

# Milli dil korpusunda Azərbaycan dilinin tədqiqi məsələləri

Rəna Mahmudova

AMEA Nəsimi adına Dilçilik İnstitutu, Bakı, Azərbaycan

*rena.memmedova.1991@inbox.ru*

*Xülasə* — Hazırda bir çox dünya dillərini əhatə edən çoxlu miqdarda elektron formatda olan korpuslar mövcuddur. Korpus müəyyən prinsiplərə müvafiq olaraq toplanmış, işarələnmiş, linqvistik təhlilin müəyyən pillələri üçün xülasələşdirilmiş mətnlər toplusudur. Korpus dilçiliyinin inkişafı, eləcə də korpusun qurulması müasir dilçiliyin aktual problemlərindən biridir. Hal-hazırda korpus dilçiliyi onun yaranmasına səbəb olan texnologiya və metodları uğurla inkişaf etdirir. Azərbaycan dilinin tədqiq edilməsi prosesini asanlaşdırmaq və yeni texnoloji vasitələrin dilə tətbiq edilməsi məsələləri dövrün aktual məsələlərindəndir. Bu istiqamətdə çox sayda araşdırmalar aparılmış və bəzi uğurlara imza atılmışdır. Bunlardan biri də Azərbaycan dilinin elektron lüğətlər korpusudur. Korpus tam sistem halına salınmamış, lüğətlərin əksəriyyəti yerləşdirilməmişdir. Hal-hazırda korpusun təkmilləşdirilməsi üzərində işlər aparılır. Gələcəkdə dilçiliyin bütün sahələrinin Azərbaycan dilinin milli korpusunda bir vahid şəkildə yerləşdirilməsi nəzərdə tutulur.

*Açar sözlər* — *milli korpus; dil korpusları; konkordans; altkorpusla; lüğətlər bloku*

## I. GİRİŞ

Korpus dilçiliyinin inkişafı, eləcə də korpusun qurulması müasir dilçiliyin aktual

problemlərindən biridir. Hal-hazırda linqvistik tədqiqatların əksəriyyətində korpusdan istifadə aparıcı rol oynayır. Baxmayaraq ki, 20-ci əsrin birinci yarısında korpusları yalnız əl ilə yaratmaq mümkün idi. Çox vaxt, səy və xərc tələb edirdi. Buna görə də, çox sayda insanın bununla maraqlandığı halda mətnlər korpusuna müraciət minimal idi. Lakin korpus dilçiliyinin inkişafı sayəsində korpusun istifadəsi hər yerdə mümkün olmuşdur. Korpusun yaradılması ilə materialı strukturlaşdırmaq və sinxronlaşdırmaq çox asanlaşdı və xərclər kəskin şəkildə azaldı.

Korpus dilçiliyinin inkişafının əhəmiyyəti şübhəsizdir. Onda humanitar və texniki elmlərin bir çox müsbət xüsusiyyətləri birləşdirilir. Lakin nisbətən yaxın dövrlərdə yarandığı üçün tam inkişaf etməmişdir və ayrıca elm sahəsi kimi statusu hələ də müəyyən edilməmişdir. Beləliklə, bir çox alimlər korpus dilçiliyini ənənəvi dilçiliyin alt sahəsi hesab edirlər. Onların fikrincə, korpus linqvistikası dil materiallarının toplanması və işlənməsinin təkmilləşdirilmiş üsulu, yeni informasiya resursudur.

## II. MİLLİ DİL KORPUSLARI

Hal-hazırda korpus dilçiliyi onun yaranmasına səbəb olan texnologiya və metodları uğurla inkişaf etdirir. Məsələnin

texniki tərəfini də unutmaz. Kompüter və informasiya texnologiyalarının inkişafında kəskin sıçrayış baş verdi. Bu imkanlardan dilçilikdə uğurla istifadə olunmağa başlandı. Qlobal internetin inkişafı və populyarlaşması sayəsində müxtəlif ölkələrdən olan çoxlu sayda istifadəçi korpusun məlumatlarından istifadə etməyə başladı. Bundan əlavə, korpusların tamlığı və genişləndirilməsi problemi aradan qaldırıldı. Korpus dilçiliyi sahəsində tədqiqatların əksəriyyəti ingilisdilli mətnlərin üzərində aparılmışdır. Bunun iki əsas səbəbi var: birincisi, Şimali Amerikada və Qərbi Avropada kompüter texnologiyalarının aktiv inkişafı, ikincisi, XX əsrin 60-80-ci illərində Britaniya dilçiliyinin inkişafı üçün əlverişli şəraitin olmasıdır [2].

Korpus dilçiliyinin yaradılmasının iki başlıca səbəbi vardır: Birinci, bu, peşəkar dilçilər üçün əvəzsiz vasitədir. Onlar bu və ya digər şəkildə dilin faktlarının tədqiqi ilə məşğul olurlar. Bu faktların dəqiq və sistemli şəkildə toplanması üçün qısa zamanda dəqiq nəticə verən bir vasitəyə ehtiyac vardı. İkincisi, milli dil korpusunun yaradılması proqramçılar üçün də çox faydalıdır.

Təbii ki, linqvistik korpus həyatını sözlə bağlanmış insanlar, məsələn, yazıçılar, qəzet və jurnalların redaktorları üçün də faydalıdır. Belə ki, onlar üçün hər hansı bir çətin məqamları korpusa istinad edərək yoxlaya bilərlər.

Hazırda bir çox dünya dillərini əhatə edən çoxlu miqdarda elektron formatda olan korpuslar mövcuddur. Məsələn, rus dilində linqvistik araşdırmalarda istifadə olunan

çoxlu sayda korpusların olduğu göstərilir. Bu baxımdan müəyyən mövzu ilə əlaqədar mövcud olan istənilən mətn kolleksiyası korpus hesab olunur. Korpus müəyyən prinsiplərə müvafiq olaraq toplanmış, işarələnmiş, linqvistik təhlilin müəyyən pillələri üçün xülasələndirilmiş mətnlər toplusudur. Korpus haqqında bu izahat milli dillərin elektron bazası çərçivəsində toplanmış mətnlərə də aid edilə bilər.

Korpus dilçiliyi mətn korpuslarının yaradılması və istifadə olunması ilə məşğul olan dilçilik sahəsidir.

### III. AZƏRBAYCAN DİLİNİN MİLLİ KORPUSU

Azərbaycan dilinin riyazi-statistik metodlarla və yeni texnoloji vasitələrlə öyrənilməsi məsələlərinə ötən əsrin ikinci yarısından başlanılmışdır. Həmin vaxtlarda dünyanın bir çox aparıcı mərkəzlərində dilin riyazi-statistik metodlarla elektron hesablayıcı maşınlarının köməyi ilə öyrənilməsi sahəsində maraqlı tədqiqatlar aparılırdı. Keçmiş SSRİ-nin bir çox elmi mərkəzlərində, o cümlədən, Moskva, Leningrad (indiki Sankt-Peterburq), Kiyev, Minsk və s. şəhərlərində də bu qəbildən olan tədqiqatlara böyük maraq var idi.

İstənilən konkret bir dilin lüğət tərkibini əhatə edən mətnlərin toplanması məqsədini daşıyan axtarış sistemi Milli korpusdur. Bu axtarış sistemində dilin bütün səviyyələri üzrə məlumatlar yerləşdirilməlidir və istifadəçinin axtarışına cavab olaraq müraciət formaları korpusun qayda-qanunlarına uyğun nizamlanmalıdır. Axtarış zamanı istifadəçi vikipediya məlumatlarından fərqli olaraq informasiyanı

konkretləşdirilə və onlara lazım olan informasiyanı əldə etmək imkanı yaradılmalıdır.

Azərbaycan dilinin milli korpusu sahəsində hazırlıq işləri türk dillərinin milli korpusu çərçivəsində aparılırdı. Bu işləri Azərbaycan dilinin milli korpusunun yaradılmasının başlanğıc mərhələsi hesab etmək olar.

#### IV. DİL KORPUSLARINDA LÜĞƏTLƏR BLOKU

Lüğətçilik elminin tarixi dərin keçmişə dayanır. Ünsiyyət vasitəsi olan dilin ifadə vasitələri sözlərdir ki, onlar nə qədər zəngin olarsa, ifadə ediləcək fikir o qədər aydın və dəqiq olar. Dillərarası, mədəniyyətlərarası əlaqənin yaradılmasında lüğətçilik elminin mühüm rolu vardır.

Lüğətlərin yaradılması heç də asan əmələ gələn proses deyil. Belə ki, ənənəvi lüğətçilikdə lüğət tərtib etmək üçün çox sayda işçi qüvvəsinə və uzun zamana ehtiyac var idi. Toplanılan sözlər kartoçkalara yazılır, daha sonra onların sistemləşdirilməsi, qruplaşdırılması prosesləri həyata keçirilirdi.

Kompüter lüğətçiliyində isə bu proseslər qısa zaman ərzində öz həllini tapır. Lüğətlərin tərtibi kompüter lüğətçiliyi, hesablama lüğətçiliyi, maşın lüğətçiliyi, avtomatik leksikoqrafiya terminləri adı altında aparılır. Bu terminlərin hamısını kompüter dilçiliyinin predmeti hesab edən “Kompüter dilçiliyinin əsasları” kitabının müəllifi Y.N.Marçuk bunların əslində eyni mənanı verən sinonim terminlər olduğunu bildirir [3]. Y.N.Marçukun bu fikrinə əsaslanaraq deyə bilərik ki, bu sadalanan

terminlərin tədqiqat obyektini kompüter vasitəsi ilə tərtib olunan lüğətlərdir.

Korpus dilçiliyinin özəyini lüğətlər təşkil edir. Belə ki, korpus ona yerləşdirilən mətnlərdəki sözlərin mənasını, hansı nitq hissəsinə mənsub olmasını və s. öyrənmək üçün mətndəki sözləri avtomatik işarələməklə həmin sözə aid bütün məlumatları əldə etmək imkanı verir. Bunun ən effektiv üsulu dilçilik korpuslarına konkordansların tətbiqidir.

Konkordans bu və ya digər konkret dildə işlənən söz və ifadələrin işləndikləri bütün kontekstlərlə və aid olduqları mənbəyə istinadla verilən siyahıdır. Fərz edək ki, elektron məkanda Azərbaycan yazıçılarının əsərlərindən ibarət mətnlər toplusu yerləşdirilmişdir. Biz hər hansı bir söz və ifadəni seçib axtarışa veririk. Axtarış sistemi həmin sözün işləndiyi bütün kontekstləri tapır və dəqiq yerləri göstərməklə siyahı şəklində istifadəçiyə təqdim edir. Təbiidir ki, hər hansı konkret söz və ifadə müxtəlif kontekstlərdə fərqli mənalarda (məna çalarlarında) işlənmiş olur. Beləliklə, istifadəçi (tədqiqatçı) söz və ifadənin mətnlər massivində əks olunmuş bütün məna müxtəlifliklərini (məna çalarlarını) qısa zaman kəsiyində dəqiq və əhatəli şəkildə əldə etmiş olur.

Leksikoqrafiya altkorpusunda bütün lüğətlərin elektron variantları arasında əlaqənin yaradılması da mühüm şərtlərdəndir. Leksikoqrafiya altkorpusu istifadəçinin hər hansı bir söz haqqında tam linqvistik məlumat əldə edə bilməsini təmin etməlidir. Tam linqvistik məlumata sözün izahlı lüğətdəki mənaları, tələffüzü, düzgün yazılışı, transkripsiyası, etimologiyası, dialektoloji lüğətlərdəki məna variantları,

frazeoloji vahidlərdə işlənmə məqamları, başqa dillərə tərcüməsi və s. aid edilə bilər.

Azərbaycan dilinin milli korpusunda lüğətlər altkorpusunun yaradılması ilə bağlı görülmüş ilk iş “Azərbaycan dilinin lüğət tərkibinin inkişafı, dilin normalarına əməl olunmasının ictimai monitorinqi və dilin verilənlər korpusunun hazırlanması üzrə inteqrasiya olunmuş elektron sistemin və lüğətin yaradılması” layihəsini göstərmək olar [1]. Bu layihədən sonra “Azərbaycan dilinin elektron lüğətlər korpusu” [4] yaradılmışdır və üzərində təkmilləşdirmə işləri aparılır. Korpusda bir neçə lüğət yerləşdirilmişdir. Gələcəkdə Azərbaycan dilinin lüğət tərkibini əhatə edən bütün lüğət növlərinin korpusda yerləşdirilməsi nəzərdə tutulur.

## NƏTİCƏ

Qloballaşan dünyada hər hansı bir milli dil öz varlığını sübut etmək, mövcud olmaq, nüfuz qazanmaq istəyirsə, internet məkanında təmsil olunmalıdır. İndi dünya bir bilik mənbəyi, örnək kimi daha çox internetdən istifadə edir, ondan faydalanır. Azərbaycan dili də öz varlığını, mövcudluğunu, nüfuzunu internetdə geniş və hərəkətli təmsil olunmaqla sübut etməlidir. Azərbaycan dilinin milli korpusunda Azərbaycan dilinin bütün mövcud lüğətləri (izahlı, orfoqrafiya, orfoepiya, dialektoloji, tarixi, etimoloji, terminoloji, statistik, ikidilli, çoxdilli və s.), Azərbaycan dilinin məşin tərcüməsi sistemi,

elmi, bədii ədəbiyyatı, publisistikası, qrammatikası, təhlil alqoritmləri və s. yerləşdirilməlidir. Qeyd etmək lazımdır ki, məşin tərcüməsi və məlumat axtarışı sistemlərində geniş istifadə olunan mətnin formal təhlili və sintezi, həmçinin səs tanınması (səsin mətnə, mətnin səsə çevrilməsi) sistemləri də Azərbaycan dilinin milli korpusunun tərkib hissələri, komponentləri hesab olunur. Azərbaycan dilinin milli korpusu elə qurulmalıdır ki, axtarış sistemində “Azərbaycan” yazmaqla onun dili, ədəbiyyatı, tarixi, incəsənəti, mentaliteti, coğrafiyası, təbiəti və s. haqqında hərtərəfli və dolğun məlumat əldə oluna bilsin. Azərbaycan dilinin milli korpusunun yaradılması bu baxımdan ən vacib, əhəmiyyətli addım olardı. Hazırda respublikamızda Azərbaycan dilinin milli korpusunun yaradılması sahəsində “Dilmanc” layihəsi çərçivəsində bir sıra işlər görülmüşdür.

## İSTİNADLAR

- [1] M.Ə.Mahmudov, Türk dillərinin milli korpusu / M.Ə.Mahmudov. – Bakı: Elm və təhsil, – 2018, – 392 s.
- [2] A.A.Васнецова, Корпусная лингвистика: исторический и лингводидактический аспекты. ISSN 2414–4452. PHILOLOGY. 2021. № 3 (33), с. 48–54
- [3] Ю.Н.Марчук, Оновы компьютерной лингвистики // Ю.Н.Марчук. – М.: МГОУ, – 2002, –234 с.
- [4] <http://korpus.azerbaycandili.az/>