

МРНТИ 16.01.29

З.А.Сироазитдинов

РҒА УҒО лингвистика және ақпараттық технология институтының меңгерушісі,
филология ғылымдарының кандидаты, аға ғылыми қызметкер

АСҚАР ҚҰДАЙБЕРГЕНҰЛЫ ЖҰБАНОВ – ҒАЛЫМ ЖӘНЕ ТӘЛІМГЕР

Аннотация: Асқар Құдайбергенұлы ХХ ғасырдың 60-жылдары ғылыми шығармашылығын жиілік сөздіктер құрастыру мен соның базасында қазақ мәтіндері құрылымын статистикалық зерттеумен бастады. А.Жұбанов өзінің іргелі зерттеулері (200 ғылыми еңбек, оның ішінде 5 монография, 4 оқу құралы және 5 жиілік сөздік) арқылы қазақ қолданбалы тіл білімінің бірнеше бағыттарының дамуына үлес қосты. Олар: ұлттық мәтін мазмұнын формалдау және квантитативтік-статистикалық зерттеу, жазба мәтінді автоматты талдау, компьютерлік лексикография, қазақ тілінің лингвистикалық деректер базасын әзірлеу және т.б. Асқар Құдайбергенұлы көп жылдар бойы түркі республикаларының тілші-әріптестеріне ақпараттық технологияны меңгеру ісі бойынша кеңес беріп, үнемі қолдау көрсетіп отырды. Қазақ тіліне қатысты жарық көрген жиілік сөздіктердің басым бөлігі А.Қ.Жұбановтың қатысуымен құрастырылғанын атап айту қажет.

Тірек сөздер: компьютерлік лингвистика, лингвистикалық база, лингвистикалық корпус.

З.А.Сироазитдинов

кандидат филологических наук, старший научный сотрудник, заведующий институтом лингвистики и информационных технологий ННЦ РАН

АСКАР КУДАЙБЕРГЕНОВИЧ ЖУБАНОВ - УЧЕНЫЙ И НАСТАВНИК

Аннотация: Асқар Құдайбергенович начал свой научный путь с 60-х годов ХХ века с составлением частотных словарей и статистическим исследованием структуры казахских текстов на его базе. А. Жубанов внес свой вклад в развитие нескольких направлений казахского прикладного языкознания через свои фундаментальные исследования (200 научных трудов, в том числе 5 монографий, 4 учебных пособия и 5 частотных словарей). Это формализация и квантитативно-статистическое исследование содержания национального текста, автоматический анализ письменных текстов, компьютерная лексикография, разработка лингвистической базы данных казахского языка и др. Асқар Құдайбергенович на протяжении многих лет консультировал и поддерживал коллег-корреспондентов тюркских республик по изучению информационных технологий. Следует отметить, что большинство частотных словарей, изданных на казахском языке составлены с участием А. К. Жубанова.

Ключевые слова: компьютерная лингвистика, лингвистическая база, лингвистический корпус

Z.A.Sirazitdinov

PhD in Philology, Senior Researcher, Head of the Institute
of Linguistics and Information Technologies, NSC RAS

**ASKAR KUDAYBERGENOVICH ZHUBANOV – A SCIENTIST
AND TUTOR**

Abstract: The article is devoted to scientific works of my teacher, research supervisor Askar Kudaybergenovich Zhubanov. For many years Askar Kudaybergenuly advised and supported correspondents of the Turkic republics in the field of information technologies. It should be noted that the majority of Kazakhstan frequency word books was compiled with A.K. Zhubanov's participation. The article is devoted to the scientific work of my teacher, scientific advisor Askar Kudaibergenovich Zhubanov. For many years, Askar Kudaibergenuly advised and supported the correspondents of the Turkic republics in the field of information technology. It should be noted that most of the Kazakhstani frequency dictionaries were compiled with the participation of A.K. Zhubanov. Askar Kudaibergenovich began a scientific publication of the 60s of the twentieth century with the compilation of frequency dictionaries and a statistical study of the structure of Kazakh texts on its basis. A. Zhubanov contributed to the development of several directions of

Kazakh applied linguistics through his fundamental research (200 scientific works, including 5 monographs, 4 textbooks and 5 frequency dictionaries). These are formalization and quantitative-statistical research of the content of the National text, automatic analysis of written texts, computer vocabulary, development of a linguistic database of the Kazakh language, etc. Askar Kudaibergenovich for many years consulted and supported fellow correspondents of the Turkic republics on the study of information technologies. It should be noted that most of the frequency dictionaries published in the Kazakh language were compiled with the participation of A.K. Zhubanov.

Keywords: computational linguistics, linguistic base, linguistic cover.

Ұстазым, ғылыми жетекшім Асқар Құдайбергенұлы Жұбановтың ғылыми шығармашылығы әр жақты. А.Жұбанов өзінің іргелі зерттеулері (200 ғылыми еңбек, оның ішінде 5 монография, 4 оқу құралы және 5 жиілік сөздік) арқылы қазақ қолданбалы тіл білімінің бірнеше бағыттарының дамуына үлес қосты. Олар: ұлттық мәтін мазмұнын формалдау және квантитативтік-статистикалық зерттеу, жазба мәтінді автоматты талдау, компьютерлік лексикография, қазақ тілінің лингвистикалық деректер базасын әзірлеу және т.б.

КСРО-дағы квантитативтік-статистикалық лингвистиканың негізін қалаушылардың бірі Қалдыбай Бектаевтың шәкірті болған Асқар Құдайбергенұлы XX ғасырдың 60-жылдары ғылыми шығармашылығын жиілік сөздіктер құрастыру мен соның базасында қазақ мәтіндері құрылымын статистикалық зерттеумен бастады. Қоғамдық ғылымдарға есептеуіш мәшинелер ене бастаған жылдары мәтіндермен жұмыс істеу мен сол сияқты нәтижелерді математикалық талдау үшін арнайы бағдарламалық пакеттер мен қосымшалар жоқ болатын. Жас ғалымның үлес қосуымен М.Әуезовтің «Абай жолы» романы цифрланып, оның материалдары ғалым әзірлеген алгоритмі мен бағдарламасының негізінде Минск-22 ЭЕМ-де зерттелді. Зерттеу нәтижелері жас ғалымның ғылыми мақалаларында жарияланып, кандидаттық диссертациясының негізін қалады.

Асқар Құдайбергенұлының қазақ мәтіндері бойынша зерттеулері нәтижесінде лингвистикалық зерттеулердегі норманы таңдау (серия көлемі (K), серия саны (n), және таң-

дау көлемі (N)) зерттеліп отырған лингвистикалық бірліктің үлестірілім сипатына едәуір әсер ететіні туралы болжам дәйектелді.

Осы зерттеулерде ғалым нормалаудың кіші мәнінде лингвистикалық бірліктердің үлестірілім түрі зерттеліп отырған мәтіннің функционалдық стиліне де байланысты болатыны анықталды.

Бұл зерттеулер А.Қ.Жұбанов пен Қ.Б.Бектаевтың жиілік сөздіктер жасау мен түркі тілдері мәтіндерін статистика-квантитативті талдаулар жүргізуде бірдей ұстанымға сүйенуге мүмкіндік берді (нормалау $K=200$, $n=1000$, $N=200000$).

Ақпараттық технологияның бастапқы кезеңінде КСРО-да операциялық жүйелерде тек ағылшын және орыс тілі ғана болды, тіпті компьютер деген сөздің өзі қолданыста жоқ болатын. Оның орнына ЭЕМ аббревиатурасы деп қолданылды. Кеңес үкіметінің көптеген тілдері үшін лингвистикалық зерттеулерде цифрлық техниканы қолдану ерекше міндет болатын. А.Қ.Жұбановтың түркі тілдерінің жазба мәтіндерін ЭЕМ-де кодтау мәселесіне арналған тұңғыш еңбегі және соның негізінде құрастырылған жиілік, кері-жиілік, әліпбилі-жиілік сөздіктер, әріптер тіркесімінің, сөзтіркесімнің жиілік тізімі алгоритмі түркологиядағы цифрлық техниканы қолдану мүмкіндігін туғызды. Ғалым ұсынған түркі мәтіндерін өңдеу технологиясында цифрлық техниканы қолдану нәтижелері өзбек, қарақалпақ және башқұрт ғалымдары еңбектерінде қолданыс тапты.

Асқар Құдайбергеноұлы көп жылдар бойы түркі республикаларының тілші-әріптестеріне ақпараттық технологияны меңгеру ісі бойынша кеңес беріп, үнемі қолдау көрсетіп отырды. Мұны әсіресе, кирил қаріптері негізінде ұлттық клавиатурада әріптерді орналастыру көптеген түркі тілдерінде қазақ тіліне өте ұқсайтынын да байқауға болады. Қазақ тілі бойынша жарық көрген жиілік сөздіктердің басым бөлігі А.Қ.Жұбановтың қатысуымен құрастырылғанын атап айту қажет. Бұл сөздіктер құнды зерттеу материалы ғана емес, қазақстандық лингвистика ғылымын биікке шығарған сөздіктер болды. Мысалы, Абай поэзиясы тілінің жиілік сөздігі КСРО-дағы Пушкин лирикасы сөздігінен кейінгі екінші сөздік болды. Ал орхон жазуы мен «Кодекс куманикус» тарихи ескерткіші мәтіндері бойынша құрастырылған сөздіктер әлемдік деңгейдегі бірегей сөздіктер қатарынан орын алады. Ғалым бұл ғылыми бағытын қазіргі кезде де жалғастырып отыр, айталық, Асқар Жұбанов қазақ тілінің үлкен жиілік сөздігін дайындауға жетекшілік жасады, бұл сөздік 2016 жылдың аяқ шенінде жарық көрді.

Ғалым шығарған жиілік сөздіктер, мәтінге статистикалық талдау жасау әдістемесі Асқар Құдайбергеноұлының авторлық лексикография мен қазақ жазушыларының жеке стилін зерттеуде қолданыс тапты. Автордың біраз еңбектері қазақ мәтінінің графемдік құрылымын зерттеуге арналды. Ғалым қазақ сөйленісінде үнді және жуысыңқы шұғыл дыбыстардың жоғары жиілікте қолданылатынын анықтады. Осы бағыттағы жұмыстары ғалымның қазақ мәтіндеріне автоматты транскрипция жасауға мүмкіндік берді. А.Қ.Жұбанов жазба мәтіндерді фонетикалық мәтіндерге автоматты түрде көшіру алгоритмін ұсынды.

Қазақ мәтіндерінің сөзтүрлендіруші және форматудырушы категориялары бойынша статистикалық материалдарды жинақтауы ғалымның қазақ сөзформасының автоматты синтездеу мәселелерінің теориялық аспектілерін зерттеуге мүмкіндік берді. Синтез-анализ мәселесі бір-бірімен тығыз байланысты болған-дықтан, оның бірі шешілсе, екіншісін де шешуге болады. Сөзформаны анализ жә-не синтездеу мәселесі тілдің автоматты морфологиялық талдауышын (анализатор) жасауға база болады. 2000 жылдардың басында ғылыми байланыстың кемдігінен осы мәселе-ге түркітанушылардың назарын аударып, бірігіп істеуге кедергі болды. Есім және етістік сөзформаларды жеке-дара синтездеу негізінде осы мәселені шешуде А.Қ.Жұбановтың өте қырағы ғылыми пайымы маңызды

болды деуге болады. Ғалым бұл ғылыми әдісті басқа түркітанушы ғалымдардан бұрын аңғарған еді. Осыдан кейін бұл мәселе башқұрт және хакас тілдерінде қолданыс тапты.

А.Қ.Жұбановтың қазақ қолданбалы тіл біліміне үлес қосқан іргелі еңбегі, әрине, ғалымның қазақ мәтінін көпдеңгейлі модельдеу мәселесіне арналған докторлық диссертациясы еді. Диссертациядағы негізгі тұжырымдар кейін монография ретінде жарық көрді. Мәтінің семантика-синтаксистік мазмұнын формалдау компьютерге түсінікті жасанды тілді қолдануды қажет етеді. Автор бұл ретте өзінің ғылыми жетекшісі құрастырған СЕМСИИТ тілін қолданады. Бұл тіл жеке сөздің, сөйлемнің, абзацтың және тұтас мәтінің мазмұнын сипаттауға мүмкіндік береді. СЕМСИИТ мәтіндік бірліктер арасындағы семантикалық және синтаксистік қатынастарды сипаттай алағын қағидағтар бойынша жұмыс істей алады. Ал бұл мәтінді семантика-синтаксистік формулаға салуға мүмкіндік береді, соның негізінде нақты бір мәтінің жалпылама семантика-синтаксистік формуласын шығаруға болады. Мұндай жұмыс түркітануда алғаш рет зерттелгендігін атап өту қажет. Олар:

- қазақ тіліндегі атауыш сөздердің семантикалық классификациясы ұсынылды;
- қазақ тілінің үш функционалды стиліндегі (көркем, публицистикалық, ғылыми) заттық-логикалық мазмұндары бойынша бастапқы, ортаңғы және соңғы абзацтар типтері анықталды;
- тірек сөздер типтері анықталды;
- мәтіндегі орналасуына байланысты абзацтың функционалды-мағыналық типтері зерттелді;
- лексика-семантикалық формулалар түрінде әртүрлі функционалды стиль мәтіндерінің модельдері құрастырылды.

Бұл алынған нәтижелер болашақта ұлттық интеллектуалдық ақпараттық жүйені жасауда міндетті түрде пайдаланылатын құнды, маңызды материал болып табылады. Атап айтқанда, ақпаратты автоматты жинау мен классификация жасауда, мәтінді автоматты түрде туындауда т.б.

Филологиялық компьютерлік мәліметтер базасын жаңа ғылыми-теориялық жаңалықтар ашудың дереккөзі, сондай-ақ көпшілік пайдаланушылардың автоматтандырылған анықтамалық құралы ретінде маңыздылығын жақсы түсінген Асқар Құдайбергеноұлы 2003 жылы IT технологиясының мамандары мен филологтардың осы мәселеде бірігуі қажет екендігі туралы баспа беттерінде арнайы көтеріп, мақала жариялаған болатын. Осы кездерден бастап А.Қ.Жұбанов өзінің бар күшін, білімін, тәжірибесін қазақ тілінің лингвистикалық деректер базасын жасауға жұмсап келеді.

Лингвистикада лингвистикалық деректер базасы және толық мәтіндер базасы бар. Лингвистикалық базалар әртүрлі лингвистикалық бірліктер туралы ақпараттарды жинақтайды. Мәселен: лексика, фонетика, морфология, айталық, Қазіргі ағылшын тілінің халықаралық компьютерлік архиві (International Computer Archive of Modern English – ICAME), орыс терминологиясының «РОСТЕРМ» деректер базасы және т.б. Ал толық мәтіндер базасын мәтіндер корпусы құрайды. Мысалы, «Ағылшын тілінің Британ ұлттық корпусы», Ресейдегі ең ірі «Орыс тілінің ұлттық корпусы» немесе «XX ғасыр соңындағы орыс газет мәтіндерінің компьютерлік корпусы».

Автордың осы бағыттағы зерттеулері барысында ғалымның «Тіл – қазына» атты лексикографиялық базасының құрылымы жасалды. Бұл лексикографиялық база әртүрлі компьютерлік бағдарламалармен жабдықталған, пайдаланушыға қажетті материалды іздеп табуға ыңғайлы етіп жасалған, жан-жақты ақпараттарды жинақтаған көпқырлы құрылымнан тұрады. Мұндай лексикографиялық деректер базасын жасаудағы негізгі қиындық, ол сөздіктердің кітап нұсқаларының базасынан жасалатындығы болды. Оларды электронды пішінге келтіру және программамен жұмыс істеуге келтіру, яғни ақпараттық

базаға салу көп еңбекті қажет етеді. А.Қ.Жұбанов қазақ тіл біліміндегі барлық сөздіктерді бір жүйеге келтіріп, оларды компьютерлік лексикографиялық базадағы деректер қорына салудың бірегей құрылымын жасады.

Қазіргі кезде «Тіл – қазына» базасында: Тіл білімі институтының лексикографиялық картотекасы, көптомдық сөздік пен қазақ тілінің басқа да сөздіктерінің электронды нұсқалары бар. База институт картотекасы, түсіндірме, синонимдер, орфоэпиялық сөздіктер, екі тілді сөздіктер, терминологиялық сөздіктердегі сөздер мен формаларды қажетке қарай іздеп табатын арнайы программамен жұмыс жасайды. Мысалы, картотекалық базадан пайдаланушы өзіне қажетті сөзді, авторды, жылы бойынша, кітап атауы бойынша немесе абзацтар, сөйлемдер бойынша тауып ала алады. Бұл база Институт қызметкері үшін септік байланыста орнатылған, осы база институттың көптеген лингвистикалық зерттеулері мен лексикографиялық жұмыстары үшін құнды материал болып отыр. Айталық, осы база қазақ тілінің онбестомдық түсіндірме сөздігінің қысқа мерзімде жарық көруіне септігін тигізді.

Асқар Құдайбергенұлы өзінің бар күшін ғылыми зерттеуге арнай отырып, тілді оқыту мәселесінде жаңа инновациялық әдістемелерді жасауға да уақыт тапты. Ғалымның тіл оқытуда жиілігі жоғары K ($K = F \cdot m / N \cdot n$) сөздерді тірек сөз ретінде алуға қатысты компьютерлік бағдарламасы мектептер мен жоғары оқу орындарында кең қолданыс тапты.

Ғалымның көпжылғы зерттеулер нәтижесі қазақ тілінің ұлттық корпусын жасауға да негіз болды. Бұл корпус қазіргі уақытта А.Байтұрсынұлы атындағы Тіл білімі институтында ғалымның жетекшілігімен жасалып жатыр. Ғалымның ұлттық корпус құрылысын жасауға қатысты мәселелері ғылыми еңбектерінде жарық көрді. Ұлттық корпус туралы зерттеулері мен ғалымның қазақ сөзформасын автоматты анализ және синтездеу міндетін шешуге қатысты зерттеулері, «Тіл – қазына» деректер базасы, сондай-ақ ғалымның ғылымға деген жанашырлығы мен ерен еңбегі институт ұжымының қазақ тілінің ұлттық корпусын сәтті жасап шығуына мүмкіндік беретіні сөзсіз.

Асқар Құдайбергенұлы – Қалдыбай Бектаевтың ең талантты және ең мықты шәкірті. Сондықтан да Қазақстанға математикалық лингвистиканың негізін қалаған белгілі ғалым Қ.Бектаев өзі институтта ашқан қолданбалы лингвистика тобының жетекшісі ретінде Асқар Құдайбергенұлына сеніп тапсырды. Өткен ғасырда бұл топ «Автоматтандыру және лексикография» лабораториясы болды. Кейін өз алдына «Қолданбалы лингвистика» бөлімі болып қайта құрылды. Осы жылдары Асқар Құдайбергенұлы Қ.Бектаевтың математикалық лингвистика мектебін ғана дамытып қойған жоқ, ол өзінің компьютерлік лингвистика мектебінің негізін қалады.

А.Жұбанов көптеген кандидаттық, докторлық диссертациялардың жетекшісі, кеңесшісі және ресми оппоненттері болды. Ол қазақ қолданбалы лингвистикасында ғана емес, башқұрт тіл білімінде де осы саланың мамандарын дайындады. Осы мақала авторы 1983-1984 жылдары А.Жұбановтың жетекшілігімен институттың «Автоматтандыру және лексикография» тобында тағылымдамадан өткен болатын. Осыдан кейін Асқар Құдайбергенұлы кандидаттық диссертацияның ғылыми кеңесшісі болды.

Ғалым шәкірттерінің диссертациялары тақырыптары қолданбалы лингвистиканың әртүрлі бағытын қамтиды. Бұл ғалымның математика, информатика және лингвистика салалары бойынша кең ауқымды, терең білімді ғалым екенінің дәлелі. Бүгінде ғалымның шәкірттері:

1. Сиразитдинов З.А. «Лексика романа «Кан» Д. Юлтыя» (квантитативно-стилистический анализ). Алма-Ата, 1989;
2. Төкебаева Ж.Ә. Абай шығармалары тілінің статоллингвистикалық аспектілері. Алматы, 2001;

3. Қалыбеков Б.Е. 50-жылдардағы бастауыш сынып оқулықтарының лексикалық жүйесі мен морфологиялық құрылымының статистикасы. Алматы, 2003;

4. Абиқеева Б.К. Қазіргі қазақ тіліндегі жай сөйлемдердің құрылымдық-статистикалық модельдері (қолданбалы лингвистика аспектісінде). Алматы, 2009;

5. Нұрлыбаев Н.М. Балаларға арналған мәтіндердегі жай сөйлемдер құрылымының лингвостатистикалық сипаты. –Алматы, 2010 және т.б.

Ғалым шәкірттері қазіргі кезде ғылыми жетекші ізімен қолданбалы лингвистика саласын дамытуға үлестерін қосып жүр.

Ғалым, әсіресе, ЖОО да сабақ беріп жүргенде көп шәкірт тәрбиеледі. Өзі қызмет істеген жоғары оқу орындарында қолданбалы лингвистика саласы бойынша дәріс оқыды, атап айтқанда, әл-Фараби атындағы Қазақ ұлттық университетінде, Абай атындағы Мемлекеттік ұлттық университетте студенттер, магистранттар, PhD докторанттарға арнайы курстар және практикалық сабақтар өткізді.

Асқар Кұдайбергенұлы қолданбалы лингвистика мамандарын дайындауда оларға білім беріп қана қойған жоқ, қазақ тіліне, ғылымға деген құрмет пен сүйіспеншілікті ұялатты. Туған тіліне деген сезімі мен жанашырлығы оған әкесі – қазақ лингвистикасының бірінші профессоры Кұдайберген Қуанұлынан қанмен тараған керемет қасиеті болса керек.

ӘДЕБИЕТТЕР ТІЗІМІ:

[1] Джубанов А.Х., Зубов А.В. Автоматизация некоторых лингвистических процессов // Вестник АН КазССР. Исследования молодых ученых. – Алма-Ата, 1968. №9. –С. 31-36.

[2] Джубанов А.Х., Бектаев К.Б. Вероятностно-статистическое моделирование тюркских текстов//Статистика казахского текста. –Алма-Ата: Наука, 1973. –С.299-328.

[3] Джубанов А.Х. Вероятностные законы распределения классов слов казахского текста//Труды VII Всесоюзной школы-семинара «Автоматическое распознавание слуховых образов». –Алма-Ата, 18-23 сентября, 1972 г. –Алма-Ата: Наука, 1973. –С.139-145.

[4] Лукьяненко К.Ф. Использование схем Пуассона и Гаусса в исследовании распределения лингвистических единиц текста //Вопросы лингвостатистики и автоматизации лингвистических работ. Труды ЦНИИПИ, –Вып. 3. – М., 1969. –С. 5-14.

[5] Джубанов А.Х., Зубов А.В. Автоматизация некоторых лингвистических процессов // Вестник АН КазССР, №9. 1968. – С.31-36.

[6] Джубанов А.Х. Бектаев К.Б., Зубов А.В. Автоматическое построение час- тотных словарей (прямого и обратного) //Вестник АН КазССР, №3. –1970. – С.48-53.

[7] Жубанов А.К. Основные принципы формализации содержания казахского текста. Автореф. дисс. док. филол. наук. –Алматы, 2002. – 32 с.

[8] Сиразитдинов З.А. Моделирование грамматики башкирского языка. Словоизменительная система. –Уфа: Гилем, 2006. – 160 с.