

Ғ.М. Аязбаев<sup>1</sup> , Н.М. Әшімбаева<sup>1</sup> , А.З. Бисенғали<sup>1</sup> , С.Қ. Құлманов<sup>1\*</sup> ,  
М. Нұрлан<sup>1</sup> 

<sup>1</sup>Ахмет Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан  
\*e-mail: k.k.sarsen@mail.ru

## ҚАЗАҚ ЖӘНЕ ОРЫС ТІЛДЕРІНІҢ ПАРАЛЛЕЛЬ ІШКОРПУСЫ: ӘЗІРЛЕМЕСІ, ҚОЛДАНЫЛУЫ ЖӘНЕ МӘСЕЛЕЛЕРІ

**Аннотация.** Мақалада лингвистикалық корпусар жасаудың тарихына қысқаша шолу жасалып, олардың әртүрлі сипаттарына қарай жіктелуіне, параллель ішкорпустардың түрлеріне сипаттама берілді. Қазақ тілінің ұлттық корпусының құрамында әзірленіп жатқан параллель ішкорпуста М.Әуезовтің «Абай жолы» роман-эпопеясының қазақ тіліндегі түпнұсқа мәтіні мен оның А.Ким жасаған орыс тіліндегі мәтінінің аудармасы абзац (сөйлем) деңгейінде қолмен теңестірілді.

Параллель ішкорпусты әзірлеу барысында Microsoft Office Excel, Notepad++, Python, Django, MySQL бағдарламалық құралдары қолданылды. Параллель ішкорпустың бағдарламалық архитектурасын және жұмыс істеу тәртібін төмендегіше көрсетуге болады: 1) екі тілдегі мәтіндер Excel офистік бағдарламасы арқылы жинақталып, абзац (сөйлем) деңгейінде қолмен теңестірілді; 2) теңестірілген мәтіндер Excel файлынан тікелей MySQL мәліметтер қорын басқару жүйесіне жүктелді; 3) жүктелген мәтіндер Notepad++ мәтіндік процессор бағдарламасы арқылы сұрыпталып, олардың статистикасы алынды; 4) сұрыпталған мәтіндерді интернетке жариялау және пайдаланушының сұрауларын қамтамасыз ету үшін Django веб сервері қолданылды; 5) Django веб серверін MySQL мәліметтер қорын басқару жүйесімен байланыстыру үшін Python тілінде жазылған, іздеу функциясымен қамтамасыз етілген Processing.py бағдарламасы қолданылды; 6) параллель ішкорпустың бағдарламалық архитектурасы client-server және MVC (Model-View-Controller) технологиялары қолданыла отырып әзірленді.

Параллель ішкорпус теңестірілген мәтіндер базасынан, белгіленім, метабелгіленім және іздеу жүйесінен және ішкорпуста енгізілген мәтін туралы ақпарат (метабелгіленім) мынадай параметрлерден тұрады: *автор, аудармашы, шығарма атауы, аударма атауы, шығарманың басылған мерзімі, аударылған мерзімі, түпнұсқа тіл, аударма тіл.* Іздеу жүйесі ізделетін сөзді сөз, сөз тіркес, сөйлем, бас әріп (қазақ және орыс тілдеріндегі) параметрлері арқылы табуға мүмкіндік береді.

Мақалада қазақ және орыс тілдеріндегі параллель ішкорпустың интерфейсіне және іздеу параметрлерінің бірі арқылы ізделетін сөзді іздегеннен кейінгі нәтижелер интерфейсіне сипаттама беріліп, екі тілдегі мәтінде қолданылған жалпы және қайталанбайтын сөз саны, сөйлем саны және екі тілде ең жиі қолданылған он сөзқолданыстың сандық және пайыздық көрсеткіші анықталды.

Сонымен қатар роман-эпопеяның қазақ тіліндегі түпнұсқа мәтінін орысша аударма нұсқасымен абзац (сөйлем) деңгейінде теңестіру барысында мынадай ерекшеліктер анықталды: 1) тұрпат межесі тұрғысынан, яғни абзацта (сөйлемде) қолданылған сөздер сан жағынан шамамен сәйкес келеді; 2) мазмұн межесі тұрғысынан шамамен сәйкес келеді; 3) тұрпат және мазмұн межесі тұрғысынан сәйкес келмейді: қазақ тіліндегі түпнұсқа мәтіндегі кейбір абзацтар (сөйлемдер) орыс тіліне бұрыс немесе үстүрт, қысқа аударылған, олардың жуық мағынасы берілген.

**Тірек сөздер:** корпус, лингвистика, параллель корпус, роман-эпопея, теңестіру, бағдарлама, белгіленім.

*\*Мақала BR11765619 «Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі қазақ тілінің ұлттық корпусын әзірлеу: ғылыми-зерттеу және оқыту интернет-ресурсы» атты бағдарламалық-нысаналы қаржыландыру бойынша зерттеу аясында жазылды.*

Ғ.М. Аязбаев<sup>1</sup>, Н.М. Ашимбаева<sup>1</sup>, А.З. Бисенғали<sup>1</sup>, С.К. Құлманов<sup>1\*</sup>, М. Нұрлан<sup>1</sup>

<sup>1</sup>Институт языкознания имени Ахмета Байтурсынұлы, Алматы, Казахстан  
\*e-mail: k.k.sarsen@mail.ru

### ПАРАЛЛЕЛЬНЫЙ ПОДКОРПУС КАЗАХСКОГО И РУССКОГО ЯЗЫКОВ: РАЗРАБОТКА, ФУНКЦИОНИРОВАНИЕ И ПРОБЛЕМЫ

**Аннотация.** В статье дан краткий обзор истории создания лингвистических корпусов, описана их классификация по различным признакам и типы параллельных подкорпусов. Оригинальный казахский текст романа-эпопеи М. Ауэзова «Абай жолы» и его русский перевод, выполненный А. Кимом, были вручную выровнены на уровне абзаца (предложения) в параллельном подкорпусе, разрабатываемом в составе Национального корпуса казахского языка.

В ходе разработки параллельного подкорпуса использовались программные средства Microsoft Office Excel, Notepad++, Python, Django, MySQL. Программную архитектуру и порядок функционирования параллельного подкорпуса можно представить следующим образом: 1) тексты на двух языках были собраны с помощью офисной программы Excel и выровнены вручную на уровне абзаца (предложения); 2) выровненные тексты загружались непосредственно из файла Excel в систему управления базами данных MySQL; 3) загруженные тексты были отсортированы с помощью программы текстового процессора Notepad++, была получена их статистика; 4) для публикации отсортированных текстов в интернете и обеспечения запросов пользователей использовался веб-сервер Django; 5) для подключения веб-сервера Django к системе управления базами данных MySQL была использована программа Processing.py, написанная на Python и снабженная функцией поиска; 6) программная архитектура параллельного подкорпуса разработана с применением технологий client-server и MVC (Model-View-Controller).

Параллельный подкорпус состоит из базы выровненных текстов, разметок, метаразметок и поисковой системы, информация о тексте, введенного в подкорпус (метаразметка), включает следующие параметры: автор, переводчик, название произведения, название перевода, дата публикации произведения, срок перевода, язык оригинала, язык перевода. Поисковая система позволяет найти искомое слово по параметрам: слово, фраза, предложение, заглавные буквы (на казахском и русском языках).

В статье дается описание интерфейса параллельного подподкорпуса на казахском и русском языках и интерфейса результатов после поиска искомого слова через один из параметров поиска, было определено общее и неповторяющееся количество слов, используемых в тексте на двух языках, количество предложений, а также числовые и процентные значения десяти наиболее часто используемых словупотреблений на обоих языках.

Кроме того, в процессе выравнивания оригинального казахского текста романа-эпопеи с русской переводной версией на уровне абзаца (предложения) были выявлены следующие особенности: 1) с точки зрения структуры, то есть слова, используемые в абзаце (предложении), примерно равнозначны по количеству; 2) с точки зрения содержания примерно совпадают; 3) не совпадают по структуре и содержанию: некоторые абзацы (предложения) в оригинальном тексте на казахском языке переведены на русский язык неправильно, поверхностно либо кратко, дан их приблизительный смысл.

**Ключевые слова:** корпус, лингвистика, параллельный корпус, роман-эпопея, выравнивание, программа, разметка.

G.M. Ayazbaev<sup>1</sup>, N.M. Ashimbaeva<sup>1</sup>, A.Z. Bisengali<sup>1</sup>, S.K. Kulmanov<sup>1\*</sup>, M. Nurlan<sup>1</sup>

<sup>1</sup>A. Baitursynuly Institute of Linguistics, Almaty, Kazakhstan

\*e-mail: k.k.sarsen@mail.ru

#### PARALLEL CORPUS OF THE KAZAKH AND RUSSIAN LANGUAGES: DEVELOPMENT, OPERATION AND PROBLEMS

**Annotation.** The research paper gives a brief overview of the history of the creation of linguistic corpora, describes their classification according to various criteria and types of parallel subcorpora. The original Kazakh text of M. Auezov's epic novel «Abai Zholy» and its Russian translation, made by A. Kim, were manually aligned at the level of a paragraph (sentence) in a parallel subcorpus being developed as part of the national corpus of the Kazakh language.

During the development of the parallel subcorpus, Microsoft Office Excel, Notepad++, Python, Django, MySQL software tools were used. The software architecture and the order of operation of the parallel subcorpus can be represented as follows: 1) texts in two languages were collected using the Excel office program and aligned manually at the paragraph (sentence) level; 2) aligned texts were loaded directly from an Excel file into the MySQL database management system; 3) the downloaded texts were sorted using the Notepad++ word processor program, their statistics were obtained; 4) the Django web server was used to publish the sorted texts on the Internet and provide user requests; 5) the Processing.py program written in Python and equipped with a search function was used to connect the Django web server to the MySQL database management system; 6) the parallel subcorpus software architecture was developed using client-server and MVC (Model-View-Controller) technologies.

The parallel subcorpus consists of a *database of aligned texts, markups, metamarkups* and a *search engine*, information about the text entered into the subcorpus (metamarkup) includes the following parameters: *author, translator, work title, translation title, publication date of the work, translation period, original language, translation language*. The search engine allows users to find the desired word by parameters: word, phrase, sentence, and capital letters (in Kazakh and Russian).

The paper describes the interface of the parallel subcorpus in Kazakh and Russian and the interface of the results after searching for the desired word through one of the search parameters, the total and non-repeating number of words used in the text in two languages, the number of sentences, as well as numerical and percentage values of the ten most commonly used words in both languages were determined.

In addition, in the process of aligning the original Kazakh text of the epic novel with the Russian translated version at the paragraph (sentence) level, the following features were identified: 1) from the point of view of structure, that is, the words used in the paragraph (sentence) are approximately equivalent in number; 2) from the point of view of content, they approximately coincide; 3) do not coincide in structure and content: some paragraphs (sentences) in the original text in Kazakh are translated into Russian incorrectly, superficially or briefly, their approximate meaning is given.

**Keywords:** corpus, linguistics, parallel corpus, epic novel, alignment, program, markup.

## Кіріспе

Кез келген тілдің мәтіндер базасынан және оларға жасалған тілдік талдаулардан (белгіленімдерден) тұратын іздеу жүйесі бар ақпараттық ресурс саналатын тілдік корпус жасау тәжірибесі ағылшын тілінің Лондон-Лунд (The London-Lund Corpus) ауызша тіл корпусынан (1960) басталып (Svartvik, Quirk, 1980), Браун корпусы (The Brown Standard Corpus of American English) (1963) (Francis), Ланкастер-Осло/Берген (LOB), Браун корпусының британдық ағылшын нұсқасы, Frown және FLOB, Brown және LOB баламалары және т.б. (Hundt; Leech, 2005) корпусстар пайда болды.

Тіл білімінде корпусстар әртүрлі сипаттарына байланысты былайша жіктеледі: 1) *құрылымдық дәлгейіне қарай*: электрондық мұрағат (өңделмеген мәтіндер), электрондық кітапхана (өңделген мәтіндер), мәтіндер корпусы (мәтіндердің өңделген жиынтығы); 2) *тілдердің қамтылуына қарай*: біртүлді, екітүлді, көптүлді; 3) *пайдаланылу мақсатына қарай*: зерттеу, иллюстративтік; 4) *қолданыстық сипатына қарай*: статикалық, динамикалық; 5) *индекстелуіне қарай*: белгіленім жасалған, белгіленім жасалмаған; 6) *формасына қарай*: 1) жазба, ауызекі сөйлеу, аралас және т.б.; 7) *жанрына қарай*: әдеби, фольклорлық, драматургиялық, публицистикалық және т.б.; 8) *қолжетімділік сипатына қарай*: коммерциялық және көпшілікке қолжетімді тегін корпусстар және т.б.

Қазіргі кезде көпшілікке қолжетімді *Ұлттық корпус* дегеніміз – белгілі бір тілдің табиғатын толық бейнелей алатын, тілдегі барлық стиль мәтіндерін қамтитын, көлемді, теңгерімді (репрезентативті) корпус. (Жұбанов, Жаңабекова, 2017: 48).

*Қазақ тілінің ұлттық корпусында* (бұдан әрі – ҚТҰК) қазақ тілінің 30 миллион сөзқолданыстан тұратын электронды мәтіндер қоры жинақталған (<https://qazcorpus.kz/>). ҚТҰК-ына енгізілген мәтіндер қорына *морфологиялық, лексикалық, семантикалық, фонетика-фонологиялық белгіленімдер* қойылған. Мысалы, корпусстан «арман» сөзін теріп іздегенде осы сөз кездесетін 21519 құжат табылады. Табылған «арман» сөзін *морфологиялық белгісіне қарай зат есім, семантикасына қарай жалпы, дара, түбір, дерексіз, лексикалық мағынасын «көксеген мақсат, тілек, қиял», фонетикасын «арман» – жуан тембр, «ар» – тұйық буын, «ман» – бітеу буын, фонологиясын «а» – ашық буын, езулік, тіларты, жуан дауысты фонема, «р» – тіл ұшы бос жуысыңқы көмей қуысты ауыз жолды діріл, үнді дауыссыз фонема, «м» – ерін-ерін бос тоғысыңқы ауыз қуысты мұрын жолды, үнді дауыссыз фонема, «н» – тіл ұшы, босаң тоғысыңқы көмей қуысты мұрын жолды, үнді дауыссыз фонема* деп талдап көрсетеді.

Қазіргі кезде ҚТҰК құрамында негізгі корпуспен қатар *диалектологиялық, тарихи, ауызша, параллель ішкорпустар* да жасалып жатыр. Ішкорпустар негізгі корпусстан *мәтіндік базасы, белгіленім, метабелгіленім, іздеу жүйесі* бойынша ерекшеленеді.

ҚТҰК-ының базасында параллель ішкорпустарды құрудың нәтижесінде қазақ тіліндегі мәтіндер мен басқа тілдердегі теңестірілген аударма мәтіндер базасын қалыптастыру арқылы тіл білімі (контрастивті лексикология, екітүлді және көптүлді лексикография, терминология, аударма теориясы, лингвомәдениеттану, когнитивті лингвистика), әдебиеттану (мәтіндердің аудармаларына текстологиялық талдау жасау), мәтінді автоматты тану (компьютерді мәшинелік аударма), әдістеме (қазақ тілін оқытудың лингвистикалық платформасын жасау, шет тілдерін оқыту) және т.б. салалардағы зерттеулерге тілдік материал жинақталады. Осы ретте бұл мақала ҚТҰК-ның құрамына енетін *параллель ішкорпустың* ерекшеліктеріне арналып отыр.

## Материал және әдістер

Мақалада қазақ және орыс тілдеріндегі параллель ішкорпусты әзірлеудің алғашқы сатысында екі тілдегі *тілдік материалды электронды пішінге келтіру*, электронды пішінге келтірілген мәтінді мазмұндық-семантикалық жағынан *теңестіру* (выравнивание), лингвистикалық белгіленімдер әзірлемесін жасауда және пайдаланушыға қажетті сөзді жылдам табуға мүмкіндік беретін «Іздеу» жүйесі интерфейсіні жетілдіруде *компьютерлік-бағдарламалық* әдістер, сондай-ақ қазақ және орыс тілдеріндегі параллель ішкорпусты әзірлеу барысында Microsoft Office Excel, Notepad++, Python, Django, MySQL бағдарламалық құралдары, сондай-ақ client-server және MVC (Model-View-Controller) технологиялары қолданыла отырып әзірленді.

## Әдебиетке шолу

«Лингвистикалық корпус» терминіне осы саланың көрнекті маманы В.П.Захаров мынадай анықтама береді: «Лингвистикалық немесе тілдік корпус дегеніміз – нақты лингвистикалық міндеттерді шешуге арналған үлкен, электронды түрдегі, біріздендірілген, құрылымданған, белгіленім қойылған тілдік деректердің жиынтығы» (Захаров, 2005: 3).

Қазіргі кезде қазақ тілінің мәтіндер корпусын жасаудың жалпы лингвистикалық теориялық және практикалық проблемаларына (Жұбанов, 2009), мәтіндерге әртүрлі белгіленімдер қою және оны автоматтандыру (Жаңабекова, 2012; Жубанов, 2015; Жаңабекова, Пирманова, 2019; Құлманов, Жаңабекова, Әшімбаева, Бисенғали, Шүленбаев, Қордабай, 2022; Койбағаров, Мусабаев, Кулманов, 2012) және т.б. мәселелерге арналған ғылыми зерттеулердің нәтижелері жарық көрді.

Лингвистикалық корпустар тілдердің қамтылуына қарай *біртiлді*, *екiтiлді*, *көптiлді* *iшкорпустар* болып бөлінетіні жоғарыда айтылды. Біртiлді iшкорпустар белгілі бір тілдің материалына негізделсе, екітiлді және көптiлді iшкорпустардың екі түрі болады: *салыстырмалы* және *параллель*.

*Салыстырмалы iшкорпус* әртүрлі тілдердің мәтін, форма, тақырып және т.б. параметрлердің бірнешеуі бойынша ұқсас және салыстыруға жататын мәтіндерінен тұрғандықтан, мұндай iшкорпусқа материал жинау оңай, сондай-ақ бұл iшкорпус түрін жасауда ұлттық корпус материалын пайдалануға болады. Салыстырмалы iшкорпустар, мысалы, шарттық құқық, гендік инженерия сияқты салаларда арнаулы мақсаттарда тілдерді салыстырмалы талдау үшін қолданылған (Lauridsen, 1996).

*Параллель iшкорпустар* негізінен *екiтiлді* және *көптiлді* болып бөлінгенімен, бұл iшкорпустың тілдік диалектілер салыстырылатын *біртiлді* түрі де кездеседі. *Екiтiлді* және *көптiлді* iшкорпустарды негізгі екі түрге бөлуге болады: 1) қандай да бір тілдің жазылған түпнұсқа мәтіндерінен және олардың бір немесе бірнеше тілге аударылған мәтіндерінен тұратын iшкорпустар; 2) екі немесе бірнеше тілде жазылғанына қарамастан, бір ғана тақырыптық саланың мәтіндерін біріктіретін iшкорпустар. *Екiтiлді* iшкорпустарға каталан-испан (<https://catalog.elra.info/en-us/repository/browse/ELRA-W0053/>), хорват ағылшын (<https://www.clarin.si/repository/xmlui/handle/11356/1058>), ағылшын-чех (<https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-1932>) сияқты iшкорпустарды жатқызуға болады. Аталған iшкорпустар бір-біріне жақын, туыстас тілдердің параллель мәтіндерінің негізінде жасалған. Iшкорпустардың базасына екі тілде шығарылатын журналдар, саяси партиялардың дебат материалдары, діни кітаптар мен көркем шығармалар және т.б. мәтіндер алынады. Әдетте параллель iшкорпустардың базасы корпустың мақсатына сай жинақталады. Ал *көптiлді* параллель iшкорпус – iшкорпустың кең тараған түрі саналады. Олардың негізгілерін атап өтсек: Виго университетінің лингвистикалық корпусы (CLUVI) (<http://sli.uvigo.es/CLUVI/index.php?corpus=21&tipo=8&lang=gl>), чех тілінің ұлттық корпусы (UČNK) (<https://treq.korpus.cz/>), орыс тілінің ұлттық корпусы (НКРЯ) (<https://ruscorpora.ru/new/search-para-en.html>). Бүгінгі күні аталған iшкорпустардан басқа да көптеген көптiлді iшкорпустар бар.

«Параллель iшкорпус» ұғымына ғалымдар әртүрлі анықтамалар береді. Солардың бірнешеуін келтіре кетейік: «параллельные корпусы являются своего рода сборниками стратегий и эквивалентов перевода, которыми руководствовались и которые придумывали переводчики» (Захаров, 2020: 48); «параллельный корпус можно определить как корпус, содержащий исходные тексты и их переводы» (Добровольский, 2015: 47); «параллель корпустар – қайсыбір негізгі тілде жазылған түпнұсқа мәтіннің және осы негізгі мәтіннің бір немесе бірнеше басқа тілдерге аударма мәтіндерінің жиынтығының электрондық қоры» (Жұбанов, Жаңабекова, 2017: 117). Көріп отырғанымыздай, берілген анықтамалардан параллель iшкорпустың түпнұсқа мәтін және оның аудармаларының жиынтығы екенін көреміз.

Параллель iшкорпустарды әртүрлі лингвистикалық зерттеулерде (контрастивті лексикология, екітiлді және көптiлді лексикография, терминология, аударма теориясы, лингвомәдениеттану, когнитивті лингвистика), салғастырмалы әдебиеттану (мәтіндердің аудармаларына текстологиялық талдау), мәтінді автоматты тану (компьютерді мәшинелік аударма), әдістеме (шет тілдерін оқыту) және т.б. салалардағы зерттеулерде тиімді қолдануға болады. (Добровольский, 2009; Steinberger, Mohamed, Alexandros, Manuel, Patrick, Przybyszewski,

2014).

Параллель ішкорпусты пайдаланушылардың сөздерді сөздік арқылы іздеп жатпай, мәнмәтіннен тауып, оның басқа сөздермен тіркесімін қарауына, тілдердің грамматикасын, семантикасын және лексикасын салыстырмалы зерттеуге мүмкіндік береді. Алайда ғалымдар параллель ішкорпустың кемшілігі ретінде мәтіндер көлемінің шағындығын және жанрлар бойынша шектеулілігін атайды. Сонымен қатар параллель ішкорпустың сапасына аудармашының қабілеті мен стилі ықпал етеді. (Lauridsen, 1996).

*Параллель ішкорпустардың негізгі түрлері*

*Орыс тілінің ұлттық корпусының параллель ішкорпустары* мынадай тілдерді қамтиды: орыс, ағылшын, неміс, француз, испан, итальян, польша, украин, белорусс, армян, болгар, эстон және латыш. (Добровольский, 2009: 383-401).

*PARASOL корпусы* (PARAllel corpus of Slavic and Other Languages) – Регенсбург университетінде жасалған параллель корпус [Waldenfels 2006].

*InterCorp корпусы* – Чех ұлттық корпусының бір бөлігі ([www.korpus.cz/intercorp](http://www.korpus.cz/intercorp)).

*PARUS корпусы* (PARalelní RUSko-Slovenský korpus) – Штур атындағы Тіл білімі институты жасаған параллель корпус (<http://korpus.sk:8099/manatee.ks/index>).

*ASPAC корпусы* – Амстердам университеті жасаған көркем мәтіндердің және олардың аудармаларының көптілді корпусы (ASPAC – Amsterdam Slavic Parallel Aligned Corpus) (<http://www.uva.nl/over-de-uva/organisatie/medewerkers/content/b/a/a.a.barentsen/a.a.barentsen.html>).

Параллель ішкорпустар *метаақпараттан* (әрбір мәтін туралы ақпарат: қашан жасалды, кім жасады, көлемі және т.б.) және *белгіленімнен* (әрбір сөздің лингвистикалық «төлқұжаты») тұрады.

Параллель ішкорпустардағы белгіленімнің маңызды элементі – түпнұсқа тілдегі әрбір тарауға, абзацқа, сөйлемге аударма тілдегі тараудың, абзацтың, сөйлемның сәйкес келуін *теңестіру*. Осылардың ішінде сөйлем бойынша теңестіру тиімдірек әрі пайдалырақ болады. Бұл жұмыс қиын да ұзаққа созылады әрі қымбатқа түседі, сондықтан әлемде сөйлем деңгейінде теңестірумен жасалған параллель корпустар саны аз.

Кейбір елдердің, мысалы, Орыс тілінің ұлттық корпусында параллель ішкорпустардағы мәтіндерді теңестіру үшін арнаулы бағдарлама жасалған. Алайда мұндай бағдарлама арқылы теңестірудің техникалық тұрғыдан қиындығы да бар. Автоматты түрде теңестіру сөйлем деңгейінде, таза құрылымдық (сөздердің, сөйлемдердің ұзындығы, саны) және статистикалық (сөйлемді құрайтын сөздердің жиілігі) әдістермен жасалатындықтан, теңестіру процесі қайта тексеріліп, «қолмен өңдеу» жүргізіледі. Мысалы, ағылшын және неміс тілдеріндегі корпусты жасаушылар (А.А. Кретов, И.Е. Воронина, Д.Спесивцев) қолданатын «ПарТекс», сондай-ақ басқа параллель корпустарды теңестіру үшін «Евклид» (Д.В. Сичинава, Т.А. Архангельский) бағдарламасы бойынша теңестіру сөйлем деңгейінде жасалады. Теңестірудің айтарлықтай қиындығының бірі – автордың мәтінді сөйлемге және абзацқа бөлудің аудармашының бөлуімен сәйкес келе бермеуі. (Сичинава, 2015). Яғни аудармашы сөйлем шегін ешқашан сақтамайды десе де болады, түпнұсқадағы сөйлем екі немесе бірнеше сөйлеммен және керісінше аударылуы мүмкін.

### **Нәтижелер және талқылау**

ҚТҰК-ның қазақ мәтіндері мен олардың орыс тіліне аударылған нұсқаларының параллель ішкорпустың жасау үшін алғашқы тәжірибе ретінде М.Әуезовтің «Абай жолы» (Әуезов, 2009а; Әуезов, 2009ә; Әуезов, 2009б; Әуезов, 2009в) роман-эпопеясы және оның А.Ким аударған орыс тіліндегі мәтіні алынды (Ауэзов, 2012а; Ауэзов, 2012б; Ауэзов, 2012в; Ауэзов, 2012г).

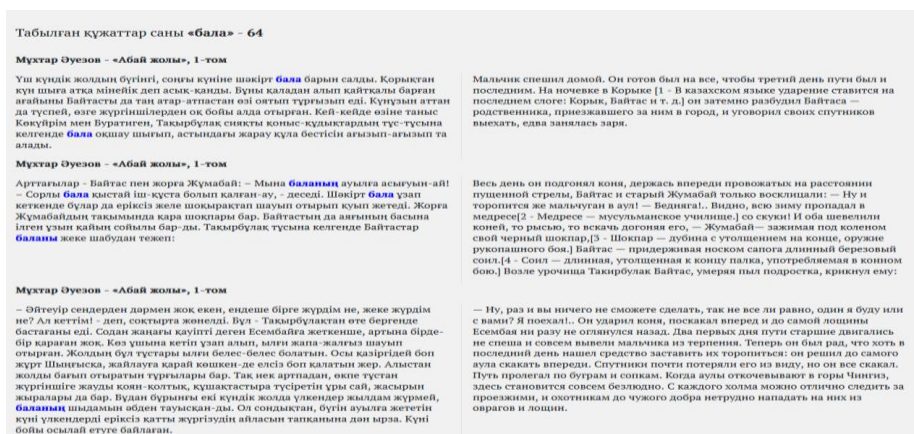
*ҚТҰК-ының параллель ішкорпустың сипаты*

Әзірленіп жатқан параллель ішкорпус теңестірілген мәтіндер базасынан, белгіленім, метабелгіленім және іздеу жүйесінен тұрады. Бұл ішкорпуста енгізілген мәтін туралы ақпарат (метабелгіленім) мынадай параметрлерден тұрады: *автор, аудармашы, шығарма атауы, аударма атауы, шығарманың басылған мерзімі, аударылған мерзімі, түпнұсқа тіл, аударма тіл*. (Автор: М.Әуезов, Аудармашы: А.Ким, Шығарма атауы: Абай жолы, Аударма атауы: Путь Абая, Мерзімі: 1942, 1947, 1952, 1956, 2009, 2013 жыл, Аударылған мерзімі: 2012 жыл, Түпнұсқа тіл: Қазақ, Аударма тіл: Орыс). Іздеу жүйесі қандай да бір сөзді *сөз, сөз тіркес, сөйлем, бас әріп* (қазақ және орыс тіліндегі) параметрлері арқылы табуға мүмкіндік береді.



1-сурет – Параллель ішкорпустаң интерфейсі

1-суретте қазақ және орыс тіліндегі параллель ішкорпустаң интерфейсі көрсетілді. Парақшада көрсетілген іздеу параметрлерінің бірі арқылы қажетті сөзді іздегеннен кейін нәтижелер интерфейсі ашылады.



2-сурет – Параллель ішкорпустаң сөз іздеу нәтижесінің интерфейсі

2-суретте параллель ішкорпустаң «бала» сөзін іздегеннен кейінгі іздеу нәтижесі көрсетілген. Мұнда кестенің сол жағында түпнұсқа, ал оң жағында аударма мәтін орналасқан. Ішкорпустаң іздеген «бала» сөзі көк бояумен ерекшеленген. Нәтижелер интерфейсіндегі мәтіндерді «көшіріп алу» функциясы жұмыс істейді.

М.Әуезовтің «Абай жолы» роман-эпопеясының қазақ тіліндегі түпнұсқа мәтіні мен оның А.Ким жасаған орыс тіліндегі мәтінінің аудармасы параллель ішкорпуста абзац (сөйлем) деңгейінде **қолмен теңестірілді**. Мұндай қолмен теңестіру жұмыс күшін және уақытты талап еткенімен, кейін параллель ішкорпустаң базасын жинақтап, өңдеуге арналған жасанды интеллектіні үйретуге негіз болып табылады.

Бұл параллель ішкорпустаң әзірлеу барысында Microsoft Office Excel, Notepad++, Python, Django, MySQL бағдарламалық құралдары қолданылды.

Параллель ішкорпустаң бағдарламалық архитектурасын және жұмыс істеу тәртібін төмендегіше көрсетуге болады (3-сурет):

1) екі тілдегі мәтіндер Excel офистік бағдарламасы арқылы жинақталып, абзац (сөйлем) деңгейінде қолмен теңестірілді;

2) теңестірілген мәтіндер Excel файлынан тікелей MySQL мәліметтер қорын басқару жүйесіне жүктелді;

3) жүктелген мәтіндер Notepad++ мәтіндік процессор бағдарламасы арқылы сұрыпталып, олардың статистикасы алынды;

4) сұрыпталған мәтіндерді интернетке жариялау және пайдаланушының сұрауларын қамтамасыз ету үшін Django веб сервері қолданылды;

5) Django веб серверін MySQL мәліметтер қорын басқару жүйесімен байланыстыру үшін Python тілінде жазылған, іздеу функциясымен қамтамасыз етілген Processing.py бағдарламасы қолданылды;

6) паралель ішкорпустың бағдарламалық архитектурасы client-server және MVC (Model-View-Controller) технологиялары қолданыла отырып әзірленді.



**3-сурет** – Паралель ішкорпустың бағдарламалық архитектурасы

#### *Паралель ішкорпус бойынша статистикалық мәлімет*

Паралель ішкорпусқа енгізілген «Абай жолы» роман-эпопеясының мәтінде қолданылған жалпы сөз саны: қазақ тілінде – 482150, орыс тілінде – 445491; қайталанбайтын сөз саны: қазақ тілінде – 59816, орыс тілінде – 59952, сөйлем саны: қазақ тілінде – 41798, орыс тілінде – 37792 екені анықталды. Екі тілдегі мәтінде ең жиі қолданылған сөзқолданыстың ондығына *Абай* сөзі кіреді: қазақ тіліндегі мәтінде қолданылу жиілігі бойынша 4-орында, орыс тіліндегі мәтінде – 10-орында.

**1-кесте** – «Абай жолы» роман-эпопеясының қазақ тіліндегі түпнұсқа мәтінде және орыс тіліндегі аударма мәтінде жиі қолданылған он сөзқолданыстың сандық және пайыздық көрсеткіші

Қазақ тіліндегі мәтінде			Аударма мәтінде		
Сөз	Кездесу жиілігі	%	Сөз	Кездесу жиілігі	%
да	5682	1,18	и	18356	4,12
деп	4554	0,94	в	10769	2,42
де	4487	0,93	не	8408	1,89
Абай	4091	0,85	с	7907	1,78
бір	3937	0,82	что	5882	1,32
осы	3490	0,72	он	4530	1,02
мен	3147	0,65	его	4065	0,91
бар	3039	0,63	к	3694	0,83
бұл	2872	0,60	по	3298	0,74
жоқ	2417	0,50	Абай	2820	0,63

Роман-эпопеяның қазақ тіліндегі түпнұсқа мәтінін орысша аударма нұсқасымен теңестіру барысында мынадай ерекшеліктер байқалды:

1. Роман-эпопеяның қазақ тіліндегі түпнұсқа мәтін мен орыс тіліндегі аударма мәтіндегі абзацтар (сөйлемдер) тұрпат межесі тұрғысынан, яғни абзацта (сөйлемде) қолданылған сөздер сан жағынан шамамен сәйкес келеді. Мысалы:

1) – *Ойбай, бізде не қауқар бар? Біз екеу-ақ... ± – Ойбой! А что мы сможем с ними сделать? Нас двое...*

2) – *Әй, балам-ау, мына жер – жау жатағы. Бұл ұрының ойнағына кеп алып, жаман ырым бастағаның не қылғаның?- деді. ± – Ой, сынок, накликаешь ты беду своими шутками! Нашел место – в самом воровском логове, – сказал он с досадой.*

3) – *Е, тастай алмай жүрген жаным жоқ! Терге демесем, жанымды ала зой деп кепілге келіп пе екемін,- деп томсарып қалды. ± – У меня тоже нет души, которую некуда бы девать! Я сказал лишь, что надо проверить. Не для того я приехал, чтобы заложить свою душу, – проговорил он мрачно.*

2. Роман-эпопеяның қазақ тіліндегі түпнұсқа мәтіні мен орыс тіліндегі аударма мәтіндегі абзацтар (сөйлемдер) мазмұн межесі тұрғысынан шамамен сәйкес келеді. Мысалы:

1) *Баланың қытығына тиетіні осы жері. ± – Но эти слова только подзадорили мальчика.*

2) *Содан жаңағы қауіпті деген Есембайға жеткенше, артына бірде-бір қараған жоқ. Көз ұшына кетіп ұзап алып, ылғи жапа-жалғыз шауып отырған. ± – Он ударил коня, поскакал вперед и до самой лоцины Есембая ни разу не оглянулся назад.*

3) – *Одан да, не күтсеңдер де осы жаман қарадан күтсеңдерші! - деді. ± – Если уж чего-нибудь ждатель – так ждите только от этого черномазого мальчугана!*

4) *Кірпігін де сирек қағады. ± Он даже моргает редко.*

5) – *Ендеше, өлімнен ұят күшті. Ел көрмеген сұмдыққа, ел көрмеген жаза керек! – деп Құнанбай байлауын айтты. ± – Честь – выше смерти. Беспрецедентный грех должен получить и беспрецедентное возмездие, – заключил Кунанбай.*

3. Роман-эпопеяның қазақ тіліндегі түпнұсқа мәтіні мен орыс тіліндегі аударма мәтіндегі абзацтар (сөйлемдер) тұрпат және мазмұн межесі тұрғысынан сәйкес келмейді: қазақ тіліндегі түпнұсқа мәтіндегі кейбір абзацтар (сөйлемдер) орыс тіліне бұрыс немесе үстүрт, қысқа аударылған, олардың жуық мағынасы берілген. Мысалы:

1) *Үй іші Абай айтқан жаңалықтарды қызыға тыңдап, көңілденіп қап еді. ≠ Старшие чувствовали себя неловко: они были смущены своим незнанием.*

2) *Бұл үлкендер осы түнде қырық-елудей жігіт-желеңді және бірер ақылшы кәріні қастарына алып, өліктің артын күтетін қамға кірісті. ≠ – Они создали на совет еще несколько стариков.*

3) *Оспан мен Смағұл Тәкежан қасына қатар жалынып, жүгіріп барып еді. ≠ – Оспан и Смағұл продолжали приставать к нему.*

4) *Жұрттың бәрі амалсыз иіріліп тұра қалып, ақ таймен Оспанның арпалыс майданына қарады. ≠ Все невольно залюбовались необыкновенным единоборством.*

5) *Абай қып-қызыл боп кетіп, шұғыл ойланып отырып қалды. ≠ – Абай вспыхнул.*

6) *Байдалы мына сөзге әбден қанағат қылды. Бала көргені – бала емес тәрізді. ≠ Его слова понравились Байдалы.*

7) *«Ақ етің аппақ екен атқан таңдай!» дегеннен басқасы есіне әзір қайта түспеді. ≠ В памяти живет только одна строка: Ты встаешь а моем сердце, рассвет любви.*

8) *Шолпысының ендігі сылдырынан өзі қысылған тәрізді. ≠ Казалось, что каждый шаг ее звенит серебряными переливами дорожного украшения.*

9) *Ырғызбай ішінен, әсіресе, осы Қарашоқыға қызығушылар көп болатын. Бұнда отырған ауылдар Борсақ, Жексен ауылы. Ол, Борсаққа әлдеқандай көрінгенмен, өзге жұрттың келесінде бұта құрым болатын. ≠ У многих из рода Иргизбай давно уже глаза разгорались на Карашоқы, где находился аул Жексена.*

10) *Абай да, Жұмабай да қызығып тыңдасты. Өзі сері, әнші, сұлу Байтасқа екеуі де – бірі кәрі, бірі жас бала – қызғана да, тамашалай да қараған еді. ≠ Байтас был признанным певцом и красавцем. Абай посмотрел на него с нескрываемым восхищением.*

11) *Үлкендер бала мінезіне бір күліп, бір таңданып сүйсініп қалды. Қара сұр бала көптің көзі бір өзіне қадалғанда енді қысылыңқырап, қызарып еді. Бірақ, көзінде жайнай жанған от білінеді. Өзге балаларының ажарынан бөлек, қызулы, саналы оты бар сияқты. ≠ Шутка мальчугана вызвала и смех и восхищение взрослых.*

12) – *Бәсе, Шанишар? ≠ – Ну, конечно, в нем течет кровь Шанишар.*

Мұндай мысалдарды көптеп келтіруге болады. Алайда мұндай «ерекшеліктер» бір тілдегі көркем әдеби туындыны құрылымы өзгеше екінші бір тілге аудару кезінде қолданылатын оңайлату (аударма мәтін орташа алғанда түпнұсқаға қарағанда лексикасы және синтаксисі жағынан «қарапайымдандырылады»), нормаландыру (аудармашы аударма тілдің нормаларына сай тілдік бірліктерді қолданумен қатар шығыс және аударылатын тілдің иесі халықтың мәдени ерекшеліктеріне байланысты «өзгерістер енгізеді», «трансформация жасайды»), симпликация (лексикалық (аз сөз қолдану), синтаксистік (сөйлемнің синтаксистік құрылымын оңайлату) және стилистикалық (фразеологиялық оралымдарды жай сөз тіркестерімен беру және т.б.; түпнұсқа мәтіндегі ұзын және күрделі сөйлемдер аударма тілде қысқа және жай сөйлемдерге бөлінеді) (Тао,



Захаров, 2015), *эксплицитация* (*кеңейту* (addition) – аударма мәтінге жаңа элементтер қосу (сипаттау, күшейту және т.б.). және *нақтылау* (specification) – түпнұсқа мәтіндегі имплицитті ақпараттың аударма мәтінде эксплицитті түрде берілуі, бұл түпнұсқа мәтінмен салыстырғанда аударма мәтіннің көлемінің ұлғаюына алып келеді) (Englund-Dimitrova, 1999) тәсілдерінің бірін қолдануға және т.б. себептерге байланысты болады.

М.Әуезовтің «Абай жолы» роман-эпопеясының орыс тіліне аударылу мәселелерін әдебиеттану теориясы тұрғысынан зерттеген ғалымдардың қатарынан Г.Бельгер, М.М. Әуезов, Қ.Алпысбаев, А.Анастасьев, С.Ананьева, А.Жақсылықов, Н.Сағындықова, Г.Қазақбаева, Г.Қазыбек, Ж.Каракузова, Б.Канапьянов, С.Елеуенов, Г.Болатова және т.б. атауға болады. Мысалы, көптеген зерттеушілер М.Әуезовтің «Абай жолы» роман-эпопеясының А.Ким жасаған аудармасына оң баға береді (Анастасьев, 2007; Куттыкадам, 2007; Ананьева, 2009).

Көрнекті аудармашы Г.Бельгер «Абай жолы» роман-эпопеясының орыс тіліндегі аудармаларына (А.Никольская, Л.Соболев, Т.Нуртазин, Я.Смеяков, В.Бугаевский, А.Жовтис, З.Кедрина) талдау жасай келе, А.Кимнің аудармасына қатысты: мәтіннің орыс тіліне кейде өлең жолдарымен аударылған тұстары байқалады; қазақтың сөзсаптамасына тән интонация, салыстыру, мақал-мәтелдерді, ұйқасты сөз орамдарын және т.б. жақсы түсініп, көбінесе дәл бере алған; аудармада солғын көпсөзділікке ұрынған тұстар кездеседі; қазақ халқының шешендік сөз өнеріне қатысты қолданыстар көбінесе «краснобай» («сөзуар») сөзімен аударылған; кейде аудармада орыс тілінде орашалақ, ыңғайсыз, өңсіз «сипаттағы» сөздер қолданылған; кейбір сөздер барабар баламасымен аударылмаған; аудармада тұрмыстық, этнографиялық сипаттағы «ағаттықтар» кездеседі; ономастикалық атаулардың транслитерациялануында сәйкессіздіктер кездеседі (*Үдері – Одели, Барлыбай – Барылбай, Алмағамбет – Алмагабет, Тайлақбай – Гайлакбай, Жаманқарин – Жаман-Гарин, Жаман-гарин; Семей – Семипалатингск*); түпнұсқада берілмеген жекелеген біртепті құрылымдар көзге түседі; кейбір сөздер дұрыс аударылмаған (*томага – подстава*); Мағрипаның ішкі көңілі сипатталатын бір бет аударылмай кеткен (Бельгер, 2009: 130-142) деген пікір айта отырып, мынадай тұжырым жасайды: «Надеюсь, новый перевод «Пути Абая», осуществленный А.Кимом, вдохновит переводчиков этого творения на иностранные языки» (Бельгер, 2009: 143).

Ғалым А.Жақсылықов роман-эпопеяның А.Ким жасаған аудармасы туралы: «стилі толық, экспрессивті бояуы жақсы берілген» деп баға бере келе, мынадай кемшіліктерді топтап көрсетеді: «а) аудармашы қазақ халқының ұлттық менталитетін өрнектейтін этномәдени және этнопсихологиялық алғышарттарды ескермейді (мысалы, жылқының жасына, түсіне, жынысына қатысты сөздер); ә) қазақ ру-тайпаларының шежіресін, құрылымы мен қатынастарын шатастырып алады, б) аудармашы ХІХ ғасырдағы қазақ қыздарының лексикасында модернизмдерді (сленглер мен урбанизмдерді) қолданады (Жақсылықов, 2009: 167-178).

Зерттеуші Г.Болатова М.Әуезовтің «Абай жолы» роман-эпопеясының А.Ким жасаған аудармасына қатысты зерттеулерге шолу жасай келе: «М.О.Әуезовтің роман-эпопеясын орыс тілінде толығымен қайта жаңғырту мәселесі әлі де болса түбегейлі шешімін тапқан жоқ. Осылайша, роман-эпопеяның орыс тіліндегі аудармалары әлі күнге сапа жағынан жетілдіру қажеттігін талап етеді. Демек, оның басқа тілдердегі аудармалары да толыққанды талдау жасауды қажет етеді, өйткені басқа тілдерге орыс тілінен аударылғаны белгілі», - деген тұжырым жасайды (Болатова, 2017: 323).

Осы орайда «Абай жолы» роман-эпопеясының қазақ тіліндегі түпнұсқа мәтіні мен орыс тіліндегі аудармасын теңестіру барысында төмендегідей жайттардың да байқалғанын атап өткізіміз келеді:

1. Роман-эпопеяның қазақ тіліндегі түпнұсқа мәтіндегі лексикалық бірліктер орыс тіліне қате аударылған. Мысалы:

1) *Өзі Абайдың 16 жасында туып ерте ересек боп қалған соң және ата қолында оңаша үйде жалғыз өскен кенже есепті болған соң. ≠ Так как Абай был старше своего первенца всего на семнадцать лет, и тот рос с младенческого возраста в доме деда, Акылбай не признавал отца и не испытывал к нему сыновних чувств.*

2) *Қазір жасқа толған, нұрлы көздері көп уақыт кірпік қақпай, жігітке ұзақ тұнжырап, қарап қалды. ≠ Она отошла к окну и долго вглядывалась в темноту наступившей ночи глазами, полными слез...*

2. Роман-эпопеяның орыс тіліндегі аудармасында сөз ортасындағы қысаң «ы» әрпі түсіріліп жазылған. Мысалы, *ауыл (қаз) – аул (орыс.), атрап, баур (қаз) және т.б.*

3. Роман-эпопеяның орыс тіліндегі аудармасында топонимдерді, этнонимдерді, антропонимдерді транслитерациялауда әртүрлі сәйкессіздіктер кездеседі: емле қателері, ру атаулары қазақ тіліндегі түпнұсқа мәтінде бас әріппен, орыс тіліндегі аударма мәтінде кіші әріппен беріледі, бір ғана атау орыс тіліндегі аударма мәтінде бірнеше түрде жазылады. Мысалы, *Керімбала – Коримбала, Есембай – Осембай, Көжекбай – Кожекбай, Ойке – Ойкал, Керей – керей, Уақ – уаки, Қорық – Корых, Корык және т.б.*

Осы ретте академик-жазушы М.Әуезовтің әлемдік деңгейдегі «Абай жолы» роман-эпопеясының қазақ тіліндегі түпнұсқа мәтіні мен оның А.Ким жасаған орыс тіліндегі аударма мәтінін параллель ішкорпусқа енгізу барысында қолданылған теңестіру әдісі арқылы негізінен екі тілдегі мәтіннің абзац (сөйлем) деңгейінде мазмұндық және құрылымдық тұрғыдан сәйкес келуі басшылыққа алынғандықтан, роман-эпопеяның аудармасының сапасын анықтау, яғни лингвистикалық (стилистикалық, орфографиялық және т.б.), әдеби (көркемдік) және т.б. талдау мақсаты қойылмағанын атап өтеміз.

### **Қорытынды**

Қазақ тілінің ұлттық корпусының құрамында әзірленіп жатқан М.Әуезовтің «Абай жолы» роман-эпопеясының қазақ тіліндегі түпнұсқа мәтіні мен оның А.Ким жасаған орыс тіліндегі мәтінінің аудармасы параллель ішкорпуста абзац (сөйлем) деңгейінде қолмен теңестірілді. Сонымен қатар параллель ішкорпусқа Ж.Аймауытовтың «Ақбілек», Т.Ахтанов «Боран», С.Мұқановтың «Мөлдір махаббат», Ш.Айтматовтың «Боранды бекет» романының, С.Елубаевтың «Ақ боз үй» роман-трилогиясының, С.Мұратбековтің «Жусан исі» повесінің, М.Жұмабаевтың «Шолпанның күнәсі» әңгімесінің және т.б. көркем шығармалардың қазақ тіліндегі түпнұсқа мәтіндері мен орыс тіліндегі аударма мәтіндерін теңестіру қолға алынып отыр.

Параллель ішкорпустың бағдарламалық архитектурасын және жұмыс істеу тәртібін төмендегіше көрсетуге болады: 1) екі тілдегі мәтіндер Excel офистік бағдарламасы арқылы жинақталып, абзац (сөйлем) деңгейінде қолмен теңестірілді; 2) теңестірілген мәтіндер Excel файлынан тікелей MySQL мәліметтер қорын басқару жүйесіне жүктелді; 3) жүктелген мәтіндер Notepad++ мәтіндік процессор бағдарламасы арқылы сұрыпталып, олардың статистикасы алынды; 4) сұрыпталған мәтіндерді интернетке жариялау және пайдаланушының сұрауларын қамтамасыз ету үшін Django веб сервері қолданылды; 5) Django веб серверін MySQL мәліметтер қорын басқару жүйесімен байланыстыру үшін Python тілінде жазылған, іздеу функциясымен қамтамасыз етілген Processing.py бағдарламасы қолданылды; 6) параллель ішкорпустың бағдарламалық архитектурасы client-server және MVC (Model-View-Controller) технологиялары қолданыла отырып әзірленді.

Қазақ және тілдерінің орыс параллель ішкорпусы теңестірілген мәтіндер базасынан, белгіленім, метабелгіленім және іздеу жүйесінен және ішкорпусқа енгізілген мәтін туралы ақпарат (метабелгіленім) мынадай параметрлерден тұрады: *автор, аудармашы, шығарма атауы, аударма атауы, шығарманың басылған мерзімі, аударылған мерзімі, түпнұсқа тіл, аударма тіл.* Іздеу жүйесі қандай да бір сөзді сөз, сөз тіркес, сөйлем, бас әріп (қазақ және орыс тіліндегі) параметрлері арқылы табуға мүмкіндік береді.

Әзірленген қазақ және орыс тілдерінің параллель ішкорпусы қазақ тілі мен өзге тілдердің теңестірілген мәтіндері арқылы қазақ тілінің қолданысын кеңейтетін цифрлық платформаның бір үлгісі, тіл білімі, әдебиеттану, әдістеме және т.б. салалардағы зерттеулерге қажетті мәтіндер базасы бола алады. Сонымен қатар бұл параллель ішкорпус қазіргі кезде көптеген тілдердегі түрлі стильдегі мәтіндерді қамтитын <https://opus.nlpl.eu/> және т.б. платформалар сияқты түркі тілдерінің мәтіндерін салыстыруға арналған осындай платформаларды, яғни параллель ішкорпустарды жасаудың бастамы деп санаймыз.

## Әдебиеттер

- Svartvik J., Quirk R. (1980) A corpus of English Conversation. – Lund: Gleerup, 1980. – 284 p.
- Francis W. (2022) Brown Corpus Manual: Manual of Information to Accompany a Standard Corpus of Present-Day Edited American English for Use with Digital Computers. [Электрон. ресурс] – URL: <http://icame.uib.no/brown/bcm.html> (қаралған күні – 01.02.2022).
- Hundt, Marianne. (2022) Manual of Information to Accompany the Freiburg-Brown Corpus of American English (FROWN). [Электрон. ресурс] – URL: <http://khnt.hit.uib.no/icame/manuals/frown/INDEX.HTM> (қаралған күні – 01.02.2022).
- Leech C. (2005) Extending the possibilities of corpus-based research on English in the twentieth century: A prequel to LOB and FLOB. ICAME Journal. – Geoffrey & Nicholas Smith, 2005. № 29. – P. 83-98.
- Жұбанов А., Жаңабекова А. (2017) Корпустық лингвистика. – Алматы, 2017. – 318 б.
- Захаров В.П. (2005) Корпусная лингвистика: Учебн.-метод. пособие. – СПб., 2005. – 48 с.
- Жұбанов А.Қ. (2009) Қазақ тілі мәтіндер корпусының компьютерлік базасын құрудың алғышарттары // Әдеби тіл және қазақ тілінің өміршеңдігі. / Жауапты ред. Қ.Күдерінова. – Алматы: «КИЕ» лингвоелтану инновациялық орталығы, 2009. –175-179-бб.
- Жаңабекова А.Ә. (2012) Қазақ тілі мәтіндеріне морфологиялық белгіленім қоюдың ғылыми-тәжірибелік негіздері // Научно-практическая конференция «Языки меньшинств в компьютерных технологиях: опыт, задачи и перспективы». – Уфа, 2012. – Б. 42-50.
- Жубанов А.К. (2015) Принципы автоматизации морфологической разметки текстов Национального корпуса казахского языка (НККЯ) // Материалы международной научно-практической конференции «Контрастивные исследования и прикладная лингвистика». – Минск: МГЛУ, 2015. – С. 111-119.
- Жанабекова А., Пирманова К. (2019) Технология программы полуавтоматической метаразметки казахского национального корпуса // Актуальные проблемы современной лингвистики и гуманитарных наук: сборник статей XI Всероссийской научно-методической конференции с международным участием. – М.: РУДН, 2019. – С. 465-474.
- Құлманов С., Жаңабекова А.Ә., Әшімбаева Н.М., Бисенғали А.З., Шүленбаев Н.Қ., Қордабай Б.Қ. (2022) Корпусқа енгізілетін мәтіндердегі сөздерге морфологиялық белгіленім қою және оларды компьютерлік бағдарламаға енгізу мәселелері // Л.Н. Гумилев атындағы Еуразия ұлттық университетінің Хабаршысы. «Филология ғылымдары» сериясы. № 3(140)/2022. – 103-113-бет (DOI: <https://doi.org/10.32523/2616-678X-2022-140-3-103-113>) <https://bulphil.enu.kz/index.php/main/article/view/29/9>.
- Койбағаров К.Ч., Мусабаяв Р.Р., Құлманов С.К. (2012) Разработка алгоритмов автоматического анализа словоформ казахского языка // «Қазіргі қазақ тіл білімі: қолданбалы лингвистиканың өзекті мәселелері»: халықаралық ғылыми-теориялық конференция материалдарының жинағы. – Алматы, А.Байтұрсынұлы атындағы Тіл білімі институты, 2012. – С.83-87.
- Lauridsen, Karen. (1996) Text Corpora and Contrastive Linguistics: Which Type of Corpus for which Type of Analysis? In: Aijmer, Karin /Altenberg, Bengt /Johansson, Mark (eds) Languages in Contrast. Papers from a Symposium on Text-based Cross Linguistic Studies. Lund: Lund University Press. – 1996. –P. 63-72.
- Захаров В.П. (2020) Корпусная лингвистика. – Петербург, 2020. – 234 с.
- Добровольский Д.О. (2015) Лингвоспецифичная лексика в корпусах параллельных текстов / Д.О.Добровольский // Речевые жанры современного общения. Тезисы докладов междунар. конф. «11-е Шмелевские чтения (23-25 февр. 2015)». – М.: ИРЯ РАН, 2015. – С. 47-49.
- Добровольский Д.О. (2009) Корпус параллельных текстов в исследовании культурно-специфичной лексики // Национальный корпус русского языка: 2006-2008. Новые результаты и перспективы. – СПб.: Нестор-История, 2009. – С. 383-401.
- Steinberger Ralf, Mohamed Ebrahim, Alexandros Poulis, Manuel Carrasco-Benitez, Patrick Schlüter, Marek Przybyszewski & Signe Gilbro. (2014) An overview of the European Union's highly multilingual parallel corpora. Language Resources and Evaluation Journal (LRE) 2014. DOI: 10.1007/s10579-014-9277-0.
- Waldenfels R. (2006) Compiling a parallel corpus of Slavic languages. Text strategies, tools and the question of lemmatization in alignment // Beitrage der Europäischen Slavistischen Linguistik (POLYSLAV). 9. Munchen, 2006. –S. 123-138.
- Сичинава Д.В. (2015) Параллельные тексты в составе национального корпуса русского языка: новые направления развития и результаты // Труды Института русского языка РАН. – М., 2015. – С.194-234.
- Тао Ю., Захаров В.П. (2015) Разработка и использование параллельного корпуса русского и китайского языков // НТИ. Сер. 2. Информ. Процессы и системы. 2015. №4. – С. 18-27.
- Әуезов М. (2009а) Абай жолы: Роман-эпопея. Бірінші кітап. – Алматы: «Жазушы», 2009. – 376 бет. – «Қазақтың 100 романы» сериясы
- Әуезов М. (2009ә) Абай жолы: Роман-эпопея. Екінші кітап. – Алматы: «Жазушы», 2009. – 432 бет. – «Қазақтың 100 романы» сериясы
- Әуезов М. (2009б) Абай жолы: Роман-эпопея. Үшінші кітап. – Алматы: «Жазушы», 2009. – 384 бет. – «Қазақтың 100 романы» сериясы
- Әуезов М. (2009в) Абай жолы: Роман-эпопея. Төртінші кітап. – Алматы: «Жазушы», 2009. – 400 бет. – «Қазақтың 100 романы» сериясы
- Ауэзов М. (2012а) – Путь Абая. / Перевод А. Кима. – Алматы: ИД «Жибек жолы», 2012. Кн. 1. – 568 с.
- Ауэзов М. (2012б) – Путь Абая. / Перевод А. Кима. – Алматы: ИД «Жибек жолы», 2012. Кн. 2. – 556 с.
- Ауэзов М. (2012в) – Путь Абая. / Перевод А. Кима. – Алматы: ИД «Жибек жолы», 2012. Кн. 3. – 504 с.

- Ауэзов М. (2012) – Путь Абая. / Перевод А. Кима. – Алматы: ИД «Жибек жолы», 2012. Кн. 4. – 528 с.
- Englund-Dimitrova В. (1999) Tolkens roll: perspektiv från nyare forskning och implikationer för tolkutbildning // Материалы Научно-практического семинара по преподаванию устного перевода (шведский язык/русский язык). – М., 1999. – С. 36-47.
- Анастасьев Н. (2007) Третье свидание // Эдебиег айдыны. 2007. – № 44 (141). – С. 9.
- Куттыкадам С. (2007) Заветы Мухтара // Мысль. – 2007. – № 11. – С. 2.
- Ананьева С. (2009) «Путь Абая» М.О. Ауэзова в переводе А. Кима // Керуен. – 2009. – № 2. – С. 136-157.
- Бельгер Г. (2009) Жизнь – эпопея (эссе-триптих, статьи). – Алматы: ИД «Жибек жолы», 2009. – 144 с.
- Жаксылыков А.Ж. (2013) Особенности нового перевода романа-эпопеи М.О. Ауэзова «Путь Абая» на русский язык // Художественный перевод и литературный процесс (избранные лекции и исследования). – Алматы, 2013. – С. 167-178.
- Болатова Г.Ж. (2017) Жаңа «Абай жолы»: А. Кимнің аудармасындағы кейбір ерекшеліктер // ҚазҰУ Хабаршысы. Филология сериясы. №2 (166). 2017. – 320-324-бб. (<https://philart.kaznu.kz/index.php/1-FIL/article/view/2357/2262>)

### References

- Svartvik J., Quirk R. (1980) A corpus of English Conversation. – Lund: Gleerup, 1980. – 284 p. (in English)
- Francis W. (2022) Brown Corpus Manual: Manual of Information to Accompany a Standard Corpus of Present-Day Edited American English for Use with Digital Computers. [Electron. resource] – URL: <http://icame.uib.no/brown/bcm.html> (date of review – 01.02.2022). (in English)
- Hundt, Marianne.(2022) Manual of Information to Accompany the Freiburg-Brown Corpus of American English (FROWN). [Electron. resource] – URL: <http://khnt.hit.uib.no/icame/manuals/frown/INDEX.HTM> (date of review – 01.02.2022). (in English)
- Leech C. (2005) Extending the possibilities of corpus-based research on English in the twentieth century: A prequel to LOB and FLOB. ICAME Journal. – Geoffrey & Nicholas Smith, 2005. № 29. – P. 83-98. (in English)
- Zhubanov A., Zhanabekova A. (2017) Korpustyq lingvistika. [Corpus Linguistics] – Almaty, 2017. – 318 b. (in Kazakh)
- Zakharov V.P. (2005) Korpustnaia lingvistika: Uchebn.-metod. posobie. [Corpus linguistics: Textbook.-method. stipend]. – SPb., 2005. – 48 s. (in Russian)
- Zhubanov A.Q.(2009) Qazaq tili matinder korpustynyn kompiuterlik bazasyn qurudyn algysharttary // Adebi til zhane qazaq tilinin omirshendigi. [Prerequisites for creating a computer base of the corpus of texts of the Kazakh language] / Zhauapty red. Q.Kuderinova. – Almaty: «KIE» lingvoeltanu innovatsialyq ortalygy, 2009. –175-179-bb. (in Kazakh)
- Zhanabekova A.A. (2012) Qazaq tili matinderine morpologikalыq belgilenim qoiudyn gylыmyi-tazhiribelik negizderi [Scientific and practical foundations of morphological designations of texts of the Kazakh language] // Nauchno-praktisheskaia konferensia «Iazyki menshenstv v compiuternыq tehnologiah: opyt, zadachi i perspektivi». – Ufa, 2012. – B. 42-50. (in Kazakh)
- Zhubanov A.K.(2015) Prinsipy avtomatizatsi morpologicheskoi razmetki tekstov Natsionalnogo korpusta kazaqskogo iazyka (NKKIA) [Principles of automation of morphological markup of texts of the National corpus of the Kazakh language (NKKYA)] // Materialy mezhdunarodnoi nauchno-praktisheskoi konferensii «Kontrastivnye issledovania I prikladnaia lingvistika». – Minsk: MGLU, 2015. – Ss. 111-119. (in Russian)
- Zhanabekova A., Pirmanova K. (2019) Tehnologia programmy poluavtomatisheskoi metarazmetki kazaqskogo natsionalnogo korpusta [Technology of the program of semi-automatic meta-marking of the Kazakh national corpus] // Aktualnye problem sovremennoi lingvistiki i gumanitarnыq nauk: sbornik statei XI Vserossooskoi nauchno-metodicheskoi konferensii s mezhdunarodnym uchastiem. – M.: RUDN, 2019. – S. 465-474. (in Russian)
- Qulmanov S.Q., Zhanabekova A.A., Ashimbayeva N.M., Bisengali A.Z., Shulenbayev N.K., Kordabay B.K. (2022) Korpustqa engiziletin matinderdegi sozderge morpologialыq belgilenim qoiyu zhane olardy kompiuterlik bagdarlamaga engizu maseleleri [Problems of morphological designations of words in texts included in the Corps and their implementation in a computer program] // L.N. Gumilev atyndagy Evrasia ultтыq universitetinin Qabarshysy. «Filologia gylыmdary» seriasy № 3(140)/2022. – 103-113-bet (DOI: <https://doi.org/10.32523/2616-678X-2022-140-3-103-113>) (<https://bulphil.enu.kz/index.php/main/article/view/29/9>) (in Kazakh)
- Koibagarov K.Ch., Musabaev R.R., Kulmanov S.K. (2012) Razrabotka algoritmov avtomatisheskogo analiza slovoform kazaqskogo iazyka [Development of algorithms for automatic analysis of word forms of the Kazakh language] // «Qazirgi qazaq til bilimi: qoldanbaly lingvistikanyn ozekti maseleleri»: qalyqaralyq gylыmyi-toerialыq konferensia materialdatynyn zhyinagy. – Almaty, A. Baitursynuly atyndagy Til bilimi instituty, 2012. –S. 83-87. (in Russian)
- Lauridsen, Karen (1996) Text Corpora and Contrastive Linguistics: Which Type of Corpus for which Type of Analysis? In: Aijmer, Karin /Altenberg, Bengt /Johansson, Mark (eds) Languages in Contrast. Papers from a Symposium on Text-based Cross Linguistic Studies. Lund: Lund University Press. – 1996. –P. 63-72. (in English)
- Zakharov V.P. (2020) Korpusthaia lingvistika. [Corpus linguistics] – Peterburg, 2020. – 234 s. (in Russian)
- Dobrovolsky D.O. (2015) Lingvospesifichnaia leksika v korpustaq parallelnыq tekstov / D.O.Dobrovolsky [Linguospecific vocabulary in the corpus of parallel texts] // Rechebye zhanry sobremennogo obshenia. Tezisy dokladov mezhdunar.konf. «11-e Shmelevskie chtenia (23-25 febral 2015)». – M.: IRIA RAN, 2015. – S. 47-49. (in Russian)
- Dobrovolsky D.O. (2009) Korpust parallelnыq tekstov v issledovanii kulturno-spetsifichnoi leksiki [The corpus of parallel texts in the study of culturally specific vocabulary] // Natsionalnyi korpust russkogo iazyka: 2006-2008. Nobye rezultaty i perspektivy. –SPb.: Nestor-Istoria, 2009. – S. 383-401. (in Russian)

Steinberger Ralf, Mohamed Ebrahim, Alexandros Poulis, Manuel Carrasco-Benitez, Patrick Schlüter, Marek Przybyszewski & Signe Gilbro (2014) An overview of the European Union's highly multilingual parallel corpora. Language Resources and Evaluation Journal (LRE) 2014. DOI: 10.1007/s10579-014-9277-0. (in English)

Waldenfels R. (2006) Compiling a parallel corpus of Slavic languages. Text strategies, tools and the question of lemmatization in alignment // Beitrage der Europäischen Slavistischen Linguistik (POLYSLAV). 9. München, 2006. –S. 123-138. (in English)

Sichinava D.V. (2015) Parallelnye teksty v sostave natsionalnogo korpusa russkogo iazyka: novye napravleniya razbitia i rezultaty [Parallel texts as part of the national Corpus of the Russian language: new directions of development and results] // Trudy Instituta russkogo iazyka RAN. – M., 2015. – S.194-234. (in Russian)

Tao Yu., Zakharov V.P. (2015) Razrabotka i ispolzovanie parallelnogo korpusa russkogo i kitaiskogo iazykov [Development and use of a parallel corpus of Russian and Chinese languages] // NTI. Ser. 2. Inform. Prosessy i sisremy. 2015. № 4. – S. 18-27. (in Russian)

Auezov M. (2009a) Abai zholy: *Roman-epopeia*. Birinshi kitap. [The Path of Abai: A novel-epic. The first book] – Almaty: «Zhazushy», 2009. – 376 bet. – «Qazaqtyn 100 romany» seriasy (in Kazakh)

Auezov M. (2009ä) Abai zholy: *Roman-epopeia*. Ekinshi kitap. [The Path of Abai: A novel-epic. The second book] – Almaty: «Zhazushy», 2009. – 432 bet. – «Qazaqtyn 100 romany» seriasy (in Kazakh)

Auezov M. (2009b) Abai zholy: *Roman-epopeia*. Ushinshi kitap. [The Path of Abai: A novel-epic. The third book] – Almaty: «Zhazushy», 2009. – 384 bet. – «Qazaqtyn 100 romany» seriasy (in Kazakh)

Auezov M. (2009v) Abai zholy: *Roman-epopeia*. Tortinshi kitap. [The Path of Abai: A novel-epic. The four book] – Almaty: «Zhazushy», 2009. – 400 bet. – «Qazaqtyn 100 romany» seriasy (in Kazakh)

Auezov M., (2012a) – Put Abaia. [Way Abay] / Perevod A. Kima. – Almaty: ID «Zhibek Zholy», 2012. Kn. 1. – 568 s. (in Russian)

Auezov M., (2012b) – Put Abaia. [Way Abay] / Perevod A. Kima. – Almaty: ID «Zhibek Zholy», 2012. Kn. 2. – 556 s. (in Russian)

Auezov M., (2012v) – Put Abaia. [Way Abay] / Perevod A. Kima. – Almaty: ID «Zhibek Zholy», 2012. Kn. 3. – 504 s. (in Russian)

Englund-Dimitrova B. (1999) Tolkens roll: perspektiv från nyare forskning och implikationer för tolkutbildning // Materials of the Scientific and Practical Seminar on teaching Interpretation (Swedish/Russian). – M., 1999. – C.36-47. (in Swedish)

Anastasiev N.(2007) Trete svidanie [The third date] // Adebiet aidyny. 2007. –№ 44 (141). – S. 9. (in Russian)

Kuttikadam S. (2007) Zabety Muhtara [Kuttikadam S. Mukhtar's Precepts] // Mysl. – 2007. – № 11. – S. 2. (in Russian)

Ananyeva S. (2009) «Put Abaia» M.O.Auezova v perevode A.Kim [«The Way of Abai» by M.O.Auezov translated by A. Kim] // Keruen. – 2009. – № 2. – S. 136-157. (in Russian)

Belger G. (2009) Zhizn – epopeia (esse-triptiq, stati). [Life – epic (essay-triptych, articles)] – Almaty: ID «Zhibek Zholy», 2009. – 144 s. (in Russian)

Zhaksylykov A.Zh. (2013) Osobennosti novogo perevoda romana-epopei M.O.Auezova «Put Abaia» na russkii iazyk [Features of the new translation of M.O.Auezov's epic novel «The Way of Abai» into Russian] (in Russian)// Qudozhestbennyi perevod i literaturnyi prosess (izbrannye leksii i issledovania). – Almaty, 2013. – S. 167-178.

Bolatova G. Zh. (2017) Zhana «Abai zholy»: A.Kimnin audarmasyndagy keibir erekshelikter [New «The Way of Abai»: some features in the translation of A. Kim] // QazUU Habarshysy. Pilologia seriasy. №2 (166). 2017. – 320-324-bb. (<https://philart.kaznu.kz/index.php/1-FIL/article/view/2357/2262>) (in Kazakh)