

КОРПУСТЫҚ ЛИНГВИСТИКА:
НЕГІЗГІ ТЕРМИНДЕР МЕН
ТҮСІНІКТЕРДІҢ
ОҚУ СӨЗДІГІ

Алматы
«Қазақ университеті»
2018

ӘОЖ 811.512.122'1

КБЖ 81.2 қаз-2-923

К 60

*Баспаға ал-Фараби атындағы Қазақ ұлттық университеті
филология және әлем тілдері факультетінің
Ғылыми кеңесі және Редакциялық-баспа кеңесі
шешімімен ұсынылған
(№4 хаттама 29 желтоқсан 2017 жыл)*

Пікір жазғандар:

филология ғылымдарының докторы, профессор **Б.А. Ахатова**
филология ғылымдарының докторы, доцент **Ж.Қ. Ибраева**

Құрастырғандар:

Г.Б. Мәдиева, С.Б. Бектемірова, Н.А. Исмаилова

К 60 **Корпустық лингвистика: негізгі терминдер мен түсініктердің оқу сөздігі / құраст.: Г.Б. Мәдиева, С.Б. Бектемірова, Н.А. Исмаилова. – Алматы: Қазақ университеті, 2018. – 40 б.**

ISBN 978-601-04-3269-7

Аталмыш сөздік қазақ тіліндегі корпустық лингвистика бойынша оқу сөздігін құрастыру тәжірибесін ұсынатын бакалавр студенттері, магистранттары, PhD докторанттары және корпустық лингвистика мәселелері бойынша қызығатын барлық қауым үшін оқу мақсатында дайындалды. Сөздікте зерттеу мақсатында қолданатын, корпустық лингвистика мәселесін зерттеуге, корпус құрастыру, сонымен қатар студенттердің өзіндік жұмысын орындауда, магистранттар мен докторанттардың осы пәнге сәйкес сұрақтар бойынша базалық терминдер мен түсініктер қамтылды. Курс бойынша корпустық лингвистика мәселелеріне арналған негізгі терминдер мен түсініктерді меңгеруде әртүрлі тілдердегі ұлттық корпус-тарды құрастырудағы теориялық-қолданбалы мәселелермен айналысатын қазақ-стандық ЖОО үшін жана болып табылатын лингвистикалық бағыттын түсініктік-терминологиялық аппаратты қолдануға және меңгеру құзырлылығын қалыптасты-руға септігін тигізеді.

Автордың редакциясымен шығарылды.

ӘОЖ 811.512.122'1

КБЖ 81.2 қаз-2-923

ISBN 978-601-04-3269-7

© Құраст.: Мәдиева Г.Б., Бектемірова С.Б.,
Исмаилова Н.А., 2018

© Әл-Фараби атындағы ҚазҰУ, 2018

АЛҒЫСӨЗ

Әл-Фараби атындағы Қазақ ұлттық университетінің филология мамандығының бакалавр, магистратура және докторантура бөлімдерінде бірнеше жыл бойына қазіргі тіл білімінде корпустық лингвистиканың мәселелерін белсенді дамушы бағыт ретін қарастыратын *Лингвистикадағы корпустық лингвистика және инновациялық технологиялар*, *Корпустық лингвистика және тілді оқытудағы компьютерлік құралдар*, *Структурализмнен корпустық лингвистикаға дейін*, *Қазақ тілінің Ұлттық корпусы*, *British National corpora*, *Корпустық лингвистиканың теориясы мен практикасы*, *Корпустық лингвистиканың теориялық мәселелері мен корпус жасаудың технологиясы және т.б.* пәндері оқылады.

Бұл пәндердің негізгі мақсаты – тілді менгерудің эмпирикалық тұрғысында корпустық лингвистика негіздерімен және корпустық лингвистика тұжырымдамасы мен әртүрлі тілдердің корпусын құрастыру жөнінде студенттерге, магистранттарға, докторанттарға таныстыру. Бұл пәндердің міндеттеріне әртүрлі құзырлылықтардың қалыптасуы жатады: корпустық технологиялардың негіздерін білуі, арнайы, мамандандырылған және ұлттық корпустармен жұмыс істеуге дағдылануы, пән орнын анықтай алуы, корпустық лингвистика сауалдарын лингвистикалық циклдің пәндері арасында қарастыруы, корпустың теориялық және практикалық маңыздылығын лингвистикалық зерттеулердің дәлелді лингвистика аясында анықтай алуы, корпустық лингвистиканың метатілін білуі. Осыған сәйкес оқу сөздігі материалының нәтижесінде оның метатілінің құраушы компоненті ретінде корпустық лингвистиканың терминологиялық аппаратын жүйелі түрде сипаттауды қажет етеді. Аталмыш оқу сөздігін құрастыруда фундаменталды теориялық негізіне шетелдік және қазақстандық авторлардың еңбектері арқау болды:

Atkins S., Clear J., Ostler N., Гвишиани Н.Б., В.П. Захаров, М.А. Даниэл, Ludeling A., Kyto M., В.В. Плу́нган, Е. Рахилин, Сичинав Д.В., А.У. Тукеев және т.б.

Қазіргі зерттеулердің функционалды бағыты нақты сөйлеу әрекетін сипаттау мақсатында, белгілі бір контекстте әртүрлі тілдік бірліктерді қолдануда тілді/тілдерді әртүрлі коммуникативті жағдайларда зерттеудің қажеттілігін туғызды. Бұл айтарлықтай күрделі мәселе кең ауқымды иллюстративті материал негізінде қарастырылуы тиіс, яғни әртүрлі тілдердің корпустары мен интернет-ресурстарының жаңа ақпараттық технологиялар көмегімен алуға болады. Мұндай фактілер компьютерлік технологиялардың қарқынды дамуымен және олардың әлеуеттің барлық салаларында, сонымен қатар қазақ тілінің арнайы зерттеулері мен оқытулуында кең қолданылуымен байланысты. Осыған сәйкес Қазақ тілінің Ұлттық корпусын дайындаудың қажеттілігі туды.

Корпустың маңыздылығы турасында В.М. Плу́нган¹, орыс тілінің корпусын құрастырушылардың бірі, мынадай анықтама береді: «Тілдің Ұлттық корпусы атауына сай келетін корпуста материалдардың қамтылуы толық әрі көлемі жағынан үлкен болса, бұл өте тиімді және пайдалы құрал (тар саладағы мамандармен қатар басқа да сала мамандары да қолдана алатын) болары анық».

Кейбір тілдердің корпусы – бұл, бір жағынан, берілген тілдегі электронды формадағы жәе ғылыми аппаратпен жабдықталған мәтіндердің жинағы. Бірақ бұл тек қана электронды кітапхана сияқты мәтіндерді жинақтау ғана емес, сонымен қатар ол мәтіндердің деректер базасы болып табылады. Сөз туралы әртүрлі ақпаратты беруде, оның әртүрлі жанрда, стильде, хронологиялық кезеңде, нақты автордың мәтіндері материалдары контекстінде көрінетін сөз турасында тиімді

¹ Плу́нган В.М. Общество в зеркале языка //ж. Отечественные записки, М., 2005, № 2; Зачем нужен Национальный корпус русского языка? Неформальное введение.

және жылдам ақпаратты алуға мүмкіндік береді».² Бұл мәселе шетел лингвистикасында шешімін табуда (Ұлыбритания, АҚШ, Франция, Германия, Жапония, Румыния, Чехия және т.б.): 60 жылдардан бастап бұл елдерде, сәйкесінше, ұлттық корпустар жасала бастады: Браун университетінің ағылшын тілі корпусы, Британ ұлттық корпусы, Ұлттық американдық корпус, француз тілінің корпусы және т.б. толық масштабты және көп функционалды ұлттық тілдердің фондтарын жасау басты міндет болып табылады. Нәтижесінде өзінің дамуын мәтін корпусын қолдану және оны жасаумен айналысатын корпусық лингвистика (XX ғ. 60 ж.) ретінде ғылыми-қолданбалы бағытын айқындады. Корпусық лингвистиканың пайда болуы келесі міндеттермен, яғни ауызша және жазбаша күйінде берілетін барлық нақты тілдік деректерді жинақтауға болатын ұлттық тілдің мәтіндерінің толық фондын құраумен байланыста болды. Мұндай міндет айтарлықтай қиындыққа ие болды, есептегіш техниканың дамуына әсер етті (XX ғ. 80 ж.). Құрастырылған компьютерлік фондтар 600 мыңнан 100 миллионнан аса сөзқолданысын қамтыды. Бірақ бұл корпустардың кеңеюі мен толықтырылуының тек қана алғашқы кезеңі ғана болып табылады. Қазақстандық

² Наука о корпусах (а таковая давно уже существует и называется «корпусная лингвистика») – это прежде всего наука о том, как сделать хорошую разметку корпуса». По его практическому замечанию, изменения в русском языке наглядно прослеживаются на основе корпуса. Так, развитие языка, рано или поздно должно сказаться на слабых зонах русской грамматической системы. Грамматика пока держится... наиболее уязвимо в русском языке склонение. Тексты Корпуса показывают заметный рост — особенно в устной речи — именительного падежа во всех позициях. Так, числительные уже практически перестали склоняться (совсем не трудно услышать что-нибудь вроде остался без триста двадцать пять тысяч рублей). На наших глазах лишаются форм склонения и географические названия типа Бородино: в повседневной речи фактически единственным вариантом становится рядом с Бородино, живу в Бородино и т. п. Не желают склоняться, что характерно, и новые заимствования, а заимствования никогда не подчиняются тем законам, которые в самом принимающем языке действуют слабо. Ждет ли русский язык судьба болгарского — потерявшего все падежи, функции которых взяли на себя предлоги, — конечно, ни один серьезный специалист предсказывать не берется. Но «что-то может произойти». Теперь у нас есть Корпус. Так что получить ответ, может быть, будет проще [www.strana-oz.ru].

лингвистика осы бағытта сондай-ақ белгілі бір кадамдар жасалуда: қазақ тілі мәтінің мәліметтер қорында әртүрлі лексикографиялық қайнаркөздер құрылды. Сонымен қатар, қазір интернет сайттарда қазақ тіліндегі мәтіндердің көп болуы парсинг үшін бірден бір мүмкіндік – автоматтандырылған үдеріс «материалдарды бір сайттан (немесе бірнеше) екінші сайтқа көшіру (немесе мәліметтер қорына жеке сайтқа салу немесе сағу үшін көшіру)» [<https://ruseo.net/chto-takoe-parsing-ili-sparsit-t9447.html>]. Алайда, корпустық лингвистиканы дамыту және ұлттық корпусты (копустарды) жасау – қазіргі қазақ тіл білімінің келешегі. Өз кезінде бір сұхбатта былай делінген: «қазіргі өндірістік интеграциялық платформа қорында тілдер комитетінің аппаратты-бағдарламалық жиынтығы жасалынатын болды, Қазақстанның мемлекеттік және басқа да халықтар тілдерін қолдау бойынша кең қызмет топтарын және сервисті көрсету үшін пайдаланушылар үшін «кірудің бірегей нүктесі» принципін қамтамасыз етуге міндеттелген мықты республикалық бірегей орталық негізделді» [<http://www.zakon.kz/87167-nacionalnyj-korpus-nacii-intervju.html>].

Қомитет қызметінің негізгі бағытының бірі қазақ тілінің ұлттық корпусын жасау базасында мемлекеттік тілдің қазіргі жағдайын сараптау жүйесін құру. Үкіметтің қолдауымен Тілдер комитеті «Қазақстан Республикасының мемлекеттік тілін дамыту және ақпараттық жүйені қолдауды жасау» үш жылдық бағдарламаны жүзеге асыруға кірісті. Компьютерлік бағдарлама кешенін жасау жобалануда, басқаша айтқанда, қазақ тіліндегі кең масштабтағы инфраструктура, софт, қазақ тілін интернет-кеңістігіне енгізу, бұрынғы жылдар мен ақпараттық технологияның қазіргі дамуы есебімен Қазақстанның мемлекеттік тілін он-лайндық сервис сериясын, тұтас және ықпалды жүйесін дамыту» [www.bimash.kz].

Бұл тапсырмалардың ойдағыдай енгізіліп жатқанын қуана айтса болады, өмірде көптеген интернет-сөздіктер пайда болды, қазақ тілін меңгеру және қазақ тілінде әртүрлі

ақпараттар берілген әртүрлі веб-сайттар жасалды. Қазіргі уақытта, келесідей сайттар, платформалар, порталдар, блогтар, білім беретін сайттар жасалды және іске асырылды: *Уикипедия* (Википедия), *Викибілім*, *Сөздік.кз*, *Тұған тіл* және т.б.; көңіл көтеретін сайттар: *Мәссаған*, *Қазақша KBH*, *ІргеТас* және т.б.; танымдық сайттар: *www.tanym.kz* және т.б.; діни бағыттағы сайттар: *Дін ислам*, *діни сайты* және т.б. Біз қазақ тілін тасымалдаушыларға арналған және қаза тілі мен басқа халықтар тілі мен мәдениетіне қызығатын және меңгеретіндерге арналған интернет-ресурстардың пайда болуы мен дамуына куәгер бола аламыз, себебі осындай сайттар қазақ тілінде әртүрлі ақпарат береді.

Осындай сайттардың, веб-парақшалардың, платформалардың, блогтардың толығыуы мен өзіндік сипаты: қарапайым мазмұннан маңызды, өзіндік дизайн түрімен ерекшеленген ақпараттық толықтыруға дейінгі сипатымен ерекшеленеді. Қазақ тілінің әрқилы корпустары да жасалынды, ол өз кезегінде олардың жақсы динамикасы мен дамуын көрсетеді. Осыған байланысты корпустық лингвистиканың терминологиясын меңгеру маңызды. Жобалық режимдегі оқыту мақсатында жасалған осы сөздік игеру үшін қажетті терминологиялық аппараттың ядролық қорын құрайтын, корпустық лингвистиканың тиісті терминдерін жүйелі беру тапсырмасын көздейді.

Оқу сөздігіндегі материалдар оқыту түріне қатысты сөздіктердегі терминдерді лексикографиялау дәстүрлі принципке сәйкес жасалынды. Корпустық лингвистика мәселесіне қызығушылар іздеген терминін табу, дефиницияны білу, өз білімін толықтыру, зерттеу жұмысына қажетті терминологиялық аппаратты жасау мүмкіндігі бар. Сөздіктің соңында корпустарды сипаттауда пайдаланылатын термин және терминдік сөзтіркестерінің тізімі беріледі, сондай-ақ сөздікте берілген терминдер таңдалып алынған, ғылыми дереккөздер, әдебиет тізімдері мен қалыптасқан корпустар тізімін келтіреді.

МАЗМҰНЫ

Алғы сөз.....	3
Терминдер сөздігі	8
Корпустық лингвистиканың негізгі терминологиялық тіркестерінің тізімі	29
Қолданыстағы ұлттық корпустардың тізімі	30
Библиографиялық тізім	35

Оқу басылымы

КОРПУСТЫҚ ЛИНГВИСТИКА: НЕГІЗГІ ТЕРМИНДЕР МЕН ТҮСІНІКТЕРДІҢ ОҚУ СӨЗДІГІ

Құрастырғандар:

Мәдиева Гүлмира Баянжанқызы
Бектемірова Сәуле Бекмұхамедқызы
Исмайлова Нұргүл Ахметхановна

Компьютерде беттеген және
мұқабасын безендірген *Ғ. Қалиева*

ИБ №11797

Басуға 19.03.2018 жылы кол қойылды. Пішімі 60x84 1/16.
Көлемі 2,5 б.т. Офсетті қағаз. Сандық басылым. Тапсырыс №1139.

Таралымы 70 дана. Бағасы келісімді.

Әл-Фараби атындағы Қазақ ұлттық университетінің
«Қазақ университеті» баспа үйі.

050040, Алматы қаласы, әл-Фараби даңғылы, 71.

«Қазақ университеті» баспа үйі баспаханасында басылды.