

Milli dil korpuslarının dilin inkişafında rolu

Rəna Hüseyn qızı Məmmədova

fəlsəfə doktoru proqramı üzrə dissertant,

AMEA Nəsimi adına Dilçilik İnstitutu

E-mail: rena.memmedova.1991@inbox.ru

Rəyçilər: filol.ü.e.d. K.A. Vəliyeva,
filol.ü.f.d., dos. K.Ə. Həbibova

Açar sözlər: korpusun yaradılması, dünya təcrübəsi, milli korpus, dil korpusu, struktur özəllikləri

Ключевые слова: создание корпуса, мировой опыт, национальный корпус, языковой корпус, структурные особенности

Key words: corpus creation, world experience, national corpus, language corpus, structural features

Yaşadağımız əsr informasiya əsri və cəmiyyətimiz informasiya cəmiyyəti adlanır. Buna səbəb informasiya texnologiyalarının inkişafı və elmin bütün sahələrində əsaslı rola malik olmasındadır. İnformasiya cəmiyyəti anlayışına texnoloji nöqtəyi-nəzərdən yanaşsaq, görürük ki, burada kompüter, internet əsas əsas texniki vasitə olaraq informasiyanın toplanması, emalı, ötürülməsi və digər mühüm funksiyaları həyata keçirir (1, s. 5).

Humanitar elmlər sahəsinə aid olan dilçilik elminin inkişafında da informasiya texnologiyalarının müstəsna əhəmiyyəti vardır. Xüsusilə dövrümüzdə inkişaf etmiş dünya ölkələrində də dil korpuslarının yaradılması dilin internet məkanında dünyaya tanınması, eləcə də tədqiqatçılara lazım olan informasiyaların asanlıqla əldə edilməsi baxımından ən optimal vasitə hesab olunur. Dilçiliyin ən yeni sahəsi - kompüter dilçiliyinin müasir istiqamətlərindən biri olan korpus dilçiliyinin əhatə dairəsi, tədqiqat obyektı, ayrı-ayrı bölmələri haqqında mütəxəssislər arasında vahid fikir hələ formalaşmamışdır. Bəzi tədqiqatçılar şifahi nitq korpuslarına daha çox əhəmiyyət verir və şifahi nitqin tədqiqi məsələlərini korpus dilçiliyinin prioritet istiqaməti kimi irəli sürürlər. Tədqiqatçıların digər qrupu isə belə hesab edir ki, korpus dilçiliyinin tədqiqat obyektı yazılı mətnlər olmalıdır. Əslində korpus dilçiliyi hər hansı bir dili təmsil edə biləcək mətnlərin, eləcə də şifahi danışığın 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.

Əgər biz dil haqqında tam və hərtərəfli dolğun linqvistik məlumat əldə etmək istəyiriksə təbiidir ki, yalnız yazılı mətnlərlə və ya yalnız şifahi danışığı əhatə edən audio yazılarla kifayətlənə bilmərik. Başqa sözlə, dilçilik təkcə qrammatika demək deyil, burada danışığ səsələrinin tədqiqi ilə məşğul olan fonetika da kənarda durmamalıdır. Bu baxımdan yanaşdıqda bir daha əmin oluruq ki, dil korpusları həm yazılı mətnləri, həm də şifahi danışığ nümunələrini əhatə etməlidir. Yalnız belə olduqda biz dil haqqında hərtərəfli linqvistik məlumat əldə edə bilərik. Ola bilsin ki, dil korpuslarında bu istiqamət eyni səviyyədə təmsil olunmasın. Yəni elə dil korpusları ola bilər ki, orada yazılı mətnlər audio yazılara nisbətən üstünlük təşkil etsin. Şifahi danışığ üslubunun tədqiqi ilə daha uğurlu məşğul olan və maraqlı nəticələr əldə edə bilən korpuslar da ola bilər. Bütün bunlar belə bir nəticəyə gəlməyə əsas verir ki, bu və ya digər korpusda dilin müəyyən xüsusiyyətləri daha qabarıq əks edilə bilər.

Başqırd dilində audiokorpusun materialları həmin dilin şərqi dialekti əsasında formalaşmışdır. Həmin materialların fonetik transkripsiyası bütövlükdə bütün fonetik korpusa, o cümlədən digər türk dillərinin fonetik korpusuna tətbiq oluna bilər. Bu transkripsiya əsasında konkret türk dilləri üçün, o cümlədən, Azərbaycan dili üçün də fonetik korpus yaradıla bilər və başqırd dilinin ənənələrindən faydalanmaq olar. Həmin transkripsiya müasir başqırd dilinin orfoqrafiyasına qismən uyğunlaşdırılmışdır. Digər dilçilərin də həmin bazaya girişi mümkündür və başqırd dilinin ərazi dialektlərinin öyrənilməsində həmin baza örnək rolunu oynaya bilər.

Məqsəddən asılı olaraq dünyada müxtəlif tipli korpuslar mövcuddur. Məsələn, bir müəllifin korpusu, İncilin korpusu, rus dilinin korpusu və s. Milli korpus isə mütəlif janrları, üslubları, regional və sosial vəziyyəti özündə əks etdirən mətnlərin külliyyatıdır (5, s. 7).

Dünya dilçiliyində, o cümlədən rus dilçiliyində korpus elektron daşıyıcıda saxlanan müxtəlif dil hadisələri və aspektləri ilə bağlı linqvistik tədqiqatlara material verən, müəyyən nizamla düülmüş təbii dil mətnləri çoxluğudur. Korpusu təşkil edən mətnlər və onlara müraciət müəyyən qayda ilə aparılır [4, s. 448]. Korpuslar əhatə etdikləri material baxımından bir-birindən fərqlənir. Elə korpuslar var ki, orada daha çox (və yaxud, tamamilə) yazılı mətnlərdən istifadə olunur. Korpusların digər növündə şifahi audio, bunlardan əlavə, qarışıq tipli korpus materialları üstünlük təşkil edir. Orada həm şifahi audiomaterial, həm də yazılı material olur, amma onlar nisbət baxımından fərqlilik təşkil edə bilər.

Dünya miqyasında əksər korpuslar yazılı mətnlərə əsaslanır. Son zamanlar şifahi nitq korpuslarına da maraq artmaqdadır. Məsələn, ingilis, çin, alman, yapon dilləri üçün şifahi nitq korpusları hazırlanmışdır. Rusiyada Saratov dialektlərinin korpusu, Arxanges dialektlərinin korpusu və rus dilinin milli korpusunda dialekt altkorpusu və s. mövcuddur. Məsələn, Rusiyada ural və altay dillərinin audiyozalılış korpusları hazırlanmışdır. Başqırdıstanda başqırd dilinin dialektləri əsasında fonetik korpus hazırlanır. Həmin korpusda informantları müxtəlif sosial qruplara, cinsinə və yaşa görə toplayırlar. Başqırd dilinin fonetik korpusu 3 mərhələdə yerinə yetirilir. 1-ci mərhələdə başqırd dilinin şərqi dialektləri tədqiq olunur. Bu mərhələdə müəyyən mövzular üzrə informantların danışmalarının diktafona yazılışı həyata keçirilir. Sonrakı mərhələdə həmin audiyozalılış frazalarına və sintaqmlara ayrılır. Transkripsiya olunur, annotasiyalaşdırılır və s. Dialekt materialı ədəbi dillə müqayisə olunur və rus dilinə tərcümə edilir. Transkripsiya yarım orfoqrafik yazılış qaydasında aparılır. Belə qayda nitq korpuslarının transkripsiyasında geniş istifadə olunur. Bu transkripsiya fonematik yazılışa yaxındır və maksimal olaraq müasir başqırd dilinin orfoqrafiyasına yaxınlaşdırılmışdır. Dialekt xüsusiyyətlərinin ifadəsi üçün vahid qrafik sistem işlənib hazırlanmışdır.

Başqırd dilində ədəbi dildən fərqli səslər xüsusi işarələrlə verilir və onlara nümunələr gətirilir. Qeyd etmək lazımdır ki, bu göstərilən işarələr yalnız başqırd dilinin transkripsiyasında istifadə olunur, çünki həmin spesifik səslər başqırd dilinə məxsus səslər hesab olunur.

Başqırd dilinin korpusunda tək-cə sözlər deyil, eyni zamanda nitqin paralingvistik elementləri də təsvir olunmuşdur. Məsələn, gülüş, öskürək, inilti, ağlaşma, ah çəkmə. Hətta nitq prosesində pauzalara da xüsusi diqqət yetirilmişdir.

Pauzalar insan haqqında heç də sözdən az şey demir. Çünki danışmaq prosesində bəzən vaxtın 40-50 faizini pauza təşkil edir. Məsələn, susmaq razılıq əlamətidir (Atalar sözü).

Pauzalar bəzən o vaxt yaranır ki, danışan öz fikrini ifadə etməkdə çətinlik çəkir. Nitq prosesində informantlar bəzən qısa, bəzən isə uzun pauzalardan istifadə edir.

Dünyada minlərlə dil korpusları mövcuddur və onların sayı ilbəil artır. Korpuslara belə böyük maraq onunla əlaqədardır ki, hazırda dilçilikdə elmi tədqiqatların istiqaməti və mahiyyəti yeni məzmun kəsb etməkdədir. Korpuslar tək-cə dilçilikdə deyil, ədəbiyyatşünaslıqda da

uğurla istifadə oluna bilər. Məsələn, yazıçı üslubunun müəyyənləşdirilməsi, dilin semantik bazasının müəyyənləşdirilməsi və s. məsələlərdə korpuslar əvəzsiz mənbə rolunu oynayır.

Sözün bütün mənalarını müəyyənləşdirmək üçün hökmən konkordanslardan istifadə etmək lazım gəlir. Konkordanslar müxtəlif təyinatlı lüğətlərin, o cümlədən izahlı lüğətlərin tərtibi üçün etibarlı baza rolunu oynaya bilər (3, s.151). Əgər izahlı lüğətdə sözün 20-30 mənası göstərilmişdirsə, konkordanslar əlavə olaraq onlarla yeni məna çalarlarını əldə etməyə imkan verir. Məlum olduğu kimi izahlı lüğətlərdə bu və ya digər sözün məna çalarlarını göstərmək üçün bədii ədəbiyyat nümunələrinə müraciət olunur. Klassik ədəbiyyatdan və ya müasir ədəbiyyatdan bu və ya digər sözün mənasını açıb göstərən nümunələr verilir. Konkordanslar isə əksinə materialdan sözün işləndiyi cümlələr çoxluğundan yeni mənaların müəyyənləşdirilməsi imkan verir.

Lüğət – konkordanslar korpus dilçiliyinin mühüm hissəsi kimi dilçilik və ədəbiyyatşünaslıq sahəsində aparılan araşdırmalarda müasir və effektiv tədqiqat vasitəsi kimi əhəmiyyətlidir. Azərbaycan dilinin milli korpusunda ilk növbədə lüğət altkorpusunun funksiyasını, fəaliyyətini qurmaq əhəmiyyətli hesab olunmuş və korpusun yaradılmasında ilk növbədə lüğətlərdən başlanılmışdır. Lüğət-konkordansın yaradılması üçün ölkəmizdə nəşr olunmuş və olunan bütün lüğətlərin elektron variantları yaradılmalıdır.

Korpus mətnlərinin konkordansının tərtibi ilə bağlı xüsusi proqramlar mövcuddur. Həmin proqramlar “konkordanser” adlanır. Bu proqramlar ixtiyari korpus mətnindən bu və ya digər dil vahidinin tezliyini müəyyənləşdirmək, həmin vahidlərin işləndiyi kontekstlərin siyahısını almağa imkan verir. Eyni zamanda həmin proqramlarla kontekstləri açar sözlərə görə çeşidləmək mümkündür. Söz-formaları yaxın mətnlərdə həmin proqramların köməyi ilə növlərə ayırmaq olar.

Akademik N.N.Kazanski “Korpus dilçiliyi və filologiya” məqaləsində korpus dilçiliyinin ənənəvi filologiyada istifadəsi yollarından bəhs edir. Müəllif göstərir ki, korpus dilçiliyi sintaktik konstruksiyaların, qrammatik kateqoriyaların, feili idarənin ayrı-ayrı məsələlərinin izahında istifadə olunduğu kimi ənənəvi filologiyada mətnlərin izahı prosesində də istifadə oluna bilər.

Qeyd etmək lazımdır ki, korpus dilçiliyi qrammatikanın bütün sahələrində eyni dərəcədə uğur gətirmir. Məsələn, leksik semantikanın təsvirində korpus dilçiliyindən səmərəli şəkildə istifadə etmək mümkün olmur. Nəzərə almaq lazımdır ki, leksik semantika ənənəvi təsvirdə də çətin formalaşdırıla bilər, amma ənənəvi filologiyada mətnlərin izahı zamanı korpus dilçiliyində yeni metodlarından istifadə zəruridir. Bu həm də ona görə vacibdir ki, çox vaxt bu və ya digər mətn korpus dilçiliyindən kənar da müxtəlif variantlarda şərh olunur. Məsələn, Puşkinin “Yevgeniy Onegin” əsərinin 3 müxtəlif şərh (N.L.Brodski, V.V.Nabokov və Y.M.Lotman) çap olunmuşdur (2, s. 6-7).

Ayrı-ayrı, kiçik-kiçik əsərlər haqqında da linqvistik izahatlar mövcuddur, amma həmin şərhlər korpus dilçiliyində istifadə olunmur. Mətnə şərh yazılması mürəkkəb vəzifədir və korpus dilçiliyi indiki səviyyədə onun öhdəsindən gələ bilmir. Bu təkcə yazılı mətnlərə aid deyil. Məsələn, şifahi xalq nümunələrinin çoxlu variantları olur ki, onların heç ənənəvi filologiyadakı şərhələrini də məqbul hesab etmək olmaz. Deməli, filoloji korpus o yerdə tərtib olunmalıdır ki, onun tərtibi hər hansı bir şifahi xalq ədəbiyyatı nümunəsinin son variantını almağa imkan versin.

Belə nəticəyə gəlmək olar ki, korpus dilçiliyi təkcə dil tədqiqatlarında deyil, ümumi filoloji mətnlərin tədqiqatında da istifadə oluna və səmərə verə bilər.

Milli dil korpuslarının ümumi məsələləri ilə bağlı daha bir cəhətə diqqət yetirək. Məlum

olduğu kimi dil korpusları müəyyən məqsədlərlə yaradılır. Dil korpuslarında təkcə dilçilik məsələləri deyil, ədəbiyyatşünaslıq, məntiq, üslubiyyat, linqvostatistika, tədris məsələləri və s. istiqamətlər tədqiqat obyekti kimi seçilə bilər. Elə korpuslar var ki, onlar birmənalı olaraq tədris məqsədi ilə yaradılır və orada qurulmuş axtarış sistemləri tədrisə istiqamətlənir. Məlum olduğu kimi ayrı-ayrı yazıçıların və şairlərin dillərinin linqvostatistik və üslubi baxımdan öyrənilməsi məsələləri də korpus dilçiliyinə daxildir. Məsələn, kordansların özlərini də milli dil korpuslarının komponentlərindən biri kimi səciyyələndirmək olar.

Məqalənin aktuallığı. İnformasiya vasitələrinin geniş yayıldığı indiki dövrümüzdə elektron resurslar vasitəsilə elmi tədqiqat aparmaq və məlumat əldə etmək proseslərinin asanlaşdırılması elmin qarşısında qoyulan əsas vəzifələrdəndir. Dilçilik elmi üçün inkişaf etmiş ölkələrdə dil korpuslarının yaradılması əsas vəzifə kimi qarşıya qoyulmuşdur. Bütün türk dilləri üçün məqbul ola biləcək vahid işarələnmə sistemlərinin hazırlanması olduqca aktual məsələdir. Azərbaycan dilinin milli korpusu üzərində də görülməli işlər çoxdur. Bütün bunlar məqalənin dövrün ən önəmli və aktual mövzusunda həsr olunduğunu göstərir.

Məqalənin elmi yeniliyi. Məqalədə ilk dəfə dilin müxtəlif bölmələri üzrə korpus dilçiliyində görülən işlər və onların Azərbaycan dilinin milli korpusuna tətbiq etmə yolları araşdırılmışdır.

Məqalənin praktik əhəmiyyəti və tətbiqi. Dil korpuslarının yaradılması dilin inkişafını izləməkdə, yeniliklərin əlavə edilməsi və köhnələrlə əvəzlənməsi prosesinin asan üsulla yerinə yetirilməsində böyük əhəmiyyətə malikdir. Bu da onu göstərir ki, dildə baş verən dəyişikliklərin tez bir zamanda tətbiq edilməsi üçün hər bir dilin korpusunun yaradılması əsaslı məsələlərdəndir.

Ədəbiyyat

1. Əliquluyev R., Valehov C., Mahmudov R.Ş. İnternet jurnalistikasının formalaşmasının bəzi aspektləri. Ekspres-informasiya. İnformasiya cəmiyyəti seriyası. Bakı: İnformasiya texnologiyaları, 2008.

2. Казанский Н.Н. Корпусная лингвистика и филология / Müasir dilçiliyin aktual problemləri beynəlxalq elmi konfransın materialları. Sumqayıt, 2016.

3. Mammadova R.H. The concordances in the researches of the language. (Alatoo academic studies). Bishkek: Ala-Too, 2019, № 3

4. Məmmədova, R.H. Milli korpusun yaradılmasında dünya təcrübəsi. Akademik A.Axundovun 85 illiyinə həsr olunmuş "Ağamusa Axundov və Azərbaycan filologiyası" mövzusunda beynəlxalq elmi konfrans. Bakı: Elm, 2017.

5. Mirzəliyeva M., Vəliyeva K. Maşın tərcüməsinin nəzəri problemləri. Bakı: Elm və təhsil, 2016.

6. <http://korpus.azerbaycandili.az>

Р.Г. Мамедова**Роль национального языкового
корпуса в развитии языка****Резюме**

В данной статье исследуются причины образования национальных языковых корпусов и их роль в развитии лингвистики. Чтобы поместить в разные области лингвистики в единый корпус, необходимо изучить корпуса мировых языков и прийти к определенному мнению и создать оптимальную структуру. Языковые корпуса могут служить информационным, обучающим и исследовательским целям. Чтобы создание корпуса национального языка дало прекрасные результаты, прежде всего, необходимо определить, для какой цели он был создан.

R.H. Mamadova**The role of the national linguistic corpus in
the development of the language****Summary**

This article examines the reasons for the formation of national language corpora and their role in the development of linguistics. To put in different areas of linguistics in a single corpus, it is necessary to study the corpuses of world languages and come to a certain opinion and create an optimal structure. Language corpora can serve informational, teaching and research purposes. In order for the creation of a corpus of the national language to give excellent results, first of all, it is necessary to determine for what purpose it was created.

Redaksiyaya daxil olub: 06.11.2020