

ЕТІСТІКТІҢ МАҒЫНАЛЫҚ ТОПТАРЫНЫҢ ІШКОРПУСТА БЕРІЛУ ЖОЛДАРЫ

*Солтанбекова А.А.¹, Рамазанов Т.², Жолшаева М.С.³, Қалибекұлы Т.⁴

¹филол.ғ.к., А. Байтұрсынұлы атындағы Тіл білімі институты,
Алматы, Қазақстан, e-mail: madina-258@mail.ru,

²2-курс докторанты, А. Байтұрсынұлы атындағы Тіл білімі институты,
Алматы, Қазақстан, e-mail: take_kzz@mail.ru,

³филол.ғ.д., профессор, С. Демирель университеті,
Алматы, Қазақстан, e-mail: maira_zhania@mail.ru,

⁴филол.ғ.к., профессор Абылай хан атындағы Қазақ халықаралық
қатынастар және әлем тілдері университеті,
Алматы, Қазақстан, e-mail: tolkyn.kalibek@mail.ru

Аңдатпа. Бүгінгі таңда қазақ тіліндегі зерттеулер, ғылыми еңбектер, тіл білімінің жекелеген салалары бойынша жұмсалатын тілдік категориялар қатарын цифрландыру кеңістігінде пайдалану мүмкіндігіне ие боп отырмыз. Лингвистикалық білімдер жүйесін, тілтанымдық ресурстарды автоматтандыру жұмыстарын яғни корпуста енгізу мәселелерін А.Байтұрсынұлы атындағы Тіл білімі институтының ғалымдары дайындаған болатын. Қазір институттың сайтында «Қазақ тілінің ұлттық корпусы» деген платформа жұмыс істейді. 21 млн. сөзқолданыстан тұратын мәтіндер базасында тіл білімінің морфология, сөзжасам, лексика, фонетикаға қатысты білімдер жүйесі автоматтандырылған.

Ұлттық корпусың өзі іштей ішкорпуста бөлінеді. Сондай ішкорпустадың бірі сөз таптары бойынша лексика-семантикалық белгіленімін енгізу жұмыстары қарқынды жүріп жатыр. Мақалада етістіктердің лексика-семантикалық, лексика-грамматикалық белгі-кодтары анықталып ірі және ұсақ мағыналық топтарға жіктеу жұмыстары жөнінде айтылады. Мысал ретінде алынған етістіктерге лексика-семантикалық белгіленім қою арқылы қалай автоматтандырылады деген мәселе төңірегінде жазылады. Етістіктің лексика-семантикалық ішкорпусын дайындау үдерісі ең алдымен жинастырылған етістіктердің семантикалық ерекшелігіне қарай макро мен микро топтарға бөлуден басталады. Одан кейін барып лексика-грамматикалық, морфологиялық, сөзжасамдық, коннотациялық белгіленімдер әзірлемесін дайындаудың жолдары көрсетіледі.

Сонымен қатар осы уақытқа дейін ірі грамматист ғалымдардың етістіктің лексика-семантикалық топтарға жіктеу үлгілері сараланады. Қазіргі таңда соңғы сұрыпталған классификация үлгісі ғалымдардың мағына бөлінісімен салыстырыла отырып белгіленім енгізу үшін толықтырылады. Лексика-семантикалық белгіленім бағдарламасын жасауға қажетті етістіктің мағыналық топтарының классификациясы жасалып, сол арқылы Институтта етістік сөз табының лексика-семантикалық белгіленімдерін ішкорпуста енгізу жолдары сөз болады.

Мақалада лексика-семантикалық белгіленім ішкорпусына енетін етістіктерге лингвистикалық белгіленімдер әзірлемесін жасауда компьютерлік-бағдарламалық әдістер, корпустық лингвистиканың теориялық және практикалық жаңа білімдер жүйесімен толықтыруда сипаттамалы әдістер, сондай-ақ ішкорпуста енгізілетін етістіктерге морфологиялық, сөзжасамдық, лексикалық талдау жасау әдісі кеңінен қолданылды. Ұсынылған зерттеу жұмысын келешекте кез келген тіл үйренуші, студент, тіл маманы, жалпы қауым етістіктің лексика-грамматикалық, сөзжасамдық қызметін анықтау үшін және

лексика-семантикалық тобын анықтау үшін пайдалануға болады. Мақалада осы мақсатқа сәйкес белгіленім қою жолдары берілді. Бұл жұмыстың құндылығын арттыра түседі.

Тірек сөздер: ұлттық корпус, метабелгіленім, белгі-код, етістік, лексика-семантикалық белгіленім, лексика-грамматикалық ерекшелік, цифрландыру, ішкорпус.

Мақала BR11765619 «Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі Қазақ тілінің Ұлттық корпусын әзірлеу: ғылыми-зерттеу және оқыту интернет ресурсы» атты зерттеу жобасының аясында жазылды.

Негізгі ережелер

Лексика-семантикалық ішкорпус базасына енгізуге 18 200 етістік жинақталып, мағыналық реңктері анықталады. Gmail платформасында етістіктердің лексика-семантикалық белгіленімін әзірлеу үшін алты түрлі белгі-код енгізу жұмыстарынан ақпарат беріледі. Бірінші: морфологиялық белгіленім бойынша сөздің құрамы мен сөз табы анықталады, екінші: сөзжасамдық сипатына қарай дара, күрделі; негізгі, туынды, үшінші: лексика-грамматикалық белгіленім негізінде салт, сабақты; болымды, болымсыз; төртінші: коннотативтік сипатына қарай жағымды, жағымсыз, бейтарап деп жіктеледі, бесінші: етістіктердің лексика-семантикалық бөлінісі іштей ірі және кіші топтарға бөлінеді, алтыншы етістіктің мағынасын дұрыс анықтау мақсатында мысалы яғни иллюстрациясы қоса беріледі.

Жекелеген сөз таптарына лексика-семантикалық белгіленім енгізу тақырыбы қазіргі қазақ тілінің салалық дамуында, қазақ тілін автоматтандыруда, қазақ тілінің ұлттық корпусын оның ішінде жекелеген ішкорпустар базасын құрудың негізінде өзектілігін жоғалтпайды.

Кіріспе

Қимыл-қозғалыстың, іс-әрекеттің, қалыптың жай күйдің, көңіл-күйдің динамикалық үдерісін, түрлі құбылыстардың қозғалысын білдіретін сөз табын етістік деп атаймыз. Кез келген құбылыс, әрекет мезгілсіз, уақытсыз жүзеге аспайды. Уақыт, субъекті, объекті және қимылдан құралған құбылыстар күнделікті айналамызда болып жатады. Кез келген сөз табы секілді етістік те дара күйінде жалаң тұрып ешқандай грамматикалық та, лексикалық та мағына бермейді. Белгілі бір сөзбен тіркесіп, сөйлем ішінде, құрмалас сөйлемдерді байланыстыруда түрлі грамматикалық мағынаға ие болады. Сөз таптарының ішіндегі грамматикалық мағына мен формаға бай сөз табының бірі болып саналады. Етістік басқа сөз таптарына қарағанда күрделісі, қолданыс аясы мен өрісі өте кең, себебі етістіктің лексика-семантикалық және лексика-грамматикалық сипатын анықтауда қатысатын субъектінің іс-әрекеті ғана емес субъектіден тыс қоршаған ортадағы, қоғамдағы тілдік аядағы, ситуациялық жағдаяттарда орын алып жатқан сан түрлі қоғалыс, қимыл, әрекет, құбылыс, жай, күй сияқты үдерістер толық қатысады, сонымен қатар әр етістіктің семантикасын ашуда мезгіл, шақ ұғымдарын тыс қалдыруға болмайды. Себебі кез келген іс-әрекет белгілі бір уақыт шеңберінде жүзеге асып жатады. Ал лексика-грамматикалық сипатын анықтауда етістіктің грамматикалық формалары арқылы туындайтын грамматикалық мағыналарды айтамыз.

Мақалада етістіктердің лексика-семантикалық, лексика-грамматикалық белгі-кодтары анықталып ірі және ұсақ мағыналық топтарға жіктеу жұмыстары жөнінде айтылады. Мысал ретінде алынған етістіктерге лексика-семантикалық белгіленім қою арқылы қалай автоматтандырылады деген мәселеге төңірегінде сөз болады.

Материалдар мен әдістерді сипаттау

Мақалада лексика-семантикалық белгіленім ішкорпусына енетін етістіктерге морфологиялық, сөзжасамдық, лексика-грамматикалық белгіленім қою принциптеріне қатысты лингвистикалық материалды gmail платформасында электронды пішінге келтіру, лингвистикалық белгіленімдер әзірлемесін жасауда компьютерлік-бағдарламалық әдістер, корпусдық лингвистиканың теориялық және практикалық жаңа білімдер жүйесімен толықтыруда сипаттамалы әдістер, сондай-ақ ішкорпусқа енгізілетін етістіктерге морфологиялық, сөзжасамдық, лексикалық талдау жасау әдісі кеңінен қолданылды.

Әдебиетке шолу

Тіліміздегі метабелгіленім, белгіленім корпус, ішкорпус ұғымдары корпусдық лингвистика саласының негізгі ұғымдары саналады. Жалпы корпус мәселесі әлемдік тіл білімінде неше жыл бұрын зерттеліп кеткен. Корпус ұғымына анықтама беріп статикалық талдау жүргізген шетелдік ғалымдар Ч.Ф. Мейердің, М. Кито мен М.А. Риссаненнің, Мак Энери және Э. Вильсонның еңбектерінен көруге болады [1], [2], [3]. Корпусты қолдану, құрастыру жайында мәліметті Э. Финеганның зерттеулерінен [4], корпус ұғымын тар мағынада жұмсалуды В.П. Захаровтың еңбегінен [5], корпусдық лингвистиканың негізін мәтін құрайтындығын В.В. Рыковтың зерттеулерінде [6] сөз болады. Ал қазақ тілінің ұлттық корпусының негізін салған А. Жұбановтың зерттеулерінен [7] компьютерлік лингвистика, метабелгіленім, белгі-код деген негізгі ұғымдарды көреміз. Ғалым қазақ сөздерінің картотекалық базасының негізінде автоматтандырылған сөз қорын жасауда, қазақ тілінің статистикалық мәліметтерін сөздіктер түрінде шығаруда, қазақ тілінің ұлттық корпусының негізін салуда аянбай еңбек етті. А. Жұбановтың бастамасын ары қарай дамытып қазіргі ұлттық корпустың ішкорпустарын автоматтандыру ісіне белсене қатысып, корпусқа енетін мәтіндерге метабелгіленім енгізу жұмыстарын ұйымдастырып, белгіленім қою мәселесіне зерттеу жүргізіп жүрген ғалым, профессор А. Жаңабекованың еңбектерінде [8], осы мәселелер кеңінен сипатталады. Сондай-ақ, жас ғалым докторант М. Иманғазина корпусқа қатысты мақалаларында [9], семантикалық белгіленім мен лексика-семантикалық белгіленімнің арасын ажыратып, күрделі тұстарын анықтап береді.

Нәтижелер

Қазақ тілінің грамматикасының тарихында етістіктің жұмсалуды өрісін, қолданысын, лексика-семантикалық топтарын анықтауда, зерттеуде

А. Байтұрсынұлынан бастау алып, кейіннен А. Аманжолов, Н. Сауранбаев, Ы. Маманов, М. Оразов, А. Ысқақов, Н. Оралбаева, А. Қалыбаева т.б. ғалымдардың еңбектерінде егжей-тегжейлі жан-жақты қарастырылып зерттелді.

Қимыл-қозғалыстың қалпын, қарқындылығын, өту сипатын, анықтау үшін грамматикалық формалардың қатары да көп деп айтамыз. Сондықтан етістік сөз табы грамматикалық мағынаны анықтайтын грамматикалық формаларға өте бай сөз табы.

Жоғарыда аты аталған ғалымдардың зерттеулерінде етістіктің лексика-семантикалық топтарын мағынасына қарай бөліп берудегі негізгі ғылыми ұстаным түрліше сипат алды. Ал 1967 жылғы ұжымдық еңбек «Қазақ грамматикасында» [10], мағынасына қарай етістіктерді 6 ірі топқа бөледі және іштей кіші мағыналық бөліністерге бөледі. Ескертетін жайт, бұл ұжымдық еңбекте іс-әрекеттің, амалдың, жай-күй мен қалыптың қарқындылығын білдіретін етістіктер түбір күйінде еш қосымшасыз салт пен сабақтылық категориясына қарай ажыратып берілген және көмекші етістіктерді де мағыналық топтың алтыншы түріне енгізген екен. Жалпы саны 6 топтан тұратын етістіктің лексикалық әрі мағыналық жігін ажыратып көрсеткен.

Қазақ тілінде етістіктер әртүрлі қалыпты, жағдайды, алуан түрлі әрекетті, қозғалысты білдіреді. Белгілі бір етістіктің мағынасын тек мәтін арқылы анықтай аламыз. Бір етістіктің өзі мәтіндегі мағыналық қолданысына қатысты әртүрлі ұғымды тура мағынасымен қатар ауыспалы мағынада жұмсалатынын білеміз. Мысалы *кел* етістігінің лексика-семантикалық тобын анықтайық: *Жарқыным, айып емес кел дегенің, Өзің келсін деген соң, мен де келдім. Келсем де досты танып, дұшпан көріп, Қасымнан алыста тұр, келме дедім, (Айтыс). Жастар, сендердің білімдерің көп. Оған дауым жоқ. Дегенмен біз алпысқа келдік. Сендер жиырмаға келдіңдер, соны ұмытпаңдар (Ж. Молдағалиев). Жас балалар үлкен адамдарға қарағанда әлдеқайда намысшыл, менменшіл келеді («Халық мұраты»).* *Үндемей кіріп келдің тайпақ қара, Өзің қандай, мен қандай байқап қара (Жамбыл, Шығармалар жинағы).* Бірінші сөйлемде *кел* етістігі тура мағынасында белгілі бір орыннан бері қарай жүру, жақындау мағынасында яғни қимылдың субъектіге қатысын, қозғалысын білдіретін топқа жатады. Екінші сөйлемде *кел* етістігі ауыспалы мағынада, адам жасының белгілі бір мөлшерге жетуі, толуы сияқты ұғымды білдіреді. Келесі сөйлемде *бол* етістігінің мағынасында ауыспалы мағынада жұмсалса, соңғы сөйлемде негізі етістіктің жетегінде жұмсалып көмекшілік, қимылдың сипатын білдіріп тұр. Яғни тілде етістіктер өз мағынасынан бөлек ауыспалы мағынада да жұмсала береді, етістіктің лексика-семантикалық топтарын анықтау өте қиын, мүмкін емес, өйткені қимылды білдіретін сөз табы лексикалық мағынаға өте бай. Сондықтан ғалымдар ортақ мағынадан туатын сөздерді лексика-семантикалық сипатын анықтай отырып ірі топтарға, ірі топтарды ішінен микро топтарға ажыратып берген нұсқаларынан әртүрлілікті байқаймыз. Өйткені барлық етістіктерді мағынасына қарай ажырату өте

қиынға соққан, дегенмен басты ерекшеліктерін анықтай отырып топтарға бөліп көрсеткен.

1962 жылы жарық көрген «Қазақ тілі» мен 1967 жылы жарық көрген «Қазақ тілінің грамматикасы» еңбектерінде етістіктің лексика-семантикалық топтарға бөлінісі бірдей десек болады. Тек атауларында ғана өзгешелік бар. 1967 жылғы еңбекте етістіктің мағыналық бөлінісін салт пен сабақтылық қатынас негізінде анықтаса, 1962 жылғы еңбекте етістіктің жай-күйіне, қалпына, нысанына қатысты анықтағандығын және бұл еңбекте 7 топқа бөлгендігін байқаймыз [11]. Екі еңбекте де мағына жағынан бөлініс бірдей, тек 1962 жылғы еңбектегі *дерексіз мәнді етістіктер* 1967 жылғы *объектімен тікелей байланысты іс-әрекетті білдіретін сабақты етістіктермен* пара пар келеді. Екі еңбектегі етістіктердің қолданысы, семантикасы бірдей.

Ы. Маманов етістік категориясын негізгі етістік, функциялық етістік тұлғалар деп екіге бөледі. Негізгі етістіктің өзін өзара түбір етістік пен модификациялы етістіктер деп тағы екіге бөледі. Түбір етістік деп барлық етістіктің грамматикалық формаларына негіз болатын, лексикалық мағынаға ие сөздер, ал жасалу жолына қарай тағы да бірнеше түрге бөлінетіндігін айтады. Ғалым «Етістіктерді мағыналық ерекшеліктеріне қарай топтастырып бөлу өте қиын, тіпті мүмкін де емес. Дегенмен, оларды мағынасы жағынан басты ерекшеліктеріне қарай топтастыруға болады деп 8 топқа бөліп береді (1-кесте). Бірақ түбір күйінде етістіктерді мағынасына қарай топтастырып берудің грамматика үшін ешқандай практикалық мәні жоқтығын алға тартады. «Грамматика үшін түбір етістіктерді сыртқы тұлғалық бірлігіне қарай топтастырып бөлу қажет. Өйткені олар сыртқы тұлғалық бірлігі әртүрлі типтес сөздер тобын жасайды» – дейді [12]. Яғни түбір етістік өзі ғана тұрып лексика-семантикалық топ құрай алмайды, оған жалғанатын қосымшалардың негізінде («морфлогиялық және синтаксистік тәсілдер») грамматикалық мағынаның үстіне лексикалық мағына жасау арқылы ғана мағыналық топқа бөлуге болатындығын айтады.

1-кесте. Ы.Мамановтың бөлінісі

1	Іс-әрекеттік қимылды білдіретін етістіктер	жазу, сызу, қазу, жұлу, көтеру, тазалау, кесу, шашу
2	Қозғалыс қимылын білдіретін етістіктер	жүгіру, еңбектеу, домалау, жорғалау, өрмелеу, қозғалу, жылжу, қашу, жүзу
3	Қалып-күй процесін білдіретін етістіктер	ауыру, ұйықтау, тұру, отыру, жату, тынығу, жүдеу, жабығу, арықтау, семіру т.б.
4	Көңіл күйін білдіретін етістіктер	қайғыру, мұңаю, қуану, ренжу, күлу, жылау, ашулану, қызғану, т.б.
5	Сапалық белгінің өзгеру процесін білдіретін етістіктер	ұзару, қысқару, сарғаю, ағару, көбею, азаю, жұқару, қалыңдау, т.б.
6	Еліктеу, бейнелеу етістіктері:	жымындау, қылмындау, пысылдау, ыңқылдау, гүрсілдеу, жылтылдау, бұртию, күржию, т.б.
7	Туу, өсу қимылын білдіретін етістіктер:	балалау, құлындау, қоздау, боталау, күшіктеу, жұмыртқалау, т.б.
8	Субъективтік реңді етістіктер. Бұлар сөйлеушінің қимылға субъективтік көзқарасын көрсетеді.	азсыну, көпсіну, кісімсу, ұлықсыну, білгішсіну, т.б.

М. Оразов «Қазақ тілінің семантикасы» атты еңбегінде сөздерді лексика-семантикалық жүйеге түсіру үшін олар парадигмалық қатынасқа негізделеді, осы топты құрайтын әрбір мүше (тілдік бірліктер) өзіндік мағыналық дербестікке ие болып, сол жүйеде белгілі бір қызмет атқару қажет деп жазады [13]. Ғалым сема деген ұғымды қолданады. «Әр сөздің мағыналық құрамы іштей семалардан тұрады. Лексика-семантикалық топ құрау үшін, топ мүшелерінің лексикалық мағынасын анықтауда семаларға сүйенеміз» дейді. В.Г. Гак семаларды үшке бөледі: архисема, дифференциалды сема, потенциалды сема. Архисема сөздерді белгілі бір лексика-семантикалық топқа біріктірсе, дифференциалды сема сөздердің лексикалық мағынасына негіз болады деп көрсетеді. М. Оразов аты аталған семаларға қатысты қағидатты қолдай отырып, сөздерді лексика-семантикалық топқа топтастыруда сөздердің ортақ семасы жалпылауыш сема (архисема) дегенді, ал сөздерді лексика-семантикалық топтың басқа мүшелерінен ажыратып тұратын, сөздердің жеке дербестігін қамтамасыз етіп тұратын семаны айырушы (дифференциалды) сема деген атауды қолдайтындығын жазады.

Жалпылаушы семалар мен айырушы семалар сөздің лексикалық мағыналық тобын құрайды дедік, ал кейбір тілдік жағдаяттарда біреуі жетекші сема, екіншісі қосалқы көмекші сема қызметін атқарады дейді ғалым. Яғни М. Оразов сол кездің өзінде функционалды грамматика ұғымдарын, теориялық мәселелерін білген. Ядро, периферия сияқты ұғымдарды негізге алған. Сөз таптарын лексика-семантикалық топтарға топтастырғанда негізгі өзек (негізгі сема), периферия деп (қосалқы сема) деп қарағандығын байқаймыз. Ғалым етістік сөз табын лексика-семантикалық топтарға бөлу мәселесінде осы жалпылауыш семаны негізге алғандығын көреміз. «Жалпылаушы сема тематикалық, лексико-семантикалық топтарға сөздерді топтастыруға негіз болады, ал айырушы сема сол топтың мүшелерінің жеке-дербес жасауына жағдай жасайтын сема болып есептелінеді» – дейді. [13]. М. Оразов аты аталған еңбегінде қазақ тіліндегі етістіктерді жалпылауыш семаларына қарай 12 топқа бөледі, атауыштық семасы қарай 45 топқа бөледі (2-кесте).

2-кесте. М.Оразовтың топтастыруы

	Ірі мағыналық топтар	Ішкі мағыналық топтар
1	Амал-әрекет етістіктері	Қоректенуді білдіретін етістіктер: іш, же, ұртта, аса, дәм тат, сорапта, жалма, т. б. Жасау, өндіру етістіктері: жаса, соқ, жарат, тік, сал, құрастыр, қабы, т. б. Қирату, бүлдіру етістіктері: қират, бүлдір, бұз, сындыр, күл талқан ет, күлін көкке ұшыр т. б. Бөлу, ажырату етістіктері: бөл, ажырат, айыр, сұрыпта, т. б. Өлшеу етістіктері: метрле, қарыста, білемде, литрле, адымда, т. б. Белгілі бір құралмен орындалатын амал-әрекетті білдіретін етістіктер: балтала, бауызда, бұрғыла, пышақта т. б.
2	Қозғалу етістіктері	Жердегі қозғалысты білдіретін етістіктер: жүр, кел, еңбекте, сырғана, жорғала, т. б. Судағы қозғалысты білдіретін етістіктер: жүз, сүңгі, малты, жалда, т. б.

		Ауадағы қозғалысты білдіретін етістіктер: ұш, самға, қалықта т. б.
		Белгілі бір бағытқа қаратылған қозғалысты білдіретін етістіктер: кел, кет, бар, жақында, т. б.
		Бағытсыз қозғалысты білдіретін етістіктер: арала, шарла, кез, қыдыр т. б.
		Объектіден алыстауды білдіретін етістіктер: кет, алыста, ұза, без, т. б.
3	Қарым-қатынас етістіктері	Теңестіру мәнді етістіктер: қос, теңестір, салыстыр, бала, т. б.
		Меншіктеу, тәуелдеу етістіктері: иемде, меншікте, жіп тақ, ие бол, т. б.
		Амандасу, қоштасу етістіктері: амандасу, сәлемдесу, көрісу, танысу, қоштасу, қайырласу, т. б.
		Құрметтеу, қадірлеу етістіктері: құрметте, қадірле, ізетте, құрмет көрсету, т. б.
		Бастыру, бағындыру етістіктері: басқар, әмір жүргізу, сөз өткізу, т. б.
		Бағыну етістіктері: көну, бас иіу, ыркына көну (журу), бағыну, т. б.
4	Сезіну етістіктері	Жалпы сезіну мәнді етістіктер: сөз, байқа, абайла, аңғар, пайымда, т. б.
		Көру етістіктері: көр, көз сал, үңіл, т. б.
		Дәм сезу етістіктері: кермектену, ащылтымдану, т. б.
		Иіс сезу етістіктері: бықсы, мұңкі, аңқы т. б.
		Терімен байланысты сезіну етістіктері: сыздау, ашу, қышу, тызылдау, т. б.
		Есіту етістіктері: тыңда, есіт, құлақ сал, т. б.
5	Ойлау етістіктері	жалпы ойлау мәнді етістіктер (ойла, байымда, ой жүгірт т. б.), жорамалдау, болжау етістіктері (болжа, бағдарла, тұспалда, межеле, т. б.), қорытындылау етістіктері (қорытындыла, шешімге кел, бел байла, тәуекел ет, т. б.) сенім етістігі (сен, дос деп сана, дос көр, жақын сана, т. б.) еске сақтау етістіктері (еске ал, зерденде сақта, жадына алу, т. б.
6	Сөйлеу етістіктері	хабарлау етістіктері (айт, жеткіз, хабарла, т. б.), сөйлеу етістігі (сөйлесу, әңгімелесу, әзілдесу, т. б.), сұрау-жауап етістіктері (сұра, жауап бер, жауап қайтар, т. б.) эмоционалды сөйлеу етістіктері (мыңқылда, көкі, отта, т. б.),
7	Дыбыс-сес етістіктері	тарсылда, сатырла, тырс етті, гүрсілде.
8	Эмоционалды етістіктер	Таңдану етістіктері: таңдану, таңырқау, аузы ашылу, т. б. Қуану етістіктері: шаттану, қуану, мәз болу, т. б. Қанағаттану етістіктері: рахаттану, қанағаттану, көңлі толу, т. б. Қайғы-мұң етістіктері: қапалану, қайғыру, қабырғасы қайысу, т. б.
9	Қалып-сапа етістіктері	Созылыңқы амалды білдіретін етістіктер: отыр, тұр, жатыр, жүр, т. б. Заттардың түсінің өзгеруімен байланысты қолданылатын етістіктер: қарай, сарғай, қызар, т. б. Заттардың көлемінің өзгеруімен байланысты қолданылатын етістіктер: шөк, тапалдану, мықыраю т. б. Заттардың сандық мөлшерінің өзгеруімен байланысты қолданылатын етістіктер: азай, көбей, молай т. б. Адамның не заттардың ішкі қасиеттерінің өзгеруімен байланысты қолданылатын етістіктер: ірі, шірі, үгіл, тоз, т. б. Әдеттенуге байланысты етістіктер: әдеттену, үйрену, дағдылану т. б.

10	Өсіп-өну етістіктері	ту, төлде, қозыла, ботала, т. б
11	Субъектив реңктерді білдіретін етістіктер	кісімсі, адамсы, аспанда, баласын т. б.
12	Табиғи құбылыстарға байланысты етістіктер	азына, гуілде, борандату, найзағай түсу, т. б

Етістіктің лексика-семантикалық топтастыру жайында зерттеу жүргізіп, ғылыми негіздеп берген профессор А. Ысқақовтың тұжырымдамасын ерекше атауға болады. Ғалым түбір етістіктердің семантикалық жүйесін мағына жақындығына сай 11 топқа бөледі (3-кесте). Қимыл атауын білдіретін етістікті лексика-грамматикалық белгілері жағынан бірыңғай жеке сөз табы ретінде саналатындықтан олардың өздерін іштей бір-біріне мағына жағынан жақындықтарына, өзара қызметі жағынан байланысына орай топ-топқа бөлуге болатындығын жазды.

А. Ысқақов та профессор Ы. Мамановтың сөзін қолдағандай. Себебі етістік те басқа сөз таптары сияқты мағыналық бөлініске бөлуге болады. Бірақ грамматика үшін етістікті лексикалық мағынасына қарай бөлу «сондайлықты ұтымды бола қоймайды» деген еді. Себебі етістік жалаң күйінде, дара, түбір күйінде тұрып, ешбір синтетикалық, аналитикалық тәсілдерсіз оның грамматикалық формасы мен мазмұны осы тәсілдер арқылы анықталмай, семантикалық құрылымына орай жіктеп алмай лексика-семантикалық топқа бөлуге болмайтындығын айтады. Өйткені грамматикалық мағына грамматикалық формалар арқылы жасалатындықтан әрі сөздің құрылысы мен құрамына байланысты болғандықтан лексикалық семантикасын анықтау мүмкін емес. Ы. Маманов секілді А. Ысқақов та түбір күйіндегі етістіктерді мағына мен өзара қызметі жағынан орайластырып топтастырады [14].

3-кесте. А. Ысқақовтың бөлінісі:

1	амал-әрекет етістіктері	(босат, көтер, күрес, ки, сыз, өлше т. б.
2	қимыл-қозғалыс етістіктері	ауна, аудар, домала, жыт, қаш, секір т. б.
3	қалып-сапа етістіктері	жат, жантай, тұр, тос, күт, ұлғай т. б.
4	ойлау-сөйлеу етістіктері	айт, сөйле, де, ескер, жатта т. б.
5	өсу-өну етістіктері	балала, жапырақта, гүлде, өс, өн т. б.
6	бағыт-бағдар етістіктері	бар, кел, кет, найт, әпер, әкет т. б.
7	көңіл күйі етістіктері	жыла, қайғыр, өкін, күл, нуан т. б.
8	бейнелеу-еліктеу етістіктері	жарқыра, күркіре, дүркіре, тарсылда, сыр- тылда, сырқыра т. б.),
9	дыбыс-сес етістіктері	
10	көру-есту етістіктері	
11	мінез-құлық етістіктері	

Зерттеуші А. Хасенова да етістіктерді әр жақты қарап, түбір етістік дейтін межеге сыйғызуға келетін сөздерді іс-әрекет, амал, қалып процесінің динамикасын білдіруі жағынан төмендегіше 6 лексика-семантикалық топқа жіктеп көрсеткен [15]. Олар: *объектімен тікелей байланысты іс-әрекетті білдіретін етістіктер; субъектінің қозғалысының беталысын, бағытын*

білдіретін етістіктер; субъектінің қалпының сапасының өзгеру процесімен байланысты айтылатын етістіктер; бейнелеу етістіктері; ішкі объектілі салт етістіктер; көмекші етістіктер. Жоғарыда аты аталған ғалымдардың бөлінісі бір-бірімен мазмұндас. Лексика-семантикалық ішкорпусқа етістіктің мағыналық бөлінісін осы ғалымдардың классификациясын салыстыра отырып, іріктеп, макро және микро топтарға жіктеу арқылы белгіленім әзірлемесіне материал ретінде дайын сұрыпталған нұсқасын ұсынамыз.

Талқылау

Қазіргі таңда етістіктің лексика-семантикалық тобын анықтауда Тіл білімі институтының ғалымдары дайындаған, зерттеу жүргізе отырып, бұрынғы ғалымдардың еңбектері негізінде етістікті мағына мен қызметі жағынан 16 топқа бөліп метабелгіленімін енгізуде жұмыс жасап жатыр. Осы бөліністі негізге ала отырып «Етістіктің лексика-семантикалық ішкорпусы» атты міндет бойынша етістіктер тобын макро мен микро мағыналық топтарға жіктей отырып белгі-кодтарын, метабелгіленімін енгізу жұмыстары жасалу үстінде.

А. Байтұрсынұлы атындағы Тіл білімі институты 2022-2023 жылдар аралығында «Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі Қазақ тілінің ұлттық корпусын әзірлеу: ғылыми-зерттеу және оқыту интернет-ресурсы» атты бағдарламалық-қаржыландыру жобасы аясында Қазақ тілінің ұлттық корпусының ішкорпустарын жасап жатыр. Грамматика бөлімі 18200 етістіктің лексика-семантикалық белгіленімін енгізу жұмыстарын қолға алды. Лексика-семантикалық белгіленім бағдарламасын жасауға қажетті қазақ тіліндегі етістіктердің лексика-семантикалық топтарының тізімін жинақтап берген профессор Б. Момынованың 72 түрлі классификациясын негізге ала отырып етістіктерді морфологиялық, сөзжасамдық, лексика-грамматикалық, лексика-семантикалық белгі-кодтарға жіктеп, автоматты бағдарламасын әзірлеу үстінде. Мысалы, «ызалану, ашуға булығыу» мағынасында жұмсалатын *доңайбаттан* етістігінің лексика-семантикалық метабелгіленімін ішкорпусқа енгізіп, лингвистикалық автоматтандыру жұмысының тілтанымдық ресурсын жасап көрсетсек. Жалпы етістіктің лексика-семантикалық белгіленімін 6 параметрлер бойынша анықталады.

1. *Морфологиялық белгіленім: сөз табы – етістік.*

2. *Сөзжасамдық белгіленімнің өзі іштей төрт белгі-код арқылы анықталады:*

а) *сөздердің морфемалық құрамы: доңайбат+тан;*

ә) *сөзжасамдық тәсіл: синтетикалық;*

б) *тұлғасына қарай – дара;*

в) *құрамына қарай – туынды.*

3. *Лексика-грамматикалық белгіленімнің өзі іштей екі белгі-код арқылы жасалады:*

а) *іс-әрекеттің болымды/болымсыздығы – болымды;*

ә) *салттылық/сабақтылық қатынасқа қатысты – салт.*

4. *Лексика-семантикалық белгіленім бойынша екі белгі-кодпен өлшенеді:*

а) *макротоп – көңіл күй етістіктері;*

ә) *микротоп – адамның сезім-күйін білдіретін етістіктер* тобын құрайды.

5. Коннотациясы – жазымсыз.

6. *Иллюстрациясы – «Қанша елемейін деп сырттай доңайбаттанғанмен, іші құрмағыр тағат таба алмайды» («Қазақ әдебиеті»).* Осындай үлгіде әр етістіктің лексика-семантикалық, лексика-грамматикалық белгіленімін енгізіп шығамыз. Етістік күрделі сөз табы екендігін ескере отырып, көпмағыналы қимыл атауларының метабелгіленімін қою ерекшелігін қарастырсақ. Мысалы, үш түрлі мағынада жұмсалатын *егіл* етістігінің лексика-семантикалық белгі-кодтарын кесте түрінде берсек. 1. Бір нәрсенің отырғызылып, өсірілуі мағынасындағы *егіл* етістігінің жұмсалуды. 2. Белгілі бір аурудың таяқшасын қанға жіберіп індеттен сақтандыру деген мағынадағы медициналық термин ретінде кездесетін *егіл* етістігінің ерекшелігін атауға болады. 3. Іші-бауыры елжіреу, көңілі босау, езілу ұғымында жұмсалатын *егіл* етістігінің үшінші қоданысын атап кетуге болады.

4-кесте. Көп мағыналы етістіктердің ішкорпуста берілу үлгісі

Белгіленім түрлері	Ішкеі белгі-кодтар	ЕГІЛ 1	ЕГІЛ 2	ЕГІЛ 3
Морфологиялық белгіленім	Сөз табы	<i>етістік</i>	<i>етістік</i>	<i>етістік</i>
	Сөздің морфемалық құрамы	<i>Ек+іл</i>	<i>Ек+іл</i>	<i>Егіл</i>
Сөзжасамдық белгіленім	Сөзжасамдық тәсіл	<i>синтетикалық</i>	<i>синтетикалық</i>	<i>синт</i>
Лексика-грамматикалық белгіленім	Тұлғасы	<i>туынды</i>	<i>туынды</i>	<i>негіз</i>
	Құрамы	<i>дара</i>	<i>дара</i>	<i>дара</i>
	Іс-әрекеттің болымды/болымсыздығы	<i>болымды</i>	<i>болымды</i>	<i>боль</i>
	Салттылық/сабақтылық	<i>салт</i>	<i>салт</i>	<i>салт</i>
Лексика-семантикалық белгіленім	Макротоп	<i>күнделікті тұрмыста орындалатын әрекеттер</i>	<i>ауру, емдеу//емделуге байланысты етістер</i>	<i>адамның күйі, етістіктер</i>
	Микротоп	<i>Құрал-сайман, жабдықтар арқылы орындалатын әрекеттер мен мөлшерді білдіретін етістіктер</i>	<i>Ем-дом қабылдау</i>	<i>Адамның күйден екі ауысуы, екінші біреу күйге түсуі қалыптың</i>
	Коннотациясы	<i>бейтарап</i>	<i>бейтарап</i>	<i>бейтарап</i>
	Иллюстрация	<i>Ауа райы бүгін шынында да рақат жайлы еді. Емхана ауласындағы асфальт жолдар бойына егілген қара ағаштар қанатын жайып, айналасына салқын саясын түсіріп тұр (О.Сәрсенбаев, Жиде).</i>	<i>Мұның үстіне бруцеллезге қарсы шығарылған вакцина адамға мезгіл-мезгіл егіліп тұрса, нұр үстіне нұр (Ә.Дайырбеков, Малдан адамға жұғатын аурулар).</i>	<i>Елжіреу егіліп оған паранаға парана оған жеткендіктен жаралмаған (С.Хайдаров сыры).</i>

Бірнеше мағына білдіретін етістіктер қатарын 1, 2, 3 деген рет санын бере отырып, белгіленімін сол рет бойынша жеке-жеке енгіземіз (ЕГІЛ 1, ЕГІЛ 2,

ЕГІЛ 3). Ең бірінші өзінің тура өз мағынасында жұмсалыуына мән береміз келесі екінші, үшінші ұғымдарын мағына жақындығына сәйкес ажыратамыз .

Қорытынды

Байқағанымыздай, етістіктердің лексика-семантикалық классификация мен белгі-кодтар арқылы одан ары тереңдей анықтау үшін әр етістіктің макро, микро топтары яғни семантикасы анықтала отырып, олардың лексикалық мағынасын, омонимдік, ауыспалы, синонимдік қатарын ажырату мақсатында иллюстрациялық материал бір сөйлеммен берілді. Морфологиялық белгіленімде етістіктің құрамын (түбір мен қосымшасын) анықтау, тұлғасына қара дара/күрделі, болымды/болымсыз, негізгі/туынды, субъектінің қатысына қарай салт/сабақты, коннотациясына қарай жағымды/жағымсыз, бейтарап екендігін анықтау әрі сол негізде белгіленім қою бұрын-соңды атқарылмаған жұмыс. Тек біраз етістіктерге морфологиялық талдау жасалынынып ұлттық корпусқа белгіленімі енген болатын. Бұл жаңа белгіленім түрі болып саналады. Етістіктер тек мағыналық жағынан ажыратылып, бір-біріне жақын ұғымдағы, бірдей қызметтегі қимыл атаулары ірі және кіші мағыналық топтарға жіктеліп қана қойылмайды, сонымен қатар әр етістіктің лексика-грамматикалық белгіленімі, морфемдік белгіленімі, құрамы мен тұлғасына қарай бөлінісі, етістің қатыстылығына қарай яғни субъектінің ырықсыз, ортақтық, өзгелік семасының берілуіне байланысты субъектінің тікелей өзі не субъектісіз атқарылуына қарай белгіленім қойылады. Сөзжасамдық белгіленім де қойылады. Әріп ретімен берілген етістіктердің синтетикалық, аналитикалық тәсіл арқылы жасалуы да ескеріледі.

Грамматика бөлімінде лексика-грамматикалық→лексика-семантикалық→сөзжасамдық сипаттағы етістіктің белгі-кодтарын қою арқылы жалпы ұлттық корпусның ішкорпусында автоматты бағдарламасы дайындалып жатыр. Бұл тілді цифрландыру, автоматтандыру мақсатында үлкен жетістік саналады. Лексика-семантикалық белгіленім енгізу жалпы тілші қауымға, ЖОО ұстаздары мен мектеп мұғалімдері, филологтарға грамматикалық білімдер жүйесін жан-жақты зерттеуге мүмкіндік туғызары анық.

ӘДЕБИЕТ

- [1] Meyer Ch. F. English Corpus Linguistics & An Introduction. Cambridge, 2002. P.168.
- [2] Kytö M., Rissanen M. A Language in transition: the Helsinki Corpus of English texts, ICAME Journal, 1992. Pp.7-27.
- [3] McEnery T., Wilson A. Corpus Linguistics. Edinburgh, 2001. 235 p.
- [4] Finegan E. Language: its structure and use. N.Y.: Harcourt Brace College Publishers, 2004.
- [5] Lüdeling A., Kytö M., eds. Corpus Linguistics. An International Handbook. Volumes 1,2. Berlin & New York: Walter de Gruyter, 2008. – <http://alknyelvport.nytud.hu/muhelyek/elte.../HSK-Corpus-Linguistics.../file>.
- [6] Рыков В.В. Корпус текстов как реализация объектно-ориентированной парадигмы // Труды Международного семинара Диалог-2002. – М.: Наука, 2002. – С. 12-13

- [7] Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие. – СПб., 2005. – 48 с.
- [8] Жұбанов А. Компьютерлік лингвистикаға кіріспе: Оқу құралы. – Алматы: Қазақ университеті, 2007. – 204 бет.
- [9] Жұбанов А.Қ., Жаңабекова А.Ә. Корпустық лингвистика. – Алматы, 2017. – 336 б.
- [10] Қазіргі қазақ тілі. – Алматы, 1962. – 510 бет.
- [11] Маманов Ы. Қазақ тіл білімінің мәселелері. – Алматы: «Абзал-Ай» баспасы, 2014. – 640 бет.
- [12] Хасенова А. Етістіктің лексика-грамматикалық сипаты. – Алматы: Ғылым, 1971. – 307 б.
- [13] Қазақ тілі грамматикасы. Морфология. – Алматы, 1967. – 647 бет.
- [14] Ысқақов А. Қазіргі қазақ тілі. – 2-басылымы. Филология факультеттері студенттеріне арналған оқулық. – Алматы: Ана тілі, 1991. – 384 бет.
- [15] Оразов М. Қазақ тілінің семантикасы. – Алматы, Рауан, 1991. – 216 бет.

REFERENCES

- [1] Meyer Ch. F. English Corpus Linguistics & An Introduction. Cambridge. 2002. P.168.
- [2] Kytö M., Rissanen M. A Language in transition: the Helsinki Corpus of English texts, ICAME Journal, 1992. P.7-27.
- [3] McEnery T., Wilson A. Corpus Linguistics. Edinburgh. 2001. 235 p.
- [4] Finegan E. LANGUAGE: its structure and use. N.Y.: Harcourt Brace College Publishers. 2004.
- [5] Lüdeling A., Kytö M., eds. Corpus Linguistics. An International Handbook. Volumes 1, 2. Berlin & New York: Walter de Gruyter, 2008. – <http://alknyelvport.nytud.hu/muhelyek/elte.../HSK-Corpus-Linguistics.../file>.
- [6] Rykov V. V. Korpýs tekstov kak realizaciya obektno-orientirovannoi paradigmy (Corps of texts as the implementation of the object-oriented paradigm). Trudy Mezhdunarodnogo seminaru Dialog-2002. M.: Nauka, 2002. 12-13 s. [In Rus.]
- [7] Zakharov V. P. Korpýsnaiа lingvístika (Corpus linguistics). Uchebno-metod. posobie. SPB. 2005. 48 s. [In Rus.]
- [8] Zhubanov A. Kompýterlik lingvístikaға kirispe (Introduction to computer linguistics: a manual). Oqu qūraly. Almaty: Qazaq universiteti, 2007.204 b. [In Kaz.]
- [9] Zhubanov A. K., Zhanabekova A. A. Korpýstyq lingvístika (Corpus Linguistics). Almaty. 2017. 336 b. [In Kaz.]
- [10] Qazirgi qazaq tili (Modern Kazakh language). Almaty. 1962. 510 b. [In Kaz.]
- [11] Mamanov Y. Qazaq til biliminiń máseleleri (Problems of Kazakh linguistics). Almaty: «Abzal-Ai» baspasy, 2014. 640 p. [In Kaz.]
- [12] Khassenova A. Etistiktiń leksika-gramatikalyq sıpaty (lexico-grammatical nature of the verb). Almaty: Nauka, 1971. 307 b. [In Kaz.]
- [13] Qazaq tili gramatikasy (Grammar of the Kazakh language. Morphology). Almaty. 1967. 647 b. [In Kaz.]
- [14] Isakov A. Qazirgi qazaq tili (Modern Kazakh language).2-basylymy.Filologia fakülterteri studentterine arnalğan oqulyq. Almaty: Ana tili, 1991. 384 b. [In Kaz.]
- [15] Orazov M. Qazaq tiliniń semantikasy (Semantics of the Kazakh language). Almaty. Rauan. 1991. 216 b. [In Kaz.]

ПУТИ ПЕРЕДАЧИ СЕМАНТИЧЕСКИХ ГРУПП ГЛАГОЛОВ В ПОДКОРПУСЕ

*Солтанбекова А.А.¹, Рамазанов Т.², Жолшаева М.С.³, Калибекулы Т.⁴

*¹к.фил.н., Институт языкознания имени А. Байтурсынулы,
Алматы, Казахстан, e-mail: madina-258@mail.ru,

²докторант 2 курса, Институт языкознания имени А. Байтурсынулы,
Алматы, Казахстан, e-mail: take_kzz@mail.ru,

³д.фил.н., профессор, Университет Сулеймана Демиреля,
Алматы, Қазақстан, e-mail: maira_zhania@mail.ru,

⁴к.фил.н., профессор, КазУМОиМЯ имени Абылай хана, Алматы, Казахстан,
e-mail: tolkyn.kalibek@mail.ru

Аннотация. В статье автором рассматриваются возможности использования языковых категорий, в частности лингвистических разметок, в современном цифровом пространстве, которые могут использоваться в отдельных областях языкознания, в научных трудах и исследованиях на казахском языке. Внедрение системы лингвистических разметок и все работы по автоматизации языковых ресурсов подготовлены учеными Института языкознания имени А. Байтурсынова. Следует отметить, что сейчас на сайте института работает платформа «Национальный корпус казахского языка». Автоматизированы база текстов, состоящая из 21 миллиона словоупотреблений, и вся система языкознания, состоящая соответственно из таких разделов, как морфология, словообразование, лексика, фонетика.

В статье рассказывается о работе по классификации глаголов на макро- и микросмысловые группы с определением лексико-семантических, лексико-грамматических разметок-кодов. В качестве примера можно привести автоматизацию лексико-семантических разметок заимствованных глаголов. Процесс подготовки лексико-семантического подкорпуса глагола начинается прежде всего с деления собранных глаголов на макро- и микрогруппы в зависимости от их семантической специфики. Далее идет подготовка разработки лексико-грамматических, морфологических, словообразовательных, коннотационных разметок.

В то же время дифференцируются модели классификации глаголов на лексико-семантические группы крупными учеными-грамматистами. В настоящее время последняя отсортированная модель классификации дополняется для введения разметок с учетом значений, выделенных различными учеными. Разработана классификация семантических групп глаголов, необходимая для создания программы лексико-семантических разметок, необходимой для введения Институтым лексико-семантических разметок классов глагольных слов в подкорпус.

В статье широко использовались компьютерно-программные методы разработки лингвистических обозначений глаголов, входящих в подкорпус лексико-семантического обозначения, описательные методы в дополнение к новой системе теоретических и практических знаний корпусной лингвистики, а также метод морфологического, словообразовательного, лексического анализа глаголов, вводимых в подкорпус. Предлагаемая исследовательская работа может быть использована в дальнейшем любым изучающим язык студентом, лингвистом для определения лексико-грамматической, словообразовательной деятельности глагола и его лексико-семантической группы. В статье даны способы постановки разметки в соответствии с указанными целями. Это повышает ценность работы.

Ключевые слова: национальный корпус, метаразметка, знак-код, глагол, лексико-семантическая разметка, лексико-грамматическая специфика, цифровизация, подкорпус.

Статья написана в рамках исследовательского проекта BR 11765619 «Разработка Национального корпуса казахского языка как информационно-инновационной базы государственного языка: научно-исследовательский и обучающий интернет-ресурс».

WAYS OF TRANSMITTING SEMANTIC GROUPS OF VERBS IN THE INTERNAL CORPUS

*Soltanbekova A.A.¹, Ramazanov T.², Zholshayeva M.S.³, Kalibekuly T.⁴

*¹Candidate of Philological Sciences, The Institute of Linguistics named after A. Baitursynuly, Almaty, Kazakhstan, e-mail: madina-258@mail.ru,

²2st-year doctoral student, The Institute of Linguistics named after A. Baitursynuly, Almaty, Kazakhstan, e-mail: take_kzz@mail.ru,

³Doctor of Philology, professor, Suleiman Demirel University, Almaty, Kazakhstan, e-mail: maira_zhania@mail.ru,

⁴Candidate of Philological Sciences, professor, KazUIR&WL named after Ablai Khan, Almaty, Kazakhstan, e-mail: tolkyn.kalibek@mail.ru

Abstract. In the article, the author examines the possibilities of using language categories, in particular linguistic markup, in the modern digital space, which can be used in certain areas of linguistics, in scientific papers and research in the Kazakh language. Implementation of the linguistic markup system all works on the automation of language resources have been prepared by scientists of the A.Baitursynov Institute of Linguistics. It should be noted that now the platform "National Corpus of the Kazakh language" is working on the institute's website. The database of texts consisting of 21 million word uses and the entire system of linguistics, consisting respectively of such sections as morphology, word formation, vocabulary, phonetics, has been automated.

The article describes the work on the classification of verbs into macro and micro semantic groups with the definition of lexico-semantic, lexico-grammatical markup codes. As an example, we can cite the question of how this is automated by labeling lexico-semantic markings on borrowed verbs. The process of preparing the lexico-semantic subcorpus of a verb begins primarily with the division of the collected verbs into macro- and micro-groups, depending on their semantic specificity. Next are the methods of preparing the development of lexico-grammatical, morphological, word-formation, connotational markings.

At the same time, first of all, the models of verb classification into lexico-semantic groups are differentiated by major grammarians. Currently, the latest sorted classification model is being supplemented to introduce markup with a comparison with the values of scientists. A classification of semantic groups of verbs has been developed, which is necessary to create a lexico-semantic markup program, through which the Institute will discuss ways to introduce lexico-semantic markup of a class of verb words into the subcorpus.

The article widely used computer-software methods for the development of linguistic designations of verbs included in the subcorpus of lexico-semantic notation, descriptive methods in addition to a new system of theoretical and practical knowledge of corpus linguistics, as well as the method of morphological, word-formation, lexical analysis of verbs introduced into the subcorpus. The proposed research work can be used in the future to determine the lexico-grammatical, word-formation activity of any language learner, student, linguist, universal verb and to determine the lexico-semantic group. The article describes the ways of setting the markup in accordance with this goal. This increases the value of the work.

Keywords: national corpus, meta-markup, sign-code, verb, lexico-semantic markup, lexico-grammatical specificity, digitalization, subcorpus.

The article was written within the framework of the research project BR11765619 «Development of the National Corpus of the Kazakh language as information-innovation state language base: research and training internet resource».

Статья поступила 21.02.2023

