

UOT-811.11-112

Səhifə:84-93

<https://doi.org/10.59849/2663-8967.2025.1.84>

Könül Həbibova
AMEA Nəsimi adına Dilçilik İnstitutu
filologiya üzrə fəlsəfə doktoru
habibovakonul@mail.ru
ORCID ID: [0000-0001-8502-229X](https://orcid.org/0000-0001-8502-229X)

KORPUS DİLÇİLİYİNDƏ TƏMƏL YANAŞMALAR VƏ AZƏRBAYCAN DİLÇİLİYİNDƏ BU SAHƏ İLƏ BAĞLI PERSPEKTİVLƏR

XÜLASƏ

Bu məqalə müasir dilçilikdə dil korpusu və korpus dilçiliyi anlayışlarının nəzəri əsaslarını və onların tədqiqat sahələrində tətbiqini araşdırır. Dil korpusu müxtəlif dillərə aid yazılı və şifahi materialların strukturlaşdırılmış məlumat bazası kimi təqdim edilir, korpus dilçiliyi isə bu məlumatlardan istifadə edərək dilin müxtəlif aspektlərini empirik üsullarla öyrənən elmi sahə kimi izah olunur. Məqalədə L.Lemnitzer və H.Zinsmeyster, C.Sinclair, M.Stabbs, V.P.Zaxarov, N.N.Kazanski, M.Mahmudov kimi tədqiqatçıların dil korpusuna yanaşmaları müqayisə edilir və onların nəzəri-metodoloji əsasları təhlil olunur. Araşdırmada dil korpuslarının leksikoqrafiya, semantika, sintaksis, tərcümə prosesi, maşın tərcüməsi və nitq texnologiyaları kimi sahələrdə rolu vurğulanır. Bununla yanaşı, Azərbaycan dilinin milli korpusunun yaradılması, bu sahədə aparılan tədqiqatlar və texnoloji imkanlar da müzakirə edilir. AMEA-nın İ.Nəsimi adına Dilçilik İnstitutunun Kompüter dilçiliyi şöbəsinin bu istiqamətdə apardığı tədqiqatlar, milli korpusun inkişafı üçün atılan addımlar haqqında da məqalədə məlumat verilir. Məqalə korpus dilçiliyinin təkcə kompüter dilçiliyinin bir qolu kimi deyil, həm də müstəqil empirik tədqiqat metodu kimi əhəmiyyətini vurğulayır. Nəticədə korpus dilçiliyinin filoloji tədqiqatlarla əlaqəsi, onun dilin semantik və sintaktik xüsusiyyətlərinin daha dərinlən öyrənilməsinə töhfəsi və gələcək inkişaf perspektivləri araşdırılır. Bu tədqiqat dil korpuslarının təkcə elmi-lingvistik tədqiqatlar üçün deyil, həmçinin pedaqoji və texnoloji tətbiqlər üçün də mühüm əhəmiyyət daşıdığını göstərir.

Açar sözlər: *dil korpusu, korpus dilçiliyi, məlumat bazası, elektron lüğət, nitq texnologiyaları.*

Giriş: Müasir dilçilikdə dil korpusları və korpus dilçiliyi anlayışları arasında fərqlər və onların qarşılıqlı əlaqəsi tədqiqatçıların diqqət mərkəzindədir. Dil korpusu müxtəlif dillərə aid yazılı və şifahi materialların sistemlik şəkildə toplanmış və strukturlaşdırılmış məlumat bazasıdır, korpus dilçiliyi isə bu məlumatlardan istifadə edərək dilin müxtəlif aspektlərini empirik üsullarla öyrənən elmi sahədir. Korpus əsaslı tədqiqatlar dilçiliyin müxtəlif istiqamətlərində — semantika, sintaksis, diskurs təhlili, nitq aktlarının öyrənilməsi, maşın tərcüməsi və dil texnologiyalarının inkişafında mühüm rol oynayır.

Korpus dilçiliyinin kompüter dilçiliyi kontekstində daha aydın və başa düşülən olmasını əsas gətirən S.Məmmədzaadə dilçiliyin bu sahəsinin kompüter dilçiliyinin bir qolu kimi tərif edir [4, s.105]. Əvvəllər “milli korpus” anlayışının “maşın fondu” kimi işləndiyi söyləyən tədqiqatçı qeyd edir ki, hazırda hər iki termin paralel olaraq işlədilir [4, s.105].

S.Məmmədzaadənin yanaşması müəyyən qədər məntiqli və əsaslandırılmış olsa da, korpus dilçiliyinin yalnız kompüter dilçiliyinin bir qolu kimi təqdim edilməsini biz mübahisəli məsələ kimi dəyərləndiririk. Bunun əsas səbəbi korpus dilçiliyinin yalnız kompüter texnologiyalarına əsaslanmaması, həm də empirik dil tədqiqatlarının bir metodu kimi daha geniş çərçivədə tətbiq olunmasıdır. Korpus dilçiliyi dilin real istifadəsinə əsaslanan mətn və nitq nümunələrinin sistemli şəkildə toplanması, strukturlaşdırılması və təhlilini əhatə edir. Kompüter dilçiliyi isə bu

məlumatlardan istifadə edərək avtomatlaşdırılmış dil emalı alqoritmləri və proqram təminatları yaratmağa istiqamətlənmiş daha geniş bir sahədir. Bu baxımdan korpus dilçiliyi kompüter dilçiliyinin əsas metodoloji bazalarından biri kimi qəbul olunsada, onu sırf kompüter dilçiliyinin bir alt sahəsi kimi qiymətləndirmək mübahisəli ola bilər.

Korpus dilçiliyinin tarixinə nəzər saldıqda görmək olar ki, bu sahə kompüter texnologiyalarından əvvəl də mövcud olub. Məsələn, XX əsrin əvvəllərində əl ilə tərtib edilən konkordanslar və dilin statistik təhlili metodları korpus əsaslı yanaşmanın ilkin formaları kimi qəbul edilə bilər. C.Sinkler, L.Lemnitzer və M.Stabbs kimi tanınmış dilçi alimlər korpus dilçiliyini müstəqil empirik tədqiqat metodu kimi dəyərləndirirlər. Onların fikrincə, korpus əsaslı təhlillər dilin leksik, qrammatik və sintaktik xüsusiyyətlərini obyektiv şəkildə tədqiq etməyə imkan yaradır və bu, yalnız kompüter texnologiyalarının tətbiq edildiyi bir sahə deyil. Müasir texnologiyalar korpus dilçiliyinin tətbiq sahələrini genişləndirsə də, onu tamamilə kompüter dilçiliyinin bir alt sahəsi kimi təqdim etmək düzgün olmazdı.

S.Məmmədzaadənin yanaşmasının digər bir tərəfi “milli korpus” və “maşın fondu” anlayışlarının paralelliyi ilə bağlıdır. Onun fikrincə, əvvəlki dövrlərdə “milli korpus” anlayışının ekvivalenti kimi “maşın fondu” ifadəsi işlədilmişdir və bu gün hər iki termin paralel olaraq istifadə olunur. Lakin bu iki anlayış arasında fərqlər mövcuddur. “Milli korpus” daha geniş bir çərçivədə dilin müxtəlif aspektlərini əhatə edən, strukturlaşdırılmış elektron resurs kimi qəbul edilir. Digər tərəfdən, “maşın fondu” termini, əsasən sovet dövründə istifadə edilmiş və dil materiallarının kompüter əsaslı tərcümə və lüğətçilik sahələrində istifadəsi ilə bağlı olmuşdur. Müasir dövrdə “milli korpus” konsepsiyası daha çox elmi-lingvistik tədqiqatlar, leksikoqrafiya, diskurs təhlili, maşın tərcüməsi və nitq texnologiyaları üçün nəzərdə tutulan sistemli bir məlumat bazasıdır.

Beləliklə, S.Məmmədzaadənin korpus dilçiliyini kompüter dilçiliyinin bir qolu kimi izah etməsi müəyyən dərəcədə məntiqli olsa da, bu yanaşma tamamilə doğru hesab edilə bilməz. Korpus dilçiliyi kompüter dilçiliyinin əsas metodlarından biri olmaqla yanaşı, eyni zamanda müstəqil bir empirik tədqiqat sahəsi kimi də fəaliyyət göstərir. Onun tətbiq dairəsi yalnız kompüter əsaslı dil emalı sistemləri ilə məhdudlaşmır, eyni zamanda, ənənəvi filoloji araşdırmalarda, dilin struktur və semantik analizində də mühüm rol oynayır. Buna görə də korpus dilçiliyini yalnız kompüter dilçiliyinin bir hissəsi kimi deyil, onunla sıx əlaqəli, lakin müstəqil bir sahə kimi dəyərləndirmək daha məqsəduyğundur.

Məhz bu cəhətdən qeyd edilməlidir ki, dil korpusunun tərfi və onun struktur xüsusiyyətləri ilə bağlı dilçilikdə müxtəlif yanaşmalar mövcuddur. L.Lemnitzer və H.Zinsmeyster (2006), C.Sinkler (1991) və M.Stabbs (2001) kimi dilçi alimlər dil korpusunun tərfini fərqli aspektlərdən izah ediblər. L.Lemnitzer və H.Zinsmeyster dil korpusunu müəyyən bir dildən və ya dil növündən toplanmış, sistemləşdirilmiş və annotasiya edilmiş yazılı və ya şifahi materiallar bazası kimi təsvir edirlər [5]. C.Sinkler isə korpusun real dil istifadəsini əks etdirməli olduğunu və onun kollokasiya, kontekst və tezlik əsaslı tədqiqatlar üçün mühüm vasitə olduğunu vurğulayır [6]. M.Stabbs korpusları leksik semantikanın və kontekstual tədqiqatların aparılması üçün fundamental resurs hesab edir [7].

Dünya dilçiliyində dil korpuslarına yanaşmaların müqayisəli təhlili göstərir ki, bu anlayışın tərfi yalnız onun struktur və məzmun xüsusiyyətlərinə deyil, həm də tədqiqat metodologiyasına əsaslanır. V.P.Zaxarov (2020) korpusları elektron formatda təqdim olunan, strukturlaşdırılmış, annotasiya edilmiş və filoloji baxımdan kompetent məlumat toplusu kimi xarakterizə edir [8]. L.Lemnitzer isə korpusların tərkibində yalnız mətnlər və metadataların deyil, həm də leksik annotasiyaların olması zərurətini vurğulayır [5, s.11].

Bu məqalədə dil korpusunun müxtəlif alimlər tərəfindən verilən tərifləri və onların elmi-metodoloji əsasları araşdırılması nəzərdə tutulur. Eyni zamanda, korpusların lingvistik tədqiqatlarda, tərcümə prosesində, maşın tərcüməsində, nitq tanıma texnologiyalarında və dil tədrisində tətbiqi imkanları müzakirə ediləcəkdir. Məqalənin məqsədi, dil korpusu və korpus dilçiliyi anlayışlarının qarşılıqlı əlaqəsini izah etməklə yanaşı, dünya dilçiliyində formalaşan təməl yanaşmaları müqayisəli şəkildə təhlil etməkdir.

Metod və materiallar. Bu araşdırmada dil korpusu və korpus dilçiliyi ilə bağlı əsas nəzəri yanaşmaların müqayisəli təhlili aparılmışdır. Tədqiqat çərçivəsində fərqli dil korpusu modelləri, onların struktur xüsusiyyətləri və tətbiq sahələri nəzərdən keçirilmiş, müxtəlif alimlərin, o cümlədən L.Lemnitser və H.Zinsmeyster, C.Sinkler, M.Stabbs, V.P.Zaxarov, N.N.Kazanski və M.Mahmudov kimi tədqiqatçıların yanaşmaları müqayisə edilmişdir. Tədqiqatda müxtəlif metodoloji yanaşmalardan istifadə olunmuşdur. İlk olaraq, təsviri və müqayisəli təhlil vasitəsilə fərqli nəzəri yanaşmaların təhlili və onların bir-biri ilə müqayisəsi aparılmışdır.

Araşdırmada əsas materiallar kimi müxtəlif nəzəri və praktiki mənbələrdən istifadə edilmişdir. Elmi-nəzəri mənbələr sırasında dil korpusu və korpus dilçiliyi ilə bağlı fundamental elmi əsərlər və tədqiqat materialları yer alır. Azərbaycan dilinin milli korpusu ilə bağlı mövcud tədqiqatlar və həyata keçirilən layihələr də araşdırmanın əsas materiallarını təşkil etmişdir. Bu çərçivədə M.Mahmudov və digər tədqiqatçıların korpus dilçiliyi ilə bağlı apardığı araşdırmalar nəzərdən keçirilmişdir. Bundan əlavə, AMEA-nın İ.Nəsimi adına Dilçilik İnstitutunun Kompüter dilçiliyi şöbəsi tərəfindən aparılan tədqiqatlar, milli korpusun formalaşdırılması istiqamətində atılan addımlar, leksikoqrafiya altkorpusunun yaradılması və elektron lüğət bazalarının formalaşdırılması ilə bağlı aparılan işlər qiymətləndirilmişdir.

Tədqiqatın əsas hissəsi. Hazırda dil korpusu ilə bağlı bir neçə tərif mövcuddur. L.Lemnitser və H.Zinsmeysterin “Korpuslinguistik: Eine Einführung” adlı əsərində dil korpusu geniş şəkildə nəzərdən keçirilmiş və onun linqvistik araşdırmalar üçün əhəmiyyəti vurğulanmışdır [5]. Bu tədqiqatda müəlliflər dil korpusunun müəyyən bir dildən və ya dil növündən alınmış yazılı və şifahi materialların sisteməlik şəkildə toplanmış və strukturlaşdırılmış məlumat bazası olduğunu vurğulayırlar. Korpus dilçiliyində bu bazanın keyfiyyətli və təmsilçilik qabiliyyətinin yüksək olması mühüm rol oynayır. Əsərdə müxtəlif korpus növləri: sinxron, diaxron, monolinqvistik, çoxdilli, ümumi dil korpusları nəzərdən keçirilir. Müxtəlif tədqiqat məqsədləri üçün korpusların seçilməsi və istifadə üsulları müzakirə olunur.

Korpusun formalaşdırılması prosesində metoddakı diqqətlik, dil məlumatlarının toplanması, təmizlənməsi və annotasiyasının (etiketlənməsi) əhəmiyyəti xüsusi qeyd olunur. L.Lemnitser və H.Zinsmeyster korpusda istifadə olunan meta məlumatların və annotasiya sistemlərinin standartlaşdırılmış olması zərurətini vurğulayırlar [5, s.67]. Əsərdə dilin müxtəlif aspektlərinin (söz ehtiyatı, qrammatika, semantika, sintaksis və s.) öyrənilməsində korpusdan istifadə metodları təhlil olunur. Korpus materialları vasitəsilə statistik təhlil, tezlik təhlili, kollokasiya və ya kontekstual tədqiqat kimi yanaşmalar təsvir edilir, korpusların yalnız nəzəri linqvistik tədqiqatlarda deyil, həmçinin tərcümə, maşın tərcüməsi, nitq tanıma texnologiyaları, lüğətçilik və dil tədrisi kimi praktiki sahələrdə də mühüm rol oynadığı vurğulanır.

Bu əsər korpusların dil araşdırmaları üçün necə strukturlaşdırılmalı və təhlil edilməli olduğu barədə nəzəri və metodoloji baxımdan fundamental biliklər təqdim edir. Həmçinin, korpusun istifadəsindəki texniki və nəzəri çətinliklər də müzakirə edilir.

C.Sinklerin “Corpora, Concordance, Collocation” adlı əsərində dil korpusuna tərif geniş nəzəri və praktik çərçivədə təqdim edilir. C.Sinkler korpusu müəyyən dil və ya dillə bağlı real istifadə hallarını təmsil edən, yazılı və ya şifahi mətnlərdən ibarət strukturlaşdırılmış məlumatlar toplusu kimi tərif edir [6]. Bu toplusu müəyyən bir dil fenomenini öyrənmək üçün elmi tədqiqatlarda əsas vasitədir. Müəllif vurğulayır ki, *“korpus süni şəkildə yaradılmış deyil, real dünyada insanların dildən necə istifadə etdiyini dəqiq şəkildə əks etdirən materiallardan ibarət olmalıdır”* [6, s.54]. Yəni, dil haqqında nəzəriyyələr yalnız real dil istifadəsi əsasında qurula bilər. C.Sinkler bir korpusun müəyyən bir dil və ya dil növünün əsas xüsusiyyətlərini təmsil etmək qabiliyyətinə malik olmasının vacibliyini qeyd edir. Təmsilçilik korpusun elmi tədqiqatlar üçün etibarlılığını təmin edir. Bu əsərdə korpusların elektron formatda olması və onların avtomatlaşdırılmış vasitələrlə təhlilinin əhəmiyyəti xüsusi vurğulanır. Bu yanaşma korpusdakı məlumatların geniş həcmdə tezlik, kontekst və kollokasiya baxımından təhlil edilməsinə imkan yaradır. C.Sinkler korpusu yalnız material toplusu kimi deyil, *“dilə öyrənilməsi və müxtəlif aspektlərinin tədqiqi üçün sisteməlik şəkildə hazırlanmış bir vasitə”* [6, s.20] kimi nəzərdən keçirir. Korpusun hazırlanmasında tədqiqat məqsədinin dəqiq müəyyənləşdirilməsi əsasdır. Onun yanaşmasında korpus yalnız məlumat

saxlama vasitəsi deyil, dilin təbiəti, xüsusilə kollokasiya (söz birləşmələrinin istifadəsi) və kontekst anlayışlarını başa düşmək üçün bir vasitədir.

C.Sinklerin bu əsərində korpus əsaslı dil tədqiqatlarının nəzəri əsaslarını qurmaqla yanaşı, korpusların dilçilik üçün metodoloji əhəmiyyəti də dərinlən təsvir edilir. *“Korpus yalnız dil məlumatlarının saxlanıldığı bir resurs deyil, həm də dil haqqında yeni biliklərin əldə edilməsi üçün bir modeldir”* [6, s.100].

M.Stabsın *“Words and Phrases: Corpora Studies of Lexical Semantics”* əsərində korpuslara tərif, əsasən leksik semantika və kontekstual tədqiqatlara əsaslanaraq təqdim edilir. M.Stabs korpusu real dil istifadəsini əks etdirən mətnlərin geniş və sisteməlik şəkildə toplanmış bazası kimi tərif edir. O vurğulayır ki, *“korpus dilin necə işlədiyini öyrənmək üçün etibarlı və empirik əsaslar təmin edir”* [7, s.35]. Korpusların dilin semantik və pragmatik xüsusiyyətlərini, xüsusilə sözlərin və ifadələrin mənalarnın kontekstdə necə formalaşdığını öyrənmək üçün əsas vasitə olduğunu qeyd edir. M.Stabs tez-tez istifadə olunan söz birləşmələri və tezlik modellərinin öyrənilməsi üçün korpusun əhəmiyyətini vurğulayaraq korpusları leksik-semantik araşdırmalarda mühüm bir alət kimi təqdim edir. Xüsusilə, o, kollokasiyaların (sözlərin bir-biri ilə sıx əlaqədə işlənməsi) və tipik dil nümunələrinin öyrənilməsinin dil mənasının anlamaq üçün vacib olduğunu vurğulayır. Korpusun müəyyən bir dil və ya üslubu dəqiq təmsil etməsi üçün onun balanslı və müxtəlif dillərdən materialları əhatə etməsinin vacibliyi qeyd olunur. M.Stabsın yanaşmasında, düzgün təmsilçilik və balans, dilin ümumi qanunauyğunluqlarını öyrənmək üçün əsasdır. M.Stabs korpusları yalnız leksik tədqiqatlar üçün deyil, həm də diskurs təhlili üçün bir vasitə kimi dəyərləndirir. O, qeyd edir ki, *“korpuslar söz və ifadələrin sosial və mədəni kontekstdə necə istifadə edildiyini göstərmək üçün geniş imkanlar yaradır”* [7, s.48]. M.Stabs korpusların rəqəmsal formatda olmasını və onların təhlilində kompüter texnologiyalarından istifadə edilməsinin vacibliyini vurğulayır. Bu, məlumatların geniş miqyasda və daha dəqiq təhlilinə imkan verir.

M.Stabsın əsərində korpuslar yalnız məlumat toplusu kimi deyil, dilin dərin mənalarnın, kollokasiyaların, semantik nüansların və diskursiv nitqi təcrübənin araşdırılması üçün əvəzolunmaz bir elmi vasitə kimi təqdim edilir. Bu, korpusların dilçiliyin müxtəlif sahələrində, xüsusilə leksik semantika və kontekstual dil tədqiqatlarında əsas rolunu aydın şəkildə göstərir.

Beləliklə, bu əsərlər haqqında təqdim etdiyimiz məlumatlara istinadən, qeyd etmək olar ki, L.Lemnitser və H.Zinsmeysterin əsərində dil korpusu, yazılı və şifahi materiallardan ibarət, sisteməlik şəkildə toplanmış və strukturlaşdırılmış məlumat bazası kimi təsvir olunur. Bu yanaşma, dil korpusunun keyfiyyətli, təmsilçilik qabiliyyətinə malik və müxtəlif tədqiqat məqsədlərinə uyğun olaraq formalaşdırılmasının əhəmiyyətini vurğulayır.

C.Sinklerin tədqiqatlarında isə korpusun real dil istifadəsini təmsil edən və elmi tədqiqatlar üçün əsas vasitə olduğu vurğulanır. C.Sinkler korpusların yalnız məlumat toplusu deyil, dilin təbiətini, kollokasiya və kontekst anlayışlarını öyrənmək üçün mühüm rol oynadığını qeyd edir. O, korpusların elektron formatda hazırlanmasını və onların avtomatlaşdırılmış vasitələrlə təhlilini dilçiliyin yeni biliklər əldə etməsi baxımından vacib hesab edir [6, s.53].

Ümumilikdə bu üç yanaşmada korpusların linqvistik tədqiqatlarda əhəmiyyətli bir vasitə olduğu, onların strukturlaşdırılması, təmsilçilik qabiliyyəti və rəqəmsal texnologiyalarla dəstəklənməsi vurğulanır. Müəlliflər korpusların həm nəzəri, həm də tətbiqi sahələrdə dilin öyrənilməsi və təhlili üçün əvəzsiz bir resurs olduğunu göstərirlər.

Rus dilçiliyində V.P.Zaxarov korpusu *“böyük həcmli, elektron formatda təqdim olunan, vahidləşdirilmiş, strukturlaşdırılmış, işarələnmiş və filoloji baxımdan kompetent dil məlumatları toplusu, konkret dilçilik tapşırıqlarının həlli üçün nəzərdə tutulmuş baza”* kimi təyin edir. [8, s.11]. V.P.Zaxarovun korpus anlayışına verdiyi bu tərif dil korpusunun elmi və praktiki əhəmiyyətini, həmçinin onun tədqiqat məqsədləri üçün necə istifadə oluna biləcəyini dərinlən izah edir. İlk növbədə, korpusun böyük həcmli və elektron formatda təqdim olunması diqqət çəkir. Elektron formatın seçilməsi, məlumatların saxlanması və emalının daha asan və sürətli həyata keçirilməsinə imkan verir. Bununla yanaşı, geniş həcmli məlumatlar dilin müxtəlif aspektlərini öyrənməyə şərait yaradır və fərqli tədqiqat məqsədləri üçün zəngin material təmin edir.

V.P.Zaxarov korpusun “vahidləşdirilmiş və strukturlaşdırılmış olması” [8, s.67] ifadəsi ilə onun elmi tədqiqat üçün daha əlçatan və sistematik bir vasitə kimi təşkil olunduğunu göstərir. Vahidləşdirmə korpusun daxilindəki məlumatların standartlara uyğun şəkildə hazırlanmasını təmin edir və istifadəçilər üçün vahid yanaşmanı mümkün edir. Strukturlaşdırılmış quruluş isə məlumatların məntiqi qaydaya uyğun təşkil edilməsini ifadə edir. Bu, dil vahidlərinin müəyyən qaydalara əsasən təsnifləşdirilməsi, mətnlərin sintaktik və ya semantik bölünməsi kimi xüsusiyyətləri əhatə edir.

Korpusun işarələnmiş olması, onun daha dərinlən tədqiq edilməsinə imkan yaradır. İşarələmə prosesi zamanı korpus daxilindəki məlumatlara qrammatik, sintaktik və ya semantik xarakteristikalar əlavə olunur. Bu, tədqiqatçıların müxtəlif dil hadisələrini daha dəqiq və məqsədyönlü şəkildə təhlil etməsinə şərait yaradır. Məsələn, sözlərin cümlə daxilindəki funksiyası, qrammatik quruluşları və kollokasiyaları haqqında məlumat işarələmə vasitəsilə təqdim olunur ki, bu da dilin daha dəqiq anlaşılmasına kömək edir.

Filoloji baxımdan müvafiq və konkret dilçilik tapşırıqlarını həll etmək üçün nəzərdə tutulmuş korpus tədqiqatçıların xüsusi problemləri və sualları araşdırması üçün unikal bir vasitədir. Bu yanaşma korpusun sıradan bir məlumat bazası deyil, dilçilik tədqiqatlarını irəlilətmək üçün xüsusi olaraq hazırlanmış bir resurs olduğunu vurğulayır. V.P.Zaxarovun yanaşması göstərir ki, dil korpusu elmi araşdırmalar üçün əsaslı bir alət olmaqla yanaşı, həm də dilin sosial, tarixi və mədəni aspektlərini araşdırmaq üçün zəngin bir mənbədir. Bu, dilin müxtəlif aspektlərini öyrənmək və yeni biliklər əldə etmək üçün çoxşaxəli imkanlar yaradır.

L.Lemnitser isə vurğulayır ki, “korpusun tərkib hissəsi olan mətnlər məlumatlar və metadatalar ilə yanaşı, onları leksik annotasiyalarla da təşkil edir” [5, s.13]. Deməli, korpusun tərkib hissəsi olan mətnlər yalnız əsas dil məlumatlarını və metadataları deyil, həm də leksik annotasiyaları ehtiva etməklə daha strukturlaşdırılmış və istifadəyə yararlı hala gətirilir. Bu yanaşma korpus daxilindəki mətnlərin yalnız məzmun olaraq deyil, dilin leksik və semantik xüsusiyyətlərini əks etdirən əlavə məlumatlarla zənginləşdirildiyini göstərir. Leksik annotasiyalar vasitəsilə sözlərin mənası, funksiyası və kontekstdəki rolu daha dəqiq müəyyənləşdirilir, bu da dilin dərin tədqiqatını və analizini asanlaşdırır.

“Korpusun qurulmasında “redaktə edilməmiş” mətnlərdən istifadə və dili təbii şəkildə təqdim etmək əsas prinsiplərdən biridir, çünki bu yanaşma dilin danışmada necə təzahür etdiyini əks etdirməyə imkan verir. Bundan əlavə, dilin müxtəlif variantlarının olması və “nümunələr” və “təlimatlar”nın olmaması da vacibdir” [6, s.19].

Azərbaycan dilçiliyində dil korpusu və korpus dilçiliyi məsələləri ilə təməldən məşğul olan dilçilərdən M.Mahmudovun adını xüsusilə qeyd etmək lazımdır. O qeyd edir ki, “korpus dilçiliyi hər hansı konkret dili təmsil edən mətnlərin qabaqcadan müəyyənləşdirilmiş qaydalar üzrə sistemli şəkildə elektron formatda toplanmasına istiqamətlənmiş məlumat axtarışı sistemlərinin yaradılması və onların tədqiqat prosesində istifadəsi məsələlərini öyrənir” [2, s.335].

Bu yanaşma dil korpuslarının yalnız mətn toplusu deyil, həm də empirik tədqiqatlar üçün nəzəri və metodoloji baza yaratdığını göstərir. M.Mahmudovun fikirləri korpus dilçiliyinin əsas funksiyasının yalnız dil materiallarının toplanması və strukturlaşdırılması ilə məhdudlaşmadığını, eyni zamanda, süni intellekt və avtomatik dil emalı sahələrində tətbiqinin vacibliyini vurğulayır. O, dil korpuslarının tədqiqatçıların dilin müxtəlif aspektlərini, xüsusilə morfolojiya, sintaksis, semantika və pragmatika sahələrində sistemlik şəkildə araşdırmasına imkan yaratdığını qeyd edir. Bu kontekstdə Azərbaycan dilçiliyində korpus əsaslı tədqiqatların genişlənməsi və texnoloji inkişaflarla inteqrasiyası dilin funksional və struktur xüsusiyyətlərinin daha dərinlən öyrənilməsinə töhfə verə bilər.

“Əslində, korpus dilçiliyi hər hansı bir dili təmsil edən biləcək mətnlərin, eləcə də şifahi danışmaların audio yazılarının sistemli şəkildə toplanması və onların əsasında axtarış sistemlərinin yaradılması deməkdir. Həmin axtarış sistemləri elə qurulmalıdır ki, biz dil haqqında istənilən məlumatı əldə edə bilək” [3, s.49]. Rus dilçisi N.N.Kazanski “Korpus dilçiliyi və filologiya” adlı tezisində korpus dilçiliyinin filoloji araşdırmalardakı rolunu və onun klassik humanitar elmlərlə qarşılıqlı əlaqəsini araşdıran fundamental tədqiqatlardan biridir. N.N.Kazanski qeyd edir ki, “korpus

dilçiliyi təkcə struktur dilçiliyin və kompüter əsaslı analizlərin bir sahəsi kimi deyil, eyni zamanda filoloji tədqiqatların əsas metodoloji aləti kimi də nəzərdən keçirilməlidir” [9, s.6]. O, xüsusilə mətnlərin dil quruluşu, tarixi inkişafı və üslubi xüsusiyyətlərinin öyrənilməsində korpus metodlarının əhəmiyyətini vurğulayır.

Bu yanaşma M.Mahmudovun fikirləri ilə səsleşir, çünki hər iki tədqiqatçı korpus dilçiliyinin yalnız texnoloji və statistik bir metod kimi deyil, həm də dilin məna strukturlarının, semantik və pragmatik aspektlərinin dərinədən tədqiqinə imkan yaradan vasitə kimi dəyərləndirilməsinin vacibliyini bildirir. N.N.Kazanski, həmçinin, filoloji tədqiqatlarda klassik mənbələrin və ədəbi mətnlərin korpus formatında təqdim edilməsinin vacibliyini qeyd edir və bunu mətnlərin qrammatik, leksik və sintaktik xüsusiyyətlərinin sistemli təhlili üçün mühüm resurs hesab edir [9, s.7].

Azərbaycanda korpus dilçiliyi ilə bağlı genişmiqyaslı tədqiqat aparan R.Məmmədova korpus dilçiliyinin tətbiq sahələrini və onun məhdudiyyətlərini təhlil edərək, xüsusilə leksik semantika və mətn şərhləri ilə bağlı bəzi problemləri vurğulayır. R.Məmmədovanın fikrincə, *“korpus dilçiliyi qrammatikanın bütün sahələrində eyni effektivlikdə nəticə vermir*” [3, s.51]. O, leksik semantikanın təsvirində korpus metodlarının yetərincə effektiv olmadığını bildirir. Bu, müəyyən qədər doğrudur, çünki leksik semantika yalnız tezlik və kollokasiya əsaslı statistik metodlarla tam izah edilə bilən bir sahə deyil. Sözlərin məna çaraları, kontekstdən asılı olaraq dəyişən məna münasibətləri, metaforik və idiomatik istifadələr bəzən korpus vasitəsilə tam dəqiq şəkildə təhlil edilə bilmir.

Bu məqam M.Stabbsın tədqiqatları ilə müəyyən qədər uyğunluq təşkil edir. M.Stabbs da qeyd edir ki, *“korpus əsaslı yanaşma müəyyən semantik sahələrdə güclü analiz imkanı versə də, bəzi abstrakt semantik anlayışları yalnız statistik göstəricilər əsasında dərk etmək çətinidir*” [7, s.21]. Xüsusilə sinonimlik, antonimlik və polisemiya kimi semantik əlaqələrin tam şəkildə modelləşdirilməsi üçün əlavə kontekstual məlumatlara və dərin filoloji şərhlərə ehtiyac duyulur.

R.Məmmədovanın vurğuladığı digər mühüm məqam isə korpus dilçiliyi ilə ənənəvi filologiya arasındakı fərqlər və onların bir-birini tamamlaması zərurətidir. Müəllif bildirir ki, *“ənənəvi filologiya mətnlərin kontekstual şərhini daha yaxşı təqdim edə bilir*” [3, s.51], halbuki korpus dilçiliyində bəzi mətnlərin çoxvariantlı interpretasiyası çətinlik yaradır.

Bu, N.N.Kazanskiyə irəli sürdüyü fikirlərlə səsleşir. N.N.Kazanski qeyd edirdi ki, *“korpus dilçiliyi filoloji tədqiqatları tamamilə əvəz etməməli, əksinə, onları tamamlamalıdır*” [9, s.6]. R.Məmmədovanın nümunə kimi verdiyi Puşkinin “Yevgeni Onegin” əsərinin bir sıra dilçilər tərəfindən müxtəlif şəkildə şərh edilməsi faktı bunu təsdiq edir. Bu şərhlər mətnin fərqli kontekstlərdə fərqli ideoloji və üslubi yanaşmalarla dəyərləndirilməsini göstərir ki, bunu korpus dilçiliyinin hazırkı vəziyyətində tam modelləşdirmək çətinidir.

Bu kontekstdə korpus metodologiyasının ənənəvi filoloji metodlarla birlikdə işləməsi daha məqsədəuyğundur. Məsələn, korpus təhlili mətnin tezlik əsaslı strukturunu, sintaktik və leksik xüsusiyyətlərini üzə çıxara bildiyi kimi, filoloji şərh də mətnin tarixi, kontekstual və üslubi xüsusiyyətlərini daha dərin şəkildə təhlil edə bilər.

R.Məmmədovanın araşdırmasında toxunulan başqa bir mühüm məsələ *“şifahi xalq ədəbiyyatının tədqiqində korpus dilçiliyinin yetərsizliyi*” [2, s.51] ilə bağlıdır. Şifahi ənənənin spesifik xüsusiyyətləri, yəni müxtəlif variantların mövcudluğu, dəyişkənlik və sabit bir “son variant”ın olmaması korpus dilçiliyinin standartlaşdırılmış məlumat emalına əsaslanan metodları üçün çətinlik yaradır.

Burada müəllifin irəli sürdüyü *“filoloji korpusun müəyyən hallarda son variantı formalaşdırmağa imkan verməli olması*” [3, s.51] fikri mübahisəli görünə bilər. Çünki müasir korpus yanaşmalarında müxtəlif variantların qorunması və paralel təqdim olunması daha məqsədəuyğun hesab edilir. Bu baxımdan, CLARIN və ELRA kimi beynəlxalq korpus layihələrində şifahi ənənəyə aid nümunələr də çoxvariantlı şəkildə saxlanılır, çünki tək bir “doğru variant” müəyyən etmək elmi baxımdan problem yarada bilər. Bu məsələyə bir qədər aydınlıq gətirək:

CLARIN (Common Language Resources and Technology Infrastructure) Avropa miqyaslı bir təşəbbüsdür və əsas məqsədi dil texnologiyaları və resurslarını tədqiqatçılar üçün əlçatan

etməkdir. CLARIN şəbəkəsi müxtəlif dillərdə olan mətn və nitq korpuslarını, annotasiyalı dil resurslarını, lüğət bazalarını və digər dil texnologiyalarını tədqiqatçılar üçün açıq şəkildə təqdim edir. Bu layihə xüsusilə akademik tədqiqatlar və dilin rəqəmsal emalı sahəsində istifadə olunur. CLARIN çərçivəsində fərqli dillərdə korpuslar yaradılır və tədqiqatçılar üçün çoxvariantlı mətnlər təqdim edilir.

ELRA (European Language Resources Association) dil resurslarının toplanması, saxlanması və yayılması ilə məşğul olan Avropa təşkilatıdır. ELRA, əsasən avtomatik tərcümə, nitqin tanınması, süni intellekt və digər dil texnologiyaları sahəsində istifadə olunan dil resurslarını təmin edir. Buraya müxtəlif dillərdə mətn və nitq korpusları, terminoloji lüğətlər, annotasiyalı məlumat bazaları və digər resurslar daxildir. ELRA-nın məqsədi dil texnologiyaları sahəsində innovasiyaları təşviq etmək və tədqiqatçılar üçün keyfiyyətli dil resursları təqdim etməkdir.

CLARIN və ELRA kimi beynəlxalq layihələr dil korpuslarının müxtəlif variantlarını saxlayır və paralel təqdim edir. Bu yanaşma, xüsusilə şifahi xalq ədəbiyyatı, tarixi mətnlər və dialektoloji materiallar üçün vacibdir, çünki bu tip mənbələrin yalnız bir “doğru variant” ilə məhdudlaşdırılması elmi obyektivliyə uyğun olmaya bilər. Bu baxımdan, beynəlxalq təcrübə müxtəlif versiyaların saxlanmasını və tədqiqatçılar üçün açıq olmasını daha məqsədəuyğun hesab edir.

M.Mahmudov və digər müəlliflər vurğulayırlar ki, *“milli korpusun bütün altkorpusları eyni vaxtda deyil, tədricən formalaşır və texnoloji təminat bazası yarandıqca fəaliyyətə keçir”* [1, s.68]. Bu yanaşma korpus dilçiliyində geniş qəbul edilən “modulyar inkişaf modeli” ilə uzlaşır. Beynəlxalq təcrübədə də (məsələn, İngilis Dilinin British National Corpora və Rusiyanın Milli Korpusu) ilkin mərhələdə əsas funksional altkorpuslar yaradılır və tədricən inkişaf etdirilir.

“Mövcud altkorpuslar arasında qarşılıqlı əlaqə, istifadəçi tələbatı və yeni məlumat bazalarının formalaşması digər altkorpusların yaradılmasına təkan verə bilər” [1, s.68]. Bu, uyğunlaşdırılmış və çevik yanaşma kimi qiymətləndirilə bilər, çünki bu metod yeni ehtiyaclar yarandıqca, korpusun tərkibi və funksionallığı genişləndirmək imkanı verir. Bu yanaşma təcrübədə bir çox qabaqcıl korpus layihələrində istifadə olunur. Məsələn, Fransanın “Frantext”, Almaniyanın “DWDS” və Çinin “Sinica Corpora” layihələri də lüğət və terminologiya bazalarını elektron sistemə daxil edərək onları tədqiqatçılar və istifadəçilər üçün açıq resursa çeviriblər.

Müəlliflər bildirirlər ki, *“leksikoqrafiya altkorpusu istifadəçiyə tam linqvistik məlumat təqdim etməlidir”* [1, s.68]. Bu, olduqca mütərəqqi bir yanaşmadır və korpus dilçiliyinin müasir multidissiplinar tədqiqatlarda istifadəsini genişləndirə bilər.

Hazırda AMEA-nın İ.Nəsimi adına Dilçilik İnstitutunun Kompüter dilçiliyi şöbəsində Azərbaycan dilinin milli korpusunun tərtibi istiqamətində bir sıra işlər görülməkdədir. Bunların təməlinə konkordansların tərtib edilərək dil korpusu üçün əsaslı bazanın təşkil edilməsi dayanır. Şöbə əməkdaşları ayrı-ayrı illərdə M.Müşfiq, H.Cavid kimi görkəmli ədiblərin əsərlərinin, eləcə də Azərbaycan xalqının möhtəşəm dil abidəsi hesab edilən “Kitabi-Dədə Qorqud dastanları”nın konkordanslarını tərtib edərək hazırlamışlar. Bu məqamla bağlı olaraq vurğulamaq da yerinə düşər ki, “Kitabi-Dədə Qorqud dastanları”nın konkordansı tərtib edilən zaman bu dil abidəsinin orijinal variantının latın qrafikalı versiyasından istifadə edilmiş və konkordans bazasında sağır nun [ŋ] işarəsi ilə qeyd edilmişdir.

Bu istiqamətdə aparılan təcrübə işləri və araşdırmalar göstərir ki, milli korpusların inkişafı təkcə strukturlaşdırılmış mətn bazasının yaradılması ilə məhdudlaşmır, həm də bu bazaların dinamik şəkildə genişlənməsi və multidissiplinar tədqiqatlarda istifadəsini təmin edən texnologiyaların tətbiqi ilə sıx bağlıdır. Bu metodoloji yanaşma həm beynəlxalq təcrübəyə əsaslanmaqla, həm də Azərbaycan dilinin rəqəmsal inkişafına uyğun bir çərçivəyə uyğunlaşdırılmaqla milli dil korpusunun yaradılması ilə bağlı olan işlərin həyata keçirilməsini şərtləndirir. Müasir korpus layihələrində (Frantext, DWDS, Sinica Corpora və s.) olduğu kimi, Azərbaycan dilinin milli korpusunun da altkorpuslar üzərində qurulması, hər bir altkorpusun spesifik bir linqvistik bazanı əhatə etməsi və eyni zamanda, onların integrasiya olunmuş şəkildə fəaliyyət göstərməsi gələcək tədqiqatlar üçün əlverişli şərait yaradır. Bu kontekstdə leksikoqrafiya

altkorpusunun mərkəzi rol oynadığı və onun digər altkorpuslarla qarşılıqlı əlaqədə olması zərurəti xüsusi qeyd olunmalıdır.

Xüsusilə, Azərbaycan ədəbiyyatının klassik nümunələrinin (M.Müşfiq, H.Cavid, “Kitabi-Dədə Qorqud dastanları”) konkordanslarının tərtib edilməsi milli dil korpusunun formalaşdırılmasında mühüm addımlardan biridir. Azsaylı işçi (tədqiqatçı və proqramçı) kontingenti ilə kiçik bir şöbənin bu işi öz öhdəsinə götürməsi təqdirəlayiqdir. Konkordansların tərtib edilməsi mətnlərin strukturlaşdırılması və xülasələndirilməsi ilə yanaşı, bu mətnlərin dil baxımından sistemli şəkildə araşdırılmasını və tarixi, semantik və sintaktik aspektlər üzrə təhlilini asanlaşdırır.

Bu işlərin davamlı inkişaf etdirilməsi üçün rəqəmsal lüğət bazalarının, annotasiya sistemlərinin və avtomatik təhlil vasitələrinin yaradılması böyük əhəmiyyət kəsb edir. Milli dil korpusunun tədricən genişlənməsi və fərqli altkorpusların inteqrasiyası Azərbaycan dilinin semantik və sintaktik xüsusiyyətlərinin daha dərinlən tədqiq edilməsinə, eyni zamanda, dilin praktik istifadəsi üçün daha səmərəli texnologiyaların inkişafına şərait yarada bilər.

Bütün bu proseslər göstərir ki, Azərbaycan dilinin milli korpusunun yaradılması yalnız dilçilik araşdırmaları üçün deyil, həm də tədris, tərcümə, süni intellekt və maşın tərcüməsi kimi sahələrdə də geniş tətbiq imkanlarına malik olacaqdır. Beləliklə, altkorpusların formalaşması və onların qarşılıqlı əlaqəsi təkcə elmi deyil, həm də texnoloji və pedaqoji perspektivlər baxımından mühüm əhəmiyyət kəsb edir.

Nəticə: Məqalə çərçivəsində təqdim edilən bu kiçik araşdırma göstərir ki, dil korpusu və korpus dilçiliyi müasir dilçiliyin əsas istiqamətlərindən biri kimi yalnız strukturlaşdırılmış mətn bazalarının yaradılması ilə məhdudlaşmır, həm də dilin müxtəlif aspektlərinin empirik metodlarla araşdırılması üçün yeni imkanlar yaradır. Müxtəlif tədqiqatçılar tərəfindən irəli sürülən nəzəri yanaşmalar (Lemnitzer və Zinsmeister, Sinkler, Stabbs, Zaxarov və digərləri) göstərir ki, dil korpusları yalnız lingvistik tədqiqatlar üçün deyil, həmçinin leksikoqrafiya, diskurs analiz, maşın tərcüməsi, nitq texnologiyaları və dilin normativ monitorinqi kimi sahələrdə də mühüm rol oynayır.

Bu tədqiqatın əsas nəticələrindən biri ondan ibarətdir ki, müasir korpus dilçiliyi yalnız statistik və avtomatlaşdırılmış metodlarla məhdudlaşmamalıdır, əksinə, ənənəvi filoloji metodlarla birlikdə işləyərək mətnlərin dərin semantik və kontekstual təhlilini mümkün etməlidir. Xüsusilə, korpusların yalnız kəmiyyət göstəricilərinə əsaslanan analizlərlə məhdudlaşmaması, eyni zamanda, mətnlərin filoloji və kontekstual şərhini dəstəkləməsi mühüm bir vəzifə kimi qalır.

Azərbaycan dilçiliyində milli korpusların yaradılması istiqamətində həyata keçirilən layihələr (AMEA-nın İ.Nəsimi adına Dilçilik İnstitutunun Kompüter dilçiliyi şöbəsində aparılan tədqiqatlar və M.Mahmudovun irəli sürdüyü model) bu sahənin inkişafı üçün perspektivli addımlardır. Xüsusilə leksikoqrafik altkorpusun yaradılması və milli dilin elektron lüğətlər bazasının formalaşdırılması, həm akademik, həm də praktiki tətbiq baxımından böyük əhəmiyyət daşıyır.

Beynəlxalq təcrübə göstərir ki, modulyar və adaptiv yanaşma, yəni altkorpusların tədricən formalaşdırılması və inkişaf etdirilməsi milli korpusların daha effektiv və istifadəyə yararlı olmasını təmin edir. Azərbaycan dilinin milli korpusu üçün bu modelin tətbiqi istifadəçilərin ehtiyaclarına uyğun olaraq yeni altkorpusların yaradılmasına və dil texnologiyalarının inkişafına mühüm töhfə verə bilər.

Nəticə etibarilə, dil korpusu və korpus dilçiliyi yalnız akademik tədqiqat sahəsi kimi deyil, həm də pedaqoji, texnoloji və tətbiqi lingvistik istiqamətində inkişaf edən bir platforma kimi dəyərləndirilməlidir. Gələcəkdə bu sahədə daha geniş miqyaslı tədqiqatların aparılması, Azərbaycan dilinin korpus əsaslı tədqiqatlarının genişləndirilməsi və rəqəmsal dil texnologiyalarına inteqrasiyası mühüm istiqamətlərdən biri olaraq qalır.

Ədəbiyyat siyahısı

1. Mahmudov, M və b. İnteqrasiya olunmuş elektron lüğətlər sistemi Azərbaycan dilinin milli korpusunun yaradılmasında bir mərhələ kimi / M.Mahmudov, İ.Tahirov, K.Ayda-zadə, S.Talıbov // *Türkologiya*, 2019, №1, — s. 66-80.
2. Mahmudov, M. Süni intellektin lingvistik problemləri / M.Mahmudov. — Bakı: Elm və təhsil, — 2024. — 376 s.
3. Məmmədova, R. Milli dil korpuslarının dilin inkişafında rolu // Bakı Qızlar Universitetinin Elmi əsərləri, — 2020. № 4(44), cild 11, — s. 49-53.
4. Məmmədzadə, S. Korpus dilçiliyi kompüter dilçiliyinin əsası kimi // *Filologiya məsələləri*, — 2015. № 1, — s. 103-107
5. Lemnitzer, L. Korpuslinguistik: Eine Einführung / L. Lemnitzer, H. Zinsmeister. — Tübingen, 2006. — 213 S.
6. Sinclair, J. Corpora, Concordance, Collocation / J. Sinclair. — Oxford: Oxford University Press, — 1991. — 179 p.
7. Stubbs, M. Words and Phrases: Corpora Studies of Lexical Semantics / M. Stubbs. — Oxford: Oxford University Press, — 2001. — 288 p.
8. Захаров, В.П. Корпусная лингвистика / В.П. Захаров, С.Ю. Богданова. — Санкт-Петербург, 2020. — 234 с.
9. Казанский, Н.Н. Корпусная лингвистика и филология // “Müasir dilçiliyin aktual problemləri” Beynəlxalq Elmi Konfransın materialları, — 24-25 noyabr, Sumqayıt, 2016-cı il, s. 6-7.

Габибова Кёнул

Основные подходы в корпусной лингвистике и перспективы этой области в азербайджанском языкознании

Аннотация

Данная статья рассматривает теоретические основы понятий языкового корпуса и корпусной лингвистики в современной лингвистике, а также их применение в научных исследованиях. Языковой корпус представляется как структурированная база данных письменных и устных материалов на различных языках, а корпусная лингвистика объясняется как научная область, изучающая различные аспекты языка с использованием эмпирических методов. В статье проводится сравнительный анализ подходов к языковым корпусам таких исследователей, как Л. Лемнитцер и Х. Цинсмайстер, Дж. Синклер, М. Стаббс, В. П. Захаров, Н. Н. Казанский, М. Махмудов, а также рассматриваются их теоретико-методологические основы. Исследование подчеркивает роль языковых корпусов в таких областях, как лексикография, семантика, синтаксис, переводческий процесс, машинный перевод и технологии распознавания речи. Помимо этого, обсуждаются создание национального корпуса азербайджанского языка, проводимые в этой области исследования и технологические возможности. В статье также охватываются исследования, проводимые Отделом компьютерной лингвистики Института языкознания имени Насими Национальной академии наук Азербайджана, шаги, предпринятые для развития национального корпуса, а также создание электронных словарных баз. В работе подчеркивается, что корпусная лингвистика является не только одним из направлений компьютерной лингвистики, но и независимым эмпирическим методом исследований. В результате исследуются связь корпусной лингвистики с филологическими исследованиями, ее вклад в глубинное изучение семантических и синтаксических характеристик языка, а также перспективы дальнейшего развития. Данное исследование показывает, что языковые корпуса важны не только для научных лингвистических исследований, но также имеют педагогическое и технологическое применение.

Ключевые слова: языковой корпус, корпусная лингвистика, база данных, электронный словарь, речевые технологии.

Habibova Konul

Fundamental Approaches in Corpora Linguistics and Perspectives of This Field in Azerbaijani Linguistics

Abstract

This article explores the theoretical foundations of the concepts of language corpora and corpora linguistics in modern linguistics, as well as their application in research fields. A language corpora is presented as a structured database of written and spoken materials in various languages, while corpora linguistics is defined as a scientific discipline that empirically studies different aspects of language. The paper compares the approaches to language corpora proposed by L.Lemnitzer & H.Zinsmeister, J.Sinclair, M.Stubbs, V.P.Zakharov, N.N.Kazanski, and M.Mahmudov, analyzing their theoretical and methodological foundations. The study highlights the role of language corpora in fields such as lexicography, semantics, syntax, translation processes, machine translation, and speech technologies. Additionally, it discusses the creation of the Azerbaijani National Corpora, research conducted in this area, and technological opportunities. The paper also covers studies carried out by the Department of Computational Linguistics at the Nasimi Institute of Linguistics of the Azerbaijan National Academy of Sciences, the steps taken towards developing the national corpora, and the creation of electronic dictionary databases. The article emphasizes that corpora linguistics is not only a subfield of computational linguistics but also an independent empirical research methodology. As a result, it examines the relationship between corpora linguistics and philological studies, its contribution to the in-depth study of semantic and syntactic characteristics of language, and future development perspectives. This research demonstrates that language corpora are crucial not only for scientific linguistic research but also have significant pedagogical and technological applications.

Keywords: *language corpora, corpora linguistics, database, electronic dictionary, speech technologies.*

Çapa tövsiyə edən: *AMEA Nəsimi adına Dilçilik İnstitutu*

Rəyçilər: *professor Məsud Mahmudov
fil.e. dok. Nərmin Əliyeva*