Ғылыми мақала
ӘОЖ 81.11, 81`44
https://doi.org/10.31489/2024Ph3/14-23

Мақаланың редакцияға түскен күні: 10.04.2024 ж.
Қабылданған күні: 27.05.2024 ж.

Н.Н. Аитова[1], Г.Ж. Байшукурова[2], А.Б. Иргебаева[3]

[1]Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан;
[2,3]Абай атындағы Қазақ ұлттық педагогикалық университеті, Алматы, Қазақстан
(e-mail: [1]nurlykhan.an@gmail.com; [2]baishukurova@mail.ru; [3]kaldanov70@mail.ru)

А. Байтұрсынұлы шығармаларын корпустық зерттеу («Қырық мысал», «Маса» жинақтары негізінде)

Мақалада Ахмет Байтұрсынұлының «Маса» және «Қырық мысал» жинақтарындағы мәтіндеріне корпустық әдіспен квалитативті талдау жасалған. Корпустық әдіспен мәтіндерді зерттеу жалпы әлем тіл білімінде жиі қолданылғанмен, қазақ тіл білімінде дәстүрге айнала қойған жоқ. Жалпы мәтінді корпустық талдау тіл мен әдебиетті зерттеудің әдіснамалық маңызды құралы саналады. Корпустық талдау сөзформалар мен сөзқолданыс жиілігін сандық сипаттауға, мәтіннің лексикалық, құрылымдық ерекшелігін айқындауға мүмкіндік береді. Зерттеу жұмысының мақсаты — А. Байтұрсынұлының мәтіндеріне корпустық талдау жүргізу арқылы автордың сөзқолданыс, сөзформа жиілігін анықтау, қазіргі корпус базаларындағы мәтіндер жиілігімен салыстыру, сол бойынша автор қолданысы мен қазіргі қазақ тілінің қолданыс динамикасы жөнінде болжам жасау. Зерттеудің ғылыми, практикалық маңыздылығы да осыдан туындайды. Мәтіндердің белгілі бір тарихи кезеңдер мен қазіргі жай-күйіндегі сөздердің жиілігін автоматты сараптау біріншіден, белгілі кезеңдердегі сөзқолданыс жиілігінің сандық ақпаратын білуге мүмкіндік береді. Екіншіден, тілдік қолданыстардың белгілі бір уақыт аралығындағы теңгерімділігін, қоғамның түрлі саласында өмір сүру қабілетін, т.с.с. сапалық, сандық өзгерістерді анықтауға, тәпсірлеуге және болжауға септігі тиеді. Зерттеу барысында мәтінді корпустық зерттеу, сөзформалардың жиілігін талдау әдістері қолданылды. Арнайы бағдарламалық инструменттер көмегімен автордың шығармалары өңделіп, леммаларды сөзтізбелеу жүзеге асырылды. Зерттеуде талдаулар негізінде қол жеткізілген нәтижелерді А. Байтұрсынұлы шығармаларының жиілік сөздігін жасауда қолдану қарастырылған. Алынған нәтижелердің шығармалардың стилдік қырын танытуда, тақырыптық ерекшеліктерін танымдық, т.б. аспектілерде зерттеуде зор көмегі бар және сөздіктер құрастыру тәжірибесінде, оқу процесінде, тілдік, әдеби жобалар жасауда пайдалануға жарақты.

Кілт сөздер: А. Байтұрсынұлы, корпустық зерттеу, жиілік сөздік, сөзформа, сөзқолданыс, ұлттық корпус, мәтін, ішкорпус.

Кіріспе

Қазақ тіл білімінде корпустық зерттеуге кейінгі уақытта ден қойыла бастады. Бұл қазақ тілінің мәтіндік корпустарының түрлі ұйымдар мен жекелеген зерттеушілер тарапынан қолға алынып, зерттеуші-ғалымдар үшін корпустық базаларды зерттеу мақсатына қолдануға тиімді етіп құруымен, программалауды, корпус интерфейстерін жетілдірумен де байланысты болып отыр.

Қазақ тіліндегі мәтіндерді корпустық зерттеу бойынша алынған өнім ретінде «Жалпы білім берудегі қазақ тілінің жиілік сөздігін» атауға болады [1]. Бұл еңбек А. Байтұрсынұлы атындағы Тіл білімі институты (ТБИ) тарапынан 2016 жылы жарық көрді. Сондай-ақ, Ахмет Байтұрсынұлы еңбектерін корпустық зерттеу мәселесі де алғаш Тіл білімі институтында жүзеге асырылып, А. Байтұрсынұлы мәтіндерінің ішкорпусы жасалды [2]. Бұдан кейінгі зерттеу — Абай атындағы ҚазҰПУ базасында гранттық қаржыландыру негізінде 2023 жылы басталған *«А. Байтурсыновтың конкордансы. Қазақша-орысша параллель корпус»* ғылыми жобасы (жоба жетекшісі — Г.Ж. Байшукурова). Зерттеу тақырыбының таңдалуы Ахмет Байтұрсынұлының шығармашылығының мәдени, әдеби ортада құндылығымен, ұлттық мұра ретіндегі маңыздылығымен де байланысты. Зерттеудің өзектілігі А. Байтұрсынұлы шығармашылығына қызығушылықпен ғана байланысты туындап отырған жоқ. Ұлт ұстазы, біртуар ғалым А. Байтұрсынұлының еңбектері қазіргі зерттеу әдістерімен талдануы бұл уақытқа дейін қолға алынбаған. Аталған ғылыми жоба аясында орындалған зерттеу жұмысы бұл олқылықтың орнын толтырып, көптеген мәселелерді шешуге де септігін тигізеді. Қазіргі уақытта жоба бойынша А. Байтұрсынұлының қазақша-орысша параллель корпусының сайты жасалды, екі тілдегі шығармалар корпус талабына сәйкес өңделіп, базаға енгізілді. Сөзтізбесі әліпбилік ретпен түзіліп, мәтіндерді сәйкестендіру (туралау) жұмыстары жасалуда. Сайт

жабық режімде жасақталу үстінде [3]. Қазіргі кезде корпустық зерттеу әдісімен А. Байтұрсынұлының жекелеген шығармалар жинақтары бойынша алғашқы жиілік сөздіктер түзіліп, нәтижесі осы мақалада көрініс тауып отыр.

Зерттеудің негізгі мақсаты — А. Байтұрсынұлының шығармаларына корпустық әдіспен терең талдау жасау, мәтін талдаудың қазіргі әдістері мен инструменттерін қолдана отырып, автордың өзіндік тілдік қолтаңбасын, ерекшелігін қазақ тілінің қазіргі ахуалымен салыстыра отырып айқындау. Бұл жұмыс автордың шығармашылығын зерттеуде заманауи тәсілдерді қолданатын кейінгі ғылыми ізденістерге де жол ашады деп сенеміз.

Зерттеу әдістері мен материалдар

Зерттеу материалы ретінде Ұлт ұстазы, мемлекет және қоғам қайраткері, түрколог-ғалым, ақын, аудармашы Ахмет Байтұрсынұлының 2013 жылы шыққан «Алты томдық шығармалар жинағының» 1-томына енген «Маса» (1911, 1922) және «Қырық мысал» (1909, 1913, 1922) жинақтары [4] алынды. А. Байтұрсынұлының шығармаларын жаңа әдіснамалық құралдармен зерттеуді тек екі жинаққа енген материалдар негізінде берудің негізгі себебі зерттеу жұмысы көлемінің шектеулі болуымен де байланысты. Зерттеушілер тарапынан бұл жұмыстар әрі қарай терең жүргізіліп, шығармалары толық зерттелу үстінде. Автордың зерттеуге алынған материалдарының көлемі талдау негізінде қорытынды жасауға жеткілікті.

Қазіргі уақытта «Til alemi.kz» сайтында электрондық кітапханада орналастырылған А. Байтұрсынұлының «Таңдамалы шығармаларының» қазақ, орыс тіліндегі нұсқалары [5], [6] және басқа да аудармалары негізінде әзірленген А. Байтұрсынұлының қазақ-орыс параллель корпусы базасы бойынша дайындалған әліпбилік жиілік те талдауда ескерілді. Бұл жұмыс зерттеу жобамыздың бір бөлігі болып табылады. Зерттеу барысында біз параллель корпусқа корпустық зерттеудің нақты құралы ретінде жүгінеміз. Сонымен қатар Ш. Шаяхметов атындағы «Тіл–Қазына» ұлттық ғылыми-практикалық орталығы әзірлеген «Қазақ тілінің ұлттық корпусының кіші корпустары» базасы [7], Қазақ тілінің ұлттық корпусындағы А. Байтұрсынұлы мәтіндерінің ішкорпустары [2] базасы да дереккөз ретінде пайдаланылды.

Талдау барысында зерттеудің келесі әдістері мен инструменттер қолданылды: А. Байтұрсыновтың мәтіндерінен тілдік материалды корпустық іріктеу әдісі; статистикалық әдіс, салыстыру, сөзформаларды жиілік талдау әдісі.

Нәтижелер мен талқылау

А. Байтұрсынұлының «Маса» және «Қырық мысал» жинақтарына енген шығармалардың сөздік құрамын корпустық талдау мәтіндердің тарихи мәдени, тілдік мәнмәтінде өзіндік мазмұнға ие бола отырып, бірегейлігін сақтайтынын көрсетті.

Зерттеу кезеңдері мәтіндерді жинақтау, корпус базасына дайындап, өңдеу, корпустық талдау, нәтижелерді салыстыра зерттеу және тәпсірлеу сатыларынан тұрады.

Біздің зерттеу нысанымызға алынған «Қырық мысал» жинағында жалпы 9939 сөзқолданыс, 4182 сөзформа бар. Демек, сөзтұлғаның бірегейлігі жалпы мәтіннің 42,08 %-ын құрайды. Бір ғана рет қолданылатын сөздер саны — 2769 сөз. Ол сөзтұлға санының 66,22 %-ына, ал жалпы мәтіндегі сөзқолданыстың 27,8 %-на тең. Төмендегі кестеге қараңыз (1-2-кесте):

1 - к е с т е

«Маса» жинағындағы сөзқолданыстардың қайталану жиілігінің көрсеткіші

Жалпы саны	Сөз саны/ пайыздық үлесі	Сөз саны/ пайыздық үлесі	Сөз саны/ пайыздық үлесі	Сөз саны/ пайыздық үлесі
Қайталану дүркінділігі	*1 рет*	*2 рет*	*3 рет*	*4-тен жоғары/*
8090 сөзқолданыс	38,22 %	13,18 %	7,78 %	40,82 %
4208 сөзтұлға	3092/73,49 %	533/ 12,68 %	210/4,99 %	372/8,84 %

«Маса» жинағында жалпы — 8090 сөзқолданыс, сөзформа саны — 4208 сөз. 1 ғана рет қолданылған сөздер саны — 3092, яғни сөзтұлғаның — 73,46 %-ы.

2 - к е с т е

«Қырық мысал» жинағындағы сөзқолданыстардың қайталану жиілігінің көрсеткіші

Жалпы саны	Сөз саны/ пайыздық үлесі	Сөз саны/ пайыздық үлесі	Сөз саны/ пайыздық үлесі	Сөз саны/ пайыздық үлесі
Қайталану дүркінділігі	*1 рет*	*2 рет*	*3-рет*	*4-тен жоғары/*
9939 сөзқолданыс	27,8 %	12,7 %	7,7 %	51,8 %
4182 сөзтұлға	2769/66,22 %	623/ 14,9 %	258/6,17 %	532/12,72 %

«Қырық мысал» жинағында жалпы — 9939 сөзқолданыс, сөзформа саны — 4182 сөз. 1 ғана рет қолданылған сөздер саны — 2769, барлық сөзтұлғаның — 66,22 %-ына тең.

Кестелерден көріп отырғанымыздай, екі жинақ мәтіндеріндегі сөзтұлға санының арасында аса көп айырмашылық жоқ. «Қырық мысалда» сөзқолданыс саны 1849 сөзге артық болғанмен, «Маса» жинағында сөзформа қолданысы 26-ға артық. Екі жинақта да 1 рет қана қолданыс тапқан сөзтұлға жалпы сөзформалар пайызының орта есеппен 66–73 %-ын құрайды. Бұл — Ахмет Байтұрсынұлының тілді қолдану ерекшелігін танытады.

Енді жиіліктегі бірінші 20 сөздің екі шығармада берілуін қарастырайық (3-кесте).

3 - к е с т е

А. Байтұрсынұлының жиілік сөздігіндегі және басқа дереккөздер бойынша алғашқы 20 сөзформа

	сөзформа	қолданыс жиілігі	сөзформа	қолданыс жиілігі	сөзформа	қолданыс жиілігі	сөзформа	қолданыс жиілігі
№	«Маса» жинағы		«Қырық мысал» жинағы		qazcorpora.kz (23 млн аса сөзқолданыс)		Жалпы жиілікті-әліпбилі сөздік	
1	*деп/ ет*	95	*деп /ет*	123	*мен/ (ес), шл)*	181935	бол/ет	108133
2	*жоқ/ мд*	91	*бір/ (ес), (са)*	119	*деп/ ет*	151235	ол/ ес	62796
3	*бар/ мд*	65	*мен/ (ес), (шл)*	83	*да/ шл*	145206	*мен/ шл*	58406
4	*да/ шл*	59	*да/ шл*	78	*бір/ (ес), (са)*	131437	*де/ ет*	46200
5	*бір/ (ес), (са)*	58	*жоқ/ мд*	65	керек/ мд	127421	ал/ ет	43359
6	*не/ (ес), (шл)*	52	*де/ (ет), (шл)*	63	және/ шл	120780	бер/ ет	40295
7	тұр/ ет	45	*бар/ мд*	62	*бар/ мд*	119345	және/ шл	39765
8	*де/ (ет), (шл)*	37	*бұл/ ес*	57	*деген/ ет*	110429	кел/ ет	34526
9	*мен/ (ес), (шл)*	37	*не/ (ес), (шл)*	45	*бұл/ ес*	109653	*де/ шл*	33943
10	*көп/мд*	35	еді / ет	41	болады/ ет	106573	*бұл/ ес*	32850
11	хан/ зт	35	*ақ/ (сн), (шл)*	37	*де/ (ет), (шл)*	103541	*да/ шл*	32109
12	шал /зт	35	*емес/ ет*	36	ол/ ес	103095	жыл/ зт	30882
13	сөз /зт	31	енді/ ет	36	қазақ/ зт	90837	*бір/ ес*	30288
14	*бұл/ ес*	28	*көп/ мд*	36	үшін/ шл	89612	өз/ ес	29594
15	*деген/ ет*	28	*болса/ ет*	33	*жоқ/ мд*	89005	*тұр/ ет*	28643
16	алып/ ет	25	келді/ ет	33	осы/ ес	87932	*сөз/ зт*	23107
17	*деді/ ет*	25	*соң/ шл*	33	*емес /ет*	86472	осы/ ес	25995
18	*емес /ет*	25	*деді/ ет*	32	сол/ ес	77952	адам/ зт	25614
19	*ма /шл*	24	*ма/ шл*	32	ал/ (ет), (шл)	75545	қазақ/ зт	25069
20	жұрт/ зт	22	сол/ ес	32	*не/ (ес), (шл)*	73622	айт/ ет	23421

«Маса» және «Қырық мысал» мәтіндеріндегі жоғары жиілікте қолданылатын сөзформалардың алғашқы 20 қатардағы өзара сәйкестігі — 65 %, 13 сөзформа *(сәйкестіктер курсивпен көрсетілген).* Ең жоғары жиіліктегі *«деп»* сөзі екі мәтінде де бірінші қатарда екенін көреміз. «Масада» бұл сөзформа 95 рет, «Қырық мысалда» 123 рет қолданылған. Сонымен қатар екі жинақта да бірінші рет-

те ***деп***, төртінші ретте ***да*** шылауы орналасқан (кестеде қою курсивпен белгіленді). Сөз таптарына қатысы бойынша «Маса» мәтіндерінде 3 зат есім, 8 етістік, оның ішінде «де» сөйлеу етістігі 4 түрлі тұлғада келеді (*де, деп, деді, деген*), 3 модаль сөз, 4 шылау, 1 сан есім, 3 есімдік қолданылса, «Қырық мысалда» 7 етістік, оның ішінде «де» сөйлеу етістігі 3 түрлі тұлғада келеді (*де, деп, деді)* және *еді* көмекші етістігі 2 түрлі тұлғада *(еді, емес*), 3 модаль сөз, 7 шылау, 1 сан есім, 4 есімдік, 1 сын есім кездеседі.

Ал енді бұл шығармалар мәтіндерінің танымалдылығы жоғары сөзформалар жиілігін 23 миллионнан аса сөзқолданысы, 1014102 сөзформасы бар «Қазіргі қазақ тілінің кіші корпустары» [7] жиілігімен және 36265 сөз тізбекті «Жалпы жиілікті-әліпбилі сөздік» [1;735] жиілігімен салыстырайық: А. Байтұрсынұлы мәтіндеріндегі өте жоғары жиіліктегі 20 сөзформа «Қазақ тілінің ұлттық корпусының кішікорпустары» жиілігіндегі бірінші қатарлардағы 20 сөзқолданыспен 55 % сәйкестікте, ал «Жалпы жиілікті-әліпбилі сөздікпен» 35 % сәйкестік байқалады. Айта кету керек, «Жалпы жиілікті-әліпбилі сөздік» түбір негізді сөздермен шектелетіндіктен де сәйкестік салыстырмалы төмен деуге болады.

Бұл жерден шығарылатын қорытынды А. Байтұрсынұлының мәтіндерінің жиілігі қазіргі қолданыста бар сөйлеу тілі жиілігіне мейлінше теңгерімді.

Мақала көлемін ескере отырып біз алғашқы 20 қатардағы сөздерді салыстырумен шектелдік. Бұл қатардағы сөзформалардың кейбірі (*мен, бір, не*) омонимдік қатар түзетіндіктен, олар айырым белгімен (/) бөліп көрсетілді. Зерттеу барысында омоним сөздер мәнмәтінде қолдану мәніне қарай ажыратылды.

«Масада» 37 рет қолданылған «мен» сөзі 5 ретінде шылау, 32 ретінде жіктеу есімдігі, ал «Қырық мысалда 83 рет қолданылып, 47 рет шылау, 36 рет жіктеу есімдігі ретінде кезігеді. Жалпы есеппен қазақ тілінің ұлттық корпустары жиілігінде де бұл сөздің абсолютті жиілігі алғашқы 10 қатарда кезігеді.

Төмендегі кестелерде корпустық талдаудан ішінара мысалдар берілді. «Мен» омонимінің шылау ретінде қолданылуы курсивтеліп көрсетілді (4-кесте).

4 - к е с т е

А. Байтұрсынұлы мәтіндеріндегі«мен» сөзінің тіркесімде жұмсалу мысалынан

«Маса» жинағынан

Өздерің біліп тұрсаңдар,	мен	несін айтайын» —
деді, Мініп ап,	мен	бастайын сені, —
айтып, жөнді сілтеп,	мен	отырсам, Көзіңнің
таразыға тартылмаған? Дегендер «	мен	жақсымын» толып
Не пайда, өнерің	*мен*	біліміннен, Тиісті
көппен көрген жалғыз	мен	бе? — Деп
Еңбекке егіз, тіл	*мен*	жақ, Ерінбесең,
емес, Орынсыз күйзелейін	мен	несіне?! ТІЛЕК
жақ, Өзіңе аян:	мен	нақақ. Аққа
ем, Құдай-ау,	мен	қазаққа? Мұбтала

«Қырық мысал» жинағынан

жұртты алдайды. ТҮЛКІ	*мен*	ҚАРАШЕКПЕН Түлкіге
пен бақ. КІСІ	*мен*	КӨЛЕҢКЕ Айтайын
қайтарып ап. ҚАРАШЕКПЕН	*мен*	ҚОЙ Шақырды
жоқ па? ҚАЙЫРШЫ	*мен*	ҚЫДЫР Қайыршы
келер сырттан. ӨЗЕН	*мен*	ҚАРАСУ Қарасу
ақырында. МАЛШЫ	*мен*	МАСА Бір
текке қарап жатпадым	мен.	Толтырып ой-
қарап жатып таппадым	мен.	Бойыммен осы
неміз бар?» — деп,	Мен-	дағы ойлаушы
Бәленің түрін көрген	мен -	Сары маса,
қоңырайма, құрбыларым! Бабы жоқ жұмыстағы	мен	бір арық, Күй қайда

Қазіргі уақытта мәтінді корпустық талдаудың дамуымен компьютерлік стилистика бағыты да жетіле түскенін білеміз. Мысалы, Масей Эдер, т.б. ғалымдар мәтін жазу стилін жоғары деңгейде талдауға арналған R стилометрия (есептеу стилистикасы) технологиясын қолданып жазу стилін сандық тұрғыдан зерттейтін еңбектерімен белгілі [8]. Зерттеу нәтижелері мәнмәтіндегі әлеуетті, лингвистикалық сараптамаға қатысты мәтіндерді, тарихи зерттеулерді, т.б. жүзеге асыруда аса қолайлы боп отыр. Соңғы кездері қазақ тілі материалдарын корпустық зерттеушілер де заманауи лингвистикалық технологиялардың қарқынды дамуының түрлі құрылымды тілдердің тілдік деректерін жинау мен талдауды жеңілдететінін, ал корпустық әдістің тілдің статистикалық мәліметтерін алудың ыңғайлы құралы екендігін мойындайды [9; 81].

Бұл бағдарламалық технологиялар автордың жазу стилін танытуда аса пайдалы. Бұның цифрлық гуманитарлық ізденістерде мәтін өңдеуге бағытталған зерттеу жұмыстарын оңтайландыруға септігі тиеді. Осы зерттеу жұмысында да біз корпустық талдау, басқа да арнайы мәтін өңдеу

бағдарламаларына сүйеніп, А. Байтұрсынұлының мәтіндеріне түрлі сүзгіде өңдеу жасадық. Нәтижелері алдағы зерттеу жарияланымдарында беріліп отырады.

Авторлық қолданыстағы жиіліктерді айқындауда Ципф заңын қолдану дәстүрі бар. Ғалымдар кілт сөздерді анықтаудың семантикалық, аргументтік талдауларына қоса, Ципф заңын да бір әдіс ретінде дәйектейді [10; 80].

Ципф заңын мәтіндерді корпустық талдауда қолдану негізгі кілт сөздерді, тақырыптық аяны анықтауға, автордың лексикалық қорының әртүрлілігі мен байлығын бағалауға, сондай-ақ, мәтіннің құрылымы мен ұйымдастырылуы туралы қорытынды жасауға мүмкіндік береді. Ципф заңын авторы американдық ғалым Джордж Ципф 1935 жылы тұжырымдаған, 1949 жылдан бері белсенді танылған бұл заңдылық эмпирикалық сипатқа ие, мәтіндік ақпараттағы сөздердің жиілігіне қарай таралуындағы статистикалық заңдылықты айқындайды. Бұл заң бойынша қандай да бір ақпараттың сандық ұйымдасуында, мәселен тілдегі/мәтіндегі сөздік қордың аз ғана бөлігі жиі, көп бөлігі сирек қолданысқа ие деп тұжырымдалады. Оған қоса әр қатардағы сөздер қолданысының қайталану дүркінділігі белгілі бір тәртіппен құрылады. Нақты айтқанда, екінші рангідегі сөздің белсенділігі бірінші рангіге қарағанда шамамен екі есе, одан кейінгі қатарына қарағанда үш есе төмен жиілікте қолданылады. Ципф заңына сәйкес, мәтіндердегі сөздерді олардың жиілігінің төмендеуіне қарай орналастырғанда, n сөз жиілігі оның сөзтізбедегі реттік санына (рангісіне) кері пропорционал болатыны айтылады[11; 25]. Бұл заң қалалардағы халықтың орналасу жиілігін, адамдардың табысы мөлшерін салыстыруда қолданылып, XX ғасырдың басында бірқатар елдерде шындыққа сәйкестігі дәлелденген. Ципф заңы қарапайым формуламен өрнектеледі:

$$P_n = P_1/n.$$

Мұндағы P_n — n қатардағы (ранг) қала тұрғындарының саны. Мәтіндердегі сөз жиілігіне қатысты Ципф формуласы төмендегідей өрнектеледі:

$$F \times R = C.$$

Мұндағы F — сөздің жиілігі; R — *дәреже* — сөзтізбедегі сөздің реттік нөмірі [11; 25].

Көптеген зерттеуші Ципф заңын мәтінді корпустық талдауда, әсіресе жиілік сөздіктер құрастырғанда пайдаланып, предлогтар, көмекші сөздер (шылау, көмекші етістіктер мен көмекші есімдер) және есімдіктердің жиі, ал керісінше, арнайы лексика, жалқы есімдер, негізгі терминдер және т.б. төмен жиілікте қолданылатынын дәлелдейді.

Біздің зерттеуімізде де осы Ципф заңын қолданудың нәтижесі жоғарыда айтылғандармен негізінен сәйкес түседі. Мысалы, А. Байтұрсынұлы мәтіндерінде шамамен 69 % сөзформа бір-екі реттен қолданылады. Бірінші жүздікте кездесетін сөздердің талдауы төмендегідей:

«Маса» мен «Қырық мысал» мәтіндері сөзтізбесіндегі алғашқы жүз сөздің бір-бірімен сәйкестігі 53 %-ды құраса, корпус базасындағы алғашқы 100 сөзбен салыстырғанда, жиілік сәйкестігі — 48 %. Ал автордың екі жинағына енген мәтіндеріндегі сөзформа жиілігінің баламалығы — 32,13 %. Нақты айтсақ, «Маса» мәтіндерінде 4208 сөзформаның 1348-ы әртүрлі жиілікте «Қырық мысалда» да қолданылған. Осы жерде айта кетсек, қытай қазақтарының оқулықтарына жасалған «Қазақша сөздерді қолдану жиілігінің статистикалық зерттеулері» тақырыбындағы жұмыстарында ғалымдар жоғары жиіліктегі сөздердің барлық оқулықтарда бірдей екендігін, тек жиіліктерінде айырмашылықтар барын дәлелдеп, қазақ тілінің барынша тұрақты жиілікте қолданылатыны туралы тұжырым жасайды [12, 21]. Бұл тұжыммен біз де келісеміз.

Төмендегі кестеден А. Байтұрсынұлы мәтіндеріндегі сөз жиілігінің теңгерімділігін (релеванттылығын) көруге болады (5-кесте).

5 - к е с т е

А. Байтұрсынұлы мәтіндері жиілік сөздігінің теңгерімділік көрсеткіші

Сөзтізбедегі реті	*«Маса» мәтіндерінде*					*«Қырық мысал» мәтіндерінде*				
	Type (сөз)	*Rank (ранг)*	*Freq (жиілігі)*	*Norm Freq (қалыпты жиілік)*	*№*	*Сөзтізбедегі реті*	Type (сөз)	*Rank (ранг)*	*Freq (жиілігі)*	*Norm Freq (қалыпты жиілік)*
1	2	3	4	5	6	7	8	9	10	11
1	*деп*	*1*	95	11742.892	1	1	*деп*	*1*	123	12375.490
2	*жоқ*	*2*	91	11248.455	2	2	*бір*	*2*	119	11973.036
3	*бар*	*3*	65	8034.611	3	3	*мен*	*3*	83	8350.941
4	*да*	*4*	59	7292.954	4	4	*да*	*4*	78	7847.872

5-кестенің жалғасы

1	2	3	4	5	6	7	8	9	10	11
5	*бір*	*5*	58	7169.345	5	5	*жоқ*	*5*	65	6539.893
6	*не*	*6*	52	6427.689	6	6	*де*	*6*	63	6338.666
7	*тұр*	*7*	45	5562.423	7	7	*бар*	*7*	62	6238.052
8	де	8	37	4573.548	8	8	*бұл*	*8*	57	5734.983
10	көп	10	35	4326.329	9	9	*не*	*9*	45	4527.618
13	сөз	13	31	3831.891	10	10	*еді*	*10*	41	4125.163
14	бұл	14	28	3461.063	11	11	*ақ*	*11*	37	3722.709
16	алып	16	25	3090.235	12	*12*	*емес*	*12*	*36*	*3622.095*
19	ма	19	24	2966.625	13	15	болса	15	33	3320.254
20	жұрт	20	22	2719.407	14	18	деді	18	32	3219.640
21	ақ	21	21	2595.797	15	22	айтты	22	28	2817.185
25	соң	22	20	2472.188	16	26	адам	26	27	2716.571
26	болар	26	19	2348.578	17	28	болып	28	26	2615.957
29	аз	29	17	2101.360	18	30	алмай	30	25	2515.344
32	бәрі	32	16	1977.750	19	31	келіп	31	24	2414.730
38	алмай	38	15	1854.141	20	32	екен	32	23	2314.116
47	айтып	47	14	1730.532	21	34	боп	34	22	2213.502
53	адам	53	13	1606.922	22	38	арыстан	38	21	2112.889
59	енді	59	12	1483.313	23	42	еткен	42	20	2012.275
67	ат	67	11	1359.703	24	47	ба	47	19	1911.661
73	ал	73	10	1236.094	25	49	болды	49	18	1811.047
86	Алек	86	9	1112.485	26	55	аз	55	17	1710.434
98	атын	98	8	988.875	27	61	артық	61	16	1609.820
116	айтады	116	7	865.266	28	71	ала	71	15	1509.206
155	адамға	155	6	741.656	29	82	ашып	82	14	1408.592
199	айдап	199	5	618.047	30	95	жатқан	95	13	1307.979
266	адал	266	4	494.438	31	106	біреу	106	12	1207.365
374	адамдар	374	3	370.828	32	118	аман	118	11	1106.751
584	абыройын	584	2	247.219	33	131	адамдар	131	10	1006.137
1117	сап	1117	1	123.609	34	152	ай	152	9	905.524
4208	**өңі**	**1117**	**1**	**123.609**	**35**	171	айтып	171	8	804.910
					36	202	анау	202	7	704.296
Ескерту – Осы кестеде барлық сөзформалардың жиілігі екі мәтін үшін де қысқартылып берілді. Қысқарту ұстанымы: Сөзтізбелік қатарда кездесу жиілігі қайталанған жағдайда бірінші кездескені көрсетілді (мысалы, аз, алды, ау, дүниеде жатыр, көріп сөзформаларының реті (рангісі) — 17, яғни әрқайсысы 17 рет қолданылған, осы жағдайда тек біріншісін көрсеттік: аз. Сондай-ақ, ең соңғы реттегі сөзформа кестенің соңғы жағында берілді					37	239	алыстан	239	6	603.682
					38	302	алатын	302	5	503.069
					39	384	адамның	384	4	402.455
					40	533	адамдарға	533	3	301.841
					41	791	адал	791	2	201.227
					42	1414	ала	1414	1	100.614
					43	**4182**	**өңім**	**1414**	**1**	**100.614**

5-кестеден көретініміз, танымалдығы жоғары бастапқы жиілікте қолданылатын сөзформалар сөзтізбелік қатардың өсу ретімен орналасқан: «Масада» 1–7-қатар, ал «Қырық мысалда» бірінші 1–11-қатар (курсивтелген) бір-бір реттен қолданыс тапқан. Толық жиілік сөзтізбеде одан әрі қарайғы әр қатар екі не одан да көп қайталанып отырғандықтан бірдей реттік нөмірмен беріліп отырады (мысалы, 8-ретте тұрған — *де*, одан кейін келетін сөз — *мен*. Екеуі де сөзтізбеде 8-ранг, жиілігі — 37, т.с.с.). Сөзтізбенің реттік қатарларындағы 1–2 сөздердің жиілік айырымы 4-ке («Масада» 1-сөз 95, 2-сөз 91 рет; «Қырық мысалда» 1-сөз 123, 2-сөз 119 рет) тең болса, 3-сөздердегі айырымы («Масада» 2-сөз 91 рет; 3-сөз 65; «Қырық мысалда» 2-сөз 119, 3–-сөз 83 рет) «Масада» 26-ға; «Қырық мысалда» 36-ға бірден төмен түседі. Ципф заңдылығы бойынша сөз жиілігі мен оның сөзтізбедегі ретінің (рангісінің) кері пропорционалдығы даусыз, ал сөз қатарындағы сөздердің жиілігі арасында 2 еседей айырым болатыны дәлелденбеді, мысалы 1–2 сөздер арасында қатты айырма жоқ, 2–3 сөздер арасындағы жиілік біршама жауықтайды. Талдау нәтижелерінің Ципф заңына тура келіңкіремеуінің өзіндік себептері авторлық қолданыс ерекшелігінен бе, әлде мәтіндердің аса ауқымды болмауынан ба, болмаса, тілдің өзгеше ұйымдасуынан ба деген сауалға нақты дәлел айтуға ертерек. Ол үшін автордың толық мәтіндерінің талдауын ескеру керек.

А. Байтұрсынұлының ғылыми-академиялық жинақтарындағы мәтіндерге жасалған талдауымен салыстырғанда, 29292 сөзформасы бар жиілік тізбеде бірінші тұрған *«бір»* сөзі 1300 рет қолданылып,

55-реттегі 226 рет қолданылған *«сөзді»* сөзформасына дейін 1-55-ретті қатар бойында әр қатар бір реттен жиілік түзеді де, 56-дан бастап қайталанатын қолданыс саны бірнеше реттік болып теңгеріле бастайды (мысалы 56-рангте 222 рет қайталанатын, 66-рангтегі 199 рет қайталанатын қатарлар екеуден, 92-рангтегі 162 рет қайталанатын қатар үшеу т.б.).

«Маса» мәтіндеріндегі сөзформалардың 4-28-ретке дейінгілері 2–8 сөз айырыммен жиілік жасап, 28-ден ең соңғысына дейінгілерінің қайталанымдылығы 1 бірліктен ғана кеми беретінін көреміз (17, 16, 15, 14, ...1). Ал «Қырық мысалда» 4–21-реттегі сөзформаларға дейін 1–7 сөз айырмашылығымен секірмелі төмендеп, 22-сөзформадан бастап ең соңғысына дейінгілері 1 бірліктен ғана кеми береді (28, 27, 26, 25,..., 1). Көріп отырғанымыздай, А. Байтұрсынұлы мәтіндерінің жиілігі бастапқы 1–28-реттегі сөздерге дейін белсенділігі жоғары боп келіп, кейінгілері 28-ден 1 реттік қолданысқа дейін бірден сирей бастайды. Бұдан сөздер жиілігінде ранг түзілуінің жоғарыда аталған заңдылықтарына қайшы келмейтінін байқаймыз. Өте жоғары жиілікте қолданылатын сөздер жалпы мәтіннің шамамен ширек бөлігін ғана құрайтынын талдау нәтижелері көрсетіп тұр.

Біз қарастырып отырған мәтіндердегі жоғары жиіліктегі сөздердің А. Байтұрсынұлының ішкорпусы мәтіндерде кездесуін талдағанда, деп*/ ет — 204;* жоқ*/ мд — 194;* бар*/ мд –215;* да*/ шл — 234;* бір*/ (ес), (са) — 235;* не*/ (ес), (шл)* — 155; тұр/ ет — 200; де*/ (ет), (шл) — 249;* мен*/ (ес), (шл) — 182;* көп*/мд — 171;* хан/ зт — 14; шал /зт — 57; сөз /зт — 124; *бұл/ ес — 191;* деген*/ ет — 161;* алып*/ ет — 136;* деді*/ ет — 39;* емес */ет — 138;* ма */шл* — 224; жұрт/ зт — 125 құжат көрсетеді [2]. Алайда бұл сөзформалар нақты ізделген формасында ғана емес, басқа да тұлғалануы, сөз ішінде осы бөлшектердің кездесуіне қарай барлық сөздерді қамти көрінетіндіктен жоғарыдағы кестелерде салыстыруға алынбады.

А. Байтұрсынұлының қазақша-орысша параллель корпусына енгізілген «Таңдамалы шығармалары» мәтінінде 12282 сөзформа, 52496 сөзқолданыс бар. Жоғары жиілікте қолданылған бірінші 20 сөзтұлға төмендегідей: *сөз/(зт) — 490, деген/(ет) — 427; да/(шл) — 405; болады/ет — 355; деп/(ет) — 332; де/(ет) — 318; мен/ (ес), (шл) — 314; не/ (ес), (шл) — 312; бұл/(ес) — 286; бір/(ес), (са) — 277; болып/(ет) — 261; болса/(ет) — 254; керек/(мд) — 243; бар/(мд) — 221; жоқ/(мд) — 202; сөйлем/(зт) — 196; екі/(са) — 191; сөздер/(зт) — 191; тұр/(ет) — 191; үшін/(шл) — 190* рет қолданылған [3]. Бұл таңдамалы шығармалар мәтіні 3 тілге аудару мақсатында қысқартылып ықшамдап құрастырылған. Автордың осы мәтініндегі сөзформалардың сапалық ерекшелігіне назар аударсақ, 3 зат есім *(сөз, сөйлем, сөздер);* 7 етістік *(деген, болады, деп, де, болып, болса, тұр*); 5 шылау *(да, де, мен, не, үшін),* 4 есімдік *(мен, не, бұл, бір*); 3 сан есім *(бір, екі);* 3 модаль сөз *(керек, бар, жоқ)* жиілік түзген. Берілген жиілікте оминимия ажыратылмаған. Десек те, осы көрсеткіш таңдамалы еңбектегі мәтіндердің сұрыпталу деңгейін болжауға мүмкіндік береді. Сонымен бірге бұл академиялық жинақ мәтінінің жиілігі көркем мәтіндеріндегі жиілікпен толық сәйкес келеді.

А. Байтұрсынұлы мәтіндері сөзформа қолданысы тұрғысынан әр алуандығы мәтіндерде қолданылған сөзтұлғалар бірегейлігінен танылады. Оның ішінде танымалдығы жоғары сөздер автордың қолданысында семантикалық өрістің өзекті бөлігін, сирек кезігу жағдайлары шеткері аймағын қамтиды деуге болады.

Бұған қоса, мәтіндердегі сөзформалардың сөз таптарына қарай ең жиі қолданылған құрылымдары анықталды. Автордың талданған мәтіндерде етістікті құрылымдардан жоғары жиілікте қолданатын сөздері: *де (де, деп, деген, деді), еді (емес), бол (бол, болса, болады, боп, болып, болар), кел (келді, келіп), айт (айтты), ал (алып, алмай), қара (қарап), сал (салып), ет (еткен, етіп), тұр (тұрған), қал (қалды), жат (жатыр, жатты), бер (берді, беріп);* модаль сөздерден: *бар, жоқ, аз, көп*; есім тұлғалардан: *бір (бірі), мен (мені, маған), бұл, не, ақ, сөз, адам, түлкі, сен (саған), бәрі, жалғыз, жан, арыстан, кім, ол (оған, онан, оны), қасқыр, құдай, жаман, осы, сол (соған), аю, дүние (дүниеде) артық, жұрт (жұртқа), күн (күні), шақ (шақта), ала, бай, қара, хан, шал, жақын, інім, қайда, алыс, балық;* сондай-ақ, шылаулар мен үстеулерден: *да, де, мен, ба, ма, түгіл, -ақ, не* сөзформалары. Сөйлеу етістіктерінің ішінде *де* етістігі түбір тұлғада, есімше, көсемше, өткен шақ тұлғасында, *бол* көмекші етістігі жіктік жалғауында, шартты рай, көсемше, есімше тұлғаларында жиі кезігеді. Автор қолданысында етістіктерді көсемше, есімше, жедел өткен шақ тұлғаларында қолдану басым. Қалып етістіктерінен *жатыр, тұр* тұлғалары белсенді қолданысқа ие. Есім сөздерді барыс септігінде жұмсаудың жоғары жиілігі байқалды.

Зерттеуге алынған А. Байтұрсынұлының мәтіндеріндегі ең ұзын сөз таңбалық тұрғыдан он алты әріптен тұрады: *білмейтіндеріңді.* Оның өзі бір ғана рет қолданылған, жалпы 9–16 таңбадан құралатын сөздер өте төмен жиіліктегі түрленген сөздер боп келеді. Көп қолданылған жиілігі де жоғары сөздер 4–

8 таңбадан тұрады. Қазіргі қазақ тіліндегі мәтіндермен салыстырғанда, 16–22 таңбадан тұратын сөздер мүлдем кездеспейді. Бұл ретте тағы да Г. Алтынбек, Кс.Л. Вангтың қытай қазақтарының оқулықтарын салыстырған корпустық талдау нәтижелеріне оралсақ, зерттеушілер жиі қолданылатын сөздердің 4–8 символдан тұратынын, ол жалпы мәтіндердің 96,67 %-ын құрайтынын көрсетеді [12, 21]. Нәтижелердің салыстырмасы көрсетіп отырғандай, қазіргі қазақ тілінің сөздік жиілігі мен автордың сөздік жиілігі, танымалдығы жоғары сөздерінің дыбыстық, морфемдік құрамы шамалас түседі.

Қорытынды

Мәтіндерді корпустық әдіспен зерттеу табиғи тілді өңдеудің ең күрделі әрі жаңа аспектісі болып отыр. Мақалада А. Байтұрсынұлының екі жинағына енген мәтіндердің корпустық статистикасы талдауға алынды. Автор қолданысы мен қазіргі қолданыстағы табиғи тілдің бір ғасырдан астам уақыт аралығындағы ресурсында тұтастай алғанда алшақтық жоқ. Бұл қазақ тілінің сөздік қоры мен грамматикалық құрылысының салыстырмалы тұрақтылығын, табиғатының төл ерекшелігін дәлелдейді. Сөз құраушы дыбыс таңбаларының сандық сипатын салыстыра отырып, көп таңбалы сөздердің қолданыс жиілігі төмен болғанын, шұбалаңқылық мүлдем болмағанын анықтадық. Жоғары жиіліктегі сөздер екі мәтінде әртүрлі жиілікпен бір-біріне сәйкес келеді. Осы зерттеу барысында А. Байтұрсынұлының шығармалары корпустық өңдеуден өткізіліп, әр сөз тегтеліп, автордың еңбектерінің қазақша-орысша нұсқалары теңгеріліп (тураланып), параллель корпустың басқы нұсқасы әзірленіп, мәтіндердегі сөздердің әліпбилік жиілік сөздігі жасалды. Корпус сайтын дамыта отырып, қосымша бағдарламалық құралдар көмегімен терең статистикалық талдаулар жүргізілді. Мәтіндердегі сөз жиілігінің Ципф заңына сәйкестігі тексерілді. Ал алдағы уақытта дыбыстық құрам сапасын талдау үшін тұпнұсқаға сай терілген мәтіндерді қарастыру, толық шығармаларының статистикалық дерегін әзірлеу көзделіп отыр.

Бұл зерттеу жұмысы Ғылым және жоғары білім министрлігі тарапынан гранттық қаржыландырылған ЖТН AP19676988 «А. Байтурсыновтың конкордансы. Қазақша-орысша параллель корпус» ғылыми жобасы (2023–2025 ж.) аясында орындалды.

Әдебиеттер тізімі

1 Жалпы білім берудегі қазақ тілінің жиілік сөздігі/ Жалпы ред. басқ.: Е.З. Қажыбек, А.М. Фазылжан. — Алматы: Дәуір, 2016. — 1472 б.

2 А. Байтұрсынұлы мәтіндерінің ішкорпусы. — [Электрондық ресурс]. — Қолжетімділігі: // https://qazcorpus.kz/find-ahmeti/

3 А. Байтұрсынұлының қазақша-орысша параллель корпусы. — [Электрондық ресурс]. — Қолжетімділігі: http://baitursynuly-corp.kz/

4 Байтұрсынұлы А. Алты томдық шығармалар жинағы [Электрондық ресурс] / А. Байтұрсынұлы. — Алматы: «Ел-Шежіре», 2013. — Т. 1. — 384 б. — Қолжетімділігі: **Ошибка! Недопустимый объект гиперссылки.**Ахмет Байтұрсынұлы. Таңдамалы шығармалары [Электрондық ресурс] / Байтұрсынұлы Ахмет. Құраст.: Е. Тілешов, Н. Аитова (жауапты ред.), О. Жұбай, А. Қадырхан. — Астана, 2022. — 240 б. — Қолжетімділігі: https://tilalemi.kz/viewer/viewer.php?file=/books/8178.pdf

5 Ахмет Байтурсынулы. Избранные труды / Ахмет Байтурсынулы; сост.: Е. Тлешов, Н. Аитова (отв. ред.), О. Жубай, А. Кадырхан; пер. с каз. Г.Ж. Байшукуровой, А.Б. Иргибаевой. — Астана, 2022. — 240 с. — [Электронный ресурс]. — Режим доступа: https://tilalemi.kz/viewer/viewer.php? file=/books/8175.pdf.

6 Қазақ тілінің ұлттық корпусының кіші корпустары. — [Электрондық ресурс]. — Қолжетімділігі: https://qazcorpora.kz

7 Eder M. Stylometry with R: A Package for Computational Text Analysis [Electronic resource] / M. Eder, J. Rybicki, M. Kestemont. — Access mode: //https://journal.r-project.org/archive/2016/RJ-2016–007/RJ-2016–007.pdf

8 Бижкенова А.Е. Семантика глаголов движения в русском, казахском и английском вариантах (с опорой на корпусные и словарные данные) / А.Е. Бижкенова, Р. Кенжебекова // Вестн. Караганд. ун-та. Сер. Филология. — 2024. — Т. 29, № 1(113). — С. 80–95. — DOI: https://doi.org/10.31489/2024ph1/80–95

9 Akizhanova D.M. et al. The Zipf's Law and Other Ways of Identifying Culture-Specific Linguistics Units / D.M. Akizhanova // Space and Culture. — India, 2018. — 6:2. — Page 78. //https://doi.org/10.20896/saci.v6i2.363

10 Zipf G.K. Human Behavior and the Principle of Least Effort / G.K. Zipf // Cambridge, MA: Addison-Wesley Press. — 1949. — P. 585.

11 Altenbek G. A corpus-based frequency statistic of Kazakh language [Electronic resource] / G. Altenbek, X.L. Wang // Bulletin of the Karaganda University. Philology Series. — 2017. — No. 2(86). — P. 14-23. — Access mode: http://rep.ksu.kz/handle/data/2181

Н.Н. Аитова, Г.Ж. Байшукурова, А.Б. Иргебаева

Корпусное исследование произведений А. Байтурсынова (на основе сборников «Кырык мысал» и «Маса»

В статье проведен квалитативный анализ текстов корпусным методом в сборниках «Маса» и «Кырык мысал» Ахмета Байтурсынова. Изучение текстов корпусным подходом не применяется широко в казахском языкознании, хоть и часто используется в языкознании многих стран. В целом корпусный анализ текста считается методологически важным инструментом изучения языка и литературы. Корпусный анализ позволяет количественно охарактеризовать словоформы и частоту словоупотребления, а также определить лексическую и структурную специфику текста. Цель исследовательской работы — определение частоты словоупотреблений, словоформ автора путем проведения корпусного анализа текстов А. Байтурсынова, сравнение с частотой текстов на современных корпусных базах, по которым делается прогноз по динамике употребления языка автора и современного казахского языка. Отсюда вытекает научная и практическая значимость данного исследования. Автоматическая экспертиза частоты слов в определенных исторических периодах и современном состоянии текстов позволяет, во-первых, узнать количественную информацию частоты словоупотреблений в определенные периоды. Во-вторых, способствует выявлению, прогнозированию качественных, количественных изменений языкового употребления в течение определенного периода времени, способности к выживанию в различных сферах жизни общества и т.д. В ходе исследования были применены методы корпусного *анализа* текста, анализа частоты словоформ. С помощью специальных программных инструментов были обработаны произведения автора, осуществлено составление реестра лем. В исследовании предусмотрено использование результатов, достигнутых с помощью анализа, при составлении частотного словаря произведений А. Байтурсынова. Вместе с тем полученные результаты оказывают большую помощь в проявлении стилистической грани произведений, изучении тематических особенностей в познавательных и других аспектах. Результаты исследования нашли применение в практике составления словарей, в учебном процессе, в разработке языковых, литературных проектов.

Ключевые слова: А. Байтурсынов, корпусное исследование, частотный словарь, словоформа, национальный корпус, текст, словоупотребление, подкорпус.

N.N. Aitova, G.Zh. Baishukurova, A.B. Irgebayeva

Corpus-based study of the works of A. Baitursynov (based on the collections «Kyryk mysal» (“Forty examples”) and «Masa» (“Mosquito”)

In the article a qualitative analysis of literary texts by the corpus method in the collections «Masa» (“Mosquito”) and «Kyryk mysal» (“Forty examples”) of Akhmet Baitursynov was provided. The study of texts by the corpus approach is not widely used in Kazakh linguistics, although it is often used in the linguistics of many countries. In general, corpus analysis of a literary text is considered a methodologically important tool for studying language and literature. Corpus analysis allows us to quantify word forms and the frequency of word usage, as well as to determine the lexical and structural specifics of the text. The purpose of the research work is to determine the frequency of word usage, word forms of the author by conducting a corpus analysis of the two named collections of works by A. Baitursynov, comparing them with the frequency of texts on modern corpus databases, according to which a forecast was made on the dynamics of the use of the author's language and the modern Kazakh language. This implies the scientific and practical significance of this study. Automatic examination of the frequency of words in certain historical periods and the current state of texts allows, firstly, to find out quantitative information about the frequency of word usage in certain periods. Secondly, it helps to identify and predict qualitative and quantitative changes in language use over a certain period of time, the ability to survive in various spheres of society, etc. In the course of the study, the methods of corpus analysis of the text and analysis of the frequency of word forms were applied. With the help of special software tools, the author's works were processed, lemmatization was carried out, and a dictionary was compiled. In the study the use of the results achieved through analysis in compiling a frequency dictionary of A. Baitursynov's works was provided. At the same time, the results obtained are of great help in the manifestation of the stylistic facet of the works, the study of thematic features in cognitive and other aspects. The results of the research were applied in the practice of compiling dictionaries, in the educational process, in the development of language and literary projects.

Keywords: A. Baitursynov, corpus research, frequency dictionary, word form, national corpus, text, word usage, sub-corpus.

References

1 Kazhybek, E.Z., & Fazylzhan, A.M. (Eds.). (2016). Zhalpy bilim berudegi qazaq tilining zhiilik sozdigi [Frequency dictionary of the Kazakh language in general education]. Almaty: Dauir [in Kazakh].

2 A. Baitursynuly matinderinin ishkorpusy [The corpus of A. Baitursynuly's texts]. Retrieved from: https://qazcorpus.kz/find-ahmeti [in Kazakh].

3 A. Baitursynulynyn qazaqsha-oryssha parallel korpusy [Kazakh-Russian parallel corpus of Baitursynuly]. Retrieved from //http://baitursynuly-corp.kz [in Kazakh].

4 Baitursynuly, A. (2013). Alty tomdyq shygarmalar zhinagy [A collestion of works in six volumes]. Vol. I. Almaty: "El-Shezhire". Retrieved from: //baitursynuly.kz/books/8000.pdf [in Kazakh].

5 Tleshov, E., Aitova, N., Zhubay, O., & Kadyrhan, A. (Comp.). (2022). Akhmet Baitursynuly. Tangdamaly shygarmalary [Selected works]. Astana. Retrieved from: https://tilalemi.kz/viewer/viewer.php? file=/books/8178.pdf [in Kazakh].

6 Tleshov, E., Aitova, N., Zhubay, O., & Kadyrhan, A. (Comp). (2022). A. Baitursynuly. Izbrannye sochineniia [Selected Works]. (G.Zh. Baishukurova, A.B. Irgibayeva, Trans). Astana. Retrieved from: //https://tilalemi.kz/viewer/viewer.php? file=/books/8175.pdf [in Russian].

7 Qazaq tilining ulttyq korpusynyng kishi korpustary [Minor corpora of the national corpus of the Kazakh language]. Retrieved from https://qazcorpora.kz [in Kazakh].

8 Eder, M., Rybicki, J., & Kestemont, M. Stylometry with R: A Package for Computational Text Analysis. Retrieved from //https://journal.r-project.org/archive/2016/RJ-2016–007/RJ-2016–007.pdf

9 Bizhkenova, A.E., & Kenzhebekova, R. (2024). Semantika glagolov dvizhenia v russkom, kazakhskom i angliiskom variantakh (s oporoi na korpusnye i slovarnye dannye) [Semantics of verbs of motion in Russian, Kazakh and English versions (based on corpus and dictionary data)]. *Vestnik Karagandinskogo universiteta. Seriia Filologiia — Bulletin of Karaganda Unicersity. Series "Philology",* 29, 1(113), 80–95. DOI: https://doi.org/10.31489/2024ph1/80–95 [in Russian].

10 Akizhanova, D.M. (2018). The Zipf's Law and Other Ways of Identifying Culture-Specific Linguistics Units. *Space and Culture*, 6:2, Page 78. India. //https://doi.org/10.20896/saci.v6i2.363

11 Zipf, G.K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley Press, 585 p.

12 Altenbek, G., & Wang, X.L. (2017). A corpus-based frequency statistic of Kazakh language. *Bulletin of Karaganda University. Philology series,* 2(86), 14-23. Retrieved from / http://rep.ksu.kz/handle/data/2181

Information about the authors

Aitova, Nurlykhan Nurullaevna — Candidate of philological sciences, associate professor, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan. E-mail: nurlykhan.an@gmail.com

Baishukurova, Gulnur Zholaushybayevna — Candidate of philological sciences, associate professor, Abai Kazakh National Pedagogical University, Almaty, Kazakhstan. E-mail: baishukurova@mail.ru

Irgebayeva, Akerke Bayanovna — Candidate of philological sciences, senior lecturer, Abai Kazakh National Pedagogical University, Almaty, Kazakhstan. E-mail: kaldanov70@mail.ru