдебиеттер тізімі әдеби шолуға жеткілікті. Зерттеу нәтижесінде пост-редакциялау тақырыбына сәйкес Скопус мәліметтер қорына енгізілген журналдар мен кітап басылымдарда мақалалар жарық көрген.</p>
9	Практикалық құндылық принципі	<p>9.1 Диссертацияның теориялық маңызы бар:</p> <p>1) <u>ия</u>;</p> <p>2) жоқ</p>	<p>9.1 Диссертацияның теориялық маңызы бар, өйткені зерттелген тақырып табиғи тілдерді өңдеу саласында маңызды тақырыптардың бірі болып саналады. Сонымен қатар қазақ тіліндегі пост-редакциялау технологиясы алғаш рет ұсынылып отыр.</p>
		<p>9.2 Диссертацияның практикалық маңызы бар және алынған нәтижелерді практикада қолдану мүмкіндігі жоғары:</p> <p>1) <u>ия</u>;</p> <p>2) жоқ</p>	<p>9.2 Диссертацияның практикалық маңызы бар және алынған нәтижелерді практикада қолдану мүмкіндігі жоғары. Ағылшын-қазақ машиналық аударманың көмегімен қазақ тіліндегі қателіктерді түзеу мақсатында аударма сапасын жақсарту үшін алғаш рет ұсынылған технологияны практикалық қолдану өте маңызды болып саналады.</p>
		<p>9.3 Практикалық ұсыныстар жаңа болып табылады?</p> <p>1) <u>толығымен жаңа</u>;</p> <p>2) жартылай жаңа (25-75% жаңа болып табылады);</p> <p>3) жаңа емес (25% кем жаңа болып табылады)</p>	<p>9.3 Диссертациялық жұмыста қолданылған практикалық ұсыныстар жаңа болып табылады. Алғаш ұсынылып отырған технологияда автоматты түрде синонимдер каталогын қалыптастыру кезінде қолданылатын онлайн сөздік, семантикалық текшені құру кезіндегі мәліметтерді өңдеу (оқыту) мен толық түзетілген(пост-редакцияланған) сөйлемді шығару кезінде толығымен жаңа практикалық ұсыныстар көрсетілген.</p>

10.	Жазу және ресімдеу сапасы	Академиялық жазу сапасы: 1) жоғары; 2) орташа; 3) орташадан төмен; 4) төмен.	Изденуші Шормакова Асем Ноябрьрнаның «Ағылшын тілінен қазақ тіліне машиналық аударманың пост-редакциялау үлгілерін, әдістерін және программалық құралдарын зерттеу және өңдеу» тақырыбындағы диссертациялық жұмысы ғылыми тілде, сауатты жазылған, техникалық тұрғыдан дұрыс рәсімделген және өзектілігі, қойылған мақсаттары мен міндеттері ғылыми жаңалығы тәжірибелік маңыздылығы жағынан «6D070300 – Ақпараттық жүйелер» мамандығы бойынша философия докторы (PhD) дәрежесін алудың «Дәрежелер беру қағидалары» талаптарына толығымен сай келеді. Диссертациялық жұмыс бойынша айтылған ұсыныстар жұмысты әрі қарай жетілдіруге нұсқаулық ретінде көрсетілген және олар диссертациялық жұмыстың құндылығын төмендетпейді. Шормакова Асем Ноябрьрн «6D070300 – Ақпараттық жүйелер» мамандығы бойынша философия докторы (PhD) дәрежесін алуға лайықты деп санаймын.
-----	---------------------------	--	---

Ресми рецензенттер пікірлерінде келесі шешімдердің бірін көрсетеді:

- 1) философия докторы (PhD) немесе бейіні бойынша доктор дәрежесін беру;
- 2) диссертацияны пысықтауға жіберу (диссертация мақалалар сериясы нысанында қорғалған жағдайлардан басқа);
- 3) философия докторы (PhD) немесе бейіні бойынша доктор ғылыми дәрежесін беруден бас тарту

Ресми рецензенттер пікірлерінің көшірмелері докторантқа диссертация қорғауға дейін кемінде 5 (бес) жұмыс күнінен кешіктірілмей беріледі.

Ресми рецензент:
Л.Н.Гумилев атындағы Еуразия ұлттық университетінің
«Ақпараттық жүйелер» кафедрасының доценті,
PhD, қауымдастырылған профессор

Самбетбаева М.А.

