

А. Жанабекова^{1*}, **Ж. Алпысбай²**

¹А. Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан;

²№15 мектеп-гимназия, Тараз, Қазақстан

*e-mail: aiman_miras@mail.ru

СӨЗДЕРДІҢ МОРФОЛОГИЯЛЫҚ ҚҰРАМЫН АВТОМАТТЫ ЛИНГВИСТИКАЛЫҚ ТАЛДАУ ЖАСАУ ТЕХНОЛОГИЯСЫ МЕН ПРОБЛЕМАЛАРЫ

Аннотация. Мақалада Қазақ тілінің ұлттық корпусындағы морфологиялық белгіленім программасын жүзеге асыратын морфологиялық анализатордың жұмыс істеу технологиясы қарастырылады. Морфологиялық белгіленім енгізу Ұлттық корпусындағы ең негізгі әрі маңызды лингвистикалық талдау болып табылады. Сонымен бірге аралық қабаттағы тілдік бірліктерді біз жасап отырған негізсөздер сөздігі мен грамматикалық формалар (сөзформалар) сөздігінде қалай ажыратылатыны туралы сөз еттіледі. Сөздердің құрамын түбір мен қосымшаға автоматты түрде бөлу және сөздердің сөз табына қатысын және сөзгертім қосымшаларының грамматикалық белгісін сипаттау компьютер үшін оңай емес. Алайда қазақ тілі сияқты жалғамалы тілдерде сөздердің құрамын автоматты түрде бөлу және сөз бөлшектеріне автоматты талдау жасау флективті тілдерге қарағанда жеңілдірек. Себебі жалғамалы тілдерде қосымшалар белгілі бір жүйемен жалғанады. Сөзформалардың формальды моделі анық боп келеді. Мақалада қазақ тіліндегі сөздерді автоматты түрде бөлшектеу мен талдаудың кейбір қиындықтары сөз болады, қазақ тілінің морфологиялық жүйесін модельдеудегі қиын мәселелер мен грамматикалық сөздікке енгізілетін категориялар мен шартты белгілер қарастырылады.

Тірек сөздер: корпус, белгіленім, анализатор, түбір, қосымша

А. Жанабекова¹, Ж. Алпысбай²

¹Институт языкознания имени А.Байтұрсынұлы, Алматы, Казахстан;

²№15-школа-гимназия, Тараз, Казахстан

*e-mail: aiman_miras@mail.ru

ТЕХНОЛОГИЯ И ПРОБЛЕМЫ АВТОМАТИЧЕСКОГО ЛИНГВОАНАЛИЗА МОРФОЛОГИЧЕСКОГО СОСТАВА СЛОВ

Аннотация. В статье рассматривается технология работы морфологического анализатора, реализующего программу морфологической разметки в Национальном корпусе казахского языка. Введение морфологической разметки является самым основным и важным лингвистическим анализом Национального корпуса. Также рассматривается как языковые единицы в промежуточном слое дифференцируются в словаре основных слов и словаре грамматических форм (словоформ), которые разрабатываем. Компьютеру сложно автоматически разделять слова на корни и аффиксы и описывать отношение слов к частям речи и грамматические характеристики аффиксов. Однако, в агглютинативных языках, таких как казахский, легче автоматически разделять слова и автоматически анализировать состав слов, чем в флективных языках. Это связано с тем, что в агглютинативных языках аффиксы прибавляются в определенной системе. Формальная модель словоформ более понятна. В статье рассматриваются некоторые трудности автоматического разбора и анализа слов в казахском языке. А также в статье

рассматриваются проблемные вопросы функциональных аффиксов в моделировании морфологической системы казахского языка, также вопросы, касающиеся категорий и условно-принятых кодов, входящих в грамматический словарь.

Ключевые слова: корпус, разметка, анализатор, корень, аффикс

A. Zhanabekova¹, Zh. Alpysbay²

¹A. Baitursynuly Institute of Linguistics, Almaty, Kazakhstan;

²15-school-gymnasium, Taraz, Kazakhstan

*e-mail: aiman_miras@mail.ru

TECHNOLOGY AND PROBLEMS OF AUTOMATIC LINGUOANALYSIS OF THE MORPHOLOGICAL COMPOSITION OF WORDS

Abstract. The article discusses the technology of the morphological analyzer, which implements the program of morphological markup in the National corpus of the Kazakh language. The introduction of morphological markup is the most basic and important linguistic analysis of the National Corpus. It is also considered how language units in the intermediate layer are differentiated in the dictionary of basic words and the dictionary of grammatical forms (word forms) that we are developing. It is difficult for a computer to automatically divide words into roots and affixes and describe the relation of words to parts of speech and grammatical characteristics of affixes. However, in agglutinative languages, such as Kazakh, it is easier to automatically separate words and automatically analyze the composition of words than in inflectional languages. This is due to the fact that in agglutinative languages affixes are added in a certain system. The formal model of word forms is more understandable. The article discusses some difficulties of automatic parsing and analysis of words in the Kazakh language. The article also discusses problematic issues of functional affixes in the modeling of the morphological system of the Kazakh language, as well as issues related to categories and conditionally accepted codes included in the grammatical dictionary.

Keywords: corpus, markup, parser, root, affix

Кіріспе

Қазіргі жаһандану тұсында ғылым салаларының барлығында дерлік жаңа технологияны пайдалану арқылы тез әрі сапалы түрде ғылыми нәтижелер алу қажеттігі туып отыр. Тіл ғылымы саласында да қазіргі кезде ғылыми-зерттеу жұмыстарын автоматтандыру ісіне үлкен мән берілуде. Осы орайда тіл ғылымын одан әрі дамыту үшін оны зерттеудің жаңа технологияларын қарастыру қажеттігі туды. Бүгінгі тіл зерттеушісі үшін ғылыми-зерттеу жұмыстарын қысқа уақыт ішінде орындап шығуға мүмкіндік беретін дайын ақпараттық құрал қызметін атқаратын тілдік корпустар жасаудың маңыздылығы дау тудырмайды. Шетел және орыс тіл білімінде орасан зор нәтиже беріп отырған тілдік корпустарды қазақ тіл білімінің материалдары негізінде жасау бүгінде үлкен сұранысқа ие.

Мұндай аса ауқымды тілдік ақпараттармен қамтамасыз етілген корпустар жасау жұмысының барысы бірнеше сатыдан тұрады. Компьютер жадына қаншалықты ақпарат енгізілсе, компьютердің жад арқылы өңдеп шығарып беретін ақпараты да соншалық болады. Сондықтан корпус құрастырушылар жадқа енгізілетін мәліметтерді өз сұраныстарына қарай икемдейді. Демек, компьютер адамның саналы ой-әрекетінің нәтижесі болып табылады.

Компьютерге арнайы тапсырма беру үшін, компьютер сол тапсырманы орындауға қажетті шарттарды толық білуі керек. Компьютер жадқа салынған арнайы формулалар, модельдер, шарттар арқылы жұмыс істейді. Сондықтан тілдік корпустар жасауда да тілдік материалдар арнайы модельдерге салынады.

Тілдік корпустар дегеніміз – үлкен массивтен тұратын лингвистикалық ақпаратпен қамтамасыз етілген автоматты жүйе. Басқаша айтқанда, корпустар әрі автоматты түрде жұмыс істеу керек, екіншіден, лингвистикалық ақпараттармен қамтамасыз етілу қажет. Корпусқа салынатын лингвистикалық ақпараттар тілдік жүйедегі тұрақты деңгейлер бірліктеріне негізделеді, яғни жүйелі тілдік қағидаттарға бағынады. Корпус құрастырушылардың компьютер жадына салатын жұмыс істеу шарттары да осы тіл болмысынан шығады. Сондықтан тілдік корпустардың барлық әрекеті, қимыл-қозғалысы лингвистикалық талаптарға байланысты болады.

Корпуста сұралатын лингвистикалық ақпараттың бірі – тілдің морфологиялық құрылысына қатысты мәліметтер. Мұндай мәліметтер алу үшін корпус жадына осы морфологиялық талдаулар туралы ақпараттар енгізіледі. Олар белгілі бір модель, формулаға түсіріледі. Корпустық лингвистикада мұндай ақпараттандырылған модельдерді (белгі-код) «разметка» деп атайды. Тіл деңгейлеріне қарай мұндай «белгіленімдер» түр-түрге бөлінеді. Морфологиялық деңгей бірліктерін таңбалаушы – морфологиялық белгіленім, синтаксис бірліктерін таңбалаушы – синтаксистік белгіленім т.б. аталады. Демек, корпусқа енгізілетін белгіленімдер де тілдік жүйеге салынады.

Тілдің морфологиялық жүйесін автоматты түрде бөлшектеу компьютер жадына бірнеше сатылы шарттардың орындалуы арқылы іске асады. Олар:

1) Сөздердің морфологиялық құрылымы екі бөліктен тұрады: біреуі – түбір сөздер және 2) оған жалғанатын қосымшалар. Түбір сөздер екіге бөлінеді: негізгі түбір және туынды түбірлер. Олардың екеуі біріктіріліп, негіз сөздер деп аталады. Компьютер ең алдымен мәтінде кездескен жекелеген сөздерді морфологиялық тұрғыдан бөлшектегенде негіз сөздерді ажыратып алуы қажет. Ол үшін жадқа негіз сөздердің тізімі берілуі керек. Бұл – бір.

2) Екіншіден, сөздердің морфологиялық құрылымының екінші бөлігі – оған жалғанатын қосымшалар. Компьютер жадына қосымшаларды бөлшектеп, олардың тілдік жүйедегі грамматикалық мағынасын білдіретін шартты белгіленімдер енгізілуі қажет. Жадқа осылай алдын ала дайындалған қосымшалардың тізімі мен шартты белгілері салынады. Бұл – екі.

3) Осы жұмыстармен қатар корпустардың қай түрі болмасын мәтіндер жиынтығынан тұратындықтан, тілдің әртүрлі стилін қамтитын электронды пішіндегі мәтіндер енгізіледі. Өйткені жоғарыдағы компьютер жадына салынған шарттарға сәйкес программа жүзеге асырылуы үшін сөздердің морфологиялық құрамын талдауға қажетті мәтіндер қажет. Жалпы корпустардың қуаттылығы да осы үлкен массивтерді қамтумен де байланысты болып келеді. Орыс және шетел тіл білімінде жасалған корпустар саны миллиондаған сөзқолданысты қамтиды. Көлемді массивті корпустар сөздердің морфологиялық құрамын бөлшектеу жұмысын компьютер көмегімен автоматты түрде атқарады. Ал мұндай программалар жұмысы үшін компьютер жадына жоғарыда аталған негізсөздер тізімі мен олардың грамматикалық түрленісі болып табылатын қосымша морфемалардың кодталған тізімі енгізілуі қажет.

Біздің бұдан бұрын жариялаған мақалаларымызда тілдің сөздік құрамын құрайтын негіз сөздер сөздігін жасаудың, яғни тілдегі сөздерді тізімдеп, түсіндіретін түсіндірме сөздік реестрлерін автоматты түрде теріп алудың механизмі қарастырылды. Бұл мақаламыз соның жалғасы ретінде сөз таптарына ажыратылған негізсөздерді тізімдеудегі проблемалық мәселелер туралы болмақ.

Материал және әдістер

Тілдердің барлығында негіз сөздер сол тілдің әртүрлі сөздіктерінде тізімделіп отырады. Сөздіктердің ішінде сөзді қамтуы жағынан түсіндірме сөздіктердің құзыреттілігі үлкен болып келеді. Сондықтан кез келген тілде жинақталған сөздік қор түсіндірме сөздіктерде көрініс табады. Осымен байланысты корпусқа морфологиялық белгіленім енгізуде сол тілдің түсіндірме сөздіктерінде берілген реестр сөздер негізге алынады. Қазақ

тілінде үлкен көптомдық түсіндірме сөздік: онтомдық және соңғы жарық көрген онбестомдық түсіндірме сөздіктер болып табылады. Осы орайда біздің корпусқа қажетті морфологиялық белгіленім енгізу үшін арнайы компьютерлік бағдарламалар көмегімен жүз мыңға жуық сөзді қамтыған онбестомдық «Қазақ әдеби тілінің сөздігінен» реестр сөздерді бөліп алу жұмысы орындалды (Қазақ әдеби тілінің сөздігі, 2011).

Негізсөздер сөздігінің жұмысы 15 томдық түсіндірме сөздіктің реестрлері автоматты жолмен алынғаннан кейін де жалғасады. Олар:

а) Корпусқа қажетті негіз сөздер сөз табына толығымен ажыратылуы қажет болғандықтан, сөз табы қойылмаған сөздерге арнайы сөз табына қатысты белгіленімдер қойып шығу қажет. Яғни реестр сөздерді өңдеу жұмысы жүргізіледі. Мысалы, 15 томдық сөздікте еліктеуіш сөздердің бірсыпырасы жеке реестр ретінде мағыналары ашылып беріледі де, кейбіреулері қос нүкте қойылып, олардың етістіктерімен тіркесі бірден беріліп кетеді. (Жарқ: жарқ етті.). Ал негізсөздер сөздігінде сөздер сөз табына ажыратылып тұру қажет. Сондықтан еліктеуіш сөздердің барлығында қос нүктелер алынып тасталып, еліктеуіш (ел.немесе елік) деген белгіленім қойылды. Осы сияқты түсіндірме сөздікте мағынасы тіркес құрамында ғана ашылған біраз сөздер кездеседі. Олардың барлығы біздің жұмыс позициямызға икемделіп, мүмкіндігінше белгілі бір сөз табына телініп берілді.

ә) Реестр сөздерді тізімдеу, өңдеу барысында оларға сөз таптарын қоюда бірлі-жарым қате кеткен тұстар кездеседі. Мұндай қате көбінесе сын есімнің де зат есімнің де қосымшасы болып табылатын *-лық/лік* қосымшалы сөздерде жиі кездеседі. Мысалы: *реттілік, түрпілік, өршімелік* сөздеріне *сын.* деген белгіленім қойылады. Алайда бұл сөздерге берілген дефиниция заттық мағына болып ашылады, сонымен қатар иллюстрациялық материалдар да зат есім сөзді дәйектейді. Сол сияқты *типологиялық* сөзіне *зат* деген белгіленім қойылады. Бұл сөздің еш мысалсыз-ақ сын есім екендігі белгілі. Осы сипаттас біршама сөздер өңдеу барысында түзетіліп отырды.

б) Негіз сөздерді өңдеуде атқарылатын тағы бір амал – реестр сөздер тізіміндегі бір сөз табына қатысты омоним сөздердің бірін ғана қалдыру. Себебі компьютер бағдарламасы мұндай бір сөз табына қатысты омоним сөздерді олардың мағыналық ерекшелігіне қарамай, бірдей таңбалайды.

в) Реестрді өңдеудің келесі бір сатысы сөзжасамдық жұрнақтар мен форма тудырушы жұрнақтардың аражігін айырып алу болып табылады. Түсіндірме сөздіктер сол тілдің сөз байлығы болғандықтан, олар негізінен тілдегі негізгі және туынды түбір сөздерді ғана қамтуы тиіс. Ал сөздерге грамматикалық мағына үстейтін қосымшалар түсіндірме сөздіктердің сөзтізбесіне енбеуі қажет. Алайда қазақ тілінің түсіндірме сөздігінде кейбір грамматикалық тұлғалы сөздер сөзтізбеден орын алады. Айталық, қимыл атау тұлғалары, салыстырмалы шырай тұлғалары. Бұл грамматикалық формалар түсіндірме сөздіктерде де шартты түрде алынып отыр. Корпусқа морфологиялық белгіленім енгізу үшін құрастырушылар сөздердің морфологиялық шегін компьютер жадына нақты айырып көрсету қажет. Егерде түсіндірме сөздік реестрінде берілгендіктен қимыл атау тұлғалы етістік сөздерді негіздер қатарында қалдырсақ, онда компьютер корпусқа енгізілген мәтіндердегі сөзформаларды морфологиялық бөліктерге бөлшектегенде, *у* тұлғалы етістіктердің барлығын түбір сөз деп көрсетеді де, бұл қосымша сөзжасамдық жұрнақ ретінде танылады. Алайда *у* қимыл атау формасының грамматикалық бірлік екендігі белгілі. Сондықтан біздің өңдеу жұмысымызда қимыл атау тұлғалы етістіктердің барлығы дерлік сөзтізбеден алынып тасталды. Мұны компьютер автоматты түрде атқарады. Алайда тағы бір ескеретін мәселе кейбір қимыл атау формасындағы етістіктердің заттық мәнді иеленуі. Мысалы, *қамсыздандыру, ақпараттандыру, әртараптандыру, хабарландыру* сияқты сөздер. Компьютер мұндай сөздерді оларға қойылған *зат есім* деген белгіленіміне қарай ажыратады да, оларды сөзтізбеде қалдырады. Сол сияқты түсіндірме сөздікте күшейтпелі шырайда тұрған сын есім сөздер түгелге жуық беріледі. Мысалы: *қып-қызыл, жап-жақын, ап-алыс* т.б. Біздің негізсөздер тізімінде де бұл сөздер негіздер қатарына енгізілді. Өйткені сөздерді

автоматты түрде бөлшектегенде күшейтпелі шырайда тұрған сөздерді компьютер арқылы ажырату қиынға соғады. Ал салыстырмалы шырай тұлғалы сөздер түсіндірме сөздікте біршама қамтылған. Алайда біздің негізсөздер тізімінен *-рақ/рек* пен *-лау/леу* тұлғалы салыстырмалы шырай тұлғалы сын есімдер алынып тасталды. Өйткені олар сын есімнің шырай категориясын құрайтын категориялық мағынаға ие грамматикалық бірліктер болып табылады. Сондықтан бұл формалар сөздердің түрленісіне арналған грамматикалық формалардың тізіміне кетеді, грамматикалық мағына ретінде кодталады.

Онбестомдық түсіндірме сөздікте етіс тұлғалы етістіктер де сөзтізбеде беріліп, мағыналары ашылады. Етіс жұрнақтары тілдік жүйеде етістіктің грамматикалық категориясы бола тұра, онтомдыққа да, онбестомдыққа да енгізілген. Өйткені тілімізде етіс тұлғалы етістіктер өте жиі кездеседі. Етістіктің статистикасын зерттеген ғалым, профессор С.Мырзабектеннің көрсетуінше, онтомдық сөздіктің жартысынан көбі тек етіс тұлғалы етістіктерден тұрады (Қазақ тексінің статистикасы, 1973). Сондықтан мұндай көлемдегі сөздік бірліктерді сөзтізбеден алып тастау үлкен ойсыраушылық туғызатыны анық. Екіншіден, жалпы түркологиялық еңбектердің көпшілігінде, соның ішінде етіс қосымшалары сөзжасамдық жұрнақтар ретінде танылады. Осы критерийлер етіс тұлғалы етістіктерді түсіндірме сөздіктердің сөзтізбесіне енгізілуіне негіз болған. Алайда етіс қосымшаларын таза тілдік теория тұрғысынан сөзжасамдық жұрнақ па, әлде форма тудырушы жұрнақ па деген мәселелер әлі де дау туғызып келе жатыр. Біздің негіз сөздер тізіміне етіс тұлғалы сөздердің барлығы енгізілді. (Яғни *азайт, азайтқыз, азайттыр* сияқты етіс тұлғалы етістіктердің барлығы негіздер тізімінде берілді). Өйткені етіс қосымшаларының таза сөзжасайтын қызметі де баршылық. Ал бұл олардың морфологиялық аражігін ажыратуда көп қиындық келтіреді. Сол себепті морфологиялық белгіленім енгізудің алғашқы сатысында әзірге етіс тұлғаларын негіздер қатарына жіберуді жөн көріп отырмыз. Тілдік корпустар құрастыруда оларға келешек түзету енгізу мүмкіндігі ескерілетіндіктен, етіс мәселесі кейінге қалдырылды.

Әдебиетке шолу

Негіздер тізімін алуда жоғарыда аталғандардан басқа да проблемалық мәселелер бар. Олар негізінен тілдік жүйенің өзіндегі аралық қабаттардан туындайды, яғни кейбір тілдік бірліктердің көпқызметтілігіне байланысты болып келеді. Айталық, *-дай/дей* қосымшалы сөздердің сөз түрлендіруші жұрнақтардан кейін жалғануы, сол сияқты *-сыз/сіз* тұлғалы сөздердің сөйлеу бірлігі ретінде қолданылуы, *-лық/лік* жұрнағының есімше тұлғаларынан кейін жалғануы т.б. Бұлардың барлығы тілдік жүйеде сөзжасамдық қызмет атқарғанмен, қолданыста сөйлеу бірлігі болатындығын ешкім жоққа шығармайды. Башқұрт тілінің зерттеушісі, компьютерлік лингвистика маманы З.Сиразитдинов бұл қосымшаларды сөзжасамдық жұрнақтар деген атаумен грамматикалық сөздікке енетіндігін айтады (Сиразитдинов, 2006). Демек, З.Сиразитдиновтың пікірінше де, мұндай көпфункционалы қосымшалар, атауына қарамастан, грамматикалық тұлға болып табылады.

Түркі тілдерінде осы негіз формалар мен туынды формаларды ажыратуда дау туғызып жүрген мәселе – сөзтудырушы қосымшалар мен сөз түрлендіруші немесе форма тудырушы қосымшалардың аражігін айыруға байланысты қайшылықтар. Сөздердің морфологиялық құрамын автоматты бөлшектеу мәселесінде де осы қайшылықтың шешімін табу аса қажет. Өйткені тілдік жүйеде шешімін таппаған тілдік құбылыстар оларды компьютер арқылы автоматты бөлшектеу ісінде де алдымыздан шығып отырады. Компьютер адамның ми қызметі арқылы салынған бағдарламаға сүйенеді, яғни адам тілдік бірліктерді қалай кодтайды, компьютер бағдарламаны жүзеге асыратын аймақты (мәтінді) сол кодтар бойынша бөлшектеп, талдап шығады.

Сөздің морфологиялық құрамын бөлшектеу үшін сөзжасамдық қосымшалар мен форма тудырушы қосымшалардың аражігін нақты ажыратып алу қажет. Түркі тілдері жалғамалы тілдерге жататындықтан сөздің құрам бөліктері флективті тілдерге қарағанда анағұрлым айқын. Солай бола тұра, түркі тілдерінің барлығына ортақ сипат – кейбір

қосымшалардың сөзжасамдық және сөзтүрлендірушілік қызметтерінің айқын ажыратылмауы. Тіпті қосымшалар классификациясында да бірізді жіктеме жоқ. Негізінен қосымшалар жігі сөзжасамдық, формажасаушы, сөзтүрлендіруші болып ажыратылады. Алайда түркі тілдерінің әрқайсысында ғана емес, бір тіл көлемінің өзінде зерттеуші ғалымдар оларды түрліше бөледі. Қосымшалар жіктемесіне қатысты проблемалар біздің кандидаттық диссертациямызда кеңінен қарастырылған болатын (Жаңабекова, 2001). Біз осы мақалада тікелей зерттеу объектімізге қатысты болғандықтан, тағы да қысқаша тоқталғалы отырмыз.

Қазақ тілінде алғаш жарық көрген грамматикаларда қосымшалар орыс тілінің ізімен сөз тудырушы, форма тудырушы, сөз түрлендіруші болып жіктеледі. Форма тудырушы қосымшалардың орнына сөз түрлендіруші терминін қолданып, жалғауларды сөз байланыстырушы деп көрсетушілер де бар. Мектеп оқулықтарында да осы үш түрге бөлінген классификация беріліп отырды. Қазақ тіл білімінде форма тудырушы мен сөз түрлендіруші қосымшаларды біріктіріп, оларды ортақ форма тудырушы қосымшалар деп танитын ғалым – Ы. Маманов. Ғалымның ізімен оның шәкірті С. Исаев та қосымшалар жігін осылай ажыратады (Исаев, 1998). Соңғы жарық көрген «Қазақ грамматикасына» С. Исаевтың пікірлері енген (Қазақ грамматикасы, 2002). Осыған ұқсас екі топтан тұратын жіктеме түркі тілдерінде, айталық, башқұрт тілінің академиялық грамматикасында көрініс тапқан. Онда қосымшалар сөзжасамдық және грамматикалық аффикстерге бөлінеді.

Біздің сөздердің морфологиялық құрамын автоматты бөлшектеуге қатысты қиындық тудыратын мәселе түркі тілдерінің барлығында бір шешімін таппаған аралық қабатта тұрған бір топ қосымшалардың (-дай/дей, -сыз/сіз, -лық/лік, -дағы/дегі, -ша/ше т.б.) тілдік табиғатынан туындайды. Олар қазақ тіл білімінде Қ. Шаяхметовтің зерттеуінде екі функциялы, қос функциялы қосымшалар деген атаумен белгілі (Шаяхметов, 1973). Татар тілінде мұндай қосымшаларды В.М. Насилов «кірістіруші аффикстер» (аффиксы включения) деп атаса (Насилов, 1958), олардың бірсыпырасын татар ғалымы Ф. Ганиев септіктер парадигмасына қосады (Ганиев, 1970). Бұлардан басқаша түркі тілдері бойынша оларды «промежуточные аффиксы», «көпфункционалы аффикстер», «функционалды жұрнақтар» т.б. атаулармен атайды.

Нәтижелер және талқылау

Аралық қабатта тұрған аталмыш қосымшалар сөздердің морфологиялық құрамын автоматты бөлшектеу кезінде қиындық тудырмау үшін олардың сөзжасамдық және сөзтүрлендірушілік немесе форма жасамдық қызметтерін нақты ажыратып алу қажет. Себебі автоматты бөлшектеуде сөздер осы екі дереккөз бойынша ажыратылады. Бірі – негіз сөздер, екіншісі – сол сөздерден өрбитін грамматикалық формалар, яғни сөз түрлендіруші қосымшалар. Автоматты жіктеме үшін компьютерге салынған жадта сөздің осы екі құрам бөліктерінің шекарасы айқын болу керек. Өйткені компьютер тек қана формалдық таңбаға ғана сүйенеді. Бұл ретте аталған «қос функциялы» қосымшалар арқылы жасалған дербес лексикалық бірліктерді бір жағынан негіз сөздер тізімінде көрсетіп, сол сияқты олардың сөз түрлендіруші жұрнақтардан кейін жалғанғандағы позицияларын олардың тілдік табиғатына, мазмұнына сәйкес атау беріп, шартты белгілер ойлап табуымыз қажет. Тілімізде осы аталған қосымшалар арқылы жасалған және дербес лексикалық бірлік болғандықтан, түсіндірме сөздіктерде реестр сөздер ретінде берілетін, біздің зерттеу нысанымыз бойынша негіз сөздер тізіміне енетін сөздер де бар, сол сияқты олардың зат есімнің немесе етістіктің сөз түрлендіруші жұрнақтарынан кейін келіп, форма тудырушы қызмет атқаратын кездері де бар. Мәселен, *-лық/лік* жұрнағы зат есім түбірлеріне жалғанып бірде зат есім тудырушы, бірде өлшемдік, қатыстық мән үстеп, сын есім тудырушы қызмет атқарады (адамдық, жақсылық, типографиялық, сындық т.б.). Сол сияқты бұл жұрнақ есімше тұлғаларынан кейін де жалғанып жұмсалады. Мысалы: *сөйлегендік, сатқандық, жазғандық, айтарлық, берерлік* т.б қолданыстарда сөйлеу бірлігі болып қолданылады. Өйткені бұл сөздер дербес лексикалық бірлік ретінде түсіндірме

сөздіктердің реестрлік қатарына енбейді. Бұл олардың әмбебаптық сипатымен де байланысты. Яғни кез келген есімше формалы етістікке *-лық/лік* қосымшасын жалғап уақытша субстантивтерді туындата беруге болады. Статистикалық мәліметтер бойынша, басқа қос функциялы қосымшаларға қарағанда *-лық/лік* қосымшасының етістік пен зат есімдерге жалғанып, сөз түрлендіруші/форма тудырушы қызмет атқаруы өте жиі кездеседі. Есімше тұлғаларынан кейін жалғанып жұмсалуына қатысты кезінде түркітанушы ғалым Н.А. Баскаков оны субстантив форма деп атаған болатын (Баскаков, 1979). Біз сөздерді автоматты бөлшектеу мен оларды кодтауға қатысты жұмысымызда бұл қосымшаның негізсөздер сөздігіне енбейтін, сөз түрлендіруші қосымшалардан кейін жалғанатын позицияларында оны осы өз мағыналық реңкіне сәйкес арнайы терминмен атауды жөн көріп отырмыз.

Қазақ тіл білімінде осы даулы қосымшалардың тілдік табиғатын сол кездің өзінде-ақ терең тани білген ғалым Ы. Маманов Н.А. Баскаковтың пікіріне сәйкес *-лық/лік* қосымшасының зат есімдерге жалғанып, арнаулық, өлшемдік мағына білдіретін жағдайында (көйлектік мата, терезелік ағаш т.б.) зат есімнің атрибут анықтауыш формасы деп атайды. Ал есімше (*-ған/ген*), қимыл атауларынан кейін жалғанғандағы жағдайын субстантив форма дейді. Сөздің морфологиялық құрамын автоматты бөлшектегенде *-лық/лік* қосымшасы үш түрлі позицияда тұрады. Біріншісі негіз сөздер тізіміне енген лексикалық бірліктер құрамында тұрады, бұл жағдайда компьютер тізімде бар *-лық/лік* формалы сөздерді дербес лексикалық единица ретінде танып, негіз сөздер орналасқан аймаққа қарай лақтырады. Екінші жағдайда тізімде жоқ *-лық/лік* формалы сөздердің түбірлерін негіз сөздерге жібереді де, *-лық/лік* қосымшасының зат есім түбірлеріне жалғанған кездегі қолданысын арнаулық, өлшемдік мәнді білдіретін атрибут анықтауыш форма деп таниды. Үшінші жағдайдағы *-ған/ген* есімше формаларынан кейін жалғанатын күрделенген формадағы қолданысында (*-ғандық/гендік*) оларды етістіктерді заттандырушы субстантив форма деген таңбамен белгілейді. *-лық/лік* қосымшасы *-ар/ер* есімше формасынан кейін жалғанғанда (барарлық, көрерлік, айтарлық) зат есімдерді субстантивтендірмей, керісінше зат есім түбірлеріне жалғанғандай арнаулық, өлшемдік мағына туғызып, атрибут анықтауыш форма туғызады.

Сол сияқты *-сыз/сіз* қосымшасының тұйық етістіктен кейін жалғанып жұмсалатын немесе зат есім түбірлеріне жалғанған жағдайдың өзінде дербес лексикалық бірлік тудырмай, жоқтық, болымсыздық мәнді уақытша формалар тудыратын кезі жиі кездеседі. Мысалы: *барусыз қал, кітапсыз, қаламсыз қалу, оқусыз қалу* т.б. Мұндай қолданыстардағы *-сыз/сіз* қосымшасын түркі тілдерінде М.А. Хабичев жоқтық, болымсыздық форма деп атайды (Хабичев, 1989). Қазақ тіл білімінде Қ. Шаяхметов аталған қосымшаның зат есімдердің түбіріне тікелей жалғанғандағы қызметін – сөз тудырушылық, ал форма тудырушы жұрнақтардан жалғанған кезде форма тудырушы қызмет атқарады дейді. Зерттеуші өзі зерттеген екі функциялы аффикстердің барлығын дерлік осы жүйемен бірізді етіп береді. Алайда *-сыз/сіз* қосымшасы зат есім түбірлеріне жалғанғанда әрқашан сөз тудыра бермейді. Ол жоқтық, болымсыздық мәнді сөйлеу бірліктерін де тудырады. Яғни сөйлеу бірлігі дегеніміз – ол сөздің грамматикалық формасын ғана тудырады деген сөз. Мысалы *-сыз/сіз* аффиксі *мейірімді, көңілді* сөздерінде туынды сын есімдер тудырған болса, *кітапсыз қалу, орамалсыз келу* деген сияқты тіркес құрамында сөйлеуде ғана қолданылатын, дербес лексикалық единица ретінде түсіндірме сөздіктерге енбейтін сөз формаларын жасайды. Демек, бұл қосымша форма тудырушы немесе сөз түрлендіруші жұрнақтардан кейін жалғанғанда ғана емес, сөздің түбіріне жалғанғанда да екі түрлі функция атқарады. Бұл қызметі жағынан олар омонимдік қатар құрайды. Егер компьютер жадына кез келген зат есімге *-сыз/сіз* қосымшасын жалғап сын есім тудыр немесе жоқтық, болымсыздық мәнді форма тудыр деген әмір берсек, автомат осы қосымшаны жалғап сөзформаларды туындата бергенмен, олардың лексикалық единица болу-болмауына қарамайды. Бұл жағдайда компьютерлік бағдарлама лексикалық бірліктер тізімінде, яғни негіз сөздер сөздігінде берілген *-сыз/сіз*

тұлғалы сөздерді сын есімдер деп белгілеп, оларды негіз сөздер тобына шығарады. Ал негіз сөздер сөздігінде берілмеген *-сыз/сіз* тұлғалы зат есімдерді туынды форма деп танып, аталған қосымшаға өзіміз кодтаған «жоқтық форма» деген белгіленім қояды.

Осыған ұқсас сын есім тудырушы, оның ішінде қатыстық сын есім тудырушы - *дай/дей* қосымшасы да *-сыз/сіз* қосымшасы сияқты зат есім түбірлеріне жалғанып та, етістік формаларынан кейін жалғанып та салыстыру мәнді сөйлеу бірліктерін туғызады. Сондықтан біздің зерттеу жұмысымызда аталған қосымшаның негіз сөздер тізіміне енбейтін жағдайларын түгелдей «салыстыру формасы» деп атадық. Мысалы: *-ғандай/гендей*, *-ымдай/імдей*, *-дағыдай/дегідей* т.б. Жоғарыда аталған қосымшалар сияқты *-дай/дей* қосымшасы зат есім түбірлеріне тікелей жалғанған жағдайдың бәрінде сөз тудырушы қызмет атқара бермейді. Мысалы: *кітаптай*, *қаламдай*, *үстелдей*, *пердедей* т.б. Мұндай кез келген сөзге жалғанған *-дай/дей* тұлғалы зат есімдер сөйлеу бірліктері болғандықтан, олардың барлығын бірдей түсіндірме сөздіктердің реестрінде беру мүмкін емес. Өйткені бұл қосымша түбірге жалғанып қолданылғанмен, түбірге тіркесіп қолданылатын «сияқты, тәрізді» септеулік шылауларымен бірдей мағыналық реңкті білдіреді, яғни жалғанған сөзіне салыстыру мәнін үстейді. Бұл жағдайда да автомат *-дай/дей* тұлғалы сөздердің негіз сөздер тізімінде барларын ғана негіз сөздер аймағына жатқызады да, қалған тізімде жоқ қолданыстарын екінші тізімнен, грамматикалық формалар, яғни қосымшалар тізімінен іздейді. *-дай/дей* қосымшасы *-ған/ген*, *-ар/ер* қосымшаларына жалғанғанда, сондай-ақ тәуелдік, көптік формаларына жалғанып (*-ымдай/імдей*, *-лардай/лердей*) жұмсалғанда жалғанған сөзіне салыстыру мәнін үстеп, форма тудырушы қызмет атқарады. Бұл жағдайларда аталған туынды формалар сөздердің грамматикалық формалары тізімінен іздестіріліп, алдын ала дайындалған морфологиялық белгіленімдер бойынша бөлшектеледі.

Осы жоғарыда қарастырылған қосымшалар сияқты тілдің синтетикалық жүйесінде әртүрлі қызмет атқаратын біршама қосымшалар бар. Мәселен, *-дағы/дегі*, *-шы/ші*, *-ша/ше*, *-лы/лі* сияқты қосымшалар.

Башқұрт тілінің зерттеушісі, компьютерлік лингвистика маманы З.Сиразитдинов жоғарыда қарастырылған функционалды қосымшаларды, біздің пікірлерімізге сәйкес, грамматикалық сөздікке енгізеді. Бірақ олардың сөзжасамдық қызмет атқаратын функциясын негізге алып, «сөзжасамдық жұрнақтар» деген атаумен атайды (Сиразитдинов, 2006). Осы атауды қолдана тұра, оларды сөзтүрлендіру тізімінде, яғни грамматикалық формалар сөздігінде қарастыратынын атап өтеді. Ғалым грамматикалық қосымшаларды сөзжасамдық қосымшалардан айырудың бір жолы олардың сөздерге талғамай жалғану шартымен де байланысты екендігін айтады. Бұл ретте қазақ тіл білімінде Ы. Маманов форма тудырушы немесе сөз түрлендіруші қосымшалар сөзге талғамай жалғанып, грамматикалық абстракция жасайды деген тұжырым жасаған болатын. Осы критерийге сәйкес аталған функционалды қосымшалар форма тудырушы қызмет атқарғанда, сөздерге жиі жалғанып жұмсалады, сөйтіп, грамматикалық абстракция жасайды. Аталған функционалды қосымшалардың осы ерекшелігі біздің оларды грамматикалық қосымша ретінде танып, грамматикалық формалар сөздігінде қарастыруымызға себеп болды.

Қорытынды

Демек, аталмыш қосымшалардың көпфункционалды қызметі формальдық тұрғыдан қиындық туғызғанмен, олардың әрқайсысының тілдік табиғатына сәйкес оларды сөзтүрлендіру аясында қарастырған жөн деп санаймыз.

Түрленуге бейім сөз таптары зат есім мен етістік екендігі белгілі. Сол себепті туынды формалардың дені осы екі сөз табынан туындайды. Енді грамматикалық сөздікке енетін категорияларды санамалап көрсетсек:

Зат есімнің грамматикалық категорияларынан:

1) *Көптік жалғау категориясы (КЖ)*: (-лар/лер, -дар/дер, -тар/тер);

2) *Септік категориясы:*

а) атау септік (ӨАС) – грамматикалық формасы жоқ болғандықтан, нөльдік формамен (Ø) таңбалаынады;

ә) ілік септік (ІС):

б) барыс септік (БС):

в) табыс септік (ТС):

г) жатыс септік (ЖС):

д) шығыс септік (ШС):

е) көмектес септік (КС):

3) *Жіктік категориясы (ЖЖ)*4) *Тәуелдік категориясы*

а) Тәуелдік жалғау (ТЖ):

ә) Тәуелдік форма (ТФ): -нікі/дікі/тікі

5) *Функционалды қосымшалар:*

а) Салыстыру формасы (СФ): -дай/дей, -тай/тей;

ә) Ұқсату формасы (ҰС): -ша/ше;

б) Жоқтық (болымсыздық) форма: -сыз/сіз;

в) Иелік форма (ИФ): -лы/лі, -ды/ді, -ты/ті;

г) Қатыстық форма (ҚФ): (-да) +ғы/ (-де) +гі, (та) +ғы/ (те) +гі;

д) Атрибутут-анықтауыш форма (АФ): -лық/лік, -дык/дік, -тық/тік (терезелік ағаш, көйлектік мата).

Зат есімнің сөзтүрленіміне жататын аталған грамматикалық категориялардан басқа да формалар автоматты белгіленім қою кезінде табылуы мүмкін. Оларды жұмыс барысында қарастыра береміз.

Етістіктің грамматикалық категорияларынан:

1) *Болымсыз етістік категориясы (БЖ):* -ма/ме, -ба/бе, -па/пе;

2) *Рай категориясы:*

а) Ашық рай (ӨАР): өткен шақ, келер шақ, осы шақ тұлғалары;

ә) Бұйрық рай (БР):

б) Қалау рай (ҚР):

в) Шартты рай (ШР):

3) *Көсемше категориясы (К):*4) *Есімше категориясы (Е);*5) *Тұйық етістік категориясы (ТЕ);*

б) *Функционалды қосымшалар:* -шы/ші (бар-шы, айтсын-шы); -сы/сі (көрген-сі, отырған-сы) т.б. Етістіктің түрленіміне қатысты функционалды қосымшаларға атау беру мәселесі алдағы жұмыстарымызда қарастырылмақ.

Бұл мақалада аралық қабаттағы тілдік бірліктерді біз жасап отырған негізсөздер сөздігі мен грамматикалық формалар (сөзформалар) сөздігінде қалай ажыратқанымыз туралы сөз еттік. Өйткені тілдік жүйедегі мұндай жайттардың табиғатын толық ажыратып алмай, сөздердің морфологиялық құрылымын автоматты түрде тани алмаймыз. Сол сияқты түрленім (грамматикалық) сөздігіне енетін грамматикалық категориялар мен грамматикалық формалар тізімін жалпылама атап өттік.

Тілдік жүйедегі мұндай жайттардың табиғатын толық ажыратып алмай, сөздердің морфологиялық құрылымын автоматты түрде бөлшектей алмаймыз.

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

Әдебиеттер

- Қазақ әдеби тілінің сөздігі. (2001) – Алматы, 2011.
- Қазақ тексінің статистикасы. (1973) Жинақ. – Алматы: Ғылым, 1973. – 731 б.
- Сиразитдинов З.А. (2006) Моделирование грамматики башкирского языка. – Уфа, 2006.
- Жаңабекова А. (2001) Сөз формаларын жасаудағы қосымшалардың функциялық ерекшеліктері: филол.ғыл.канд. дисс. – Алматы, 2001.
- Исаев С. (1998) Қазіргі қазақ тіліндегі сөздердің грамматикалық сипаты. Алматы: Рауан, 1998. – 303 б.
- Қазақ грамматикасы. (2002) – Астана: Астана полиграфия, 2002. – 784 б.
- Шаяхметов Қ. (1973) Екі функциялық аффикстер: филол. ғыл. Канд. дисс. – Алматы, 1973.
- Насилов В.М. (1958) Аффиксы включения // Сб. Вопросы языка и литературы стран востока. – М. 1958.
- Ганиев Ф.А. (1970) О синтетических и аналитических падежах в татарском языке. // Сб. Вопросы тюркологии. – Казань, 1970.
- Баскаков Н.А. (1979) Историко-типологическая морфология тюркских языков. – М.:Наука, 1979.
- Хабичев М.А. (1989) Именное словообразование и формообразование куманских языках. – М.:Наука, 1989. – 217 с.

References

- Qazaq adebi tilinin sozdigi. (2001) – Almaty, 2011. [Dictionary of the Kazakh literary language. (2001) - Almaty, 2011.] (in Kazakh)
- Qazaq teksinin statistikasy. (1973) Zhinaq. – Almaty: Gylym, 1973. – 731 b. [Statistics of the Kazakh text. (1973) Collection. – Almaty: Science, 1973. – 731p.] (in Kazakh)
- Sirazitdinov Z.A. (2006) Modelirovanie grammatiki bashkirskogo yazyka. – Ufa, 2006. [Sirazitdinov Z. A. (2006) Modeling the Grammar of the Bashkir Language. - Ufa, 2006.] (in Russian)
- Zhanabekova A. (2001) Soz formalaryn zhasaudagy Kosymshalardyn funkciyalyk erekshelikteri: filol.gyl.kand. diss. – Almaty, 2001. [Zhanabekova A. (2001) Functional Features of Applications in the Development of Word Forms: philol.science.cand. diss. – Almaty, 2001.] (in Kazakh)
- Isaev S. (1998) Qazirgi qazaq tilindegi sozderdin grammatikalyk sipaty. Almaty: Rauan, 1998. – 303 b. [Isaev S. (1998) The Grammatical Nature of Words in the Modern Kazakh Language. Almaty: Rauan, 1998. – 303 p.] (in Kazakh)
- Qazaq grammatikasy. (2002) – Astana: Astana poligrafiya, 2002. – 784 b. [Kazakh grammar. (2002) – Astana: Astana polygraphy, 2002. - 784 p.] (in Kazakh)
- Shayakhmetov K. (1973) Eki funkciyalyk affikster: filol. gyl. kand. diss. – Almaty, 1973. [Shayakhmetov K. (1973) Two-function Affixes: philol. science. cand. diss. - Almaty, 1973.] (in Kazakh)
- Nasilov V.M. (1958) Affiksy vklyucheniya // Sb. Voprosy yazyka i literatury stran vostoka. – M. 1958. [Nasilov V.M. (1958) Affixes of inclusion // Collection of Questions of the language and literature of the countries of the East. – M. 1958] (in Russian)
- Ganiev F.A. (1970) O sinteticheskix i analiticheskix padezhah v tatarskom yazyke. // Sb.Voprosy tyurkologii. – Kazan', 1970. [Ganiev F.A. (1970) On synthetic and analytical cases in the Tatar language. // Sat. Questions of Turkology. – Kazan, 1970.] (in Russian)
- Baskakov N.A. (1979) Istoriko-tipologicheskaya morfologiya tyurkskih yazykov. – M.:Nauka, 1979. [Baskakov N.A. (1979) Historical and typological morphology of the Turkic languages. – M.:Nauka, 1979.] (in Russian)

Xabichev M.A. (1989) Imennoe slovoobrazovanie i formoobrazovanie kumanskih yazykax. – M.:Nauka, 1989. – 217 s. [Khabichev M.A. (1989) Nominal word formation and formation of Cuman languages. – M.:Nauka, 1989. – 217 p.] (in Russian)