KORPUSLıQ LINGVISTIKA TILDI IZERTLEWDIŃ ZAMANAGÓY METODOLOGIYASı SıPATıNDA: TEORIYASı, METODLARı HÁM QARAQALPAQ TILI USHıN KELESHEGI

Darbaeva Aygul Muxtarovna
Ájiniyaz atındağı NMPI magistrantı
lunniycvetok19@gmail.com
Doi: https://doi.org/10.5281/zenodo.23264618

Annotaciya. Bul maqalada korpuslıq lingvistika házirgi empirikalıq lingvistikanıń áhmiyetli baǵdarlarınıń biri hám tildi arnawlı túrde dúzilgen tábiyiy tillik maǵlıwmatlar jıynaǵı tiykarında izertlew metodologiyası sıpatında qarastırıladı. Til korpusı hám korpuslıq lingvistika túsiniklerine tiyisli tiykarǵı kózqaraslar talqılanıp, korpustıń reprezentativligi, teńgerimliligi hám annotaciyalanıwı máseleleri qarastırıladı. Sonday-aq korpuslıq analizdiń tiykarǵı usılları – jiyilik analizi, konkordans, kollokaciyalardı, gilt sózlerdi, tillik birliklerdiń bólistiriliwin hám statistikalıq baylanısların izertlew usılları sıpatlanadı. Korpuslıq usıldıń leksika, morfologiya, grammatika, sintaksis, semantika hám diskurstı izertlewdegi imkaniyatlarına ayrıqsha dıqqat awdarıladı. Bólek bólimde qaraqalpaq tiliniń korpuslıq lingvistikası hám házirgi sanlı resursları máselesi qarastırılıp, qaraqalpaq tiliniń mektep sabaqlıqları korpusı, veb-tekstler hám sóylew korpusları sıyaqlı resurslar tallanadı. Qaraqalpaq tili ushın reprezentativli, annotaciyalanǵan hám kóp funkciyalı korpustı rawajlandırıwdıń zárúrligi tiykarlanadı. Sonday-aq korpuslıq usıldıń qaraqalpaq tiliniń sintaksislik sistemasın, sonıń ishinde qospa gáplerdi izertlewdegi keleshek imkaniyatları kórsetiledi.

Tayanısh túsinikler: korpuslıq lingvistika, til korpusı, korpuslıq analiz, elektronlıq korpus, konkordans, kollokaciya, jiyilik, sintaksis, diskurks, qaraqalpaq tili, az resurslı til, sanlı lingvistika.

Kirisiw. Házirgi til bilimi tildi izertlewdiń empirikalıq bazasın keńeyttirip, sanlı usıllardı ilimiy izertlewge belsene engiziw menen sıpatlanadı. Bul procestiń eń áhmiyetli nátiyjeleriniń biri – korpuslıq lingvistikanıń rawajlanıwı. Korpuslıq lingvistika arnawlı túrde jıynalǵan, sistemalastırılǵan hám kompyuterlik texnologiyalar járdeminde islengen úlken kólemdegi tábiyǵıy tillik tekstler jıynaǵı tiykarında tildi izertlewge múmkinshilik beredi.

Korpuslıq usılǵa bolǵan qızıǵıw, eń aldı menen, ayırım mısallar yakı izertlewshiniń tillik intuiciyasına súyengen analizden anıq til qollanılıwında ushırasatuǵın kóp muǵdardaǵı maǵlıwmatlardı sistemalı túrde izertlewge ótiw múmkinshiligi menen baylanıslı. G. Kennedi XX ásirdiń aqırında-aq «large, computerized bodies of text» (úlken kompyuterlestirilgen tekstler jıynaǵı)n qollanıw lingvistikalıq izertlewlerdiń tez rawajlanıp atırǵan baǵdarlarınıń biri bolǵanın atap ótedi. [1].

Keyingi dáwirlerde korpuslıq metodologiya keńeyip, tek leksikografiya hám ingliz tilin izertlew menen sheklenbey, grammatika, sintaksis, diskurs analizi, sociolingvistika, awdarmatanıw,til úyretiw hám kompyuterlik lingvistika tarawlarında da keńnen qollanıla basladı.

S.T.Gris korpuslıq lingvistikanı házirgi til biliminde tez rawajlanıp atırǵan metodologiyalardıń biri sıpatında qarastırıp, reprezentativlik hám teńgerimlilik, razmetkalaw hám annotaciyalaw sıyaqlı túsiniklerdiń áhmiyetin atap ótedi. Ol tiykarǵı usıllar qatarında «frequency lists, concordances, collocations» [2, 1225-b.] – jiyilik dizimleri, konkordanslar hám kollokaciyalardı kórsetedi. Demek, korpuslıq usıl tek tekstlerdi jıynaw menen sheklenbey, olardı sistemalı túrde sanlıq hám sapalıq jaqtan analizlewdi de talap etedi.

Bul izertlewdiń aktuallıǵı tillik resursları dúnya tillerine salıstırǵanda sheklengen tillerdi házirgi korpuslıq usıllar tiykarında izertlew zárúrligi menen belgilenedi. Bul tárepten qaraqalpaq tili ayrıqsha qızıǵıwshılıq tuwdıradı.

Til korpusı hám korpuslıq lingvistika túsinikleri. Házirgi lingvistikada «korpus» túsinigi tek tekstlerdiń ápiwayı jıynaǵın ańlatpaydı. Korpustıń qalay dúziliwi, onıń reprezentativligi, strukturası, metamaǵlıwmatları, izlew imkaniyatları hám razmetka sisteması onıń ilimiy qunın belgileytuǵın tiykarǵı belgiler bolıp esaplanadı.

A.Stefanovich korpuslıq lingvstikanı baqlawǵa tiykarlanǵan metodologiyalıq usıllardıń biri sıpatında qarap, izertlew dizaynınıń, maǵlıwmatlardı ajıratıp alıwdıń hám olardı statistikalıq jaqtan bahalawdıń sistemalı túrde alıp barılıwınıń zárúrligin atap ótedi [3]. Bul kózqaras ayrıqsha áhmiyetli, sebebi kóp muǵdardaǵı tekstlerdiń jıynalıwı óz-ózinen olardı ilimiy korpusqa aylandırıp qoymaydı. Izertlewshi tekstlerdi tańlaw kriteriyaların, tekst túrlerin, janrlıq hám waqıtlıq shegaralardı, razmetkalaw usılların hám analiz proceduraların anıqlawı kerek.

T. Makeneri hám E. Xardi korpuslıq lingvistikanı tildi izertlewge arnalǵan proceduralar sisteması menen baylanıstıradı [4]. Korpuslıq lingvistikanıń rawajlanıwı tillik maǵlıwmattıń tábiyǵıy qollanılıwına, jiyilikke, forma hám funkciya arasındaǵı baylanıslarǵa tiykarlanǵan jańa teoriyalıq kózqaraslardıń payda bolıwına alıp keldi.

D. Biber, S. Konrad hám R.Reppen korpuslıq usıldı kompyuterde saqlanǵan «large databases of real language examples» [5] (real tillik mısallardıń úlken bazaları)n paydalanıw menen baylanıstıradı. Bul usıldıń áhmiyetli ózgesheligi tillik birliklerdi tábiyǵıy kontekstte qalay qollanılatuǵının baqlaw imkaniyatı menen túsindiriledi.

Demek, til korpusın izertlew maqsetine sáykes sistemalı túrde jıynalǵan, reprezentativli yaki arnawlı baǵdarlanǵan tekstlik yamasa sóylew maǵlıwmatları jıynaǵı dep túsiniwge boladı. Korpus izertlewshi ushın tillik birliklerdi izlewge, salıstırıwǵa hám sanlıq jaqtan analizlewge múmkinshilik beretuǵın túrde shólkemlestiriledi.

2. Korpuslıq lingvistika neni izertleydi? Korpuslıq lingvistika tillik sistemanıń tek bir dárejesin ǵana emes, onıń túrli dárejeleri hám aspektlerin izertlewge múmkinshilik beredi.

2.1. Leksika hám frazeologiya. Korpuslıq izertlewdiń eń erte hám nátiyjeli baǵdarlarınıń biri – leksikanı izertlew. Korpus járdeminde sózlerdiń jiyiligin, olardıń bólistiriliwin hám qollanılıw kontekstlerin anıqlawǵa boladı. Ayrıqsha áhmiyetke iye kollokaciyalardı – sózlerdiń nızamlı túrde birge qollanılıwın izertlewge iye. Dj. Sinkler korpuslıq usıldıń rawajlanıwın korpus, konkordans hám kollokaciya túsinikleri menen tıǵız baylanıstırdı [6].

M.Stabbs korpuslıq maǵlıwmatlar sózlerdi jeke-jeke emes, olardıń real tekstlerde hám sóz dizbeklerinde qollanılıwı arqalı izertlewge múmkinshilik beretuǵının kórsetedi [7]. Sonlıqtan korpuslıq semantika leksikalıq birliklerdiń bir-biri menen baylanısın, semantikalıq tallawların hám turaqlı qollanıw úlgilerin izertlew menen tıǵız baylanıslı.

2.2. Morfologiya hám grammatika. Korpuslıq usıl morfologiyalıq formalardıń jiyiligin, grammatikalıq kategoriyalardıń bólistiriliwin hám túrli grammatikalıq formalardıń qollanılıw ózgesheliklerin anıqlawǵa múmkinshilik beredi. Mısalı, izertlewshi belgili bir formanıń qanshelli jiyi qollanılatuǵının, qanday leksemalar menen baylanısatuǵının, qanday grammatikalıq konstrukciyalarda ushırasatuǵının hám jazba hám awızeki sóylewde ayırmashılıqlardıń bar-joǵın anıqlay aladı.

2.3. Sintaksis. Korpuslıq usıldıń ayrıqsha keleshekli baǵdarlarınıń biri – sintaksisti izertlew. Korpus járdeminde ayırım abstraktlı modellerdi emes, real tekstlerde júzege asqan sintaksislik konstrukciyalardı izertlew múmkin. Atap aytqanda, tómendegi máselelerdi izertlewge boladı:

- sintaksislik modellerdiń jiyiligi;
- konstrukciyalardıń variantlılıǵı;
- komponentlerdiń jaylasıw tártibi;
- predikativlik bóleklerdiń baylanısı;
- baǵınıńqı konstrukciyalardıń túrleri;
- sintaksislik konstrukciyalardıń janr hám stil menen baylanısı;

- sintaksislik forma hám semantikanıń ózara qatnası.

Qospa gáplerdi korpuslıq usılda izertlew ayrıqsha áhmiyetli, sebebi bul usıl zamanagóy tilde qanday konstrukciyalardıń ónimli ekenin, olardıń qanday kontekstlerde qollanılatuǵının hám túrli tekstlerde qalay bólistiriletuǵının anıqlawǵa múmkinshilik beredi.

2.4. Diskurs. Korpuslıq lingvistika diskurstı izertlewde de keńnen qollanıladı. P. Beyker korpuslardı diskurstağı jiyilik, dispersiya, konkordanslar, kollokaciyalar hám gilt sózlerdi analizlew ushın paydalanıw múmkinshiliklerin kórsetedi [8].

Sonlıqtan korpuslıq analiz tek tildiń strukturasın emes, til arqalı sociallıq, mádeniy hám kommunikativlik qubılıslardıń qalay sáwlelenetuǵının da izertlewge múmkinshilik beredi.

3. Korpuslıq izertlewdiń tiykarǵı usılları. Korpuslıq lingvistikadağı eń kóp qollanılatuǵın usıllardıń biri – **jiyilik analizi**. Bul usıl belgili bir tillik birliktiń yamasa konstrukciyanıń absolyut hám salıstırmalı jiyiligin anıqlawǵa múmkinshilik beredi. Jiyilik maǵlıwmatları túrli tillik qubılıslardı, janrlardı, stillerdi hám til variantlardı salıstırıwda qollanıladı.

Ekinshi tiykarǵı instrument – **konkordans analizi**. Konkordans belgili bir tillik birliktiń korpustağı qollanılıwların onıń konteksti menen birge kórsetedi. Bul izertlewshige tillik birliktiń real kontekstegi leksikalıq hám sintaksislik baylanısların kóriwge múmkinshilik beredi.

Úshinshi áhmiyetli usıl – **kollokaciyalardı analizlew.** Onıń járdeminde izertlenip atırǵan birlik penen statistikalıq yamasa funkcionallıq jaqtan baylanıslı bolǵan sózler anıqlanadı.

Bunnan basqa, zamanagóy korpuslıq lingvistikada tómendegi usıllar da qollanıladı:

- gilt sózlerdi analizlew;
- leksika-grammatikalıq modellerdi izertlew;
- dispersiyanı analizlew;
- túrli korpuslardı salıstırıw;
- statistikalıq testler;
- klasterlik analiz;
- maǵlıwmatlardı vizualizaciyalaw.

S.T.Gris jiyilik dizimleri, konkordanslar hám kollokaciyalardı korpuslıq lingvistikanıń tiykarǵı usılları qatarında kórsetedi [2].

Stefanovich bolsa korpuslıq izertlewdiń izbe-iz alıp barılıwın, yaǵnıy máseleni anıqlaw, izertlew dizaynın dúzetiw, maǵlıwmatlardı ajıratıw hám statistikalıq interpretaciyalawdı zárúr dep esaplaydı [3].

4. Korpustıń reprezentativligi hám annotaciyalanıwı. Korpuslıq lingvistikadağı eń quramalı metodologiyalıq máselelerdiń biri – **reprezentativlik** máselesi. Korpus izertlew maqsetine sáykes bolıwı kerek. Barlıq til túrlerin birdey dárejede sáwlelendiretuǵın universal korpus dúziliwi múmkin emes.

T. Makeneri, R. Syao hám Yu. Tono korpus dizaynında reprezentativlik hám teńgerimlilik máselelerinıń áhmiyetin atap ótedi [9]. Izertlew maqsetine qaray korpus ulıwmalıq, arnawlı, janrlıq, oqıw, tariyxıy, dialektlik, parallel yamasa kóp modallıq korpus bolıwı múmkin.

Korpus ushın annotaciyalawda ayrıqsha áhmiyetke iye. Annotaciya morfologiyalıq, sintaksislik, semantikalıq, diskurslıq hám basqa da maǵlıwmatlardı qamtıwı múmkin. Sintaksisti izertlewde sintaksislik razmetka ayrıqsha áhmiyetli bolıp, belgili bir konstrukciyalardı avtomatlıq yamasa yarım avtomatlıq túrde ajıratıp alıwǵa múmkinshilik beredi.

Demek, korpus jaratıwdıń ózi de arnawlı ilimiy izertlewdi talap etetuǵın process bolıp, onda tekstlerdi tańlaw principleri, tekstlerdi standartlastırıw, metamaǵlıwmatlardı kirgiziw hám maǵlıwmat sapasın baqlaw máseleleri sheshiliwi kerek.

5. Korpuslıq lingvistika hám qaraqalpaq tili. Qaraqalpaq tili ushın korpuslıq lingvistikanıń rawajlanıwı ayrıqsha áhmiyetke iye. Dástúriy qaraqalpaq til biliminde fonetika, morfologiya, sintaksis hám stilistika máseleleri boyınsha kóplegen ilimiy izertlewler ámelge asırılǵan. Qaraqalpaq tiliniń

sintaksislik sistemasın izertlewde A. Nájimovtıń “Хәзирги қарақалпақ тили синтаксиси” mazmunındağı miyneti áhmiyetli orın iyeleydi [10]. Sonday-aq, A. Dáwletov, M. Dáwletov hám M. Qudaybergenovtıń “Хәзирги қарақалпақ әдебий тили. Синтаксис” miyneti de qaraqalpaq tili sintaksisin izertlewdiń tiykarǵı dereklerinińiń biri bolıp tabıladı [11].

Biraq dástúriy til sıpatlaması menen korpuslıq izertlew bir-birinen ózgeshe máselelerdi sheshedi. Dástúriy grammatika tillik qubılıslardı sistemalastırıp, olardıń sapalıq sıpatlamasın berse, korpus real qollanıwǵa tiykarlanǵan kóp kólemdegi maǵlıwmatlar arqalı bul nızamlılıqlardı tekseriwge múmkinshilik beredi.

Qaraqalpaq tiliniń zamanagóy sanlı resurslarınıń payda bolıwı bul baǵdardıń keleshekli ekenin kórsetpekte. 2023-jılı K.Madatov, S. Bekchanov hám Y. Vichich ***Karakalpak Language School Corpus (KAASC)*** dep atalatuǵın resurs tuwralı maǵlıwmat berdi. Biul korpus qaraqalpaq tilindegi 23 mektep sabaqlıǵınan dúzilgen bolıp, avtorlar onıń tiykarında stop-wordlardı avtomat túrde ajıratıw ushın TF-IDFtiń unigram, bigram hám collocation usılların paydalanǵan [12]. Bul maǵlıwmatlar jıynaǵınıń DOI nomeri: 10.1016/j.dib.2023.109111.

Bul resurs qaraqalpaq tiliniń sanlı lingvistikası ushın áhmiyetli qádem bolıp tabıladı. Biraq sabaqlıq tekstleriniń korpusın tolıq zamanagóy tildiń ulıwmalıq korpusına teńlestiriwge bolmaydı. Sebebi sabaqlıq tekstler kórkem ádebiyat, publicistika, ilimiy stil, rásmiy-isbası stili, awızeki sóylew hám sanlı kommunikaciya sıyaqlı basqa tekst túrleriniń tolıq reprezentaciyasın bere almaydı.

Sońǵı izertlewlerde qaraqalpaq tiliniń veb-tekstlerin jıynaw máselesine de dıqqat awdarılmaqta. 2026-jılda jarıq kórgen izertlewde az resurslı tiller ushın veb-saytlardan tekstli maǵlıwmatlardı avtomatlıq túrde jıynaw usılları qarastırılıp, qaraqalpaq tili mısalında kólemli veb-maǵlıwmatlar bazasın dúziw imkaniyatı kórsetilgen.

Sonday-aq qaraqalpaq tiliniń sanlı resursları tek jazba tekstler menen sheklenbey, awızeki sóylew maǵlıwmatların da qamtıy basladı. Bul qaraqalpaq tili ushın avtomatlıq sóylewdi tanıw hám basqa NLP-texnologiyalardı rawajlandırıw ushın áhmiyetli qádem bolıp tabıladı.

Sonıń menen birge, ayırım korpuslar hám maǵlıwmatlar jıynaǵınıń bar ekenligi tolıq túrde milliy qaraqalpaq tili korpusınıń jaratılǵanın ańlatpaydı. Sistemalı korpuslıq izertlew ushın túrli janrlar hám dáwirlerdi qamtıytuǵın, metamaǵlıwmatları tolıq, standartlastırılǵan razmetkası hám izlew instrumentleri bar resurs zárúr.

Ayrıqsha keleshekli baǵdar – **qaraqalpaq tiliniń sintaksislik annotaciyalanǵan korpusın jaratıw.** Bunday korpus gáptiń strukturasın, sintaksislik baylanıs túrlerin, komponentlerdiń jaylasıw tártibin, túrli modellerdiń jiyiliginhám olardıń funkcionallıq-stillik bólistiriliwin sistemalı túrde izertlewge múmkinshilik beredi.

6. Qaraqalpaq tili sintaksisin korpuslıq izertlewdiń keleshekleri. Sintaksis qaraqalpaq tiline korpuslıq usıldı qollanıwdıń eń keleshegi bar baǵdarlarınıń biri bolıp tabıladı. Zamanagóy izertlewlerde qospa gáplerdi funkcionllıq hám struktura-semantikalıq jaqtan izertlewge ayrıqsha dıqqat berilmekte.

Korpuslıq metodologiya bunday izertlewlerdi sanlıq hám kontekstlik maǵlıwmatlar menen tolıqtırıwǵa múmkinshilik beredi. Máselen, qospa gáplerdi izertlew barısında tómendegi máselelerdi anıqlawǵa boladı:

1. Baǵınıńqı gáp túrleriniń jiyiligi;
2. Ayırım strukturalıq modellerdiń ónimdarlıǵı;
3. Dánekerler hám basqa baylanıs qurallarınıń jiyiligi;
4. Konstrukciyalardıń funkcionallıq stiller boyınsha bólistiriliwi;
5. Struktura hám semantika arasındaǵı baylanıs;
6. Kórkem, publicistikalıq, ilimiy hám internet tekstlerindegi konstrukciyalardıń qollanılıw ózgeshelikleri.

Ayrıqsha keleshekli baǵıt – **struktura-semantikalıq, funkcionallıq hám korpuslıq usıllardı ózara baylanısta qollanıw**. Bul jaǵdayda sintaksislik konstrukciya bir waqıtta formal struktura, belgili bir mánini bildiretuǵın tillik birlik hám tekstte real qollanılatuǵın konstrukciya sıpatında izertlenedi. Qaraqalpaq tili ushın bunday usıl dástúriy sıpatlawǵa tiykarlanǵan modelden empirikalıq tekserile alatuǵın, yaǵnıy reprezentativli korpus maǵlıwmatlarına súyenetuǵın jańa izertlew modeline ótiwge járdem beredi.

7. Korpuslıq lingvistikanıń zamanagóy tendenciyaları hám jasalma intellekt. Zamanagóy korpuslıq lingvistika kompyuterlik lingvistika hám jasalma intellekt texnologiyaları menen tıǵız baylanısta rawajlanbaqta. Úlken kólemdegi tekstlik maǵlıwmatlardı avtomatlıq túrde qayta islew tillik birliklerdi izlewdi, klassifikaciyalawdı hám statistikalıq analizlewdi tezlestiriwge múmkinshilik beredi.

Bul másele az resurslı tiller ushın ayrıqsha áhmiyetli. Úlken kólemdegi annotaciyalanǵan korpustıń joqlıǵı til modellerin úyretiw, mashinalıq awdarma, avtomatlıq sóylewdi tanıw hám basqa NLP-texnologiyalardı rawajlandırıw imkaniyatların shekleydi.

Zamanagóy ilimiy izertlewler korpuslıq lingvistikanıń jasalma intellekt penen ózara baylanısınıń kúsheyip baratırǵanın kórsetedi. H.Curry hám T.McEnery 2025-jılǵı izertlewinde jasalma intellektti korpusqa baǵdarlanǵan oqıtıwda paydalanıwdı korpuslıq izertlewdiń keleshekli baǵdarlarınıń biri sıpatında kórsetedi [16].

Biraq avtomatlastırıw lingvisttiń ornın baspaydı. Avtomatlıq túrde alınǵan maǵlıwmatlar tekseriliwi, lingvistikalıq jaqtan túsindiriliwi hám sapası baqılanıwı kerek. Sonlıqtan eń keleshekli model – kompyuterlik usıllar úlken kólemdegi maǵlıwmatlardı qayta islew ushın paydalanılıp, al lingvistikalıq analiz olardıń mazmunın hám tillik nızamlıqların túsindiretuǵın model.

Juwmaq. Korpuslıq lingvistika – arnawlı túrde dúzilgen tábiyǵıy tillik maǵlıwmatlar jıynaǵın sistemalı túrde analizlewge tiykarlanǵan zamanagóy empirikalıq lingvistikalıq metodologiya. Onıń tiykarǵı instrumentlerine jiyilik analizi, konkordans, kollokaciyalıq analiz, gilt sózlerdi izertlew, bólistiriliwdi analizlew hám statistikalıq usıllar kiredi.

Korpuslıq usıl leksika, frazeologiya, morfologiya, grammatika, sintaksis, semantika hám diskurs sıyaqlı tildiń túrli deńgeylerin izertlewge qollanıladı. Onıń eń tiykarǵı artıqmashılıǵı – sanlıq hám sapalıq analizdi bir-biri menen baylanıstırıw múmkinshiligi.

Qaraqalpaq tili ushın korpuslıq lingvistika ayrıqsha ilimiy keleshekke iye. Búgingi kúnde qaraqalpaq tiliniń mektep oqıwlıqları korpusı, veb-tekstler hám sóylew maǵlıwmatları sıyaqlı sanlı resursları qáliplesip atır. Biraq túrli janrlar hám stillerdi tolıq qamtıytuǵın, reprezentativli, kóp dárejeli hám annotaciyalanǵan ulıwmalıq korpus jaratıw zárúrligi saqlanıp qalmaqta.

Ayrıqsha áhmiyetli másele – sintaksislik annotaciyalanǵan korpus jaratıw. Bunday resurs qaraqalpaq tiliniń sintaksislik sistemasın jańa basqıshta izertlewge, sonıń ishinde qospa gáplerdi korpuslıq usılda analizlewge ilimiy tiykar jaratadı. Dástúriy struktura-semantikalıq analizdi korpuslıq hám statistikalıq usıllar menen birlestiriw zamanagóy qaraqalpaq tilinde sintaksislik konstrukciyalardıń haqıyqıy qollanılıw ózgesheliklerin tereńirek ashıp beredi.

Demek, qaraqalpaaq tili korpuslıq lingvistikasınıń rawajlanıwı tek teoriyalıq emes, al ámeliy áhmiyetke de iye. Bul baǵdar elektronlıq sózliklerdi, avtomatlıq morfologiyalıq hám sintaksislik analizdi, mashinalıq awdarmanı, sóylewdi tanıwdı hám basqa da tábiyǵıy tildi qayta islew texnologiyaların rawajlandırıwǵa xızmet etedi. Keleshekte korpuslıq usıl qaraqalpaq tilin kompleksli izertlew hám onıń sanlı keńisliktegi rawajlanıwınıń áhmiyetli qurallarınıń birine aylanıwı múmkin.

Paydalanılǵan ádebiyatlar

1. Kennedy G. An Introduction to Corpus Linguistics. London: Longman, 1998. 315 p.
2. Gries S. Th. What is Corpus Linguistics? // Language and Linguistics Compass. 2009. Vol. 3. No. 5. P. 1225–1241. DOI: 10.1111/j.1749-818X.2009.00149.x.

3. Stefanowitsch A. Corpus Linguistics: A Guide to the Methodology. Berlin: Language Science Press, 2020. DOI: 10.5281/zenodo.3735822.
4. McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice. Cambridge: Cambridge University Press, 2012. DOI: 10.1017/CBO9780511981395.
5. Biber D., Conrad S., Reppen R. Corpus Linguistics: Investigating Language Structure and Use. Cambridge: Cambridge University Press, 1998.
6. Sinclair J. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991.
7. Stubbs M. Words and Phrases: Corpus Studies of Lexical Semantics. Oxford: Blackwell, 2001.
8. Baker P. Using Corpora in Discourse Analysis. London: Continuum, 2006.
9. McEnery T., Xiao R., Tono Y. Corpus-Based Language Studies: An Advanced Resource Book. London: Routledge, 2006.
10. Нәжимов А. Ҳәзирги қарақалпақ тили синтаксиси. Нөкис: Қарақалпақстан, 1990.
11. Дәўлетов А., Дәўлетов М., Кудайбергенов М. Ҳәзирги қарақалпақ әдебий тили. Синтаксис. Нөкис: Билим, 2009.
12. Madatov K., Bekchanov S., Vičič J. Dataset of Karakalpak language stop words // Data in Brief. 2023. Vol. 48. Article 109111. DOI: 10.1016/j.dib.2023.109111.
13. Uteuliev N. U., Abdalieva G. R., Kalmuratov B. K., Asenbaeva D. A. Collecting Textual Data from Websites for Low-Resource Languages // Science and Society. 2026.
14. Uteuliev N., Khudaybergenov K., Kudaybergenov J., Kudaybergenov T. Karakalpak Speech Corpus: The First Benchmark Dataset for Automatic Speech Recognition. 2026.
15. Сейдуллаева Д. Ю. Функционально-семантические типы сложноподчинённых предложений // Наука и общество. 2026. № 2-2.
16. Curry N., McEnery T. Corpus linguistics for language teaching and learning: A research agenda // Language Teaching. 2025. Vol. 58. No. 2. P. 232–251. DOI: 10.1017/S0261444824000430.
17. Gries S. Th. Quantitative Corpus Linguistics with R: A Practical Introduction. New York: Routledge, 2009.
18. Lüdeling A., Kytö M. Corpus Linguistics: An International Handbook. Vol. 1. Berlin/New York: De Gruyter Mouton, 2008.
19. O'Keeffe A., McCarthy M., Carter R. From Corpus to Classroom: Language Use and Language Teaching. Cambridge: Cambridge University Press, 2007.
20. Kilgarriff A., Grefenstette G. Introduction to the Special Issue on the Web as Corpus // Computational Linguistics. 2003. Vol. 29. No. 3. P. 333–347.