

NƏRGİZ ƏLƏKBƏROVA
Azərbaycan Milli Elmlər Akademiyası
Nəsimi Dilçilik İnstitutunun
Hüseyn Cavid Prospekti, 31 5 mərtəbə
narguizli@gmail.com

MÜXTƏLİF SİSTEMLİ DİLLƏRDƏ TEZLİK LÜĞƏTLƏRİNİN STRUKTUR XÜSUSİYYƏTLƏRİ

Xülasə

Son dövrdə müxtəlif sistemli dillərin tədrisi informasiya texnologiyalarının inkişafı ilə əlaqədar çox yüksək səviyyədə və innovativ üsullarla həyata keçirilir. Yüzlərlə elektron və kitab şəklində hazırlanan lüğətlər dillərin tədrisində çox mühüm və xüsusi rol oynayır. Bu kimi yararlı missiyanın həyata keçirilməsində sözsüz ki, tezlik lüğətlərinin oynadığı başlıca rolu qeyd etmək lazımdır. Tezlik lüğəti tədqiq olunan leksik vahidlərin dildə yayılma tezliyini göstərən həmin vahidlərin siyahısıdır. Bu lüğətlər sözlərin düzəlmə sırası və yerləşdirilməsindən asılı olaraq müxtəlif növlərə bölünürlər. Həmin faktorları nəzərə alaraq tezlik lüğətləri tezlik, əlifba-tezlik və əks-əlifba tezlik lüğətlərinə bölünürlər.

Müxtəlif sistemli dillərin quruluşu fərqli olduğu, üçün tezlik lüğətlərinin tərtibində də müəyyən fərqlər meydana çıxır. Bununla yanaşı, bəzi indekslər bütün lüğətlər üçün eynilik təşkil edir. Bunlara leksik vahid, onun rəngi, tezliyi və mətnlərdə rastlanma sayını nümunə gətirmək olar.

Bəzi digər göstəricilər dilin mənsub olduğu ailədəki dil strukturuna görə dəyişə bilər. Tezlik lüğətləri lüğət minimumlarının hazırlanması, aktiv və passiv leksikanın, hər bir üsluba məxsus sözlərin kəmiyyət statistikasının müəyyənləşdirilməsində, dillərin tədrisi prosesində əvəzsiz mənbə rolunu oynayır. Hazırda Qlobal dünyanın ayrılmaz hissəsi olan məşin tərcüməsi tezlik lüğətlərinin bazası əsasında fəaliyyət göstərir.

Məqalədə Azərbaycanda və ölkə xaricində işlənmiş fərqli tezlik lüğətləri tədqiq olunmuşdur. Burada Azərbaycan lüğətçilik sahəsində dəyərli mənbə olan “Azərbaycan dilinin tezlik lüğəti” və “Azərbaycan qəzet dilinin tezlik lüğəti” nin tərtib üsulları, leksik vahidlərin tezliyi araşdırılmışdır. Yerli lüğətlərdən əlavə, məqalədə türk dilçiliyinə böyük töhfə olan Qazax dilinin tezlik lüğəti və ikinci beynəlxalq dil status almış İspan dilinin tezlik lüğəti də ətraflı araşdırılmışdır.

Bütün bu araşdırmalardan bir daha belə nəticəyə gələ bilərik ki, tezlik lüğətləri bir informasiya axtarış sistemində ən yararlı mənbə olmaqla yanaşı, həm də dillərin tədrisində əvəzsiz mənbə rolunu oynayır.

Açar sözlər: tezlik, lüğət, söz-forma, rəng, leksik vahid, üslub, registr

Müəyyən dil hadisələrini öyrənmək üçün həmin dilə tətbiq olunan riyazi metodlar lingvostatistika termini və ya statistik lüğətçilik adı altında birləşir. Dilçilikdə bu statistik metodların tətbiqi bəzə nitqdə və ya mətnlərdə tez-tez və ya az işlənən dil vahidlərini aşkarlamaq üçün çox böyük imkanlar verir. Bu vahidlərin işlənmə tezliyinin tədqiqi müəyyən nitq ifadələrinin əmələgəlmə qanunauyğunluğunu müəyyənləşdirməyə yardım edir. Dil vahidlərinin tezliyini müəyyənləşdirmək üçün ilk təcrübələr tezlik lüğətlərinin tərtibi ilə başlamışdır. Son dövrdə bu cür lüğətlərin tərtibi çox böyük maraq doğurur. Bu cür lüğətlər, leksikanın statistik modeli rolunu oynayıb dilin lüğət tərkibinin özülünü, aktiv və passiv lüğət fondunu müəyyənləşdirməyə imkan verir.

Tezlik lüğətləri elə lüğət növləridir ki, burada sözlərin, söz formaların və söz birləşmələrinin dildə işlənmə tezliyi rəqəmlərlə göstərilir. Növündən asılı olaraq, bəzi tezlik lüğətlərində söz forma (məs: ayaqdan, ayağa, ayağın), bəzi lüğətlərdə yalnız şəkilçisiz sadə söz (ayaq), bəzilərində isə mürəkkəb söz və ya söz birləşməsi (ayağüstü) lüğət vahidi kimi əsas götürülür. Hazırda şəkilçilərin təmsil olunduğu tezlik lüğətlərinə də rast gəlmək mümkündür.

Bu lüğətlər sözlərin düzülüş üsulundan, tezliyin çoxalma və ya azalma sırasından asılı olaraq müxtəlif növlərə bölünür. Belə ki, əlifba sırasına əsasən tərtib olunmuş bu kimi lüğətlərdə istənilən sözün tezlik dərəcəsini həmin sözün əlifbadakı yerinə əsasən dərhal əldə etmək olar.

Bununla yanaşı, qeyd etmək lazımdır ki, tezlik lüğətlərinin ərsəyə gəlməsi uzun və mürəkkəb prosesdir. Bu prosesin ən çətin mərhələsi lüğətin bazasını təşkil edən sözlərin tədqiqat obyektini kimi çıxış edən və tezliklərin araşdırılması üçün istifadə olunan, ədəbi dili əks etdirən mətnlərin seçimidədir. Belə ki, tezlik lüğətləri bədii əsərlər, qəzet materialları, elmi əsərlər, tədris materialları və s. əsasında hazırlanır. Əgər həmin seçilmiş mətnlər kiçik olarsa, bu zaman onlar dili tam təmsil edə bilmədiyi üçün müəyyən sözün obyektiv olaraq tezliyini hesablamaq mümkün olmayacaq. Məhz bu vəziyyətlə qarşılaşmamaq üçün və dilin lüğət tərkibini tamamilə əhatə etmək üçün seçmələr iri həcmli mətnlər olmalıdır.

Təcrübə göstərir ki, tezlik lüğətləri ən çox elmi-texniki mətnlər və qəzet materiallarından götürülmüş seçmələr əsasında tərtib olunur.

Bundan əlavə, tezlik lüğətləri tərtib olunma üsullarına görə də əllə tərtib olunan, maşınla və ya qarışıq üsulla tərtib olunan lüğətlərə bölünür.

“Tezlik lüğətlərinin tərtibi prosesi ümumi şəkildə aşağıdakı mərhələlərdən ibarətdir:

- mətndən bütün müxtəlif sözlər (söz formalar) götürülür;
- hər bir sözün tezliyi hesablanır;
- sözlərin əldə edilmiş məcmusu tezlik sırası ilə düzülür.

Tədqiqatın məqsədindən asılı olaraq müxtəlif siyahılar və statistik göstəricilər alınır. Məsələn, çox vaxt tezlik siyahısında əlavə sözlərin əlifba tezlik siyahısı da verilir. Lazım gəldikdə lüğətdə hər bir sözün sıra nömrəsi, mütləq tezliyi və s. göstərilir.”(4, s. 4)

Əl üsulu ilə tərtib olunan tezlik lüğətlərində mətnlərdə vahidlərin seçilməsindən tezliklərin hesablanmasına bütün proseslər insan tərəfindən həyata keçirilir. Maşınlar vasitəsi ilə tərtib olunan tezlik lüğətlərində isə yalnız yuxarıdakı əməliyyatın birinci hissəsi, yəni mətnlərdən lüğət vahidlərinin seçilməsi prosesi insan tərəfindən həyata keçirilir. Onların maşınlara daxil edilməsi də əvvəllər çap qurğusu vasitəsilə həyata keçirilirdisə, elm və texnologiyanın inkişafı sayəsində indi bu proses 2-3 saniyə ərzində həyata keçirilə bilər. Məsələn, bu sahədə xüsusi ixtisaslaşmış proqramlardan birinə “Textgrabber” adlanan yeni programı misal gətirmək olar. Belə ki, onun köməyi ilə lüğət tərtibçiləri çox böyük vaxt və zəhmət itkisindən azad olunublar. Məsələn, əvvəlki üsulla 300 min sözdən ibarət lüğəti ərsəyə gətirmək üçün tədqiqatçı 3 il vaxt sərf edirdisə, sonradan bu rəqəm texnologiyanın inkişafı sayəsində maşınların köməyi 3 ay yarım, indi isə yeni mütləqkil proqramların köməyi ilə 3 günə həyata keçirilə bilər. İndiki texnologiyanın sayəsində maşınlar nəinki verilənləri təhlil edir, əməliyyat nəticəsində vahidlərin tezliyini müəyyənləşdirib əlifba sırasına əsasən nizamlaya bilirlər.

Tezlik lüğətlərinin dilçilikdə oynadığı rolu və əhəmiyyəti, o cümlədən tətbiq dairəsi çox geniş sahəni əhatə edir.

Linqvistik statistikanın tətbiq olunduğu ən yararlı sahələrdən biri müqayisəli tarixi dilçilik olub dilin tarixi inkişafı ilə bağlı baş verən dəyişiklikləri izləməyə imkan verir. Əsas leksik bazanın nə kimi dəyişikliyə uğradığı statistik lüğətlərin təhlili aparılan zaman açıq-aşkar meydana çıxır. “Qlotoxronologiya bu statistik metodlardan biri olub qohum dillərin lüğət bazasının tarix boyu nə kimi dəyişikliyə uğradığını araşdırır. Bu sahədə araşdırma aparən dilçi alim Morris Svodeşin hipotezasına görə qohum dillərin baza lüğətinin tarix boyu dəyişikliyə uğraması demək olar ki, eyni sürətlə gedir.”²

Tezlik lüğətləri tədris prosesində və dillərin öyrədilməsində müəllim və tələbələr üçün əvəzsiz mənbədir. Xüsusilə, xarici dillərin tədrisi zamanı müəllimlər tezlik lüğətlərində vahidlərin çıxarılmış tezlik sırasına əsasən aktiv leksikanı -yəni dildə daha çox işlənən və sabit -sözləri müəyyən edib ilk növbədə həmin sözlərin tədrisindən başlamalıdır. Məsələn, bu hər hansı əcnəbi şəxsə dili daha tez qavrayıb, yerli xalqla qaynayıb qarışmağa, cəmiyyətə daha tez inteqrasiya olmağa yardımçı olur.

Müasir Azərbaycan dilinin ilk tezlik lüğəti qəzet leksikası əsasında yaranmışdır ki, bu da 1970-ci illərə təsadüf edir. Həmin illərdə bu cür lüğətlərin hazırlanması çox ağır və mürəkkəb proses idi.

Belə ki, elektron hesablama maşınlarının yalnız dəqiq elmlərlə bağlı araşdırmalarda istifadə olunduğu bir dövrdə onların dilçilik sahəsində də istifadəsinin mümkünlüyü ağılasıgmaz gəlirdi. Belə yanaşma “Azərbaycan qəzet dilinin tezlik lüğəti” və “Əks əlifba lüğəti”nin hazırlanmasında uzun müddət maneə olmuşdur. Psixoloji faktor olaraq, qısa desək kibernetika elmi də hələ həmin illərdə dilçiliyin həmin sahəyə nüfuz etməsinə hazır deyildi. Bundan əlavə digər bir faktoru da qeyd etmək lazımdır ki, maşın texnologiyası o illərdə indiki kimi yüksək inkişaf səviyyəsinə qədəm qoymamış və dilçilər hər hansı bir mətni maşınlara daxil etmək üçün perfolent və ya perfokartlar adlanan qurğulardan istifadə edirdilər. Bunlar, üzərində xüsusi dəlikləri olan kağızdan hazırlanmış xüsusi informasiya daşıyıcıları idilər. Mətnlərin bu cür qurğular vasitəsi ilə maşınlara daxil edilməsi çox böyük zaman alırdı.

Tezlik lüğətləri dildə normanın sabitləşməsində, nitq fəaliyyəti ilə bağlı araşdırmalarda çox böyük əhəmiyyət kəsb edir. Bu lüğətlər, həmçinin, danışq kitabçalarının və minimumların hazırlanmasında da əvəzsiz mənbə rolunu oynayır. O cümlədən, bu lüğətlər nəzəri və tətbiqi dilçiliyin, ümumi leksikologiyanın qarşılaşdığı problemləri həll edir. Dilin tezlik lüğətinin hazırlanmasında əsas rol mətnlərin seçilməsi, yəni seçmə oynayır və bu da çox vaxt qəzet materialına əsaslanır. Qəzet böyük bir kütlə tərəfindən oxunduğu kimi, kütlə tərəfindən yazıldığından dildə müəyyən sabit normanın yaranmasında çox böyük rol oynayır. Bundan əlavə, qəzet materialı, həmçinin, gündəlik danışq dilində meydana gələn yeni vahidləri izləməyə, nitqdə qeyri-normativ halları aşkarlamağa və yeni istifadə formalarını izləməyə imkan verən ən yaxşı tədqiqat materialıdır. Aydındır ki, bu cür lüğətlərin tərtibi zamanı seçmənin həcmi nə qədər böyük olarsa oradan əldə edilən məlumat da bir o qədər dolğun olar.

Belə zəngin qəzet materialları əsasında yaradılan azərbaycan dilinin ilk tezlik lüğəti AMEA Dilçilik İnstitutunun Elmi Şurasının qərarı ilə akademik Əli Abbasov, akademik Ağamusa Axundovun təşəbbüsü ilə ərsəyə gəlmişdir. “Lüğətin Azərbaycan dilçiliyində ilk təcrübə olması, texniki imkanların məhdudluğu və qısa müddətdə bu sahədə müəyyən elmi nəticə olmaq zərurəti ilə əlaqədar olaraq seçmənin həcmi 10000 söz-forma qəbul edilmişdir.” (1, s.129) Amma indiyə kimi tədqiqat zamanı tərkibi 20000 dən tutmuş 20 milyona qədər sözü əhatə edən tezlik lüğətləri ilə rastlaşdıq. Sözügedən lüğətin tərtibi üçün istifadə olunan seçmələr Respublikamızda 1976 ci ildə çap olunmuş “Kommunist”, “Bakı”, “Azərbaycan gəncləri” kimi qəzetlərdən götürülmüş materiallara əsaslanır. Bu lüğətin hazırlanması üçün həmin ilə aid olan, yuxarıda adları çəkilən qəzetlərin hər səhifəsindən 800 söz götürülüb işlənmişdir. Qəzet materialları ifadə etdiyi məzmunundan asılı olaraq təbii şəkildə müxtəlif janrlara bölünür. Bunlara əsasən qəzətdə baş məqalə, daxili xəbərlər (bunlar xüsusilə, ölkədaxilində baş verən elm və texnologiya, sənaye, kənd təsərrüfatı, biznes, ölkənin daxili siyasəti ilə bağlı yenilikləri, əhatə edir.), xarici xəbərlər, şou və idman sahəsini əhatə edir.

“Azərbaycan Qəzet dilinin Tezlik Lüğəti”

2004-cü ildə işıq üzü görmüş müasir “Azərbaycan qəzet dilinin tezlik lüğəti” Kamilə Vəliyeva, Məsud Mahmudov və Valeri Pinesin gərgin səyi nəticəsində meydana gəlmiş və hal-hazırda lüğət sahəsində çox böyük əhəmiyyət daşıyır.

Bu lüğət iki hissəyə bölünüb seçmələrdən əldə edilən sözləri əliba sırası ilə nizamlanmasını əks etdirən əlifba-tezlik sözlüyündən və tezlik sözlüyündən ibarətdir. Hər iki hissədə sözlərin mətnə işlənmə tezliyini əks etdirən vahidlər və həmin sözün işləndiyi mətnlərin sayını əks etdirən vahidlər əks olunub.

Əlifba-tezlik sözlüyü işlənmiş 81778 seçmədən qeydə alınmış 6532 müxtəlif sözü əks etdirir.(3,s.4)

Burada tezliyi müəyyən olunan leksik vahidlər əlifba sırasına əsasən düzülmüşdür. Sözlərin bu cür sıralanması bizə istədiyimiz sözün əlifba sırası ilə tez bir zamanda tapılmasına yardım edir.

Tezlik sözlüyü isə eyni leksik vahidlərin yüksək tezlikdən aşağı tezliyə doğru sıralanmasına əsasən tərtib olunmuşdur. Buradakı leksik vahidlərin iki kəmiyyət göstəricisi qeyd olunmuşdur. Belə ki, bunlara sözlərin mətnə işlənmə tezliyini əks etdirən vahidlər və həmin sözün işləndiyi mətnlərin sayını əks etdirən vahidlərə bölünür. Lüğətdə ən yüksək tezliyi ilə fərqlənən sözlər

bunlardır: və (tezlik -2535 yayılma 100), etmək (tezlik- 1256 yayılma-100), bu (tezlik-1040 yayılma-100), bir (tezlik-770 , yayılma 100) o (tezlik 719- yayılma 100).

Qəzet dilinin tezlik lüğətini araşdırarkən maraqlı məqamla üzləşirik. Belə ki, kitabdakı sözləri nəzərdən keçirərkən belə qənaətə gəlirik ki, bəzi leksik vahidlərin işlənmə tezliyi ilə işləndiyi mətnlərin sayını göstərən tezlik bəzən eyni, bəzən isə fərqli ola bilər. Məsələn, “davamlı” və “çəkiliş” sözlərində bu göstəricilər üst-üstə düşür və müvafiq olaraq birinci sözdə tezlik və yayılma vahidi 1-1 olduğu halda, ikinci sözdə bu göstərici 3-3 kimi təzahür edir. Bir çox hallarda isə, leksik vahidlərin tezliyi ilə rast gəlinəndiyi mətnlərin sayı uzlaşmır. Məsələn, “buraxılan” və “çempionat” sözləri- hər ikisi eyni tezliyə malikdir və lüğətdə tezlik göstəriciləri 14 dər. Lakin birinci söz 9 mətnədə , ikinci söz isə 4 mətnədə müşahidə olunur. Onu da qeyd etmək lazımdır ki, hər hansı sözün hələ yüksək tezliyə malik olması onun ümumişlək olmasına dəlalət etmir. Ümumişlək leksik vahidin həm də rast gəlinəndiyi mətnlərin say göstəricisi də yüksək olmalıdır. Belə ki, yuxarıda göstərilən nümunədə “buraxılan” sözü ümumişlək sayıla bildiyi halda , bunu “ çempionat ” sözünə şamil etmək olmaz.

“Azərbaycan dilinin tezlik lüğəti”

Son dövrdə azərbaycan dili lüğətçilik sahəsində ən zəngin xəzinə “Azərbaycan dilinin tezlik lüğəti” hesab oluna bilər. Leksik-statistik model olaraq bu lüğət nəzəri və tətbiqi dilçilikdə, dilin öyrənilməsi ilə bağlı bir sıra problemlərin həllində əvəzsiz mənbə rolunu oynayır.

Bu lüğət, həmçinin, dilin aktiv və passiv lüğət bazasının araşdırılmasında, dilin lüğət tərkibinin tarix boyu uğradığı dəyişikliyin müşahidə edilməsində, fərdi üslub da daxil olmaqla elmi, işguzar, məişət, publisistik və s. kimi üslubların kəmiyyət nöqtəyi-nəzərindən araşdırılmasında çox böyük elmi əhəmiyyətə malikdir.

Digər tezlik lüğətləri kimi, bu lüğət də xarici dillərin və ana dilinin tədrisi prosesində oxucular üçün çox yararlı mənbədir. Bu lüğətin köməyi ilə lüğət minimumlar və ya danışiq kitabçaları hazırlayıb dilin tədrisi prosesində faydalanmaq olar.

Leksik statistik model kimi tezlik lüğəti dilin sözmələgətirmə və sözdəyişmə imkanlarının izləməyə və tədqiq etməyə imkan verir.

“Azərbaycan Qəzet dilinin Tezlik Lüğəti ” qəzet üslubunu özündə əks etdirirdisə, Azərbaycan dilinin tezlik lüğəti özündə elmi, bədii, işguzar, publisistik, məişət və s. üslubları cəmləşdirir. Bundan əlavə, bu lüğətin tərtibi üçün çox zəngin leksik bazadan -50 milyona yaxın söz-formadan təşkil olunmuş mətnlərdən istifadə olunmuşdur. “Həmin bu söz formaları kompüterlər vasitəsi ilə lingvostatistik baxımdan təhlil olunduqdan sonra istifadə edilmişdir. Gələcəkdə həmin baza əsasında müxtəlif tipli tezlik lüğətlərinin (əsasların, söz formalarının , söz köklərinin tezlik, əlifba-tezlik və əks əlifba-tezlik siyahıları, qrammatik və leksik şəkilçilərin , qrafem və qrafem birləşmələrinin, durğu işarələrinin tezlik səciyyələrinin əks etdirən siyahılar və s.) hazırlanması nəzərdə tutulur.” (4,s.3)

Yuxarıda adı çəkilən baza əsasında yaradılan “Azərbaycan dilinin tezlik lüğəti ” 1-ci cildi 2010 cu ildə tərtibçilər M.Mahmudov, Ə.Fətullayev, S.Məmmədova ,R.Fətullayev, R.Hüseynli, B.Talıbov, G.Əzimova ,N.Abdullayevin gərgin səyi və İnformasiya və texnologiya Nazirliyinin dəstəyi ilə ərsəyə gəlmişdir.

Bu lüğətin üç cildə hazırlanması nəzərdə tutulmuşdur. Birinci cild söz-köklərinin tezliyi əsasında hazırlanmışdır.

“Lüğətdə sözün kökü dedikdə, sözün şəkilçiləri ayrıldıqdan sonra qalan və müəyyən əşyavi məna bildirən hissəsi başa düşülür. Burada söz kökləri müasir inkişaf vəziyyəti baxımından araşdırılır, tarixi, etimoloji aspekt nəzərə alınmır. Ünsiyyət prosesində müəyyən morfonoloji dəyişmələrə məruz qalmış söz kökləri ilkin qrammatik formaya gətirilmir. Məsələn : "gedir" söz-formasında şəkilçi atıldıqdan sonra qalan hissə "ged" kök hesab olunur, əslində, sözün kökü ilkin formada "getmək" hesab olunur.” (4,s.4)

Lüğətdə cəmlənən söz köklərinin tezliyi üç kateqoriyaya əsasən qruplaşdırılmışdır : Tezlik siyahısı, əlifba -tezlik siyahısı və əks əlifba tezlik siyahısı .

Tezlik siyahısında dilin müxtəlif üslublarında işlənen sözlər yüksək tezlikdən azalan tezliyə doğru sıralanmışdır.

Əlifba -tezlük siyahısında həmin leksik vahidlər tezliyə görə deyil, əlifba sırasına əsasən düzülüb.

Əks -əlifba tezliyində isə statistik hesablamağa cəlb olunan söz-kökləri sözün sonundakı hərfa görə əlifba sırası ilə yer almışdır.

Lüğətdə “Azərbaycan Qəzet dilinin Tezlük Lüğəti” ndə olduğu kimi ən yüksək tezlük göstəricisini “və”, “ol”, “bu”, “o” kimi söz kökləri nümayiş etdirir. Lakin həmin lüğətdən fərqli olaraq burda cəmlənən sözlər bütün üslubları əhatə etdiyindən həcme daha böyükdür, belə ki, ən yüksək tezlük 1388871 “və” qoşmasına məxsus olmuş, ən alçaq tezlük isə -6 dır. Azərbaycan Qəzet dilinin Tezlük Lüğəti” ndə isə bu göstərici müvafiq olaraq 2500-1 aralığında dəyişir.

Qazax dilinin tezlük lüğəti

Tezlük lüğətlərindən söz açarkən müasir türk dilçiliyində bu günə qədər işlənmiş ən zəngin və irihəcmli lüğət sayılan “Qazax dilinin tezlük lüğəti”ni xüsusi vurğulamaq lazımdır. Belə ki, lüğət 61 cilddən ibarət olub məzmun və formasına görə XXI əsrin ən nadir tezlük lüğəti hesab olunmaqdadır. Lüğətin bütün cildləri bir DVD diskinə yerləşdirilib və iki lüğət böyük bir kitab şəklində dərc edilmişdir.

Bu iri həcmli, zəngin söz ehtiyatına malik olan lüğətin ərsəyə gəlməsində Qazaxıstan Respublikasının Elm və Təhsil Nazirliyinin, Əhməd Baitursinuly adına Dilçilik İnstitutunun 30-dan çox tədqiqatçısı iştirak etmişdir.

Lüğətin hazırlanmasında ümumi redaksiya heyətinə professor Erden Hacıbek və filologiya elmləri namizədi Anar Fazılcanova rəhbərlik etmişdir.

Sözündən tezlük lüğəti 7 milyondan artıq sözün istifadəsi əsasında yaradılmışdır ki, bu da öz növbəsində, qazax dilinin 30 milyon ümumişlək söz bazasından seçilmişdir. Qeyd edək ki, bu 61 cilddən ibarət tezlük lüğəti "Qazax dilinin Milli Korpusu" adlı elmi proqramın bir hissəsidir və növbəti iki il ərzində bu bazanı təşkil edən sözlərin 100 (yüz) milyona qədər artırılması nəzərdə tutulur.

“Qazax dilinin tezlük lüğətləri” aşağıdakı üç növə bölünür:

- Əlifba tezliyi lüğəti;
- Tezlük lüğəti;
- Əks əlifba tezlük lüğəti.

Lüğətə daxil olan söz bazası ayrı-ayrı üslublar şəklində əks olunmuşdur. Bunlara: elmi üslub; rəsmi-ışğuzar üslubu; bədii üslub; publisistik üslub; məişət üslubu və s. daxildir ki, hər bir üslub da öz növbəsində tezliyin ifadə olunma növünə görə 3 cildə hazırlanmışdır.

Yuxarıda adları çəkilən üslublara əlavə olaraq, lüğətdə həm də ümumi baza yer almışdır.

Bundan əlavə bəzi lüğətlər qazax dilinin əsas nitq hissələrinin və köməkçi nitq hissələrinin tezliyini özündə əks etdirir. Bunlara: isim; sifət; say; əvəzlilik, feil, zərf; ara sözlər, qoşma, modal sözlər; köməkçi fe'llər daxildir.

Lüğətlə bağlı digər maraqlı bir məqam orada dilin fəal lüğətinin bir hissəsi olan omonimlərin də yer almasıdır.

“Qazax dilinin tezlük lüğəti”nin tərtibi zamanı dil vahidi olaraq söz-forma əsas götürülmüşdür. Bu dil vahidinin əsas götürülməsi qazax tədqiqatçılarna qrafik sistemin öyrənilməsi və sözdəyişdirmə prosesini izləməyə geniş imkanlar verəcək.

Ümumi lüğət - 36 265 leksik vahidi ehtiva etmişdir və 3 hissəyə bölünmüşdür. 1-ci lüğətdə söz-formaların tezliyi əlifba sırasına əsasən düzülüb, onların rəngi, mütləq tezliyi və rast gəlinməsi mətnlərin sayı göstərilmişdir.

Ümumi tezlük-əlifba lüğətində isə leksik vahidlər yüksək tezlükdən alçaq tezliyə doğru sıralanmışdır. Yuxarıda adı çəkilən bölgülər bu lüğətdə də aparılmışdır. Bundan əlavə, leksik vahidlərin əlifba-tezlük lüğətindəki sıraları da qeyd olunmuşdur.

Ümumi əks-əlifba tezlük lüğətində isə leksik vahidlərin rəngi, mütləq tezliyi, hansı nitq hissəsinə məxsusluğu və rast gəlinməsi mətnlərin sayı göstərilmişdir. Ümumi tezlük lüğətində tezliyi ən yüksək olan sözlər siyahısında : бол (olmaq) -108133, ол (o) - 62796, мен (və) – 58406 kimi sözlər qeydə alınmışdır ki, bunlar da “Azərbaycan dilinin tezlük lüğəti” və “Azərbaycan qəzet dilinin tezlük lüğəti” siyahısında yüksək tezlikli vahidlər sırasına daxil edilmişdir.

1. Ümumi əlifba-tezlik lüğəti;
2. Ümumi tezlik-əlifba lüğəti;
3. Ümumi əks-əlifba tezlik lüğəti.

Çox zəngin söz ehtiyatına malik olan bu lüğətdə üslubların distributiv tezliyi yer almışdır ki, o da öz növbəsində fərdi olaraq hər bir üsluba daxil olan leksik vahidlərin tezliyini göstərir və müvafiq olaraq elmi, rəsmi-işguzar, ədəbi üslub, publisistik üslub, məişət üslubu, üslublar cəmi, ortaqlıq baza, ümumi tezlik adlı başlıqlara bölünmüşdür. Həmin yarımbaşlıqlarda leksik vahidlərin həmin üslublardakı tezlik vahidi yazılmışdır.

4. Üslublara görə distribütiv əlifba-tezlik lüğəti;
5. Üslublara görə distribütiv tezlik- əlifba lüğəti;
6. Üslublara görə distribütiv əks əlifba -tezlik lüğəti;

7. Elmi üslubun əlifba-tezlik lüğəti; (22159 leksik vahidi ehtiva etmişdir. Bu tezlik lüğətində leksik vahidin rəngi, özü, mütləq tezliyi və yazılma dərəcəsi göstərilmişdir. Burada ən yüksək tezlik göstəricisinə malik vahidlər “və” (tezlik–2535, yayılma-100), “etmək” (tezlik- 1256, yayılma-100), “bu” (tezlik-1040, yayılma-100), “bir” (tezlik-770, yayılma 100), o (tezlik 719 - yayılma 100) və s.

8. Elmi üslubun tezlik-əlifba lüğəti;
9. Elmi üslubun əks-tezlik lüğəti;

10. Rəsmi-işguzar üslubun əlifba-tezlik lüğəti (7488 leksik vahidi ehtiva etmişdir. Burada leksik vahid, onun rəngi, mütləq tezliyi və yayılma dərəcəsi göstərilmişdir. Ən yüksək tezlik göstəricisinə malik vahidlər: жыл (il) - 14582, республика (respublika) - 5674, ол (o) – 5271;

11. Rəsmi-işguzar üslubun tezlik-əlifba lüğəti;
12. Rəsmi-işguzar üslubun əks-əlifba tezlik lüğəti;

13. Ədəbi üslubun əlifba-tezlik lüğəti; (12289 leksik vahidi ehtiva etmişdir. Burada həmin vahidlər, onların rəngi, mütləq tezliyi və yayılma dərəcəsi qeyd olunmuşdur);

14. Ədəbi üslubunun tezlik-əlifba lüğəti;
15. Ədəbi üslubunun əks-əlifba tezlik lüğəti;

16. Publisistik üslubun əlifba-tezlik lüğəti (15079 leksik vahidi ehtiva etmişdir. Ən yüksək tezlik göstəricisinə malik vahidlər: бол (olmaq) - 9549, ол (o) - 7039, бер (vermək) - 5708 və s.;

17. Publisistik üslubun tezlik-əlifba lüğəti ;
18. Publisistik üslubun əks-əlifba tezlik lüğəti;

19. Şifahi üslubun əlifba-tezlik lüğəti ;(9370 leksik vahidi ehtiva etmişdir. Ən yüksək tezlik göstəricisinə malik vahidlər: бол (olmaq) - 7380, ол (o) - 4358, ал (odur) - 4321, айт (deyirəm) – 3896;

20. Şifahi üslubun tezlik-əlifba lüğəti;
21. Şifahi üslubun əks-əlifba tezlik lüğəti;

22. Ümumi mətn verilənlər bazasının əlifba-tezlik lüğəti; (24365 leksik vahidi ehtiva edir. Həmin vahidlərin rəngi, özü mütləq tezliyi, yayılma dərəcəsi verilmişdir. Ən yüksək tezlik göstəricisinə malik vahidlər: бол (olmaq) 47880, ол (o) 30274, мен (və) 23411, да (çox) 21256;

23. Ümumi mətn verilənlər bazasının tezlik -əlifba lüğəti;
24. Ümumi mətn bazasının əks-əlifba tezlik lüğəti;

25. İsmi əlifba-tezlik lüğəti;
26. İsmi tezlik-əlifba lüğəti;
27. İsmi əks-əlifba tezlik lüğəti

28. Sifətin əlifba-tezlik lüğəti;
29. Sifətin tezlik- əlifba lüğəti;
30. Sifətin əks-əlifba-tezlik lüğəti;

31. Sayın əlifba-tezlik lüğəti;
32. Sayın əks-əlifba tezlik lüğəti;
33. Sayın tezlik əlifba lüğəti;

34. Əvəzlilərin əlifba-tezlik lüğəti;

35. Əvəzlərlərin tezlik-əlifba lüğəti; (176 leksik vahid ehtiva etmişdir. Ən yüksək tezlik dərəcəsinə malik sözlər : ол (o) -62796, будур (budur) -32850 , bir (bir) -30288 və s.
36. Əvəzlərlərin əks-əlifba tezlik lüğəti;
37. Feillərin əlifba-tezlik lüğəti;
38. Feillərin tezlik-əlifba lüğəti;
39. Feillərin əks-əlifba lüğəti;
40. Zərflərin əlifba-tezlik lüğəti;
41. Zərflərin tezlik-əlifba lüğəti;
42. Zərflərin əks-əlifba tezlik lüğəti;
43. Köməkçi nitq hissələrinin əlifba-tezlik lüğəti;
44. Köməkçi nitq hissələrinin tezlik əlifba lüğəti;
45. Köməkçi nitq hissələrinin əks-əlifba tezlik lüğəti;
46. Xitabların əlifba-tezlik lüğəti;
47. Xitabların tezlik -əlifba lüğəti;
48. Xitabların əks-əlifba tezlik lüğəti;
49. Təqlidi sözlərin əlifba-tezlik lüğəti;
50. Təqlidi sözlərin tezlik-əlifba lüğəti;
51. Təqlidi sözlərin əks-əlifba tezlik lüğəti;
52. Modal sözlərin əlifba-tezlik lüğəti;
53. Modal sözlərin tezlik-əlifba lüğəti;
54. Modal sözlərin əks-əlifba tezlik lüğəti;
55. Köməkçi feillərin əlifba-tezliyi lüğəti;
56. Köməkçi feillərin tezlik-əlifba lüğəti;;
57. Köməkçi feillərin əks-əlifba tezlik lüğəti;
58. Omonimlərin üsluba distributiv tezlikli lüğəti;
59. Söz formalarının əlifba-tezlik lüğəti;
60. Söz formalarının tezlik-əlifba lüğəti;
61. Söz formalarının əks-əlifba tezlik lüğəti;

“Qazax dilinin tezlik lüğəti” dilin tədrisi, yeni lüğətlər yaradılması, kompüter hesablama dilçiliyi sahəsində, dilin linqvistik tipoloji sahəsində tədqiqatlar üçün istifadə olunması nəzərdə tutulur.

Hazırda Qazaxıstan Elm və Təhsil Nazirliyinin əmri ilə yaradılmış bu lüğətdən rusdilli şagirdlərin məktəbdə qazax dilinin mənimsəmələri üçün istifadə olunması nəzərdə tutulmuşdur. Belə ki, dilin bu üsulla öyrədilməsi Pareto prinsipinə əsaslanıb dil tədrisində olduqca effektiv nəticə verir. Belə ki, qazax dilinin tədrisində şagirdlər üçün leksik baza minimumu yazılı nitqin 85 %-ni təşkil edir . Bu minimumu mənimsəmək üçün məktəbli şagirdə 1 tədris ili ərzində hər gün 8 söz öyrənmək kifayət edir.

Ümumiyyətlə, lüğət bir çox ünsürlərə görə Azərbaycan dilinin tezlik lüğəti ilə çox böyük oxşar cəhətlərə malikdir. Bunlara lüğətin tərtib sistemini (əlifba -tezlik, tezlik lüğəti, əks-əlifba tezlik lüğəti), lüğət vahidi olaraq söz-formanın əsas götürülməsini, ən yüksək tezlikli vahidlərin eynilik təşkil etməsi kimi məqamları misal göstərmək olar.

İspan dilinin tezlik lüğəti

Apardığımız tezlik lüğətləri ilə bağlı araşdırmalarda diqqətimizi ispan dilinin tezlik lüğəti çəkdi. Belə ki, lüğət hazırlanma tərzinə görə yuxarıda sadalanan lüğətlərdən nisbətən fərqli olub , özünəməxsus xüsusiyyətləri ilə maraq doğurur.

İspan dilinin tezlik lüğəti, İspan dilini öyrənənlər üçün əvəzolunmaz vasitə olmaqla bu dildə ən çox istifadə olunan 5000 sözün siyahısını özündə əks etdirir və 20 milyonluq söz bazasından istifadə əsasında meydana gəlmişdir. Həmin söz bazası-korpus ispan və latın Amerika mətnlərindən götürülmüş şifahi, bədii və gündəlik mətnlər əsasında formalaşmışdır. Nəzərə alsaq ki, ispan dili istifadəçilərinin sayı 470 milyon olmaqla Çin dilindən sonra dünyada ikinci beynəlxalq dil statusu almışdır, bu lüğətin hazırlanması ilə Brigham Young Universitetinin Linqvistika Şöbəsinin dil tədqiqatçıları bu dildən yararlanmaq istəyən çox böyük oxucu kütləsinə töhfə vermişdir. Bunu da

qeyd etmək olar ki, ispan dili əsasən İspaniya və bir sıra Latın Amerikasına ölkələrində istifadə olunur. Lüğətin yaranmasında böyük zəhməti keçən Brigham Young Universitetinin Linqvistik Şöbəsinin dosenti Mark Davies lüğətlə bağlı düşüncələrinin belə ifadə edir :

“Bir tezlik lüğətinin müəllimlər və dil öyrənənlər üçün dəyəri nədir? Niyə ancaq hər bir dərslərdə olan lüğət siyahısından yararlanmayaq? Bu sualın ən qısa cavabı bundan ibarətdir ki, ənənəvi dərslər kitabı hər fəsildə müəyyən mövzu ilə bağlı lüğəti təqdim edir (məs:qida, xəstəlik, nəqliyyat, geyim və s.), lakin həmin siyahıda şagirdin danışdığı zamanı və gündəlik mətnlərdə bu sözlərdən ən çox hansı ilə qarşılaşacağı barədə heç bir göstərici olmur. Əslində, bəzi sözlər o qədər nadir hallarda işlənir ki, şagird onlarla həmin fəslə aid testdən kənarında, “real həyatda” heç bir yerdə bir daha qarşılaşmır.”(5,s.1)

Sözlük, həmçinin, ayrıca müxtəlif mövzular üzrə ən çox işlənən leksik vahidlərin 30 tematik siyahısını özündə əks etdirir. İspaniyanın tezlik lüğəti, bütün səviyyələrdə olan tələbələrə, ispan lüğətlərinin öyrənilməsinə effektiv və cəlbədicə bir şəkildə, maksimum dərəcədə sadələşdirilmiş üsulla mənimsəməyə yardım edir.

“Bəzən xarici dil öyrənən tələbələr qəzet və ya bədii ədəbiyyatdan onlara tanış olmayan sözləri seçib onları öyrənməyə çox böyük zaman ayırırlar. Bu tezlik lüğəti bəzən çox böyük yararı olmayan sözlərin öyrənilməsinə sərf olunan zaman israfının qarşısını almağa yardım edəcək.”(5,s.2)

Bəzən müəllimlər də orta dil səviyyəsi qruplarına (intermediate groups) tələbələrin dil bazasının həmin səviyyə üçün kifayət qədər dolğun olmadığını sezdikdə bu lüğətlərdən tələbənin lüğət ehtiyatındakı boşluqları doldurmaq üçün istifadə edə bilirlər.

Sözügedən lüğətdə leksik vahidlərin rəngi, leksik vahidin özü, onun ingilis dilində ekvivalenti, nümunə olaraq işləndiyi cümlə , mətnlərdə yayılma dərəcəsi, tezliyi və hər hansı leksik layəyə məxsusluğu ifadə olunmuşdur. Hind-avropa dil ailəsinə məxsus ispan və ingilis dillərində azərbaycan dilindən fərqli olaraq 3 registr (dil lüğət tərkibinin üslubi leksik layələri) mövcuddur. Bunlara spoken-şifahi, fiction- bədii, non-fiction -ədəbi registrlər daxildir. Birinci və ikinci registrlər azərbaycan dilində şifahi və bədii üslublara uyğun gəlsə də, sonuncu, yəni “non-fiction”ədəbi registra azərbaycan dilindəki publisitik, elmi, rəsmi-ışguzar üslubun sözləri daxildir.

Bunu lüğətdən götürülmüş konkret nümunə əsasında daha aydın izah etmək olar. Lüğətdə yer alan söz “bruja” (cadugər) sözünü nəzərdən keçirək :

4305 bruja nf witch, hag

- había una leyenda de una bruja que se montaba en una escoba
- 61-251 +f -nf

Yuxarıdakı göstəricilərə nəzər salsaq, ilk rəqəm **4305** leksik vahidin rəngini -yəni həmin tezlik lüğətində sırasını, **bruja** -leksik vahidin özünü , **nf-** (noun -isim ,femine-qadın cinsi)-hansı nitq hissəsi və cinsə məxsusluğunu , **witch, hag-**isə həmin leksik vahidin ingilis dilində ekvivalentini ifadə edir. Sözlük kontekstə daha yaxşı anlaşılması üçün **bruja** -cadugər cümlədə də işlədilir. Belə ki, “**había una leyenda de una bruja que se montaba en una escoba**” –“Süpürgə üstündə uçan bir cadugar var idi.”- kimi tərcümə olunur.

Növbəti rəqəmlərə gəldikdə isə 61 -bruja sözünün mətndəki yayılma dərəcəsinə, 251 isə həmin sözün tezliyinin göstərir. Nəhayət, +f isə sözün daha çox bədii ədəbiyyatda işləndiyinə, -nf isə rəsmi üslubda rast gəlinmə ehtimalının çox aşağı olduğuna dəlalət edir.İndi isə, fikrimizə aydınlıq gətirmək üçün bu indeksləri bir-bir nəzərdən keçirək.

Nitq hissələri

Qeyd etmək lazımdır ki, azərbaycan dilindəki kimi ispan dilində də bəzi kateqoriyalar ikili xarakter daşıyır. Məsələn, isimlər , xüsusilə , dini mənsubiyyət və milliyəti ifadə edən sözlərdə (católico-katolik, americano-amerikan) sifətləşir, və ya sifətlər (todos-hamısı) əvəzləşir. Xüsusilə, bu dildə isimlər müxtəlif cins kateqoriyaları qəbul etdiyindən bunu da lüğətdə qeyd etmək

zəruri olur. Əksər isimlər ya kişi cinsinə məxsus olub lüğətdə “nm” (noun masculine; año- il, libro- kitab), ya da qadın cinsinə məxsus olub “nf” kimi (noun feminine; tierra-torpaq, situación-vəziyyət) qeyd olunmuşdur. Bir çox hallarda peşə bildirən isimlər lüğətdə nm/f kimi (autor-müəllif, director- rejissor) qeyd olunur , ancaq lüğətdə kişi cinsi kateqoriyası əks olunmuş , lakin həmin vahidin qadın cinsinə məxsus forması da (autora, directora) tezlik lüğətinin statistikasında həmin forma ilə bir yerdə nəzərə alınmışdır. Bir neçə sözün isə həm qadın , həm kişi cinsi forması vardır, lakin belə sözlər də bir neçə mənə ifadə edir. (cometa = comet (m)- kometa (k) ; kite (f)- çərpələng (q)).

İngilis dilində ekvivalentlik

İspan dilinin tezlik lüğəti hər bir sözün bütün mümkün mənalarını ifadə edən ikidilli lüğət olmadığından burada hər bir sözün ingilis dilindəki yalnız ən sadə və ən çox işlək olan ekvivalenti yer almışdır. Həmçinin, lüğətdə elə birləşmələrə rast gəlinir ki, eyni mənəni ifadə edib cüzi yazılış tərzləri ilə fərqlənir. Bunlar üçün də ümumi tezlik hesablanmış və belə vahidlər lüğətdə əllə sayılacaq qədər azdır :180 embargo nm sin embargo: nevertheless (baxmayaraq ki), 333 medida nf measure; a medida que: to the extent

Konteksts daxili birləşmələr

Butun bu cümlə və birləşmələr İspan dili korpusundan (Corpus del Español (www.corpusdelespanol.org)) götürülmüşdür və əsas məqsəd tədqiqat obyektini olan leksik vahidləri daha dolğun ifadə edən minimum sözlərdən təşkil olunmuş söz birləşmələri və cümlələri tapmaq idi.Yeni “uydurulmuş” qısa cümlələrlə sözsüz ki, leksik vahidlərin əsas mənəsini daha aydın ifadə etmək olardı, lakin bu orijinaldan kənar olardı. Son olaraq, orijinaldan yoxluğu ilə cümlələrin mənəsinə heç bir xələl verməyəcək sözləri ixtisar etməklə bəzi uzun cümlələr qısaldılmış və anlaşıqlı olmaq üçün sadələşdirilmişdir.

Registr müxtəlifliyi

Hər bir leksik vahidin tezliyindən əlavə leksik lay göstəricisinin yanında simvollar [+ o -o + f -f + nf-nf] qeyd olunur və buradakı (+) və ya (-) işarələr həmin laylarda leksik vahidin çox və ya nadir hallarda işləndiyinə dəlalət edir.

Buradakı

[+ o ; -o] -oral-şifahi registri, (leksik lay)

[+ f ; -f] -fiction -bədi registri, (leksik lay)

[+ nf ; -nf] non-fiction -ədəbi (leksik lay) registri ifadə edir.

Bu göstərici yalnız sözün tezliyi digər iki registrlərdən 8-9% yüksək və ya 8-9% aşağı göstəriciyə malik olduğu zaman qoyulur. Onu da qeyd etmək lazımdır ki, [+o] işarələnən bəzi sözlər-dilin şifahi korpusunda çox rastlansa da, gündəlik danışmada çox rastlanmaya da bilər. Bu ona əsaslanır ki, dilin şifahi korpusu özündə hər gün keçirilən mətbuat konfranslarını, siyasi çıxışları, siyasətçilərdən alınan intervyuları özündə birləşdirir.

Tematik lüğət

Yuxarıda bəhs etdiyimiz tezlik lüğətində, həmçinin, 30 fərqli mövzu üzrə leksik vahidlərin tezliyi qeydə alınmışdır. Bu siyahıya bədən üzvləri, qida, ailə, hava, sənət-peşə, milliyyət, rənglər, hisslər, hərəkət və ünsiyyət fe'lləri və digər mövzularla bağlı leksik vahidlər aiddir. Bundan əlavə, tezlik lüğətini ərsəyə gətirən dilçilər ispan dilini yeni öyrənən və orta dil biliklərinə malik tələbələrə qrammatikada ən çox üzləşdikləri çətinlikləri də nəzərə almışlar. Məsələn, lüğətdə ən çox işlənən sifətin kiçiltmə “ito” (gato/gatito; cat/kitty; pişik /balaca pişik) və böyütmə “isimo” (alto/altísimo; tall/really tall; hündür/çox hündür) bildirən şəkilçiləri , isim düzəldən şəkilçilər, və ən çox işlənən fellər yer almışdır. Nəhəyət, lüğətdə ən maraqlısı isim, sifət , fel və zərflərin hər 3 leksik layda -registrdə (şifahi, bədi, ədəbi) yeri əks olunmuşdur. Buna əsasən, son 100-200 ildə dilə daxil olan ən yüksək tezlikli leksik vahidləri müəyyən etmək olar.

Yaxın gələcəkdə zəngin baza əsasında yaradılacaq yeni azərbaycan dili tezlik lüğətlərində leksik vahidlərin ingilis dilində uyğun ekvivalentlərini qeyd etmək, və onları nümunə olaraq sadə cümlələrdə işlətmək zənnimcə , çox faydalı iş olardı. Belə ki, bu son dövrdə ölkəmizə səyahət edən əcnəbi turist və tələbələr üçün bu dili tez bir zamanda mənimsəmək üçün əvəzedilməz mənbə rolunu oynaya bilər.

Ədəbiyyat siyahısı

- 1.M.Mahmudov “Komp dilçiliyi” Bakı,”Elm və təhsil”, 2013, s.121
- 2.<https://ru.wikipedia.org/wiki/%D0%93%D0%BB%D0%BE%D1%82%D1%82%D0%BE%D1%85%D1%80%D0%BE%D0%BD%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D1%8F>
- 3.K.Vəliyeva, M.Mahmudov, V.Pines, “Azərbaycan qəzet dilinin tezlik lüğəti”, Bakı, Nurlan, 2004- 3 s .
- 4.Azərbaycan dilinin tezlik lüğəti (tərtibçilər:M.Ə.Mahmudov ,Ə.Fətullayev və b.), Bakı , Elm, 2010,-464 s
5. Mark Davies “A Frequency Dictionary of Spanish” , first published 2006 by Routledge 270 Madison Ave, New York, NY 10016 2006
- 6.Жалпы білім берудегі қазақ тілінің жиілік сөздігі / Жалпы редак. басқарғандар: Қажыбек Е.З., Фазылжанова А.М. – Алматы: Дәуір, 2016. – 1472 б.
7. Ляшевская О. Н., Шаров С. А. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). — М.: Азбуковник, 2009. — 1087 с. — ISBN 978-5-91172-024-7.
- 8.Adam Kilgarrieff. “Putting Frequencies in the Dictionary” // International Journal of Lexicography. — 1997. — № 10(2). — P. 135—155

N.Alakbarova

Structural features of frequency dictionaries in various systemic languages Summary

Recently, the teaching process of languages of various systems has been carried out in a very high and innovative way, with the advent of information technology. Hundreds of electronic and book dictionaries play a very important and special role in teaching languages. It is important to note the key role played by frequency dictionaries in realizing such a useful mission. These dictionaries consist of a list of lexical units that show the frequency of those lexical units in a language. Frequency dictionaries are divided into different types depending on the sequence and the frequency of words. According to these factors, they are divided into frequency, alphabetical and reverse-alphabetical frequency dictionaries.

There are also differences in the structure of various system languages, which affects the creation of frequency dictionaries. However, some indexes remain the same for all dictionaries. These include lexical unit, its rank, frequency, and number of entries in the texts.

Some other indicators may vary depending on the language structure, to which language family the language belongs. Frequency dictionaries are an indispensable source of language teaching, preparation of minimum dictionaries, defining the number of active and passive vocabulary, determining quantitative statistics for each style of the word. Currently, machine translation, which is an integral part of the global world, works on the bases of frequency dictionaries.

This article has been studied various frequency dictionaries in Azerbaijan and abroad. The article also discusses compiled methods for creating the “Frequency dictionary of the Azerbaijani language” which is a valuable source in the field of the Azerbaijani dictionary and “The frequency dictionary of the Azerbaijani newspaper language” and the frequency of lexical units in both of them. In addition to local dictionaries, the dictionary of the Kazakh language, which made a great contribution to Turkish linguistics, the Spanish dictionary was also carefully studied. In all of these studies, we can conclude that frequency dictionaries are the most useful source in an information retrieval system, but also play an indispensable role in learning a language.

Keywords: frequency, dictionary, word-form, rank, lexical unit, register

**Структурные особенности частотных
словарей на разных языках
Резюме**

В последнее время обучение языков различным системам было проведено очень высоко и инновационно, с появлением информационных технологий. Сотни электронных и книжных словарей играют очень важную и особую роль в обучении языкам. Важно отметить ключевую роль, которую играют частотные словари в реализации такой полезной миссии. Эти словари представляет собой список тех единиц, которые показывают частоту изучения лексических единиц в языке. Частотные словари делятся на различные типы в зависимости от последовательности и расположения слов. Учитывая эти факторы они разделены на частотные, алфавитные и обратно-алфавитные частотные словари. Существуют также различия в структуре различных системных языков, которое влияет на создание частотных словарей. Однако некоторые индексы одинаковы для всех словарей. К ним относятся лексическая единица, ее ранг, частота и количество вхождений в текстах. Некоторые другие показатели могут различаться в зависимости от языковой структуры, к которой языковой семье принадлежит язык. Частотные словари служат незаменимым источником преподавания языка, сводя к минимуму минимальное количество активной и пассивной лексики, определяя количественную статистику для каждого стиля слова. В настоящее время машинный перевод, являющийся неотъемлемой частью глобального мира, работает на основе словаря частотных словарей.

В этой статье было изучена различные частотные словари в Азербайджане и за рубежом. В статье, так же рассматриваются скомпилированные методы создания «Частотный словарь азербайджанского языка» и «Частотный словарь азербайджанского газетного языка» и частота лексических единиц, которая является ценным источником в области азербайджанского словаря. Помимо местных словарей, словарь казахского языка, который внес большой вклад в турецкую лингвистику, Испанский словарь также был тщательно исследован. Во всех этих исследованиях мы можем заключить, что частотные словари являются наиболее полезным источником в информационной поисковой системе, но также играют незаменимую роль в обучении языку.

Ключевые слова: частота, словарь, словоформа, ранг, лексическая единица, стиль, регистр

Redaksiyaya daxil olma tarixi: 18.04.2019

Çapa qəbul olunma tarixi: 23.04.2019

Filologiya elmləri doktoru, professor Məsud Mahmudov tərəfindən çapa tövsiyə olunmuşdur.