ҒТАМР 16.31.61.49.38.49 https://doi.org/10.26577/EJPh.2022.v187.i3.09

К.Қ. Пірманова[1]*, А.Ә. Жаңабекова[2], А. Барменқұлова[2]

[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Қазақстан, Алматы қ.
[2]А. Байтұрсынұлы атындағы Тіл білімі институты, Қазақстан, Алматы қ.
*e-mail: kunsulu.pirmanova@mail.ru

ҰЛТТЫҚ КОРПУСТАРҒА НЕГІЗДЕЛГЕН ЛИНГВИСТИКАЛЫҚ ЗЕРТТЕУЛЕР ЖҮРГІЗУ (Қазақ, орыс, ағылшын тілі материалдары негізінде)

Мақалада лингвистикалық зерттеулерде корпустық материалдарды пайдалану технологиясы қарастырылады. Тіл корпустарын зерттеу мен құру сала мамандары үшін ғана емес, қоғамдық-әлеуметтік проблема ретінде де маңызды. Шетелдік және ресейлік ғалымдардың еңбектерінде корпус құрылғаннан кейін академиялық сөздіктер мен грамматикалар қайта қаралып, қайта жазылғандығы туралы ақпарат бар. Корпустық лингвистика саласының зерттеу әлеуеті зор. Сондықтан зерттеу жұмыстары үшін қазақ әдеби тілінің мәтіндері негізінде тілдік корпустарды пайдалану мүмкіндіктерін қарастыру және осы мүмкіндіктерді іске асырудың ғылыми-теориялық негіздемесін әзірлеу қажет. Өйткені, тілдік корпусты құрастыруда, атап айтқанда әртүрлі сөздіктерді құрастыруда, ғылыми грамматиканы қайта жазуда немесе белгілі бір лингвистикалық құбылыстарды анықтауда алынған нәтижелерді қолдану корпустық лингвистиканың болашақ дамуы үшін ғана емес, сонымен қатар ғылыми зерттеулердің жаңа технологиясын қалыптастыру үшін де маңызды. Корпус келесі нәтижелерге қол жеткізуге мүмкіндік береді: тілдің лексикасы, грамматика саласында зерттеулер жүргізу үшін бай тілдік материал ретінде үйрену; тілдің бірнеше кезеңдік тарихына қатысты ақпарат беру мүмкіндігі; жиілік шығару; белгілі бір кезеңге, өзгеру сатысына, дамуға қатысты тілдік бірліктердің семантикалық өрісі және т. б. туралы ақпарат; лингвистиканың барлық салаларына қатысты анықтамалық қызмет, яғни функцияларды орындау мүмкіндігі көзі; жиілік сөздіктерін автоматты түрде құру мүмкіндігі; автоматты түрде аударма негізінде екі тілді (параллель, аударылған) корпустар және т. б.; екі тілді түсіндірме сөздіктерді құру мүмкіндігі; грамматикаларды қайта жазу мүмкіндігі; тілді оқыту немесе үйрену мүмкіндігі, яғни оқулықтар мен оқу бағдарламалары үшін негіз құру жөніндегі қызмет және т.б. Мақала тақырыбы бойынша жасалған зерттеу жұмысы осы жоғарыда санамаланып көрсетілген нәтижелерді жүзеге асырудың ғылыми-теориялық және инженерлік технологиясын айқындап береді. Корпус материалдары бойынша грамматикалық, дискурстық, лексикографиялық зерттеулер жүргізу мүмкіндіктері сипатталады.

Түйін сөздер: корпус, лексикография, грамматика, дискурстық зерттеу, жиілік, эмперикалық қолдау.

K.K. Pirmanova[1]*, A.A. Zhanabekova[2], A. Barmenkulova[2]

[1]Al-Farabi Kazakh National University, Kazakhstan, Almaty
[2]A. Baitursynov Institute of Linguistics, Kazakhstan, Almaty
*e-mail: kunsulu.pirmanova@mail.ru

Conducting linguistic research based on national corps (based on the materials of the Kazakh, Russian, English languages)

The article discusses the technology of using corpus materials in linguistic research. The study and creation of language corpora is important not only for industry specialists, but also as a socio-social problem. The works of foreign and Russian scientists contain information that after the creation of the corpus, academic dictionaries and grammars were revised and rewritten. The field of corpus linguistics has a great research potential. Therefore, for research work, it is necessary to consider the possibilities of using language corpora based on the texts of the Kazakh literary language and develop a scientific and theoretical justification for the implementation of these possibilities. After all, the use of the results obtained in the compilation of a language corpus, in particular in the compilation of various dictionaries, rewriting scientific grammar or identifying certain linguistic phenomena, is important not only for the future development of corpus linguistics, but also for the formation of a new technology of scientific research. The corpus allows you to achieve the following results: the study of language as a rich language material for research in the field of vocabulary, grammar; the ability to provide information related to

several periodic histories of the language; identification of frequency; information about the semantic field of language units related to a certain period, stage of change, development, etc.; background information related to all areas of linguistics, i.e. the ability to perform functions; the ability to automatically create frequency dictionaries; bilingual (parallel, translated) corpus based on automatic translation, etc.; the possibility of creating bilingual explanatory dictionaries; the possibility of rewriting grammars; the possibility of teaching or learning a language, i.e. activities to create the basis for textbooks and curricula, etc. The research work on the topic of the article defines the scientific-theoretical and engineering technology for the implementation of the above results. Based on the materials of the corpus, the possibilities of conducting grammatical, discursive, lexicographic research are described.

Key words: corpus, lexicography, grammar, discourse research, frequency, empirical support.

К.К. Пирманова*[1], А.А. Жанабекова[2], А. Барменқулова[2]

[1]Казахский национальный университет имени аль-Фараби, Казахстан, г. Алматы
[2]Институт языкознания имени А. Байтурсынова, Казахстан, г. Алматы
*e-mail: kunsulu.pirmanova@mail.ru

Проведение лингвистических исследований на основе национальных корпусов (на основе материалов казахского, русского, английского языков)

В статье рассматривается технология использования корпусных материалов в лингвистических исследованиях. Изучение и создание языковых корпусов важно не только для специалистов отрасли, но и как общественно-социальная проблема. В трудах зарубежных и российских ученых содержится информация о том, что после создания корпуса были пересмотрены и переписаны академические словари и грамматики. Область корпусной лингвистики имеет большой исследовательский потенциал. Поэтому для исследовательской работы необходимо рассмотреть возможности использования языковых корпусов на основе текстов казахского литературного языка и разработать научно-теоретическое обоснование реализации этих возможностей. Ведь использование полученных результатов при составлении языкового корпуса, в частности при составлении различных словарей, переписывании научной грамматики или выявлении тех или иных лингвистических явлений, важно не только для будущего развития корпусной лингвистики, но и для формирования новой технологии научных исследований. Корпус позволяет достичь следующих результатов: изучение языка как богатого языкового материала для проведения исследований в области лексики, грамматики; возможность предоставления информации, относящейся к нескольким периодическим историям языка; выявление частотности; предоставление информации о семантическом поле языковых единиц, относящихся к определенному периоду, этапу изменения, развитию и т. д.; справочной информации, относящейся ко всем областям лингвистики, т. е. возможность выполнения функций; возможность автоматического создания частотных словарей; двуязычные (параллельные, переводные) корпуса на основе автоматического перевода и др.; возможность создания двуязычных толковых словарей; возможность переписывания грамматик; возможность обучения или изучения языка, т. е. деятельность по созданию основы для учебников и учебных программ и т.д. Исследовательская работа по теме статьи определяет научно-теоретическую и инженерную технологию реализации перечисленных выше результатов. По материалам корпуса описываются возможности проведения грамматических, дискурсивных, лексикографических исследований.

Ключевые слова: корпус, лексикография, грамматика, дискурсное исследование, частота, эмпирическая поддержка.

Кіріспе

Тілдік корпустар жасау мәселесі қазақ тіл білімінде күні бүгінге дейін толық шешімін таппай келе жатқан өзекті мәселелердің бірі. Әлем тілдеріндегі компьютерлік (қолданбалы) лингвистика жетістіктерін ұлттық тіліміздің қажетіне пайдалану осы сала мамандарының алдында тұрған жауапты іс.

Тілдік корпустарды зерттеу мен жасау сала мамандары үшін ғана емес, қоғамдық-әлеуметтік мәселе ретінде де аса маңызды. Шетел, орыс ғалымдары еңбектерінде корпус жасалғаннан кейін академиялық сөздіктер мен грамматикаларының қайта құрастырылып, қайтадан жазылып шыққандығы туралы мәліметтер берілген. Корпустық лингвистика саласының ғылыми-зерттеу әлеуетін орасан зор. Сондықтан қазақ әдеби тілі мәтіндері бойынша тілдік корпустарды ғылыми-зерттеу жұмыстарына пайдаланудың мүмкіндіктерін қарастырып, сол мүмкіндіктерді жүзеге асырудың ғылыми-

теориялық негіздемесін жасау қажет. Өйткені тілдік корпустарды құрастыру барысында алынған нәтижелерді қолданысқа енгізу, нақты айтқанда, әртүрлі сөздіктер құрастыру ісінде, ғылыми грамматикаларды қайта жазуда немесе қандай да бір тілдік құбылыстарды айқындауда пайдалану корпустық лингвистиканың болашақ дамуы үшін ғана емес, ғылыми-зерттеу жүргізудің жаңа технологиясын қалыптастыру үшін де өзекті.

Корпустарды пайдаланушыларды, ең алдымен, тілшілерді, әдетте, нақты мәтіндердің мазмұнынан гөрі, олардың метамәтіндік ақпараты және қайсыбір тілдік элементтері мен олардың құрылымдық қолданыстарының мысалдары көбірек қызықтырады. Корпустар арқылы жүргізілген ең алғашқы тілдік зерттеулер әртүрлі тілдік элементтердің мәтіндегі қолдану жиіліктерін анықтаумен ғана шектелетін. Статистикалық әдістер машиналық аударма, сөйлеу тілін танып білу (разпознавание) мен оны синтездеу, орфография мен грамматиканы тексеретін құралдар және т.б. осы сияқты күрделі лингвистикалық мәселелердің шешімін табуда қолданылады. Мәселен, корпус материалында статистикалық әдістерді қолданып, қай сөздер әрдайым бірге қолданатынын білуге болады, бұл оларды тұрақты сөз тіркестеріне жатқызуға болатындығының айғағы деуге болады. Семантикалық жағынан алғанда, тұрақты сөз тіркестері тұтас сипаттағы (бөлінбейтін) мағыналық бірлік, ал мұндай жағдай лексикография саласы мен мәтінді автоматты өңдеу жүйелерінде ескерілуі өте маңызды.

Сонымен қатар лексикография мен грамматиканы зерттеуде корпустар аса бай дереккөз болып табылады. Лексикографиялық зерттеулермен семантика саласындағы зерттеулер өте тығыз байланыста болып келеді. Корпустағы қайсыбір лингвистикалық бірліктің қоршауын бақылау негізінде ондай тілдік бірлікті сипаттайтын семантикалық белгілерін анықтауға болады.

Тілші-теоретиктер корпустарды өз болжамдарын тексеруге және теорияларын дәлелдеуге қажетті құрал ретінде қолданады. Қолданбалы бағытта жұмыс істейтін лингвистер (мұғалімдер, аудармашылар және т.б.) компьютерлік корпустарды тілдерді үйретуге және өздерінің кәсіби мәселелерін шешуге пайдаланады. Корпустарды пайдаланушылардың айрықша тобын компьютерлік лингвистер құрайды: олар компьютерлік тілдік модельдерді құрастыру үшін мәтінде орын алатын статистикалық және лингвистикалық заңдылықтарды айқындау мен пайдалануды мақсат етеді. Тілге қатысты басқа да мамандар (әдебиетшілер, редакторлар) корпус арқылы өздерін қызықтыратын сұрақтар бойынша қанағаттандырарлық жауап ала алады. Қоғамдық ғылымдар салаларының мамандары (тарихшылар, социологтар) өздерінің зерттеу нысандарын тіл арқылы, яғни кезең, автор немесе жанр деп аталатын мәтін параметрлерін қолдану арқылы зерттеулеріне мүмкіндігі бар. Әдебиетшілер корпусты стильдердің ерекшеліктерін зерттейтін ғылыми зерттеулер үшін пайдаланады. Корпустар әртүрлі автоматтанған жүйелерді (машиналық аударма, сөйлеу тілін тану, ақпараттық іздеу) зерттеу үшін пайдаланылады.

Зерттеу материалы және әдістері

Мәтінге грамматикалық белгіленім енгізу әдістері, лингво-статистикалық, талдау және жинақтау әдістері, логика-семантикалық, диструбутивтік, алгоритмдер теориясы мен компьютерлік база құруға қатысты әдістер және т.б.

Корпустық деректер теориялық лингвистикаға қандай септігін тигізеді деген сұраққа келсек, әрине, олар тілшілердің лексика мен грамматика жайындағы пайымдауын да алмастыра алмайды, бірақ олар мамандарға бай репрезентативтік эмпирикалық материал береді. Түптеп келгенде, корпустар тілдік зерттеулер барысында пайдалануға болатын үш типті деректер бере алады: *эмпирикалық қолдау, жиілігі бойынша ақпарат, экстралингвистикалық ақпарат (метаақпарат).* Осы аталған деректер типтерін мүмкіндігінше толығырақ қарастырайық.

1. Эмпирикалық қолдау. Көптеген тілшілер корпусты «мысалдар банкі» ретінде пайдаланады, яғни корпустан өз зерттеулеріндегі болжамдарына, қағидаттары мен ережелеріне эмпирикалық қолдау табуға ұмтылады. Табылған мысалдар, әрине, ойдан шығарылуы немесе кездейсоқ табылуы мүмкін, бірақ корпустық лингвистика тәсілі тілдік материалдың репрезентаттығы мен теңгерімділігін, сондай-ақ қандай да бір корпусты пайдаланушының таңдау мүмкіндігін туғызатын іздеу құралын қамтамасыз етеді.

Сондай-ақ кейбір ғылыми гипотезалардың кераған екенін дәлелдейтін материалдар алуға мүмкіндік береді, яғни ғылыми теориялардың шынайылық сипатын дәйектейді.

Орыс тілінің ұлттық корпусын пайдалана отырып, ғалым Н.В. Перцов орыс тілінің беделді тілшілерінің ой-пікірлерін былайша теріске шығарады: «...Следует признать, что возможно-

сти корпусов все-таки еще недостаточно усвоены лингвистической общественностью вообще и лингвистами в частности. Обращение к корпусным данным еще не стало столь же привычным и обязательным при формулировке и проверке тех или иных утверждений относительно фактов языка, как обращение к грамматикам и словарям, к работам коллег» (Перцов, 2006: 318). Тіпті корпустарды кең қолданып жүрген авторлардың жұмыстарының өзінде тілдік деректер жайындағы тұжырымдар корпустық деректерге қайшы келіп жатады.

Әрбір тілдік деңгейдегі, сөз дыбыстарынан бастап, тұтас әңгіме мен мәтінге дейінгі жасалған болжамдарды анықтайтын дәлелдер корпустардан табылуы мүмкін. Өзін-өзі бақылау кезінде мүмкін болмайтын жағдайларды корпус материалдары бойынша құрылым ішінен қайта талдап, нәтижелерді қайта жаңғыртуға болады (Перцов, 2006).

2. Жиілігі бойынша ақпарат. Корпусты қолданудың сапалы әдісі эмпирикалық қолдау түрінде көрініс табады да және, сонымен бірге, корпустар сандық зерттеулер жүргізу үшін сөздердің, фразалар мен сөз тіркестердің қолдану жиілігі туралы ақпарат бере алады. Сандық зерттеулер (әрине олар көбінде сапалық талдауға негізделеді) теориялық және компьютерлік лингвистиканың көптеген аясында қолданыс табады. Олар әртүрлі сөйлеушілер топтарының немесе әртүрлі мәтіндер типтерінің арасындағы ұқсастық пен айырмашылықтарды көрсетеді, психолингвистикалық және тағы да басқа зерттеулер үшін деректердің қолдану жиілігін айқындауға мүмкіндік береді.

3. Метаақпарат. Тілдік контекстке қосымша ретінде, мәтіндер корпусы сөйлеушінің не жазушының жасы немесе жынысы, мәтін жанры, мәтіннің пайда болуы жайындағы уақыттық немесе кеңістіктік және т.б. туралы экстралингвистикалық ақпарат немесе метаақпарат береді. Мұндай метаақпарат әртүрлі мәтіндер типтері мен әртүрлі сөйлеушілер топтарын салыстыруға мүмкіндік жасайды.

Көптеген ғалымдардың пайымдауынша, корпустық лингвистика – лингвистиканың жеке парадигмасы дегеннен гөрі, оның әдіснамасы (әдістемесі) деген дұрыстау келеді. Мәселен, ағылшын тілінің көптеген белгілі корпустары тіл ғылымының әртүрлі бағыттарының өкілдері жүргізетін арнайы зерттеулер үшін құрастырылған және қолданылған. Мысалы, *CHILDES* корпусы әртүрлі коммуникативтік жағдаяттағы балалардың ауызша тілі транскрипті арқылы балалардың тілді игеру қабілетіне қызығушылық білдіретін ғалымдардың психолингвистика саласындағы зерттеулерінде кең қолданыс табады (McWhinney, 2000). Ағылшын тілінің алғашқы кезеңдерінен бастау алатын жазба мәтіндерінің әртүрлі типтерінің *Хельсинкалық корпусы* тіл тарихының даму үдерісін зерттеу үшін қолданылады (Kytö, 1992). Лондондық жасөспірімдердің *ағылшын тілінің Бергендік корпусы* COLT (The Bergen Corpus of London Teenage Language) 13-17 жастағы жасөспірімдердің сөйлеу тілі мәтіндерін жинақтаған және ол әлеуметтік лингвистика саласы бойынша белгілі жастар топтарының тілдерін зерттеу үшін пайдаланылады (Stenström, 1996). Өз зерттеулерін сапалы нәтижелер алуға мүмкіндік беретін «нақты» тілдік материалдар бойынша жүргізілетін тілдік талдаулардың ұтымдылығы корпустарды пайдаланатын тілшілердің қызығушылығын арттырады (Meyer, 2002).

Әдебиетке шолу

Зерттеу жұмысы әлемдегі тілді компьютерлік технология бойынша зерттеудің бірнеше бағытын әдіснамалық негіз етіп алады. Атап айтқанда, 1) шетел және орыс ғалымдарының компьютерлік және корпустық лингвистика саласына қатысты теориялық бағыты (Л.Н.Засорина, А.П.Ершов, В.Б.Касевич, А.В.Венцов, Е.В.Грудева, Е.В.Ягунова, А.Н. Баранов, С.А.Шаров, И. М.Богуславский т.б. еңбектері, сонымен қатар ең алғаш 1960 жылы Америкада жасалған Браундық корпус, 1980 жылы Швецияда Уппсал университетінде жасалған мәтіндік корпус, Прагада Карл университетінде жасалған Чех ұлттық корпусы, 2001 жылы басталып, 2004 жылдары интернет-сайтқа енгізілген «Орыс әдеби тілінің ұлттық корпусы» т.б); 2) Қазақ тіл білімінде морфология саласына қатысты лингвостатистикалық зерттеулер жүргізген Қ.Б.Бектаев, А.Ахабаев, А.Қ.Жұбанов, С.Мырзабеков, Д.Байтанаев, Қ.Молдабеков, А.Белботаев т.б. ғалымдар негізін салған лингвостатистикалық бағыт және жиілік сөздік құрастыру тәжірибесі; 3) лексикографиялық зерттеулер тәжірибесіне қатысты И.Хамдамова, А.А.Акабиров, В.Месгудов, С.Алтаев, Д.Тачмурадова, М.Исмаилова, С.Омуралиева, Н.Суфьянова, М.Равшанов, Ж.М.Гузевтің сияқты түркітанушылардың, С.Жиенбаев, І.Кеңесбаев, К.Аханов, Ә.Болғанбаев, Б.Сүлейменова, Б.Қалиев, С.Бизақов, М.Малбақов сияқты қазақ

лексикограф-мамандарының лексикография мәселелеріне арналған ғылыми еңбектерін атаған жөн.

Нәтижелер және талқылау

Корпустарға негізделген лексикографиялық зерттеулер

Лексикографиялық зерттеулер біріншіден, сөздіктер құрастыру үшін қажет болса, екіншіден, дескриптивті және қолданбалы лингвистика үшін де қажет. Мәселен, зерттеу жүргізудің алдында лексикографтардың қандай ақпаратқа мұқтаж екендігін біліп алу қажет. Мысалы, орыс тілінің академиялық түсіндірме сөздігін құрастырудағы авторлардың негізгі сұранысы мынадай ақпараттарды тауып алу қажеттігін тудырған:

- белгілі бір кезеңдерде пайда болған жаңа сөз;
- сөздің бастапқы формасы;
- мағыналары белгілі сөздерді ашатын мысалдар (цитаталар);
- сөздікте мысал келтірілмеген мағыналарға мысал табу;
- қайсыбір мағынаға қосымша берілетін жаңа мысалдар;
- лексикалық және синаксистік тіркесімдердің жаңа түрлері;
- жаңа фразеологизмдер;
- арнайы терминдердің қазіргі кезге сай жаңа ғылыми түсініктемесі (Гришина, 2008).

Грамматикалық және лексикографиялық модельдер жүйелі түрде өзара байланыста болады. Мысалы, дәстүрлі әдістер синонимдік сөздер тобын анықтай алатын болса, лексикографиялық зерттеулер корпустық тәсілдерді қолдану арқылы өзара қатынастағы сөздердің әртүрлі жағдаятта және әртүрлі контекстерде қалай қолдануға болатындығын көрсету мүмкіндігі бар.

Д. Байбера, С. Конрад, Р. Реппендердің «*Corpus Linguistics*. Investigating language structure and use» атты оқулығында (Савчук, 2005) корпустық тәсіл арқылы жұмыс істейтін зерттеуші-лексикографтардың алдарында тұрған негізгі алты мәселені жеке қарастырады:

1. Нақты сөз бойынша қандай мағыналар ассоциацияланады?
2. Белгілі бір сөздің қолдану жиілігі оған жақын тұрған басқа сөздердің жиілігіне қарағанда қандай деуге болады?
3. Берілген сөздің қандай тілдік емес модельдері бар (регистрге, тарихи кезеңдерге, диалектілерге және т.б. қатысты)?
4. Берілген сөзбен қандай сөздер әдетте бірге кездеседі және әртүрлі регистрлерде олардың тіркесіп қолдану реттілігінің үлестірімі қаншалықты?
5. Сөз қолданысының мағынасы мен типтері қалай үлестірілген?
6. Синоним сөздер қалай пайдаланылады және қалайша әртүрлі контекстерде өз орнында қолданылады? (Biber, 1998).

Лексикографиядағы корпустық тәсілмен зерттеудің ұтымды жағы – корпустан берілген сөз қолданылатын контекстер жиынын кең түрде алуға болатындығы. Әрі қарай конкорданстық тізімдерді (KWIC) қолдана отырып, контекстерден сол сөзді ассоциациялайтын әртүрлі мағыналарды бөліп алуға болады.

Корпустарды құрастырудың тағы бір мақсаты – зерттеушінің өз зерттеу аясындағы еңбегін жеңілдету. Мысалы, корпус тек күрделі зерттеу саласы мәтіндерінің жиынтық бөлігі қана емес, сонымен бірге ол, мүмкіндігінше, көлем жағынан да айтарлықтай айырым табуы қажет.

Көлемді корпустардың конкорданстары да өте үлкен болады. Соған сәйкес өте көп деректер бере алады. Конкорданстар көлемі үлкен корпустарда көлемі бірнеше мың бетке жетуі мүмкін, тіпті бір мысалға жүздеген конкорданстар сәйкес келуі мүмкін (Баранов, 2007). Мысалы, көлемі 8 миллиондық Лонгман-Ланкастер подкорпусында (ішкі корпусында) *deal (дело)* сөзіне 1500-ден аса қолданысты (мысалды) ұсынған, ал мұндай көп мәлімет оларды мағынасына қарай топтауды немесе маңыздылығына қарай сұрыптауды қиындатады. Мұндай жағдайда қосымша құралдар мүмкіндігін пайдаланылған. Айталық, программалық конкорданстардың көпшілігі сөздердің қолдану жиілігі бойынша сөздер тізімін жасай алады. Ондай тізімдер әдетте әліппи тәртібімен, кездесу тәртібімен немесе жиіктерінің кему не өсу тәртібімен жасалады. Одан басқа көпшілік қауымға белгілі **Sketch Engine** жүйесі құрылымдық-синтаксистік модель бойынша реттелген шектеулі сөзтіркестерінің статистикалық жиынын (коллокацияларды) тауып бере алады. Енді *Corpus Linguistics* оқулығы бойынша *deal* сөзімен мағынасы ұқсас сөздерді табуға мысалдар келтірейік.

Сөзердің мағынасын талдауға қиындық тудыратын жайт, ол – ағылшын тілінде көптеген сөзформалардың грамматикалық қызметтерінің көп болуы. Мәселен, *deals* сөзформасының 3-ші жақтағы жекеше формадағы етістік ретінде қолданылуы және көптік формадағы

зат есім түрінде де қолданылуы. *Deal* және *dealing* сөзформалары етістік ретінде де және зат есім түрінде де қолданыла алады. Аннотация жасалмаған корпус деректері бойынша құрастырылған жиілік сөздіктердің тиімділігі шектеулі болуының себебі – олардың қай сөздердің грамматикалық қолданысының жиі, ал қай қолданысының сирек екендігін ажырата бермеуінде.

Deal сөзформасының зат есім ретінде неше рет кездесетінін және неше рет етістік ретінде қолданылатынын анықтау үшін ең алдымен олардың контекстегі формасына қарай, сөзформалардың грамматикалық категорияларын анықтап алу қажет, содан кейін барып әрбіреуінің қолдану жиілігін есептеп шығаруға болады.

Ондай корпуста әрбір сөзформаға грамматикалық категориялары бойынша шартты түрдегі арнайы белгіленім жүргізілгендіктен, оларды автоматты түрде ажыратып, жиіліктерін есептеуге мүмкіндік жасалады. Мұндай әр сөз табына жататын омоним сөздердің екі түрлі позицияда қолданысын қолдап санап шығу қиын. Мысалы, қазақ тіліндегі «қара» сөзінің белгілі бір көлемдегі мәтінде неше рет етістік, неше рет сын есім мәнінде қолданылғанын анықтау үшін мәтінге алдымен автоматты морфологиялық белгіленім қою бағдарламасы іске қосылады. Табылған омонимдер қолдап ажыратылады немесе белгілі бір контекстік қоршауларға негізделіп, омоним ажыратудың формальдық межесі анықталады. Содан кейін барып жиілікті анықтауға мүмкіндік туады.

Лонгман-Ланкастер корпусында басқа базаларға қарағанда ***deal*** мен ***deals*** айтарлықтай жиі кездеседі және бұл қомақты база бұл сөздердің қолданысын талдауға мүмкіндік береді. Аталған кестедегі деректер бойынша, жиілік көрсеткіштері қызықты модельдер табуға негіз болады. Біріншіден, мәтіндердің барлық мысалдары үшін нормалданған есептеулер (толық қамтитын есептеулер) бойынша, етістік ретіндегі ***deal/deals*** зат есім ретіндегі осы сөздердің қолдану жиілігінен тек аз ғана айырым табады (етістік ретінде – 1 млн.-ға шаққанда 119 сөз, ал зат есім ретінде – 1 млн.-ға 90 сөз). Егер ол сөздердің кездесу жиілігін регистр бойынша қарастыратын болсақ, басқа көрініс байқаймыз. Мысалы, ғылыми әдебиет мәтіндерінде ***deal/ deals*** сөздер етістік ретінде қолдануы олардың зат есім ретінде қолдануымен салыстырғанда екі рет жиі қолданылатынына көзімізді жеткізуімізге болады (етістік – 176, ал зат есім – 74 рет 1 млн. сөзге шаққанда). Ал көркем проза мәтіндеріндегі ***deal/deals*** сөздерінің қолданысы жоғарыда айтылғандарға қарағанда, қарамақарсы модельдің көрінісін бейнелейді, яғни ол сөздердің зат есім ретіндегі қолданысы етістік ретіндегі қолданысына қарағанда анағұрлым жиі (зат есім – 107, ал етістік – 63 рет 1 млн. сөзге шаққанда).

DEAL сөзінің қолдану моделі корпусты құрастыруға қатысты басқа да бір маңызды жағдайды сипаттайды: тек бір ғана регистрмен шектелген корпус тілді басқа регистрлерде бейнелей алмайды. Мысалы, бір регистрдегі материалдар бойынша құрылған модель негізінде басқа регистрден табылған сәйкес материалдар үшін жалпылама қорытынды жасауға болмайды. ***DEAL*** сөзінің қолдануына қатысты мысалдар негізінде мынадай қорытынды жасауға болады: ***deal*** сөзінің зат есім және етістік ретінде ғылыми әдебиетте қолдануының қатынастық жиіліктері олардың көркем прозадағы қолдануының қатынастық жиілігіне қарағанда толығымен қарамақарсы сипатта деуге болады. Кез келген осы регистрлермен шектелген корпус, басқа регистр бойынша табылған мәліметтерді тіпті де көрсете алмас еді және ол сөздің тілдік қолданысының моделі дұрыс құрылмаған болар еді.

Регистр бойынша мағыналардың (мәндерінің) үлестірімі. Корпустар конкорданстарды пайдалана отырып, сөздердің мағынасын ашуға байланысты зерттеулер жүргізуге мүмкіндік тудырады. Сөздер мағынасын ашуға байланысты зерттеуді олардың ***коллокаттарын*** талдаудан бастауға болады, яғни талдауға тиісті сөздің жиі тіркесетін сөзінен бастайды. Әрбір коллокация үшін бір мәнмен немесе мағынамен ассоциациялаудың күшті үрдісі орын алады (бірнеше сөзтіркестері бір мағынамен ассоциациялануы мүмкін). Сондықтан, сөздің ең жиі коллокациясын бөліп алып, сөз мағыналарына тиімді және дәлелді (сенімді) талдау жүргізуге болады. Әрі қарай корпустағы зат есім ретіндегі DEAL сөзінің, оның сөздіктегі дефинициялары-мен (тұжырымды анықтамасымен) бірге, коллакаттарын талдау нәтижелерімен салыстыру қажет болады.

Ғылыми әдебиетте, әлбетте, ***good deal*** және ***great deal*** коллокациялары үлкен мөлшердегі нәрсені немесе бизнестегі қайсыбір операцияны білдіреді. Оң жақты коллокаттарды қарастырған кездерде ***deal*** сөзінің бір нәрсенің санына қатысты мағыналары ең көп кездесетіні анықталды. Мұны ***of*** сөзінің оң жақты коллокатының жиілігі де растайды (1 млн. сөзге

39 рет). Келесі коллокаттың жиілігі анағұрлым кіші: ***more*** (7 рет 1 млн. сөзге), ***in*** және ***to*** (3 рет 1 млн. сөзге). Сонымен, зат есім ретіндегі ***deal*** сөзі ***a good/great deal of*** сөз тіркесіндегідей сандық мағынада ең жиі қолданылады.

Сонымен, лексикографиялық жұмыс келешектегі екі мүмкіндікті де қамтуы қажет: мағыналардың барлығын да бөліп алып қарастыру, бірақ олардың регистрлік қатыстылығына қарай, ең жиілеріне немесе аса маңыздыларына сілтеме жасап отыру.

Қазақ тілінің онтомдық түсіндірме сөздігін жасауда алғашқыда картотекалық қорда жинақталған 5 миллион кәртішке пайдаланылған. Кейін 15 томдық сөздікте бұрынғы қол жұмысы жартылай автоматтандырылып, сөздік жасау жұмысын әлдеқайда тездеткен болатын. Онтомдық сөздікте 2 мағынасы ғана көрсетілген көптеген сөздің 15 томдық сөздікте 10-нан аса мағынасы ашылып беріледі. Мұндай мағыналық толықтырулар мысалдар контекстері арқылы ашылған. Енді корпус материалдары бойынша жинақталатын конкорданстар түсіндірме сөздіктердегі сөз саны мен сөз мағыналары ашуда теңдесі жоқ қуатты құрал болатыны даусыз.

Жоғарыда ағылшын тіліндегі әртүрлі стильдер бойынша ізделген сөздердің оң жақ және сол жақ контекстік қоршаулары бойынша тіркесімділігі анықталып, мағыналары айқындалған. Осы сияқты қазақ тілінде Ұлттық корпус материалдары негізінде әрбір стиль бойынша сөз мағынасын ашудың әдістәсілдерін табуға болады. Ал бұл болашақ лексикографиялық жұмыстарды автоматтандыру дегенге саяды.

Коллокацияларды статистикалық әдістермен бөліп алу. Лексикалық тіркесімді талдау үшін корпустық әдістерді қолдану сөздіктердің жаңа түрлерін құрастыруға, соның ішінде тұрақты сөз тіркестерінің сөздігін құрастыруға мүмкіндік береді. Корпустарды пайдалану лексикалық бірліктердің тіркесіп қолдануы жайлы деректерді, яғни олардың тіркесімдік, меңгерімдік және т.б. ерекшеліктерін алуға мүмкіндік жасайды. Қазіргі кездегі қолданыста бар тұрақты сөз тіркестерінің сөздіктерінде, біріншіден, сөз тіркестері толық түрде қамтылмаған, екіншіден, оларды айтарлықтай жүйелі түрде жасалған деуге келмейді. Сол себепті қазірге кезде сөздіктің жаңа типін, яғни тұрақты сөз тіркестері сөздіктерінің біріктірілген немесе коллокация сөздігі деп аталатын сөздігін жасауға қажеттілік туындайды. Мұндай сөздік, шындығында да, әртүрлі тұрақты сөз тіркестерін қамтиды (фраземалар, немесе түсіндірме сөздіктерде ромбадан кейін берілетін сөз тіркестері) (Хохлова, 2008).

Қазіргі кезде тіл білімінде коллокацияның қайсыбір бөліктерінің байланыстылық дәрежесін анықтайтын бірнеше тәсілі бар. Ондай статистикалық өлшем ретінде ассоциация өлшемін (*MI*, *t-score*, *log-likelihood*) таңдап алуға болады. Олар көбінде корпустағы сөз тіркестері құрамдарының өзара жақындық дәрежесін есептеу кезінде қолданылады. Биграмма (берілген сөздің сол жақ немесе оң жағындағы сөздермен тіркесімі) бөліктерінің сызықтық жақындығы тұрақты тіркесімді табу әрекетінде, сол сияқты мәтіндегі коллокацияларды және сөз тіркестерінің басқа да түрлерін іздеп табу кезінде аса маңызды алғышарттық рөл атқаруы мүмкін.

Орыс тілі үшін статистикалық әдістерді қолдануға болатындығын тексеру үшін және жоғарыда көрсетілген шаралар негізінде коллокацияларды бөліп алу мүмкіндігін айқындау мақсатымен бірнеше тәжірибе жүргізілген. Зерттеу жұмысы CQP (http://corpus1.leeds.ac.uk/ ruscorpora.html) корпус-менеджердің көмегімен орыс тілінің газеттер мәтіні (The corpus of Russian newspapers) корпусы базасы негізінде жүзеге асқан. Аталған корпус С.А.Шаровтың ғылыми жетекшілігімен Лидса университетінде (Ұлыбритания) 2001-2004 жылдары жарық көрген және 78 млн. сөзқолданыстан тұратын көлемдегі газеттер мәтіні негізінде құрастырылады. Зерттеу материалына 19 зат есім сөздерден тұратын коллокация негіз болған және олар келесі аталатын қағидат бойынша таңдалып алынады. Ең алдымен С.А. Шаровтың орыс тілінің электрондық жиілік сөздігінен бірінші мыңдыққа енетін ең жиі жиіліктегі зат есім сөздер таңдап алынады. Әрі қарай, Шағын (кіші) академиялық сөздік (Малый академический словарь – МАС) (Словарь русского языка 1981-1984) бойынша таңдалып алынған сөздердің қолдану жиіліктерін бұрмалайтын омонимдік сыңарларының бар-жоғы тексеріледі (мысалы, екі мағынадағы *брак* сөзі; *друг другa* деген сияқты сөздер леммаға келтіру кезінде бір лемма түрінде көрініс табады). Омонимдік сыңарлары бар сөздер тізімнен алынып тасталып, тәжірибеде қарастырылмаған. Сосын тізімде қалған зат есім сөздер Е.Г.Борисованың орыс тілінің коллокациялар сөздігімен (Баранов, 2003) салыстырылып қаралады. Егер қарастырылып отырған сөзге қатысты оның тіркесімдігі жайлы сөздік мақаласында ақпараттың жоқтығы неме-

се шектеулігі байқалса, ондай сөздер де тізімнен алынып тасталған. Сонымен, тірек сөздердің мынадай тізімі алынған: *власть*, *внимание*, *возможность*, *война*, *вопрос*, *дождь*, *жизнь*, *закон*, *любовь*, *место*, *мнение*, *мысль*, *ночь*, *ответ*, *помощь*, *радость*, *слово*, *случай*, *смысл*. Төменде 1-кестеде *война* тірек сөзімен бірге алғашқы 10 коллокацияның (106-дан) деректері келтірілген. *Война* тірек сөзі MI өлшемінің мағынасы (ортақ ақпарат көлемі) бойынша сұрыпталған. Мұндағы *Joint* – корпустағы берілген коллокацияның абсолютті жиілігі; *Freq1* – биграмманың бірінші сөзінің абсолютті жиілігі, яғни *война* сөзі үшін сол жақтық коллокат; *LL score, MI, T-score* – log-likelihood өлшем мәні, берілген коллокация үшін MI және t-score. Тірек сөздердің тізімінен мынандай жайтты байқауға болады, бір жағынан, тізімде тұрақты сөз тіркестері орын алса, екіншіден, барынша жоғары MI өлшемінің көрсеткішіне де ие екендігіне көз жеткізуге болады. Зерттеу нәтижелері бойынша, MI өлшемінің мәндері 0-ден 1-ге дейінгі аралықта тұрақты сөз тіркестеріне жатқызуға болатын тіркестер кездеспеген. Бұдан мынадай қорытынды жасауға болады, MI ассоциациясының өлшемдік мәні аталған интервал арасына түсетін тіркестер статистикалық тұрғыда онша мәнді емес деуге болады. Барлық жинақталған тіркестер бойынша бәріне бірдей мынадай үрдіс байқалады: өлшемнің мәні неғұрлым аз болса, соншалықты ықтималдық көп болады және орыс тілі сөздіктерінде мұндай сөз тіркестері тұрақты деп бекітілмеген. Сондықтан, сөздіктерде берілген тіркесімдік туралы деректер ассоциация өлшемі негізінде алынған деректермен сәйкес келеді. Сөздіктерде келтірілген көптеген коллокациялар (фраземалар) ассоциация өлшемі негізінде түзілген тізімнің жоғарғы жағында орналасады екен. Ал бұл жайт – коллокация деректерінің байланыстылық көрсеткіші мәнінің жоғары екендігінің айғағы.

1-кесте – «Война» сөзіне қатысты ассоциация өлшемінің мәндері (сол жақ контекст)

Коллокация	***Joint***	***Freq1***	***LL score***	***MI***	***T-score***
необъявленный война	9	76	30,19	11,03	3,00
междоусобный война	4	54	12,43	10,35	2,00
партизанский война	45	728	135,77	10,09	6,70
рельсовый война	6	100	18,00	10,05	2,45
победоносный война	9	174	26,31	9,84	3,00
вялотекущий война	6	142	16,92	9,54	2,45
позиционный война	5	128	13,90	9,43	2,23
холодный война	171	4747	469,90	9,31	13,06
грянуть война	14	457	37,19	9,08	3,73
финляндский война		148	10,37	8,90	2,00

Жүргізілген тәжірибенің нәтижесінде бірде-бір сөздікте кездеспеген сөз тіркестердің мәтіндерден бөлініп алынғаны белгілі болды, ал бұл жайт аса маңызды екендігін айта кетпекпіз. Мұндай тіркестерді талдаудың нәтижесінде байқалғаны, ол тізімнің бас жағында орналасқан биграммалар (қайсыбір өлшемнің кему тәртібі бойынша сұрыпталған), біршама ықтималдықтар үлесін құрайтын тіркесім сөздер тұрақты тіркестерге жататындығы анықталды, сол себепті оларды сөздікке енгізуге болады. Сөз тіркестері тізімінің төменгі жағында орналасқан сөз тіркестері көпшілік жағдайда, еркін сөз тіркестері қатарына жататындығы белгілі болды. Түсіндірме сөздіктерде ромбадан кейін келтірілген сөз тіркестер қайсыбір дәрежеде тұрақтылыққа ие болғанымен, оларды толыққанды тұрақты сөз тіркестері қатарына жатқызуға болмайды. Жүргізілген тәжірибе нәтижесі бойынша шығатын қорытынды, ол біріншіден, сипатталған статистикалық тәсілдерді лексикография практикасында қолдануға болады, екіншіден, олар қолданыстағы сөздіктердің тұрақты сөз тіркестерін толық қамтымайтынын көрсетеді.

Мамандандырылған корпустарда коллокацияларды айқындау аса үлкен практикалық

маңыздылыққа ие болуы мүмкін. Мысалы, корпус негізінде алынған Н.В. Гогольдің жазған хаттар корпусы негізіндегі деректерді жалпы тілдік корпустар деректерімен салыстырылады, нәтижесінде көптеген жағдайда, авторлық сөзқолданыстың ерекшеліктерін бейнелейтін сөз тіркестерінің елеулі айырмашылықтарын байқауға болады. Сонымен, жоғарыда сипатталған әдіс-тәсілдер жазушы тілін зерттеу мен оның сөздігін құрастыруда және әртүрлі стильдер шеңберінде немесе кезеңдік шығармалардағы тіркесімдіктің ерекшеліктерін айқындауда тиімді әдіс болып табылады.

Бірнеше тілдің корпустарындағы биграммаларды демоверсия режимінде (тәртібінде) іздеуді *http://www.aot.ru/cgi-bin/bigrams.cgi* сайты бойынша жүзеге асыруға болады.

Қазақ тілінің 15 томдық сөздігін құрастыру кезінде де көптеген мысалдар контекстері арқылы жаңа тіркестер анықталып, сөздікке енгізілгендігі белгілі. Болашақта қазақ тілінде де лексикографиялық жұмыстарға корпустың берері мол. Корпус арқылы жаңа сөздер, сөз мағыналарын, жаңа тіркестер, оккозионал қолданыстарды жүздеп табуға болады.

Корпустарға негізделген грамматикалық зерттеулер. Грамматиканы зерттеу тілдің құрылымын түсінумен байланысты. Лексикографияға қарағанда грамматиканы эмпирикалық зерттеулер бойынша қарастыру дәстүрі жоқ. Тіл иелерінің өз грамматикалық қорын барынша пайдалана бастағанына көп уақыт өте қойған жоқ. Дәстүрлі грамматикалық зерттеулердің корпустық деректерге негізделіп жүргізілуі, яғни сөз, сөйлем, дискурс деңгейлерінде жүргізілуі олардың айтарлықтай ерекшелігіне айнала бастады.

Әрбір регистрге шаққанда осы сияқты грамматикалық құбылыстар заңдылығын қазақ тілі корпустары бойынша да анықтауға болады. Мұндай зерттеулер әрине бұған дейін де жүргізілген. Алайда корпус массиві мұндай зерттеулердің мүмкіндігін бұрынғыдан мыңдаған есе арттырады. Көлемді регистрлер бойынша анализ жасауға мүмкіндік береді.

Ғылыми әдебиет абстрактылы жағдайларды, үдерістер мен нысандарды сипаттауға назар аударса, көркем проза нақты адамдар орындайтын жеке сипаттау мен әрекеттерге көбірек көңіл бөледі. Сондықтан зат есімдер мен етістік сөздердің үш регистрде де (ғылыми әдебиет, көркем проза, ауызша тіл) қолдану жиіліктерінде айырмашылық болуы керек. Шындығында, зерттеушілер белгілі авторлардың стилін сипаттаудың кейбір жағдайларында зат есімдер мен етістіктердің қолдану жиіліктерінің салыстырмалы шамаларын пайдаланады. Мұндай санақтар, *Corpus Linguistics* оқулығы авторларының пікірінше, регистр бойынша сөз таптарына қатысты қолдану айырмашылықтарын көрсете алады.

Корпустарға негізделген дискурстық зерттеу

Көптеген лексикалық және грамматикалық құбылыстардың тілдегі қолданылу заңдылығын толық түсіну тек көлемді дискурсивті контекстегі олардың атқаратын қызметін талдау (анализдеу) арқылы ғана мүмкін болмақ.

Дискурс дегеніміз – экстралингвистикалық, яғни оқиғалық аспектіде алынған прагматикалық, әлеуметтік-мәдени, психологиялық және тағы да басқа факторлармен қоса есептегендегі жүйелі (қисынды) түрдегі мәтін; мақсатты бағыттағы әлеуметтік әрекет кезінде қарастырылатын сөйлеу, адамдар арасындағы әрекет кезінде және олардың санасындағы механизмдерінде (когнитивтік процестерде) орын алатын компонент (ЛЭС, 1990). Дискурстың элементтері: баяндалатын оқиғалар, оларға қатынасушылар, перформативтік ақпарат және «оқиға еместер», яғни а) оқиғаға себепші жағдайлар; ә) оқиғаға түсінік беретін фон; б) оқиғаға қатысушылардың бағасы; в) дискурстың оқиғаларға қатысын айқындайтын ақпарат (Демьянков, 1982). Корпустық лингвистика әдістері дискурстық талдаудың дамуына айтарлықтай үлес қосуы мүмкін. Өйткені корпус дискурстың белгілі бір типтерінің ерекшеліктерін және жеке мәтіннің қарастырып отырған регистрдегі дискурс моделіне қаншалықты сәйкес келетіндігін де сипаттап бере алады (Баранов, 1993).

Дыбыстық корпус материалы бойынша дискурстық зерттеу. Орыс тілінің материалы бойынша энантиосемиді зерттеу жүргізілген, яғни сөздегі қарама-қарсы мағыналардың бірге келуін немесе «ішкі антонимия» мәселесін зерттеу (Маркасова, 2008). Бұл құбылысты күнделікте тілдесім көрініс тапқан орыс тілінің дыбыстық корпусы материалы арқылы (*Бір сөйлесу күні* – БСК; Один речевой день – ОРД) зерттеу авторды мынадай ойға әкелді: тілшілердің бұрын ескере қоймаған энантиосеми түрінің (типінің) бар екендігін дәлелдеу мақсатында сөйлеу тілінде жиі қолдананатын риторикалық энантиосемияны зерттеуді қолға алды.

ОРД (немесе БСК) корпусындағы белгілі фрагменттерінің лексикасын зерттеу мынадай жайтқа әкелді, И19 арқылы шартты белгіленген ақпарат беруші (30-35 жастағы әйел адам) көңіл

қалдырар ештеңе айтпайды, тек сұхбаттасушыға табыс тілейді, тіпті мақтайды (*азамат, тамаша, жақсы, өте жақсы).* Бірақ баланың аты және басқа да жақсы көру, еркелету маркерлері (орыс тіліндегі *солнышко, заинька, зайка, котик, умница, дорогой, милый, миленький және т.б.*) балаларға айтылатын жағымды дыбыстармен, қажетті эмоциямен айтылмайды.

Корпустарға негізделген дискурстарды зерттеу мәселесі мынадай 4 салаға бөлінуі мүмкін:

1) дискурсты құрастыру және мәтін құрылымы;

2) қарым-қатынас жасаудың дискурсивті-прагматикалық аспектілері;

3) мәтіндік және прагматикалық коллокациялар;

4) мәтін мен дискурстегі вариативтілік (Крейдлин, 2000).

Ғалым Т. Виртаненнің пікірі бойынша, ең соңғы екі саланың корпус арқылы жүргізілетін кең ауқымды зерттеулердің келешегі зор. Сондай-ақ Қазақ тілінің ұлттық корпусында ауызша корпустар жасалған жағдайда дискурсты зерттеудің жаңа мүмкіндіктері туатындығы даусыз.

Қорытынды, тұжырымдар

Зерттеудің мақсаты мен міндеттеріне байланысты мынадай нәтижелер алынады:

- электронды тілдік корпустарды лексикографиялық зерттеулерде пайдалану мүмкіндіктері, яғни оны әртүрлі сөздіктер (терминологиялық, этнографиялық, аймақтық, түсіндірме т.б.) құрастыру тәжірибесінде пайдаланудың лингвистикалық және инженерлік технологиясы, әдістәсілдері айқындалады; сөздік құрастыру ісін автоматтандырудың ғылыми методологиясы жасалады;

- тілдік корпустардан қазақ тілінің әртүрлі жиілік сөздіктерін алудың ғылыми-инженерлік технологиясы жасалады;

- тілдік корпус материалдары негізінде қазақ тілінің құрылымдық жүйесіндегі тілдік заңдылықтар қайта қарастырылып, бұрын-соңды жасалған ғылыми тұжырымдар нақты тілдік фактілер бойынша тиянақталады немесе ғылыми жаңартпалар енгізіледі; лингвистикалық зерттеулер жүргізудің әдіс-тәсілдері айқындалады;

- таңдама мәтіндерге жасалған морфологиялық, синтаксистік, семантикалық т.б. белгіленімдер қою тілдік жүйедегі осындай талдаулар жүйесімен тығыз сабақтастықта қарастырылып, тілдік талдаудың автоматты түрі енгізілетін болады т.б.

Қорыта айтқанда, зерттеу нәтижелері тіл білімінің барлық салаларында қолданылу мүмкіндігі бар. Нақты айтқанда, әртүрлі сөздіктер (жиілік, түсіндірме, аймақтық, фразелогиялық, этнографиялық) құрастыруда, ғылыми грамматикалар жазуда, оқыту жүйесінде және оқулықтар құрастыруда, әдістемелік құралдар жазуда, аударма жұмыстарында т.б. Тілдік корпустар компьютерлік базаға салынып, бір орталықты басқару жүйесі бойынша жұмыс істейтіндіктен, тілдік зерттеулердің барлығын дерлік нақты фактологиялық материалдармен қамтамасыз етеді. Тілдік корпустармен жұмыс істеу ғылыми-зерттеу жұмыстарын автоматтандырып, ғылымды дамытуға орасан зор үлес қосатындығы сөзсіз.

Мақала *BR11765619 «Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі қазақ тілінің ұлттық корпусын әзірлеу: ғылыми-зерттеу және оқыту интернет-ресурсы»* жоба тақырыбы бойынша зерттеу аясында жазылған.

Әдебиеттер

Перцов Н.В. О роли корпусов в лингвистических исследованиях // Труды международной конференции «Корпусная лингвистика – 2006». – СПб.: Изд-во С.-Петерб. ун-та; Изд-во РХГА, 2006. – С. 318-331.

McWhinney B. The CHILDES Project: Tools for Analyzing Talk. – Mahwah, NJ. Lawrence Erlbaum Associates. Third Edition, 2000. – Vol. 1.

Kytö M., Rissanen M. A language in transition: The Helsinki Corpus of English texts, ICAME Journal, 1992. – 16: 7-27.

Stenström A-B., Andersen, G. More trends in teenage talk: A corpus-based investigation of the discourse items cos and innit // C. Percy, C. Meyer & I. Lancashire (eds). Synchronic corpus linguistics. Amsterdam: Rodopi, 1996. – Pp. 189-203.

Meyer Ch. F. English Corpus Linguistics: An Introduction. Cambridge: Cambridge University Press, 2002. – Xvi + 168.

Гришина Е.А., Савчук С.О. Национальный корпус русского языка как инструмент для изучения вариативности грамматических норм // Труды международной конференции «Корпусная лингвистика – 2008» 6-10 октября 2008 г. – СПб., 2008. – С. 161-169.

Савчук С.О. Орыс тілінің ұлттық корпусындағы метамәтіндік белгіленімдер: базалық ұстанымдары мен негізгі функциялары // Орыс тілінің ұлттық корпусы: 2003-2005. Нәтижелері және болашағы. – М., 2005.

Biber D., Conrad S., Reppen R. Corpus Linguistics. Investigating language structure and use. Cambridge University Press, 1998.

Баранов А.Н. Введение в прикладную лингвистику. – М., 2007.

Хохлова М.В. Экспериментальная проверка методов выделения коллокаций // Slavica Helsingiensia 34. Инструментарий русистики: Корпусные подходы. – Хельсинки, 2008. – С. 343-357.

Баранов А.Н. Корпусная лингвистика // Баранов А.Н. Введение в прикладную лингвистику: Учебное пособие. – М.: Едиториал УРСС, 2003. – С. 114.

ЛЭС – Лингвистический энциклопедический словарь / Под ред. В.Н. Ярцевой. – М.: Сов. энциклопедия, 1990.

Демьянков В.З. Англо-русские термины по прикладной лингвистике и автоматической переработке текста. Вып. 2. Методы анализа текста // Всесоюзн. центр переводов. Тетради новых терминов, 39. – М., 1982.

Баранов А.Н., Плунгян В.А., Рахилина Е.В. Путеводитель по дискурсивным словам русского языка. – М., 1993.

Маркасова Е.В. Риторическая энантиосемия в корпусе русского языка повседневного общения «Один речевой день» // Компьютерная лингвистика и интеллектуальные технологии. Выпуск 7 (14). По материалам ежегодной международной конференции «Диалог» (2008) / Гл. ред. А. Е. Кибрик. – М., 2008. – С. 352-355.

Крейдлин Г.Е. Голос и тон в языке и речи / Отв. ред. Н.Д. Арутюнова // Язык о языке. – М.: Языки русской культуры, 2000. – С. 453-501.

References

Baranov A.N. (2003) Korpusnaya lingvistika [Corpus linguistics] // Vvedenie v prikladnuyu lingvistiku: Uchebnoe posobie. – M.: Editorial URSS, – S. 114 s. (In Russian)

Baranov A.N. (2007) Vvedenie v prikladnuyu lingvistiku. [Introduction to Applied Linguistics] – M. (In Russian)

Baranov A.N., Plungyan V.A., Rakhilina E.V. (1993) Putevoditel’ po diskursivnym slovam russkogo yazyka. [A guide to discursive words of the Russian language] – M. (In Russian)

Biber D., Conrad S., Reppen R. (1998) Corpus Linguistics. Investigating language structure and use. Cambridge University Press. (In Russian)

Demyankov V.Z. (1982) Anglo-russkie terminy po prikladnoj lingvistike i avtomaticheskoj pererabotke teksta. [English-Russian terms on applied linguistics and automatic text processing] Vyp. 2. Metody analiza teksta // Vsesoyuzn. centr perevodov. Tetradi novyh terminov, 39. – M. (In Russian)

Grishina E.A., Savchuk S.O. (2008) Naczional`ny`j korpus russkogo yazy`ka kak instrument dlya izucheniya variativnosti grammaticheskikh norm [The National Corpus of the Russian Language as a tool for studying the variability of grammatical norms] // Trudy` mezhdunarodnoj konferenczii «Korpusnaya lingvistika – 2008» 6-10 oktyabrya 2008 g. – SPb., – S. 161-169. (In Russian)

Khokhlova M.V. (2008) Eksperimental’naya proverka metodov vydeleniya kollokacij [Experimental verification of collocation isolation methods] // Slavica Helsingiensia 34. Instrumentarij rusistiki: Korpusnye podhody. – Hel’sinki, – S. 343-357. (In Russian)

Kreidlin G.E. (2000) Golos i ton v yazyke i rechi [Voice and tone in language and speech] // Yazyk o yazyke / Otv. red. N.D. Arutyunova. M.: Yazyki russkoj kul’tury. – S. 453-501. (In Russian)

Kytö M., Rissanen M. (1992) A language in transition: The Helsinki Corpus of English texts, ICAME Journal, – 16: 7-27.

LES – Lingvisticheskij enciklopedicheskij slovar’ (1990) [LED – Linguistic Encyclopedic Dictionary] / Pod red. V.N. Yarcevoj. – M.: Sov. enciklopediya. (In Russian)

Markasova E.V. (2008) Ritoricheskaya enantiosemiya v korpuse russkogo yazyka povsednevnogo obshcheniya «Odin rechevoj den’» [Rhetorical enantiosemy in the corpus of the Russian language of everyday communication “One speech day”] // Komp’yuternaya lingvistika i intellektual’nye tekhnologii. Vypusk 7 (14). Po materialam ezhegodnoj mezhdunarodnoj konferencii «Dialog» / Gl. red. A. E. Kibrik. – M., 2008. – C. 352-355. (In Russian)

McWhinney B. (2000) The CHILDES Project: Tools for Analyzing Talk. – Mahwah, NJ. Lawrence Erlbaum Associates. Third Edition, – Vol. 1.

Meyer Ch. F. (2002) English Corpus Linguistics: An Introduction. Cambridge: Cambridge University Press, – Xvi + 168.

Pertsov N.V. (2006) O roli korpusov v lingvisticheskikh issledovaniyakh [On the role of corpora in linguistic research] // Trudy` mezhdunarodnoj konferenczii «Korpusnaya lingvistika – 2006». – SPb.: Izd-vo S.-Peterb. un-ta; Izd-vo RKhGA, – S. 318-331. (In Russian)

Savchuk S.O. (2005) Orys tilinin ulttyq korpusyndagy metamatindik belgilenimder: bazalyq ustanymdary men negizgi funkciyalary [Meta-text designations in the national corpus of the Russian language: basic principles and main functions] // Orys tilinin ulttyq korpusy: 2003-2005. Natizheleri zhane bolashagy. – M. (In Russian)

Stenström A-B., Andersen, G. (1996) More trends in teenage talk: A corpus-based investigation of the discourse items cos and innit // C. Percy, C. Meyer & I. Lancashire (eds). Synchronic corpus linguistics. Amsterdam: Rodopi, – Pp. 189-203.