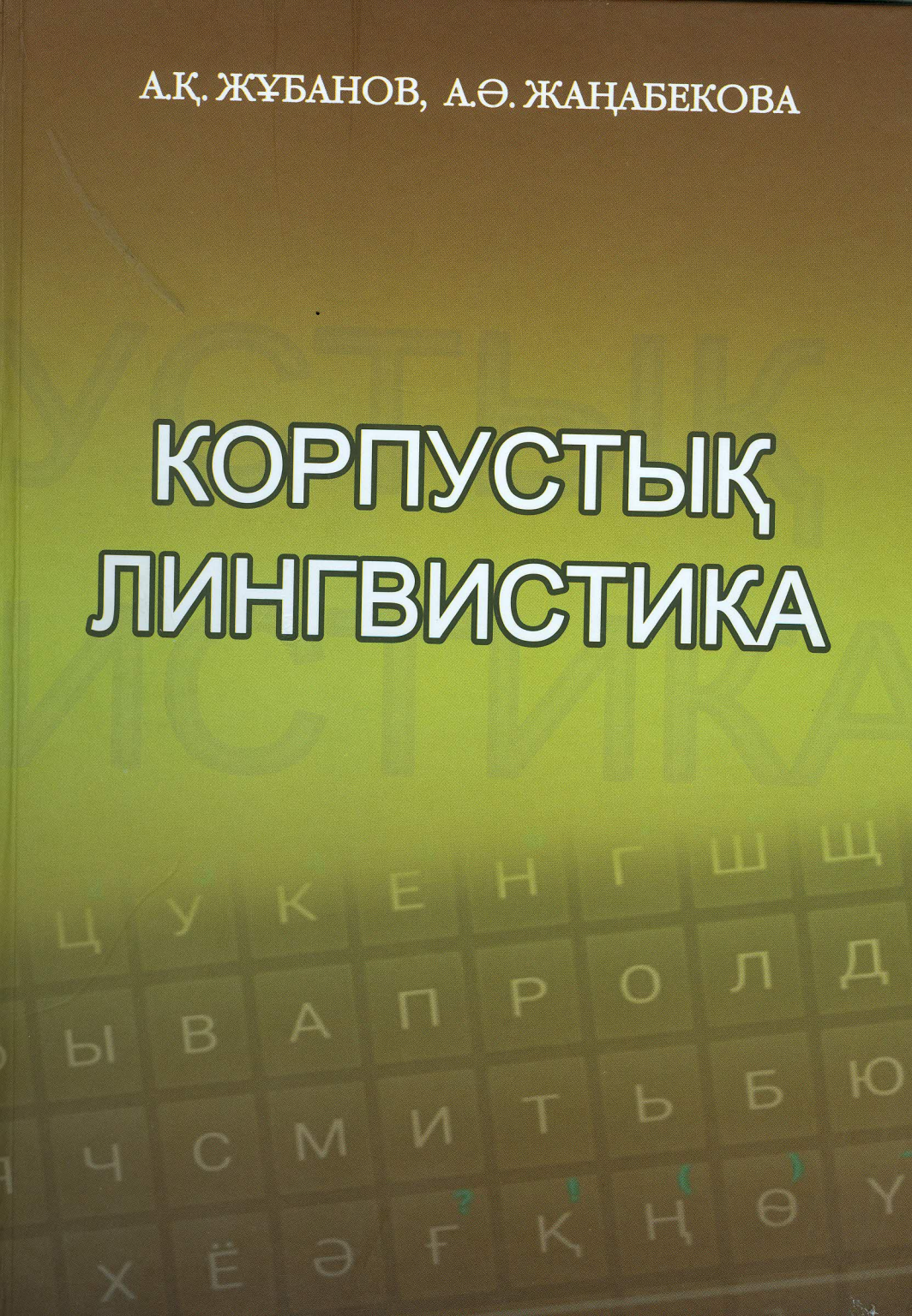
А.Қ. ЖҰБАНОВ, А.Ә. ЖАҢАБЕКОВА

КОРПУСТЫҚ ЛИНГВИСТИКА



ӘОЖ 811.512.122
КБЖ 81.2 Қаз
Ж 80

А.Байтұрсынұлы атындағы Тіл білімі институтының
Ғылыми кеңесінде мақұлданған

Жауапты редактор: филология ғылымдарының докторы, профессор **М. Малбақов**

Пікір жазғандар: филология ғылымдарының докторы, профессор **Н. Уәли**

филология ғылымдарының кандидаты **А. Фазылжанова**

Жұбанов А.Қ., Жаңабекова А.Ә.
Ж 80 Корпустық лингвистика/ А.Қ. Жұбанов, А.Ә. Жаңабекова. – Алматы: «Қазақ тілі» баспасы, 2017. – 336 б.

ISBN 978-601-7293-43-7

Оқырман назарына ұсынылып отырған оқу құралы корпустық лингвистиканың негізгі ұғымдарымен таныстыра отырып, корпустық технологиялардың негіздерін меңгеруге, корпустармен жұмыс істеу дағдысын қалыптастыруға және корпустық лингвистика пәнінің ақпараттық технологиялар қатарынан алатын орнын айқындауға мүмкіндік тудырады.

Мұндай зерттеу жұмысы қазақ тіл білімінде бұрын-соңды қолға алынбағандықтан, біріншіден, зерттеу нәтижесі корпустық лингвистика бойынша алғашқы теориялық тәжірибелердің негізін қаласа, екіншіден, қазақ тілінің ұлттық корпусын құрастыруға бағыт-бағдар береді.

Оқу құралы филология мамандықтарындағы студенттерге, магистранттар мен PhD докоранттарға, компьютерлік лингвистика, корпустық лингвистика пәнінен дәріс оқитын оқытушыларға және ақпараттық технология саласындағы мамандарға арналған, сонымен қатар корпусты құрастырушылар мен оны пайдаланушыларға әдістемелік құрал бола алады.

ӘОЖ 811.512.122
КБЖ 81.2 Қаз

ISBN 978-601-7293-43-7

АЛҒЫ СӨЗ

Оқу құралының мақсаты – оқырманды корпустық лингвистиканың негізгі ұғымдарымен, корпустармен жұмыс істеу жолдарымен, әдіс-тәсілдерімен таныстыру, корпустық технологиялар негіздерін меңгеруге мүмкіндік жасау және бұл пәннің басқа да ақпараттық технологиялар қатарынан алатын орнын айқындап беру.

Мақсатқа сай, оқу құралында мынадай міндеттер көзделеді: оқырмандарды корпустық зерттеулер тарихымен таныстыру; оқырмандарды корпустық лингвистиканың тілдік және программалық құралдарын зерттеуге бағыттау; оқырмандарды корпустық лингвистиканың программалық құралдарымен және ақпараттық ресурстарымен жұмыс істеу дағдысын қалыптастыру; оқырмандарды корпустық деректерге негізделген зерттеулермен таныстыру және нақты тілдік материалдармен корпустар арқылы жұмыс істеудің тез және тиімді жүзеге асатындығын көрсету.

Оқу құралы төрт бөлімнен тұрады. «Корпустық лингвистикаға кіріспе» атты *бірінші бөлімде* корпустық лингвистиканың негізгі ұғымдары, терминдері, тіл білімінің жеке саласы ретінде қалыптасу тарихы, тіл білімінің басқа салаларымен байланысы және қолданыстағы корпустар түрлері туралы мәлімет берілген.

«Корпустар құрастыру мәселелері» деп аталатын *екінші бөлімде* лингвистикалық белгіленім (разметка) түрлері, сонымен қатар корпустарды жобалауға, іріктеуге және тілдік материалдарды өңдеуге, белгіленім тәсілдеріне, стандарттауға қатысты технологиялық үдерістер сипатталады.

«Корпустарды пайдалану» атты *үшінші бөлімде* корпустардан қажетті мәліметтерді іздестіруді қамтамасыз ететін корпустық менеджерлер мен корпус материалдары негізінде жүргізілген әртүрлі лингвистикалық зерттеулердің (лексикография, грамматика, дискурс) нәтижелері сипатталады.

«Қазақ тілінің Ұлттық корпусын құрастыру тәжірибесінен» деп аталатын *төртінші бөлімде* Қазақ тілінің ұлттық корпусын құрастырудағы алғашқы тәжірибе нәтижесі, морфологиялық белгіленімнің лингвистикалық әзірлемесі, мәтіндерге морфология-

лық белгіленім қоюды автоматтандыру, аннотацияланған корпустардың ғылыми-практикалық маңызы туралы баяндалады.

Бұл оқу құралын жазуға авторлардың «Компьютерлік лингвистика» мен «Корпустық лингвистика» пәндері бойынша оқыған дәрістері және «Қазақ тілінің Ұлттық корпусын құрастыруға» қатысты теориялық зерттеулерінің (мақалаларының) материалдары мен корпус құрастыру тәжірибесі негіз болды. Сонымен бірге, аталған пән бойынша орыс тілінде жарық көрген филолог-ғалымдар, доцент Виктор Павлович Захаров (Корпусная лингвистика: Учебно-метод. пособие. – СПб., 2005) пен профессор Светлана Юрьевна Богдановамен бірігіп жазған (Корпусная лингвистика: учебник для студентов гуманитарных вузов. – Иркутск: ИГЛУ, 2011. – 161 с.) еңбектері және т.б. осы сала бойынша жазылған ғылыми еңбектер материалдары да негіз болды. Әсіресе, Қазақ тілінің ұлттық корпусы жасалу үстінде болғандықтан, үшінші тарауда ағылшын тілі негізінде жасаған В.Захаровтың корпусты пайдалануға қатысты материалдары түпнұсқа күйінде алынды.

Сонымен бірге осы оқу құралының авторының бірі А.Қ.Жұбановтың қазақ тілін зерттеуге арналған жаңа сала «Қолданбалы лингвистика» бойынша жүргізген ғылыми зерттеу жұмыстары да ұсынып отырған кітапқа теориялық және практикалық жағынан негіз болды деуге болады.

Атап айтқанда, автордың «Қолданбалы тіл білімінің мәселелері» атты монографиялық жұмысы ҚР БжҒМ Ә.Байтұрсынұлы атындағы Тіл білімі институты, ҚР Мәдениет және Ақпарат министрлігі Тіл комитетінің қолдауымен және «Арыс» қорының «Қазақ ғылымының озық үлгілері» атты мегажобасының аясында жарық көрді (Алматы: «Арыс» баспасы, 2008. – 640 бет). Аталған монографияның «Қолданбалы лингвистика: қазақ тілінің статистикасы» деп аталын бірінші бөлімінде қазақ тілі мәтіндеріндегі сөзтұлғалардың лексика-морфологиялық құрылымын статистикалық әдіспен зерттеу үшін шартты белгі-кодты сәйкестендіру (белгіленім) бағдарламасы көрініс тапқан. Мәтін бірліктеріне тән белгі-кодтарды сәйкестендіру «ұялы принципке» негізделіп, негізгі сөз таптары ретінде қазақ тілінің зат есім, етістік, сын есім сөздері тиісті белгі-кодқа сәйкестік кестелер арқылы берілген.

Сол сияқты, аталған монографияның ең өзекті саналатын екінші бөлімі «Основные принципы формализации содержания казахского текста» деп аталып, мұнда «Қазақ тілінің атаушы сөз таптарын семантикалық топтастыру» мәселесі және қазақ мәтінінің семантикалық-мәнді бірліктері мен қисындасу ережелерін тандап алу мәселесі де сөз болады.

А.Қ.Жұбановтың ғылыми зерттеулері қазақ тілінің мәтіндік корпусын жасау мәселесіне теориялық тұрғыдан да, практикалық тұрғыдан да құнды негіз болмақ. Оқу құралын жазу барысында пайдаланылған автордың мұндай ғылыми жұмыстары қазақ тіл білімінде бұрын-соңды қаралмағандықтан, біріншіден, зерттеу нәтижесі корпустық лингвистика бойынша алғашқы теориялық тәжірибелердің негізін қаласа, екіншіден, тілдік материалды модельдеу нәтижесінде көптеген тілдік ақпараттарды алып, оларды тіл білімінің кез келген саласында нақты тілдік дерек ретінде пайдалануға, ұстануға мүмкіндік жасайды.

Қазақ тілінің мәтіндік корпусын зерттеуге арналған бұл жұмыстың нәтижелерін жоғарғы оқу орындарында оқытылатын «Қазақ тілінің функционалды грамматикасы», «Компьютерлік лингвистика», «Корпустық лингвистика», «Морфология», «Мәтін лингвистикасы» т.б. салалар бойынша дәрістер мен семинар сабақтарында пайдалануға болады.

Филолгия ғылымдарының докторы,
профессор
Мырзаберген МАЛБАҚОВ

МАЗМҰНЫ

3. КОРПУСТАРДЫ ПАЙДАЛАНУ

4. ҚАЗАҚ ТІЛІНІҢ ҰЛТТЫҚ КОРПУСЫН ҚҰРАСТЫРУ ТӘЖІРИБЕСІНЕН