

МРНТИ 16.31.02

**Э.Д. Сүлейменова**

әл-Фараби атындағы Қазақ Ұлттық университетінің профессоры, филология  
ғылымдарының докторы  
Алматы қаласы, Қазақстан

### **ҚАЗАҚ ТІЛІ ҮШІН ҰЛТТЫҚ КОРПУС КЕРЕК ПЕ?**

**Аннотация.** Мақалада Ұлттық корпус мәселелері қарастырылады. Сондай-ақ Ұлттық корпусның қажеттілігі мен оның жасалуы жайында айтылады.

**Тірек сөздер:** Ұлттық корпус, репрезентативтілік, метамәтін, анықтамалық-ақпараттық жүйе.

**Э.Д. Сүлейменова**

Профессор Казахского Национального университета имени аль-Фараби,  
доктор филологических наук,  
Алматы, Казахстан

### **НУЖЕН ЛИ НАЦИОНАЛЬНЫЙ КОРПУС ДЛЯ КАЗАХСКОГО ЯЗЫКА?**

**Аннотация.** В статье рассматриваются проблемы, задачи и отличительные черты Национального корпуса. Обосновывается необходимость и актуальность его создания.

**Ключевые слова:** Национальный корпус, репрезентативность, метатекст, справочно-информационная система.

**E. D. Suleimenova**

Professor of the Al-Farabi Kazakh National University,  
Doctor of Philology,  
Almaty, Kazakhstan

### **IS THE NATIONAL CORPUS NECESSARY FOR THE KAZAKH LANGUAGE?**

**Annotation.** The article deals with the problems, tasks and distinctive features of the National Corpus. The necessity and relevance of its creation are justified.

**Keywords:** National corpus, representativeness, metatext, reference and information system.

#### **Ұлттық тіл корпусы деген не?**

Корпус – бұл мәтіндерді электронды түрде жинауға негізделген ақпараттық-анықтамалық жүйе. Анықтамадан айқындалып тұрғандай, біріншіден, корпус – бұл мәтіндерді электронды (цифрлендірілген) түрде жинау; екіншіден, корпус анықтамалық-ақпараттық жүйе ретінде қолданыла алады. Алайда мұндай анықтама тым жалпылама сипат алып, ақпараттық-анықтамалық жүйе деген не және оның көмегімен қандай деректер (ақпарат, анықтама) алуға болады, сондай-ақ неге корпус ұлттық деп аталады деген сұрақтарға жауап ашық күйінде қалып қояды.

Мәтіндерді электронды жинаудың әрқандай түрлері белгілі, мысалы, электронды кітапхана, білімдік ресурстар, түрлі мамандықтар бойынша кітапханалар, энциклопедиялар мен әртүрлі сөздіктер т.б. Корпустың мәтіндерді мұндай виртуалдық жинаудан басты ерекшеліктері төмендегідей:

– Корпуста цифрлендірілген мәтіндерді жинаудың мөлшері, әдетте виртуалдық кітапханалар көлемінен айтарлықтай асып түседі, миллионға, кейде миллиард сөзқолданысқа дейін жететін ұлттық корпустар бар (айталық, ағылшын тілінің Оксфорд корпусы қазіргі таңда 2 млрд сөзқолданысқа бейімделген). Қазіргі таңда ұлттық деп белгілі бір елдің тілін тұтас сипаттайтын ең үлкен көрнекті корпустар аталып жүр (мысалы, америкалық, австралиялық ағылшын тілінен бөлек британ ұлттық варианты үшін жасалған Британ ұлттық корпусы – British National Corpus, BNC <http://www.natcorp.ox.ac.uk/>).

– Корпус көрнектілігімен /шамаластығымен (теңдестірілгендігімен, тіл дамуының әртүрлі кезеңдеріндегі жазбаша және ауызша мәтіндердің барлық типтерінің тең көлемділігімен) және репрезентативтілігімен (тілдің өмір сүруінің қазіргі немесе оның өмір сүруінің барлық кезеңдеріндегі мәтіндердің саналуан жанрлары, стильдері, аумақтық және әлеуметтік варианттары т.б. жазбаша немесе ауызша түрде беріледі, демек корпус үшін көркем шығармалар да, публицистика да, ауызша сөйлеу жазбалары да, іскери қағаздар да, жарнама да – тілді зерттеу және тілді оқыту тұрғысынан пайдалының барлығы маңызды болып табылады) сипатталады.

– Цифрлендірілген мәтіндердің базасы корпус белгісінің көмегімен арнайы өңделеді, ал электронды кітапханалар мұндай өңдеуден өтпейді. Дәл осы белгі корпусты түбегейлі тілдің қолданылуы мен өзгеруі жайлы ақпараттың жаңа және объективті дереккөзіне айналдырады. Белгі (аннотация, код) – мәтіндердің және олардың бірліктеріне тән қасиеттер жайлы арнайы лингвистикалық ақпарат беретін корпустың негізгі сипаттамасы. Корпустың белгісі дәлірек және әрқилырақ болған сайын оның әлеуметтік, ғылыми және оқыту құндылығы жоғарырақ болады. Кейбір корпустар азғана белгімен қанағаттанса, кейбіреулерінде белгі ұдайы жетілдіріліп отырады, мысалы, Орыс тілінің ұлттық корпусы (<http://www.ruscorpora.ru/>) метамәтіндік (мәтіндерге авторы, тақырыбы, жанры, жазылған уақыты және орны т.б. жайлы атрибуттарды қосып жазу, олар сұрыптау өлшемдерін беруге мүмкіндік береді, мысалы, ер адам жазған мәтіндер, нақты бір кезеңде жазылған мәтіндер, көбелек жайлы жазылған мәтіндер және т.б., сондай-ақ корпустың шағын бөліктерімен жұмыс істеуге жағдай жасайды), морфологиялық (омонимияны алу, әрбір бірлікке арнайы жасалған таңбалар жүйесінің көмегімен морфологиялық белгілерді қосып жазу, мысалы, сөз таптарына тиесілілігі, сөзформасының грамматикалық белгілері (септеу, көптік, вид, шак, рай және т.б.), акценттік (корпус бірлігіне екпін және басқа просодиялық сипаттамаларды қосып жазу), лексика-семантикалық (бірлікке семантикалық белгілерді қосып жазу, мысалы, 'лицо', 'вещество', 'пространство', 'скорость', 'движение', 'обладание', 'свойство человека', 'отглагольное имя' және т.б., сөйтіп бір бірлік бірнеше кластың фасеттік сипаттамаларын алуы мүмкін), синтаксистік белгілерге (тәуелдік ағашы, синтаксистік қатынас атаулары және т.б.) негізделген.

Мәтіндердің белгісі интерактивті режимде бірнеше кезеңде автоматты түрде немесе жартылай автоматты түрде арнайы құрылған программалардың көмегімен және толықтырылып отыратын жұмыстық сөздіктерді қолдана отырып жүзеге асырыла алады.

Арнайы құрылған белгімен жасақталған көрнектілік, шамаластық және репрезентативтілік – барлығы бірге – ұлттық тіл корпусының негізгі қасиеттерін құрайды. *Ұлттық корпус* – қандай да бір елдің тілінде бар барлық жазбаша және ауызша дискурстарды (жарнамадан бастап көркем әдебиет мәтініне дейін) толық, теңбе-тең және шамалас көрсететін, тілдің нақты қолданылуы мен өзгеруі жайлы мәліметтердің (оның ішінде статистикалық) түбегейлі жаңа дереккөзі болып қызмет ететін көлемі жағынан ең үлкен корпус.

### **Ұлттық тіл корпусы кім үшін жасалады?**

Ұлттық корпустардың негізгі тұтынушылары/пайдаланушылар төмендегілер болып табылады: а) әртүрлі мамандану саласындағы зерттеуші тіл мамандары, ең алдымен, корпустық лингвистика мамандары, теоретиктер және тіл тарихшылары, әлеуметтік және психоллингвистика, лингвистикалық сараптама мамандары және т.б.; б) әдебиеттанушылар, тарихшылар, журналистер, саясаттанушылар, әлеуметтанушылар, яғни іс жүзінде гуманитарлық білім саласының барлық өкілдері; в) берілген тілді бірінші және екінші тіл ретінде меңгеру саласының мамандары: біздің заманымызда көптеген оқулықтар мен оқу бағдарламалары барған сайын корпуста сүйеніп отырады, себебі оқулықтарға, мамандардың басым көпшілігінің есептеуінше, қандай да бір дискурста қолданылатын тілдік жиілік құбылыстарды ғана енгізу керек; г) мектеп оқушылары, студенттер, мұғалімдер, оқытушылар, редакторлар, жазушылар – таныс/танымас емес сөздің немесе грамматикалық форманың қолданылу ерекшелігін тез әрі тиімді тексеруді қажет ететіндердің барлығы; д) сөздіктер, корпустар немесе басқа да электронды мәтін жинағын жасаушы программалармен қамтамасыз ететіндер.

Ұлттық корпус кәсібіне немесе айналысатын ісіне байланысты тілдің қолданылуы мен құрылымының ерекшелігін түсінуге ұмтылғысы келетіндердің барлығына қажет.

### **Тілдердің Ұлттық корпусы не үшін жасалады?**

Бұл сұраққа жауап беру оңай емес, ол корпустың мәнінде жатыр.

Біріншіден, Ұлттық тіл корпусы – бұл тіл мамандары мен лексикографтардың көптеген өкілдерінің еңбегімен жасалған сөздіктер мен грамматикаларды қоса алғандағы мүмкіндіктерінен асып түсетін тіл туралы тамаша, бірегей, жаңаша анықтама. Өзінің мағынасы бойынша тілді және оның дамуын тану үшін корпусты А.Е.Кибриктің айтуынша микроскоп, телескоп, рентген аппараты, томография және т.т. ойлап табумен салғастыруға болады. Ұлттық корпустар пайда болғанға дейін тіл туралы деректер лексикография, грамматика және лексикология мамандарының көп жылғы табанды еңбектерінің арқасында жиналды. Корпус сұраныс бойынша болар-болмас секундтардың ішінде тіл туралы жан-жақты толық ақпарат береді. Сөйтіп корпус эмпирикалық базаны кеңейтеді және лингвистикалық зерттеулерге айрықша түсіндірме күш беріп, тың жаңа міндеттер қоюға, тілдің өмір сүруінің неғұрлым дәлірек бейнесін айқындауға, құбылыстардың іс жүзінде тілде қолданылу дәрежесін анықтап дәлелдеуге мүмкіндік береді.

Екіншіден, Ұлттық корпус мәнмәтіндердің және қандай да бір сөздің, оның лексика-семантикалық/графикалық/орфографиялық вариантының, грамматикалық формасының және грамматикалық мағынасының, сөз тіркесі, сөйлем түрлерінің бірнеше ғасырлар бойы немесе жекелеген жанрларда, нақты авторлардың мәтіндерінде т.б. қолданылу тарихын қадағалауға жағдай жасайды. Тек корпустың көмегімен ғана тілдің лексикасы мен грамматикасын ғылыми тұрғыдан зерттеуді қамтамасыз етуге, сонымен қатар түрлі кезеңдерде (бір ғасырдан екі ғасырға дейінгі) тілде болып жататын тілдік өзгерістердің нәзік, бірақ үздіксіз үдерістерін тіркеуге болады.

Үшіншіден, Ұлттық корпусты тура және кері жиілік сөздіктерді жасауда қолдануға болады, олардың бірліктері тілдің әртүрлі даму кезеңдеріндегі мәтіндерде немесе әртүрлі дискурстар мен қосалқы тілдердің қазіргі мәтіндерінде қолданылатын лексикалық бірліктер, грамматикалық формалар, сөзжасам модельдері, синтаксистік модельдер және т.б. болуы мүмкін.

Төртіншіден, тілдің грамматикалық құрылысын ғылыми сипаттау, сонымен бірге беделді академиялық сөздіктер осы тілдік бірліктердің нақты қолданысын объективті түрде көрсететін тілдердің корпусы негізінде құрастырылуы тиіс.

Бесіншіден, дәл осы тіл корпусы тілді қалыптандыру мен стандарттаудың дереккөзі ретінде қызмет ете алады, себебі корпуста жазбаша және дыбысталатын тіл тіркеледі.

Корпустың мүмкіндіктерін көрсету үшін тағы да Орыс тілінің ұлттық корпусына жүгінеміз, бүгінгі таңда Орыс тілінің ұлттық корпусында мынадай шағын корпустар бар: терең аннотацияланған корпус, мұнда әрбір сөйлемге толық морфологиялық және синтаксистік құрылым (тәуелдік ағашы) құрылған; мәтіндердің параллель орысша-ағылшынша корпусы, мұнда белгілі бір орыс немесе ағылшын сөзінің не сөз тіркесінің балық аудармаларын табуға болады; диалектілік мәтіндер корпусы, мұнда Ресейдің әртүрлі аймақтарының грамматикалық ерекшелігі сақталған диалектілік сөздері жазылған; поэтикалық мәтіндер корпусы, мұнда тек қана лексикалық және грамматикалық белгілер бойынша ғана емес, өлеңге тән арнайы белгілер (сонет, эпиграмма, белгілі бір ырғақпен түрі бар амфибрахиямен жазылған өлеңдер және т.б.) бойынша да іздеу мүмкіндігі бар; орыс тілінің оқыту корпусы, мұның белгісі орыс тілінің мектеп бағдарламасына бейімделген; ауызша сөйлеу корпусы, көпшілік және жеке ауызша сөйлеудің магнитофондық жазбалары, сондай-ақ 1930-2000 жылдардағы кинофильмдердің транскриптілері бар.

Қазақстан Республикасының Мәдениет және ақпарат министрлігі Тіл комитетінің сайтында «Қазақ тілінің корпусы» деген сілтеме бар, онда корпусстың мақсатміндеттері былайша анықталған: «Қазақ тілінің корпусы қазақ тіліндегі барлық ақпарат түрлерін жинауға және оны өңдеу, топтастыру, берілген өлшемдер бойынша талдау жөніндегі бағдарламаларды енгізуге...», «қазақ тілінің лексикасы мен грамматикасын ғылыми тұрғыдан зерттеуді қамтамасыз етуге...», «Қазақстан Республикасының мемлекеттік тілінің қазіргі жай-күйін талдаудың керемет жүйесін қалыптастыруға» арналған. Корпустың міндеттері жалпылама түрде түйінделгенін, өкінішке орай, толыққанды Ұлттық корпус құруға бейімдемейтінін көріп отырмыз. Тіл комитетінің Ұлттық корпусы жасау жөніндегі жұмыстары әлі де жүргізіліп жатқандығы айқын, сондықтан сілтемелердің бір де біреуі қазіргі таңда жұмыс істемейді және Қазақстан азаматтарының маңызды қажеттіліктері әзірге қанағаттандырылмай отыр.

Қазақ тілінің Ұлттық корпусын жасаудың басты мақсаты – оның нормаландырылуы мен қалыптандырылуына қызмет ету әлі толық тұжырымдалмаған күйі қалып отыр.

Дәл осы нормаланған және қалыптандырылған тілдің болуы оның мемлекеттік тіл ретінде табысты қолданылуының басты шарты болып табылады. Бұған қазақ тілін дискурстардың корпусық түрлері бойынша үйрету, оның фонетикалық, лексика-семантикалық, грамматикалық, стилистикалық және т.б. нормаларын дамыту және қалыптастыру арқылы ғана қол жеткізуге болады. Тек теңдестірілген, көрнекті көлемі бар, мүмкіндігінше толық және ынғайлы барлық белгі түрлерімен жасакталған қазақ тілінің Ұлттық корпусы ғана тілдік нормалардың бекітілу қарқынының және оларды жүйелі әрі стандартты түрде дұрыс жазу ережелерінде, сөздіктерде, грамматикада, оқулықтарда көрсету, солайша неологиялық, грамматикалық және ортологиялық үдерістерді тіркеу, тілдік бірліктердің семантикасы мен тіркесімділігіндегі сипатталмайтын өзгерістерді, түрлі түлғалардың жиілігін, тіл бірліктерінің функционалды емес түрленуін және т.б. көру мүмкіндігінің толық бейнесін береді, яғни узушта (тұрақты қолданыста) не-ғұрлым жиі қолданылатынын анықтауға байқауға болады.

Корпус – бұл тек «тілді зерттеудің күшті қаруы ғана емес, сондай-ақ жаңа идеология» (В.А.Плунгян), ол тілдің нақты қолданылуын зерттеуге, корпусық тілдік жоспарлауды жүзеге асыруға ғана емес, сонымен бірге қазақ әдеби тілін нығайтуға көмектеседі.

Сонымен, тіл корпусының мүмкіндіктері қызықты әрі әсерлі екендігін байқадық. Олай болса бізге қазақ тілінің корпусы керек пе? Бұл сұраққа жауап айқын деп ойлаймын: Қазақ тілінің Ұлттық корпусы – маңызды мемлекеттік міндеттердің бірі.