

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ БІЛІМ ЖӘНЕ ҒЫЛЫМ МИНИСТРЛІГІ
ҒЫЛЫМ КОМИТЕТІ
АХМЕТ БАЙТҰРСЫНҰЛЫ атындағы ТІЛ БІЛІМІ ИНСТИТУТЫ

ЖАЛПЫ БІЛІМ БЕРУДЕГІ ҚАЗАҚ ТІЛІНІҢ ЖИІЛІК СӨЗДІГІ

АЛМАТЫ
2016

ӘОЖ 811.512.122(038)

КБЖ 81.2-4

Ж 23

*Қазақстан Республикасы Білім және ғылым министрлігінің тапсырмасы бойынша
2016 жылғы «Қазақ тілінің ұлттық корпусын әзірлеу және жасау» зерттеу
жобасы аясында дайындалды*

Жалпы редакциясын басқарғандар:

филология ғылымдарының докторы, профессор **Қажыбек Е.З.**,
филология ғылымдарының кандидаты **Фазылжанова А.М.**

Арнаулы ғылыми редактор:

филология ғылымдарының докторы, профессор **Жұбанов А.Қ.**

Жауапты редактор:

филология ғылымдарының докторы **Жаңабекова А.Ә.**

Жауапты шығарушылар:

филология ғылымдарының кандидаты **Исаева Г.С.**,
филология ғылымдарының кандидаты **Жұмабаева Ж.Т.**,
филология ғылымдарының кандидаты **Сейтбекова А.А.**

Ғылыми кеңесшілер:

филология ғылымдарының докторы, профессор **Уәли Н.**,
филология ғылымдарының докторы, профессор **Сүлейменова Э.Д.**

Пікір жазғандар:

филология ғылымдарының докторы, профессор **Алдаш А.М.**,
филология ғылымдарының докторы, профессор **Ахметжанова З.Қ.**,
филология ғылымдарының докторы, профессор **Манкеева Ж.А.**,
филология ғылымдарының докторы **Сүйеркүл Б.М.**,
филология ғылымдарының кандидаты **Сәрекенова Қ.**

Ж 23 Жалпы білім берудегі қазақ тілінің жиілік сөздігі / Жалпы редак. басқарғандар:
Қажыбек Е.З., Фазылжанова А.М. – Алматы: Дәуір, 2016. – 1472 б.

ISBN 978-601-217-569-1

7 миллионнан астам сөзқолданысы бар мәтіндік (тілдік) база негізінде жасалған «Жалпы білім берудегі қазақ тілінің жиілік сөздігі» – мемлекеттік тілді оқытуда, әсіресе оқулықтарды, оқу құралдары, оқу сөздіктерін, тілдік (лексикалық, грамматикалық) минимумдарды жасауда бірден-бір қажетті құрал. Сөздіктің тілдік базасы оқушының мектепті бітіргенше күнделікті қоғамдық және тұрмыстық өмірінде коммуникацияға түсетін түрлі дискурстағы мәтіндерден құралды. Сөздікте қазақ тілінің бес стилінен (ғылыми, көркем әдебиет, публицистикалық, ресми-іскери, сөйлеу) және әртүрлі стиль мәтіндері қамтылған, бірақ статистикасы жеке шығарылмаған аралас мәтін базасынан алынған 36265 сөзден тұратын сөзтізбенің жиілігі көрсетілді. Әрбір стиль, сөз табы бойынша да сөзтізбе жеке алынып, жиілігі берілді. Тұңғыш рет 1000-нан астам грамматикалық омоним мен омографтардың жиілігі ажыратылды. Сонымен бірге алғаш рет сөзтізбесі 250 000 түрленімге жуық сөзформалардың жиілік сөздігі жасалды. Сонымен, жалпы жиілік, стильдер бойынша үлестірім жиілік, әрбір стильдің жеке жиілігі мен әр сөз табының жеке жиілігі, сол сияқты сөзформалардың жиілігі үш түрлі құрылымда: әліпбилі-жиілік, жиілікті-әліпбилі және керіәліпбилі-жиілік түрінде жасалып, пайдаланушылардың назарына ұсынылып отыр. Бұлардың ішінде негізсөз сөзтізбесіне негізделген «Стильдер бойынша үлестірілім әліпбилі-жиілікті сөздік» пен бір топ сөздің мәтінді қамту пайызы көрсетілген «Жалпы жиілікті-әліпбилі сөздік» кітап түрінде қағаз нұсқада, сөздіктің аталған басқа 59 түрленімі кітапқа қосымша DVD дискіде электронды нұсқада берілді.

Сөздік көпшілік қауымға – білім, оқу-ағарту, оқушы жастарға, ғылым, баспасөз саласы мамандарына, ана тілін терең білуге, қазақ тілін үйренуге талпынған барша оқырманға арналады.

ӘОЖ 811.512.122(038)

КБЖ 81.2-4

© А. Байтұрсынұлы атындағы
Тіл білімі институты, 2016;

ISBN 978-601-217-569-1

СӨЗДІКТІ ҚҰРАСТЫРУШЫЛАР:

филология ғылымдарының докторы **Жаңабекова А.Ә.**,
филология ғылымдарының докторы, доцент **Рысберген Қ.Қ.**;
филология ғылымдарының кандидаты **Аманбаева А.Ж.**,
филология ғылымдарының кандидаты **Әмірбекова А.**,
филология ғылымдарының кандидаты **Әміржанова Н.С.**,
филология ғылымдарының кандидаты **Әшімбаева Н.М.**,
филология ғылымдарының кандидаты **Жонкешов Б.С.**,
филология ғылымдарының кандидаты **Жұмабаева Ж.**,
филология ғылымдарының кандидаты **Исаева Г.С.**,
филология ғылымдарының кандидаты **Қаршығаева А.А.**,
филология ғылымдарының кандидаты **Құлманов С.Қ.**,
филология ғылымдарының кандидаты **Мамырбекова Г.М.**,
филология ғылымдарының кандидаты **Оңғарбаева Н.Б.**,
филология ғылымдарының кандидаты **Өтебаева Э.Ә.**,
филология ғылымдарының кандидаты **Пашан Д.М.**,
филология ғылымдарының кандидаты **Сейтбекова А.А.**,
филология ғылымдарының кандидаты **Солтанбекова А.Б.**,
филология ғылымдарының кандидаты **Фазылжанова А.**
филология ғылымдарының кандидаты **Хабиева А.А.**;
аға ғылыми қызметкер **Тұрғынбаева Ж.Т.**;
ғылыми қызметкер **Бармешова Н.Т.**;
PhD докторант **Карбозова Б.Д.**,
PhD докторант **Шормақова А.Б.**;
филология магистрі **Барменқұлова А.С.**,
филология магистрі **Қожахметова А.Қ.**,
филология магистрі **Құсмолданова М.Қ.**,
филология магистрі **Серікбаева А.Д.**,
филология магистрі **Тлегенова Г.Б.**;
филология магистранты **Қазақбаева Ш.Ш.**,
филология магистранты **Мейірбекова А.Ж.**;
IT-бағдарламашы **Тоқмырзаев Д.О.**

Нұрсұлтан НАЗАРБАЕВ
Қазақстан Республикасының
Президенті – Елбасы

АЛҒЫ СӨЗ

Қазақ тілі – өмірдің сан саласында қолданылуға әлеуеті әбден жететін түркілік түбірін сақтаған сөздік қоры жағынан әлемдегі ең бай, саламатты әрі жаңғырған тілдердің бірі. Ұлтымыздың ұлы қазынасын тіліміз арқылы ұрпақтан ұрпаққа жеткізіп отыру – кез келген қоғам үшін парыз. Бұл – білім беру жүйесінің басты мақсаты.

Тілде сақталған ұшан-теңіз білімдер жүйесіне болашақ қазақстандықтардың барлығы дерлік қол жеткізе алуы үшін алдымен оларды кез келген қоғамдық ортада жаппай мемлекеттік тілде сөйлеуге баулуымыз керек. Ол үшін еліміздегі қазақтілді емес мектеп оқушыларына тілді үйретуді жалпы қолданыстағы белсенді сөздерден бастаған абзал. Мұндай сөздер сөйлеушінің дiттеген ойын жеткізуге, орындатуға, сәтті тілдік қарым-қатынас орнатуға көмегін тигізеді.

Қолдарыңызға ұстап отырған «Жалпы білім берудегі қазақ тілінің жиілік сөздігі» – мемлекеттік тіліміздің оқу-әдістемелік жүйесінің тілтанымдық базасын жаңғыртуға, тілді оқыту үдерісінің тиімділігін арттыруға, оқулық, оқу кешендері мен құралдарының тілдік материалын жүйелеуге, оларды заманауи талапқа сай құрастыруға септігін тигізетін бірегей басылым.

Қазақстан Республикасы Тәуелсіздігінің 25 жылдық мерейтойына сөздікті құрастырушы ғалымдар мен тіл мамандарының жасаған тартуы болып отырған бұл еңбек тілді үйренуші және тілді үйретуші барша қауымның игілігіне жарайды деп білеміз.

КІРІСПЕ

«Жалпы білім берудегі қазақ тілінің жиілік сөздігі» Қазақстан Республикасы Білім және ғылым министрі *Е.К.Сағадиевтің* тапсырмасы бойынша 2016 жылғы «Қазақ тілінің ұлттық корпусын әзірлеу және жасау» зерттеу жобасы аясында дайындалып, баспаға ұсынылды.

«Жалпы білім берудегі қазақ тілінің жиілік сөздігі» (әрі қарай – Сөздік) – осы типтегі сөздік жасаудың озық тәжірибесіне сүйеніп, жалпы саны 1000-ға жуық грамматикалық омонимдер мен кейбір омографтар ажыратылған, сол сияқты мәтіндік базасы мектеп оқушысының күнделікті қоғамдық және тұрмыстық өмірінде коммуникацияға түсуі мүмкін түрлі дискурстағы мәтіндердің 7 миллионнан астам сөзқолданыстан тұратын ауқымды көлемін қамтыған алғашқы тілтанымдық лексикографиялық еңбек. Сөздікте қазақ тілінің бес стилінен (ғылыми, көркем әдебиет, публицистикалық, ресми-іскери, сөйлеу) және әртүрлі стильден теңгеріммен жинақталған ортақ мәтіндер базасынан алынған 36265 бірліктен тұратын негізсөз сөзтізбесінің жиілігі көрсетілді. Олардың әрбір стиль, сөз табы бойынша да жиілігі берілді. Тұңғыш рет 1000-ға жуық грамматикалық омоним мен омографтар ажыратылып, олардың әрқайсысының жиілігі көрсетілді. Сонымен бірге алғаш рет 200 000-нан астам бірлікті қамтыған сөзтізбеден тұратын сөзформалардың жиілік сөздігі жасалды. Сонымен, жалпы жиілік, стильдер бойынша үлестірілім жиілік, әрбір стильдің жеке жиілігі мен әр сөз табының жеке жиілігі, сол сияқты сөзформалардың жиілігі үш түрлі құрылымда: әліпбилі-жиілік, жиілікті-әліпбилі және керіәліпбилі-жиілік түрінде жасалып, жиілік сөздіктің 61 түрленімі нұсқасы ұсынылып отыр.

Бұлардың ішінде «Стильдер бойынша үлестірілім әліпбилі-жиілікті сөздік» пен бір топ сөздің мәтінді қамту пайызы көрсетілген «Жалпы жиілікті-әліпбилі сөздік» кітап түрінде қағаз нұсқада, сөздіктің аталған басқа 61 түрленімі, аса көлемді болғандықтан, кітапқа қосымша DVD дискіде электронды нұсқада берілді.

Сөздік Ахмет Байтұрсынұлы атындағы Тіл білімі институты тілтанушылары мен басқа да мекеме мамандары қатыстырылған ғылыми-шығармашылық ұжымының бірлескен еңбегінің арқасында жарыққа шықты.

Сөздікті жасаудың теориялық негіздері бойынша арнаулы ғылыми редакторлықты атқарған, компьютерлік бағдарлама бойынша жиілікті анықтау жұмыстарын ұйымдастырып, жетекшілік жасаған филология ғылымдарының докторы, профессор *А.Қ.Жұбановтың* еңбегі ерекше болды.

Сөздіктің құрылысы мен құрамы, мәтіндер базасының сапасы мен құрамы, тілтанымдық сараптамасы бойынша ғылыми кеңесшілік жұмыстарын филология ғылымдарының докторы, профессор *Н.Уәли*, филология ғылымдарының докторы, профессор *Э.Д.Сүлейменова* (әл-Фараби атындағы Қазақ ұлттық университеті) жүзеге асырды.

Сөздікті құрастыру барысында туындаған теориялық және практикалық мәселелер бойынша филология ғылымдарының докторы, профессор *Ә.Жүнісбек*, филология ғылымдарының докторы, профессор *Р.Шойбеков*, филология ғылымдарының докторы, профессор *Б.Қ.Моминова* ғылыми кеңес беріп отырды.

Сөздікті тілтанымдық сараптамадан өткізіп, сапасын жақсарту бойынша филология ғылымдарының докторы, профессор *А.М.Алдаш*, филология ғылымдарының докторы, профессор *З.Қ.Ахметжанова* (С.Демирел университеті), филология ғылымдарының докторы, профессор *Ж.А.Манкеева*, филология ғылымдарының докторы *Б.М.Сүйерқұл*, филология ғылымдарының кандидаты *Қ.Сәрекенова* (Л.Гумилев атындағы Еуразия ұлттық университеті) байыпты пікір берді.

Сөздікке белгілі ғалымдар, атап айтқанда, қоғам қайраткері, әдебиеттанушы, филология ғылымдарының докторы, профессор *У.Қалижан*, әлеуметтану ғылымдарының докторы, профессор, ҰҒА корреспондент-мүшесі *З.К.Шаукенова* пікір-ұсыныстарын айтты.

Сөздіктің идеясын ұсыну, құрамы мен құрылысын анықтау, мәтіндік базасын жинақтау, өңдеу, сөздердің жиілігін алу, сөзтізбесін құру, сөздерді леммалаудан бастап, бірнеше мәрте редакциялау, коррекциялау т.б. осы сияқты әзірлеу, құрастыру, баспаға дайындауға қатысты сан түрлі жұмыстарына «Сөздіктің құрамы, құрылысы мен статистикасы» бөлімінің авторы, филология ғылымдарының кандидаты *А.М.Фазылжанованың* жалпы редактор ретіндегі ұйымдастыруымен, филология ғылымдарының докторы *А.Ә.Жаңабекованың* жауапты редакторлығымен А.Байтұрсынұлы атындағы Тіл білімі институты тілтанушы-мамандарының ұжымы тұтастай дерлік жұмылдырылды.

Атап айтқанда, Сөздіктің теңгерімді әрі балансталған мәтіндік базасының жинақталып, компьютерлік әрі тілтанымдық өңделуіне жауапты филология ғылымдарының кандидаты *Г.С.Исаева* сөздікті құрастыру, баспаға жариялау барысында алға қойылған тапсырмалардың орындалуының барлығына ерекше атсалысты.

Мәтіндік базасын жинап, баспадан жарық көргенінше Сөздікті әзірлеп, жариялаудың әрбір кезеңіндегі күрделілігі жағынан сан түрлі жұмыстарға түгелдей белсене қатысқан филология ғылымдарының докторы, доцент *Қ.Қ.Рысбергелінің*, филология ғылымдарының кандидаты *Н.Б.Оңғарбаеваның*, филология ғылымдарының кандидаты *Н.М.Әшімбаева*, филология ғылымдарының кандидаты *А.А.Хабиева*, филология ғылымдарының кандидаты *А.Ж.Аманбаева*, филология ғылымдарының кандидаты *Ж.Т.Жұмбаева*, филология ғылымдарының кандидаты *Г.М.Мамырбекова*, филология ғылымдарының кандидаты *А.А.Сейтбекова*, филология ғылымдарының кандидаты *А.Б.Әмірбекова*, филология ғылымдарының кандидаты *Н.С.Әміржанова*, филология ғылымдарының кандидаты *Э.Ә.Өтебаева*, филология ғылымдарының кандидаты *Д.М.Пашан*; аға ғылыми қызметкер *Ш.М.Жамықбаева*, аға ғылыми қызметкер *Ж.Т.Тұрғынбаева*,

аға ғылыми қызметкер *Н.Т.Бармешова*, PhD докторант *Б.Д.Карбозова*, филология магистрі *А.С.Барменқұлова*, филология магистрі *А.Қ.Қожахметова*, филология магистрі *Г.Б.Тлегенова*, филология магистрі *А.Д.Серікбаева*, филология магистранты *А.Ж.Мейірбекова*, филология магистранты *Ш.Ш.Қазақбаеваның* еңбегін ерекше атап өту керек.

Сол сияқты Сөздікті құрастырудың әртүрлі жұмыстарына Институт мамандарынан филология ғылымдарының кандидаты *С.Қ.Құлманов*, филология ғылымдарының кандидаты *Б.С.Жонкешов*, филология ғылымдарының кандидаты *А.Б.Солтанбекова*, филология ғылымдарының кандидаты *А.А.Қаршығаева*, аға ғылыми қызметкер *Г.Ө.Қалықова*, ғылыми қызметкер *А.А.Хисамтдинова*, ғылыми қызметкер *З.Б.Садырбаева*, кіші ғылыми қызметкер *Г.А.Әбдіхалықова*, кіші ғылыми қызметкер *М.Е.Қарбозова*, PhD докторант *Е.Маралбек*, PhD докторант *А.Б.Шормақова*, филология магистрі *М.Қ.Құсмолданова*, филология магистрі *А.А.Серікбаева*, филология магистрі *Е.Б.Бесіров* белсене қатысты.

Тұтас мәтіннен сөздіктің сөзтізбесін құрау, сөздердің жиілігін алудың компьютерлік бағдарламасын әзірлеу, сөздіктің стильдер, сөз таптары, сол сияқты жалпы және стильдерге үлестірілім түрлері бойынша әліпбилі-жиілік, жиілікті-әліпбилі, керіәліпбилі-жиілік түрленімін түгел жасап шығу ісі IT-маманы, бағдарламашы *Д.О.Тоқмырзаевтың* еңбегімен жүзеге асырылды.

Халықтың сөз қазынасы ұшан-теңіз. Оны жинап, тәртіпке салып, жүйелеп, санын шығару мүлде мінсіз болатындай оңай іс емес. Сондықтан қолдарыңыздағы еңбектің сапасын жақсартатын, кемтігін толтыратын кеңестеріңіз бен сын-ескертпелеріңіз Сөздік үшін аса пайдалы болмақ, бұған сөздікті құрастырушы шығармашылық ұжым Өздеріңізге алдын ала зор алғысын білдіреді.

Ақыл-кеңестеріңіз бен сын-ескертпелеріңізді мына мекенжайға жіберулеріңізге болады: 050010 Алматы қ., Құрманғазы көш., 29 үй, Ахмет Байтұрсынұлы атындағы Тіл білімі институты, tilbilimi@bk.ru

Е.Қажыбек

СӨЗДІКТІҢ ҚҰРАМЫ, ҚҰРЫЛЫСЫ МЕН СТАТИСТИКАСЫ

1. СӨЗДІКТІҢ МАҚСАТЫ

«Жалпы білім берудегі қазақ тілінің жиілік сөздігі» ең алдымен мемлекеттік тілді оқытуда, оқулық, оқу құралдарын, оқу сөздіктерін, тілдік (лексикалық, грамматикалық) минимумдарды ғылыми негізде жасауда аса қажетті осы сипаттағы еңбектердің ішінде омонимдердің ажыратылғандығымен, мәтіндік базасының ауқымдылығымен ерекшеленетін тың тілтанымдық-лексикографиялық еңбек болып табылады. Сонымен бірге сөздік морфологиялық, фонетикалық дублет, жарыспалы сөздердің қай нұсқасын әдеби нормаға алуды анықтауда, кірме сөздер, терминдер, неологизмдердің тілдің жалпыхалықтық лексикадағы орнын анықтауда, басқа да түрлі сөздіктер жасауда, лексикографиялық теориялық мәселелерді шешуде, тілдің лексикалық байлығы мен стильдік ерекшелігін, түрлі стильдердегі сөзқолданыс сипатын анықтауда, типологиялық зерттеулер жүргізуге керек материал жинауда, сөздер мен сөз таптарының статистикалық деректерін алуда бірден-бір қажетті құрал болып табылады.

Сөздіктің негізгі мақсаты – тілдегі белсенді сөздік қор шеғарасын, сол арқылы қазақ тілін, әсіресе екінші тіл ретінде үйренетіндерге арналған тиімді оқулықтар мен оқу құралдарын, оларға база болатын лексикалық-грамматикалық минимум әзірлеуге қажетті лексикалық материалды анықтап беру.

Бұл Сөздікте қазақстандық оқушының бірінші сыныптан бастап мектеп бітіргенінше қарым-қатынасқа түсетін тілдік орта, тілдік жағдаяттарындағы мәтін түрлері мейлінше толық қамтылды. Дегенмен де оны тек мектеп қабырғасында игерілетін білім деңгейін ғана қамтыған мәтіндер базасындағы сөздік деуге келмейді, себебі, тілдік материалының көп бөлігін оқу дискурсындағы мәтіндер алғанымен, қоғамдық өмірдің алуан саласындағы күрделілігі жағынан оңайдан қиынға дейінгі әртүрлі сипат-мазмұндағы типі мен жанры алуан түрлі жалпы көлемі ауқымды мәтіндер қамтылды. Бұл мектеп қабырғасын аттаған баланың оны бітіріп шыққанша зияты мен тілдік қарым-қатынастағы білім-біліктілік-дағдысы қоса жетіліп отыратынын ескерумен, қазақ тілін жетік түрде игерген тілдік тұлғаны қалыптастыруға, оны дамытуға қажетті бола алатын қоғамдық өмірдің сан қырлы саласындағы тілдесім әрекетіне тән мәтіндерді барынша қамту ниетімен байланысты Сөздік базасының ерекшелігі болып табылады.

Жоғарыда айтылғандар Сөздіктің оқу-білім ғана емес, қазіргі қазақ тілінен ақпарат іздеген ғылыми-зерттеулер саласында да құнды құрал бола алатынын көрсетеді. Себебі жиілік сөздіктің базасына алынған мәтіндер қазіргі стиль

қоғамның сан саласында өмір сүріп отырған алуан түрлі тілдесім әрекетіне қатысты болуы арқылы тіліміздің бүгінгі тынысын, айырым ерекшелігін, онда орын алған заманауи трендтер мен тенденциялардан сандық көрсеткіштер арқылы дәйектелген мәлімет бере алады. Тілдегі құбылыс – санадағы өзгерістің бейнесі, санадағы трансформациялар тілде айнаға түскен кескіндей анық бейнеленеді, олай болса, Сөздік қоғамдық санадағы құбылыстарды зерттеушілер үшін де таптырмас құрал бола алады.

2. ЖИЛІК СӨЗДІК НЕ ҮШІН КЕРЕК?

Осы сөздіктің әліпбилі-жиілік түрін ашып, кез келген жерінен сөзтізбедегі сөздерді оқысақ, олардың тілдегі орны, «танымалдығы» әртүрлі екені аңғарылады. Мысалы, *абадан, абажадай... абақ, абақай, ...абат, абаттан* сөздері қатарындағы *абадан* (қасқырдың топтанған әуегінің (үйір)көсемі, күшті де мықты бастаушысы), *абақ* (1. кәсіби. Ұзындығы мың метрдей, көлденеңі бес жүз метрдей болатын, теңізден балық сүзетін жылымның бір түрі. 2. Үлкен өзендер мен көлдерден балық сүзетін аудың түпкі бөлігі, ау түбіндегі дорба, қалта)¹ сөздерінің мағынасын барлығы бірдей біле бермейді. Тіпті қазақ тілінде сөйлеп жүріп, көптеген сөздің мағынасын білмей өту күнделікті өмірімізге еш зиянын тигізбеуі де әбден мүмкін. Алайда өмір бойы қазақша сөйлеп, қазақи ортада өсіп, *мен, бол, сен, және, ақ, жүр, жатыр* сияқты сөздердің мағынасын білмеу мүмкін емес. Демек, сөздердің барлығы бірдей «танымал», әрі олардың тілдік қарым-қатынастағы арқалаған жүгі, атқаратын рөлі тең емес. Бір қызығы, бәріне таныс сөздер кез келген жиілік сөздікте жиілігі жоғары болып келеді. Олай болса, сөздердің ішіндегі белсенділерін қайдан көреміз, олардың жиіліктерінің арақатынасын қалай анықтаймыз деген сұрақтың жауабын статистика, сан арқылы жиілік сөздіктер бермек.

Жиілік сөздік сонда ең алдымен тіл үйрету үшін қажет. Алайда оның тілді ана тілі ретінде үйретудегі пайдасы екінші тіл ретінде үйретуге қарағанда өзгеше. Өзге тілдің иелмені үшін қазақ тілін екінші тіл ретінде оқытуда, сөз жоқ, жиілігі жоғары, яғни белсенді сөздік қордағы сөздерді анықтап, түрін түстеп алудың маңызы зор. Ал қазақ тілін ана тілі ретінде игерген тұлға үшін белсенді сөздерден гөрі, жиілігі аз сөздерді игеру пайдалырақ. Сондықтан бұл мәселеге байыппен қарау керек.

Сонымен қатар жиілік сөздіктің мәліметтері заманауи ақпараттық-технологиялық құралдар арқылы жасалатын табиғи тілді автоматтаушы, мәтінді өңдеуші компьютерлік қондырғы, бағдарлама, қосымшалар, интернет-ресурстар және т.б. әзірлеу үшін аса маңызды болып отыр. Жаһаният заманында тіл жазбаша формада тек қағазда өмір сүрмейді, электрондық нұсқадағы жазбаша мәтіндердің үлесі жылдап емес, күн санап артып келеді. Мәтінді компьютерде теру, өңдеу жұмысын жеңілдетудің түрлі it-бағдарламалары, әсіресе интернеттегі лингвистикалық технологиялар қарқынды дамып отыр. Мәселен, электрондық мәтіннің емлесін автоматты түрде тексеретін редактордың компьютерлік

¹ Қазақ әдеби тілінің сөздігі. 1-т. 30 және 32-66. Алматы: Дәуір, 2011.

бағдарламасын әзірлеу үшін жиілік сөздіктегі ақпарат аса құнды болып табылады. Мысалы, компьютерді пайдаланушы «маға» деген сөзді теріп қойса, мәтін редакторы оның қате екенін сөзформалар тізбесінен анықтап, оған дұрыс балама іздейді, алдымен жиілік сөздік негізінде алынған жиілігі жоғары «маған» есімдігін ұсынады, ал осыған ұқсас «мағаш» сияқты жиілігі аз диалект сөз, қолданылу ықтималдығы төмен болғандықтан, кейінгі кезекте ұсынылады. Компьютерлік емле редакторын сөздердің сөзжасамдық және формажасамдық түрленімін түгендейтін аса ауқымды мәтіндік база негізінде жасалған сөзформалар сөздігінің көмегінен сапалы жасап шығару мүмкін емес.

Жиілік сөздік сол сияқты тілтанымдық ізденістер үшін аса қажет. Бір қызығы, жиілік сөздіктегі математикалық тәсілдермен анықталған сандық мәліметтер тек дәстүрлі-құрылымдық зерттеулер үшін ғана емес, тарихи-салыстырмалы, сонымен бірге антропоцентристік жаңа парадигмадағы когнитивтік модульдегі зерттеулер үшін де таптырмайтын ақпарат көзі болып табылады.

Мәселен, осы жиілік сөздікті жасаушы тілтанушылар алғаш рет ауқымды мәтіндік базаны өңдеудің арқасында тілдегі көптеген жаңа үрдістерді байқау мүмкіндігіне ие болды. Солардың бірі – тілдік нормадағы узусқа, яғни тілдік жүйеде қалыптасқан әдеттегі нұсқаға, логикалық жолға сәйкес келмейтін дәстүрлі жолмен қалыптасқан дағдылы нормаға қатысты мәліметтер. Узустың мысалына *туылдым* (жүйе, логика бойынша) емес, *тудым* (дағды, дәстүр бойынша), *менге* емес, *маған*, т.б. сөздерді келтіруге болады, дәл осы сөздердің, яғни узустық сипаты бар тілдік бірліктердің жиілігі жаппай алғанда жоғары болатыны байқалды. Сонымен қатар осы сияқты тіл теориясы тұрғысынан диахрондық аспектіде тың зерттеулерді қажет ететін тағы бір құбылыс: сөздікте ажыратылған грамматикалық омонимдердің дені бір немесе екі, ары кетсе үш буынды болып келді және олардың орташа жиілігі жаппай жоғары мәндерден көрініп отырды. Демек, мұндай омонимдер тіліміздің белсенді сөздік қорының шегарасына енеді. Мәліметтерді 1-кестеден қарауға болады.

1-кесте.

Омонимдердің әліпбилі-жиілік сөздігінен үзінді

СӨЗ/сөз табы	Жалпы жиілігі	Көнетүркілік түбірі
айт/ет	23421	*ajt-
айт/зт	63	*ajt
айт/од	7	
айтыс/ет	262	*ajtis
айтыс/зт	299	*ajtis
айыр/ет	758	*ajr-
айыр/зт	20	*ajr
айыр/сн	1057	

ак/ет	256	*ak-
ак/зт	57	*āk
ак/сн	1201	*āk
ак/шл	964	
акша/зт	2019	* ākğa (Э.В.Севортян бойынша) ²
акша/сн	202	*āk (түбірiнен)
ал/ет	43359	*al-
ал/сн	106	*āl
ал/шл	8393	
ала/зт	17	* āla
ала/сн	2969	* āla

Бұл мәселені диахрондық аспектідегі ізденістер арқылы зерттеу керектігі омонимдердің көнетүркілік лексикалық қабаттан қалыптасқандығы туралы фактіні дәйектеп тұр. Осы мәселеге назар аударған компаративист ғалымдар аз емес. Қазақ тіл білімінде мұны алғаш теориялық тұрғыда қарастырған ғалым академик Ә.Қайдар³, кейіннен оның шәкірттері болды. Дегенмен жаңа жиілік сөздіктегі тың мәліметтерге сүйене отырып, жүргізілген заманауи зерттеулер тілдегі көнетүркілік түбірлердің құрамындағы бірінші реттік (первичная) созылыңқы дауыстылардың қатысуымен, яғни гетерогенді және этимологиясы бір түбірден тараған, яғни гомогенді омонимдерге қатысты бұл мәселенің әлі де ашылмай жатқан көптеген көлеңкелі тұсына сәуле түсіретініне сенімдіміз.

Сол сияқты лингвокогнитивтік зерттеулер үшін де жиілік сөздік құнды мәліметтер бере алады. Қоғамдық санада идеологиялық, базалық концептілердің қайсысы өзектіленіп отырғанын жиілігі жоғары атауыш сөздер аңғартады. Сөздіктен адам санасына, қоғамдық санаға қатысты танымдық, эвристикалық материалды табуға әбден болады. Осы пікірді алғаш айтқан философ ғалымдардың бірі М.Эпштейн жиілік сөздікті әлемнің тілдік бейнесі ретінде зерттеу құралына айналдыруды ұсынады. Автормен келісе отырып, оның «сөздердің жиіліктік реті – бұл ұғымдардың жалпыхалықтық, стихиялық-демократиялық жүйесі»⁴ деген анықтамасының дәлдігін 21 жыл бұрын жарыққа шыққан «М.О.Әуезовтің 20 томдық шығармалар текстерінің сөздіктері» еңбегіндегі сөзтізбе мен осы Сөздіктің сөзтізбесін салыстырғандағы жиілік мәліметтері көрсетіп отырғанын алға тартамыз. Ол үшін екі сөздіктің де реестріндегі көмекшілік мәні бар сөздерді ысырып, жиілігі жағынан алғашқы жүздікте тұрған атауыш сөздерді салыстыра қарауға болады:

² Севортян Э.В. Этимологический словарь тюркских языков. М.: Наука, 1974. 120 с.

³ Кайдар А.Т. Структура односложных корней и основ в казахском языке. Алматы: Арыс, 2005.

⁴ Эпштейн М. Предлог «В» как философема. Частотный словарь и основной вопрос философии // Вопр. философии. – 2003. – № 6. – С. 86-95.

**1995 және 2016 жылғы Жиілік сөздіктердің алғашқы
100-дігіндегі зат есімдер**

Реті	«М.О.Әуезовтің 20 томдық шығармалар текстерінің сөздіктері»	«Жалпы білім берудегі қазақ тілінің жиілік сөздігі»
	СӨЗ	ЖЫЛ
	ЕЛ	СӨЗ
	ЖЕР	АДАМ
	КҮН	ҚАЗАҚ
	ҮЙ	білім
	БАС	ЕЛ
	АДАМ	ЖЕР
	ҚАЗАҚ	ТІЛ
	жол	БАЛА
	БАЛА	КҮН
	ҚОЛ	халық
	КӨЗ	жұмыс
	ат	ӨЛЕҢ
	ӨЛЕҢ	еңбек
	жан	КӨЗ
	ой	әйел
	кісі	кез
	ауыл	ҮЙ
	ақын	жол
	кез	қызмет
	қыз	ҚОЛ
	ЖЫЛ	деңгей
	ТІЛ	БАС
	жігіт	қала
	ІС	өмір
	–	республика
	–	ие
	–	түр
	–	тапсырма
	–	іш
	–	мемлекет
	–	мәселе
	–	ІС
	–	отбасы

Егер 1995 жылғы сөздікте жиілігі жоғары алғашқы жүз сөздің ішінде 25, ал 2016 жылғы сөздікте 34 зат есім бар болса, олардың 15 екі сөздікте де қайталанып тұр: *жыл, сөз, адам, қазақ, ел, жер, тіл, бала, күн, өлең, көз, үй, қол, бас, іс*. Мұның сыры неде? Неліктен соңғы, яғни тәуелсіздік алғаннан кейін шыққан сөздікте алғашқы жүздіктің қатарынан *республика, мемлекет* сөздері орын алған? – осы сұрақтардың жауабын қоғам тынысы мен сол қоғамды жасап отырған адамның танымын қарастыратын, лингвистикалық ізденістің антропоэзеттік парадигмасындағы когнитивтік тіл білімі тұрғысынан жүргізген зерттеулер ғана берері сөзсіз.

Олардың бірінде жиілік сөздіктердің мәліметтеріне сүйене отырып, қандай да бір мәтіндер базасындағы жиілігі жоғары лексемаларға талдау жүргізу әдісін доминанттық талдау деп атап, сол арқылы белгілі бір этнос, әлеуметтік топ немесе жеке тұлғаның концептуалды картинасының доминанттарын анықтауға болатыны алға тартылып жүр⁵. Мұндай әдіске әбден сенуге болады деп ойлаймыз, себебі жиілік сөздік сөйлеу практикасын неғұрлым объективті көрсететін сөздіктің бірегейі, өйткені ондағы жоғары жиілікті сөздер сөз ағымындағы көп қолданысты, яғни сол тіл иелмені жасап жатқан белгілі бір қоғамдағы неғұрлым сұранысы жоғары ұғым-түсініктерді көрсетеді. Зерттеуші доминанттық талдау арқылы дүниенің тілдік бейнесі тұрғысынан жиілігі жоғары сөздерді анықтап, одан әрі доминант сөздерден доминант концептілерді табуға жол ашады.

Демек, екі кезеңде: бірі – кеңестік, екіншісі – тәуелсіздік алғаннан кейінгі жылдары жарық көрген жиілік сөздіктердегі *жыл, сөз, адам, қазақ, ел, жер, тіл, бала, күн, өлең, көз, үй, қол, бас, іс* атауыш сөздерінің жоғары жиілікпен сипатталуы олардың қазақ ұлттық концептосферасы доминант (базалық идеологиялық және құндылықтық) концептілерінің тілдік репрезентанттары екенін көрсетсе керек. *Республика, мемлекет* сөздерінің Тәуелсіздік кезеңінде туындаған мәтінде жиілігі жағынан алғашқы жүздіктен көрінуінің себебі де осында жатыр: мұны доминант концептілердің тәуелсіздік алумен байланысты этномәдени тілдік ұжым санасындағы трансформацияның (толығының) көрінісі ретінде қарауға болады. Ұлттық санадағы өзгерістер – тілтанушылардың ғана емес, әлеуметтанушы, саясаттану, философ т.б. қоғамдық-гуманитарлық ғылым салалары мамандарының зерттеу объектісі. Олай болса, жиілік сөздік аталған сала ізденушілері үшін де деректер көзі болып табылады.

Сонымен, қорыта айтқанда, жиілік сөздік оқыту үдерісін ұйымдастырушы құралдар әзірлеу, жаңа ақпараттық-лингвистикалық технологиялар жасау және теориялық тілтанымдық ізденістер жүргізу үшін аса қажет лексикографиялық еңбек болып табылады.

3. ЦИПФ НЕМЕСЕ ПАРЕТТО ЗАҢЫ

Тілдегі сөздердің «танымалдық» деңгейін оның сөздіктегі жиілігі көрсетеді, ал олай болса, сөздердің жиіліктері арасында белгілі бір заңдылық болуы мүмкін бе, мәселен, ең жиі қолданылатын сөз бен одан кейінгі жиілікте тұрған сөз жиіліктерінің сандық көрсеткіштері арасындағы қатынаста заңдылық бар ма, олар

⁵ Хроленко А.Т. Основы современной филологии. М.: Флинта, 2013. 252 с.

геометриялық прогрессиямен өзгере ме, әлде қалай деген сияқты сұрақтарға жауап іздеген ғалымдар аз болған жоқ. Солардың бірі америкалық Джордж Кингсли Ципф ашқан сөзтізбедегі сөздер жиілігінің өзгеру заңдылығын қазіргі ғылымда Ципф заңы деп атайды. XX ғасырдың басында америкалық тілтанушыдан бұрын да осы құбылысты байқаған ғалымдар мен ізденушілер болған, алайда ғылыми қауымдастық арасында айналымға енуіне зор үлес қосқандықтан, ол Д.Ципфтің есімімен аталып кетті. Бүгінгі күннің инфометриясын, яғни ақпараттың сандық, соның ішінде семантикалық сипаттамаларын өлшеудің теориясын⁶ бұл әмбебап заңсыз елестету мүмкін емес, себебі ғалым о баста мұны қала халқының санына қатысты айтты, кейін сөздердің жиілігіне тән екенін көрсетті, сөйтіп, тіпті қоғам мүшелері арасындағы байлықтың шоғырлану мөлшеріне қатысты да осы заңдылықтың орындалатынын анықтады.

Ципф заңы бойынша, тілдегі сөздер қорының шағын ғана бөлігі жиі, ал басым бөлігі өте сирек қолданылады, бұған қоса сөздер белсенділігі бойынша мына тәртіпке бағынады: бірінші разрядтағы сөздер екінші разрядтағы сөздерден екі есе, ал үшінші разрядтағы сөздерден үш есе көп жиілікпен қолданылады⁷. Яғни сөз жиілігі оның разрядына кері пропорционал.

Ұсынылып отырған сөздіктегі сөздердің жиілігінде осы заңның сақталғанын **9-кестеде** берілген мәліметтер дәлелдейді. Кестедегі мәлімет бойынша, 108 мыңнан **астам** рет қолданылған *бол* сөзінің көрсеткіші екінші разрядта тұрған *ол* және *мен* сөздері жиілігінен шамамен екі еседей артық, ал үшінші разрядтағы 16 сөздің орташа жиілігінен үш есе **анық** артық, осы сияқты қалған разрядтағы сөздерге қатысты осы заңдылық жалпы алғанда сақталып отырады.

Бұл құбылысты экономикада италиялық әлеуметтанушы Вильфредо Паретто 1897 жылы ашқандықтан, оны көрсететін заңдылықты Паретто заңы, кейде үлестірімі деп те атайды. Паретто заңын қарапайым түрде «халықтың 10 пайызы барлық игіліктің 90 пайызына ие», «нәтиженің 80 пайызы жұмсалған күштің 20 пайызына тән» деген сияқты бүгінгі бизнес-ортада әйгілі болып кеткен тұжырымдар төңірегінде түсіндіруге болады. Ал ғылыми түрде оның негізі Ципф ашқан заңдылықпен бірдей.

Аталмыш заңдар ұсынылып отырған сөздік сөзтізбесіндегі 7 млн сөзқолданыстық мәтінде 108 113 рет жиілікпен қолданылған сөздің неге біреу-ақ (*бол* етістігі), ал есесіне небары 1 рет қана кездескен сөздердің жалпы саны 8082, бұл 36265 сөзден тұратын бүкіл сөзтізбенің 4/1 бөлігінен де артық деген сұраққа жауап береді.

4. ҚАЗАҚ ТІЛІНІҢ ЖИІЛІК СӨЗДІКТЕРІ ТӘЖІРИБЕСІНЕН

Тілдік қарым-қатынасты жүзеге асыруға жұмсалатын лексикалық бірліктердің қандай да бір уақыт кесіндісіндегі (қазіргі заман, кеңестік кезең, т.б.) және тіл иелмендері арасындағы қатынастың нақты кеңістігіндегі (ғылыми ортадағы, ресми-іскери жағдаяттағы, бейресми т.б. қарым-қатынастағы) ауқымы мен

⁶ *Комлев Н.Г.* Словарь иностранных слов. Более 4500 слов и выражений. М.: Эксмо-Пресс, 2006. (қараңыз: Инфометрия).

⁷ *Шаров С.А.* Частотный словарь, <http://www.artint.ru/projects/frqlist.php>

қолданыс қарқынын жиілік сөздіктер ерекше дәлдікпен сипаттай алады. Осыған байланысты ақын-жазушы, саясаткер, қоғам қайраткерлерінің жекелеген беделді тіл иелмендері шығармаларының жиілік сөздігі, сол сияқты саяси, экономикалық, педагогикалық т.б. түрлі дискурстағы мәтіндерден алынған салалық жиілік сөздіктер болады. Бұлардың алғашқысы бойынша жиілік сөздіктер қазақ лексикографиясында аз емес, ал соңғысына қатысты, яғни арнайы бір дискурс негізінде жасалған жиілік сөздіктің алды ұсынылып отырған осы еңбек болмақ.

Қазақ тілінің жиілік сөздіктерінің тілтанымдық теориялық негізі лингвостатистика саласының қалыптасып, дамуы арқылы жасалды. Қазақ тіл білімінде ХХ ғасырдың 70-жылдары осы саланың дамуына ерекше еңбек сіңірген ғалым Қалдыбай Бектаев болды. Белгілі математик Қ.Бектаев ЭЕМ (электресеңтеуіш мәшинелердің) дамуы кезеңінде математикалық білімдер арқылы лингвистикалық құбылыстарды модельдеу, олардың статистикасын алу ісінің ғылым-білім, тіл үйрету, тілтанымдық ізденістерді дамыту ісіне де аса қажет екенін түсінді. Сөйтіп, Ахмет Байтұрсынұлы атындағы Тіл білімі институтында Қолданбалы лингвистика бөлімінің ашылуына мұрындық болып, қазірде осы бөлімді талай жылдан бері басқарып келген, филология ғылымдарының докторы, профессор А.Қ.Жұбановпен және басқа да осы саланың мамандарымен бірлесіп, тұңғыш қазақ тілінің жиілік сөздіктерін жарыққа шығуына жетекшілік жасады.

Бұл ретте профессор Асқар Жұбановтың: «Түркі тілдері бойынша статистикалық зерттеулер жүргізіліп, жиілік сөздіктердің ХХ ғ. 70-жылдары алғаш қазақ тілінде жарық көруі – сол кезеңдегі үлкен жаңалық еді» – деген сөзін мақтанышпен келтіруге болады⁸. Шынымен де, түркі тіл білімінде статоллингвистикалық зерттеулер нәтижесінде алғашқы жиілік сөздіктерді қазақ тілтанушылары құрастырған болатын. Олардың ішінде ерекше атап өтуге тұратын еңбектер қатарында: Қ.Б. Бектаев, А.Қ. Жұбанов, С.Мырзабеков, А.Б. Белботаев құрастырған «М.Әуезовтің «Абай жолы» романының жиілік сөздігі» (1979) мен А.Қ. Құрышжанов, А.Қ. Жұбанов, А.Б. Белботаев құрастырған «Құманша-қазақша жиілік сөздік» (1978). Сол сияқты Қ.Б.Бектаев, А.Қ.Жұбанов, С.Мырзабеков, А.Б.Белботаев құрастырған «М.О.Әуезовтің 20 томдық шығармалар текстерінің жиілік сөздігі» (1995), Ахмет Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімі профессор А.Жұбановтың жетекшілігімен және А.Жаңабекова, Б.Карбозова, А.Қожахметовалардың құрастыруымен 2012-2014 жж. гранттық зерттеу бойынша поэзия, көркем проза, публицистикалық стиль, ғылыми-техникалық стиль мәтіндері негізінде жарық көрген жекелеген сөздіктер материалдары базасында 2,4 миллион сөзқолданыстан тұратын мәтінді қамтыған (оның 2,1 млн-ы М.Әуезовтің 20 томдық шығармалар жинағынан) сөзтізбесі 30515 сөзден тұратын жиілік сөздік (2016) те осы қатардағы аса қажетті еңбек.

Аталған сөздіктер ұсынылып отырған жаңа еңбектің әзірленуіне теориялық әрі практикалық бастама болды. Олардың бүгінгі күнге дейін құндылығы жоғары. Мәселен, «М.Әуезовтің «Абай жолы» романының жиілік сөздігі» ұлы жазушы шығармасының «сөз шұрайы, тіл кестесі келісті, өрнекті сөз орамдарына

⁸ Жұбанов А., Жаңабекова А., Карбозова Б., Қожахметова А. Қазақ тілінің жиілік сөздігі. Алматы: «Қазақ тілі баспасы», 2016. 5-б.

мейлінше бай теңдесі жоқ туынды» екенін сандық мәліметтер, цифрлар арқылы нақтылай⁹ түседі. М.Әуезов осы бір ғана шығармасында 16 983 сөз қолданса, 20 томдық жинақта жарық көрген еңбектерінде 29483 сөз қолданғанын жоғарыда көрсеткен жиілік сөздіктер анықтап отыр. Бұл олардың тұтынушыға беретін сан түрлі пайдалы ақпаратының тек бір жағы ғана. Сол кездегі қазақ тіліндегі түрлі құбылыстардан, мысалы, советизмдердің игерілген, игерілмеген формаларынан, тілдің сол уақыттағы белсенді сөзжасаушы жұрнақтарынан т.б. толып жатқан тілтанымдық-статистикалық ақпараттан хабардар ететін осындай еңбектер тілтанушы үшін, тіл үйренуші үшін ешқашан өз құндылығын жоғалтпайды.

5. СӨЗДІКТІҢ ДЕРЕККӨЗІ – МӘТІНДІК БАЗАСЫ (МӘТІН КОРПУСЫ)

Кез келген жиілік сөздік сөзтізбесінің көлемі мен құрылымы, ондағы сөздердің жиілігі мәтін базасына тікелей қатысты болады. Сондықтан сөздіктің барлық параметрлері өз мақсатына жауап бере алатындай болуы үшін оның мәтін базасы сол мақсатқа сәйкес жинақталуы тиіс.

Мәселен, экономика не заң саласындағы мамандарға тіл үйрету үшін қажетті лексикалық қорды анықтау мақсатында жиілік сөздік жасалатын болса, онда сөздер жиілігі экономистер мен заңгерлердің күнделікті ресми және бейресми тілдік қарым-қатынасында қолданылуы ықтимал және мазмұны олардың кәсіби қызығушылығынан қатты ауытқып кетпейтін мәтіндер базасы негізінде анықталуы керек. Бұл – жиілік сөздік мәтіндеріне қойылатын талаптың біріншісі: мәтіннің м а з м ұ н д ы қ жағының жиілік сөздік мақсатына сәйкестігі. Осы талапқа сәйкес жиілік сөздіктің мәтіндік базасына мектеп қабырғасында, мәдени орындарда, тұрмыстық қатынаста, оқушының қарым-қатынасқа түсуі әбден ықтимал ғылыми, ғылыми-көпшілік ортадағы дискурстардан мәтіндер жинақталды. Сол сияқты көркем және публицистикалық мәтіндер де алынды. Олардың дені балаларға арналған дереккөздер мен авторлардан жинақталды.

Екінші талап – мәтіндердің т е ң г е р і м д і л і г і м е н ж а н - ж а қ т ы л ы ғы . Корпустық лингвистикада «баланстылық» терминін қазақшалап «теңгерімділік», ал «репрезентативтілікті» – *жан-жақтылық* деп атау жиі қолданылады¹⁰. Бұл жерде теңгерімділігі дегеніміз жиілік сөздік базасына алынған мәтіндердің қазақ тілінің қоғамдық өмір саласындағы көріністерін, нақты айтқанда, жалпы білім берудегі дискурстарды мейлінше тең көлемде толық қамту. Ол үшін қазақ әдеби тілінің 5 функционалды стилі бойынша мейлінше көлемді мәтіндер алынды. Ал жан-жақтылықты қамтамасыз ету үшін құрастырушылар әрбір стильдегі түрлі жанр, тип, тақырыптар бойынша мәтіндерді әрі әртүрлі авторлардан аралас жинақтады. Бір автордан бір тақырыпқа қатысты мәтіндердің басымдығына жол бермеу керектігі де ескерілді. Себебі мәтін бір тақырыпқа арналған болса, онда сол тақырыпқа қатысты тірек сөз міндетті түрде бірнеше рет, тіпті аса үлкен жиілікпен кездесуі әбден мүмкін. Сондықтан мәтіндердің жан-жақтылығын ескеру тілдің объективті сипаты туралы дәйекті ақпарат алу үшін аса маңызды.

⁹ Құрышжанов Ә., Уәли Н. «Абай жолының» жиілік сөздігі // Қазақ әдебиеті, 29 февраль, 1980.

¹⁰ Құрышжанов Ә., Уәли Н. «Абай жолының» жиілік сөздігі // Қазақ әдебиеті, 29 февраль, 1980.

Үшінші талап – мәтін базасының неғұрлым ауқымды болуы. Аталмыш талапқа сай қазіргі кезде жиілік сөздіктерді тілдің ұлттық корпустарына сүйеніп әзірлеу тәжірибесі әбден орнықты. Мәселен, орыс тілінің соңғы ауқымды жиілік сөздігі Орыс тілінің ұлттық корпусының базасында 1950-2007 жж. аралығында әзірленген 92 млн сөзқолданыстық мәтін көлемін қамти отырып жасалды. Мұндай сөздіктегі жиілік мәндері, тілдегі белсенді лексикалық қордың шегарасы ерекше дәлдікпен анықталады, оның үстіне жиілік сөздік құрастырушыларына корпус-тағы метабелгіленімнің арқасында түрлі стиль, жанр, мәтін типі, мәтіннің шығу уақытына, авторына қатысты да дербес жиіліктерді анықтау, сөзтізбе бірліктерін леммалау ісі еш қиынға соқпайды, яғни мұндай ауқымды базадағы жиілік сөздіктің ақпараттылығы жоғары, әрі тұтынушының алуан түрлі сұранысына жауап бере алу қасиеті де күшті болады. Сондықтан қазақ тіл білімінде корпус-лингвистиканы дамытып, тілдің ұлттық корпусын жасауды алға тартып отырмыз. Тілдің корпусы жиілік сөздіктер жасау ісін жеңілдетіп қана қоймай, оның нақтылығы мен дәлдігінің деңгейін көтерері сөзсіз, корпус бұдан басқа да көптеген келелі мәселені шешудің инновациялық-ақпараттық құралы болмақ¹¹. Бүгінгі күні А.Байтұрсынұлы атындағы Тіл білімі институты 7 млн-дық мәтін базасы бойынша жасалған қазақ тілінің ұлттық корпусының бастапқы қанатқақты жобасын интернетке қосып отыр. Бұл шағын корпус-лингвистикалық сапасы мен ауқымы әлі де үлкен толықтыруларды қажет етеді. **Оны «Қазақ тілінің ұлттық корпусы» атты қазақ, орыс, ағылшын – үш тілдегі жеңіл іздеу механизмі арқылы онлайн-режимде тұтынушыға әп-сәтте қазақ мәтіндері бойынша тілтанымдық сипаттағы кез келген ақпаратты дереу әрі толыққанды жеткізе алатын ашық түрде қызмет етіп, үздіксіз толықтырылып, жетілдіріліп отыратын, мемлекеттік тілдің инновациялық-ақпараттық базасы ретіндегі мегажоба деңгейінде әзірлеп, жасап шығу – тілтанушылармен бірге барлық қоғамдық-гуманитарлық сала мамандары үшін алға тартылған заманауи талап.** Сөздік ұлттық корпус бұрын жасалғандықтан, оның мәтіндік базасы 7 млн сөзқолданыспен шектелді және сөзтізбедегі сөздерді леммалау (түбірге келтіру) ісі де автоматты түрде емес, құрастырушылардың қол күшімен жүзеге асырылды. Дегенмен бұл Сөздік өзіне дейінгі жиілік сөздіктермен салыстырғанда, мәтіндік базасы көлемінің ауқымдылығымен ерекшеленеді.

Төртінші талап – мәтіндердің қазіргі тілдік кезеңнен жинақталуы. Бұл тұста «қазіргі тілдік кезең» деген ұғымға мазмұнының ерекшелігі бар екеніне назар аудару керек. «Қазіргі кезеңге» ХХ және ХХІ ғасырдағы мәтіндер алынды. Олар ХІХ және одан кейінгі «тарихи кезеңдегі» мәтіндерге қарама-қарсы қойылып отыр. Себебі бұл кезеңде бүгінгі заманғы тіл иелменіне бейтаныс тілдік бірліктер көп кездесетіні анық. Мұндай қолданыстар ХХ ғасыр мәтнінде де болып жатады, дегенмен ХХ ғасырда өмір сүрген классик жазушы, ақын, ғалым, журналист, мемлекет және қоғам қайраткерлері шығармаларының қазіргі білім беру жүйесі оқыту

¹¹ Сүлейменова Э.Д. Қазақ тілі үшін ұлттық корпус керек пе? // Ғылымдағы ғұмыр: ана тілі және түркі әлемі. Халықаралық ғылыми-теориялық конференция материалдары. Алматы: «Елтаным баспасы», 2014. 393-395-бб.

құралдарында кеңінен қолданылуы осы уақытта туған мәтіндердің сөздік базасына алынуына тағы бір салмақты себеп болды. Әрине, ішінара «тарихи кезеңге» жататын мәтіндер кездесіп жатты, олар Сөздік мәтіндерінің базасына оқулықтар мен балаларға арналған оқу құралдары, хрестоматиялардағы қолданысы арқылы енді. Сөздіктің мәтіндік базасында ХХІ ғасырдың мәтіндері үлкен үлеске ие, тіліміздің барлық стильдері бойынша жинақталған бұл тілдік деректердің ішінде бүгінгі заманғы үрдіске сай БАҚ арқылы жинақталған мәтіндердің үлес салмағы едәуір болды.

Осы талаптарды ескере отырып, Сөздікке оқыту дискурсын көбірек қамтитын әдеби тілдің ғылыми, көркем әдебиет, ресми-іскери, публицистикалық, сөйлеу сияқты бес түрлі стилі бойынша топтастырылған сан түрлі типтер мен жанрдағы әртүрлі авторға тән көлемі 7 млн сөзқолданыстан тұратын ауқымды мәтіндік база жинақталды. Осы айтылғанға байланысты сөздіктің мәтіндік базасы бойынша алынған мәліметтер тек тіл үйрету практикасы мен методикасы үшін ғана емес, жалпы қазіргі тілдің ахуалына қатысты теориялық ізденістер үшін де қызықты болмақ.

Қазақ әдеби тілінің функционалды стильдері және ортақ база мәтіндерінің сипаттамасы

Ғылыми стиль бойынша 1-сыныптан бастап 11-сыныпқа дейінгі қазақ тілі пәнінен мектеп оқулықтары, кейбір жаратылыстану пәндері бойынша бастауыш сынып оқушыларына арналған оқулық, оқу құралдары мәтіндері, Қазақстан тарихы және жалпы тарих пәні оқулықтарының мәтіндері, балаларға арналған ғылыми-көпшілік еңбектер, балалар энциклопедиялары, балаларға арналған БАҚ-та, ғылыми-көпшілік басылымдарда жарық көрген мақалалар, рефераттар, тезистер, жазба пікірлер, пікірсараптар, аңдатпалар, баяндамалар, ғылыми тақырыптарды талқылаған сұхбаттар, дәрістер т.б. алынды. Сондай-ақ құрылтай, конференциялар, симпозиумдарда айтылған қоғамдық-гуманитарлық бағыттағы ғылыми-көпшілік баяндаулардың баспасөздегі мәтіндері пайдаланылды.

Ресми-іскери стиль бойынша ресми қатынаста, әсіресе мектеп өміріндегі іскерлік қарым-қатынаста, құжат жүргізуде, сол сияқты басқару мен құқықтық қарым-қатынастың басқа салаларында, ішінара халықаралық қатынаста қолданылатын заңнамалық құжаттамалар; білім беру, экономика, әскери салаға қатысты құжаттар, жарнама, ресми мекемелердегі іс жүргізу мәтін үлгілері, үкімет қызметі және т.б. іскери салаларда қамтылатын ресми сөйлеудің кең тараған мәтіндері алынды.

Публицистикалық стильдің базасына балаларға арналған радио-теледидар бағдарламаларының сценарийлері мен балалардың интернет ресурстарындағы мәтіндері, мерзімді басылым: балалар газеті, балалар журналдарындағы беделді тіл иелмендері – ақын-жазушылардың, қоғам қайраткерлері мен журналистердің, сондай-ақ балалардың өз шығармашылығының туындылары, әртүрлі қызықты танымдық мақалалар, эссе т.б. алынды, сонымен бірге кең тараған ресми ұлттық

газеттердің қоғамдық өмірге, мәдениет, әлеуметтануға, ішінара саясат пен экономикаға, экологиялық, мәдени, ахлақтық тәрбиеге т.б. алуан түрлі мәселелерге арналған мәтіндер жинақталды.

Көркем әдебиет стилінің базасына балалар жазушыларының прозалық, поэзиялық, драматургиялық шығармаларының мәтіндері, балалардың өз шығармашылығынан шыққан көркем туындылар, балабақша, мектеп қабырғасы, ЖОО-да баланың танып-білуіне ұсынылатын, ұсынылуы мүмкін рухани құндылығы жоғары ұлттық ақын-жазушылардың классикалық көркем туындылары, сол сияқты әлемдік классикалық әдебиеттің аудармаларының мәтіндері жинақталды.

Сөйлеу стиліне дидактикалық сөз (әңгімелесу, сұхбаттасу); көпшілікке арналған ауызша сөз (баяндама, лекция) мәтіндері енгізілді. Олар негізінен ауызекі тілден, сонымен бірге «Балапан» телеарнасының сценарийлері мен драмалық шығармалардан жинақталды.

Ортақ база – статистикасы стильдер бойынша алынбаған 5 стильден тұратын жеке база. Тілдік материалдарына жиілік сөздіктің алғашқы дайындығы кезеңінде арнайы стильдерге бөліп, бөлек жинақталмаған, бірақ барлық стиль мәтіндері теңгерімді түрде жинақталған қазіргі қазақ тілінің көріністерін, яғни тіліміздің қоғамдық өмірдегі негізгі салаларындағы қолданысын қамтитын теңгерімді мәтіндер берілді. Бұл мәтіндер базасының басқалардан айырым белгісі, мұнда қазіргі қоғамдық-әлеуметтік, қоғамдық-экономикалық сипаттағы мәтіндер көбірек қамтылды.

3-кесте.

Сөздіктегі стильдер және ортақ база мәтіндері көлемінің арақатынасы

Мәтіндер базасы	Үлесі, %	Сөзқолданыс саны
Ғылыми стиль мәтіндері	29,2	2 125637
Ресми-іскери стиль мәтіндері	8,3	609420
Көркем әдебиет стилі мәтіндері	10,5	746012
Публицистикалық стиль мәтіндері	9,2	669430
Сөйлеу стилі мәтіндері	6,4	465463
Ортақ база мәтіндері, <i>оның ішінде:</i>	36,4 <i>оның ішінде:</i>	2 658 927 <i>оның ішінде:</i>
ғылыми стиль мәтіндері	9,9	718840
көркем әдебиет стилі	6,6	484249
публицистикалық стиль	18,3	1342159
ресми-іскери стиль	0,9	62100
сөйлеу стилі	0,7	51579
Барлығы	100%	7 274 889

6. СӨЗДІКТІ ЖАСАУ КЕЗЕҢДЕРІ

Шартты түрде сөздіктің дайындалу үдерісін бірнеше кезеңге бөлуге болады:

1-кезең. *Мәтіндік база жинақтау.* Мәтіндік базаны жинақтауға құрастырушылар ерекше жауапкершілікпен қарады. Бұл кезеңдегі жұмыстар жоғарыда мәтін базасына қойылатын талаптарға сәйкес түрде атқарылды деуге әбден болады.

2-кезең. *Жиналған материалды қарапайым символдық талдауға икемді, элементтерінің жиілігін шығаруға қолайлы түрге келтіру.* Аталған кезеңде құрастырушылар екі мәселені шешті. Біріншісі – мәтінді элементтері талдау мен өңдеуге икемді түрге келтірілген символдар массивіне айналдыру.

Екіншісі – мәтінді икемдеу барысында суреттер мен кестелердің өзгеріп кеткен түсініксіз таңбаларынан, мағыналық жүгі жоқ символдардан, яғни «мәтіндік қоқыстан» тазарту. Бұл екі мәселе де жаппай барлық мәтіндерге қатысты бола берген жоқ. Кейбір мәтіндер керекті Word форматында қолжетімді болды. Алайда көптеген мәтін арнайы жинақталды, электрондық нұсқасы дайын күйінде қолымызда болмады, сондықтан оларды қағаз нұсқадан сканерлеу арқылы WORD, PDF форматына ауыстыру кезеңі және ауысқан мәтіндегі «қоқыстарды» тазалау ісіне құрастырушылардың біраз күші жұмсалды. **Мұнда да базасы цифрленген, электрондық нұсқаға салынған «Қазақ тілінің ұлттық корпусының» қажеттілігі қатты сезілді.**

3-кезең. *Басқа мағыналық элементтердің жиілік сипаттамасына кедергі келтіретіндіктен, мәтінді кейбір мағыналық жүгі бар элементтерден тазарту.* Жиілік шығаруға арналған базадағы тазартылуға тиісті мәтін элементтерін шартты түрде «мағынасыз» («қоқыс») және «мағыналы» деп екі түрге бөліп алдық. «Қоқыстарға» қандай таңбалалар жататыны жоғарыдағы жұмыс кезеңінің сипаттамасында айтылды. Ал «мағыналы» керексіз элементтерге мәтін ішіндегі кейбір қысқартып берілген сілтемелік қана жүгі бар сөздер, мысалы: «келесі бетті қара», «бб.»; латын қаріптерімен терілген сөздер: *SMALL, HP*, т.б.; математикалық формулалардағы латын-грек әліпбиінің әріп-таңбалары: β , ∞ , μ , π ; басқа да таңбалар: ®, ©, 1234567890, ***, т.б.

Мұндай таңбаларды мәтіннен тазалау барысында жинақталған тәжірибе келесі жиілік сөздіктер жасау ісін жеңілдетеді. Мәселен, осы таңбалардың, яғни «мағыналы», «мағынасыз» керексіз элементтердің тізімдемесін жасау аса қажет. Сонда келесі осындай жұмыстарда оларды компьютерден «іздеу» командасымен тауып алып, босаралықпен (пробел) автоматты түрде алмастырса, бұл кезеңге жұмсалатын уақыт пен күш әлдеқайда азаяды.

4-кезең. *Мәтіннен онимдер мен қысқарған сөздерді (аббревиатураларды) бөлек жинау.* Алғашқы екі кезең жұмыстарынан өткен мәтін ішінен Институттың ономастика бөлімі мамандары дайындаған «Ономастикалық стандарт» (ономастикалық атаулардың бір жүйеге түскен тізімі) негізінде табылған атаулар теріліп алынып, бөлек жиналды. Сонымен бірге қысқарған сөздер қолда бар тізімі бойынша мәтіннен алынып тасталды. Алайда қолда бар стандарт пен тізімде жоқ бұл атаулар мәтінде қалып қойғандықтан, сөзтізбеден де оларды тазалау қажет болды.

5-кезең. *Морфологиялық талдағыш арқылы сөз таптары мен омонимдерді ажырату.* Омонимдерді және сөз таптарын анықтау мақсатымен Институттың Компьютерлік лингвистика мамандары әзірлеген морфологиялық анализатор арқылы тазартылған мәтін өңдеуге салынды. Бұл процесс барысында мәтін құрамындағы морфологиялық талдағыш (анализатор) сөз таптары анықтаған сөздер мен омоним сөздерге шартты белгілер қойылды. Бұл істе талдағыштың әлі де жетілдіретін тұстарының көптігі байқалды. Себебі өңдеу барысында ол көптеген сөздердің сөз табы мен омонимдерді анықтай алмады. Төменде анализатор арқылы өңделген мәтіннің үлгісі берілді.

<Sent> Ерте-ерте ертеде, ешкі құйрығы келтеде, Қаратаудың ойында, Қарасудың бойында Қазанқап деген бір кедей болыпты. </Sent>

Ерте | Ерте <үс>

ертеде | ертеде <үс>

, |, <Punct>

ешкі | ешкі <зт>

құйрығы | құйрық <зт,ы/тжс-3+өас>

келтеде | келте <сн,де/жс>

, |, <Punct>

Қаратаудың | қаратау <гео,дың/іс>

ойында | ой <зт,ы/тжс-3+нда/жс>

, |, <Punct>

Қарасудың | қарасу <зт,дың/іс>

бойында | бой <зт,ы/тжс-3+нда/жс>

Қазанқап | Қазанқап <ғю,зт>

деген | де <ет,ген/е+ар+жжс-3>

бір | бір <ес>

кедей | кедей <сн>

болыпты | бол <ет,ып/к+ты/ар+жжс-3>

. | . <Punct>.

Сөйтіп, морфологиялық талдағышта сөз табы анықталмаған сөздер мен омонимдердің тізімі автоматпен жасалды. Оларды автоматты түрде жинап алу оңай, себебі анализатор мәтнінде мұндай сөздердің тұсында «*unknown*» – «белгісіз» деген таңба беріледі. Осы таңба арқылы олардың тізімі жасалды. Талдағышта анықталмаған («*unknown*»- «белгісіз» таңбасымен шыққан) сөздердің ішінен сөз табы белгісіздері мен ажырамаған омограф-омонимдер сұрыпталып, сөз таптары қойылды, омограф-омонимдердің сөз табына қатысы контекст арқылы қолдан ажыратылғандықтан, олардың сандық көрсеткіштерін автоматты түрде шығару мүмкіндігі туды. Бұл эксперимент жиілік сөздіктегі омонимдер жиілігін анықтауға септігін тигізді.

6-кезең. *Мәтінді сөздік пішінге келтіру.* Ол үшін алдымен автоматты түрде мәтін ішіндегі бас әріптер кіші әріптерге ауыстырылды. Сонан кейін мәтін ішіндегі

лексикалық бірліктер сөздік пішінге келтірілді, яғни бірінен соң бірі тұратындай бір бағанаға сөзтізбе форматқа түсірілді. Әр лексикалық бірлік тұсына екінші баған тіркеліп, оған сөздің жиілігі қойылды. Visual Basic for Applications (**VBA**), Excel, Notepad++, Word, Макростар бағдарламаларының көмегімен бірінші бағандағы бірдей сөздер анықталып, олар автоматты түрде біріктіріліп, қосынды жиілігі жазылды. Сөйтіп, сөздіктің сөзтізбесінің алғашқы нұсқасы дайындалып, тілтанымдық өңдеуге қайта берілді.

7-кезең. *Леммаға келтіру, сөз таптарын анықтау, омонимдер мен омографтарды ажырату.* **Сөздіктің сөзтізбесінің** алғашқы нұсқасы екі көшірмеде сақталды. Біріншісі бойынша **сөзформалар** жиілігі, ал екіншісі бойынша **негізсөздердің** жиілігі, стильдерге үлестірілген жиілік, әр стильдің жиілігі, ортақ база мәтіні жиілігі, сөз таптарының жиілігі мен омонимдер жиілігі жасалды. Негізсөз сөзтізбесінен, сөзформаларды қоспағанда, сөздіктің барлық 58 түрленімі жасалды. Сондықтан оның сөзтізбесін жасау ісі аса жауапты болды. Мұнда сөздер негіз формасына келтірілді. Сөзді **леммалау**, яғни леммаға (негізге) келтіру мен сөз табын қою үдерісі бір-бірімен тығыз байланысты болғандықтан, бірге атқарылды.

Қазақ тілінде ғана емес, әлемдік тіл білімінде жиілік сөздік құрастырудың сан алуан әдіс-тәсілі бар. Дәстүрлі реестр сөздерді тізімдейтін басқа да кез келген сөздік түрі сияқты жиілік сөздіктер де негіз сөздер бойынша құрастырылады. Сондықтан жиілік сөздіктер жасаудың қай тәсілін пайдаланса да, сөзформаларды негізге келтіру (лемматизациялау) процесіне тіреліп отырады. Леммалау әрекеті қолданбалы тіл білімі дамыған елдерде әдетте морфологиялық талдағыш көмегімен автоматты түрде жасалады. Алайда морфологиялық талдағышта сөз табы қойылмаған көптеген сөзді негізге келтіру кезінде олардың қай сөз табына жататынын анықтау қиынға соқты, тізімі алдыңғы (бесінші) кезеңде жасалып қойған мұндай сөздер құрастырушылардың қолын біраз ұстады. Мәселен, *ағын* сөзінің *ақ+ын/ТС* табыс септігіндегі субстантивтенген сын есімдік қолданысы (мысалы, *ағын таңдады*) мен *ағын* зат есімдік қолданысын (мысалы, *ағын су*) ажырату үшін сөздіктің құрастырушылары оларды мәтіннен тексеру жұмыстарымен отырды. Мұндай тексеру сөздіктің сапасын едәуір жақсартты, статистикалық мәліметтерінің дәлдік деңгейін арттырды. Сөйтіп, сөздіктегі сөздерді леммаға келтіру барысында *ал* (сын есім) – *ал* (етістік), *жаз* (етістік) – *жаз* (зат есім), *кез* (етістік) – *кез* (зат есім) сияқты **грамматикалық омонимдер**¹² мен *ағартушы* (*ағарт/ет+ушы* және *ағартушы/зт*), *адымдай* (*адымда/ет+й* және *адымдай/сн*), *айнала* (*айнал/ет+а* және *айнала/зт*) сияқты кейбір **омографтардың** (жалпы саны 230) қолмен ажыратылу ісі қатар жүргізілді. Сөзтізбеде жиілігі жағынан ажыратылған омонимдер жеке сөздік түрінде осы басылымның қосымшасындағы дискіде электрондық нұсқада берілді. Ал әрқайсысының жиілігі мәтіннен тексеріліп, анықталған кейбір омографтарды 4-кестеден көруге болады.

¹² Омонимдердің сөздігі жеке жасалды, осы кітапқа қосымша дискіде берілді.

Сөздікте мағынасы ажыратылған омографтар (бұл омографтардың бастапқы тізімі, ал жиілік сөздікте екі, үш сөз табы қойылған бұл сөздер жеке реестрде беріліп, көрсетілді)

СӨЗ/сөз табы	СӨЗ/сөз табы
абаймен/фо/үс	қиналыс/ет/зт
ағартушы/зт/ет	қисын/ет/зт
ағын/сн/зт	қоспа/зт/ет
адымдай/сн/ет	құйын/зт/ет
аза/зт/ет	құрал/зт/ет
айдай/сн/ет	қырын/сн/ет
айқындаушы/сн/ет	қызба/сн/зт/ет
айнала/зт/ет	қылмысты/сн/зт
алушы/зт/ет	қына/зт/ет
анада/үс/зт	қысқа/сн/зт
анықтау/ет/сн	қыстау/зт/ет
аңды/ет/зт	қыстық/сн/ет
ап/ел/ет	қышқылды/сн/зт
аптал/сн/ет	лақтырушы/ет/зт
аптап/зт/ет	мекенде/ет/зт
арнау/зт/ет	менгеруші/зт/ет
аршыл/сн/ет	модельде/ет/зт
арықта/ет/зт	модельдеуші/ет/зт
асқан/үс/ет	мойында/зт/ет
аспа/зт/сн/ет	мойындаушы/зт/ет
асты/зт/ет	мұңдас/сн/ет
асын/ет/зт	мүйіздей/ет/зт
асыра/үс/ет	мініс/зт/ет
атау/зт/ет	нашарлау/сн/ет
атта/ет/зт	негізде/ет/зт
аттан/од/ет/зт	оймақ/зт/ет
атты/сн/зт/ет	ойнақта/ет/зт
ауа/зт/ет	ойыс/зт/ет
ауған/зт/ет	оқтау/ет/зт
аударма/зт/ет	онда/үс/са/ес
ауыс/сн/ет	ондағы/ес/са
аштық/зт/ет	оңбаған/сн/ет

әрекеттесуші/зт/ет	орайлас/сн/ет
бейімдеу/сн/ет	оралма/ет/сн
бесті/зт/са	орама/зт/ет
борат/зт/ет	орма/зт/ет
бөгет/зт/ет	орманды/сн/зт
бөгеу/зт/ет	орталықтан/ет/зт
бөле/зт/ет	орында/ет/зт
бөлектеу/сн/ет	орынды/сн/зт
бұрау/зт/ет	орынсыз/сн/зт
бұрыс/зт/ет	орыстан/ет/зт
біліс/сн/ет	өндірі/зт/сн/ет
бітеу/сн/зт/ет	өтел/зт/ет
бітірі/зт/ет	пысықтау/ет/сн
езбе/сн/ет	пікірлес/сн/ет
езу/зт/ет	серпіліс/зт/ет
екпе/сн/зт/ет	сипат/зт/ет
елес/зт/ет	соғушы/зт/ет
еміс/үс/ет	соғысушы/зт/ет
ене /зт/ет	сойыс/зт/ет
енші/зт/ет	соқтығыс/зт/ет
есепті/сн/үс/зт	сұқпа/зт/ет
ескерту/зт/ет	сұрау/зт/ет
етекте/ет/зт	сыйлас/ет/сн
дегендей/үс/ет	сыйын/ет/зт
демеу/зт/ет	сынама/зт/ет
егеу/зт/ет	сілтеме/зт/ет
жазба /зт/ет	сіңірі/ет/зт
жақтас/сн/ет	табу/зт/ет
жақтау/зт/ет	табысты/сн/ет
жалау/зт/ет	тақау/үс/ет
жалдау/зт/ет	талды/сн/зт
жанар/зт/ет/іо	талма/зт/ет
жанды/сн/зт/ет	таңдай/зт/сн
жаппа/зт/сн/ет	тапсырма/зт/ет
жарау/зт/сн/ет	тарау/зт/ет
жаудыр/сн/ет	тасы/ет/зт
желе/зт/үс	татар/ет/зт
жеміт/зт/сн	тату/сн/ет

жетекте/зт/ет	теңеу/зт/ет
жуа/зт/ет	тоқыл/зт/ет
жұмбақта/зт/ет	тоқыма/зт/ет
жылын/зт/ет	туған/зт/ет
жырға/зт/ет	туында/зт/ет
жыртыс/зт/ет	тұжырымда/зт/ет
зерттеуші/зт/ет	тұсау/зт/ет
илеу/ет/зт	тәркілеу/зт/ет
іі/ет/зт	тұманда/зт/ет
кезде/ет/зт	түйе/зт/ет
кездік/ет/зт	түлен/зт/ет
кезеп/зт/ет	түрме/зт/сн/ет
кезік/ет/зт	түсте/зт/ет/үс
келісімді/сн/зт	түсуші/зт/ет
кеспе/зт/ет	түсіністік/зт/ет
кешпе/сн/ет	тізбе/зт/ет
көктеу/зт/ет	тілеу/зт/ет
көрші/зт/ет	тіреу/ет/зт
көрін/зт/ет	тіркес/зт/ет/сн
көсеу/зт/ет	тіркеу/ет/зт
көшір/зт/ет	ұлу/ет/зт
кіріс/зт/ет	ұлыс/зт/ет
кіріспе/зт/ет	ұмтылыс/ет/зт
қадау/зт/сн/ет	үгіт/зт/ет
қаза/зт/ет	үндеу/ет/зт
қаймақта/ет/зт	үнді/зт/сн
қайшылас/сн/ет	үстеме/зт/ет
қаласыз/сн/ет	шабыс/зт/ет
қалау/зт/ет	шаға/зт/ет
қалдық/зт/ет	шөлде/зт/ет
қалмақ/зт/ет	шұбат/зт/ет
қамырық/зт/ет	шылау/ет/зт
қанар/зт/ет	шынтақта/ет/зт
қандас/зт/ет	шыңда/ет/зт
қанды/сн/зт	шытын/ет/зт
қандық/зт/ет	шілікті/сн/зт
қаптал/зт/ет	ызылда/ет/зт
қарқынды/сн/зт	ықпалдас/сн/ет

қарман/ет/зт	ылғалда/ет/зт
қату/сн/ет	ылғалды/сн/зт
қатысты/сн/ет	ілбі/ет/зт
қашаған/сн/зт/geo	ілгерілеу/ет/сн/зт
қашау/ет/зт	іле/geo/үс/ет
қима/зт/ет	ілемк/зт/ет
	электролизде/ет/зт

Әрбір сөз табына қатысты леммалау кезіндегі қиындықтар мен мәселелер және оларды шешу жолдары туралы **сөз таптары жеке сөзтізбесінің ерекшеліктері** туралы төмендегі тараушада жан-жақты айтылды.

8-кезең. Статистикалық мәлімет алу. Сөзтізбенің құрамы осы кезеңдегі өңдеуден өткізілгеннен кейін сөздіктің статистикалық деректері мынадай формула бойынша есептелді:

Абсолютті жиіліктер қосындысы алынған мәтіндер көлеміне тең:

$$N = \sum_{i=1}^n f_i = f_1 + f_2 + f_3 + \dots + f_n;$$

F_i – абсолютті жиілік.

$$\text{Қатынастық жиілік: } F_i = \frac{f_i}{N};$$

$$\text{Жиынтық қатынастық жиілік: } \sum F_i = \sum_{i=1}^n \frac{f_i}{N} = 1.$$

Сөздік бірлігінің мәтінді қамту пайызының пайыздық салмағы:

$$\sum_{i=1}^n F_i \cdot 100\% = \sum_{i=1}^n \frac{f_i}{N} = 100\%.$$

Көрсетілген формулалар арқылы сөздіктің статистикалық деректері мен бір топ сөздің қайталана кездесу жиілігі негізіндегі олардың мәтінді қамту пайызы есептеліп, жиілік сөздікте көрсетілді.

Осы тұста **бір топ сөздің мәтінді қамту пайызы** ұғымына түсінік бере кету сөздіктің тұтынушысы үшін пайдалы болмақ. Бұл терминнің мәнін басқаша түсіндірген болса, оны сөздік сөзтізбесіндегі қатарынан тұрған сөздердің мәтінді түсінуге қатысты үлесінің көрсеткіші деуге де болады. Бұл көрсеткіш жеке сөздерге ешқандай қатысы жоқ, ол басынан бастағандағы бір топ сөзге қатысты жиынтық көрсеткіш болып табылады.

Мәселен, жалпы әліпбилі-жиілік сөздіктегі *абадан* сөзі 7 млн сөзқолданыстан астам мәтінде небары 2 рет қолданылады, демек оның реестр басында тұрғандағы мәтінді қамту пайызы 0,00003 – өте төмен. Сонымен бірге ол әліпбилі-жиілік сөздікте бірінші сөз. Яғни, басқаша айтқанда, сөздіктегі осы бірінші тұрған сөзді ғана білу мәтіннің 0,00003 пайызын түсіну мүмкіндігін береді. Ал егер осы бірінші тұрған сөзбен қоса *абажардай*, *абажурлы* 2-3-сөзді де танып-білу мәтінді түсіну пайызын сәл болса да арттыра түседі, 0,00015 болады. Бірақ бұл өте төмен

көрсеткіш, 1-ге де жетпейді, себебі бұл сөздердің жиілігі өте төмен (біреуі 2 рет, екіншісі 6 рет, үшіншісі 3-ақ рет қолданылған).

Ал жалпы жиілікті-әліпбилі сөздікте *абадан* жиілігі төмен сөз ретінде соңғы жақта 24316-ретте тұр, яғни сөздіктің бұл типіндегі сөзтізбеден бірінші тұрған, яғни *бол* сөзінен бастап осы 24316-реттегі, яғни *абадан* сөзіне дейін меңгерген адам мәтіннің 99,7 пайызын түсінеді деп жорамалдауға әбден болады. Бір топ сөздің мәтінді қамту пайызы деген көрсеткіш осындай практикалық маңызға ие. Төменде сөздіктің статистикалық сипаттамасында жалпы жиілікті-әліпбилі сөздіктің негізсөз сөзтізбесінің осы көрсеткіш бойынша кестесі берілді¹³.

9-кезең. Негізсөз және сөзформалар сөзтізбесін жүйеге келтіру. Негізсөз, яғни екінші нұсқадағы **сөзтізбе** әліпбилі-жиіліктік типте әзірленді. Оның құрамында сөйлеу стиліне, бейканондық лексикаға қатысты сөздердің орфографиялануының алуан түрлі варианттары, жаргон, диалект сөздер, яғни бейәдеби тілдің лексикалық бірліктер **сараптама** барысында өңделді, біразы сөзтізбеден алынды.

Ал сөздік **сөзтізбесінің** бірінші нұсқасында, яғни **сөзформалар** тізбесінде сөздер леммаға келтірілмеген күйде сақталды. Бұл сөзтізбені сөзформалар құрады. Оның құрамында да жоғарыда айтылған бейәдеби тілдік бірліктер болды. Сөзтізбенің екі нұсқасы да осындай бірліктерден тазаланды. Алайда сөзформалар сөздігінде ішінара қиылып тасталған өте сирек қолданыстағы тым пәс стильдегі бірліктер болмаса, бұлардың біразы сақталды. Оның себебі сөзформаларда сөзтүрленімнің барлық ерекшелігін қамту мақсат етілді. Екі сөзтізбедегі аздаған айырмашылыққа қарамастан, олар ортақ принциптерге сүйеніп, құрастырылды.

Жиілік сөздік сөзтізбелерін құру п р и н ц и п т е р і :

1) дереккөзде бірнеше нұсқада таңбаланған сөздердің дұрыс нұсқасын тандап алу үшін олар канондық және бейканондық лексикалық бірліктер түрінде екі топта қарастырылды. К а н о н д ы қ л е к с и к а ға ¹⁴ жататын сөздер, жиілік сөздік базасындағы мәтіндерде әртүрлі таңбаланғанына қарамастан, орфографиялық сөздіктің соңғы басылымы бойынша қазақ әдеби тілінің жазу нормасына сәйкес орфограммаларға келтірілді¹⁵. Мәтін базасында түрлі варианттармен мынадай сөздер қолданылды:

- ескі кітаби тілдің бірліктері (*ажым* емес, *әжім*; *ғәдауат* емес, *гадауат*; *әбілқаят* емес, *абылхаят*),
- игерілген орыс сөздері (*машина*, *мәшін* емес, *мәшине*; *режим* емес, *режім*; *костюм* емес, *кәстөм*, *кострюль* емес, *кәстрөл*),
- сөйлеу стилінің бірліктері (*барад* емес, *барады*, *кәйтін* емес *қайтып*, *үйткені* емес, *өйткені*),
- диалектілік фонетикалық варианттар (*дізе* емес, *тізе*; *дұз* емес, *тұз*; *жәман* емес, *жаман*),

¹³ 10-кестені қараңыз.

¹⁴ Канондық лексика деп жазылуы қатаң түрде орфографиялық нормаға бағынуы тиіс, олай болмаған жағдайда, мәтін ішінде жазбаформалық бейнесіне нұқсан келетін лексиканы айтамыз, бұған көбінесе толық лексикалық мағыналы атауыш сөз таптары мен грамматикалық мағынасы доминант болатын көмекші сөз таптарының бірліктері жатады.

¹⁵ Орфографиялық сөздік. Алматы: Дәуір, 2013.

- босан орфографиялық норма бірліктері (*шекара* емес, *шегара*; *көкөніс* емес, *көгөніс*, *қыркүйек* емес, *қыргүйек*),
- кейбір қатаң орфографиялық норманы БАҚ-та жаппай бұзушылықтан туған қате жазу үлгілері (*жәй* емес, *жай*; *шәй* емес, *шай*; *жәйт* емес, *жайт*) т.б.

Олардың барлығы орфографиялық нормаға сәйкес түрде берілді. Бұл жерде ескерте кететін жайт, орфографиялық сөздікте *балуан*, *палуан*, *айғай*, *айқай* сияқты т.б. жарыспалы вариант түрінде көрсетілген сөздердің варианттары мәтінде берілген күйінде түгел сақталды.

Сөздіктегі сөздерді нормаға сәйкес түзету оның тілтанушы үшін ізденіс материалы ретіндегі құндылығына нұқсан тигізетіні белгілі, себебі сөздердің дәл мәтіндегідей көрінісі сақталмайды, әсіресе жазудағы тіл бірліктерінің қате варианттарын талдауға, ортологиялық зерттеулерге қатысты мәліметтері дәйекті бола алмай қалады. Алайда сөздіктің мақсаттарының бастысы жалпы білім берудегі әдістемелік құралдарға негіз жасау болғандықтан, әдеби тіл нормаларының доминант болуы басымдыққа алынды.

2) бейканондық лексикаға¹⁶ жататын сөздердің сөзтізбедегі барлық орфограммасы орфографиялық сөздікке қатаң түрде сүйеніп берілген жоқ. Олардың кейбірі мәтін ішіндегі нұсқада таңбаланып кетті. Бұл – олардың жазудағы варианттарының өзгергенімен мағынасының өзгермейтінімен, оқырманның оны тану процесіне бұдан еш кедергі болмайтынымен байланысты болған босандық. Мысалы: *ойбу*, *ойбуй*, *үйбай* одағайлары, *бырқ-сырқ*, *дабыр-күбір* сияқты еліктеуіштер орфографиялық сөздікте берілмеген, ал *салдыр-гүлдір* еліктеуіші орфографиялық сөздікте *салдыр-күлдір* түрінде көрсетілген. Алайда бұлардың фонетикалық варианты түсінуге кедергі келтірмейтіндіктен, әрі бір компоненттерінің этимологиясы күңгірт болғандықтан, мәтіндегі нұсқасы сақталды.

3) қазақ тілінің грамматикалық нормасына қайшы, бірақ жиілігі жоғары сөздер сөзтізбеде сақталды. Мәселен, қазіргі мәтіндерде жиі кездесетін *азаматша*, *төрайым* сияқты сөзтізбедегі қолданыстар **қ а т е** екенін ескертеміз, тілімізде род категориясы жоқ болғандықтан, норма бойынша *азамат*, *төраға* түрінде қолданылуы керек. Алайда осындай теріс үрдістің көрінісін сөздіктен табу тілтанушы ізденістері үшін керек болады деген мақсатпен оларды қате болса да, сөзтізбеде қалдыруға тура келді;

4) кейбір социолектілік мәні бар сөздер сақталды. Мәселен, қарттардың тілінде қолданылатын *кәрия*, *бейіл*, *бейілді*, *бейіш* нормадағы *қария*, *пейіл*, *пейілді*, *пейіш* формасы да берілді;

5) сөзтізбеде қазақша баламасы терминком арқылы бекіген орыс сөздері берілмеді. Мысалы, *анкета* (сөздікте *сауалнама*), *администрация* (сөздікте *әкімшілік*), *нерв* (сөздікте *жүйке*), *система* (сөздікте *жүйе*). Олар сөзтізбеден бөлек алынды;

¹⁶ Бейканондық лексика деп жазылуы қатаң түрде орфографиялық нормаға бағынбаса да, мәтін ішінде жазбаформалық бейнесіне нұқсан келмейтін лексиканы айтамыз, бұған еліктеуіш сөздер мен одағайлар, қаратпа сияқты қыстырынды құрылымдар жатады.

6) алдыңғы принципке қарсы кейбір советизмдер мен баламасы бола тұра, мәтіндерде бүгінгі күнге дейін қазақшасымен жарысып, орысша нұсқада жиі қолданылып жүрген әсіресе тарихқа қатысты ұғым, түсінік атаулары сақталды. Мұндағы мақсат, біріншіден, Кеңестік кезеңдегі мәтіндердің тарихи шындығын ескеру, себебі оқушылар тарих пәнінде оларды өтпей қоймайды (мысалы, сөзтізбедегі *совхоз, стахановшы, комсомол, комсорг* сөздері), екіншіден, сөздік тұтынушысын қазіргі мәтіндердегі терминдік мәні бар, аудармасы жаппай қолданысқа енбей жатқан сөздерден де (мысалы, сөзтізбеде *территория, аумақ* та бар; *мораль, ахлақ* та бар; *космос, ғарыш* та бар) бейхабар қалдырмау;

7) қазақша баламасы жоқ, халықаралық терминдік жүйеден орыс тілі арқылы енген терминдер кеңбейімді, тарбейімді қолданыс аясына қарамастан, мәтінде кездесу көлеміне қарай сөзтізбеде түгел қамтылды, мысалы: *интонация, агробио-ценоз, адсорбент*, т.б.;

8) қазақша баламасы бекіп, кең қолданыс алған орыс тілінен келген термин сөздердің орысша түбіріне (негізіне) қазақ тілінің сөзжасамдық жұрнақтары жалғануы арқылы жасалған түрленімі сөзтізбеде сақталды: *советтік, глобальды, структуралық*, т.б.).

10-кезең. *Техникалық жөндеу, баспаға әзірлеу.* Бұл кезеңде сөздіктің 61 түрленімі түгенделіп, олардың сөзтізбесі мен статистикалық мәліметтері соңғы тексеруден өткізіліп, қолжазба баспаға тапсырылды. Бұл кезеңде сөздіктің қағаз нұсқасы беттеліп, беттелген мәтін техникалық жағынан редакцияланды. Ал электрондық нұсқадағы сөздіктің қалған 59 түрленімі дискіге жазылды.

7. СӨЗДІК ТИПТЕРІ, ТҮРЛЕРІ, ТҮРЛЕНІМІ

«Жалпы білім берудегі қазақ тілінің жиілік сөздігі» 3 түрлі типте әзірленіп, 7 түрге топтастырылған 61 түрленім бойынша ұсынылып отыр.

Сөздіктің барлық түрі мынадай үш типте жасалды:

Әліпбилі-жиілік сөздіктің сөзтізбесінде сөздер бірінші әрпінен бастап соңғы әрпіне дейін қатаң әліпби тәртібімен орналасады. Әліпби тәртібімен берілгендіктен, ізделген сөзді немесе сөзформаны табу жеңіл болады, алайда оның мәтінде қолданылу жиілігі туралы ақпарат берілгенімен, жалпы жиілікте орны қандай екені туралы мәлімет алу қиын, сондықтан жалпы жиілікті-әліпбилі сөздікте әрбір сөздің әліпбилі-жиілік түріндегі реттік нөмірі көрсетіліп отырды.

Жиілікті-әліпбилі сөздік сөздердің мәтіндегі кездесу жиілігінің кему ретімен орналасып, ал бірдей жиілікті сөздер қатаң әліпби тәртібімен орналасады. Жиі қолданыстағы сөздер сөзтізбенің басында орналасады, оларды қарым-қатынас жасау кезіндегі белсенді сөздер ретінде тануға болады. Жиілігі жоғары сөздер мен сөзформалар (сөзформалар сөздігінде) әртүрлі стильдегі күнделікті қарым-қатынастың негізгі сөздік және грамматикалық қорын құрайды деп батыл айтуға болады. Ал белсенді лексикалық қор тілді оқытуда, әртүрлі тілдік оқулықтар мен оқу құралдарын, лексикалық минимумдар, оқу сөздіктерін құрастыруда негізгі құрал қызметін атқарады, яғни коммуникативтік актіні ұйымдастырудың тілдік негізі болып табылады.

Керіәліпбилі-жиілік сөздікте керісінше сөздер соңғы әрпінен бастап алғашқы әрпіне қарай қатаң әліпби тәртібімен орналасады. Сөздікте сөзтізбедегі сөздер оң жағы (соңғы әрпі) бойынша реттеледі. Бұл сөздікте сөздер немесе сөзформалар соңғы әрпінен бастап алғашқы әрпін қоса есептегенге дейінгі аралықта қатаң әліпби тәртібімен орналасады. Керіәліпбилі-жиілік сөздікте әрбір сөзтудырушы жұрнақтың немесе сөзтүрлендіруші қосымшаның (сөзформалар сөздігінде) қолданылу жиілігі туралы мол ақпарат беріледі. Бұл сөздіктен зерттеуге алынған мәтіндердегі сөзтудырушы және форматудырушы морфемалардың құнарлы не құнарсыз, белсенді не салғырт екенін анықтауға, терминдер мен кірме сөздерге жалғанған сөзтудырушы, сөзтүрлендіруші қосымшалардың құрылымын талдауға қатысты бағалы тілтанымдық деректер алуға әбден болады.

Сөздіктің түрлері

1) ЖАЛПЫ ЖИІЛІК. Мұнда 7 млн сөзқолданыстан тұратын теңгерімді мәтін негізінде алынған түбірге келтірілген (леммаланған) сөздердің жиілігі көрсетілді. Бұл сөздіктің сөзтізбесінде 36265 сөз бар. Сөзтізбедегі сөздердің стильдерге қатысты жиіліктері үлестіріліп көрсетілмей, жалпы алынды. Жалпы жиілік сөздіктің әліпбилі, жиілікті және керіәліпбилі үш түрленімі берілді. Олардың ішінде «Жалпы жиілікті-әліпбилі сөздік» осы кітапқа еніп отыр, яғни қағаз нұсқада жарияланып отыр. Жалпы жиілікті сөздіктің қалған екі типі осы кітапқа қоса дискіде электрондық нұсқада берілді.

2) СТИЛЬДЕР БОЙЫНША ҮЛЕСТІРІЛІМ ЖИІЛІК. Бұл сөздікте сөзтізбедегі 36265 сөз қазақ тілінің ғылыми, көркем әдебиет, ресми-іскери, публицистикалық, сөйлеу – 5 фунционалды стилі бойынша таратылып көрсетілді. Сөздіктің аталған түрінің ыңғайлылығы сонда, онда әрбір сөздің осы бес стильдегі қолданыс жиілігі бір паракта беріледі, сөйтіп, сөздікті пайдаланушы әрбір сөздің түрлі стильдегі мәліметтерін салыстыра отырып, оның әрқайсысындағы «қызметі мен өмір сүру деңгейінен» құнды статистикалық ақпарат ала алады. Мәселен, *автокөлік* сөзі көркем әдебиет стилі мәтіндерінде мүлде кездеспесе, ең көбі ғылыми стильде (82 рет) қолданылған. Сол сияқты стильдер бойынша үлестірілім жиілік сөздіктің әліпбилі, жиілікті және керіәліпбилі үш түрленімі берілді.

3) ҚАЗАҚ ӘДЕБИ ТІЛІ ФУНКЦИОНАЛДЫ СТИЛЬ МӘТІНДЕРІНІҢ ЖИІЛІГІ. Сөздіктің осы түрінде әрбір стильге тән болған сөздердің жеке жиілігі үш түрлі типте берілді. Яғни ғылыми, ресми-іскери, көркем әдебиет, публицистикалық, сөйлеу стильдері мәтіндерінің әліпбилі-жиілік, жиілікті-әліпбилі және керіәліпбилі-жиілік сөздіктері топтастырылды. Бұл әсіресе оқулық авторлары үшін таптырмас қосымша құрал болмақ, себебі жекелеген стильге тән мәтін құрастыру үшін оның өзіндік сөз бедерін, лексикалық ерекше құрамын сақтау аса қажет, ал жиіліктегі сөздерді қолдану бұған көмегін тигізеді.

4) ОРТАҚ БАЗА (БАРЛЫҚ СТИЛЬДЕРДЕН ТЕҢГЕРІМДІ ТҮРДЕ ҚҰРАЛҒАН) МӘТІНДЕРІНІҢ ЖИІЛІГІ. Ортақ база мәтіндерінің жиілігі стильдер бойынша үлестірілім сөздікте де, сол сияқты өзі жеке де берілді. Өзін жеке

алғандағы мақсат барлық стильдерден теңгерімді түрде аралас жинақталған заманауи мәтіндердің ауқымды базасынан жиілікті көру. Сонымен бірге осы базадан қазіргі қоғамдық-экономикалық, қоғамдық-әлеуметтік лексикалық бірліктердің белсенділігін аңғару оңай, себебі оның базасы қоғамдық-гуманитарлық, әлеуметтік-экономикалық дискурс мәтіндеріне негізделді. Төмендегі кестеден әрбір стиль және ортақ база мәтіндерінен алынған жиілік сөздіктердің сөзтізбесіндегі сөз санын және жалпы абсолютті жиілікті көруге болады.

5-кесте.

Қазақ әдеби тілінің функционалды стильдері және ортақ база мәтіндерінің жиілік сөздігі (15)

	Сөзтізбедегі сөздер саны	1 рет қолданылған сөз саны	Жалпы абсолютті жиілік
Ғылыми стиль	22 159	4 965	1 776 024
Іскери стиль	7 488	1 804	465 324
Көркем стиль	12 289	3 146	435 033
Публицистикалық стиль	15 079	4 132	502 352
Сөйлеу стилі	9 370	2 627	301 619
	Сөзтізбедегі сөздер саны		Абсолютті жиілік
Ортақ база	24 365	5 691	2 522 467

Бір рет қолданылған сөз саны жағынан публицистикалық мәтіндер қатты ерекшеленетіні байқалып тұр, яғни бұл мәтіндерде 15 079 сөздің 4 132-сі бір рет қолданылса, сонда 10 974 сөз небары 498 220 (502 352-4 132) жалпы абсолютті жиілікпен қолданылған, бұл мәліметтер публицистикалық стильде бірегей сөзқолданыстардың көптігін, сондай-ақ дәстүрлі клишелік сөзқолданыстардың аздығын байқатады. Дәл осындай картинаны сандық мәліметтер көркем әдебиет стилі мәтіндері жиілігінен де көрсетеді.

5) СӨЗ ТАПТАРЫНЫҢ ЖИІЛІГІ. Бұл сөздіктің 33 түрленімі жасалды. Бұл түрленім **екінші сөзтізбеден**, яғни негізсөз сөзтізбесінен жасалды. Сондықтан әрбір сөз табына қатысты сөзтізбе негізсөз сөзтізбесінің ерекшелігін көрсетеді. Сөз таптарына жатпайтын көмекші есімдер мен модаль сөздер грамматикалық категориясы да осы түрге еніп кеткенін ескеру керек. Сонда жиіліктің бұл түрі бойынша **зат есім, сан есім, сын есім, есімдік, етістік, үстеу, еліктеуіш, одағай, шылау** сияқты сөз таптары және **көмекші есімдер мен модаль сөздердің**, яғни 11 аталымның әліпбилі-жиілік, жиілікті-әліпбилі, керіәліпбилі-жиілік сөздіктері жасалып, электрондық нұсқада ұсынылды.

Зат есімнің жиілік сөздігі сөзтізбесінің ерекшеліктері:

– Сөзтізбеде зат есімнің барлық семантикалық топтары қамтылды. Сол сияқты кейбір мағынасы жеке тұрғанда күңгірт сөздердің сөзтізбеде орын алуының өз

себебі бар. Мәселен, қазіргі қазақ тілінде *үлде* сөзі жеке қолданылмайды, алайда оның сөзтізбеде жеке тұрған себебі *үлде мен бүлде* фразеологизмінің компоненттері леммалау үдерісінде жеке-жеке сөз ретінде реестрге алынады, дәл осылай бір ғана денотат атауы болып табылатын, бірақ бөлек тұлғаланатын *киіз үй, ас үй, таңғы ас* сияқты лексикаланған сөз тіркестерінің, *айна қатесіз, қол алысу, көз көрген* сияқты бөлек жазылатын идиомалардың компоненттері *киіз, үй, ас, айна, қатесіз, қол, ал, көз, көр* түрінде жеке-жеке леммаға келтіріліп, әрқайсысының жиілігі бөлек саналғанын ескертеміз;

– алдындағы тармақта сөз болған бір ұғым атауы болып табылатын күрделі сөз компоненттерінің жеке тұлғаланғанына қарамастан, кейбір сөздердің тұтастығы қоснүкте арқылы беріліп, сақталды. Бұл сөздер тіркес бола тұра, оларға сөз табы көрсетілуінің себебі бірінші компоненті өзі бөлек тұрғанда тұрпаты жағынан көзшаламға мүлде түсініксіз, бейтаныс болып келетінімен байланысты, мысалы, *о: о дүние/зт, кирил: кирил жазуы/зт*. Ескерте кететін жайт, *бұ дүние* қолданысы болғанымен *бұ* жеке есімдік ретінде берілді, себебі ол басқа контекстерде, әсіресе сөйлеу стилінде өзі жеке тұрып та жиі қолданылады, ал *о* түріндегі *ол* есімдігінің редукцияланған формасы мәтін базасында тек *о дүние* контексінде кездесті, сондықтан ол қоснүкте арқылы таңбаланды;

– етістік мәнінде де, сол сияқты зат есім мәнінде де жұмсалатын тұйық етістіктің *у-тұлғалы* формасынан жасалған омоним қатары ажыратылды: *сұрау/зт, сұрау/ет*, сөйтіп *-у* тұлғалы зат есімдер мәтіндегі контексінен анықталды;

– *қазақ, орыс, қытай* сияқты ұлт аттары берілді, алайда *керей, найман* сияқты ру аттары сөзтізбеге алынбады;

– ұғым, денотат не конататқа атау болуға жарайтын, жаңа сөз жасау әлеуеті бар жергілікті сөздер сөзтізбеден алынған жоқ, мәселен, *қапақ* – ‘ішін үңгіген асқабақ’.

– контекст талғамай, кез келген етістіктің түбіріне жалғана алатын *-ушы* функционалды жұрнағы арқылы жасалған, бірақ зат есім категориясына өткен немесе аралық кезеңде тұрған кейбір сөздер зат есім ретінде таптастырылды: *ойлаушы, қолданушы, қоныстанушы, монтаждаушы, нұсқаушы*, т.б. Осы *-ушы* формасындағы зат есімдер мен есімшелерден тұратын омографтар ажыратылды, бұл үшін мәтін контексінде қолмен жұмыс жүргізілді, есімшелер етістіктің негізіне *қорға, жаса, көр* түрінде леммаға келтірілді, ал *қорғаушы, жасаушы, көруші* сияқты т.б. заттық ұғым атаулары зат есімнің сөзтізбесіне алынды.

– қазақ орфографиясында бас әріп орфограммасымен бекіген сөздерді сөзтізбеде бас әріппен бердік: *Алла, Құран* т.б.;

– *қалып* және *қалпы* сөздері де жеке берілді, мұндағы «қалпы» «қалыптың» тәуелсіздіктің 3-жағындағы фонетика-грамматикалық түрленімі емес, себебі *қалпына келу, сол қалпы* сияқты тіркесте оның «қалып» түбірінен (*қалыпқа салу - қалыбын жасау*) мағынасы бөлек; сонымен бірге құрамында тәуелдіктің 3-жағының грамматикалық тұлғасы сөзжасамдық рөл алған *қарасы/зт, қарауы/зт* сияқты сөздер де осылайша тәуелдік формасын сақтаған күйі леммаға келтірілді (мысалы, *көздің қарасы, қарауында болу*);

– орфографияда бір тұлғасы бекігеніне қарамастан (*үлгерім*), кейбір фонетикалық варианттар арнайы берілді, мысалы, *үлгерім* және *үлгірім* сөздерінің

соңғысының жиілігі 1-еу ғана, бұлар орныққан вариантты растау үшін арнайы сақталды;

– кейбір қазақша баламасы бекіген советтік кезеңнің тарихи картинасына қатысты және сол кезеңдегі қазақ мәтіндерінде түпнұсқа тілдегідей жазылуға тиіс болған орыс сөздері мәтіндегідей нұсқада сақталды: мысалы, сөзтізбеден *марксист, систематика, квартирант, этикетка, коминтерн*, сол сияқты *конфискация (тәркіле де бар), класс (сынып, тап та бар), сигнал (белгі де бар), зоопарк (зообақ та бар), революция (төңкеріс те бар), план (жоспар да бар), регулятор (реттегіш те бар), кубик (текше де бар), космос (ғарыш та бар), алфавит (әліпби де бар), ставка (мәлишерлеме де бар), стадия (кезең де бар), спутник (жерсерік те бар), армия (әскер де бар), экзамен (емтихан да бар), скиф (сақ та бар), география (жағрафия да бар)* сөздерін көруге болады;

– қазақша жазу үлгісінде игеріліп, орнығып қалған орыс сөздерінің қазақи үлгісі, орфографиядағы нормасына қайшы болғанына қарамастан, мәтіндердегідей нұсқада сақталды: сөзтізбеде *манжета* емес, *манжет, монета* емес *мәнет, папах* емес, *папах*;

– қаратпа атаулар зат есім ретінде таптастырылды: *мәртебелім/зт, қарақтарым/зт, тұлпарым/зт, ағатай/зт* т.б.;

– сөйлеу тілі арқылы игеріліп, жазуда да қазақ әріптері арқылы таңбаланып жүрген орыс сөздері сөзтізбеде мәтіндегі нұсқасымен берілді: *жүлік, пәшкі, пірәндік, итан, атбекет, доқтыр, кәртішке, банды, мөшек, кәсек* – бұлар ауызекі тілдегі шеттілдік сөздің фонетикалық жолмен игерілу үдерісінен ақпарат беретін тілтанушы ізденістері үшін қажетті мәлімет ретінде сақталды;

– кейбір бейонимденген және мифтік атаулар зат есім ретінде берілді: *мұңкір-нәңкір/зт, қожанасыр/зт, Хауана/зт*;

– ескікітаби лексикалық бірліктердің мәтіндердегі жазылу варианттары сақталды: *дүния (дүние), хақиқат (ақиқат), кәмалат (кәмелет), машқар (машар), кәлам (қалам)* т.б.

– жиілігінің төмендігіне қарамастан, ескікітаби лексиканың терминдік әлеуеті жоғары бірліктері сөзтізбеде сақталды: *әмалият* (практика мағынасында, теория мағынасындағы *назарияның* қатарында);

– XIX ғасырдағы тарихи мәтіндерінде сол кездің ауызша сөзі бойынша игерілген, бірақ бүгінде тілімізде генерал, генерал-губернатор формада бекіген *генарал, генарал-губернатор* сөздері ескі қолданысы бойынша берілді, себебі олардың мәтіндегі жиілігі жоғары болды.

– қазір өзектіленіп отырған ескікітаби лексикалық қабаттағы діни мазмұндағы ұғым, түсінік атаулары және олардың қазақша үлгідегі варианттары берілді: *мүкәрәм, мүррид, мүмін/мүмин, сүннет/сүннит, сүннеттік/сүнниттік*;

– кейбір диалект сөздер мен сөйлеу тілінің элементтері: *бұлаңшы/зт* – қытай қазақтары тіліндегі бүлікші мағынасындағы сөз, *насам* – ‘ірі, үлкен’ – мағынадағы жергілікті сөз, *ешен* – ишанның сөйлеу формасы;

– *америкалық* сөзі омоним ретінде ажыратылды, оның заттық мағынасы мәтіндерде ‘Америка мемлекетінің тұрғыны’, ал сын есімдік мағынасы ‘осы құрлыққа қатысты’ деген ұғымда жұмсалды;

– *азаматша, төрайым, әділқазы, әділқазылық* сияқты тіл мәдениетінің, ұлттық грамматиканың нормаларына қайшы **қ а т е** қолданыстардың сөзтізбеде сақталғанына ерекше назар аударамыз, мұндағы мақсат – қазіргі қазақ мәтіндерінің ортологиялық сапасына қатысты тілтанымдық ізденістерге қажетті ақпаратты көрсету.

– қазақ орфографиясындағы жарыспалы варианттар қалдырылды: *айғай, айқай*;

– сөзтізбеден жаңа сәтті терминдер мен жаңа сөздерді де көруге болады: *жақпамай, ойбайлам, ойтүйсік, ойтаным* т.б.;

– жалпы есімдерге жалқы есімдік жұрнақ жалғану арқылы жасалған тәуелсіздік алған кезден бері қоғамдық санада өзектіленген қоғамдық-саяси, қоғамдық-экономикалық құбылыстардың жаңа атаулары сөзтізбеден орын алды: *желтоқсандық, желтоқсаншы, алашордалықтар, алашордашыл, ақорда* т.б.;

– қазіргі қазақ тілінде *үлде* сөзі жеке қолданылмайды, алайда сөзтізбеде оның жеке тұрған себебі *үлде мен бұлде* фразеологизмінің компоненттері леммалау үдерісінде жеке-жеке сөз ретінде реестрге алынады, дәл осылай бір ғана денотат атауы болып табылатын, бірақ бөлек тұлғаланатын *киіз үй, ас үй, таңғы ас* сияқты лексикаланған сөз тіркестерінің, *айна қатесіз, қол алысу, көз көрген* сияқты бөлек жазылатын идиомалардың компоненттері *киіз, үй, ас, айна, қатесіз, қол, ал, көз, көр* түрінде жеке-жеке леммаға келтіріліп, әрқайсысының жиілігі бөлек саналғанын ескертеміз;

– тәуелдіктің үшінші жақ тұлғасындағы жинақтық мәндегі қос сөздердің компоненттері сын есім, модаль сөз болғанына қарамастан, осы формамен дайын күйде қолданылатындықтан (яғни тәуелдіктің үшінші жағы бұл сөздерде грамматикалық форма емес, сөз тудырушы жұрнақ қызметін алған), *ақ-қарасы, бары-жоғы, іші-сырты* т.б. зат есім ретінде таптастырылды;

– қазіргі сөйлеу нормасы тұрғысынан қолданысы бейәдеби лексикаға жататын *қатын* сөзі (жиілігі 422 рет) тарихи ескі мәтіндерде, көркем шығармаларда өзінің бұрынғы оң мағынасында қолданылғандықтан, сөзтізбеде сақталды.

Сын есімнің жиілік сөздігі сөзтізбесінің ерекшеліктері:

– сөзтізбеде сын есімнің шырай формалары жеке-жеке берілді. Күшейтпелі шырай көрсетілді: *ап-аман, өп-өтірік, тап-тарғыл*; салыстырмалы шырай сақталды: *шалғайлау, кішілеу, толығырақ, шағынырақ, шалғайырақ*. Ал аналитикалық тәсілмен жасалған асырмалы шырайдың *өте әдемі, аса биік* сияқты формаларындағы әр компонент жеке-жеке түбірге келтіріліп, *өте/үс, аса/үс, әдемі/сн, биік/сн* түрінде жиілігі бөлек саналды;

– *ған, -ген, -қан, -кен* есімше формасы арқылы етістік негізінен жасалған сындық мағынаны білдіретін сөздер сын есім ретінде таптастырылды: *антұрған, туған, арбизан-арбизан, қуарған, имиген, кербәққан, қатқан*, т.б.;

– *мың, миллион* сан есім негізіне жалғанған *-даған, -деген* формасымен келген *мыңдаған, миллиондаған* сөзі сандық мәнінен алшақтап, ‘қаптаған, көп’ деген сындық ұғымға ойысқанын мәтіндер базасындағы қолданысы көрсетті, сондықтан сөзтізбеде осы сөздер сан есім ретінде леммаланып кетті;

– есімдіктерге *-дай, -дей* теңеу мәнді қосымшаларының жалғануы арқылы жасалған сөздер лексикаланып, сапа ұғымын білдіретіндіктен, сын есімдерге таптастырылды: *осындай, андай-мұндай, сондай* т.б.;

– *қасқабас үйрек* сияқты күрделі атаулардағы анықтаушы компонент *қасқабас* сөзі бірге тұлғаланғандықтан олар сөзтізбеде сын есім ретінде таптастырылды;

– қоснүкте арқылы берілген сын есімдік мағынадағы тіркес сөздер бір сөз табымен берілді: *тыр: тыр жалаңаш; тыр: тырдай жалаңаш;*

– кейбір диалектілер сөзтізбеден орын алды: *самсақ (самсаған), пешеттеуші (мөр қоюшы)* т.б.;

– кейбір универбтенген фразеологизмдер бір ғана сипатты білдіретіндіктен, әрі мәтінде бірге таңбаланғандықтан, олардың орфографиялық сөздікте берілген бөлек таңбалану орфограммасы мәтіндерде кездеспегендіктен, сөзтізбеде де бір сөз ретінде жазылды: *қараборбай, қарасирақ;*

– жалпы есімдерге сын есім тудырушы жұрнақ жалғану арқылы жасалған советизмдер көрініс тапты: *маркшил, сталиндік* т.б.

– әдетте алдында қандай да бір анықталушының болуын міндетті түрде талап ететін, әрі онсыз жалаң қолданыла алмайтын етістік түбірге болымсыз *-сыз, -сіз* формалы сөздер сын есімдер категориясында берілді: *жеткісіз, бергісіз, көргісіз;*

– бірге жазылуға сұранып тұрған, мәтіндерде екі түрлі – бірде бөлек, бірде біріктіріліп жазылған, бірақ орфографиялық сөздікте бөлек тұлғаланатын сын есімдер қос нүкте арқылы берілді: *қара: қара кесек; сан: сан түрлі;*

– сөйлеу тіліндегі сингармофонетикалық варианттар сақталды: *сұп-сұркей (сұп-сұрқай);*

– бала тіліне тән қолданыстар кездесті: *әйей/сн;*

– *дағы, -дегі* функционалды жұрнақтары арқылы жасалғанымен (*сондағы, ағымдағы, алдындағы, жоғарыдағы, жылдағы, қатардағы, маңайдағы, ойдағы*) қатыстық сапа мағынасында тұрақты қолданысқа ие болған сөздер сын есім ретінде берілді;

– қос сөздер қатарында *-у* тұлғасымен келген, детерминанттылығы күшті, яғни анықталушының болуын міндетті түрде талап ететін, жалаң түрде қолданылмайтын кейбір күрделі сөздер сөзтізбеде сын есімдер категориясында берілді: *тарихи-ағарту/сн, спорт-сауықтыру/сн;*

– бұрынғы мәтіндерде сол кездің орфограммасымен жазылған сөздер ескіше тұрпатын сақтады: *алфавиттік/сн* (қазір: *әліпбилік*).

Сан есімдер жиілік сөздігі сөзтізбесінің ерекшеліктері:

– сөзтізбеде сан есімнің есептік, реттік, жинақтық, болжалдық, топтық разрядтарына жататын сөздер қамтылды: *бір, бірінші, ондаған, екі-үш, бір-екі, үшеу, төртеу, жетеу* т.б. Ал бөлшектік сан есімдер түбірге келтірілді: *екіден бір – екі* және *бір* болып сөзтізбеде жиілігі бөлек саналды;

– балалар фольклорында санамақ үлгілерінде жиі кездесетін күрделі сан есімдер мәтіндегідей берілді: *бір-екі-үш;*

– сөзтізбеде *-ақ* шылауы омоним ретінде *ақ* заты есімінен, сын есімінен ажыратылды және бөлек берілді, алайда *бір-ақ* тұрпатындағы сан есімде ол бөлініп

алынбады, себебі осы форманың сан есімдік қана емес, үстеулік те омонимі бар, осыны ажырату үшін, бұлар өздері жеке омонимдік қатар құрды: *бір-ақ/сан есім* (мысалы, *бір-ақ күнде жетті*, бір-ақ рет болды) және *бір-ақ/үстеу* (мысалы, *ауылға бір-ақ тартты, етекке бір-ақ түсті*).

Есімдіктің жиілік сөздігі сөзтізбесінің ерекшеліктері:

– сөзтізбеде есімдіктің барлық – жіктеу, сілтеу, сұрау, өздік, белгісіздік, болымсыздық, жалпылау түрі қамтылды: *мен, сен, сол, ана, өз, әркім, ешкім, барлық* т.б.;

– лексикалық мағынасы есімдік мәнінен алшақтаған жатыс септігінің формасы арқылы түрленген *онда, мұнда* т.б. есімдіктер үстеу ретінде берілді;

– *бірдеменің* сөйлеу тіліндегі бейнормалық *бірдеңке* қолданысы сирек (небары 2 рет) болса да, сол сияқты *ештеме* мен *ештеңе* жарыспалы варианттың бейнормалық *ештеңке* қолданысы (небары 2 рет) сөзтізбеде сақталды, бұл екіншілерінің (бейнормалық нұсқаларының) жиілігінің сиреу үдерісін аңғарту үшін берілді;

– *сонымен* есімдігі осы формада берілді, леммаға келтірілмеді, себебі осы күйінде қолданылып: *сонымен қатар, сонымен бірге, сонымен қоса* – күрделі шылау жасайды, *мұны, ондағы, содан, бұдан* есімдіктері де осы себепті *бұл, ол, сол* түріндегі түбірге келтірілмеді;

– сөзтізбеде *бұл* есімдігінің *бұ* деген сөйлеу тіліндегі редукцияланған нұсқасы берілді, себебі оның мәтіндердегі жиілігі аса жоғары – 2240 рет;

– *әрқайсы, әрқайсысы* екі түрлі нұсқада берілді, алайда орфографиялық ереже бойынша сөздікте бекітілген норма *әрқайсысы* (жиілігі 47) болуы тиіс. Бұған қарамастан мәтін базасында орфограммасы қате болып саналатын *әрқайсы* (жиілігі 804) тұлғасы ерекше жоғары жиіліктен көрінді. Сондықтан қазіргі орфографиялық нормаға қайшы *әрқайсы* нұсқасы сөзтізбеден орын алды.

– сөзтізбеде өздік есімдігінің тәуелдіктің үшінші жағындағы *өзі* (жиілігі 8848) және түбір *өз* (жиілігі 29594) формасы да берілді. Нормаға сәйкес *өзі* сөзі *өз* түрінде леммаға келтірілуі керек еді, алайда ол сөйлеуде дайын, сол күйі қолданылады, мәтіндегі контекстерден тәуелдік формасы оған сөзтудырушы тұрпат бергенін аңғару оңай: мысалы, *осы өзі дұрыс па, басшының өзі келді, мұның өзі*; сонымен қатар, *өзі* есімдігінің түбір тұлғадағы *өз* есімдігінен жиі қолданылады.

– есімдіктердің орфографиялық нормадағы жарыспалы варианттары сөзтізбеде жеке-жеке жиілікте көрсетілді: *бәлен* және *пәлен*;

– *неге* сұрағының сөйлеу тіліндегі *неғып* нұсқасы есімдік ретінде көрсетілді.

Етістіктің жиілік сөздігі сөзтізбесінің ерекшеліктері:

– сөзтізбеде етістіктің барлық семантикалық түрі қамтылды;

– бірен-сараң немесе небары бір ғана контексте етістікпен қолданылатын қимыл атауын білдіретін идеомалық тіркес құрылымдағы бірінші компонент қос нүктемен көрсетіліп, бұлар сөзтізбеде етістік ретінде берілді: *жүзе: жүзеге ас, қолды-аяқ: қолды-аяққа тұрма, сеспей: сеспей қат, сағы: сағы сыну*;

– *өмірем: өмірем қап* пен *өрем: өрем қап* – екі нұсқа да мәтіндегі қолданысына қарай берілді, алайда соңғысының жиілі (небары 2) аз;

– қазақ тіліндегі *отыр, тұр, жүр, жатыр* 4 қалып етістігі құрамындағы түбірі *жат* болып келетін соңғысының осы шақтағы *жатыр* формасы леммаға келтірілмеді, сондықтан сөзтізбеде оның *жат* түбірінің де, *жатыр* осы шақ формасының да жиілігі бөлек-бөлек есептелді. Себебі мәтінде ол *жатыр* зат есіммен омонимдес болып келеді. Сондықтан *жатыр/зт* мен *жатыр/ет* контекстегі жиіліктері бөлек саналды;

– *алып бар, солай ет, олай еттің* сөйлеу тіліндегі редуцияланған *анар, сөйт, өйт* формалары не себепті етістік ретінде сөзтізбеде берілді;

– етістіктің *-ғыр, -қыр, -гір, -кір* тұлғалы қарғыс формалары сөйлеуде дайын түрінде қолданылатындықтан, леммаға келтірілмеді, олардың жиілігі түбірінен бөлек есептеліп берілді: *атқыр, ұрғыр, жамандатқыр, құрғыр, қиылғыр, қағынғыр, көгергір* т.б.;

– кейбір сөйлеу тіліндегі етістіктер мен етістіктің нұсқалары берілді: *кедейле* (кедей болу), *құса* (ұқсау);

– көсемшенің *-й*, әсіресе *-п* формасымен дайын күйінде сөйлеуде қолданылатын кейбір сөздер қимылдың сын-сипатын білдіргенімен, үстеу категориясына жатпайтындықтан, сөзтізбеде етістік ретінде таптастырылып, түбірге (леммаға) келтірілмеді: *айтып-айтпай, асығып-үсігіп, салдыратып-күлдіретіп, тұрып-жығылып, алаулатып-жалаулатып, аптығып-үптігіп, бырысып-тырысып, кіріп-шығып, өкіріп-бақырып, сылқылдап-былқылдап, сыпырып-сиырып, таласып-тармасып* т.б. Бұлардың дені экспрессиялық бояулы сөздер екені байқалды;

– мәтінде тек *-у* тұлғасында дайын күйде қолданылатын, онсыз өз мәнінен, тұрпаттық тұтастығынан айырылып қалатын кейбір етістіктер түбірге келтірілмеді: *бақылау-қадағалау* (*бақылау-қадағала* бола алмайды), *бақылау-реттеу* (*бақылау-ретте* бола алмайды), *бақылау-тексеру* (*бақылау-тексер* бола алмайды), *ақыру-зекіру* (*ақыру-зекір* бола алмайды). Бұлардың дені терминдік мәні бар қос сөз формасында келетіні байқалды.

Үстеудің жиілік сөздігі сөзтізбесінің ерекшеліктері:

– сөзтізбеде үстеулердің барлық семантикалық түрі қамтылды;

– есімдік негізге адъективтік мән беретін сөзжасамдық қосымшалар жалғану арқылы жасалған сөздер үстеу ретінде таптастырылды: *солай, онда, мынадай, өзімен-өзі, өзіне-өзі*;

– алғашқы компоненті жалаң түрде қолданылмайтын, идиомаланған үстеулік мағынадағы тіркес қос нүкте арқылы таңбаланды: *жуыр: жуыр арада*;

– *ған (-ген, -қан, -кен)* есімше формалы сөздер үстеуленгендіктен, үстеудің сөзтізбесіне алынды: *мыңғырған, аумаған*;

– *де* етістігінің есімше формасына теңеулік *-дай, -дей* мәнді функционалды жұрнағынан жасалған *дегендей* сөзі үстеу ретінде көрсетілді;

– кейбір сөйлеу тілі мен диалектілік элементтер берілді: *сонсын – сонан соңның* сөйлеу формасы, *беталбаты* (Шығыс Қазақстан өңіріне тән жергілікті сөз);

– екі сыңары да көсемшенің *-й* формасымен жасалған етістік негізді дайын күйінде қолданылатын қос сөздер, қимылдың сипатын білдіретіндіктен, үстеу

ретінде таптастырылды: *ауырмай-сырқамай, арымай-талмай, аспай-саспай, ерінбей-жалықпай, бұзбай-шашпай, жатпай-тұрмай, жылай-жылай* т.б.;

– *ырақ, -ірек* формасындағы үстеуленген сөздер берілді: *жайырақ, нақтырақ, оқшауырақ, соңырақ, шатшаңырақ, арырақ, бұрынырақ, жоғарырақ, тезірек, әрірек, берірек, дәлірек, ертерек, кейінірек, кешірек* т.б.

Шылаулардың жиілік сөздігі сөзтізбесінің ерекшеліктері

– сөзтізбеде шылаулардың барлық түрі қамтылды;

– *-ақ, -ау, -ай* шылау сөздері негізгі сөзбен дефис арқылы ерекшеленіп, бірақ бірге тұлғаланатынына қарамастан, сөзтізбеге негізгі сөзінен бөлек, жеке шығарылды;

– *ма, ме, ба, бе, па, пе* және *ша, ше* сұраулық шылауларының фонетикалық варианттары сөзтізбеде бөлек-бөлек, әрқайсысының жиілігі жеке-жеке берілді;

– *қана, гана*, демеуліктері, *көрі, гөрі, дейін, шейін* септеуліктері, *мен, бен, пен, менен, пенен* жалғаулықтарының да фонетикалық варианттары бөлек-бөлек алынып, сөзтізбеде олардың әрқайсысының жиілігі жеке көрсетілді.

Одағайлардың жиілік сөздігі сөзтізбесінің ерекшеліктері

– сөзтізбеде одағайлардың барлық семантикалық түрі қамтылды;

– бейканондық лексикалық бірліктер ретінде қаралғандықтан, жалпы сөзтізбеде одағайлардың орфограммаларына қатаң талап қойылмады, көбі соңғы орфографиялық сөздікке сәйкес түзетілгенімен, біразы мәтіндегі нұсқамен қалдырылды, мысалы, *уахалейкүм, уағалейкүмассалам; уа, уай, уау; уәллаһи, білләһи; уф, уф-уф, ойбай, үйбай, ойбуй, ойбу; шу, шүү; ах, аһ; айда, әйда; әуп, әуіп*. Бұлардың бірінің орфограммасы сөздікте бекіген (*уағалейкүмассалам, ойбай, шүү, аһ* т.б.), екіншілері мәтіндегі таңбалануы бойынша (*уахалейкүм, үйбай* т.б.) сөзтізбеге түзетілмей алынды;

– бірнеше рет қайталанған нұсқалар да сөйлеу бірліктері ретінде берілді: *ай, ай-ай, ай-ай-ай*;

– бұл сөз табының сөзтізбесінде *-ай, -ау, -оу* демеулік шылаулары негізгі сөздерінен ажыратылмады: *шіркін-ай, шіркін-оу, япыр-ай, япыр-оу, япырым-ай, япырым-оу, апыр-ай, апыр-ау, апырым-ай, апырым-оу*.

Модаль сөздер жиілік сөздігі сөзтізбесінің ерекшеліктері

– бұл сөздер сөз таптарының құрамына жатпайды, алайда олардың мәтіндегі жиілігі жоғары екенін ескеріп, бөлек алу ұйғарылды;

– жалпы сөзтізбеде *-ақ* шылауы бөлек жиілікпен берілді, алайда *сондай-ақ* модаль сөзінде ол негізгі сөзден ажыратылмады, себебі *сондай* сөзі бұл шылаусыз есімдік негізді сын есім ретінде таптастырылды;

– болымсыздық мәндегі *беймүмкін* сөзі модаль сөз ретінде таптастырылды.

Көмекші есім жиілік сөздігі сөзтізбесінің ерекшеліктері

– бұл сөздер сөз таптарының құрамына жатпайды, алайда олардың мәтіндегі жиілігі өте жоғары екенін ескеріп, зат есімге таптастырмай, бөлек алу ұйғарылды;

– сөзтізбеде *алды, арты, қасы, жаны, ортасы, арасы, асты, үсті* т.б. мәтінде кездескен көмекші есімдер қамтылды.

6) МАҒЫНАСЫ АЖЫРАТЫЛҒАН ОМОНИМДЕРДІҢ ЖИІЛІГІ. Морфологиялық талдағыштан өткен, омонимі ажыратылған мәтіндерден жиілік сөздік алу әлдеқайда жеңіл. Олай болмаған жағдайда, омоним сөздерге келгенде автомат оларды танымайды да, әр сөз табына жататын омоним сөздердің қайсысының жиі кездесетінін нақты көрсете алмайды. (Мысалы, *ақ* сөзі әрі сын есім, әрі етістік). Омоним ажыратылған корпустары жасалған тілдер көп емес, ал болған күнде мәтін көлемі шағын болып келеді. Себебі барлық омонимдерді автоматты түрде ажырататын компьютерлік бағдарлама ешбір тілде жасалмаған деуге болады. Біздің тәжірибемізде де, морфологиялық талдағыш таныған аздағаны болмаса, омонимдер негізінен құрастырушылардың қол күшімен ажыратылды.

Сөздіктің бұл түрінде жоба барысында мәтіннен алынып, ажыратылған қазақ омонимдерінің жиілігі берілді. Олардың дені түркілік түбірден жасалған сөздер екені туралы жоғарыда айтып кеттік. Омонимдердің қатарында біреуінің ерекше жоғары, ал екіншісі төмен жиіліктен көрінетіні туралы құбылыс та аңғарылды. Сонымен бірге омонимдерді стильдер бойынша үлестірілім әліпбилі-жиілікте берудегі мақсат, олардың әрқайсысының түрлі функционалдық стильдердегі «болмысын» таныту болды. Мәселен, *қара* сөзінің сын есімдік омонимі ресми-іскери стильде белсенді емес, ал көркем әдебиет стилінде жоғары жиілік мәндерінен көрінді. Бұл сөздік оқыту мәтіндерін құрастырушылар үшін ғана емес, тіл зерттеушілері үшін де қызық мәліметтер бере алады деп ойлаймыз.

7) СӨЗФОРМАЛАРДЫҢ ЖИІЛІГІ¹⁷. «Жалпы білім берудегі қазақ тілінің жиілік сөздігінің» бір түрі ретіндегі «Сөзформалардың жиілік сөздігі» – қазақ тіл біліміндегі алғашқы тәжірибе, үлкен нәтиже. Әдетте сөзформалар сөздігі тілдегі сөздердің түрленім формасын тізімдеуді мақсат етеді. Нәтижесінде бір зат есім немесе бір етістік сөз қанша формамен түрлене алады деген статистикалық мәлімет алынады. Алайда ұсынылып отырған сөзформалардың жиілік сөздігінде сөздердің барлық түрленім парадигмасы қамтылмауы мүмкін, өйткені сөздік база-сына алынған мәтіндерде кездескен сөзформалар ғана тіркелді.

Жалпы сөзформалардың жиілік сөздігі тіл білімінің дәстүрлі зерттеулері үшін ғана емес, жаңа бағыттағы ізденістер, әсіресе ұлттық корпустар құрастыру, оның ішінде морфологиялық анализатор жұмысы үшін де құнды мәліметтер бере алады. Ұсынылып отырған сөзформалар сөздігі қазақ тілінің 5 стилі қамтылған мәтіндерден алынған теңгерімді, жан-жақты мәтін базасында әзірленгендігімен құнды.

Жиілік сөздік жасаудың бір жолы – мәтіндерден сөзформалардың жиілік сөздігін алу арқылы қолдап леммаға келтіру. «Жалпы білім берудегі қазақ тілінің жиілік сөздігі» осы әдіс арқылы жасалды. Ал осы еңбекке негіз болған «Сөзформалардың жиілік сөздігі» де тіл білімі ғана емес, оқыту, автоматты аударма жасау сияқты іргелі істерге негіз болатын маңызды еңбек деуге болады.

Сөзформалар жиілік сөздігі сөзтізбесінің, яғни бірінші сөзтізбенің **негізгі үш ерекшелігі** бар:

¹⁷ Бұл тараушаны жазуға Сөздіктің жауапты редакторы, филология ғылымдарының докторы *Айман Жаңабекова* ерекше үлес қосты.

а) сөзформалар түбірге келтірілген жоқ (леммаланбады);

ә) сөз табы қойылмады;

б) негізсөз сөзтізбедегідей әрбір стиль бойынша таратылған (распределительный) сандық мәліметтерді көрсетпейді, мұнда сөзформалар жиілігі мәтінді қамту пайызы бойынша жалпы жиілікпен ғана берілді.

Сөзформалардың жиілік сөздігін жасауда сөзформалар түбірге келтірілмеген, сөздер түрленген формада беріліп, сөз табы қойылмайтындықтан, оның сөзтізбесін өңдеу, рәсімдеу жұмысы негізінен **орфографиясы мен лексикалық құрамына** қатысты жүргізілді.

Сөзформалар сөзтізбесі аса көлемді болғандықтан, оны өңдеудің негізгі жұмыстарын жеке көрсету керек:

а) *орфографиялық норма бойынша техникалық өңдеу жұмысы жүргізілді:* сөздер мен сөзформалардың жиілігі көлемді мәтіндерден алынғандықтан, сөздердің жазылуында, әсіресе мәтіннің сканерлену кезіндегі техниканың қазақ қаріптерін танымаудан туған қателер көп болды. Сөзформалардың жиілік сөздігін рәсімдеу барысында жоғарыда аталған қателер бойынша өңдеу жұмыстары жүргізілді. Бұл жиілік сөздіктің екі сөзтізбесіне (бірінші және екінші) де тән техникалық өңдеу процесі болып табылады;

ә) *мәтінде нормасы сақталмаған қолданыстар (қос сөз, біріккен сөз, тіркес т.б.) әдеби нормаға сәйкестендірілді:* сөздік құрам үнемі өзгеріп отыратындықтан, бұрынғы орфографиялық нормамен берілген мәтіндерде кейбір сөз тіркестері қазіргі тілімізде бірге тұлғаланып, дербес лексикалық бірлікке айналды. Мұндай сөздер сөзтізбеде қазіргі орфограммасы бойынша бірге жазылды. Сол сияқты мәтінде қос сөз ретінде **қате** берілген кейбір сөздер тіркес түріндегі дұрыс нұсқаға ауыстырылып, оның компоненттерінің жиілігі жеке-жеке есептелді: *кез-келген* сөзі орфографиялық сөздікте сызықшасыз бөлек жазылады. Сонымен қатар кейбір сөздердің контексте сызықша арқылы жазылуы көп кездеседі. Мысалы, *айта-айта, жаза-жаза, асығып-сүрініп, айғайлап-ұрысып, барып-келіп* т.б. Ал осындай сөздік құрамда лексикалық бірлік ретінде танылмайтын сызықша арқылы берілген мұнай сөздер де «Сөзформалардың жиілік сөздігінде» екіге бөлініп, жеке сөз ретінде жазылды. Қос сөздер сыңарлары кейде орын ауыстырылып қолданылатыны байқалды. Мысалы; *кіші-үлкен, қара-ағын* т.б. Мұндайлар сөзтізбеде нормаға сәйкестендіріліп: *үлкен-кіші, ақ-қарасы* т.б. түрінде берілді

б) *кейбір сөйлеу варианттары әдеби нормаға сәйкестендірілді:* (*алакеуім – алагеуім, батшағар – патшағар, жекжат – жегжат, ада-күде – ада-гүде, немқұрайлы – немқұрайды* т.б.);

в) *орфографиялық сөздікте бөлек жазылған біртұтас мағынаны білдіретін кейбір сөздер, соған қарамастан, соңғы үлкен түсіндірме сөздіктегі¹⁸ тұлғалануы бойынша қос нүктемен берілді:* алғы: алғы сөз; үй: үй іші; ауа: ауа райы, әр: әр алуан, дер: дер кезінде, елді: елді мекен, жек: жек көру, жер: жер шары, жүзеге: жүзеге асыру, кез: кез келген, өмір: өмір бақи, сан: сан түрлі, тұла: тұла бойы т.б.;

г) *одағай сөздер құрамындағы -ай, -ау шылаулары негізгі сөзден ажыратылмады, ал басқа сөздерде -ақ, -ау, -мыс, -ай шылаулары бөлініп, сөзтізбеге*

¹⁸ Қазақ сөздігі. Бас редакторы. Н.Уәли, Алматы: «Дәуір», 2013. – 1488 б.

жеке шығарылды: мысалы, -ай: *балам-ай, жаным-ай*; ак: *бәлсінбей-ақ, бәріміз-ақ, бекер-ақ, бергендей-ақ, бес-ақ, болар-ақ, бөлмей-ақ, бүгін-ақ*; -ау: *бүлдірді-ау, білген-ау, дейміз-ау, достар-ау, дұрыс-ау, едім-ау*; ай: *екен-ай, елім-ай, жалғызым-ай, жануар-ай, жаратқан-ай, жатқанын-ай*; мыс: *жаттыпты-мыс, жеткізген-мыс, келіпті-мыс* т.б. сөздер құрамынан бөлініп алынды. Мұндай шылауларды сөз құрамынан бөліп алуымыздың өзіндік себептері бар. Өйткені мұндай шылаулар, біріншіден, сөздердің түрлену парадигмасына жатпайды, екіншіден, кез келген сөзге сөйлеу мақсатына қарай қолдана беруге болады, үшіншіден, жалпы шылаулар көмекші сөздер болса да, дербес лексикалық бірлік ретінде жеке сөз табын құрайды.

ә) бейқанондық лексикаға жататын көңіл күй одағайлары мәтіндерде адамның әртүрлі эмоциясын бейнелеу үшін алуан түрлі дыбыстық нұсқалармен берілді. Мысалы, *Ойбуууй, ойби, ойбааай, ойбаяй, ойбайоу* немесе *оxxx, оxx* т.б. Алайда сөзтізбеде олардың көбі өңдеу кезінде қысқартылып, мүмкіндігінше нормативтік сөздіктерде бекітілген түрге түсірілді.

д) ауызша формада редуцияланатын кейбір сөздердің сөйлеу тіліндегі варианттары қалдырылды: мысалы: *алмайм, көрем, жазам* т.б.

Тілімізде кейбір сөз тіркестері сөйлеу барысында дыбыстық құрамы кірігіп, біртұтас лексикалық бірлікке айналады. Сөз тіркестерінің универбтену процесінің көрінісі ретінде нормаға түспесе де, осындай бірлі-жарым сөйлеу тіліне тән қолданыстар берілді. Мысалы, *баратқанда, баратыр*. Мұндағы мақсат – сөйлеу тілі элементтерінің жазба мәтінде таңбалану мәдениетіне арналған ізденістер үшін құнды болып табылатын мәліметтерді сақтап қалу.

Сөзформалардың лексикалық құрамы төмендегідей рәсімделді:

а) *онимдер (мәтінде көбінесе кісі аттары мен топонимдер жиі кездесті) алынып тасталды*. Онимдер арнайы файлда бөлек жинақталды. Ал топонимдерге *-лық/лік* жұрнақтары жалғанған, зат есім, қатыстық сын есім сөздер сөздікте берілді: *алматылық, астаналық, астрахандық, атбасарлық, атыраулық* т.б.;

ә) *қысқарған сөздер (бас әріппен жазылатын буында бөлінбейтіндері) сөздіктен бөлек алынып, жеке файлда жинақталды*. Ал сөздің алдында префикс сияқты жалғанатындары сөздікте берілді. Мысалы: агро-: *агроөндіріс, агроөнеркәсіп, агроодақ, агрохимия*; авто-: *автобекет, автобизнес, автобөлшек, автокөлік*; авиа-: *авиакасса, авиабилет*, инфра-: *инфрақұрылым, инфрадыбыс* т.б.;

б) *тіліміздегі баламасы кең қолданысқа енген орыс сөздері сөзтізбеге енбеді, бөлек жиналды, ал термин болып қалыптасқандары берілді*, алайда белгілі бір кезеңдегі тіл дамуының үдерісін көрсететіндіктен, *алфавит, артист, аэропорт* т.б. сияқты жарыспалы орыс варианттары, баламасы (*әліпби, әртіс, әуежай*) бола тұра, қалдырылды. Орыс сөздерінің игерілгендері сөздікте берілді. Мысалы: *кәстөм, мәшек, кәмпит, кәртішке, сөмке, кінешке, кәстірөл* т.б.;

в) *мәтіндер әдеби тілде жазылуына қарай таңдалып алынса да, біршама жергілікті сөздер кездесті*. Сөзформалар сөзтізбесінде жергілікті сөздер, жаппай болмаса да, мағынасы түсінікті, аймақтық сөздіктерде тіркелген біразы берілді. Мысалы, *албаты (беталды), алжұмыр (помидор), алқайды (мадақтау), алқаулас*

(арқалас), бозым (айыр өркеш түйе), бұлаңшы, турамыш (тамақ түрі), боқша (сөмке), әгүршік (қияр) т.б.¹⁹;

г) жиілік сөздікке алынған мәтіндерде көне сөздер мен көне формалар да ұшырасады. Бұлардың түсіндірме сөздіктерде мағыналары берілгендіктен, сөз-тізбеде сақталды. Мысалы, албыс (албасты), беркіді (берік болды), бұғра (лауазым аты) т.б.;

ғ) жиілік сөздік базасындағы мәтіндерде мемлекет, жер т.б. географиялық нысан атауларынан жасалған *-лық/лік* тұлғалы қатыстық сын есімдердің бірлі-жарымы екі түрлі нұсқада кездесетіндігі байқалды. Мысалы, *америкалық – американдық, италиялық – итальяндық*. Бұлардың біріншісі (*америкалық, италиялық*) қазақ тілінің сөзжасамдық нормасына сәйкес болса, екіншісі (*американдық, итальяндық*; орыс тіліндегі *-н-* сына дыбысы – интерфиксі қазақ сөзінде сақталмауы тиіс) **қ а т е** қолданыс. Алайда екеуі де сөзтізбеде өз жиіліктерімен берілді, мұндағы мақсат – қазіргі жазба мәтін мәдениетіне қатысты ортологиялық зерттеулер үшін құнды болып табылатын деректерді сақтау.

Сөздіктің түрленімі

1) Жалпы жиілік

1. Жалпы әліпбилі-жиілік сөздік
2. Жалпы жиілікті-әліпбилі сөздік
3. Жалпы керіәліпбилі-жиілік сөздік

2) Стильдер бойынша үлестірілім жиілік

4. Стильдер бойынша үлестірілім әліпбилі-жиілік сөздік
5. Стильдер бойынша үлестірілім жиілікті-әліпбилі сөздік
6. Стильдер бойынша үлестірілім керіәліпбилі-жиілік сөздік

3) Қазақ әдеби тілі функционалды стиль мәтіндерінің жиілігі

7. Ғылыми стильдің жиілікті-әліпбилі сөздігі
8. Ғылыми стильдің әліпбилі-жиілік сөздігі
9. Ғылыми стильдің керіәліпбилі-жиілік сөздігі
10. Ресми-іскери стильдің жиілікті-әліпбилі сөздігі
11. Ресми-іскери стильдің әліпбилі-жиілік сөздігі
12. Ресми-іскери стильдің керіәліпбилі-жиілік сөздігі
13. Көркем әдебиет стилінің жиілікті-әліпбилі сөздігі
14. Көркем әдебиет стилінің әліпбилі-жиілік сөздігі
15. Көркем әдебиет стилінің керіәліпбилі-жиілік сөздігі
16. Публицистикалық стильдің жиілікті-әліпбилі сөздігі
17. Публицистикалық стильдің әліпбилі-жиілік сөздігі
18. Публицистикалық стильдің керіәліпбилі-жиілік сөздігі

¹⁹ **Ескерте кететін жайт**, бірінші сөзтізбеде, яғни сөзформалар сөздігінде жергілікті қолданыстар мейлінше сақталды, ал екінші, яғни негізсөз сөзтізбесінде, жалпы жиілік сөздікте олардың біразы еңбеді, себебі білім беру мәтіндерінде әдеби тілдің лексикасына қарама-қарсы қойылатын мұндай бірліктер барынша аз болуы тиіс деген принцип сақталды.

19. Сөйлеу стилінің жиілікті-әліпбилі сөздігі
20. Сөйлеу стилінің әліпбилі-жиілік сөздігі
21. Сөйлеу стилінің керіәліпбилі-жиілік сөздігі

*4) Ортақ база (барлық стильдерден теңгерімді түрде құралған)
мәтіндерінің жиілігі*

22. Ортақ база мәтінінің жиілікті-әліпбилі сөздігі
23. Ортақ база мәтінінің әліпбилі-жиілік сөздігі
24. Ортақ база мәтінінің керіәліпбилі-жиілік сөздігі жасалды.

5) Сөз таптарының жиілігі

25. Зат есімнің жиілікті-әліпбилі сөздігі
26. Зат есімнің әліпбилі-жиілік сөздігі
27. Зат есімнің керіәліпбилі-жиілік сөздігі
28. Сын есімнің жиілікті-әліпбилі сөздігі
29. Сын есімнің әліпбилі-жиілік сөздігі
30. Сын есімнің керіәліпбилі-жиілік сөздігі
31. Сан есімнің жиілікті-әліпбилі сөздігі
32. Сан есімнің әліпбилі-жиілік сөздігі
33. Сан есімнің керіәліпбилі-жиілік сөздігі
34. Есімдіктің жиілікті-әліпбилі сөздігі
35. Есімдіктің әліпбилі-жиілік сөздігі
36. Есімдіктің керіәліпбилі-жиілік сөздігі
37. Етістіктің жиілікті-әліпбилі сөздігі
38. Етістіктің әліпбилі-жиілік сөздігі
39. Етістіктің керіәліпбилі-жиілік сөздігі
40. Үстеудің жиілікті-әліпбилі сөздігі
41. Үстеудің әліпбилі-жиілік сөздігі
42. Үстеудің керіәліпбилі-жиілік сөздігі
43. Шылаудың жиілікті-әліпбилі сөздігі
44. Шылаудың әліпбилі-жиілік сөздігі
45. Шылаудың керіәліпбилі-жиілік сөздігі
46. Еліктеуіш сөздердің жиілікті-әліпбилі сөздігі
47. Еліктеуіш сөздердің әліпбилі-жиілік сөздігі
48. Еліктеуіш сөздердің керіәліпбилі-жиілік сөздігі
49. Одағайдың жиілікті-әліпбилі сөздігі
50. Одағайдың әліпбилі-жиілік сөздігі
51. Одағайдың керіәліпбилі-жиілік сөздігі
52. Модаль сөздердің жиілікті-әліпбилі сөздігі
53. Модаль сөздердің әліпбилі-жиілік сөздігі
54. Модаль сөздердің керіәліпбилі-жиілік сөздігі
55. Көмекші есімдердің жиілікті-әліпбилі сөздігі
56. Көмекші есімдердің әліпбилі-жиілік сөздігі
57. Көмекші есімдердің керіәліпбилі-жиілік сөздігі

6) Мағынасы ажыратылған омонимдердің жиілігі

58. Мағынасы ажыратылған омонимдердің стильдер бойынша үлестірілім әліпбилі-жиілік сөздігі

7) Сөзформалардың жиілігі

59. Сөзформалардың жиілікті-әліпбилі сөздігі

60. Сөзформалардың әліпбилі-жиілік сөздігі

61. Сөзформалардың керіәліпбилі-жиілік сөздігі жасалды.

Бұлардың барлығының қағаз нұсқадағы көлемі ауқымды болғандықтан, тек «Жалпы жиілікті-әліпбилі сөздік» пен «Стильдер бойынша үлестірім әліпбилі-жиілік сөздік» кітап болып шығарылды, ал қалған 58 түрленімі осы кітапқа қоса берілген дискіге жазылып, электрондық нұсқада ұсынылып отыр.

8. СӨЗДІК СТАТИСТИКАСЫ

Негізсөз сөзтізбесінің статистикасы

Негізсөз сөзтізбесінде барлығы 36265 бірлік бар. Бұл тілтанымдық сараптамадан кейін біраз сөздермен қысқартылғаннан кейінгі соңғы құрам. О бастағы сөзтізбедегі сөздер саны бұдан әлдеқайда көп болған: онимдер – топонимдер, антропонимдер, сондай-ақ жиілігі 1 ғана болған мағынасы күнгірт, диалектілік, пәс стильдік қолданыстар, ауызекі сөйлеудің өзінде сирек кездесетін жаргондар мен варваризмдер, ұлттық тілде баламасы бар орыс сөздері алынып тасталды.

Сөз басындағы әріптерге қатысты сөз статистикасы. Төмендегі кестеде әрбір әріптен басталатын сөздердің саны берілген. Іздеуге оңай болу үшін 6-кестеде әріптер әліпби ретімен орналасса, 7-кестеде сөз санының кему тәртібімен берілді.

6-кесте.

Жалпы әліпбилі-жиілік сөздіктегі әрбір әріптен басталатын сөздердің рет саны мен мөлшері:

Әріп	Әліпбилі-жиілікті сөздік реестріндегі орны	Барлығы, сөз саны	Мәтінді қамту пайызы (%)
А	1 – 2976	2976	7,72928
Ә	2977 – 3670	693	2,22661
Б	3671 – 6299	2628	12,86481
В	6300 – 6429	129	0,0801
Г	6430 – 7055	625	0,28946
Ғ	7056 – 7225	169	1,40439
Д	7226 – 8738	1512	3,67772
Е	8739 – 9549	810	4,22051
Ж	9550 – 12037	2487	8,96265
З	12038 – 12406	368	0,8877

И	12407 – 13102	696	0,63203
Й	13103 – 13110	7	0,00255
К	13111 – 15825	2714	7,45765
Қ	15826 – 19331	3505	9,57542
Л	19332 – 19621	289	0,22798
М	19622 – 21429	1807	5,24786
Н	21430 – 21879	449	1,50542
О	21880 – 22669	789	4,65901
Ө	22670 – 23208	538	2,871923
П	23209 – 24245	1036	1,25142
Р	24246 – 24823	577	1,28002
С	24824 – 28040	3216	6,95548
У	31506 – 31661	155	0,23443
Ұ	31662 – 32156	494	1,48655
Ү	32157 – 32572	415	1,49494
Ф	32573 – 33011	438	0,30424
Х	33012 – 33289	277	0,68465
Һ	33290 – 33291	2	0,00072
Ц	33292 – 33385	93	0,02925
Ч	33386 – 33402	16	0,01211
Ш	33403 – 34951	1548	2,05353
Щ	34952	1	0,00008
Ы	34953 – 35314	361	0,24372
І	35315 – 35563	248	0,86351
Э	35564 – 36243	679	0,35590
Ю	36244 – 36246	3	0,05766
Я	36247 – 36265	18	0,01634

7-кесте.

**Жалпы әліпбилі-жиілік сөздіктегі әрбір әріптен басталатын
кему ретімен берілген сөз саны:**

Әріп	Әліпбилі-жиілікті сөздік реестріндегі орны	Барлығы, сөз саны	Мәтінді қамту пайызы (%)
Қ	15826 – 19331	3505	9,57542
Т	28041 – 31503	3462	9,47815
С	24824 – 28040	3216	6,95548
А	1 1– 2976	2976	7,72928
К	13111 – 15825	2714	7,45765
Б	3671 – 6299	2628	12,86481
Ж	9550 - 12037	2487	8,96265
М	19622 – 21429	1807	5,24786
Ш	33403 – 34951	1548	2,05353
Д	7226 – 8738	1512	3,67772

П	23209 – 24245	1036	1,25142
Е	8739 – 9549	810	4,22051
О	21880 – 22669	789	4,65901
И	12407 – 13102	696	0,63203
Ә	2977 – 3670	693	2,22661
Э	35564 – 36243	679	0,35590
Г	6430 – 7055	625	0,28946
Р	24246 – 24823	577	1,28002
Ө	22670 – 23208	538	2,871923
Ұ	31662 – 32156	494	1,48655
Н	21430 – 21879	449	1,50542
Ф	32573 – 33011	438	0,30424
Ү	32157 – 32572	415	1,49494
З	12038 – 12406	368	0,8877
Ы	34953 – 35314	361	0,24372
Л	19332 – 19621	289	0,22798
Х	33012 – 33289	277	0,68465
І	35315 – 35563	248	0,86351
Ғ	7056 – 7225	169	1,40439
У	31506 – 31661	155	0,23443
В	6300 – 6429	129	0,0801
Ц	33292 – 33385	93	0,02925
Я	36247 – 36265	18	0,01634
Ч	33386 – 33402	16	0,01211
Й	13103 – 13110	7	0,00255
Ю	36244 – 36246	3	0,05766
Һ	33290 – 33291	2	0,00072
Щ	34952	1	0,00008

Жоғарыдағы кестеде қазақ әріптерінің ішінде Қ-дан басталатын сөздің көп екені байқалады. Демек, Қ дыбысынан басталатын түбірлердің сөзжасамдық аясы кең екені аңғарылады. Сол сияқты И, К, Г, П, Э, Ф дыбысынан басталатын сөз санының көп болуы бұлардың айтарлықтай бөлігін шет тілінен енген терминдер құрайтынымен байланысты.

Кестедегі қызықты көрініс ретінде С және Б дыбыстарынан басталатын сөз саны мен мәтінді қамту пайызына қатысты мәліметтерді келтіруге болады. Мұнда Б дыбысынан басталатын сөз саны жағынан кестеде алдына Қ, Т, С, А, К әріптерін жіберіп, өзі алтыншы орында (2628 сөз) тұрса да, барлық әріп ішінде осы дыбыстан басталатын сөздердің мәтінді қамту пайызы ең жоғары көрсеткішті беріп тұр – 12,8648, демек, осы дыбыстан басталған сөздердің мәтін ішіндегі қолданылу жиіліктері өз алдында тұрған саны көп дыбыстарға қарағанда жоғары болған. Яғни осы дыбыстан басталатын сөздердің ішінде Қ-дан басталатын сөздерге қарағанда белсенді лексикалық қабаттағы сөздер көбірек болса керек. Ал С әрпінен басталатын сөздер саны жағынан Қ мен Т әрпінен басталатын сөздерден кейін үшінші орында тұр (3462 сөз), алайда оларды мәтінді қамту пайызы небары 6,95548, яғни өзінің маңындағы алғашқы орындарда тұрған әріптердің көрсеткішінен азырақ,

демек, осы сөзден басталатын сөздер көп болғанымен, олардың біразының белсенді лексикалық қабатқа кірмей тұрғаны анық.

Сөздердің жиілігіне қатысты статистикасы. Жалпы жиілікті-әліпбилі сөздік (негізсөздердің) сөзтізбесіндегі 36265-нан құралған барлық сөздің біріншісінен (бол етістігінен) бастап 1060-сөзі 7 миллион сөзқолданыстан тұратын мәтін базасында 108 133 реттен бастап 1000 ретке дейінгі жиілікте қолданылған. Яғни ең жоғары жиіліктен бастап 1000 реттік жиілікпен небары 1060 сөз ерекшеленген. Бұл осы мың алпыс сөздің белсенді лексикалық қабаттың бірліктері екенін аңғартады. Осылай сөзтізбедегі сөздерді жиілік деңгейі бойынша таптастырса, мынадай мәлімет көрінеді (8-кестені қараңыз).

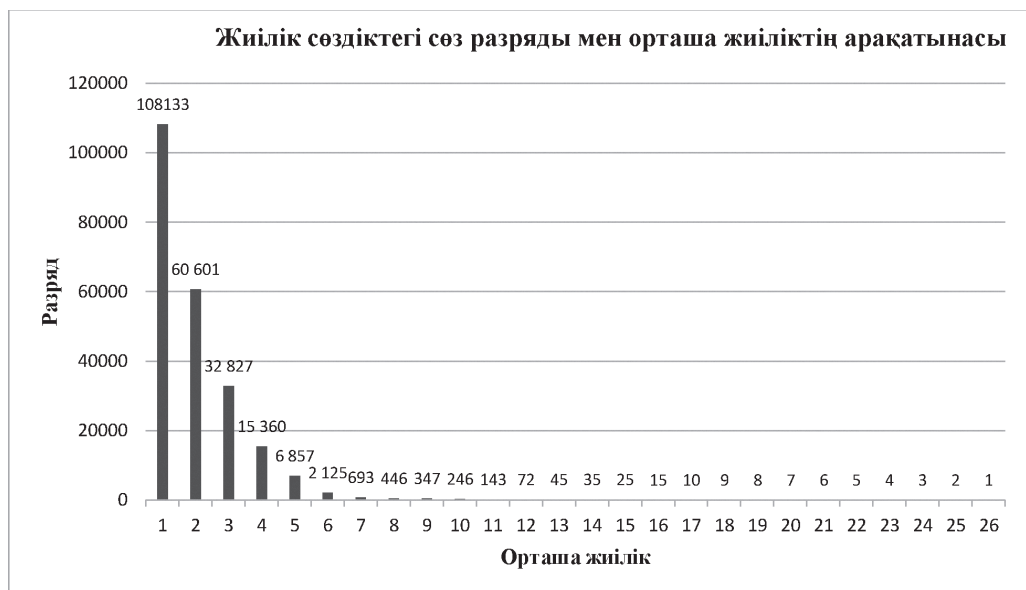
8-кесте.

Жалпы жиілікті-әліпбилі сөздік сөзтізбесіндегі сөздердің жиілігі жағынан таптастырылуы

Разряд №	Қолданыс саны, жиілігі	Жиілікті-әліпбилі сөздік реестріндегі орны	Барлығы сөз саны	Жиынтық жиілігі	Разрядтардың орташа жиілігі	Сөзтізбедегі сөздерді қамту пайызы	Мәтінді қамту пайызы
1	100 001-ден жоғары	1	1	108133	108133	0,0027	1,80137
2	50001-100000 рет	2-3	2	121202	60 601	0,00551	2,01908
3	25001-50000 рет	4-19	16	525239	32 827	0,04411	8,74986
4	10001-25000 рет	20-79	60	921621	15 360	0,16544	15,35312
5	5001-10000 рет	80-187	108	740507	6 857	0,29780	12,33597
6	1001-5000 рет	188-1058	871	1 851286	2 125	2,40176	30,84025
7	501-1000 рет	1059-1846	788	546104	693	2,17289	9,09744
8	401-500 рет	1847-2203	357	159109	446	0,98442	2,65057
9	301-400 рет	2204-2665	462	160442	347	1,27395	2,67278
10	201-300 рет	2666-3465	800	197065	246	2,20598	3,28287
11	101-200 рет	3466-5249	1784	254416	143	4,91934	4,23827
12	51-100 рет	5250-7592	2343	168118	72	6,46077	2,80064
13	41-50 рет	7593-8472	880	39863	45	2,42658	0,66408
14	31-40 рет	8473-9694	1222	43268	35	3,36964	0,72079
15	21-30 рет	9695-11596	1902	47720	25	5,24472	0,79505
16	11-20 рет	11597-15154	3558	52581	15	9,81111	0,87594

17	10 рет	15155-15741	587	5870	10	1,61864	0,09778
18	9 рет	15742-16403	662	5958	9	1,82545	0,09926
19	8 рет	16404-17148	745	5960	8	2,05432	0,09928
20	7 рет	17149-18068	920	6440	7	2,53688	0,10729
21	6 рет	18069-19111	1043	6258	6	2,87605	0,10425
22	5 рет	19112-20395	1284	6420	5	3,54060	0,10695
23	4 рет	20396-22063	1668	6672	4	4,59947	0,11114
24	3 рет	22063-24315	2252	6756	3	6,20984	0,11255
25	2 рет	24316-28183	3868	7736	2	10,66593	0,12887
26	1 рет	28184-36265	8082	8082	1	22,28595	0,13464

Бұл кесте жоғарыда сөз болған инфометриядағы басты заң – *Цифр заңының немесе Паретто принципінің*, яғни «сөз жиілігі оның разрядына кері пропорционал» деген тұжырымның ақиқат екенін растап тұр. Мұны төмендегі «Жиілік сөздіктегі сөз разряды мен орташа жиіліктің арақатынасы» атты сызбадан айқын көруге болады. Тағы бір қызығы, сөздікте небары 1 рет қолданылған сөз саны **8082** жалпы сөзтізбедегі сөз санының (36265) 4/1 бөлігін құрайды. Ал **1-10** реттік жиілікпен кездескен сөздердің барлық саны **21 111**, бұл жалпы сөзтізбенің жартысынан астамы. Демек, тілде белсенді лексикалық қордың бірліктері оның негізгі қорының жартысынан да аз болады.



Жоғарыда берілген статистикалық мәліметтерді бір топ сөздің мәтінді қамту пайызының көрсеткіші де толықтырып тұрғанын берілген кестеден байқауға болады. Қарапайым сөзбен айтқанда, 9-кестедегі мәліметтер осы жиілік сөздіктегі бірінші сөзден бастап 1822-сөзіне дейін білген тіл тұтынушысы қазақ мәтінінің шамамен 80 пайызын түсінеді дегенге саяды.

**Жалпы жиілікті-әліпбилі сөздік сөзтізбесіндегі бір топ сөздің
мәтінді қамту пайызы**

№	Жиілікті-әліпбилі сөздік реестріндегі сөздер	Мәтінді қамту пайызы %
1	1 – 13	10%
2	1 – 41	20%
3	1 – 92	30%
4	1 – 183	40%
5	1 – 333	50%
6	1 – 576	60%
7	1 – 994	70%
8	1 – 1822	80%
9	1 – 3854	90%
10	1 – 36265	100%

Сөз таптарына қатысты статистика. 10-кестеден әрбір сөз табы бойынша сөздіктің осы түріндегі жиілік сөздіктеріндегі сөзтізбедегі сөз санын, жалпы абсолютті жиілікті көруге болады.

**Сөз таптарының әліпбилі-жиілік сөздігі бойынша
(сөзтізбедегі сөз санының кему ретімен берілді)**

Р/с	Сөз табы	Сөзтізбедегі сөздер саны	Жалпы абсолюттік жиілік
1	Зат есім	17 339	2 537 662
2	Сын есім	9409	649 671
3	Етістік	7259	1 584 807
4	Үстеу	1258	216 251
5	Еліктеуіш сөздер	402	5 911
6	Сан есім	85	80 048
7	Есімдік	176	415 372
8	Одағай	175	7 437
9	Шылау	107	417 780
10	Модаль сөздер	47	78 655
11	Көмекші есімдер	9	9 234

Сөз таптарының ішінде сөзтізбедегі сөз саны жағынан ең көбі зат есім, сонымен бірге осы сөз табының абсолютті жиілігі де басқаларынан бірнеше есе артық. Бұл зат есімдердің мәтіндегі алатын үлес салмағының аса жоғары екенін аңғартады, демек, мектеп оқулықтары мәтіндерінде олардың басым болуына назар аудару керек. Сонымен бірге етістіктердің де сөз саны небары 7 259 болғанымен, олардың жиілігі мәтінде аса жоғары 1 584 807, абсолютті жиілік мәнінің жоғары болуы етістіктердің құрамында көмекшілік қызмет атқаратын грамматикалық

тұлғалардың көптігінен ақпар береді, демек жалпы білім беруге арналған тіл оқулықтарында етістіктің грамматикалық формаларына ерекше ден қою керектігі сөздік мәліметтерінен де дәйектеліп тұр.

Сөзтізбедегі сөз санының тым аздығына қарамастан, шылау, модаль сөз бен көмекші есімдердің абсолютті жиіліктерінің жоғары болуы жалпы тіл-тілдерге тән әмбебап құбылыстың қазақ тіліндегі көрінісі: кез келген мәтінде көмекші, яғни грамматикалық мағынасы басым сөздер ерекше жоғары жиіліктен көрінеді. Мәселен, орыс тіліндегі барлық жиілік сөздіктерде алдыңғы орынға *в* предлогі ие болады.

11-кесте.

Сөз таптарының мәтінді қамту пайызы

Р/с	Сөз табы	Сөзтізбедегі сөздер саны	Сөз таптарының абсолютті жиілігі	Сөз таптарының мәтінді қамту пайызы
1	Зат есім	17339	2537662	42,27444131
2	Етістік	7259	1584807	26,40100633
3	Сын есім	9409	649671	10,82274888
4	Шылау	107	417780	6,959719652
5	Есімдік	176	415372	6,919605226
6	Үстеу	1258	216251	3,602485362
7	Сан есім	85	80048	1,333504808
8	Модаль сөздер	47	78655	1,310299079
9	Көмекші есімдер	9	9234	0,153827496
10	Одағай	175	7437	0,123891606
11	Еліктеуіш сөздер	402	5911	0,098470254

Бұл кестеде сөз таптары мәтінді қамту пайызының көрсеткіші бойынша кему ретімен орналасты. 10-кестедегі, яғни сөз санының кему ретінен ауытқу байқалады: сөзтізбеде сөз саны жағынан сын есім зат есімнен кейін екінші орынды алады, бірақ оның мәтінді қамту пайызы етістіктен аз. Демек оқу мәтіндерінде етістіктердің мағынасын түсіну деңгейін қамтамасыз ету аса маңызды, себебі олар зат есімдермен қатар кез келген тілдегі ақпаратты түсіну үшін аса маңызды сөз табы болып табылады.

Бір қызығы, 7 миллион сөзқолданыстан тұратын мәтінде небәрі 107 шылау сөз қолданылған, алайда олар мәтінді қамту пайызы жағынан зат есім, етістік, сын есімнен кейінгі маңызды орында тұр. Мұндай мәлімет тіл үйренушілер мен тіл үйретушілер үшін маңызды болып табылады, себебі мәтіндегі сөздердің арасындағы грамматикалық, семантикалық күрделі қатынасты көрсететін бұл сөз табын игермейінше сол тілдегі ауызша, жазбаша сөзді түсіну толық жүзе аса алмайды.

Стильдерге қатысты статистика. Негізсөз сөзтізбесі бойынша жасалған әрбір функционалды стильдегі сөз таптарының арақатынасы туралы деректер оқу мәтіндерін құрастырушы педагогтер, оқулық авторлары үшін қажетті болып

табылады. Әсіресе қазақ тілі оқулығындағы мәтіндер стильдік әралуандықпен ерекшеленуі керек. Себебі тіл үйренуші неғұрлым көп стильдегі мәтіндерді тануды, талдауды және өзі құруды үйренсе, соғұрлым оның тілдік коммуникациядағы икемділігі, қоғамдық өмірдің әртүрлі саласында қарым-қатынас орната алу қабілеті арта түседі. Сондықтан стильдік тұрғыдан маркерленген оқу мәтіндерін таңдауда, құруда, стильдердің тілдік контентіне қатысты оқу тапсырмаларын құрастыруда осы тараушада көрсетілген жиіліктің мәліметтері құнды болмақ деп санаймыз.

12-кесте.

Әр стиль мәтіндеріндегі сөз таптарының үлес салмағы мен арақатынасы

Сөз таптары	Ғылыми стиль: сөзтізбедегі сөз саны 22 159		Ресми-іскери стиль: сөзтізбедегі сөз саны 7 488		Көркем әдебиет стилі: сөзтізбедегі сөз саны 12 289		Публицистикалық стиль: сөзтізбедегі сөз саны 15 079		Сөйлеу стилі: сөзтізбедегі сөз саны 9 370	
	Саны	Үлесі, %	Саны	Үлесі, %	Саны	Үлесі, %	Саны	Үлесі, %	Саны	Үлесі, %
Зат есім	11 324	51	3 749	50	5 105	41,5	7 210	48	4 335	46,2
Етістік	4 352	19,6	1 501	20	3 540	28,8	3 520	23,3	2 472	26,4
Сын есім	5 184	23,4	1 777	23,7	2 357	19,2	3 156	21	1 648	17,6
Сан есім	61	0,3	44	0,6	60	0,5	63	0,4	48	0,5
Есімдік	135	0,5	73	0,1	136	1,1	143	0,9	117	1,2
Үстеу	734	0,3	253	3,4	655	5,3	673	4,5	478	5
Шылау	89	0,4	62	0,8	89	0,7	87	0,6	73	0,8
Еліктеуіш	158	0,7	3	0,04	213	1,7	140	0,9	68	0,7
Одағай	77	0,3	4	0,05	92	0,75	72	0,5	89	0,9
Модаль сөз	31	0,1	14	0,2	32	0,26	41	0,3	33	0,35
Көмекші есім	9	0,04	7	0,09	7	0,06	9	0,06	7	0,07

Статистикалық мәліметтер барлық стиль мәтіндеріндегі сөз таптарының үлесі мен арақатынасына қатысты мынадай о р т а қ құбылыстарды аңғартты:

1) зат есімдер сөзтізбенің жартысына жуығын құрайды, яғни оқу мәтіндерінде зат есімдердің көп болуы шарт;

2) етістіктер мен сын есімдер кез келген стильдегі мәтінде зат есімдер санының жартысын құрайды:

– ғылыми стильде зат есім үлесі 51, етістік – 19,6, сын есім – 23,4;

– ресми-іскери стильде зат есім үлесі 50, етістік – 20, сын есім – 23,7;

– көркем әдебиет стилінде зат есім 41,5, етістік – 28,8, сын есім – 19,2;

– публицистикалық стильде зат есім 48, етістік – 23,3, сын есім – 21;

– сөйлеу стилінде зат есім үлесі 46,2, етістік – 26,4, сын есім – 17,6 пайызды құрайды;

3) кез келген стильдегі мәтіннің басым бөлігін зат есім, етістік пен сын есім құрайды:

- ғылыми стильде зат есім, етістік, сын есімнің жалпы үлесі 94%;
- ресми-іскери стильде зат есім, етістік, сын есімнің жалпы үлесі 93,7%;
- көркем әдебиет стилінде зат есім, етістік, сын есімнің жалпы үлесі 89,5%;
- публицистикалық стильде зат есім, етістік, сын есімнің жалпы үлесі 92,3%;
- сөйлеу стилінде зат есім, етістік, сын есімнің жалпы үлесі 90,2%.

Сонымен бірге оқу мәтіндерін құруда ескерілуі тиіс әр стильге тән сөз таптарына қатысты мынадай ерекшеліктер байқалды (кестеде бұл мәліметтер қою қаріппен берілді):

- 1) зат есімдердің үлесі **ғылыми стильде** басым, яғни ұғым атаулары ғылыми мәтіннің негізі болып табылады;
- 2) етістіктердің үлесі **көркем әдебиет** стилінде басым, демек көркем әдебиет мәтіндеріне динамикалық сипат тән;
- 3) бір қызығы, **ресми-іскери** стиль мәтіндерінде сын есімдердің үлесі үлкен болған, алайда олардың жиіліктері аса жоғары мәндерден көрінбеді, олар алуан түрлілігімен ерекшеленді, демек іскери қарым-қатынас мәтіндері үшін анықтаушылық қызметтегі сөздердің маңызы зор болмақ;
- 4) сан есімдер мен есімдіктердің де үлесі **ресми-іскери** стиль мәтіндерінде басым болғаны байқалды. Демек ресми сипаттағы мәтіндерге нақтылық әрі есім сөздердің орнына жұмсалатын сілтемелердің көптігі тән;
- 5) **көркем әдебиет** стиліндегі мәтіндер үстеулердің үлкен үлесімен ерекшеленді, бұл жоғарыдағы осы стильде етістіктердің үлесінің көптігі туралы фактіні растаушы, толықтырушы мәлімет десе де болады, себебі көркем шығармадағы етістіктер жалаң емес, көбіне өздерінің сын-сипатын білдіретін үстеумен тіркесе жүреді;
- 6) шылаулар мен көмекші есімдердің ең үлкен үлесі **ресми-іскери** және **сөйлеу стиліне** тән болды, сөздер арасындағы грамматикалық қатынас түрлерін білдіруге жұмсалатын бұл сөз табы жалпы барлық стильде де жоғары жиіліктен көрінді, алайда оның саны басқа сөз таптарынан әлдеқайда аз. Ал ресми-іскери стиль мен сөйлеу стилі мәтіндеріндегі оның үлесінің көптігі бұндағы коммуникацияның ерекшеліктеріне байланысты болса керек, ресми мәтінде семантика байлығынан гөрі, грамматикалық қатынас байлағының көптігін көрсетсе, сөйлеу стилінде көзбе-көз қарым-қатынас барысында сөз субъектісі, объектісін, предикатын қайталамай-ақ, олардың арақатынасын көмекші сөз таптарының көмегімен берсе де, мағынаға нұқсан келмейтінінің белгісі болып тұр;
- 7) еліктеуіш сөздердің үлесі **көркем әдебиет стиліндегі** мәтіндерге тән болып шықты. Көркем мәтін прагматикасына тән объективті ортадағы құбылыстарды шынайы, асқан деректілікпен сипаттауға қатысты бұл құбылысты оқыту мәтіндерін құрастырғанда есте ұстаған абзал;
- 8) одағайлар мен модаль сөздер **сөйлеу стиліндегі** мәтін құрамында көп кездесті, адам сезімінің алуан түрлі күйін білдіретін сөздердің бейресми қарым-қатынас мәтіндеріне тән болуы заңды да, осы ерекшелік оқу мәтіндерінде көрініс табуы тиіс.

Сөзформа сөзтізбесінің статистикасы

7 миллионнан астам сөзқолданыстық мәтіндік база көлемінен алынған 246 328 сөзтізбеден тұратын бұл сөздік үш түрлі құрылымда жасалған: 1) *Сөзформалардың әліпбилі-жиілік сөздігі*. 2) *Сөзформалардың жиілікті-әліпбилі сөздігі*. 3) *Сөзформалардың керіәліпбилі-жиілік сөздігі*.

«Сөзформалардың жиілік сөздігінің» **статистикалық ақпараты** өте көп, сондықтан негізгі ерекшеліктерді ғана сипаттап, қысқаша беру жөн саналды. Өйткені сөздік әрбір сөзформаның, қосымшаның т.б. жиілігі туралы арнайы ғылыми зерттеулер жүргізуге тұрарлық мәліметтерге аса бай.

«Сөзформалардың жиілікті-әліпбилі сөздігінің» сандық мәліметтері бойынша төмендегідей жалпылама **түйіндер** жасауға болады.

Негізсөздердің жиілік сөздігінің, яғни жиілікті-әліпбилі ретпен берілген екінші сөзтізбенің және сөзформалардың жиілік сөздігі, яғни бірінші сөзтізбенің арасында аздаған айырмашылық бар. Айталық, негізсөздердің (леммаланған) жиілік сөздігінде, яғни екінші сөзтізбеде ең жиі кездесетін сөз *бол* етістігі бірінші тұрса, сөзформалардың жиілік сөздігінде, яғни бірінші сөзтізбеде *мен* (жалғаулық шылау, түбір тұлғадағы жіктеу есімдігі) сөзі жиілігі бойынша бірінші тұр. *Бол* етістігінің сөзформалар сөздігінде бірінші орынға шықпауының өзіндік себебі бар. Негізсөздердің жиілік сөздігінде *бол* етістігінің барлық түрленісі түбірге келтіріліп, лемматизация жасалғанда сыйыстырылып, *болған, болды, бола, болмайды, болар, болмақ* т.б. формаларының барлығының жиілігі бір түбірге қосылады. Сондықтан ол жоғары жиілікті көрсетеді. Ал сөзформалардың жиілік сөздігінде *бол* етістігінің түрленім формаларының әрқайсысы бөлек-бөлек сөзтізбелік бірлік ретінде қаралатындықтан, жиілігі азаяды. Соның өзінде сөзформалар жиілік сөздігінде *болып* етістігі – 16, ал *болады* етістігі – 24, *болды* етістігі – 30, *болған* етістігі – 42, *болатын* етістігі – 111 орында тұр. Сонда *бол* етістігінің түрленім формаларының бесеуі алғашқы 100 сөздің ішінде тұр. Осы түрленім формалары қосылғанда түбір тұлғада негізсөздер сөздігінде жиілігі ең жоғары сөз ретінде бірінші орынға шығады.

Сөзформалардың жиілік сөздігінде жалғаулық шылаулардың жиілігі жоғары. *Мен* сөзінен кейінгі жиілікте *және* жалғаулық шылауы, үшінші орында да жалғаулық шылауы тұр. Бұдан мынадай заңдылықты байқауға болады. Сөздер түрленіп келгенде жиіліктері сол түрленімдерге бөлініп кетеді, ал жалғаулық шылаулар түрленбейтіндіктен, жоғары жиілікті көрсетеді, екінші жағынан, жалғаулық шылаулардың жиілігінің жоғары болуы олардың контекстегі сөзбайланыстырушылық қызметімен де байланысты болса керек. Сол сияқты мынадай шылаулар да жиілігі жоғары сөздер қатарында тұр: *туралы, не, үшін, ал, бойынша, ғана*.

Сөзформалар жиілік сөздігінде жиілігі бойынша 4 орында *деп* етістігі тұр. *Де* етістігінің *деген* формасының да жиілігі жоғары, 14 орында, *болып* етістігі – 16 орында. Бұл фактілер де *деп, боп* көмекші етістіктерінің сөзбайланыстырушылық қызметінен туындайды.

Ол, бұл, осы, сол, өз сияқты есімдіктер де жиі қолданыстағы сөздер қатарынан орын алған. Негізсөздердің жиілік сөздігінде де шылаулар мен есімдіктердің жиілігі жоғары.

Зат есімдер мен сын есімдер, сан есімдер, сол сияқты есімдіктердің, үстеулердің түрленген сөзфорларынан гөрі, түбір күйдегі тұлғасы жиі кездеседі. Мысалы, *орындық* сөзі түбір формада 147 рет кездесе, *орындыққа*– 93 рет, *орындықта* – 34 рет, *орындықтан* – 11 рет, *орындықтар* – 29 рет, *орындықтарға* – 6 рет, *орындықтарда* – 7 рет, *орындықтарды* – 12 рет, *орындықты* – 38 рет т.б. кездескен. Мұндай сандық көрсеткіш басқа да осындай сөздерге тән.

«Сөзформалардың жиілік сөздігінде» тағы бір байқалатын ерекшелік – қос сөздерден гөрі сол қос сөздерді құрайтын сөздер сыңарларының жеке тұрғанда жиі кездесетіндігі. Мысалы, *ой* – 26646, *ой-пікір* – 38, *ой-арман* – 4, *ой-қиял* – 1, *ой-өріс*– 5 т.б.

«Сөзформалар жиілік сөздігінде» берілген сандық мәліметтерден атауыш сөздердің түбір тұлғасынан кейінгі жиі қолданыстағы формалары септік жалғаулы сөздер, ал тәуелдік жалғаулары көбінесе үшінші жақта жиі кездеседі. Көптік жалғаулы сөздер де түбір тұлғамен кейде барабар болса, кейде одан сиректеу.

-лық/лік тұлғалы қатыстық сын есімдердің де басқа сөзформаларға қарағанда жиілігі әжептеуір. Мысалы, *оптимист* деген сөздің жиілігі – 6 болса, *оптимистік* – 25, *оптика* – 14, *оптикалық* – 38 т.б. Бұдан термин сөздердің қазақ тілі қосымшалары жалғанып, игерілген формалары (көпшілік жағдайда) түбір тұлғадан гөрі жиі кездесетіндігін байқауға болады.

Етістік сөзформалардың сандық мәліметтерінен *-ын/ін/н* көсемше тұлғалы сөздердің, *-ған/ген* есімше формалы сөздердің, жіктік жалғаудың 3 жақ тұлғасының (*-ды/ді*) жиі кездесетінін байқаймыз. Етістіктің қимыл атау тұлғасының көбінесе терминденіп, процесс атауына айналғандарының жиілік көрсеткіштері жоғары. Мысалы, *автоматтандыру, азарту, ақпараттандыру* т.б.

9. ҚАЗАҚ ТІЛІНІҢ ҰЛТТЫҚ КОРПУСЫ – БОЛАШАҚ ЖИІЛІК СӨЗДІКТЕРДІҢ БАЗАСЫ

Сөздердің жиілігі тілдің өмір сүріп отырған қазіргі немесе тарихи кезеңіне қарай алына алады. Сонымен бірге сөздердің жиілігіне қатысты статистикалық ақпарат мейлінше дәйекті болу үшін жиілік алынатын мәтіндер базасы неғұрлым ауқымды әрі теңгерімді, сонымен бірге тілдің қоғамдық өмірдегі барлық саласына қатысты болуы тиіс. Мәтіндер базасының осындай жоғары талаптарына жауап бере алатын тілтанымдық өнім – тіл-тілдің ұлттық корпустары. Қазіргі әлемдік лингвистикада осы жағынан алға шыққан Ұлыбритания (British National Corpus), Финляндия (Финляндияның тілдік банкі), Ресей (Орыс тілінің ұлттық корпусы) т.б. елдер бүгінгі күні кез келген тілтанымдық өнімді корпустарының базасында жасауды дәстүрге айналдырып отыр. Мәселен, орыс тілінің соңғы жиілік сөздігі ұлттық корпус базасындағы 92 млн сөзқолданыстан тұратын мәтін негізінде жасалды. Корпус жиілік сөздік үшін тек мәтін базасын ғана ұсынып қоймай, сонымен бірге базаға алынған мәтіндерге қатысты алуан түрлі таптастырылым ақпарат бере алады, омонимдік, омографтық қолданыстарды автоматты түрде ажыратуға, яғни олардың әрқайсысының жиілігін жеке көрсетуге, тілдің өмір сүру кезеңдеріндегі актив лексикалық қордың құрамындағы сапалық, сандық өзгерістер мен трансформацияларды анықтауға мүмкіндік береді.

Себебі корпус – тек электрондық нұсқаға көшірілген мәтін базасы емес, ол терең аннотацияланған, тұтынушының сан түрлі сұранысына қарай тиісті ақпаратты беруге икемделген іздеу жүйесі бар «ақылды» база. Тереңірек анықтағанда, ұлттық корпус дегеніміз – мемлекеттік тіл ретіндегі қазақ тілінің лингвистикалық базасы болатын инновациялық ақпараттық-анықтамалық жүйе, оның мәтіндері арнайы белгіленім арқылы өңделіп, іздеу тетіктерімен қамтамасыз етілген, статистикалық әрі ғылыми дәл ақпаратты барынша қысқа уақытта алуға мүмкіндік беретін онлайн жүйесінде кез келген пайдаланушыға қолжетімді мегажоба.

Қазақ тілінің ұлттық корпусы – барлық стиль, жанр және әртүрлі типтегі мәтіндер көлемін түгел қамтитын мемлекеттік тілді сипаттау мен қолдаудың заманауи тетіктерінің бірі, сондықтан Ұлттық корпусы жасау ұлттық мәдениеттің, жалпы мемлекеттік-ұлттық бірегейліктің нығаюына, дамуына септігін тигізеді. Ол кез келген тұтынушыға қазақ тілінің бүкіл байлығын игеруге көмектеседі және қазақ тілінің графикалық бөлінісі салдарын жояды: кирил, латын және араб әліпбиімен бөлінген қазақ мәтіндерін бір жерге бір әліпби жүйемен жинақтайды; қазақ тілінің тарихын жаңғырту мен стандартталуына эмпирикалық база құрайды; көне және заманауи мәтіндер жиынтығы ретінде тілдегі үздіксіз құбылыстарды тіркеуге мүмкіндік береді, сол себепті қазақ тілін кодтау, стандарттау және жаңғыртудың мықты дереккөзі болады; мәліметтерінің дәйектілігі аса жоғары түрлі типтегі қағаз және электрондық нұсқадағы заманауи қазақ сөздіктерін жасауға, қазақ тілінің сөздік қоры, терминдері мен сөздері туралы дәйекті ақпарат алуға, соның ішінде жаңа жиілік сөздіктер әзірлеуге зор мүмкіндік береді (орыс тілінің үлкен академиялық сөздігінде – 150 000, Оксфорд сөздігінде – 600 000, үлкен «Қазақ сөздігінде» 106 000 сөз бар). Демек, Қазақ тілінің ұлттық корпусы – қазіргі және болашақ сөздіктердің сөздік қорын жинақтау мен ұлғайтудың мықты құралы.

Сонымен бірге Ұлттық корпус тіл оқыту, оқулық, оқу құралдарын, түрлі типте жаттығу, онлайн-мультимедиялық сервистер әзірлеуде кең қолданыс табады, өйткені оқыту мәтіні үшін аса маңызды т.б. сөздердің жиілігі мен контексті, қолданыс табатын мәтін типі мен жанры туралы жан-жақты ақпаратты тек Ұлттық корпус қана бере алады.

Осы сөздікті құрастырушылар – Ахмет Байтұрсынұлы атындағы Тіл білімі институтының тілтанушылары, болашақ қазақ сөздіктерін, соның ішінде заманауи жиілік сөздіктер жасау үшін алдымен Қазақ тілінің ұлттық корпусын әзірлеу ісін қолға алуға дайын.

А. Фазылжанова

СӨЗДІКТЕГІ ҚЫСҚАРҒАН СӨЗДЕР

зт – зат есім
ет – етістік – глагол
сн – сын есім
са – сан есім
ес – есімдік – местоимение
үс – үстеу
шл – шылау
ел – еліктеуіш сөз – звукоподражательные слова
од – одағай
ке – көмекші есім
мд – модаль сөз

ҚАЗАҚ ӘЛПБИИ

А а	И и	Р р	Ш ш
Ә ә	Й й	С с	Щ щ
Б б	К к	Т т	Ъ ъ
В в	Қ қ	У у	Ы ы
Г г	Л л	Ұ ұ	І і
Ғ ғ	М м	Ү ү	Ь ь
Д д	Н н	Ф ф	Э э
Е е	Ң ң	Х х	Ю ю
Ё ё	О о	Һ һ	Я я
Ж ж	Ө ө	Ц ц	
З з	П п	Ч ч	