

**ҚОЛДАНБАЛЫ ТІЛ БІЛІМІ
ПРИКЛАДНОЕ ЯЗЫКОЗНАНИЕ
APPLIED LINGUISTICS**

FTAMP 16.31.02

doi.org/10.55491/2411-6076-2024-1-133-143

А.Ә. Жаңабекова¹ , Г.Б. Тлегенова^{1*} , А.Ж. Мұқатаева² , М.С. Жолшаева³ 

¹А. Байтұрсынұлы атындағы Тіл білімі институты, Қазақстан, Алматы қ.

²Satbayev University, Қазақстан, Алматы қ.

³SDU University, Қазақстан, Алматы қ.

*e-mail: gulden_20.88@list.ru

ҚАЗАҚ ТІЛІНІҢ ОҚУ КОРПУСЫН ӘЗІРЛЕУ МӘСЕЛЕЛЕРІ

Аннотация. Ұлттық корпустарды құрастыру ісі қазіргі цифрлық заманның кезек күттірмейтін аса маңызды әрі өзекті мәселесі болып отыр. Бұл ретте корпустар жасау ісі әлемдік лингвистикада өткен ғасырдың орта шенінде-ақ басталып кетті. Алғашқы ағылшын тілінің корпусын жасау тәжірибесі басқа да еуропа тілдерінің, орыс, одан қала берді түркі тілдерінің корпустарын жасауға негіз болды. Бұрын қол жұмыстары арқылы орындалған зерттеу және оқыту ісі корпустары пайдаланудың арқасында үлкен нәтижелерге қол жеткізді. Корпустар пайдаланушыны мол материалдармен қамтамасыз етті, жаңадан сөздіктер құрастырып, лингвистикалық зерттеулер жүргізуді жеңілдетті. Корпусты пайдаланушылар саны күннен күнге артты. Мұндай қажеттіліктер корпустарды пайдаланушы мақсатына қарай түрлендіріп жасауға ықпал етті. Тарихшылар немесе тарихи тіл білімін зерттеушілер үшін – Тарихи корпусты, Аудармашылар мен тіл үйренушілер үшін – Параллель корпусты, сөйлеу ерекшеліктерін зерттеушілер мен сөз мәдениетін дамыту үшін – Ауызша корпустарды жасау қажеттілігі туды. Осы ретте білім беру ісін цифрландырудың бірден-бір құралы болатын – Оқыту, білім беру корпустарын жасау да аса қажеттілік туғызды. Оқыту корпусы дегеніміз – тіл үйренушілерге арналған ақпараттық-корпустық база. Мақалада оқу корпустарын әзірлеудің әлемдік тәжірибесіне шолу жасалады. Ағылшын, орыс тілінің Оқыту корпустарының мәтіндік базасында қандай мәтіндер қамтылған, тілүйренушілер үшін қандай лингвистикалық ақпараттар берілген т.б. мәселелері зерттеледі. Сонымен қатар А.Байтұрсынұлы атындағы Тіл білімі институтында әзірленіп жатқан Қазақ тілінің оқу ішкорпусын әзірлеуге қажетті алғышарттар анықталады. Метабелгіленімдері мен тіл үйретуге/үйренуге бағытталған лингвистикалық белгіленімдері, анықтамалық құралдарды тіл үйренушінің қазақ тілін меңгеруіне қарай бейімдеу жолдары, компьютерлік-бағдарламалық әдістері қарастырылады.

Тірек сөздер: оқу корпусы, білім беру корпусы, тіл үйренуші, белгіленім, іздеу жүйесі.

А.А. Жаңабекова¹, Г.Б. Тлегенова¹, А.Ж. Мұқатаева², М.С. Жолшаева³

¹Институт языкознания имени А. Байтұрсынұлы, Казахстан, г. Алматы

²Satbayev University, Казахстан, г. Алматы

³SDU University, Казахстан, г. Алматы

e-mail: gulden_20.88@list.ru

ВОПРОСЫ РАЗРАБОТКИ УЧЕБНОГО КОРПУСА КАЗАХСКОГО ЯЗЫКА

Аннотация. Создание национальных корпусов является актуальной проблемой современной цифровой эпохи. При этом создание корпусов в мировой лингвистике началось еще в середине прошлого века. Опыт создания корпуса первого английского языка послужил основой для создания корпусов других европейских языков, а так же русского, тюркских языков. Исследовательские и учебные корпуса, которые ранее выполнялись вручную, достигли больших результатов благодаря использованию. Корпуса обеспечивали пользователя большим количеством материалов, создавали новые словари и облегчали проведение лингвистических исследований. Количество пользователей корпуса увеличивалось день ото дня. Такие потребности способствовали созданию корпусов в зависимости от назначения пользователя. Для историков или исследователей исторического языкознания – исторический корпус, для переводчиков и изучающих язык – параллельный корпус, для исследователей особенностей речи и развития культуры речи – возникла необходимость в создании устных корпусов. При этом возникла острая необходимость в создании учебных, образовательных корпусов – единственного инструмента цифровизации образования. Учебный корпус – это информационно-корпусная база для изучающих язык. В статье дается обзор мирового опыта разработки учебных корпусов. Изучаются вопросы, какие тексты содержатся в текстовой базе учебных корпусов английского, русского языков, какая лингвистическая информация предоставляется языковедам и др. Также будут определены необходимые предпосылки для разработки учебного

курса казахского языка, разрабатываемого в Институте языкознания им. А. Байтурсынова. Рассматриваются метадаанные и лингвистические разметки, направленные на обучение/изучение языка, пути адаптации справочных средств к овладению казахским языком изучающим язык, компьютерно-программные методы.

Ключевые слова: учебный корпус, образовательный корпус, изучающий язык, разметка, поисковая система.

A.A. Zhanabekova¹, G.B. Tlegenova¹, A.Zh. Mukatayeva², M.S. Zholshayeva³

¹Baitursynuly Institute of Linguistics, Kazakhstan, Almaty

²Satbayev University, Kazakhstan, Almaty

³SDU University, Kazakhstan, Almaty

e-mail: gulden_20.88@list.ru

ISSUES OF DEVELOPMENT OF THE KAZAKH LANGUAGE TRAINING CORPUS

Abstract. The creation of national corpora is an urgent problem of the modern digital era. The creation of corpora in the world linguistics began in the middle of the last century. The experience of creating a corpus of the first English language served as a basis for creating corpora of other European languages, as well as Russian and Turkic languages. Research and teaching corpora, which were previously carried out manually, have achieved great results through the use of. The corpora provided the user with more materials, created new dictionaries and facilitated linguistic research. The number of corpus users increased day by day. Such needs favored the creation of corpora depending on the purpose of the user. For historians or researchers of historical linguistics – historical corpus, for translators and language learners – parallel corpus, for researchers of speech peculiarities and development of speech culture - there was a need to create oral corpora. At the same time, there is an urgent need to create educational corpora – the only tool for digitalization of education. An educational corpus is an information and corpus base for language learners. The article gives an overview of the world experience in the development of educational corpora. The questions of what texts are contained in the text base of English, Russian language learning corpora, what linguistic information is provided to linguists, etc. are studied. The necessary prerequisites for the development of the Kazakh language training course being developed at the A. Baitursynov Institute of Linguistics will also be determined. Metadata and linguistic markups aimed at teaching/learning the language, ways of adapting reference tools to Kazakh language acquisition by language learners, computer-software methods are considered.

Keywords: training corpus, educational corpus, language learner, markup, search system.

Кіріспе

Бүкіл тіл жүйесін біртіндеп компьютер жадына енгізіп, оларды белгілі бір автоматты басқару орталығы арқылы іске қосу, әсіресе, тіл тұтынушыларының практикалық қажеттілігі үшін аса маңызды. Бұрын ғылыми зерттеулер мысал жинау, карточка жазу, оларды жинақтау, топтау сияқты қол жұмыстары арқылы атқарылып, ұзақ уақытқа созылған ғылыми ізденісті қажет еткен болса, тілдік корпустар мұндай жұмыстарды барынша жеңілдететіндігі және де тілші мамандар үшін ғана емес, журналистер, әдебиетшілер, тарихшылар, мәдениеттанушылар, тіпті қалың көпшілік қауым үшін де ақпараттық-анықтамалық құрал қызметін атқару мүмкіндігі бар екендігі анықталып отыр. Осы орайда білім беру жүйесін автоматтандыру ісі қазіргі педагогикада аса өзекті мәселелердің біріне айналды. Бұл өз кезегінде мұғалімдерді шығармашылық әрекеттен, қол жұмысынан босатып, білім беру жүйесін қарапайым механикалық амалдарға ауыстырады. Білім беру корпусы жасалса, білім беру барысындағы дәстүрлі әдіс-тәсілдерге жұмсалатын адам еңбегі компьютерге (білім беру корпусына) жүктеледі. Бұл адам баласының бір рет ғана атқаратын еңбегін ізденушілер өз қалауынша компьютердің көмегімен сан рет қайталауына мүмкіндік алады (Жұбанов, 2008: 593).

Білім беру/оқу корпусы – оқытуды дараландыру үшін компьютерлік негіздегі техникалық, оқу-әдістемелік, лингвистикалық, бағдарламалық пен ұйымдастырушылықты қамтамасыз етуге арналған ішкорпус. Корпус арқылы білім беруді «автоматтанған оқу жүйесі» құрайды (Жұбанов, 2008: 593).

Білім беру корпустары әр түрлі формада болады. Олардың әрқайсысы тілдік білім беруде белгілі бір мақсаттарға қызмет етеді. Кең таралған түрлерінің бірі – тіл үйренушілер әртүрлі оқу контексттерінде жасаған тілдік деректерден тұратын оқу корпусы. Оқушылар корпусы оқушылардың тілінің дамуын зерттеу, оқушылардың қателіктерін анықтау және оқушылардың қажеттіліктерін ескере отырып оқу материалдарын әзірлеу үшін құнды ресурстар болып табылады. Тағы бір түрі – тілді оқыту процесінде қолданылатын мәтіндер мен материалдарды, мысалы, оқулықтар, сынып сабақтары және тілді үйренуге арналған ресурстарды қамтитын білім

беру корпусы.

Білім беру корпусы оқытушыларға оқу материалдарын талдауға және бағалауға, олардың тиімділігін анықтауға және мақсатты оқыту стратегияларын жасауға мүмкіндік береді.

Білім беру корпустарын құру бірнеше кезеңдерді қамтиды: соның ішінде деректерді жинау, аннотация және талдау. Деректерді жинау оқулықтар, сынып жазбалары, онлайн ресурстар және шынайы байланыс контексттері сияқты әртүрлі көздерден жазбаша мәтіндерді, ауызша транскриптілерді немесе мультимедиялық материалдарды жинауды қамтуы мүмкін. Ал аннотация лингвистикалық талдау мен іздеуді жеңілдету үшін сөйлеу бөлігінің тегтері, синтаксистік құрылымдар және дискурс маркерлері сияқты корпусның лингвистикалық ерекшеліктерін жүйелі түрде таңбалауды және кодтауды қамтиды. Оқу корпусын талдау әртүрлі сандық және сапалық әдістерді, соның ішінде жиілікті талдауды, үйлестіруді және дискурстық талдауды пайдалана отырып, тілді пайдалану үлгілері мен вариацияларды зерттеу үшін жүргізілуі мүмкін.

Білім беру корпустары тілді оқытуда, оқуда және бағалауда әртүрлі қолданыстар табады. Тілді оқытуда корпустар материалдарды әзірлеу, тілдік курстарды жобалау және оқушылар арнайы ресурстар жасау үшін пайдаланылуы мүмкін. Білім беру корпустарын қолдана отырып тілді оқытуда оқушылардың коммуникативтік құзыреттілігі мен тілдік хабардарлығын дамыту үшін шынайы тілдік деректер мен нақты жағдайларды пайдалануға ерекше назар аударылады.

Тілді үйрену кезінде корпус студенттерге конкорданс, коллокацияны талдау және қателерді талдау сияқты әрекеттер арқылы өз бетінше оқуға, сөздік қорын және тілдік тәжірибені алуға мүмкіндік береді. Тілді меңгеру деңгейін бағалау тілді меңгеру деңгейін, оқушылардың үлгерімін бағалау үшін лингвистикалық деректерді пайдалануды қамтиды, бұл олардың күшті және әлсіз жақтарын, сондай-ақ оқу қажеттіліктерін түсінуге мүмкіндік береді.

Білім беру корпусындағы деректер:

- 1) тіл үйрену;
- 2) орфографиялық, грамматикалық, синтаксистік ережелерді меңгеру;
- 3) оқулықтар, оқу құралдарын, әртүрлі жаттығулар жинағын құрастыру;
- 4) онлайн мультимедиялық оқыту сервистерін құру;
- 5) ғылыми лингвистикалық зерттеулер жүргізу үшін аса пайдалы.

Материал және әдістер

Мақалаға зерттеу материалы ретінде Оқу корпустарын әзірлеу тәжірибесі туралы жазылған еңбектер және қазақ тілінде оқу корпустарын жасаудың алғашқы қадамдары туралы жазылған ғылыми еңбектер алынды. Мақалада бірнеше зерттеу әдістері қолданылды. Материалға шолу жасауда сипаттау тәсілі, әлемдік оқу корпустарының мәтіндік базасы, белгіленімдері, іздеу жүйесі т.б. мәселелерді баяндауда салыстыру әдісі қолданылды. Оқу корпусының мәтіндік базасын жинақтауда филологиялық-сараптама әдісі негізге алынды. Ал оқу корпусының бір ерекшелігі – омоним ажырату мәселесін қарастыруда дистрибуциялық талдау әдісі қолданылды.

Әдебиетке шолу

Шет тілін (ағылшын) үйренушілердің сөйлеу мәтіндерінің корпусын жасау өткен ғасырдың 90-жылдары басталған. Мәселен, осындай корпустардың бірінде 10 информанттан ағылшынша сөйлеуінен жазып алынған қате қолданыстар іріктелген 2000 сөзқолданыстан тұратын картотекалық қор бар. Осындай қате қолданыстар негізінде типтік қателерді анықтап, «актуальный словарь», «актуальная грамматика» тұрғысынан тіл үйренуші лексикасы мен грамматикасындағы қателіктерді түсіндіретін, жөндейтін қосымшалар жасаған, белгіленімдер енгізген. Мұндай тіл үйренушілер сөйлеуінен жинақталған корпустар материалдарын тілдің төл иелмендері мен тіл үйренушілердің сөйлеуін салыстырып, әртүрлі қорытындылар жасаған. Осы сияқты корпус мысалы – Эли Хинкель корпусы (500 мыңдай сөзқолданыстан тұрады) (Hinkel, 2002).

Осылай оқу корпусының мақсаты шет тілін екінші тіл ретінде үйретуге бағытталған болатын. Тіл үйренуге арналған корпус – белгілі бір жолмен біртұтас, құрылымдалған және белгіленген көлемді мәтіндік массив, электронды түрде қолжетімді және нақты зерттеу міндеттерін шешуге арналған өкілдік мәтіндерді немесе олардың фрагменттерін іріктеу болып

табылады.

Тілдік корпустың егжей-тегжейлі белгіленуі оны ғылыми-зерттеу және тәжірибелік-конструкторлық жұмыстардың кең ауқымын шешу үшін пайдалануға мүмкіндік береді. Соның ішінде аударма практикасында, шет тілдерін оқытуға септігін тигізеді.

Т. В. Шмелева тілдік корпусты шешуге көмектесетін тапсырмаларды қорытындылайтын корпустың «тапсырма» түрін жасады. Атап айтқанда, ол келесі міндеттерді анықтады:

- 1) лексикографиялық (мысалы, иллюстрациялық материалды іздеу, талдау, сөз семантикасындағы өзгерістер, жаңа сөздіктер жасау және т. б.)
- 2) варианттылық;
- 3) сөзжасамдық (мысалы, сөзжасамдық модельдердің өнімділігін, туынды сөздердің контекстік мүмкіндіктерін, туынды сөздерді зерттеу және т.б.);
- 4) синтаксистік (мысалы, етістікті басқарудың варианттарын зерттеу – тосковать о/тосковать по (аңсау/аңсау));
- 5) қазіргі емлеге қатысты (мысалы, бренд/брэнд) орфографиялық тапсырмалар (Шмелева, 2007).

1-кесте – Тілдік корпустардың классификациясы

Тілдік деректер түрі	Ауызша, жазбаша, мультимедиялық, аралас
Мәтіндердің тілі	Біртілді, екітілді, көптілді, аралас
Тілдік норма	Әдеби, диалект, аралас
Мәтіндердің дискурсивті-стистикалық байланысы	Әдеби, публицистикалық, ғылыми, аралас
Мәтіндердің жанры	Прозалық, поэтикалық, драматургиялық, аралас
Қолданылу мақсаты	Өртүрлі, мамандандырылған
Қолжетімділігі	Қолжетімді, коммерциялық, жабық
Мақсаты	Ғылыми, иллюстративті, аралас
Хронологиялық белгісі	Синхронды, диахронды

Ең танымалы: «International Corpus of Learner English» атты халықаралық ағылшын корпусы (1990) Бельгияда жасалған (Granger, 1990). Бұл корпусқа ағылшын тілін жетік меңгерген студенттердің эсселері енгізілген.

Тағы бір танымал оқу корпусы С.Гранже корпусында 14 тіл ұстармандарының эсселер жинағы берілген (Granger, 1998).

Гонконг корпусында ағылшын тілін үйренуші студенттердің жазба жұмыстары жинақталған т.б.

Оқу корпустарын талдаған зерттеушілердің көрсетуінше, қазіргі Оқу корпустарының көпшілігінің мәтіндік базасы 50 000-150 000 сөзқолданыстан тұрады. Алайда көлемді Оқу корпустары жоқ емес. Мысалы, the Quebecs learner corpus-ында 250 000-нан аса сөзқолданыс бар (Мальцева, 2011).

Орыс тілінің оқу корпусы 2007 жылы қолданысқа енгізілген. Бұл корпус орыс тілі мен әдебиетін мектепте оқытуға арналып жасалған. Бұл корпусты сонымен қатар жоғары оқу орындарындағы білім беру кезінде қолдануға болатыны айтылады. Әсіресе, өзге салалар мамандарына тіл үйретуде, өзге ұлт өкілдеріне тіл меңгертуде қолданады (ОТҰК). Орыс тілінде оқу корпустары негізінде оқушылар жіберетін типтік грамматикалық, стилистикалық қателіктер туралы да зерттеулер жазылған (Соснина, 2017).

Орыс ғалымдары Добрушина Н. Р., Левинзон А. И. оқу корпусының үш түрге бөлінетінін ажыратып көрсетеді: «В настоящий момент существует три основных направления применения корпуса в образовании: 1) составление заданий и упражнений с использованием корпусного материала, 2) организация самостоятельной исследовательской работы учащихся, 3) экспертиза учебников и словарей» (Добрушина Н. Р., Левинзон А. И., 2006: 336).

Қазақ тілінде Білім беру корпусын Ұлттық корпустың құрамдас бөлігі ретінде әзірлеу қажеттігі және оның мәтіндік базасы мен белгіленімдері туралы алғаш айтқан ғалым – профессор

Асқар Құдайбергенұлы Жұбанов болды. А.Жұбанов Білім беру корпусын құрастыруда келесі міндеттерді негізге алу қажеттігін айтады: «Білім беру корпусына енетін мәтіндердің метамәтіндік параметрлері компьютерлік базаға салынуы қажет. Сондықтан корпус құрастыруда метабелгіленімдер (оқулық атауы, авторы, жылы, шыққан орны, стилі, жанры, көлемі т.б.) мен грамматикалық белгіленімдерді мәтінге енгізудің әдіс-тәсілдерін, түрлерін, компьютерлік бағдарламаларын т.б. зерттеудің маңызы зор; – Өртүрлі сөздіктер құрастыруда, ғылыми грамматикалар жазуда, оқыту жүйесінде және оқулықтар құрастыруда, әдістемелік құралдар жазуда, аударма жұмыстарында, ғылыми зерттеулер жүргізуде, статистикалық мәліметтер алуда, энциклопедиялық ақпараттар алуда осы білім беру корпусы негізге алу керек; – Білім беру корпусына енген мәтіндердің құрамына тек қазақ тілін ғана емес, басқа да пәндер қосу және мәтіндерде кездесетін омоним сөздерді автоматты түрде немесе жартылай автоматты түрде ажырату қажет» (Жұбанов, 2016: 81).

Қазақ тілінде Оқу корпусын әзірлеу туралы мәселе А.Байтұрсынұлы атындағы Тіл білімі институтында А.М.Фазылжанның жетекшілігімен «Қазақ мәтінін автоматты тану: лингвистикалық модульдер мен ІТ-шешімдер әзірлемесі» атты бағдарламалық-нысаналы қаржыландыру бойынша зерттеу жобасының аясында арнайы көтеріліп, зерттеліп, жасақталып жатыр.

Нәтижелер және талқылау

Оқу корпусы – білім беруге арналған мәтіндік база және лингвистикалық білімдер жүйесі. Оқу корпусы – оқытуды оңтайландыру үшін компьютерлік негіздегі оқу-әдістемелік, лингвистикалық білімдер мен оларды техникалық, бағдарламалық жағынан қамтамасыз етуге арналған ішкорпус. Оқып-үйренуді автоматтандыру оқу барысында үйренушінің қажеттілік, қызығушылық деңгейін арттыруға мүмкіндік туындатады. Корпус электронды пішіндегі мәтіндер жинағы болғандықтан, оқу корпусына да ең алдымен білім алуға, оқып-үйренуге бағытталған мәтіндер жиынтығы болу керек. Білім беру мәселесі ең алдымен мектеп оқулықтарынан басталады.

Оқу корпустарын жасаудың көпшілігінің басты мақсаты – оқытылып отырған тілді меңгертудің әдіс-тәсілдерін анықтау мен оның тиімділігін талдауға бағытталған. Әсіресе мектеп оқушылары мен өзге тілді адамдарға тіл меңгеру ісін оңтайландыруды көздейді. Тіл үйрену кезіндегі жазба жұмыстары мен ауызша сөйлесімдер базасын жасақтау арқылы, корпустық мәтіндік база жинап, ондағы типтік қателерді анықтау арқылы тіл үйрету процесіндегі маңызды, назар аударуға тиісті нәрселерді шығарады. Сөйтіп, тіл меңгеру, білім беру бағдарламалары мен әдіс-тәсілдеріндегі кемшіліктерді анықтауға мүмкіндік береді.

Оқу корпусының функциялары:

- Бастапқы білім деңгейін, ізденушілердің іскерліктерін, олардың дара ерекшеліктерін анықтау;
- Оқу материалын даярлау: түсіндірілетін мәтіндер мен оқылатын мәселе бойынша, оқу және бақылау тапсырмалары бойынша иллюстрациялар;
- Оқу материалын көрсету, оның күрделілік деңгейі ақпаратты беру қарқыны бойынша адаптациясы.

Оқу корпустарын құрастыруда тіл меңгертудің деңгейлеріне лайықтап жасақтау мәселесі де басты назарда болуы керек. Өйткені тілді мүлде білмейтін басқа ұлт өкілі мен бұл тілде сөйлей алатын тіл үйренушінің қажеттілігі бірдей емес. Осы орайда оқу корпустарын құрастырушылар кейде оны тіл деңгейлеріне қарай бөліп жасақтайды.

Қазақ тілінің Оқу корпусын әзірлеуде де деңгейлік оқыту принципін ұстану қажет деп есептейміз. Айталық, қазақ тілін оқытуда жоғарғы оқу орындарында «Практикалық қазақ тілі», «Кәсіби қазақ тілі» пәндері бар. Бұл пәннің алғашқысы орыс аудиториясында оқитын 1-курс студенттеріне оқытылады. Ал екіншісі мамандықтары бойынша терминдерді меңгеру мен кәсіби сөйлеуге бағытталады. Сол сияқты қазіргі уақытта жоғарғы оқу орындарының филология факультеттерінде «Шешендік өнер» т.б. атаулармен қазақ тілінде шешен сөйлеуге бағытталған арнайы элективті курстар жүргізіледі. Ісқағаздарын жүргізу бойынша да пәндер оқытылады. Қазақ мектептерінде оқытылатын қазақ тілі пәні орта мектеп шеңберіндегі қазақ тілінің ғылыми

аппаратын игеруге бағытталуы тиіс. Осы тілдік деңгейлерге қарай Қазақ тілінің оқу корпусының мәтіндік базасы мен лингвистикалық құралдары төмендегі ретпен жасақталуы тиіс:

- Практикалық қазақ тілі;
- Кәсіби қазақ тілі;
- Теориялық қазақ тілі;
- Іскери қазақ тілі;
- Шешендік қазақ тілі.

Орыс тілінің оқу корпусының мәтіндік базасы бірнеше стильдер мен жанрларды қамтиды. Корпустың Іздеу жүйесі үлкен екі ұяшыққа жинақталып берілген:

- 1) Көркем мәтіндер;
- 2) Бейкөркем (нехудожественные) мәтіндер.

Көркем мәтіндер негізінен жоғары сынып оқушыларының қазақ тілі мен әдебиет хрестоматияларында берілетін және сыныптан тыс оқу ұсынылатын көркем шығармалардан жинақталған. Біздің түсінуімізше, орыстың классикалық жазушылар шығармаларын алған. Ал бейкөркем мәтіндер орыс тілі мен сөз мәдениеті пәндерінде оқытылатын басқа да функционалды стильдерден алынған. Мысалы:

- Публицистика;
- Ресми-іскери;
- Ғылыми (мектеп);
- Сөйлеу.

Жанрлық құрамы жан-жақты, негізінен мектеп бағдарламасы аясындағы мәтіндерге негізделген. Атап айтқанда, *мақала, сұхбат, күнделік, өтініш, іскери хат, ғылыми не ғылыми-көпшілік мақала, сараптама, аңдатпа, реферат, тұрмыстық хаттар, күнделікті сөйлеу т.б.*

Қазақ тілінің оқу корпусын жасауға қатысты А.Жұбанов «Білім беру корпусына – ең алдымен мектеп бағдарламасы бойынша оқытылатын оқулықтар және ЖОО студенттерінің пән бойынша оқытылатын шығармалары кіреді», – деген болатын.

Қазақ тілінің оқу корпусының мәтіндік базасына не саламыз?

● Көркем шығармалар – *Жалпы тіл үйренушілердің бәріне (1) орыс тілінде әдетте мектеп оқулықтарының хрестоматияларында берілетін классикалық шығармалар, 2) балалар әдебиеті; фольклор (ертегі, жыр, жұмбақ, жаңылтпаш, өлеңдер, мақал т.б.);*

● Тақырыптық мәтіндер (*практикалық қазақ тілін үйренушілер үшін*) (*Отбасы, Денсаулық, Саяжай, Саяхат, Дүкенде т.б.*)

● Ғылыми мәтіндер – *Кәсіби қазақ тілін меңгерушілер үшін (басқа пәндер бойынша)*

● Публицистикалық мәтіндер – *Газет-журналдарға мақала жаза алу үшін («Балдырған», т.б.)*

● Ісқағаз мәтіндері – *Ісқағаздарын жазып үйрену үшін (өтініш, хабарландыру, келісімшарт, құттықтау, қызметтік хат т.б. үлгілері)*

Жалпы шетелдік оқу корпустарында, жоғарыда айтып өттік, типтік қателерді анықтау үшін, мектеп оқушыларының эсселері, т.б. жазба жұмыстарын корпус базасына салған. Алдымен жаппай жазба жұмыстарын жинақтап, эксперименттер жүргізген. Екінші кезекте оқушылар жазбаларының қатесін корпустық менеджерлер арқылы тексеріп, типтік қателерді анықтаған. Оқу корпустарын компьютерлік қамтамасыз ету шет тілін меңгертудің тиімділігін арттыруға әсер етеді. Өйткені корпустық менеджерлер (құрылғылар) орындайтын мүмкіндіктерді қол жұмысы арқылы жүзеге асыру өте қиынға соғады. Бұл туралы М.С.Мальцева: «Как правила, учебный корпус отражает результаты типичной деятельности, представляя одинаковые письменные задания, выполненные обучаемыми их эквивалентных групп. Учебный корпус позволяет анализировать продукты учебной деятельности студентов, выявляя их успеваемость, наиболее типичные лексические, стилистические и грамматические ошибки, проблемы в знаниях, на которые следует направить особое внимание преподавателям в дальнейшей работе с обучаемыми», – деп тұжырымдайды (Мальцева, 2011).

Мұндай типтік қателер орыс тілін оқытудағы бағдарламалар мен оқыту әдістеріне өзгерістер жасауға ықпал еткен. Сөйтіп, орыс тілінде тіл үйрету мәселесін шешуде Оқу

корпустарының маңызы зор.

Ал енді қазақ тілінде оқу корпустарының осы қызметін қалай жүзеге асырамыз деген сұрақ туындайды. Бұл ретте орыс тілінің оқыту корпустарындағыдай эсселер базасын жинақтау мен ондағы типтік қате қолданыстарды анықтау үшін біздің ойымызша:

Мектеп платформаларындағы оқушы жазбалары, эсселер (мектептерден алу) (Күнделік кз.)

- *Қазақ тілінен емтихан платформаларындағы эсселер (жоғары оқу орындарынан алу)*
- *Арнайы эссе жаздыру (мектеп мұғалімдерімен байланыс орнату) (эксперимент)*

арқылы электронды жазба жұмыстарын жинақтауға болар еді.

Мұндай типтік қателерді О.А.Фирсова бірнеше түрге бөледі:

- Лексикалық қателер;
- Грамматикалық қателер;
- Стилистикалық қателер.

Лексикалық қателерге дұрыс емес сөзқолданыстарын, сөздердің тіркесіміндегі қате тіркестерді, сөз мағынасын дұрыс түсінбеу сияқтыларды жатқызады. *Грамматикалық қателер* род, число, падеждегі, олардың байланысуындағы, синтаксистік құрылымдардағы қателер. *Стилистикалық қателер* деп функционалды стиль түрлерін дұрыс, орынды қолданбауды атаған, яғни мәселен, ғылыми мәтіндерде ауызекі сөйлеу тілінің элементтерін қолдану т.б.

Қазақ тілінің Оқу корпусын жасау егер басқа шетелдік оқу корпустарындай сауаттылыққа бағытталса, онда оның жұмысын мынадай корпустық қосымшалар арқылы реттеуге болады:

- Мәтіндік редактор;
- «Сөз тексер» қосымшасы;
- Тіркесімділік сөздігі;
- Жиілік шығару программасы.

Оқу корпустарындағы типтік қателерді анықтау мәселесімен байланысты екінші бір қажеттілік – оқушылар, оқытушылар, тіл үйренушілер үшін Сауаттылық құралдарын Оқу корпусының жұмыс үстеліне енгізу болып табылады.

- Жазу сауаттылығы (Орфографиялық сөздік, емле, пунктуациялық анықтағыш)
- Оқу сауаттылығы (Орфоэпиялық сөздік, орфоэпиялық конвертор)
- Лексика-грамматикалық сауаттылық (оқулықтар, түсіндірме сөздіктер т.б.).

Лингвистикалық белгіленімдердің ішінде корпустағы грамматикалық белгіленім Оқу корпустарында маңызды рөл атқарады. Оның ішінде морфологиялық белгіленімдер Оқу корпустарының негізгі белгіленімі болып келеді. Оқу корпустарындағы морфологиялық белгіленімдер мектеп бағдарламасына негізделген дәстүрлі грамматикаға сүйенеді. Оқу корпустарында Академиялық грамматика материалдары әлдеқайда жеңілдетіліп берілген, күрделі, даулы мәселелер алынып тасталған. Бұл жайлы ғалымдар: «Морфологическая разметка в обучающем корпусе адаптирована к стандартным нормативным пособиям по русскому языку и обеспечивает именно традиционный, упрощенный грамматический разбор», – деп көрсетеді (Мальцева, 2011).

Орыс тілінің мектеп бағдарламасына бағытталған белгіленімі бар, омонимдер қатары ажыратылған корпус – орыс тілінің оқыту корпусы. Бастапқыда корпустың мақсаты лексикалық, орфографиялық және грамматикалық қателерге талдау жасау болды. Аз көлемде болса да оқыту корпусына алынған мәтіндерді омонимдерін ажыратып берген, яғни оқушы шатаспау үшін, Оқу корпусын пайдаланушылар үшін омоним сөздердің қай сөз табына жататынын, қандай грамматикалық форма екенін нақты көрсеткен. Алайда омонимдер қолмен ажыратылған. Өйткені автоматты әдіс кейбір қате талдауға ұрындыруы мүмкін. «В корпусе представлены тексты со снятой вручную омонимией» – деп мұны корпус әзірлемешілері атап көрсетеді.

А.Жұбанов корпустағы омоним мәселесі туралы: «Табиғи тілді талдау автоматты түрде айна қатесіз жүргізіледі деуге болмайды, себебі, ол, әдетте, бір лексикалық бірліктің (сөз, сөзтіркес, сөйлем) бірнеше талдау нұскасын ұсынады. Мұны *грамматикалық омонимия* деп атайды. Осындай проблемалық мәселелерді (морфологиялық, синтаксистік) жою компьютерлік лингвистика саласының ең бір маңызды және күрделі мәселелеріне жатады», – деп көрсеткен

болатын (А.Жұбанов, 2016).

Қазақ тілінің Оқыту корпусына салынған мәтіндерге түгелдей морфологиялық белгіленімді автоматты программа арқылы жүзеге асыру мүмкін емес. Өйткені тілдегі омоним, омограф, омоформалар бір сөз туралы бірнеше нұсқаны, яғни омонимдер қатарын көрсетеді. Тіл үйренуші тілді толық меңгермеген жағдайда, шатасуы мүмкін. Сондықтан автоматты морфологиялық талдаулар жасау бағдарламасына қосымша, омонимдерді қолмен ажырату қажеттігі туындайды. Мәселен, төмендегі омоним, омоформаларды компьютерлік бағдарлама автоматты түрде қаймағынада тұрғанын ажырата алмайды.

Қара жол
Бері қара
Қара танымайды

Сал боп ауырды
Көре сал
Сал адам

Көр соқыр
Көр қазу
Кино көр

Біздің ойымызша, оқу корпусына салынатын мәтіндердегі омонимдерді түгелдей ажырату мүмкін емес. Шетелдік корпусарда осы себепті оның көлемі аз боп келеді. Бұл ретте Оқу корпусына мәтіндерді *белгіленген және белгіленбеген* мәтіндер деп екіге бөліп орналастыру қажет. Олай болмаса, белгіленген мәтіндер көлемі аз болып, Оқу корпусының мәтіндік базасы азайып қалады. Білім беру корпусына енетін мәтіндердің метамәтіндік параметрлері компьютерлік базаға салынуы қажет.

Сондықтан корпус құрастыруда метабелгіленімдер (оқулық атауы, авторы, жылы, шыққан орны, стилі, жанры, көлемі т.б.) мен грамматикалық белгіленімдерді мәтінге енгізудің әдіс-тәсілдерін, түрлерін, компьютерлік бағдарламаларын т.б. зерттеудің маңызы зор. Білім беру мәселесі ең алдымен мектеп оқулықтарынан басталу қажет.

Қазақ тілінің оқу корпусы кімдерге арналып жасалады?

- Оқытушыларға (әртүрлі деңгейде қазақ тілі пәнін өткізу кезінде);
- Оқулық құрастырушыларға;
- Екінші тіл ретінде қазақ тілін оқушыларға (Практикалық қазақ тілі);
- Кәсіби тілдесім мен кәсіби терминологияны меңгерушілерге (Кәсіби қазақ тілі);
- Қазақ тілін ғылым тілі ретінде оқитын оқушылар мен студенттерге (Теориялық қазақ тілі);
- Қазақ тілінің дәстүрлі сөзсаптамын жатық меңгергісі келетіндерге (Шешендік қазақ тілі);
- Жалпы тіл (қазақ тілін) үйренушілерге.

Орыс тілінің оқыту корпусында «проект-спутники» деп аталатын қосымша жұмыс істейді. Бұл қосымша корпус материалына негізделген жаттығулардан, мектеп платформаларындағы әдістемелік құралдардан тұрады.

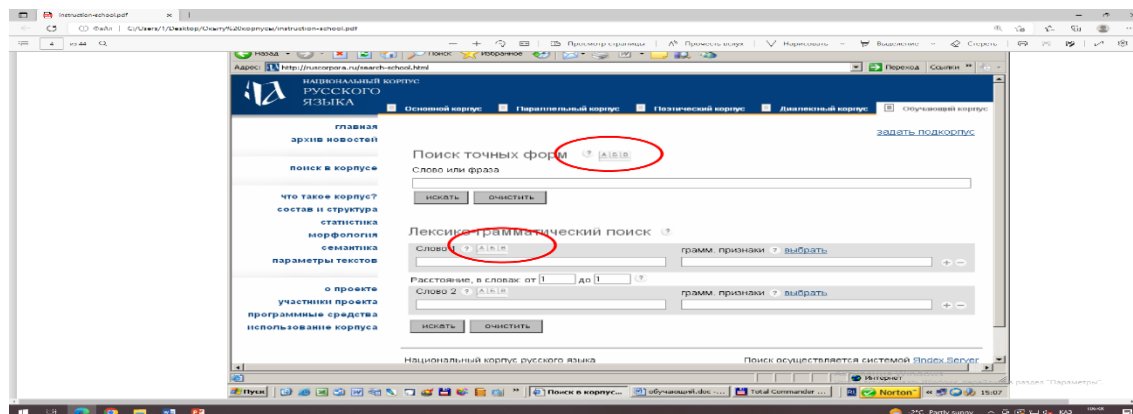
Орыс тілінде ғылыми грамматикалық талдау мектепте бекітілген талдауға қарағанда күрделі. Падеж саны көбірек. Алайда Орыс тілінің оқыту корпусы тіл мұғаліміне практикалық көмек және мектеп бағдарламасындағы тапсырмаларды оқытуға бағытталғандықтан, грамматикалық талдау дәстүрлі, жеңілдетілген болып, мектепте оқытудың тағы бір ерекшелігі – орфографияға бағытталады.

Оқыту корпусынан қазіргі ғылыми орыс тіліндегі кейбір категориялар алынып тасталған. Олар: компаратив, санау формасы (счетная форма) звательная форма, есімдік-предикатив.

Лексика-грамматикалық топтардың семантикалық ақпараты мектеп бағдарламасына сәйкес енгізілді: (*зат есім, сын есім, есімдік*).

Оқыту корпусын пайдалану бойынша нұсқаулық оқытушылар мен қатысушылардың

қажеттілігі ескере отырып арнайы жасалған. Нұсқаулық қолжетімді әрі көрнекі түрде жасалған.



1-сурет – Орыс тілінің оқу корпусы

Автоматтанған оқу жүйесі әртүрлі пәндерді оқыту барысында пайдаланылуы мүмкін. Оның қызметі сол жүйедегі программалардың оқушыға тигізетін әсері арқылы анықталады. Оқыту программаларының негізгі қызметтеріне мыналар жатады:

- Жаттығу;
- Бақылау;
- Бағалау (Жұбанов, 2008: 130).

Бұлар Білім беру корпусына тікелей қатысты программалар. Жаттығу программасының негізінде мұғалім сабақ үстінде корпустан қажетті иллюстративті материалды іздей алады немесе жаттығуларды құрастыра алады. Немесе ЖОО-да студенттер семинарда пайда болған нақты қойылған мәселелерді шешуде сұрақтарға жауап бере алады. Мұғалім немесе оқушы, студент корпустан сөздің белгілі бір формасын іздеп-тауып, солардың негізінде неше түрлі жаттығулар құрастыруына болады.

Ал бақылаушы программалардың жұмысы кезінде әрбір тапсырманың жауаптары базада сақталып отыруы қажет. Жауаптардың негізінде кеткен қателердің ақиқат және кездейсоқ екенін ажырата алады.

Бағалау программаларында қателерді әрі қарай талдап, тапсырма нәтижелерін статистикалық өңдеуден өткізеді. Осылай корпустен жұмыс істеуде студенттің, оқушының шығармашылық қабілеті ашылады. Мәселен, қолданушы корпустан сөздің белгілі бір формасын іздеп-тауып, солардың негізінде неше түрлі жаттығулар құрастыруына, мысалдар жинақтауына болады. Тапсырмаларды өз бетімен де, үй жұмысы ретінде де орындай алады.

А.Жұбанов сонымен қатар оқыту жүйесін құрастыруға қатысты мынадай ұстанымдар болу қажеттігін атап көрсетеді:

1. Көптеген автоматтанған оқу жүйесінің жоғары нәтижелілігінің жеткіліксіздігі оқып-үйрену ісіне кешенді түрдегі қатынастың жоқтығы. Оқытудың кешенділік ұғымы оқып-үйрену барысында компьютердің толық аралас жағдайы.

2. Тілді оқып-үйрету көпаспектілі болуы қажет. Яғни, оқыту программасын құрастыруда оқып-үйрету мазмұнын құрайтын ақпаратты тілдің әр деңгейіне бірдей қатысты болу керек дейді.

3. Оқыту барысының табиғилығы. Арнайы дайындықсыз диалог жүргізуді ұйымдастыру керек. Мысалы, тілді жаңадан үйреніп жатқандарға ана тілі негіз болғаны дұрыс.

4. Оқып-үйренуді автоматтандыру оқу барысында үйренушінің қажеттілік, қызығушылық деңгейін арттыруға мүмкіндік туындатады.

5. Білім беру корпусына енетін мәтіндердің метамәтіндік параметрлері компьютерлік базаға салынуы қажет. Сондықтан корпус құрастыруда метабелгіленімдер (оқулық атауы, авторы, жылы, шыққан орны, стилі, жанры, көлемі т.б.) мен грамматикалық белгіленімдерді мәтінге енгізудің әдіс-тәсілдерін, түрлерін, компьютерлік бағдарламаларын т.б. зерттеудің маңызы зор.

6. Өртүрлі сөздіктер құрастыруда, ғылыми грамматикалар жазуда, оқыту жүйесінде және оқулықтар құрастыруда, әдістемелік құралдар жазуда, аударма жұмыстарында, ғылыми-зерттеулер жүргізуде, статистикалық мәліметтер алуда, энциклопедиялық ақпараттар алуда осы білім беру корпусы негізге алу керек.

7. Білім беру корпусына енген мәтіндердің құрамына тек қазақ тілін ғана емес, басқа да пәндер қосу және мәтіндерде кездесетін омоним сөздерді автоматты түрде немесе жартылай автоматты түрде ажырату қажет (Жұбанов, 2008: 131).

Қорытынды

Қорыта келгенде, оқу корпусы әртүрлі мәтіндер корпусынан құралады. Өйткені оқу корпусының талаптары бойынша – ол мәтіндер үлкен көлемде, құрылымдалған, өлшемділікте, бірізділікте электронды формада сақталады.

Сонымен бірге Қазақ тілінің оқу корпусы лингвистикалық білімдер беретін ақпараттық жүйе ғана емес, лингвистикалық және лингводидактикалық талдаулар үшін репрезентативті база болып табылады. Әсіресе оқу корпусы оқып-үйренушілерге арналған білім беру жүйесі ретінде қазақ тілін меңгеру процесінің ақиқат бейнесін көрсетеді.

Осы ретте тіл деңгейлері бойынша мектеп оқулықтарына арналған ең адекватты ережелер мен оларды дәйектейтін мысалдар базасын жасау аса маңызды.

Сонымен қатар мектеп оқулықтарына, тіл үйренушілерге арналған лексикалық минимум мен олардың иллюстрациялық мәтіндік базасын, мектеп жасына арналған түсіндірме, екі тілді сөздіктердің автоматты нұсқасын корпусқа енгізуде, ғылыми-зерттеулер жүргізуде, статистикалық мәліметтер алуда, энциклопедиялық ақпараттар алуда да маңызды. Тіл үйренудегі бақылау, өздік жұмыстарды әлдеқайда жеңілдетеді. Оқушылардың жазба жұмыстарын, атап айтқанда, рефераттар жазу, зерттеулер жүргізу сияқты қажеттіліктерін өтейтін болады.

Оқу корпусы лингвистикалық талдаудың бірқатар функцияларын жүзеге асыруға мүмкіндік береді. Мысалы, синтаксистік модельдер, сөздерді санау, жиілікті анықтауда тиісті компьютерлік бағдарлама арқылы қамтамасыз етіледі. Ал оқу корпустарын компьютерлік қамтамасыз ету шет тілін оқыту процесін оңтайландыруға ықпал етеді, бұл корпус менеджерлерін пайдаланбай орындау мүмкін емес мәселелерді шешуге мүмкіндік береді (корпус негізіндегі талдау құралдары).

Оқу корпусы оқытушыларға, зерттеушілерге және студенттерге құнды лингвистикалық ресурстар бере отырып, тілдерді оқытуда және зерттеуде маңызды рөл атқаратынын атап өткен жөн. Өртүрлі контексттерде тілдің қолданылуы, өзгергіштігі және құрылымы туралы түсінік бере отырып тілді тиімдірек оқытуға, үйренуге және бағалауға ықпал етеді. Дегенмен, корпустарды тілдік білімге біріктіру білім берудегі корпус лингвистикасы саласын дамытуға бағытталған үздіксіз зерттеулер мен әзірлемелерді талап етеді.

Мақала BR11765619 «Қазақ мәтінін автоматты тану: лингвистикалық модульдер мен IT-шешімдер әзірлемесі» атты бағдарламалық-нысаналы зерттеу аясында әзірленді.

Әдебиеттер

- Granger S. (1990) (ed). Learner English on computer. – London: Addison Wesley Longman. – 1990. – 250 p.
- Granger S. (1998) The computer learner Corpus: a versatile new source of data for SLA research // Learner English on Computer. – L., 1998. – 16 p.
- Добрушина Н. Р., Левинзон А. И. (2006) Информационные технологии в гуманитарном образовании: Национальный корпус русского языка // Вопросы образования. 2006. – № 4. – С. 163-179.
- Жұбанов А. (2008) Компьютерлік лингвистикаға кіріспе. – Алматы: Арыс, 2008. – 640 б.
- Жұбанов А.К. (2016) Корпустық лингвистика. – Алматы: Қазақ тілі, 2016. – 336 б.
- Мальцева М.С. (2011) «Учебный корпус как база для лингвистического и лингводидактического анализа в рамках методики преподавания иностранных языков» // Социально-экономические явления и процессы. – № 9 (031), 2011. – С. 209-212
- Национальный корпус русского языка / <https://ruscorpora.ru/>
- Соснина Е.П. (2017) К анализу лингвистических ошибок учащихся на базе учебного корпуса текстов. – 2017. – С. 39-44.
- Hinkel E. (2002) Second language writers' text. – Mahwah, NJ, 2002. – 392 p.

<https://cyberleninka.ru/article/n/uchebnyy-korpus-learner-corpus-kak-baza-dlya-lingvisticheskogo-i-lingvodidakticheskogo-analiza-v-ramkah-metodiki-prepodavaniya/viewer>

Шмелева Т.В. (2007) Задачник корпуса // Национальный корпус русского языка и проблемы гуманитарного образования / под ред. Н.Р. Добрушиной. – М., 2007. – С. 25-34.

References

Granger S. (1990) (ed). *Learner English on computer*. – London: Addison Wesley Longman. – 1990. – 250 p. (in English)

Granger S. (1998) *The computer learner Corpus: a versatile new source of data for SLA research*. – *Learner English on Computer*. – L., 1998. – 16 p. (in English)

Dobrushina N. R., Levinzon A. I. (2006) *Informacionnye tehnologii v gumanitarnom obrazovanii: Nacional'nyi korpus rysskogo iazyka*. – *Voprosy obrazovaniya*. – 2006. – № 4. – S. 163-179. [Information Technologies in Humanitarian Education: National Corpus of the Russian Language. – Education issues. – 2006. – № 4. – S. 163–179.] (in Russian)

Jūbanov A. (2008) *Kömpüterlik lingvistikaga kirispe*. – Almaty: Arys, 2008. – 640 b. [Introduction to computational linguistics. – Almaty: Arys, 2008. – 640 p.] (in Kazakh)

Jūbanov A.Q. (2016). *Korpystyq lingvistika*. – Almaty: «Qazaq tili» baspasy, 2016. – 336 b. [Zhubanov A. Q. (2016). *Corpus linguistics*. – Almaty: «Qazaq tili», 2016. – 336 p.] (in Kazakh)

Mal'ceva M.S. (2011) *Ychebny korpus kak baza dlia lingvisticeskogo i lingvodidakticheskogo analiza v ramkah metodiki prepodovaniya mostrannyh iazykov*. – *Social'no-ekonomicheskie iavlenia i processy*, № 9 (031). – 2011. – S. 209-212 [Teaching corpus as a base for linguistic and linguodidactic analysis within the framework of foreign language teaching methodology. Socio-economic phenomena and processes. – № 9 (031). – 2011. – P. 209-212] (in Russian)

Nacional'nyi korpus rysskogo iazyka [National corpus of the Russian language] / <https://ruscorpora.ru/> (in Russian)

Sosnina E.P. (2017) *K analyzy lingvisticheskikh oshibok uchashchihsia na baze uchebnogo korpusa tekstov*. – 2017. – S. 39-44. [Toward analyzing learners' linguistic errors based on a training corpus of texts. – 2017. – S. 39-44] (in Russian)

Hinkel E. (2002). *Second language writers' text*. – Mahwah, NJ, 2002. – 392 p. (in English)

<https://cyberleninka.ru/article/n/uchebnyy-korpus-learner-corpus-kak-baza-dlya-lingvisticheskogo-i-lingvodidakticheskogo-analiza-v-ramkah-metodiki-prepodavaniya/viewer> (in Russian)

Shmeleva T.V. *Zadachnik korpusa // Nacional'nyi korpus rysskogo iazyka i problemy gumanitarnogo obrazovaniya* / pod red. N.R. Dobrushinoi. – M., 2007. – S. 25-34. [Shmeleva T.V. *Corpus taskbook // The National Corpus of the Russian Language and problems of humanitarian education* / edited by N.R. Dobrushina. – M., 2007. – P. 25-34.] (in Russian)