

Könül HƏBİBOVA\*

## KORPUS DİLÇİLİYİ VƏ KONKORDANS: METODOLOJİ YANAŞMALAR VƏ YENİ TƏTBİQLƏR XÜLASƏ

Bu məqalədə müasir korpus dilçiliyinin əsas anlayışları, xüsusilə konkordans və kollokasiya analizlərinin tətbiqi, elmi diskursda və poetik mətnlərin tədqiqində rolu araşdırılmışdır. Müəyyən edilmişdir ki, korpus dilçiliyinin inkişafı, milli korpusların yaradılması, avtomatik mətn emalı və süni intellekt texnologiyalarının inteqrasiyası ilə birlikdə aparılmalıdır. Konkordans və kollokasiya metodlarının elmi diskursda, media və poetik mətnlərin tədqiqində geniş tətbiqi Azərbaycan dilçiliyində yeni metodoloji yanaşmaların inkişafına töhfə verə bilər.

**Açar sözlər:** *korpus dilçiliyi, konkordans, kollokasiya, diskurs analizi, süni intellekt, milli dil korpusu.*

Konkordanslar və dil korpusları dilçilikdə mətnlərin təhlili və dil strukturlarının öyrənilməsi üçün mühüm vasitələr kimi bir-biri ilə sıx əlaqədədir. Konkordans, mətnlərdəki sözlərin və ifadələrin konkret kontekstlərini təqdim edərək, onların semantik və sintaktik əlaqələrini araşdırmağa imkan verir. Digər tərəfdən, korpuslar geniş miqyaslı yazılı və ya şifahi dil məlumatlarını ehtiva edən rəqəmsal bazalar kimi konkordansların hazırlanması üçün əsas mənbələrdən biridir. Beləliklə, korpuslar və konkordanslar bir-birini tamamlayan dilçilik alətləri kimi fəaliyyət göstərir. Konkordansların tarixi inkişafı onların korpuslarla olan əlaqəsinin güclənməsinə səbəb olmuşdur. Əvvəllər, əsasən, dini və klassik mətnlərin əlifba sırasına uyğun tərtib olunmuş konkordanslar dilin müəyyən strukturlarını öyrənmək üçün istifadə olunurdu. Lakin XX əsrdə elektron korpusların meydana çıxması ilə konkordansların tətbiq dairəsi genişləndi. “*Müasir konkordanslar korpuslarda saxlanılan böyük miqyaslı məlumatlar əsasında daha dəqiq analizlər təqdim edə bilər, bu da dilin müxtəlif aspektlərinin təhlilini asanlaşdırır*” [Əliyeva: 2023, s.201].

Rəqəmsal texnologiyaların inkişafı korpus əsaslı konkordansların yaranmasını mümkün etmişdir. Belə konkordanslar dilin təbii istifadəsini daha geniş şəkildə əhatə edən analizlər təqdim edir. Məsələn, COCA (Corpus of Contemporary American English) və BNC (British National Corpus) kimi nümunələrdən istifadə edilərək hazırlanmış konkordanslar ingilis dilində sözlərin və ifadələrin müxtəlif kontekstlərdə istifadəsini araşdırmağa imkan verir. Bu yanaşma dilçilik tədqiqatlarında korpusların və konkordansların birgə istifadəsinin effektivliyini nümayiş etdirir.

COCA – Müasir Amerika İngilis dili Korpusu, 1990-cı ildən etibarən Makr Devis (Mark Davies) tərəfindən yaradılmış və daim yenilənən böyük bir verilənlər bazasıdır. O, 500 milyondan çox sözdən ibarətdir və xəbərlər, romanlar, akademik məqalələr, şifahi nitq və digər müxtəlif janrlarda olan mətnləri ehtiva edir [Heydarov: 2019]. COCA, söz və ifadələrin tezliyini, kollokasiyalarını və sintaktik quruluşlarını araşdırmağa imkan verən bir

---

\* AMEA Nəsimi adına Dilçilik İnstitutunun Kompüter dilçiliyi şöbəsinin aparıcı elmi işçisi, fil.ü.f.d., dosent, habibovakonul@gmail.com

vasitədir. Bu korpus, dilçilik tədqiqatlarında, məşin tərcüməsində, akademik araşdırmalarda və lüğət tərtibində geniş istifadə olunur. Digər mühüm korpus olan BNC – Böyük Britaniya Milli Korpusu isə 100 milyon sözdən ibarət bir məlumat bazasıdır. Bu korpus Britaniya ingilis dilinin müxtəlif üslub və sahələrdə istifadəsini əks etdirən yazılı və şifahi mətnlərdən ibarətdir. BNC rəsmi və qeyri-rəsmi üslublar arasındakı fərqləri araşdırmaq, eləcə də ingilis dilinin leksik, sintaktik və semantik xüsusiyyətlərini öyrənmək üçün mühüm bir resurs hesab edilir.

Korpuslar və konkordansların inteqrasiyası yalnız dilin leksik və semantik təhlilində deyil, həm də tarixi və mədəni aspektlərinin öyrənilməsində mühüm rol oynayır. *“Müxtəlif dillərdə yaradılmış korpuslar vasitəsilə dillərin inkişafı, mətnlərin tarixi strukturlaşması və üslub xüsusiyyətləri üzrə konkordanslar hazırlamaq mümkündür”* [Məmmədova: 2019, s.148]. Məsələn, Osmanlı türkçəsindən müasir türk dilinə keçid prosesini öyrənmək üçün rəqəmsal korpuslara əsaslanan konkordanslar tərtib edilmişdir. Elektron korpusların həcmnin böyüməsi və onların annotasiyalarla zənginləşdirilməsi, konkordansların daha spesifik və hədəf yönümlü hazırlanmasını təmin edir. Bu, yalnız nəzəri dilçilik tədqiqatlarında deyil, həm də praktiki sahələrdə, məsələn, tərcümə və nitq tanıma texnologiyalarında istifadəyə şərait yaradır.

Müasir konkordansların hazırlanmasında istifadə olunan texnologiyalar, korpuslardakı məlumatların avtomatik analizi və tezlik hesablamalarını həyata keçirməyi mümkün edir. Bu yanaşma dilçilikdə daha sürətli və dəqiq nəticələr əldə etməyə imkan verir. Belə ki, kompüter əsaslı korpuslardan istifadə edərək hazırlanmış konkordanslar müəyyən sözlərin və ifadələrin tezlik analizini təqdim edir və bu məlumatları vizual formatda əks etdirir. Konkordanslar və korpuslar arasında qarşılıqlı əlaqə dilçilik tədqiqatlarında mühüm metodoloji əsas yaradır. Məqalə çərçivəsində təqdim edilən tədqiqatımız qeyd edilən iki vasitənin bir-birini necə tamamladığını, onların birlikdə istifadəsinin dilçilik tədqiqatlarına gətirdiyi yenilikləri və üstünlükləri araşdırmağa yönəlib. Belə bir yanaşma dilin daha dərin və hərtərəfli öyrənilməsinə töhfə verir, bu isə müasir dilçilikdə mühüm əhəmiyyət kəsb edir.

#### **Metod və materiallar**

Məqalə çərçivəsində təqdim edilən bu elmi araşdırmada dilçiliyin təsviri metodundan istifadə edilmiş, empirik materiallara müraciət olunmuşdur.

#### **Tədqiqatın əsas hissəsi**

##### *Müasir korpus dilçiliyi: konkordansların təbii və kollokasiya*

Konkordansların əsasında yaradılan dil korpusları bu və ya digər bir dilin materiallarını vahid bazada birləşdirir. Bu planda müasir dilçi-tədqiqatçılar getdikcə daha çox linqvistik korpusların məlumatlarına müraciət edirlər. XX əsrin 90-cı illərinin ilk yarısında kompüter texnologiyalarının inkişafı nəticəsində müstəqil bir dilçilik sahəsi kimi formalaşan korpus dilçiliyi sahəsi empirik materialların yığılması və təhlilində aparıcı mövqe tutur [Habibova: 2023, s.174]. Hazırda statistik məlumatlar əsasında dil korpusları tərəfindən təqdim edilən məlumatlara istinad etməklə mətnlər üzərində geniş leksikoqrafik, qrammatik, semantik və diskursiv təhlillər aparmaq mümkündür. Dilçilik araşdırmaları zamanı korpus dilçiliyinin metodoloji bazası inkişaf edir, yeni kompüter alətləri və yeni korpuslar yaradılır. Lakin dilçilik tədqiqatlarında korpusların istifadəsini məhdudlaşdıran ciddi problemlər də mövcuddur.

Hazırda dünyanın əksər əsas dilləri üçün milli dil korpusları fəaliyyət göstərir. Bəzi müəlliflər mövcud olan dil korpusları haqqında məlumatlar təqdim edirlər. Məsələn, V.P.Zaxarov milli və digər korpuslar haqqında məlumatı cədvəl şəklində təqdim edir

[Захаров: 2020, s.85-87]. M.Kopotev “Korpus dilçiliyinə giriş” adlı əsərində CLARIN [[www.clarin.eu](http://www.clarin.eu)] və ELRA [[www.elra.info](http://www.elra.info)] kimi ixtisaslaşdırılmış kataloqları qeyd edir [Копотев: 2014, s.16]. Bu kataloqlar üç mindən çox dil korpusuna dair məlumatları ehtiva edir.

M.V.Kopotev və V.P.Zaxarovun əsərlərində təqdim olunan məlumatlar, korpusların struktur xüsusiyyətləri və onların linqvistik tədqiqatlarda istifadəsi ilə bağlı mühüm elmi əsaslar ortaya qoyur. V.P.Zaxarov öz tədqiqatlarında milli və beynəlxalq korpusları strukturlaşdırılmış cədvəllər vasitəsilə müqayisəli formada təqdim edərək, onların annotasiya səviyyələri, verilənlər bazasının həcmi, dil resurslarının müxtəlifliyi və tətbiq sahələrini sistemləşdirir. Onun tədqiqatlarında əsas vurğulanan məqam korpusların təkcə linqvistik tədqiqatlar üçün deyil, həm də süni intellekt və maşın tərcüməsi sahələrində istifadəsinin vacibliyidir. O, eyni zamanda milli korpusların sosial-linqvistik xüsusiyyətlərinin öyrənilməsinə təsirini də qeyd edir və göstərir ki, müasir dil korpusları yalnız birbaşa dil tədqiqatları üçün deyil, həm də mədəni və sosial tədqiqatlar üçün dəyərli mənbə rolunu oynayır.

Digər tərəfdən, M.V.Kopotev öz əsərində CLARIN və ELRA kimi beynəlxalq platformaların mühüm rol oynadığını vurğulayır. O qeyd edir ki, *“bu sistemlərə tədqiqatçılara yalnız məlumat bazasına çıxış vermir, həm də korpusların annotasiya, metadataların idarə edilməsi və dil resurslarının paylanması üçün standartlaşdırılmış metodologiyalar təklif edir”* [Копотев: 2014, s.24]. M.V.Kopotev korpusların interaktiv istifadəsini, onların nitq texnologiyalarında tətbiqini və avtomatik təhlil imkanlarını geniş müzakirə edir. Onun fikrincə, müasir korpuslar təkcə leksik və sintaktik tədqiqatlar üçün deyil, həm də dil öyrənmə, avtomatik nitq tanıma və təbii dil emalı sahələrində tətbiq edilməlidir. Həm V.P.Zaxarovun korpusların strukturlaşdırılması və tətbiq sahələri ilə bağlı yanaşmaları, həm də M.V.Kopotevin interaktiv və texnoloji inkişaflara yönəlmiş baxışı göstərir ki, müasir dilçilikdə korpusların rolu təkcə linqvistik tədqiqatlarla məhdudlaşmır. Bu yanaşmaların sintez edilməsi, milli və beynəlxalq səviyyədə yeni tədqiqatların aparılması üçün geniş imkanlar yaradır.

Haqında bəhs edilən kataloqlar haqqında da bir qədər məlumat verək: ELRA (European Language Resources Association) [[www.elra.info](http://www.elra.info)], dil texnologiyaları və korpus dilçiliyi sahəsində istifadə edilən dil resurslarını (mətn və nitq korpusları, lüğətlər, annotasiya alətləri və s.) toplayan və yayılmasını təşkil edən Avropa təşkilatıdır. Bu kataloq vasitəsilə müxtəlif dillərdə yaradılmış mətn və nitq korpusları, onların strukturu, annotasiya səviyyələri və lisenziya şərtləri haqqında məlumat əldə etmək mümkündür. ELRA həmçinin dil resurslarının standartlaşdırılması və geniş istifadəsini təşviq edən təşkilatlardan biridir.

CLARIN (Common Language Resources and Technology Infrastructure) [[www.clarin.eu](http://www.clarin.eu)], əsasən humanitar və sosial elmlərdə istifadə olunan dil resurslarının və texnologiyalarının əlçatanlığını artırmaq məqsədilə yaradılmış Avropa tədqiqat infrastrukturunu şəbəkəsidir. CLARIN müxtəlif ölkələrin dil resurslarını birləşdirən bir texnoloji ekosistem yaradır və tədqiqatçılara bu resurslara vahid platforma vasitəsilə çıxış imkanı verir. Bu kataloq vasitəsilə müxtəlif dillər üçün hazırlanmış mətn və nitq korpusları, linqvistik annotasiya alətləri və digər dil texnologiyaları haqqında geniş məlumat almaq mümkündür.

İlk növbədə, korpus məlumatları əsasında tədqiqatlar aparan elmi araşdırma mərkəzlərinin və eksperimental dilçilik laboratoriyalarının yaradılması və uğurla fəaliyyətdə olması qeyd edilməlidir. Bu laboratoriyalar Avropanın bir çox ölkələrində fəaliyyət göstərir. Məsələn, Moskva Dövlət Dilçilik Universitetində Ye.Qolubkovanın rəhbərliyi altında korpus dilçiliyi laboratoriyası fəaliyyət göstərir. Laboratoriya korpus məlumatları əsasında

dildə baş verən yeni hadisələri və dəyişikliklərin tendensiyalarını araşdırır, multimodal korpusların yaradılması, dilçilik proqram təminatının hazırlanması və korpus dilçiliyi metodları ilə həyata keçirilən layihələrin reallaşdırılması ilə məşğuldur. Dil korpusları kontekstdə leksik uyğunluğun öyrənilməsində genişmiqyaslı imkanlar təqdim edir. Bu məsələ bir çox dilçi alimlərin maraq dairəsindədir. Məsələn, T.Yu.Pavelyeva, Böyük Britaniya Milli Korpusunda ingilis dilindəki *time* sözünün kollokasiyalarını araşdıraraq onları dayanıqlıq dərəcəsinə görə təsnif edir [Павельева: 2016, s.59].

Ümumiyyətlə, kollokasiyaların müəyyənəşdirilməsi və tədqiqi məsələsində ənənəvi dilçilik və korpus dilçiliyi arasında ciddi metodoloji fərqlər mövcuddur. Ənənəvi yanaşmaya görə, kollokasiyalar semantik və qrammatik uyğunluğa malik sabit söz birləşmələri kimi qəbul edilir və onların təhlilində, əsasən lüğətçi və ya filoloq yanaşması tətbiq olunur. Lakin korpus dilçiliyində kollokasiyalar statistik göstəricilər əsasında müəyyən edilən dayanıqlı leksik kombinasiyalar kimi nəzərdən keçirilir. Bu fərqi mahiyyəti ondan ibarətdir ki, ənənəvi yanaşma kollokasiyaların sabitliyini dilin daxili qaydaları və normativlik prinsipi çərçivəsində təhlil edir, halbuki korpus dilçiliyi kollokasiyaların tezlik, qarşılıqlı təsir və kontekstual sabitliyi üzərində dayanır.

Bu məsələyə T.Yu.Pavelyeva da öz tədqiqatında toxunmuşdur və korpus əsaslı kollokasiya analizində MI (Mutual Information – qarşılıqlı informasiya əmsalı) kimi statistik metodların istifadəsini əsaslandırmışdır. Onun yanaşması A.N.Baranov və D.O.Dobrovolski tərəfindən irəli sürülən kollokasiya və frazeoloji vahidlərin əlaqələndirilməsi məsələsinə də toxunur. Həmin alimlər hesab edirlər ki, kollokasiyalar semantik baxımdan bütünlüklə sərbəst və ya tamamilə sabit vahidlər arasında yerləşən, müəyyən dərəcədə məna məhdudiyətinə malik leksik konstruksiyalardır [Баранов, Добровольский: 2008, s.65].

Maraqlıdır ki, T.Y.Pavelyeva məhz korpus yanaşması çərçivəsində frazeologizmlərlə kollokasiyalar arasındakı fərqləri vurğulayır. O qeyd edir ki, korpus əsaslı kollokasiya analizində əsas komponent (baza) leksik mənasını qoruyub saxlayır, lakin obrazlı məna və emosional yük zəifləyir və ya tamamilə itir, bununla belə, söz birləşməsinin struktur dayanıqlılığı və kontekstual sabitliyi qorunur [Павельева: 2016, s.57-58]. Burada mühüm məqamlardan biri də konkordansların kollokasiya analizində əsas alət kimi tətbiq olunmasıdır. Konkordanslar leksik vahidin müxtəlif kontekstlərdə necə istifadə edildiyini əks etdirən mətn parçalarıdır və onların köməyi ilə sözlərin semantik əlaqələri, tezlik göstəriciləri və dayanıqlıq səviyyələri müəyyən edilir. T.Y.Pavelyevanın araşdırmasının nəticələri göstərir ki, leksik vahidlərin müxtəlif nitq hissələri ilə kombinasiyaları onların sabitlik səviyyəsini və kontekstual uyğunluğunu müəyyən etməyə imkan verir [Павельева: 2016, s.58].

Bu müzakirədən belə nəticəyə gəlmək olar ki, ənənəvi və korpus əsaslı yanaşmalar kollokasiyaların öyrənilməsində fərqli baxış bucaqları təqdim etsə də, müasir dilçilikdə statistik metodların tətbiqi kollokasiyaların daha obyektiv və empirik əsaslarla müəyyən edilməsinə şərait yaradır. Bununla belə, korpus dilçiliyinin yalnız statistik göstəricilərə əsaslanması da müəyyən məhdudiyətlər yarada bilər, çünki kollokasiyaların kontekstual semantikasi və funksional yüklənməsi yalnız rəqəmlərlə ölçülə bilməz. Bu səbəbdən, ən optimal yanaşma hər iki metodu – həm ənənəvi, həm də korpus əsaslı analiz üsullarını integrativ şəkildə tətbiq etməkdən ibarətdir.

*Korpus yanaşması və elmi diskurs: kollokasiya və konkordansın yeni aspektləri*

Əgər ənənəvi yazılı mətn korpusları, əsasən konkordans vasitəsilə söz və ifadələrin istifadəsini araşdırmaq üçün nəzərdə tutulurdusa, A.Vyaçorkovska nitq korpusları

avtomatik nitq tanıma (ASR), nitq sintezi və generativ süni intellekt modellərinin öyrədilməsini araşdıran tədqiqatında dil korpuslarının yalnız yazılı mətnlər üçün deyil, həm də nitq əsaslı tədqiqatlar üçün əhəmiyyətli olmasını xüsusilə vurğulayır. Müəllif bildirir ki, nitq korpusları yalnız insan-maşın ünsiyyəti üçün deyil, həm də lüğətçilik, avtomatik tərcümə və nitq emalı texnologiyaları üçün kritik əhəmiyyət daşıyır.

Bu yanaşma ənənəvi korpus dilçiliyində kollokasiya və konkordans anlayışlarının necə dəyişdiyini göstərir. Əgər yazılı mətnlərdə kollokasiyalar daha çox semantik uyğunluğa əsasən müəyyən edilirdisə, nitq korpuslarında bu kombinasiyalar statistika və akustik parametrlərlə analiz edilir. Bu, ənənəvi korpus dilçiliyində qəbul edilən “stabil kollokasiyalar” anlayışının nitq korpuslarında daha mürəkkəb bir quruluşa malik olduğunu göstərir [Wieczorkowska: 2025].

İ.A.Anaşına və İ.İ.Konkova istifadəçi korpuslarının tədqiqi üçün nəzərdə tutulan *AntConc* adlı proqram məhsulunun sınağını aparmış və bu proqramdan robototexnika sahəsinə aid elmi-texniki diskurs korpuslarının leksik tərkibinin xüsusiyyətlərini aşkar etmək üçün istifadə etmişlər. Korpus alətləri olan *Word List*, *Concordance* və *Keyword List*-dən istifadə edərək müəlliflər bu tip diskursu xarakterizə edən leksik vahidlərin siyahılarını təqdim etmiş, analiz olunan mətnin əsas ideyasını əks etdirən leksik vahidlərin tipik istifadəsini və əhatəsini nümayiş etdirmişlər. Daha sonra tədqiqatçılar ən çox istifadə olunan leksik vahidləri müəyyənləşdirərək onları nitq hissələrinə görə sıralamış, kollokatların statistik analizini aparmışlar [Анашкина, Конькова: 2021, s.33-35]. Qeyd edilir ki, “*tezlik təhlilindən əlavə, korpus dilçiliyi kollokasiyaların, yəni sözün tipik və sabit əhatə mühitinin tədqiqinə imkan verir. Bu tapşırığın yerinə yetirilməsi üçün korpus dilçiliyi kollokasiyaların müəyyənləşdirilməsi üçün istifadə olunan ehtimal-statistik metodlara əsaslanır*” [Анашкина, Конькова: 2021, s.34] Müəlliflər bu tip diskursda statistik tədqiqatlar aparmaq üçün korpuslardan istifadənin məhsuldarlığı ilə bağlı qənaətə gəlmişlər və bu yanaşmanın digər diskurs tiplərinin tədqiqatlarında da eyni dərəcədə səmərəli olduğunu vurğulamışlar. Bu tədqiqat *AntConc* vasitəsilə elmi-texniki diskursun leksik xüsusiyyətlərini təhlil edərək, korpus əsaslı metodların effektivliyini nümayiş etdirir. “*Word List*” və “*Keyword List*” funksiyaları göstərir ki, analiz edilən mətnlərdə isimlər, sifətlər və feli sifətlər üstünlük təşkil edir və əsas açar sözlər (məs., *robot*, *robotics*, *calibration*) sistemli şəkildə təkrarlanır.

Burada tətbiq edilən “*Concordance*” funksiyası “*robot*” termininin prepozisiya və postpozisiya mövqelərində hansı qrammatik elementlərlə birləşdiyini müəyyən edib. “*Concordance Plot*” terminlərin bərabər paylanmasını təsdiqləyərək elmi diskursun sistemli rekurrensiya prinsipinə əsaslandığını göstərir. “*Collocates*” funksiyası isə ən çox rast gəlinən kollokasiyaları müəyyən edərək, elmi üslubda terminoloji sabitliyin statistik baxımdan da sübut olunduğunu ortaya qoyur. Bu tədqiqat sübut edir ki, korpus əsaslı diskurs analizi elmi mətnlərin strukturunu, terminoloji sabitliyini və leksik inkişafını araşdırmaq üçün mühüm metodoloji alətdir. Konkordans və kollokasiya analizləri elmi üslubun leksik sabitliyini və terminlərin sistemləşdirilmiş istifadəsini müəyyən etməyə imkan verir. S.Mustafayeva qeyd edir ki, “*korpusda nəinki şifahi mətn nümunələri, həm də yazılı mətn nümunələri yer ala bilər. Korpusa balanslaşdırılmış və təmsilçi struktur vermək üçün praktiki olaraq mümkün olduğu qədər dil istifadəsinin bütün sahələrindən nümunələr tərtib edilməlidir*” [Mustafayeva: 2023, s.152].

Bu yanaşma göstərir ki, elmi diskursun leksik və sintaktik strukturunun öyrənilməsi üçün korpus əsaslı metodlar vacib rol oynayır. S.Mustafayeva tərəfindən vurğulanan balanslaşdırılmış və təmsilçi korpusların əhəmiyyəti tədqiqatın nəticələri ilə üst-üstə düşür.

Korpusların yalnız yazılı mətnləri deyil, həm də şifahi nitq nümunələrini əhatə etməsi, onların daha geniş diskurs analizlərinə imkan verməsini təmin edir. Bununla yanaşı, tədqiqat elmi mətnlərdə terminlərin bərabər paylanması və onların digər leksik vahidlərlə kollokativ əlaqələrinin sistemli xarakter daşdığını sübut edir. Bu nəticə onu göstərir ki, elmi diskurs yalnız spesifik terminlərin istifadəsinə deyil, həm də bu terminlərin sintaktik və semantik mövqələrinə əsaslanan müəyyən struktur çərçivəsinə malikdir. Bu tədqiqat elmi diskursun terminoloji sabitliyini və leksik strukturunu təhlil etməklə yanaşı, dil texnologiyalarının avtomatik mətn emalı sahəsində tətbiqini də aktuallaşdırır. Korpus əsaslı diskurs analizinin nəticələri, mətnlərin avtomatik redaktəsi və düzəlişi üçün nəzərdə tutulmuş texnoloji vasitələrlə sıx əlaqəlidir.

Bu baxımdan, R.İsmayılovun qeyd etdiyi “Windows” əməliyyat sistemində avtomatik mətn düzəliş funksionallığının mövcudluğu [İsmayılov: 2023] korpus əsaslı analizlərin praktiki istifadəsinin bir nümunəsidir. Korpuslar və konkordans alətləri yalnız akademik araşdırmalar üçün deyil, həm də mətnin avtomatik analizi, səhv aşkarlanması və düzəlişi üçün mühüm mənbə rolunu oynayır. “*Məsələn, “Key Switcher” proqramı mətnlərdəki səhvləri müəyyən edib avtomatik düzəltməklə, məhz dil vahidlərinin statistikasına və tezliyinə əsaslanan yanaşmadan istifadə edir. Bu proqram təminatı korpus əsaslı yanaşmaların praktik tətbiq sahələrindən biri kimi çıxış edir*” [İsmayılov: 2023].

Eyni zamanda, “Grammarly” kimi platformalar mətnlərin statistik və linqvistik təhlili əsasında avtomatik düzəlişlər həyata keçirir, bu da diskursun strukturunu qorumaq və akademik üslubun normativliyini artırmaq üçün mühüm vasitələrdən biridir. “Punto Switcher” isə klaviatura girişlərini avtomatik tanımaq və transliterasiya əməliyyatlarını icra etmək üçün nəzərdə tutulmuş bir sistem kimi [İsmayılov: 2023], məhz mətnlərin fonetik və qrafik uyğunluğunu tənzimləyir.

Bu nümunələr göstərir ki, elmi mətnlərdə terminlərin və linqvistik vahidlərin avtomatik təhlili və düzəlişi üçün müasir proqram təminatları korpus əsaslı metodologiyaya əsaslanır. Terminlərin kollokativ əlaqələrinin müəyyən edilməsi, diskurs daxilində onların bərabər paylanması analizi və mətnin normativliyinin yoxlanılması, korpus texnologiyalarının akademik mətnlərin redaktəsi və linqvistik emalı sahəsində tətbiqini artırır. Beləliklə, diskurs analizində istifadə olunan korpus əsaslı metodlar, avtomatik mətn emalı və redaktə sistemləri ilə birləşdirildikdə, elmi üslubun strukturunu qorumağa və terminoloji sabitliyi təmin etməyə imkan yaradır. Bu yanaşma, korpus texnologiyalarının dilçilik, süni intellekt və avtomatik mətn analizi sahələrində nə qədər geniş istifadə olunduğunu bir daha sübut edir.

Bu tədqiqat korpus əsaslı metodların dilçilik tədqiqatlarında, xüsusilə konkordans və kollokasiya analizlərində effektivliyini göstərir. Konkordansların müasir dil korpusları ilə inteqrasiyası nəticəsində mətnlərin struktur və semantik analizi daha dəqiq və sistemli şəkildə aparıla bilər. Müxtəlif korpus əsaslı tədqiqatların müqayisəsi göstərdi ki, elmi diskursda leksik sabitlik, terminoloji struktur və nitq vahidlərinin istifadəsi konkordans və kollokasiya analizləri ilə uğurla müəyyən edilə bilər.

Nitq korpusları və avtomatik mətn emalı sahəsində aparılan tədqiqatlar göstərdi ki, müasir korpus texnologiyaları yalnız yazılı mətnlər üçün deyil, həm də nitq məlumatlarının təhlili üçün əhəmiyyətli rol oynayır. A.Vyaçorkovska tərəfindən nitq tanıma və generativ süni intellekt modellərinin öyrədilməsi üçün dil korpuslarının istifadəsinə dair tədqiqatı bu sahənin inkişaf perspektivlərini nümayiş etdirir. Belə nəticəyə gəlinmişdir ki, yazılı və şifahi mətnlərin birlikdə korpus analizinə daxil edilməsi, dilin müxtəlif istifadələrini və semantik variasiyalarını daha dəqiq tədqiq etməyə imkan yaradır.

## ƏDƏBİYYAT

1. Əliyeva N. və b. Korpus dilçiliyinin təşkilində leksikoqrafiyanın rolu və vəzifələri / N.Əliyeva, R.Heydərov // Ümummilli lider Heydər Əliyevin 100 illik yubileyinə və Beynəlxalq Ana dili gününə həsr olunmuş “Azərbaycan dilinin İKT problemləri, İKT-nin Azərbaycan dili problemləri” respublika elmi-praktiki konfransı, 21-22 fevral 2023-cü il.

2. İsmayılov R. İnternet resurslarında milli dil korpusunun inkişaf etdirilməsi məsələləri 04.04.2023[14:14]/[https://azertag.az/xeber/internet\\_resurslarinda\\_milli\\_dil\\_korpusunun\\_inkisaf\\_etdirilmesi\\_meseleleri\\_arasdirma-2556842](https://azertag.az/xeber/internet_resurslarinda_milli_dil_korpusunun_inkisaf_etdirilmesi_meseleleri_arasdirma-2556842)

3. Məmmədova R. Dil korpuslarının tədrisdə istifadəsi // Sivilizasiya, 2019. C.8, №4

4. Mustafayeva S. Müasir dilçilikdə korpus dilçiliyinin yeri // İpək yolu, 2023. №3.

5. Habibova K. Virtual discourse as a communicative requirement of the modern era. Ad Alta // Journal of Interdisciplinary research. 2023. V. 13. №. 1, pp. 173-176.

6. Heydərov R. A., Jafarov A. G., Habibova K. A. Software for experimental methods in phonetic research // 1st International Scientific Conference “Modern Management Trends and the Digital Economy: from Regional Development to Global Economic Growth” (MTDE 2019). Atlantis Press,

7. Анакшина И.А. Корпусно-ориентированный метод дискурс-анализа (на материале английского языка) / И.А. Анакшина, И.И. Конькова // Вестник ПНИПУ. Проблемы языкознания и педагогики. 2021. №2.

8. Баранов А.Н. Аспекты теории фразеологии / А.Н. Баранов, Д.О. Добровольский. М.:2008.

9. Захаров В.П. Корпусная лингвистика / В.П. Захаров, С.Ю. Богданова. С.-Петербург, 2020.

10. Копотев М.В. Введение в корпусную лингвистику. Москва: Animedia Company, 2014.

11. Павельева Т.Ю. Изучение коллокаций на основе лингвистических корпусов текстов // Вестник Тамбовского университета. Серия Гуман.науки. 2016. Т. 21. №3-4.

12. ELRA — [www.elra.info](http://www.elra.info)

13. CLARIN — [www.clarin.eu](http://www.clarin.eu)

## КОРПУСНАЯ ЛИНГВИСТИКА И КОНКОРДАНС: МЕТОДОЛОГИЧЕСКИЕ ПОДХОДЫ И НОВЫЕ ПРИМЕНЕНИЯ РЕЗЮМЕ

В данной статье рассматриваются основные понятия современной корпусной лингвистики, в частности, применение анализа конкорданса и коллокаций, их роль в изучении научного дискурса и поэтических текстов. Установлено, что развитие корпусной лингвистики должно осуществляться совместно с созданием национальных корпусов, автоматической обработкой текста и интеграцией технологий искусственного интеллекта. Широкое применение методов конкорданса и коллокации в научном дискурсе, медиа и поэтических текстах может способствовать развитию новых методологических подходов в азербайджанском языкознании.

**Ключевые слова:** *корпусная лингвистика, конкорданс, коллокация, дискурсивный анализ, искусственный интеллект, национальный корпус.*

## CORPUS LINGUISTICS AND CONCORDANCE: METHODOLOGICAL APPROACHES AND NEW APPLICATIONS

### ABSTRACT

This article examines the key concepts of modern corpus linguistics, particularly the application of concordance and collocation analysis, and their role in studying scientific discourse and poetic texts. It has been established that the development of corpus linguistics should be carried out alongside the creation of national corpora, automatic text processing, and the integration of artificial intelligence technologies. The extensive use of concordance and collocation methods in scientific discourse, media, and poetic text analysis can contribute to the development of new methodological approaches in Azerbaijani linguistics.

**Keywords:** *corpus linguistics, concordance, collocation, discourse analysis, artificial intelligence, national language corpus.*