

С.Қ. Құлманов*, А.Ә. Жаңабекова
Н.М. Әшімбаева, А.З. Бисенғали
Н.Қ. Шүленбаев, Б.Қ. Қордабай

А. Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан
Байланыс үшін автор: k.k.sarsen@mail.ru

Корпусқа енгізілетін мәтіндердегі сөздерге морфологиялық белгіленім қою және оларды компьютерлік бағдарламаға енгізу мәселелері

Аңдатпа. Мақалада тіл біліміндегі корпус жасау ісінің тарихына қысқаша шолу жасалып, корпусстық лингвистика ғылымына сипаттама беріледі, морфологиялық белгіленімге қойылатын теориялық және практикалық міндеттер мен талаптар көрсетіледі.

Корпусқа енгізілетін мәтіндердегі сөздерге морфологиялық белгіленім қою алғашында қолмен жасалды. Мысал ретінде алынған сөйлемдегі жеке сөздерге жасалған морфологиялық талдаудың негізгі принциптеріне, қойылған белгіленімдерге түсініктеме берілді. Морфологиялық талдау, негізінен, мәнмәтінге байланыссыз жасалатыны белгілі. Мақалада сөз таптарының морфологиялық құрылымдарын талдау мен оларға белгіленімдер қоюда кездескен негізгі ерекшеліктер жеке-жеке көрсетіледі.

Сөздердің морфологиялық құрамын автоматты түрде ажырату компьютер жадына бірнеше сатылы шарттардың орындалуы арқылы іске асады. Олар: 1) сөздердің морфологиялық құрылымын ажырату (түбір сөз, қосымшалар); 2) компьютер жадына алдын ала дайындалған қосымшалардың тізімі мен шартты белгілерін енгізу; 3) компьютер жадына тілдің әртүрлі стилін қамтитын, морфологиялық белгіленім қойылған электронды пішіндегі мәтіндерді енгізу. Бұдан кейін компьютерлік бағдарлама арқылы мынадай жұмыстар орындалады: а) сөз табы қойылмаған кейбір сөздерге белгіленімдер қою; ә) реестр сөздерді тізімдеу, өңдеу барысында оларға сөз таптарын қоюда кеткен кейбір қателерді қолмен жөндеу; б) реестр сөздер тізіміндегі бір сөз табына қатысты омоним сөздердің бірін ғана қалдыру; в) сөзжасамдық жұрнақтар мен форма тудырушы жұрнақтардың аражігін ажырату.

Мәтіндер корпусын құрастыруға арналған компьютерлік бағдарламаны жасаудың мынадай инженерлік-лингвистикалық және техникалық проблемаларын атауға болады:

- корпусқа енгізілген мәтіндерді морфологиялық талдау принциптері жаңадан жасалып отырғандықтан, әрі қарай жетілдіруді талап етеді;
- морфологиялық белгіленім қоюда тілдің ішкі заңдылықтарына байланысты шартты түрде алынған белгілер қолданылатындықтан, компьютерлік бағдарлама жасау барысында олар да өзгерістер мен толықтырулар енгізуді қажет етеді.

Кілт сөздер: корпус, корпусстық лингвистика, мәтін, морфология, шартты белгі, белгіленім, компьютерлік программа

DOI: <https://doi.org/10.32523/2616-678X-2022-140-3-103-113>

Кіріспе

Тіл білімінде корпус жасау ісі 1960 жылғы Лондон университетінің колледжінде Р. Квирктың ғылыми жетекшілігімен «Обзор употребления английского языка» жобасы аясында алғашқы ауызша тіл корпусы болып табылатын ағылшын тілінің Лондон-Лунд (The London-Lund Corpus) корпусының бастау алады [1].

1963 жылы АҚШ-та пайда болған У. Нельсон Фрэнсис пен Г. Кучера (Браун университетінің лингвистика кафедрасы, Род-Айленд, АҚШ) әзірлеген көлемі 1 млн. сөзқолданыстан тұратын Браун корпусының (The Brown Standard Corpus of American English) құрамына әрбірі 2 мың сөзқолданысқа тең 500 мәтін қамтылды [2; 3]. 1990 жылдары Браун корпусы кеңейтіліп, Ланкастер-Осло/Берген (LOB), Браун корпусының британдық ағылшын нұсқасы, сондай-ақ Frown және FLOB, Brown және LOB баламалары жасалып, оларға арналған ақпараттық нұсқаулықтар шығарылды [4; 5]. Осы кездерде тілдік зерттеулердің басым бөлігі компьютерленген мәтіндік корпусстарға лингвистикалық талдау жүргізуге бағытталды. Осыған байланысты «корпустық лингвистика» саласының зерттеу нәтижесі тілдік зерттеулерге тәжірибе жүргізу ісінде, сондай-ақ сөздік әзірлеу, грамматикалар мен оқу құралдарын жазуда кеңінен қолданылып отыр [6; 7].

Корпустық лингвистика тіл ғылымының математикалық лингвистика, дискурстік анализ және лексикография салаларымен тығыз байланыста дамып келеді. Сондықтан, бір жағынан, мәтіндер корпусын жасау корпустық лингвистика қызметінің нәтижесі болса, екінші жағынан, лингвистиканың басқа салаларында жүргізілетін зерттеулерге эмпирикалық материал бола алады. Осыған байланысты корпустық лингвистиканың жетістіктері, бір жағынан, қазақ тіл білімінде тілші ғалымдарға зерттеуге қажетті тілдік материалды тез тауып алып, оларды өңдеуде тиімді мүмкіндік берсе, екінші жағынан, корпус мемлекеттік тілді үйрену мен үйрету ісінде тіл үйренушілерге де, тілді оқытушыларға да,

тіпті көпшіліктің күнделікті қарым-қатынас жасау барысында қажетті қазақ сөзінің бүкіл «лингвистикалық сипаттамасын» табуға, оны меңгеруге таптырмас құрал саналады.

Зерттеудің әдістемесі

Мақалада корпусқа енгізілетін мәтіндердегі сөздерге морфологиялық белгіленім қою принциптеріне қатысты *лингвистикалық материалды электронды пішінге келтіру, сәйкестендіру (идентификациялау), лингвистикалық белгіленімдер әзірлемесін жасауда және пайдаланушыға қажетті лингвистикалық және экстралингвистикалық ақпаратты жылдам тауып алуға мүмкіндік беретін «Іздеу» жүйесі интерфейсін жетілдіруде компьютерлік-бағдарламалық әдістер, отандық корпустық лингвистика ғылыми пәнінің мазмұнын корпус әзірлеу барысында жинақталған теориялық және практикалық жаңа білімдер жүйесімен толықтыруда сипаттамалы әдістер, сондай-ақ корпусқа енгізілетін мәтіндердегі сөздерге морфологиялық талдау жасау барысында құрылымдық (морфологиялық) талдау әдісі кеңінен қолданылды.*

Талқылау

Корпустық лингвистика саласындағы көрнекті ғалым В.П. Захаров «лингвистикалық корпус» терминіне: «Под названием лингвистический, или языковой, корпус текстов понимается большой, представленный в электронном виде, унифицированный, структурированный, размечанный, филологический компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач» [7, 3-б.] деген анықтама береді. Яғни корпус – мәтіндердің электронды пішінге келтірілген жинағы, ал оның қызметі осы мәтіндер жинағын зерттеу жүргізуге пайдалануға, тіпті күнделікті қарым-қатынас барысында қажетін өтеуге бағытталады.

Тіл білімінде лингвистикалық корпусстар төмендегідей топтарға бөлініп қарастырылады:

І. Құрылымдық дәрежесіне қарай: 1) электрондық архив (сыртқы есте сақтау таспадағы

өңделмеген мәтіндер); 2) электрондық кітапхана (өңделген пішіндегі мәтіндер); 3) мәтіндер корпусы (мәтіндердің стандартталған, өңделген және әр пәндік саланы бейнелейтін жиынтығы).

II. Тілдердің қамтылу санына қарай: 1) біртілді; 2) екітілді; 3) көптілді.

III. Қолдану тәсіліне қарай: 1) зерттеу корпустары (көбінесе тілдік жүйелер қызметінің әртүрлі аспектілерін зерттеуге арналған); 2) иллюстративтік (көрнекілік) корпусстар (олар зерттеу жүргізілгеннен кейін ғана құрылады; бұл корпусстардың мақсаты – алынған нәтижелерді растау және негіздеу).

IV. Динамикалық сипатына қарай: 1) тілдік жүйенің белгілі қалпын бейнелейтін статикалық корпусстар; 2) динамикалық (мониторлық) корпусстар.

V. Индекстеуге қатысты: 1) белгіленбеген жай корпусстар; 2) белгіленген (аннотацияланған, мазмұнданған) корпусстар. Қазіргі кездегі корпусстар тек аннотацияланған болып келеді, себебі тек белгіленген корпусстардан ғана қажетті ақпаратты іздестіруге, пайдалануға мүмкіндік бар.

VI. Тіл формасына (түрпатына) қарай: 1) жазба түрдегі корпусстар; 2) сөйлеу түріндегі (фонетикалық) корпусстар. [8, 175-179-бб.].

Профессор А.Жұбанов корпусстық лингвистика әдістері ішінен мынадай төрт топты жеке бөліп көрсетеді: 1) филологиялық әдістер; 2) теориялық-лингвистикалық әдістер; 3) математикалық (статистикалық) әдістер; 4) ақпараттық технологиялардың әдістері [8, 176-б.]. Ғалым бірінші топқа жанрға, мәтін авторларына, мәтіндердің жанрлық сипатына және т.б. мәселелерге қатысты әдістерді, екінші топқа – мәтін таңдаудың және корпусстағы лингвистикалық ақпараттың берілуінің жалпы және жеке лингвистикалық тәсілдерді, үшіншісіне таңдама мәтіндердің санына, оның репрезентативтілігін (тұлғалылығын) қамтамасыз ету аспектілеріне қатысты математикалық статистиканың әдістерін, төртінші топқа корпус деректерін компьютерге енгізу және оны өңдеу іс-әрекеттерін қамтамасыз ететін ақпараттану әдістерін жатқызады [8, 177-б.].

Қазіргі кезде қазақ тілінің мәтіндер корпусын жасаудың жалпы лингвистикалық теори-

ялық және практикалық [8] проблемаларына, мәтіндерге әртүрлі белгіленімдер қою және оны автоматтандыру [9; 10; 11], сөзформаларды автоматты талдау алгоритмін әзірлеу [12] және т.б. мәселелерге арналған ғылыми зерттеулердің нәтижелері жарық көрді.

Нәтижелер

Корпус әртүрлі стильден алынған сөз, сөзтіркес, сөйлем немесе толық мәтіндер жиынтығынан тұрады және осы мәтіндердегі әрбір тілдік бірліктің лингвистикалық сипаттамасы жасалады. Мұндай лингвистикалық сипаттама «аннотациялау» деп аталады. «Аннотациялау» тілдік бірлікке шартты белгілер арқылы берілетін «қысқаша мазмұн» болып табылады. Мұндай шартты белгілер қазақ тілінде «белгіленім» (А.Жұбанов) деген терминмен аталады.

Филология ғылымдарының докторы А. Жаңабекова тілдік корпус жасауда, яғни грамматикалық белгі қоюда мынадай мәселелерге назар аудару керек деп көрсетеді: «1) Шартты белгілер қою жағын ойластыру; 2) Грамматикалық белгілерді таңдау; 3) Белгіні контекст бойынша қоямыз ба, әлде жүйедегі барлық грамматикалық белгілерді қамтимыз ба, осы мәселенің басын ашып алу; 4) Мәтіндегі сөздерді бөлшектеуде олардың синтагмалық байланысын сақтап отырамыз ба, әлде әрбір сөздің формасын парадигматикалық қатарға түсіріп, жеке-жеке аламыз ба, осыны анықтап алу (мысалы, «барын салды» деген тіркесті екі форма түрінде бөлек-бөлек қараймыз ба, әлде бірге күрделі етістік ретінде аламыз ба?) және т.б.» [9, 45-б.]. Біздіңше, біріншіден, әрбір сөз табына қойылатын халықаралық белгілер бар. Алайда олар шартты түрде болғандықтан, біз сөз таптарына тіліміздің сипатына сәйкес белгі қоюымыз қажет, екіншіден, әртүрлі сөз таптарының грамматикалық формаларын біршама толық қамту үшін аталған шартты белгілерге қосымша белгі қоюымыз керек, ал үшінші және төртінші мәселе, негізінен, тілдік деректерді талдау әдісіне байланысты қойылуы керек; тілдік бірліктерге морфологиялық талдау жасалатын болса, олар (тілдік бірліктер) мән-

мәтіннен тыс, жеке-жеке алынып талданады, ал синтаксистік және семантикалық талдауда міндетті түрде мәнмәтін ескеріледі. Яғни корпусқа тілдік деректердің толық (ең алдымен морфологиялық, содан кейін синтаксистік, семантикалық) құрылысы сипатталуға тиіс.

Корпусқа енгізілетін мәтіндердегі сөздерге морфологиялық белгіленімдер қою (лингвистикалық) және оларды компьютерлік бағдарламаға енгізу (компьютерлік) мәселелерін екіге бөліп қарастыруға болады:

I Мәтіндегі сөздерге морфологиялық белгіленім қою. Бұл жұмыс алғашында қолмен жасалса, қазіргі кезде жартылай автоматты тәсілмен жүргізіледі. Мақалада М.Әуезовтің «Абай жолы» романынан алынған мысалдар талдау нысаны болды. Мысалы, *Таң алдында бір ғана сағат мызғығаны болмаса, Абай бұл түнді ұйқысыз өткізді* деген сөйлемдегі сөздер морфологиялық тұрғыдан төмендегіше талданды:

таң {**таң** – зат: жалпы, дара, түбір, дерексіз + Ø жекеше + Ø септік, ілік}

алдында {**алд** – көмекші есім: дара, түбір + Ø жекеше + тәуелдік, 3-жақ + септік, жатыс}

бір {**бір** – сан: дара, түбір, есептік}

ғана {**ғана** – шылау: дара, түбір, демеулік}

сағат {**сағат** – зат: жалпы, дара, түбір, деректі + Ø жекеше + Ø септік, атау}

мызғығаны {**мызғы** – етістік: дара, түбір, негізгі, салт + есімше + тәуелдік, 3-жақ}

болмаса {**болмаса** – шылау: дара, түбір, жалғаулық}

Абай {**Абай** – зат: жалқы, дара, түбір, деректі + Ø жекеше + Ø септік, атау}

бұл {**бұл** – есімдік: дара, түбір, сілтеу + Ø жекеше + Ø септік, атау}

түнді {**түн** – зат: жалпы, дара, түбір, дерексіз + Ø жекеше + септік, табыс}

ұйқысыз {**ұйқысыз** – сын: дара, туынды, қатыстық}

өткізді {**өт** – етістік: дара, түбір, негізгі, салт + функционалды жұрнақ + жедел өткен шақ}

Берілген сөйлемдегі жеке сөздерге жасалған морфологиялық талдаудың негізгі принциптері, қойылған шартты белгілер төмендегіше түсіндіріледі:

1) мәтіндегі сөздер (қара жазумен) жеке бөлек алынады: мысалы, **мызғығаны**;

2) одан кейін ирек жақша қойылып, осы сөздің (қою жазумен) реестрге енетін формасынан кейін «-» белгісі арқылы негізгі грамматикалық сипаттамасы беріледі: **мызғы** – етістік: дара, түбір, негізгі, салт;

3) «+» белгісі қойылып, сөзге жалғанған (жалғануы мүмкін) грамматикалық тұлға (бірнешеу болса, араларына «+» белгісі қойылады) беріледі: «+ есімше + тәуелдік, 3-жақ»;

4) сөз құрамындағы барлық грамматикалық тұлғалар көрсетілген соң ирек жақша жабылады.

Жекелеген сөздерге жасалатын морфологиялық талдау мәнмәтінге байланыссыз жасалады. Дей тұрғанмен, талдау барысында тілдің құрылымдық ерекшеліктері ескерілуі қажет. Осыған байланысты сөздердің морфологиялық құрылымын талдау және оларға белгіленімдер қою барысында кездескен ерекшеліктер төмендегіше көрсетілді:

1) Күрделі зат есімдер.

а) қосарланған зат есімдер: *Құнанбайдың алдында мысықтай үндемей отырып, былай шыға бере, бала-шаға, малы-жалшыны сықпыртып боқтап жүру оның дағдысы болған: малы-жалшыны* {**малшы-жалшы** – зат: жалпы, күрделі, қос, деректі + Ø жекеше + септік, табыс}. Келтірілген мысалдағы *бала-шаға, жалшы-малшы* сөздері бір ғана бірлік ретінде алынып, оларға жалпы грамматикалық сипаттама тұтас берілді;

ә) күрделі атаулар: «*Жәмшид*», «*Сеидбатталгази*», «*Мың бір түн*», *Табары жазған тарих*, «*Жүсіп-Зылихалар*», «*Ләйлі-Мәжнүндер*», «*Көрұғлы*» сияқты хикая дастандар да бар. Мұндағы «*Мың бір түн*» атауы былайша талданды: **Мың бір түн** {**Мың бір түн** – зат: жалқы, күрделі, тіркес, деректі + Ø жекеше + Ø септік, атау}. Бұл жағдайда компьютерлік бағдарламаның мүмкіндігіне байланысты мынадай шартты белгі қойылды, біріншіден, бағдарлама осы үш сөзден тұратын күрделі атауды бір ғана бірлік ретінде тану үшін мәтінде оларды арасындағы бос орынға «_» белгісі қойылды, екіншіден, аталған тіркестегі әрбір сөзге жеке-жеке талдау жасалды. Мысалы, **мың** {**мың** – сан: дара, түбір, есептік}; **бір** {**бір** – сан: дара, түбір, есептік}; **түн** {**түн** – зат:

жалпы, дара, түбір, дерексіз + Ø жекеше + Ø септік, атау). Ал мысалдағы «Жүсіп-Зылихалар», «Ләйлі-Мәжнүндер» атауларын бағдарлама бір ғана бірлік деп тану үшін былайша талданады: **Жүсіп-Зылихалар** {**Жүсіп-Зылиха** – зат: жалқы, күрделі, қос, деректі + көптік + Ø септік, атау}; **Ләйлі-Мәжнүндер** {**Ләйлі-Мәжнүн** – зат: жалқы, күрделі, қос, деректі + көптік + Ø септік, атау};

б) әртүрлі сөз таптарынан тұрып, бір ғана ұғымды білдіретін күрделі есімдер: *Қазір де де, ала_көлеңкеде кәрі әжесі байқап отыр: ән тоқтай қалса, кішкентай қыз қарақаттай қап-қара көздерін ашып алады.* Сөйлемдегі ала көлеңкеде тіркесі күрделі сөз ретінде қарастырылып: **ала_көлеңкеде** {**ала көлеңке** – зат: күрделі, тіркес}, сөз тіркесін құрайтын сөздер жеке-жеке талданды: **ала** {**ала** – сын: түбір, сапалық}, **көлеңке** {**көлеңке** – зат: жалпы, түбір, деректі + Ø жекеше + септік, жатыс}.

2) Күрделі сын есімдер.

а) әртүрлі сөз таптарынан біріккен сын есімдер: *Абай біртүрлі сейіліп, іштей серпіліп алып, тау жотасынан биіктеп, аспанға қарады: біртүрлі* {**біртүрлі** – сын: күрделі, біріккен, қатыстық}; *Қалжыңбас, әзілқой болатын: қалжыңбас* {**қалжыңбас** – сын: күрделі, біріккен, қатыстық};

ә) сын есімдердің қосарлануы: *Ол тап беріп шелекке жетті де, білегін сыбана беріп, кір-кір қолын қолтығына шейін бойлата қымыздың ішіне салып жіберді: кір-кір* {**кір-кір** – сын: күрделі, қос, сапалық};

в) басқа сөз таптарының (сын есім жұрнақтары жалғануы мүмкін) қосарлануы: *Абай сол күні өзінің өзгеше қалпын бақсы-құшнаштай көрген: бақсы-құшнаштай* {**бақсы-құшнаштай** – сын: күрделі, қос, қатыстық}. Бірігу және қосарлану арқылы жасалған сын есімдер бір ғана бірлік ретінде саналып, жалпы грамматикалық сипаттамасы берілді;

г) сын есімдердің (екі сөзден артық болуы да мүмкін) тіркесуі: *Аласа, ал_күрең бетеге даланы да, бөктер, беткейді де тегіс көгертіп, жасарта бастаған.* Мұндағы **ал_күрең** {**ал күрең** – сын: күрделі, тіркес, сапалық} күрделі сын есімі жеке алынып, оны құрайтын сөздер же-

ке-жеке талданады: **ал** {**ал** – сын: түбір, сапалық}, **күрең** {**күрең** – сын: түбір, сапалық};

д) сын есімнің басқа сөз таптарымен (екі сөзден артық болуы да мүмкін) тіркесуі: *Жүзі жұқалау_келген, самай шашы жып-жылтыр, қырлы мұрын, қара_торы келіншек аса бір таза, шапшаң көрінді.* Мұндағы **жұқалау_келген** {**жұқалау келген** – сын: күрделі, тіркес, сапалық} күрделі сын есімі тұтас алынып, оны құрайтын сөздер жеке-жеке талданады: **жұқалау** {**жұқа** – сын: дара, түбір, сапалық + шырай, салыстырмалы}, **келген** {**келген** – етістік: дара, түбір, көмекші + есімше}.

Сын есімдерге морфологиялық талдау жасау барысында шырай категориясы жеке көрсетілді. Мәселен, *Жүзі жұқалау_келген, самай шашы жып-жылтыр, қырлы мұрын, қара_торы келіншек аса бір таза, шапшаң көрінді: жып-жылтыр* {**жып-жылтыр** – сын: күрделі, қос, сапалық, шырай, күшейтпелі}; *Сүйіндіктің аппақ үлкен үйі жылы екен, іші тола тірелген жүк, буулы тең менен сандықтар: аппақ* {**аппақ** – сын: күрделі, біріккен, кіріккен, сапалық, шырай, күшейтпелі}; *Мал, жан, көп ауыл осы орынды қыстауға лайықтап алғанмен, әлі жаңа қыстау сала_алмай, ескілеу тар қораға сыя_алмай, күннің көзі жылынысымен киіз_үйге шығыпты: ескілеу* {**ескі** – сын: дара, түбір, сапалық + шырай, салыстырмалы}.

3) Күрделі сан есімдер.

а) сан есімдердің қосарлануы: *Екі-үш күннен бері бұл өңірде күн бұлыңғыр болған: екі-үш* {**екі-үш** – сан: күрделі, қос, болжалдық}. Қосарланған сан есімдер бір ғана бірлік ретінде алынып, жалпы грамматикалық белгілері берілді.

ә) сан есімдердің тіркесуі: *Он_бес түйені тізген бір көш өзге нөпірдің алдын ала оқшау шықты: он_бес* {**он бес** – сан: күрделі, тіркес, есептік};

б) сан есімнің басқа сөз табымен тіркесуі: *Құдасының аулына Ұлжан отыз_шақты кісімен келген* деген сөйлемдегі *отыз шақты* болжалдық сан есімі ретінде бірге алынып: **отыз_шақты** {**отыз шақты** – сан: күрделі, тіркес, болжалдық}, оны құрайтын сөздер жеке талданды: **отыз** {**отыз** – сан: түбір, есептік}, **шақты** {**шақты** – шылау: түбір, септеулік}.

4) Күрделі есімдіктер.

а) есімдіктердің бірігуі: *Абайдың өз әжесінен басқа ешкімнен естімеген жыры*: **ешкімнен** {**ешкім** – есімдік: күрделі, біріккен, болымсыздық + Ø жекеше + септік, шығыс};

ә) есімдіктердің қосарлануы: *Бірақ Абай өзі әке мінездерін сынағыш болғаннан бері, сол әкеден қорлық көрген елдің қай-қайсысының болса да, сөз ұғар деген адамдарымен сырласа сөйлесе, әке мінездері, ел күйі, көптің мұңы сияқтыны толық ұғынып, кең түсіне бастаған*: **қай-қайсысының** {**қай-қайсы** – есімдік: күрделі, қос, сұрау + Ø жекеше + тәуелдік, 3-жақ + септік, ілік}. Бірігу және қосарлану арқылы жасалған есімдіктер бір ғана бірлік ретінде саналып, жалпы грамматикалық сипаттамасы берілді.

б) есімдіктердің басқа сөз таптарымен тіркесуі: *Өз басы ренжітейін деп те ойлаған емес деген сөйлемде өз басы* тіркесі тұтас алынып: **өз басы** {**өз басы** – есімдік: күрделі, тіркес}, оны құрайтын сөздер жеке талданды: **өз** {**өз** – есімдік: дара, түбір, өздік + Ø жекеше + Ø тәуелдік, 3-жақ + Ø септік, ілік}, **басы** {**басы** – зат: жалпы, дара, түбір, деректі + Ø жекеше + тәуелдік, 3-жақ + Ø септік, атау}.

5) Күрделі үстеулер.

а) әртүрлі сөз таптарының бірігуі: *Үйге кіргелі бірталай болса да, Ұлжан үндемей гана барды тыңдап тұр екен*: **бірталай** {**бірталай** – үстеу: күрделі, біріккен, мөлшер};

ә) әртүрлі сөз таптарының кірігуі: *Бүгін Оспан кеш батарда Смағұлға барып, екеуі бір түрлі тату бола қалып еді*: **бүгін** {**бүгін** – үстеу: күрделі, біріккен, кіріккен, мезгіл};

б) әртүрлі сөз таптарының (негізінен үстеулердің) қайталануы: *Абай қазақша ай атының бәрін сұрап, қайта-қайта қайырып жаттап алды*: **қайта-қайта** {**қайта-қайта** – үстеу: күрделі, қос, қимыл-сын};

в) әртүрлі сөз таптарының қосарлануы: *Ұзын кірпігін ұйқылы-ояу қағып, «тағы айт» дегендей қыңқылдай бастайды*: **ұйқылы-ояу** {**ұйқылы-ояу** – үстеу: күрделі, қос, қимыл-сын};

г) бір сыңары мағынасыз қосарланған үстеулер: *Осындайлық бейуақ шағын Абай әрқашан үнсіз, жым-жырт қана өткереді*: **жым-жырт** {**жым-жырт** – үстеу: күрделі, қос, қимыл-сын}. Бірігу, кірігу, қайталану және қо-

сарлану арқылы жасалған үстеулер бір ғана бірлік ретінде қарастырылып, жалпы грамматикалық сипаттамасы берілді.

д) әртүрлі сөз таптарының тіркесуінен жасалған үстеулер: *Абайды әкелікке еріксіз көндірем дегендей боп, Кұлбадан түні бойы ұйқы бермей жылайды* деген сөйлемдегі *түні бойы* тіркесі тұтас алынып: **түні бойы** {**түні бойы** – үстеу: күрделі, тіркес, мезгіл}, оны құрайтын сөздер жеке талданды: **түн** {**түн** – зат: жалпы, түбір, дерексіз + Ø жекеше + тәуелдік, 3-жақ + Ø септік, атау}, **бой** {**бой** – көмекші есім: түбір + тәуелдік, 3-жақ + Ø септік, атау}.

6) Күрделі етістіктер.

Қазақ тілінде күрделі етістіктер баяндауыштың етістіктің кірігуі, негізгі етістік пен көмекші етістіктің тіркесуі, негізгі етістік пен басқа сөз таптарының тіркесуі және т.б. арқылы жасалады. Күрделі етістіктерге белгіленім қою барысында төмендегідей ерекше жағдайларға көңіл аударылды:

а) екі (немесе одан да көп) етістік түбірдің кірігуі: *Тағы бір кезде Ғабитхан әдейілеп тұрып, Қарашоқыға, Күнкенің ауылына барып, екі қоржын басы кітап әкеледі*: **әкеледі** {**әкел** – етістік: күрделі, біріккен, кіріккен, сабақты + көсемше + Ø рай, ашық + жіктік, 3-жақ, Ø жекеше}. Мұнда әкеледі сөзін морфологиялық тұрғыдан талдағанда оны бағдарламаның жеке бір ғана бірлік ретінде тануына байланысты әкел сөзі түбір ретінде алынды.

ә) етістік түбір мен басқа сөз таптарының кірігуі арқылы жасалған күрделі етістіктер: *Жылда бүйтед деймісің?* сөйлеміндегі *деймісің* сөзі былайша талданды: **деймісің** {**дейсің** **бе** – етістік: күрделі, біріккен, кіріккен, құранды} → **дейсің** {**де** – етістік: дара, түбір, көмекші + көсемше + Ø рай, ашық + жіктік, 2-жақ, жекеше}, **бе** {**бе** – шылау: дара, түбір, демеулік};

б) етістік пен етістіктің тіркесуі арқылы жасалған күрделі етістіктер: *Алғашқы бір жұмадай Ғабитханның тәпсіріне қарап, бұрынғы білген тілдерін қайта құрастырып алды* деген сөйлемдегі құрастырып алды тіркесі бір ғана бірлік ретінде талданды: **құрастырып алды** {**құрастырып ал** – етістік: күрделі, тіркес}, оны құрайтын сөздер жеке алынды: **құрастырып** {**құра** – етістік: туынды, негізгі, сабақты +

функционалды жұрнақ + көсемше}, алды {ал – етістік: түбір, көмекші + жедел өткен шақ + Ø рай, ашық + Ø жіктік, 3-жақ, Ø жекеше};

в) етістік пен басқа сөз таптарының тіркесуі арқылы жасалған күрделі етістіктер: *Көктемге шейін күндіз-түні кітап оқуға салынды*. Қою жазумен берілген күрделі етістік құрамындағы етістік жеке алып талдайтын болсақ, оның мәнмәтіндегі мағынасы ашылмай қалуы немесе басқаша, яғни толық мағыналы етістік ретінде түсінілуі мүмкін. Осыған байланысты аталмыш сөз тіркесінің арасына «_» белгісі қойылып талданды: **оқуға_салынды** {оқуға сал – етістік: күрделі, тіркес, құранды} → **оқуға** {оқу – зат: жалпы, дара, түбір, дерексіз + Ø жекеше + түйық етістік + септік, барыс}, **салынды** {сал – етістік: дара, түбір, көмекші + өздік етіс + жедел өткен шақ + Ø рай, ашық + Ø жіктік, 3-жақ, Ø жекеше}.

II Мәтіндегі сөздерге қойылған морфологиялық белгіленімді компьютерлік бағдарламаға енгізу

Тілдің морфологиялық жүйесін автоматты түрде бөлшектеу, яғни мәтіндегі сөздерге қойылған морфологиялық белгіленімді компьютер жадына енгізу бірнеше сатылы жұмыстардың орындалуы арқылы іске асады. Олар:

1) Сөздердің морфологиялық құрылымы екі бөліктен тұрады: а) түбір сөздер және ә) 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 сөздердің морфологиялық құрамын бөлшектегенде негіз сөздерді ажыратып алуы қажет. Ол үшін компьютер жадына негіз сөздердің тізімі енгізілуі керек.

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

3) Осы жұмыстармен қатар корпустардың қай түрі болсын мәтіндер жиынтығынан

тұратындықтан, тілдің әртүрлі стилін қамтитын электронды пішіндегі мәтіндер енгізіледі. Өйткені жоғарыдағы компьютер жадына енгізілген шарттарға сәйкес бағдарлама жүзеге асырылуы үшін сөздердің морфологиялық құрамын талдауға қажетті мәтіндер керек. Жалпы корпустардың қуаттылығы да осы үлкен көлемдегі мәтіндерді қамтумен де байланысты болып келеді. Көлемді массивті корпустар сөздердің морфологиялық құрамын бөлшектеу жұмысын компьютер көмегімен автоматты түрде атқарады. Ал мұндай бағдарламалар жұмысы үшін компьютер жадына жоғарыда аталған негізсөздер тізімі мен олардың грамматикалық түрленісі болып табылатын қосымша морфемалардың кодталған тізімі енгізілуі қажет.

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

а) Корпусқа қажетті негізсөздер сөз табына толығымен ажыратылуы қажет болғандықтан, сөз табы қойылмаған сөздерге арнайы сөз табына қатысты белгіленімдер қойылады. 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.) деген белгіленім қойылады. Осы сияқты түсіндірме сөздікте мағынасы тіркес құрамында ғана ашылған біраз сөздер кездеседі. Олардың барлығы біздің жұмыс позициямызға икемделіп, мүмкіндігінше белгілі бір сөз табына телініп беріледі.

ә) Реестр сөздерді тізімдеу, өңдеу барысында оларға сөз таптарын қоюда кейбір қате кеткен тұстар кездеседі. Мұндай қате, мысалы, 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 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 тілдің сөз байлығы болғандықтан, олар негізінен тілдегі негізгі және туынды түбір сөздерді ғана қамтуға тиіс.

Қорытынды

Мәтіндегі сөздерге морфологиялық талдау жасау барысында кездесетін мынадай мәселелерді атауға болады: қазақ тіл білімінде кейбір лексикалық бірліктердің сөз таптарына тиістілігі туралы әртүрлі пікірлердің кездесуі; жекелеген грамматикалық тұлғалардың тұрпат межесі мен мазмұн межесінің айқын ажыратылмауы; қазақ тілінің төл ерекшелігіне байланысты мәнмәтін мәселесі; авторлық қолданыс ерекшеліктері; корпус әзірлеуде қолданылатын компьютерлік бағдарламаның мүмкіндіктері.

Тілдің морфологиялық жүйесін автоматты түрде бөлшектеу, яғни мәтіндегі сөздерге қойылған морфологиялық белгіленімді ком-

пьютер жадына енгізу бірнеше сатылы жұмыстардың орындалуы арқылы іске асады. Олар: а) сөздердің морфологиялық құрылымын ажырату (түбір сөз, қосымшалар); ә) компьютер жадына алдын ала дайындалған қосымшалардың тізімі мен шартты белгілерін енгізу; б) компьютер жадына тілдің әртүрлі стилін қамтитын, морфологиялық белгіленім қойылған электронды пішіндегі мәтіндерді енгізу. Бұдан кейін компьютер бағдарламасы арқылы мына жұмыстар орындалады: а) сөз табы қойылмаған кейбір сөздерге белгіленім қою; ә) реестр сөздерді тізімдеу, өңдеу барысында оларға сөз таптарын қоюда кеткен кейбір қателерді қолмен жөндеу; б) реестр сөздер тізіміндегі бір сөз табына қатысты омоним сөздердің бірін ғана қалдыру; в) сөзжасамдық жұрнақтар мен форма тудырушы жұрнақтардың аражігін ажырату.

Мәтіндер корпусын құрастыруға арналған компьютерлік бағдарламаны жасаудың мынадай инженерлік-лингвистикалық және техникалық проблемаларын айтуға болады:

- корпусқа енгізілген мәтіндерді морфологиялық талдау принциптері жаңадан жасалып отырғандықтан, әрі қарай жетілдіруді талап етеді;

- морфологиялық белгіленім қоюда тілдің ішкі заңдылықтарына байланысты шартты түрде алынған белгілер қолданылатындықтан, компьютерлік бағдарлама жасау барысында олар да өзгерістер мен толықтырулар енгізуді қажет етеді.

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

Әдебиеттер тізімі

1. Svartvik J., Quirk R. A corpus of English Conversation. – Lund: Gleerup, 1980.
2. Francis W. Computational Analysis of Present-Day American English. Providence, RI: Brown University Press. – Nelson & Henry Kucera, 1967.

3. Francis W. BROWN CORPUS MANUAL: Manual of Information to Accompany a Standard Corpus of Present-Day Edited American English for Use with Digital Computers. [Electronic resource]. – URL: <http://icame.uib.no/brown/bcm.html> (Accessed: 1.02.2022).

4. Hundt, Marianne. Manual of Information to Accompany the Freiburg-Brown Corpus of American English (FROWN) [Electronic resource]. - URL: <http://khnt.hit.uib.no/icame/manuals/frown/INDEX.HTM> (Accessed: 1.02.2022).

5. Leech C. Extending the possibilities of corpus-based research on English in the twentieth century: A prequel to LOB and FLOB. ICAME Journal. – Geoffrey & Nicholas Smith, 2005. No 29. – P. 83-98.

6. Баранов А.Н. Введение в прикладную лингвистику. Учебное пособие. – М.: Едиториал УРСС, 2003.

7. Захаров В.П. Корпусная лингвистика. Учебн.-метод. пособие. – СПб., 2005.

8. Жұбанов А.Қ. Қазақ тілі мәтіндер корпусының компьютерлік базасын құрудың алғышарттары // Әдеби тіл және қазақ тілінің өміршеңдігі. / Жауапты ред. Қ.Күдерінова. – Алматы: «КИЕ» лингвоелтану инновациялық орталығы, 2009.

9. Жаңабекова А.Ә. Қазақ тілі мәтіндеріне морфологиялық белгіленім қоюдың ғылыми-тәжірибелік негіздері // Научно-практическая конференция «Языки меньшинств в компьютерных технологиях: опыт, задачи и перспективы». – Уфа, 2012.

10. Жубанов А.К. Принципы автоматизации морфологической разметки текстов Национального корпуса казахского языка (НККЯ) // Материалы международной научно-практической конференции «Контрастивные исследования и прикладная лингвистика». – Минск: МГЛУ, 2015.

11. Жаңабекова А., Пирманова К. Технология программы полуавтоматической метаразметки казахского национального корпуса // Актуальные проблемы современной лингвистики и гуманитарных наук: сборник статей XI Всероссийской научно-методической конференции с международным участием. – М.: РУДН, 2019.

12. Койбагаров К.Ч., Мусабаяев Р.Р., Құлманов С.К. Разработка алгоритмов автоматического анализа словоформ казахского языка // «Қазіргі қазақ тіл білімі: қолданбалы лингвистиканың өзекті мәселелері»: халықаралық ғылыми-теориялық конференция материалдарының жинағы. – Алматы, А.Байтұрсынұлы атындағы Тіл білімі институты, 2012.

References

1. Svartvik J., Quirk R. A corpus of English Conversation (Gleerup, Lund, 1980, 284 p.).

2. Francis W. Computational Analysis of Present-Day American English (Brown University Press, Providence, RI; Nelson & Henry Kucera, 1967, 305 p.).

3. Francis W. BROWN CORPUS MANUAL: Manual of Information to Accompany a Standard Corpus of Present-Day Edited American English for Use with Digital Computers [Electron. resource]. Available at: <http://icame.uib.no/brown/bcm.html> (Accessed: 1.02.2022).

4. Hundt, Marianne. Manual of Information to Accompany the Freiburg-Brown Corpus of American English (FROWN). [Electron. resource]. Available at: <http://khnt.hit.uib.no/icame/manuals/frown/INDEX.HTM> (Accessed: 1.02.2022).

5. Leech C. Extending the possibilities of corpus-based research on English in the twentieth century: A prequel to LOB and FLOB. ICAME Journal. – Geoffrey & Nicholas Smith, 29, 83-98 (2005).

6. Baranov A.N. Korpusnaia lingvisika [Corpus linguistics], Vvedenie v prikladnuiu lingvistiku [Introduction to applied linguistics]. Study guide (Editorial URSS, Moscow, 2003).

7. Zaharov V.P. Korpusnaia lingvistika [Corpus Linguistics] Textbook (Saint Petersburg, 2005).

8. Zhubanov A.K. Kazak tili matinder korpusynyn kompiuterlik bazasyn kurudyn algyshtartary [Prerequisites for creating a computer database of the corpus of texts of the Kazakh language], Adebil til zhane kazak tilinin omirshendigi [Artistic language and viability of the Kazakh language]. Responsible editor K.Kuderinova (Linguistic and Local History Center “KIE”, Almaty, 2009).

9. Zhanabekova A.A. Kazak tili matinderine morfologialyk belgilenim qoiudyng gylimi-tajiribelik negizderi [Scientific and practical bases of morphological markup of Kazakh language texts], Nauchno-prakticheskaja konferensia «Jazyki menshinstv v kompúternyh tehnologiah: opyt, zadachi i perspektivy» [Scientific and practical conference «Minority languages in Computer technologies: experience, tasks and prospects»] (Ufa, 2012, P. 42-50).

10. Zhubanov A.K. Prinsipy avtomatizatsii morfologicheskoi razmetki tekstov Natsionalnogo korpusa kazahskogo iazyka (NKKA) [Principles of automation of morphological markup of texts of the National corpus of the Kazakh language (NCKL)], Materialy mejdunarodnoi nauchno-prakticheskoi konferensii «Kontrastivnye issledovaniya i prikladnaia lingvisika» [Materials of the International scientific and practical conference «Contrastive research and Applied Linguistics»] (MSLU, Minsk, 2015, P. 111-119).

11. Zhanabekova A., Pirmanova K. Tehnologiya programmy poluavtomaticheskoi metarazmetki kazahskogo natsionalnogo korpusa [Technology of the program of semi-automatic meta-marking of the Kazakh national Corpus], Aktualnye problemy sovremennoi lingvisiki i gumanitarnykh nauk: sbornik statei XI Vserossiiskoi nauchno-metodicheskoi konferensii s mejdunarodnym uchastiem. [Actual problems of modern linguistics and humanities: collection of articles of the XI All-Russian Scientific and Methodological Conference with international participation] (RUDN, Moscow, 2019, P. 465-474).

12. Koibagarov K.Ch., Musabaev R.R., Kulmanov S.K. Razrabotka algoritmov avtomaticheskogo analiza slovoform kazahskogo iazyka [Development of algorithms for automatic analysis of word forms of the Kazakh language]. «Kazirgi kazak til bilimi: koldanbaly lingvisikanyng ozekti maseleleri»: halykaralyk gylymi-teorialyk konferensia materialdarynyng jinagy [«Modern Kazakh linguistics: actual problems of applied linguistics»: collection of materials of the international scientific and theoretical conference] (A. Baitursynov Institute of Linguistics, Almaty, 2012, P. 83-87).

S.K. Kulmanov, A.A. Zhanabekova, N.M. Ashimbayeva, A.Z. Bisengali, N.K. Shulenbayev, B.K. Kordabay
A. Baitursynov Institute of Linguistics, Almaty, Kazakhstan

Problems of morphological markup of words in corpus texts, and their inclusion in a computer program

Abstract. The article gives a brief overview of the history of corpus creation in linguistics, characteristics of corpus linguistics, theoretical and practical tasks and requirements of morphological markup are indicated.

Morphological markup of words in corpus texts was originally created manually. Explanations of the basic principles of morphological analysis of individual words and markings are given. It is known that morphological analysis is carried out mainly without reference to the context. The article separately highlights various features encountered in the analysis of morphological structures of parts of speech and the placement of morphological markings of words.

Automatic disassembly of the morphological system of the language is carried out by performing several stepwise conditions in the computer memory. These are: 1) identification of the morphological structure of words (single-root word, affixes); 2) entering a list and pre-prepared affixes into the computer's memory; 3) entering electronic format texts of various language styles and containing morphological markings into the computer's memory. Then, with the help of a computer program, the following works are performed: a) marking parts of speech on some words that are not placed; b) in the process of processing registry words, manually correct single errors when placing parts of speech on them; b) leave only one of the homonyms relative to one of the parts of speech in the list of registry words; c) identify differences in word-forming suffixes and formative affixes.

Keywords: corpus, corpus linguistics, text, morphology, conditional marking, markup, computer program

С.К. Кулманов, А.А. Жанабекова, Н.М. Ашимбаева, А.З. Бисенгали, Н.К. Шуленбаев, Б.К. Кордабай

Институт языкознания им. А. Байтурсынова, Алматы, Казахстан

Проблемы морфологической разметки слов в текстах корпуса и их включения в компьютерную программу

Аннотация. В статье даются краткий обзор истории создания корпуса в языкознании, характеристика корпусной лингвистики, указываются теоретические и практические задачи и требования морфологической разметки.

Морфологическая разметка слов в текстах корпуса изначально создавалась вручную. Даны разъяснения по основным принципам морфологического анализа отдельных слов и разметок. Известно, что

морфологический анализ проводится в основном без привязки к контексту. В статье отдельно выделяются различные особенности, встречающиеся при анализе морфологических структур частей речи и размещении морфологических разметок слов.

Автоматическая разборка морфологической системы языка осуществляется за счет выполнения нескольких ступенчатых условий в памяти компьютера. Это: 1) выявление морфологической структуры слов (однокоренное слово, аффиксы); 2) внесение в память компьютера списка и заранее подготовленных условных обозначений аффиксов; 3) внесение в память компьютера текстов электронного формата различных стилей языка, содержащих морфологические разметки. Затем с помощью компьютерной программы выполняются следующие работы: а) проставление разметок частей речи на некоторых словах, где они не указаны; б) в процессе обработки реестровых слов вручную исправляются единичные ошибки в указании их части речи; в) оставляется только одно из слов-омонимов относительно одной из частей речи в списке реестровых слов; г) выявляются различия словообразовательных суффиксов и формообразующих аффиксов.

Ключевые слова: корпус, корпусная лингвистика, текст, морфология, условная пометка, разметка, компьютерная программа.

Авторлар туралы мәлімет:

Құлманов С.Қ. – филология ғылымдарының кандидаты, доцент, Терминология орталығының жетекшісі, А.Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан.

Жаңабекова А.Ә. – филология ғылымдарының докторы, Қолданбалы лингвистика бөлімінің меңгерушісі, А.Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан.

Әшімбаева Н.М. – филология ғылымдарының кандидаты, Терминология орталығының жетекшісі ғылыми қызметкері, А.Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан.

Бисенғали А.З.-Г. – философия ғылымдарының докторы (PhD), Терминология орталығының ғылыми қызметкері, А.Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан.

Шүленбаев Н.Қ. – гуманитарлық ғылымдар магистрі, Терминология орталығының аға ғылыми қызметкері, А.Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан.

Қордабай Б.Қ. – гуманитарлық ғылымдар магистрі, Терминология орталығының кіші ғылыми қызметкері, А.Байтұрсынұлы атындағы Тіл білімі институты, Алматы, Қазақстан

Kulmanov S.K. – Candidate of Philology, Associate Professor, Head of the Center of Terminology, A. Baitursynuly Institute of Linguistics, Almaty, Kazakhstan.

Zhanabekova A.A. – Doctor of Philology, Head of the Department of Applied Linguistics, A. Baitursynuly Institute of Linguistics, Almaty, Kazakhstan.

Ashimbayeva N.M. – Candidate of Philological Sciences, Leading researcher of the Center of Terminology, A. Baitursynuly Institute of Linguistics, Almaty, Kazakhstan.

Bisengali Akmaral Z.-G. – Doctor of Philosophy (PhD), Research associate of the Center of Terminology, A. Baitursynuly Institute of Linguistics, Almaty, Kazakhstan.

Shulenbayev N.K. – Master of Humanities, Senior researcher of the Center of Terminology, A. Baitursynuly Institute of Linguistics, Almaty, Kazakhstan.

Kordabay B.K. – Master of Humanities, Junior researcher of the Center of Terminology, A. Baitursynuly Institute of Linguistics, Almaty, Kazakhstan.