

Korpus dilçiliyinin təşkilində leksikoqrafiyanın rolu və vəzifələri

Nərmin Əliyeva¹, Rasim Heydarov²

^{1,2}AMEA Nəsimi adına Dilçilik İnstitutu, Bakı, Azərbaycan

¹rajabova.narmin@gmail.com, ²dr.rasimheydarov@mail.ru

Xülasə — Məqalədə leksikoqrafiyanın korpus dilçiliyində təşkilindəki rolundan bəhs edilir. Əsas məqsəd həmin sahənin korpus dilçiliyinin formalaşmasındakı rolunu göstərməkdən ibarətdir. Bunun üçün təsviri metoddan istifadə olunmuşdur.

Açar sözlər — *tətbiqi dilçilik; lüğətçilik; kompüter dilçiliyi; leksikoqrafiya; tezaurus*

I. GİRİŞ

Müxtəlif formalarda olub tədrislə yanaşı digər məqsədlərə də xidmət edən lüğətlər Oksford İngilis dili lüğəti (OED) kimi tarixi lüğətlər və Oksford Biokimya və Molekulyar Biologiya Lüğəti (ODBMB) kimi yüksək ensiklopedik əsərləri əhatə edir. Qeyd edilən birinci lüğət klassik dil lüğətidir, ikinci lüğət isə daha çox ensiklopedik xarakter daşıyır və terminologiyaya həsr olunub. Bu lüğətlər arasında müəyyən oxşar cəhətlər də vardır. Belə ki, hər iki lüğətdə olan sözlər əlifba sırası ilə düzülmüşdür.

OALD (Oxford Advanced Learner Dictionary) kimi lüğətlərin adı bizə çox şey açıqlamasa da biz hamımız lüğətin necə göründüyünü “bildiyimizdən” bu, adətən problem hesab edilmir. Amma təəssüflə

qeyd olunmalıdır ki, həm ümumi, həm də metaleksikoqrafik baxımdan bəzi gizli problemlər də ola bilər. Adi istifadəçilər adətən lüğətlərin necə tərtib edildiyindən və hansı auditoriya üçün nəzərdə tutulduğundan xəbərsiz olurlar. Onlar sözlərin mənasına “sabit” münasibət bəsləməyə meyillidirlər ki, bu da onların dilin təkamülünü unutmasına və ya görməməsinə səbəb olur. Bu amillər tələbələrimizi evdə tapdıqları, artıq dərəcədə köhnəlmiş və onları xüsusi kontekstlərdə ümumi dil mənalardan istifadə etməyə məcbur edən, çox vaxt isə gülməli nəticələr verən lüğətlərə etibar etməyə vadar edə bilər. Bu problemlər çox vaxt, hətta xarici dil müəllimləri arasında da geniş yayılmış lüğətlərin növlərini və onlardan istifadəni bilməmək səbəbindən meydana gəlir.

II. ƏSAS HİSSƏ

Son illər korpus dilçiliyi ilə bağlı yazılan əsərlər artsa da bu sahədə tədqiqatların əksəriyyəti hələ də ingilisdilli mətnlərin materialı üzərində aparılır. Bunun iki əsas səbəbi var: birincisi, Şimali Amerikada və Qərbi Avropada kompüter texnologiyasının aktiv inkişafı, ikincisi, XX əsrin 60-80-ci illərində Britaniya

dilçiliyinin inkişafı üçün əlverişli ab-havanın olması.

Linqvistik korpuslar içərisində Braun korpusu ilk korpus hesab olunur. O, 1963-cü ildə Braun Universitetində V.Fransis və Q.Kuçera tərəfindən hazırlanmışdır. İlk korpusun həcmi 1 milyon söz idi ki, buraya 1961-ci ildə ABŞ-da müxtəlif janrlarda nəşr olunmuş mətnlərdən hər biri 2000 sözdən ibarət 500 fraqment: məşhur yazıçı və şairlərin bədii mətnləri, qəzet və jurnallardan məqalələr, yazılı işgüzar nitq nümunələri və dini mövzularda mətnlər daxil edilmişdi. Braun korpusunun yaradılmasının bir neçə səbəbi var idi. İlk növbədə, bu, müxtəlif janrlara aid olan ingiliscə mətnlərin sistemli şəkildə öyrənilməsinin təmin edilməsi, ikincisi, bu məlumatları müqayisə etmək üçün kifayət qədər material təqdim etmək, üçüncüsü, yeni elmi intizamın yaranmasına bir çox alimlərin marağını cəlb etmək idi. Bu, tətbiqi dilçilikdə çoxlu mübahisələrə və müzakirələrə səbəb olan bir sıçrayışa səbəb oldu. Korpusun özünün tərtibinə gəlincə, o, bir tərəfdən statistik məlumatlar əsasında, digər tərəfdən isə alimlərin təcrübəsi və intuisiyası əsasında qurulmuşdu. Obyektivliyə nail olmaq üçün sadə və şəffaf mətnlərdən istifadə etmək lazım idi.

İngilis dili ilə bağlı hazırda ən məşhur korpus isə Britaniya Milli Korpusu hesab edilir (British National Corpus, BNC). 1991-1994-cü illərdə Oksford Universiteti və Lankaster Universitetinin tədqiqatçıları tərəfindən yaradılmış bu korpusun həcmi 100 milyon sözdür və sələflərindən xeyli böyükdür. Britaniya Milli Korpusu 90%

yazılı və 10% şifahi mətnlərdən ibarətdir. Mətnlər XX əsrin sonlarına aiddir və müxtəlif janrları əhatə edir. Burada qəzet məqalələri, qeyri-bədii ədəbiyyat, işgüzar yazışma nümunələri, dini mövzularda mətnlər, qeyri-rəsmi çıxışların transkripsiya edilmiş yazıları, radio verilişləri, hökumət çıxışları və s. daxildir.

Sonralar bir çox ölkələr Britaniya Milli Korpusundan nümunə götürərək öz milli korpuslarını yaratmağa qərar verdilər. Rusiyada bu proses 2003-cü ildən 2010-cu ilə qədər çəkdi. Rusiyalı mütəxəssislər yeddi il ərzində vahid mətn bazası yaratmaq üzərində işləyirdilər. “Rus Dilinin Milli Korpusu” Assosiasiyası “Yandex” şirkətinə əməkdaşlıq təklif etdi və onların texniki dəstəyi ilə Rus Dilinin Milli Korpusunun hazırkı versiyasının yerləşdiyi internet saytı hazırlandı. Rus Dilinin Milli Korpusu XVIII əsrin ortalarından XX əsrin əvvəllərinə qədər olan dövr üçün 163 milyondan çox söz istifadəsini ehtiva edir. Korpusda müxtəlif dövrlərə aid mətnlərin olması səbəbindən burada həm müasir, həm də tarixi mətnlərdən nümunələrə rast gəlmək mümkündür. Eyni zamanda mətnlər ədəbi, danışıq, rəsmi və elmi üslublara aiddir. Rus Dilinin Milli Korpusunda həm xalq dili, həm də jarqon və dialektlərin çoxlu nümunələri var. Rus Dili Milli Korpusunun rəsmi saytında “korpus” termininin mənası, istifadə qaydaları və onun yaradılmasının məqsədi aydın şəkildə müəyyən edilmiş bir istinad məqaləsi də var.

Hazırda xarici dilin tədrisində korpus yanaşmasının ən populyar sahələrindən biri

korpusun köməyi ilə öyrənmə və ya verilənlərə əsaslanan öyrənmədir (DDL). Bu təlimin mahiyyəti ondan ibarətdir ki, öyrənənlər birbaşa korpusdan, yəni “xam” dil məlumatlarından istifadə edirlər. Bu yanaşma tələbələrin öyrənmə prosesi, müşahidə-təxmin-eksperiment modelindən istifadə etməyə təşviq edildikdə, dili daha effektiv öyrənmə bilər. Proses mütləq kompüter terminalı ilə məhdudlaşmır. Korpus axtarırlarının (konkordansların) çap formasında nəticələri asanlıqla paylama materiallarına, tədris vəsaitlərinə və s.yə inkorporlaşa bilər.

Təkdilli və ikidilli lüğətlər arasındakı fərq göz qabağındadır və sonuncunun istifadəsi dil tədrisində yaxşı məlum olsa da, biz burada onlardan bəhs etməyəcəyik. Bununla belə, qeyd etmək lazımdır ki, ikidilli lüğət də özünün birdilli qarşılığı kimi, kompüterin yaranması ilə inqilabi dəyişikliklərə məruz qalmışdır. İkidilli lüğətlərin modernləşdirilməsində pioner olan Sue Atkins bu məsələni geniş təhlil etmişdir [1]. Burada həm ümumi, həm də xüsusi birdilli lüğətlər barədə danışacağıq.

Ümumi dil lüğətlərindən fərqli olaraq, xüsusi istifadəyə həsr olunmuş əsərlər adətən terminoloji cəhətdən əsaslandırılır və iki formada təqdim olunur: əsasən tərcüməçiyə ünvanlanan, bu tədqiqatın predmeti olmayan çoxdilli terminlər və mövzu mütəxəssisləri üçün nəzərdə tutulmuş təkdilli ensiklopedik lüğətlər tələbələr üçün ixtisaslaşdırılmış lüğətin müzakirəsi üçün mühüm əsas yaradır. Lakin bu cür lüğətlərin yanaşması tələbə lüğətinin yanaşmasından tamamilə fərqlidir.

Əksər ümumi dil əsərlərində istifadə olunan klassik təriflər sistemi sözlün məqsədyönlü təhlilinə yönəlib, yəni anlayışın fərqli xüsusiyyətlərini verir, ensiklopediya məqalələri isə bundan çox-çox kənara çıxır, ətraflı genişləndirici məlumatlar verir. ODBMB kimi lüğətlərin məqsədi terminləri müəyyən etmək və izah etməkdir, onlar yalnız ana dilində danışmaq qabiliyyəti olan istifadəçilərə ünvanlanır və istifadəni göstərməyə və ya izah etməyə cəhd etmir. Bu cür lüğətlər öz təbiətinə görə göstəriş xarakteri daşıyır və bu sözlərin real həyat vəziyyətlərində istifadəsini öyrətmək məqsədi daşımır.

Müasir kompüterlərin yaranması lüğətlərin yaradılmasında inqilabi dəyişikliklərə gətirib çıxardı. Belə ki, kompüterlər məlumatların daha səmərəli işlənməsi üçün verilənlər bazası ilə yanaşı, həm də məlumatların yeni formalarına və xüsusən də tələbə lüğətlərinə bu məlumatların təqdim edilməsinin yeni üsullarına çıxış imkanı verdi. Kodlaşdırma lüğətlərində inqilabi dəyişikliklər bu tədqiqatın əsas diqqətini təşkil edir, lakin inqilabın mahiyyətini nəzərdən keçirməzdən əvvəl, tədris və lüğətçilik arasındakı əlaqəni nəzərdən keçirməliyik.

Tələbələr üçün nəzərdə tutulan lüğətlərdə sözlərin mənalarının ifadəsi kontekstlidir, sözlər yalnız kontekstdə mənalar qazanır və buna görə də lüğət bu kontekstləri faktiki istifadəni göstərməyə çalışmalıdır. Bu tip lüğətlərə hansı sözlərin daxil edilməsi problem olaraq qalır. Bu səbəbdən də söz siyahıları subyektivdir. İlk dəfə Hornby yerli olmayan tələbələr üçün

faydasız hesab etdiyi sözləri lüğətə daxil etməklə, OALD üçün əsas kimi COD istifadə etdi. Bu seçim onun intuisiyasına və dil müəllimi, tədqiqatçı və leksikoqraf kimi böyük təcrübəsinə əsaslanırdı. Təcrübəli leksikoqrafın biliyini qiymətləndirsək də istənilən halda həm sözlərin seçimi, həm də mənə sırası istər-istəməz subyektivdir. Bu maneəni kompüterlərdən istifadə etməklə aradan qaldırmaq olar.

Kompüterlərin yaranması və inkişafı lüğətlərin yaradılmasına böyük təsir göstərmiş, ənənəvi leksikoqrafiyanın kartotekalarda saxlanılan məlumatları verilənlər bazalarına yerləşdirilmişdir. İnternetin və CD-ROM-ların yaranması istifadəçinin məlumatlara sürətli çıxışı olan yeni formatların inkişafına səbəb oldu. Bir çox elektron lüğətlərin yalnız elektron formatda olan kağız lüğətlər olduğuna dair tənqidlərə baxmayaraq, onlayn və CD-ROM lüğətləri ənənəvi lüğətlərlə müqayisədə bir çox praktik üstünlüklərə malikdir. Bu cür lüğətlər öz növbəsində, leksikoqrafik məlumatların hesablanmasına kömək etdi, çünki təbii dil emalı (NLP) tədqiqatçılarınin təhlil üçün elektron materiala çıxışı var idi [4].

Kompüter texnologiyaları bütün əsas leksikoqrafik layihələr tərəfindən qəbul edilmişdir, lakin bəzi sahələrdə mənbə məlumatlarının təbiətində daha böyük bir inqilab olmuşdur. Korpus dilçiliyi leksikoqrafik təcrübənin çox hissəsini dəyişərək çoxlu sayda etibarlı məlumat əldə etməyə imkan verdi. Korpus dilçiliyi ilə bağlı ilk tədqiqatlar isə John Sinkler və

Birminqem Universitetində COBUILD komandası tərəfindən aparıldı.

Ən yaxşı lüğətlər belə bütün problemlərimizi həll etmir. Bir çox ESP (English for spesified purpose – xüsusi məqsəd üçün ingilis dili) tələbələrinin, xüsusən də təbii elmlər sahəsində təhsil alanların problemi onların lüğətlərdən az istifadə etməsidir. Ən yaxşı halda onlar ikidilli lüğətdən istifadə edəcəklər ki, bu da öz növbəsində xeyli səhvlərlə müşahidə olunur. Makkrierinin [3] ana dilində danışanların zəif lüğət bacarıqlarına malik olduğunu göstərdiyini nəzərə alsaq, qeyridil daşıyıcılarının bundan faydalanmayacağını güman etmək olar. Xüsusən də verilən nümunələr onların ixtisas sahələrinə aid deyilsə, bu zaman səhvlər daha çox nəzərə çarpır.

Bu məsələlərin çoxu digər tədqiqatçılar tərəfindən də qeyd edilmişdir. Problemin mənbəyi bəllidir; nə qədər cəhd etsək də, lüğət yalnız kontekstsizləşdirilmiş dəyəri təmsil edə bilər, dəyərin lüğət mənbəyindən real mətnə transferini idarə etmək istifadəçidən asılıdır. Buradan belə nəticə çıxır ki, istifadəçinin kodlaşdırmaq və ya deşifrə etmək istədiyi kontekstdən nə qədər uzaq olsa, anlaşılmaqlıq riski bir o qədər çox olar. Bu zaman iki həll yolu təklif edilə bilər; daha yaxşı lüğət bacarıqlarının öyrədilməsi və daha kontekstli lüğətlərin hazırlanması.

Lüğətin təkmilləşdirilməsi Kampoy Kubillo [2] tərəfindən qəbul edilmiş strategiyadır. O, kimyaçı tələbələrlə işləyərək onları öz lüğətlərini yazmağa vadar etdi. Bu, tələbələrə lüğətlərlə tanış

olmağa və tədqiqatçıya tələbələrin lüğətlərdən necə istifadə etdiyini daha yaxşı başa düşməyə imkan verdi. Sınıf razılaşmasından istifadə edərək, tələbələrə real dili öyrənərkən kiçik şəxsi lüğətlər yaratmağı öyrətmək olar. Bununla belə, kiçik miqyasda faydalı ola və pedaqoji dəyərə malik olsa da, ESP tələbələri leksikoqraf deyillər və əsas anlayışlardan başqa hər hansı leksika yaratmaq üçün vaxt və bacarıqlara malik deyillər. Bir çox tələbələr, o cümlədən dilçilik fakültəsinin tələbələri konkordans analizinə uyğunlaşmaqda çox çətinlik çəkirlər ki, bu da çox xüsusi bacarıqdır. Aydın ki, real lüğət baxımından bu, leksikoqrafın öhdəsində qalır.

Biz başa düşməliyik ki, korpus leksikoqrafiyasında məna və anlayış aydın şəkildə fərqləndirilməlidir. Məna müxtəlif anlayışları, o cümlədən ensiklopedik və semantik anlayışları əhatə edir. Lakin bu mənalara çatdırın linqvistik vasitələr bir korpusda birləşdirilə və lüğət girişlərində qruplaşdırıla bilər. Bu, istifadəçiyə deşifrə etməyə, başa düşməyə kömək edəcək, lakin mütləq kodlaşdırmağa imkan verməyəcək. Semantik təhlil kodlaşdırmaya kömək edə bilər, lakin istifadəçi tərəfindən dərhal tanınmayan insan kateqoriyasına aiddir. Semantik yanaşma sözləri iyerarxiyaya ayıran *WordNet*-də (<www.cogsci.princeton.edu/~wn/>) görünə bilər. COBUILD kimi lüğətlər məna və anlayış arasındakı xətti keçməyə çalışırlar.

İnternet resurslarında bir çox konkordanslar var ki, onların bəziləri

pulsuz, bəziləri isə pulludur. Mayk Skott tərəfindən yaradılmış *WordSmith Tools*, ehtimal ki, *Windows*-un ən çox istifadə olunan lüğətidir. İndi dördüncü versiyasında mövcud olan *WordSmith* alətləri geniş spektrli leksik analiz imkanlarını, o cümlədən *Speech-POS* və *Text Encoding Initiative* (TEI) uyğun korpusunu idarə etmək bacarığı təklif edir.

Aydın ki, konkordanslar yaradılarkən ilk addım mətnlər korpusunun qurulmasıdır. Bu zaman dizayn meyarları da böyük əhəmiyyət kəsb edir, çünki xüsusi hal yalnız xüsusi nəticələrə səbəb ola bilər. Böyük istinad korpusuna əsaslanan lüğətlər təmsilçi olduqlarını iddia edə bilsələr də, ixtisaslaşmış korpusda əsas söz əsaslandırma, korpusa nəyin daxil olduğunu və niyə olduğunu aydın şəkildə təqdim etməkdir. Növbəti addım baş sözlərin seçilməsidir. Bu, böyük ölçüdə lüğətin nəzərdə tutulan məqsədindən asılıdır, əgər məqsəd terminləri kontekstdə göstərməkdirsə, onda əl ilə və ya avtomatik termin çıxarma sistemi tələb olunacaq. PPD vəziyyətində kollokasiya şəbəkələrindən istifadə etməklə daha geniş leksik baza axtarılır [6], bu, subyektivliyi azaltmaq və ixtisaslaşdırılmış elementlərin əksəriyyətini aradan qaldıracaq sadə tezliyə həddən artıq asılılığın qarşısını almaq üçün bir vasitə kimidir. Hansı üsul qəbul olunsa da, ən vacib addım sözün fərdi mənalarnı çıxarmaq üçün müşayiətdən istifadə etməkdir.

NƏTİCƏ

Nəticə etibarilə, qeyd olunmalıdır ki, lüğətlərin tərtibi zamanı texnologiyanın həddindən artıq istifadəsi risklərə gətirib çıxarır. Bu səbəbdən də, kompüterlər təcrübəli leksikoqrafı əvəz edə bilməz. Wordsketch və digər sözlərin aydınlaşdırılması layihələri isə daha çox leksikoqrafın işini sürətləndirmək üçün nəzərdə tutulmuşdur, lakin sürət həm də maraqlı materialın diqqətdən kənar qalması mənasını verə bilər. Belə vasitələrin istifadəsi təkcə nitq hissələrinin işarələnməsinin deyil, həm də qrammatik təhlilin etibarlılığının tanınması deməkdir. Bu mənada heç bir yanaşma mükəmməl deyil və hər bir yanaşmanın öz üstünlükləri və mənfi cəhətləri vardır. Linqvistik

korpuslar hazırlanarkən bütün bu məsələlər nəzərə alınmalıdır ki, nəticə də istənilən səviyyədə olsun.

İSTİNADLAR

- [1] Atkins, B.T.S. 2002. *Bilingual Dictionaries: past, present, future*. In Corr  ard, M.-H. (ed.), 1-29
- [2] Campoy Cubillo, M.C. 2002. “Dictionaries use and dictionary needs of ESP students: An experimental approach”. *International Journal of Lexicography* 15/3, 205-228. DOI : 10.1093/ijl/15.3.206
- [3] Fontenelle, T. 2002. “Lexical knowledge and Natural Language Processing”. In Corr  ard, M.-H. (ed.), 216-229
- [4] McCreary, D.R.. 2002. “American freshmen and English dictionaries: I had aspersions of becoming an English teacher”. *International Journal of Lexicography* 15/3, 181-205. DOI : 10.1093/ijl/15.3.181