

UOT 004.82 (004.650:06.72:80)

Qurbanova Ə.M.

AMEA İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
afruz1961@gmail.com

TERMİN YARADICILIĞI İŞLƏRİNİN AVTOMATLAŞDIRILMASI METODLARININ TƏHLİLİ

Məqalədə dilin inkişafına cəmiyyətin şüurlu təsirinin nəticəsi olan termin yaradıcılığı məsələlərinə baxılmışdır. Qeyd edilmişdir ki, istənilən dilin lüğət tərkibi həm daxili, həm də xarici resurslar hesabına yenilənir və sözün dildə yaşam dövrü onun istifadəsinin aktivliyindən asılıdır. Terminologiyanın harmonikləşdirilməsi prinsipləri müəyyənləşdirilmişdir. Anlayışların de-terminləşmə prosesi öyrənilmiş, bu prosesin səbəbləri müəyyənləşdirilmişdir. Leksikanın xüsusi və ümumişlək sözlərinin qarşılıqlı münasibətinin de-terminləşmə prosesində rolu qeyd edilmişdir. Terminlərin mətnlərdə rastgəlmə tezliyi və çəki əmsalının hesablanması metodları araşdırılaraq təhlil edilmişdir.

Açar sözlər: dil siyasəti, termin yaradıcılığı, terminlərin harmonikləşdirilməsi, de-terminləşmə prosesi, terminin çəki əmsalı.

Giriş

Dilin leksik tərkibi cəmiyyətdə baş verən dəyişikliklərə çox həssasdır. Ədəbi dilin lüğət tərkibinin həcmi onun funksional imkanlarını çoxaldır. Dilin lüğət tərkibinin yenilənməsi yalnız yeni əlamətlərin yaradılması yolu ilə deyil, eyni zamanda daxili resurslar hesabına, yəni ədəbi dilə dialekt sözlərin cəlb olunması yolu ilə, dildə olan sözlərə fərqli bir mənə verməklə, əvvəllər “köhnəlməmiş” leksikanın geri qaytarılması hesabına, başqa dillərdən alınma sözlər, beynəlmiləl sözlər hesabına həyata keçirilir [1].

Dilə daxil olan yeni sözün hansı yolla gəlməsindən asılı olmayaraq, “vətəndaşlıq hüququnun” tanınması üçün həmin söz dil daşıyıcıları tərəfindən qəbul olunmalıdır, mənimsənilməlidir, başqa sözlə, həmin söz müəyyən kontekstdə müntəzəm istifadə olunmalıdır. Deməli, həmin termin seçim kriteriyalarına: terminin qrammatik xarakteri, onun dəqiqliyi, şəffaflığı, lakonikliyi, etibarlılığı, tələffüzü və nitqdə tətbiqinin asanlıığı, törəmələrini yaratmaq imkanı və s. uyğun olmalıdır.

Maraqlı faktıdır ki, dildə xarici sözlər daha sağlam yaşayır. Bu onunla əlaqədardır ki, dilin daşıyıcısı doğma dildə olan sözlərə “etibar etmir”, hesab edir ki, onlar müəyyən hadisənin, predmetin və s. mənasını verə bilmir. Bu da öz növbəsində dil daşıyıcısının dil kompetensiyasının aşağı səviyyəsinin göstəricisi ola bilər.

Dil siyasəti məsələlərinə dildə termin dövriyyəsinin nizamlanması daxildir. Bu məqsədlə hökumətlər yanında terminoloji komissiyalar yaradılır.

1995-ci ildə Azərbaycan dilinə dövlət statusu verildikdən sonra dilçiliyin bir qolu olan termin yaradıcılığı işlərində yeni mərhələ başladı [2]. Müxtəlif sahələr üzrə çoxlu sayda terminoloji lüğətlər hazırlandı. Terminlərin bir hissəsi əsasən daxili resurslar hesabına yaradılıb ki, onların da müəyyən hissəsi alınma sözlərdir. Alınma sözlər heç bir problem yaratmır, onlar müəyyən normaları aşmamalıdır. Bunun üçün dilin terminoloji inkişafını nizamlamaq zəruridir və daxili resurslar hesabına həyata keçirilən termin yaradıcılığı prosesinin dayanmasına icazə vermək olmaz. Xarici sözün doğma dildə yerini müəyyən edən zaman həmin sözün dünya dil məkanında vəziyyətini nəzərə almaq lazımdır. Dildə olan müəyyən sayda sözlərin bir çox digər dillərdə də olması onun öyrənilməsinə müəyyən dərəcədə asanlaşdırır.

Başqa dillərdə olduğu kimi, Azərbaycan dilində də terminologiyanın yaradılması və zənginləşdirilməsi üçün ən əlverişli yol dilin daxili imkanları əsasında terminlərin, termin-söz birləşmələrinin yaradılmasıdır. Belə ki, elm və texnikanın müxtəlif sahələrində meydana gələn yeni anlayışların ifadəsi üçün terminlər yaradılarkən ən əvvəl ana dilinin lüğət tərkibinə nəzər

salmaq, onun daxili imkanlarından səmərəli istifadə etmək lazımdır. Bu baxımdan uzun tarixi inkişaf yolu keçmiş Azərbaycan dilinin terminoloji sisteminin zənginləşməsində dildaxili imkanlar – leksik, morfoloji, sintaktik üsulla yaranan terminlər hər zaman üstünlük təşkil edirlər [3].

Cəmiyyətdə baş verən dəyişikliklər, yeniliklərlə bağlı həyatımıza daxil olan yeni anlayışları həm daxili imkanlar, həm də xarici resurslardan istifadə etməklə adlandırmaq zəruriyyətə çevrilib.

Terminologiyanın harmonikləşdirilməsi prinsipləri

Ənənəvi olaraq terminlər alim və mütəxəssislər (terminoqraflar) tərəfindən mətnlərdən seçilib, sistemləşdirilib və ekspertiza olunduqdan sonra kitab şəklində çap olunur. İnternet vasitəsi ilə informasiyanın sel kimi daxil olması bu işi çox çətinləşdirmişdir.

Digər tərəfdən, termin yaradıcılığı işi əvvəlki dövrə nəzərən xeyli asanlaşmışdır. Qeyd edək ki, [4]-də Azərbaycan terminoloji informasiya sisteminin konseptual modeli təklif edilmiş, sistemin yaratdığı imkanlar təhlil edilmişdir. Hər bir alim və mütəxəssis yeni termin yaradarkən bu sözün əvvəllər termin kimi qeydə alınıb-alınmaması haqqında məlumat əldə etmək üçün Milli terminoloji informasiya sisteminin (MTİS) Terminoloji reyestrinə müraciət edə bilər. Əgər axtarılan termin həmin reyestrə varsa, onda yeni termin qəbul etməyə ehtiyac qalmır. Əks halda bu termin MTİS ilə inteqrasiya olunmuş digər ölkələrə aid MTİS-də axtarılır, təhlil olunur və daha səmərəli qərar qəbul olunur, başqa sözlə, beynəlxalq harmonikləşdirmə həyata keçirilir.

Hazırkı dövrdə elm, mədəniyyət və iqtisadiyyat sahəsində beynəlxalq əməkdaşlığın gücləndirilməsi daha çox inkişaf etmiş milli dillərin terminologiyasının harmonikləşdirilməsi işlərinin inkişafını tələb edir. Terminologiyanın harmonikləşdirilməsi prinsiplərinin işlənilməsi terminoloqların beynəlxalq əməkdaşlığı sahəsində işlərin mühüm hissəsi hesab olunur.

Harmonikləşdirmənin tərkib hissəsi olan terminlərin planlı şəkildə beynəlmiləlləşməsi yerinə yetirilməlidir, başqa sözlə, terminlər arasında dəqiq uyğunluq müəyyən edilməli, həmçinin beynəlmiləl terminlərin sinonimlərinin içərisindən seçilərək, formaca yaxın olan müxtəlifdillə terminlərin mənalari razılaşıdırılmalıdır.

Terminologiyanın harmonikləşdirilməsi aşağıdakı mərhələlərdən ibarətdir:

1. Milli terminologiyaya və termin sistemlərinin sistemli müqayisəsi;
2. Müqayisə edilmiş milli terminologiyada əks olunan bütün anlayışlar nəzərə alınaraq onların klassifikasiyası sxeminin qurulması;
3. Ekvivalent milli terminlərin birmənalı anlaşılmasının müəyyən edilməsi və hazırlanması haqqında razılaşmanın işlənilməsi;
4. Beynəlmiləlləşmə.

Milli və beynəlmiləl anlayışlar və onları təmsil edən termin sistemlərinin harmonikləşdirilməsi müəyyən dillərdə standartlaşdırmanın vahid texniki dilinin hazırlanmasına istiqamətlənib. Termin sistemlərinin harmonikləşdirilməsi zamanı iki qrup faktorları nəzərə almaq lazımdır:

- linqvistik (terminin aid olduğu dilin xüsusiyyətləri ilə əlaqəli);
- ekstralingvistik (termini təsvir edən predmet sahəsi və nəzəriyyə ilə əlaqəli).

Termin sistemlərinin beynəlxalq harmonikləşdirilməsinin ekstralingvistik faktorları aşağıdakılardır:

- biliklərin inteqrasiyası;
- elm və texnologiyanın beynəlmiləlləşməsi;
- elm və texnologiyanın nəzəri və metodoloji birliyi;
- dünya sivilizasiyasının müasir mərhələsi üçün xarakterik olması.

Termin sistemlərinin beynəlxalq harmonikləşdirilməsinin linqvistik faktorlarına daxildir:

- dillərin xüsusi mədəsədlər üçün formalaşması;
- xüsusi mədəsədlər üçün dilin leksik vahidi funksiyasını daşıyan terminlərin beynəlmiləlləşməsi prosesləri.

Termin sistemlərinin beynəlxalq harmonikləşdirilməsi prinsipləri aşağıdakı hallarda

mümkündür:

- iki və ya daha çox ölkədə eyni bir bilik və ya fəaliyyət sahəsi inkişaf etdiyi halda bu mümkündür və məqsəduyğundur;
- əgər bu ölkələrdə verilən sahənin əsasında eyni və ya yaxın anlayışlar sistemi ilə eyni və ya yaxın nəzəriyyələr durursa, bu mümkündür.

Termin sistemlərinin harmonikləşdirilməsinin tətbiqi nəticələri aşağıdakılardır:

- beynəlxalq terminoloji lüğətlər və standartlar;
- çoxdilli informasiya-axtarış tezaurları.

Beynəlxalq Standartlaşdırma Təşkilatının ISO/TC37/SC1 texniki komitəsi tərəfindən ISO 860:2007 “Anlayış və terminlərin beynəlxalq harmonikləşdirilməsi” standartı işlənmişdir [5]. Bu standart konsepsiya, konseptual sistemlər, tərif və terminlərin harmonikləşdirilməsinə metodoloji yanaşmanı müəyyən edir. O, birdilli və ya çoxdilli kontekstdə milli və ya beynəlxalq səviyyədə harmonikləşdirilmiş terminologiyanın işlənilməsinə tətbiq edilir.

Anlayışların de-terminləşmə prosesi

Qeyd etmək lazımdır ki, İnternet vasitəsilə termin haqqında əlavə bilik (etimologiya, təsvirlər, şərtlər və s.) əldə etmək mümkündür. Müasir texnologiyalar terminləri yalnız mətn şəklində deyil, həm də digər multimedia təsvirlərinin (foto, audio, video) köməyi ilə də təqdim etməyə imkan verir.

İnsanların informasiyaya asan çıxışı və sosial media vasitəsi ilə yayımı biliklərin çox böyük sürətlə populyarlaşmasına şərait yaradır. Ona görə də nisbətən dar çərçivədə istifadə olunan anlayışlar (terminlər) da sürətlə de-terminləşir (terminləşmənin əksi).

Hələ elmi-texniki inqilab dövründə “informasiya-kommunikasiya proseslərinin aktivləşməsinin” nəticəsi olaraq müxtəlif bilik sahələrində terminlərin ciddi artımı və onların ümumişlək leksikaya daxil olması haqqında danışılırdı. Həmin dövrlər bu proses “terminoloji partlayış” adını almışdı. Hazırkı qloballaşma dövründə “özünəməxsus terminoloji ekspansiya” haqqında danışmaq olar ki, bunu da bir sıra linqvistlər “leksikanın intellektuallaşması” tendensiyası kimi müəyyən edirlər, bu da terminologiyanın kommunikativ rolunun yüksəlməsi ilə əlaqəlidir [6].

Müasir dünyada daha çox və tez-tez xüsusi leksika “vahid semantik sahədə gündəlik olaraq” istifadə olunur və “daimi de-terminləşmə prosesinə uğrayır”.

De-terminləşmə anlayışının özünün anlaşılması məsələsi bu günədək birqismətli həllini tapmayıb. Terminin terminlər sistemindən çıxaraq ümumişlək sözlə çevrilməsi ilə əlaqədar xüsusi mənasını itirməsi de-terminləşmədir. İnkişaf etmiş milli dildə de-terminləşmə dilin nominativ sisteminin zənginləşməsinin əsas yollarından biridir, xüsusi və ümumişlək leksikanın qarşılıqlı fəaliyyətini əks etdirir. De-terminləşmə terminin dil substrantının (alt qatının) semantik və funksional variasiyasını əks etdirir. Funksional de-terminləşmə terminin terminoloji funksiyasını itirməsidir. Semantik de-terminləşmə terminin terminoloji mahiyyətini itirməsi və terminin yeni semantikasının inkişafında meydana çıxır.

Ümumişlək leksika daimi olaraq terminlərlə zənginləşdirilir, bu isə digər dillərdə olduğu kimi, Azərbaycan dilində də uyğun anlayışların aktuallığı ilə izah olunur. De-terminləşmə prosesi kommunikativ proseslərin intellektuallaşması üsulu kimi, dünyanın informasiya şəklinin tanınması və zənginləşməsinin anlaşılmasını şərtləndirir.

Terminlərin ictimai sferada işlənmə tezliyini avtomatik olaraq müəyyən etməklə de-terminləşmə prosesini analiz etmək və qiymətləndirmək olar. Qeyd etmək lazımdır ki, anlayışların terminləşmə və de-terminləşmə göstəriciləri onların istifadə edildiyi dilin yayılma arealından, başqa sözlə, ölkə əhalisinin sayından asılıdır.

Terminləşmə – adi, hər gün işlənən sözün terminə çevrilməsi, tərs proses – terminin ümumişlək dilə daxil olması prosesidir. Bu proses 2000-ci ildə İnqrid Meyer və Kristen Makintoş tərəfindən terminologiyaya aid işlərində analiz olunub və de-terminləşmə adlandırılıb [7].

Onlar de-terminləşmənin iki kateqoriyasını təsvir edirlər:

1. Termin gerçək mənasını saxlayır, qoruyur, lakin predmetin aid olduğu sahənin ekspertləri tərəfindən istifadə olunmur. Belə halda mümkündür ki, predmet populyarlaşmağa başlayıb, insanlar termini istifadə etmək üçün anlayışı kifayət qədər yaxşı anlayırlar. Nümunə olaraq, xəstəliklərə aid tibbi terminləri göstərmək olar, belə ki, bu terminlər kifayət qədər geniş yayılıb və onlar haqqında hamının məlumatı var.

2. Söz hal-hazırda tamamilə digər konsepsiyanı təsvir edir.

Terminologiya daimi hərəkətdə olan, dinamik bir sistemdir. Belə ki, bu sistemdə terminlərin, anlayışların, təriflərin həyat dövrü başa çatır. Onlar istifadədən çıxır, köhnəlmiş və qeyri-elmi anlayış kimi ləğv edirlər.

Digər dillərdə olduğu kimi, Azərbaycan dilində də terminoloji sistemlərdən leksik vahidlərin alınmasının bir neçə dövrü (məsələn, elmi-texniki inqilab əsri və s.) seçilirdi. Hal-hazırda yeni era – kompüter texnologiyaları və İnternet erası kompüter jarqonu və texniki terminlərdən ibarət xüsusi kompüter altdilinin yaradılmasına gətirib çıxarıb və kifayət qədər zəngin terminoloji sistemi təmsil edir. Kompüter altdilinin bu və ya digər hissəsi kütləvi informasiya vasitələrində və İnternetdə aktiv təmsil olunur. Nəticədə terminlər geniş kontekstdə istifadə edilməyə başlayır.

Kompüter və İnternet şəbəkəsi istifadəçilərinin sayının artması nəticəsində terminoloji leksika yalnız mütəxəssislərə deyil, adi istifadəçilərə də məlum olur. Lakin adi istifadəçilər həmin terminlərin mənalari haqqında yalnız təxmini təsəvvürə malikdirlər.

Beləliklə, insanın müasir həyatında İnternet və kompüter texnikasının populyarlaşması gündəlik dil ünsiyyəti və professional dil ünsiyyətçilərinin sıx əlaqəsinə səbəb olur: yeni termin və anlayışlar yalnız bizim nitqimizi deyil, fikir və düşüncələrimizi də dəyişir və tamamlayır. Bütünlüklə, insan cəmiyyətində yalnız dünya və insanın özü haqqında ümumi və xüsusi biliyin ötürülməsi üçün deyil, eyni zamanda uzunmüddətli perspektivlərə hesablanan xüsusi estetik konstruksiyaların formalaşması üçün nəzərdə tutulan mürəkkəb sistemlər yaradılır [8].

Yuxarıda deyildiyi kimi, hazırda bir çox terminlər de-terminləşməyə uğrayırlar və ya yeni məna alırlar. Onların semantikasi genişlənir və ya məhdudlaşır, onlar məcazi mənada işlənir. Terminlərin semantik strukturunun genişlənməsi tendensiyası onların bir sferadan digərinə nüfuz etməsi prosesində meydana çıxır. Bu zaman ilkin məna və dəyəri atıla da bilər, saxlanıla da bilər. Linqvistik ədəbiyyatda bu cür sözlər neologizmlər – əvvəllər məlum olan, lakin yeni işlənmə dərəcəsi alan sözlər kimi klassifikasiya olunurlar.

De-terminləşmə prosesi terminin elmi dəqiqliyini itirməsinə, eyni zamanda onun reallaşdırılması sferasının genişlənməsinə səbəb olur. Qeyd etmək lazımdır ki, elmin və texnikanın sürətli inkişafı, tədqiqat prosesləri və nəticələrinin kütləvi informasiya vasitələrində aktiv əks olunması bu prosesə gətirib çıxarır. Bir çox tədqiqatçıların fikirlərinə görə, məhz bu proses ayrı-ayrı terminlərin xüsusi leksikadan ümumişlək leksikaya keçməsinə səbəb olur.

M.İ.Fominanın fikirlərinə görə, tədqiqatçıların de-terminləşməyə yüksək diqqətinin əsas səbəbləri aşağıdakılardır [9]:

- cəmiyyətin ümumi maarifləndirilməsi (informasiya təminatı);
- elm və texnologiyanın sürətli inkişafı;
- təhsilin səviyyəsinin yüksəlməsi;
- xüsusi sözün ümumi mənasının aydınlığı və şəffaflığı.

Beləliklə, de-terminləşmə prosesi aşağıdakı kimi təqdim oluna bilər:

- ilkin olaraq termin mütəxəssislər tərəfindən xüsusi kontekstdə istifadə olunur;
- sonradan termin qeyri-mütəxəssislər tərəfindən istifadə olunmağa başlayır, lakin xüsusi kontekstdə;
- terminin istifadəsi xüsusi kontekstdən çıxır.

Qeyd etmək lazımdır ki, birinci mərhələ yalnız xüsusi elmi diskurs üçün xarakterikdir. Lakin ikinci və üçüncü mərhələ elmi-populyar diskursda de-terminləşmə prosesinin bir çox xüsusiyyətlərini nümayiş etdirir.

Terminlərin gizli əlaməti onları dilin digər vahidlərindən fərqləndirməyə və bütün terminlər çoxluğunu ayırmağa imkan verir. Ümumi anlayışlar terminlərin bu gizli əlaməti ilə təyin olunurlar. Ümumi anlayışların bir neçə tipi mövcuddur, ona görə də terminlərin müxtəlif tipləri meydana çıxıb bilər.

İnsan biliyinin inkişafının hər mərhələsində, hər bir əsrdə müəyyən sayda ümumelmi və ümumtexniki anlayışlar yer alır. Onlara fəlsəfə, sistemlərin ümumi nəzəriyyəsi, kibernetika, informatika və digər metodoloji elmlərin də ümumi anlayışları qoşulur. Bu elmlərin anlayışlarının bəzi hissəsi ümumelmi anlayışlar kimi müxtəlif bilik sahələrində istifadə oluna bilər. Nəzərə almaq lazımdır ki, ümumelmi (ümumtexniki) və sahələrarası anlayışlar ümumi məzmunla malikdirlər, bu səbəbdən müxtəlif sahələrdə tətbiq oluna bilərlər [10].

Beləliklə, kommunikativ proseslərin intellektuallaşdırılması üsullarından biri olan determinləşmə prosesi reallığın anlaşılması, idrakın inkişafı və dünyanın informasiya şəklində zənginləşməsinə kömək edir.

Terminin çəki əmsalının hesablanması metodları

Ayrı-ayrı terminlərin işlənmə tezliyi elmin səviyyəsini və müəyyən mərhələdə ictimai strukturun inkişafını əks etdirir.

İstifadəçinin informasiya tələbatını ifadə edən sorğu ayrı-ayrı terminlərdən ibarət olur. Axtarış alqoritmi yerinə yetirilən zaman sorğunun terminlərinin müqayisəsi baş verir və onların yaxınlıq dərəcəsi, başqa sözlə, relevantlıq dərəcəsi müəyyənləşdirilir.

Terminin çəkisi sənəddə nə qədər çoxdursa, həmin sənəd o qədər relevantdır və axtarışın nəticələrinin siyahısında bu sənəd o qədər yüksək mövqə tutur.

Beləliklə, terminlərin ölçülməsi axtarış sisteminin əsas məsələsi – istifadəçinin relevant sənədlə təmin olunması üçün zəruridir.

Sadə və daha çox yayılan axtarış modeli olan bul modeli terminlərin ölçülməsi üçün ikili say sistemini istifadə edir [11]. Bu metod indeksləşmiş terminlərin seçimi mərhələsində həyata keçir və onunla nəticələnir ki, axtarışa daxil olan terminə 1-ə bərabər çəki verilir, digər terminlərə 0 çəki verilir. Beləliklə, axtarışa daxil olan terminlərin hamısı eynimənalı sayılır.

Bul modelinin çatışmazlıqları var. Belə ki, mütləq vahid çəkilərin istifadəsi istifadəçi üçün bir sorğuya cavab olaraq bir çox nizamlanmamış sənədləri verir, axtarışın nəticələrinin qəbulunda əhəmiyyətli çətinliklərə yol açır. Bu dəstdən müvafiq sənədlərin seçilməsi həqiqətən əhəmiyyətli çətinliklər yaradır. Vəziyyətdən çıxış yolu terminlərə fərqli çəkilərin verilməsidir. Eyni bir sənədin axtarışına daxil olan terminlər müxtəlif çəkiyə malik ola bilərlər. Eyni zamanda, eyni bir terminin çəkisi müxtəlif sənədlərdə müxtəlif ola bilər.

Bul metodundan başqa, terminlərin çəkisinin qiymətləndirilməsi üçün əsasən aşağıdakı metodlardan istifadə olunur:

- statistik tezlik modeli;
- ehtimal modeli;
- semantik analiz modeli.

Hər hansı bir sənəd üçün terminin vacibliyinin qiymətləndirilməsinin sadə üsullarından biri *TF – IDF* (ing. *TF – term frequency, IDF – inverse document frequency*) – statistik tezlik metodunun istifadəsidir. Terminlərin ağırlıq çəkisinin statistik tezlik modeli indeksləşmə tezlik metodu ilə sıx bağlıdır [11]. Prinsip ondan ibarətdir ki, əgər söz hər hansı bir sənəddə tez-tez rast gəlinirsə, bununla yanaşı digər sənədlər toplusunda hərdən rast gəlinirsə, bu söz həmin sənəd üçün böyük məna kəsb edir. Sözü çəkisi həmin sözün sənəddə istifadəsinin sayı ilə mütənasibdir, digər sənədlər kolleksiyasında həmin sözün istifadə tezliyinə tərs mütənasibdir. *TF – IDF* metodu daha çox mətnin analizi və informasiya axtarışında istifadə olunur.

TF - terminin tezliyidir, terminin sənəddə nə qədər tez-tez rast gəlindiyini ölçür. Bir sənəd çərçivəsində t_i sözünün əhəmiyyəti qiymətləndirilir. Məntiqi olaraq təsəvvür etmək olar ki, termin

uzun sənədlərdə qısa sənədlərə nisbətən daha çox sayda rast gəlinə bilər, bu səbəbdən mütləq say axtarılmır. Mətdə lazımı termin istifadə olunan zaman nisbi say tətbiq olunur. Başqa sözlə,

$$tf(t, d) = \frac{n_i}{\sum_k n_k} \quad (1)$$

$n_i - t_i$ termininin d_i sənədinə daxil olmalarının sayı, $n_k - d_i$ sənədində olan sözlərin ümumi sayıdır. Buradan görünür ki, sənəddə daha çox rast gəlinən istənilən söz bu sənədin məzmununu ən dəqiq əks etdirən vacib termin kimi qəbul oluna bilər.

Terminlərin ağırlıq çəkisinin birinci forması Hans Peter Lun (1957) tərəfindən verilib. Sənəddə rast gəlinən terminin çəkisi onun tezliyi ilə mütənasibdir [12].

TF sayılan zaman, bütün terminlər əhəmiyyətinə görə bir-birinə bərabər sayılır. Lakin praktika göstərir ki, sənəddə ən böyük rastgəlinmə tezliyinə malik olan sözlər heç də həmişə bu sənəd üçün vacib termin olmurlar. Məlumdur ki, xidməti sözlər, bağlayıcılar, əvəzlilər və s. daha tez-tez rast gəlinir, amma praktiki olaraq mətnin mənasına təsir etmir.

Sənədlərin məzmunlarının Zipf qanununa uyğun olaraq təhlilinin nəticələrinə əsasən, demək olar ki, əgər sənəddə hər hansı söz böyük rastgəlinmə tezliyinə malikdirsə, lakin termin deyilsə, onda bu söz korpusun digər bir çox sənədlərində də böyük tezliklə rast gəlinməlidir [13]. Məhz bu xassə sənədlər üçün vacib terminlərin seçilməsi prosesində istifadə oluna bilər. Ona görə də ikinci mərhələdə informasiya korpusunun bütün sənədləri təhlil olunur və terminlərin korpusun digər sənədlərində rastgəlinmə xarakteristikaları nəzərə alınmaqla, onların həmin sənəd üçün çəki əmsalı müəyyən edilir. Bu kəmiyyət sənədlərin ümumi sayının terminin rast gəlinədiyi sənədlərin sayına nisbətinin loqarifması kimi təyin olunur və əks sənəd tezliyi adlanır.

Qeyd edək ki, “əks sənəd tezliyi” (*ing. inverse document frequency, IDF*) adı altında terminoloji spesifikliyin statistik interpretasiyasını Karen Spark Cons (*ing. Karen Spärck Jones, 1972*) fikirləşib [14]. *IDF* birbaşa terminin vacibliyini ölçür və aşağıdakı kimi işarə olunur:

$$idf(t, D) = \log \frac{|D|}{|\{d_i \in D / t \in d_i\}|} \quad (2)$$

$|D|$ – korpusdakı sənədlərin sayı; $\{d_i \in D / t \in d_i\}$ D korpusunda olan t termininin rast gəlinədiyi sənədlərin sayıdır ($n_i \neq 0$).

Terminin sənəddəki tezliyi və əks sənəd tezliyi kəmiyyətlərini tezlik üzrə vahid indeksləşdirmə modeli çərçivəsində birləşdirmək olar.

$$w_{ij} = tf(t, d) \cdot idf(t, D) \quad (3)$$

(1) və (2) ifadələrini (3) düsturunda yerinə qoymaqla, informasiya fəzasının bütün digər sənədlərinin xarakteristikalarını nəzərə almaqla, t_i termininin d_i sənədindəki w_{ij} çəki əmsalının hesablanması üçün aşağıdakı düstur alınır:

$$w_{ij} = \frac{n_i}{\sum_k n_k} \cdot \log \frac{|D|}{|\{d_i \in D / t \in d_i\}|} \quad (4)$$

(4) düsturundan görünür ki, t_i termininin d_i sənədində rast gəlinmə tezliyi nə qədər böyük, eləcə də bu terminin rast gəlinədiyi sənədlərin sayı nə qədər az olarsa, onda t_i termini d_i sənədində bir o qədər böyük çəkiyə malik olar. Başqa sözlə, əgər d_i sənədi t_i termininin cəmləşdiyi sənəd olarsa, onda t_i termini bu sənəd üçün əhəmiyyət kəsb edən termin rolunu oynayır.

Qeyd edək ki, *TF – IDF* metodu terminin əhəmiyyətinin ölçülməsi üçün kifayət qədər universal metrikadır. Bu metoda əsaslanan müxtəlif formullar vardır. Onlar əmsallar, normallaşdırma, loqarifmik miqyasın istifadəsilə fərqlənirlər. Ən populyar formullardan biri *BM 25* formuludur [15].

Real informasiya ehtiyacları və terminlər arasında uyğunluğun qurulması üçün terminlərin çəkisinin qiymətləndirilməsinin ehtimal modeli işlənmişdir. Ehtimal modeli verilən sənədin verilən sorğuya relevant olması ehtimalının dəqiq qiymətləndirilməsinə əsaslanır [16].

Eyni bir termin tamamilə müxtəlif anlayışları ifadə edə bilər. Uyğun olaraq, bəzi sənəddə bu və ya digər terminin mövcudluğu sənədin sorğuya relevant olduğunu göstərmir. Təsvir olunan problemləri latent semantik indeksləşmə (LSİ) həll edir [17]. Bu yanaşmanın məğzi ondan ibarətdir ki, hər bir sənəd açıq, latent semantik struktura malikdir. Belə strukturun analizi (LSİ) hər sənədi hər hansı terminin mövcudluğu və ya olmaması, onun semantik mənası nöqtəyindən təsvir etməyə imkan verir, məsələn, sənəd onun tərkibinə daxil olmayan terminlə adekvat təsvir oluna bilər və əksinə, bəzi terminlər sənədin məğzini əks etdirmir, onların sorğunun terminləri ilə üst-üstə düşməsi sənədi relevant etmir.

Beləliklə, latent faktorların say analizinin nəticəsində terminin gücü korrektə oluna bilər və sənədin axtarışı onun məzmununa daha adekvat ola bilər. LSİ riyazi olaraq xətti cəbrin metodlarından biri olan matrisin sinqulyar ayrılması metodu ilə həyata keçirilir [18]. Müasir alqoritmlər ehtimal nəzəriyyəsi aparatından da istifadə edirlər.

Qeyd edək ki, Azərbaycanda termin yaradıcılığı işlərinin avtomatlaşdırılması məsələləri ilə bağlı müəyyən işlər görülmüşdür.

Keçən əsrin 80-ci illərində Ukrayna Elmlər Akademiyasının Kibernetika İnstitutunda Azərbaycan Elmlər Akademiyasının Kibernetika İnstitutu (indiki AMEA İdarəetmə Sistemləri İnstitutu) ilə birlikdə terminoloji leksikanın avtomatlaşdırılmış analizi sistemi (TLAAS) işlənmişdir [19]. TLAAS leksikanın bir sıra parametrlərə görə təsnifatı, bəzi leksik qruplaşmaların formalaşması, sözün və lüğətin sistem parametrlərinin hesablanması kimi funksiyaları yerinə yetirir. Sistem tərəfindən hazırlanan məlumatlar leksikanın nəzəri tədqiqatlarında və eyni zamanda terminoloji lüğətlərin korreksiyasında tətbiq olunur.

Azərbaycanda MTİS-in konsepsiyası çərçivəsində Milli Terminoloji veb-portal [20] yaradılmışdır ki, bu da termin yaradıcılığı işlərinin avtomatlaşdırılması istiqamətində mühüm addımdır.

Nəticə

Tədqiqat işində aşağıdakı nəticələrə gəlinmişdir:

- terminləşmə və de-terminləşmə, terminlərin harmonikləşdirilməsi prinsipləri və çəkisinin qiymətləndirilməsi metodlarının təhlili göstərdi ki, termin yaradıcılığı özündə dinamik hadisələri əks etdirir. Ona görə də bu sahənin mütəxəssisləri dilin funksional inkişafını dəstəkləmək zəruriyyətini nəzərə almaqla, cəmiyyətdəki yeniliyə diqqət etməlidirlər. Belə ki, cəmiyyətdəki yeniliyin adının öz ana dilində müəyyən olunması çox vacibdir;
- Azərbaycanda anlayış və terminlərin harmonikləşdirilməsi üzrə milli standartın işlənməsinə ehtiyac yaranmışdır ki, bu da terminlərin planlı şəkildə beynəlmilləşməsi, aralarında dəqiq uyğunluğun müəyyən edilməsinə səbəb olacaqdır;
- müəyyən olunmuşdur ki, hər hansı bilik sahəsi üzrə terminlərdən istifadə tezliyinin dinamikasına görə həmin sahənin populyarlaşması səviyyəsini müəyyənləşdirmək olar;
- terminin əhəmiyyətinin müəyyən olunması üçün istifadə olunan statistik tezlik modeli açar sözlərin çıxarılması üçün istifadə olunan məşhur alqoritmlərdən biridir. Bəzi daha yeni metodlar bu modeli hesablama prosesinin hissəsi kimi istifadə edir. Praktiki olaraq korpusdan istifadə edən bütün metodlar çəki funksiyasından asılıdırlar. Bu funksiya bütün korpus üçün oxşar qaydalarla konkret məndə termin və ya ifadənin bəzi xüsusiyyətlərini balanslaşdırır;

- Azərbaycanca müxtəlif sahələr üzrə hazırlanmış terminoloji lüğətlərin vahid informasiya sistemində toplanması nəticəsində termin yaradıcılığının səmərəliliyi yüksələcəkdir.

Ədəbiyyat

1. Гарипова Ф.Х. Практические аспекты языковой политики: терминологическая работа // Вестник ВЭГУ, № 3(41), 2009, с.34–39.
2. Azərbaycan Respublikasının Konstitusiyası, Maddə 21. Dövlət dili, 12 noyabr 1995-ci il. www.azerbaijan.az/portal/General/Constitution/doc/constitution_a.pdf
3. Sadıqova S. Müasir Azərbaycan ədəbi dilində termin yaradıcılığı prosesi, Bakı: “Elm”, 2010, 244 s.
4. Əliquliyev R.M., Qurbanova Ə.M. Azərbaycanca terminoloji informasiya sisteminin yaradılmasının konseptual əsasları // İnformasiya cəmiyyəti problemləri, Bakı, 2011, №1, s.3–8.
5. ISO 860:2007 (en), Terminology work – Harmonization of concepts and terms. www.iso.org/obp/ui/#iso:std:iso:860:ed-3:v1:en
6. Бычкова О.Н. Детерминологизация как функционально-коммуникативная трансформация термина // Научный Альманах: Язык. Текст. Дискурс, Ставрополь: Изд-во СГПИ, 2011, выпуск 9, с.458–462.
7. Meyer I., Mackintosh K. L’étirement du sens terminologique: aperçu du phénomène de la déterminologisation // Le Sens en Terminologie, 2000, pp.198–216.
8. Глазырина А.И., Антонова А.В. Детерминологизация в русском компьютерном подязыке: данные компонентного анализа // ИИЯ УрГПУ, г. Екатеринбург, Россия, 2013, с.20–26
9. Фомина М.И. Современный русский язык. Лексикология, 4-е изд., испр., М.: Высшая школа, 2001, 415 с.
10. Зубкова А.А. Термин как особенная лексическая единица (на материале предметной области «Логистика») / Актуальные вопросы филологической науки XXI века : сб. статей V Междунар. науч. конф. молодых ученых, Екатеринбург, Уральский федеральный университет, 2016, с.107–112.
11. Солтон Дж. Динамические библиотечно-информационные системы, Москва, Издательство «Мир», 1979, 559 с.
12. Luhn Hans Peter. A Statistical Approach to Mechanized Encoding and Searching of Literary Information // IBM Journal of Research and Development, 1957, vol.1, no.4, pp.309–317.
13. Zipf G.K. Human behavior and the principle of least effort: an introduction to human ecology. Cambridge, Addison-Wesley Press, 1949, 573 p.
14. Karen S. J. A Statistical Interpretation of Term Specificity and its Application in Retrieval // Journal of Documentation, 1972, vol.28, no.1, pp.11–21.
15. Zaragoza H., Craswell N., Taylor M., Saria S., Robertson S. Microsoft Cambridge at TREC-13: Web and HARD tracks / The Thirteenth Text Retrieval Conference (TREC-2004). Gaithersburg, Maryland, 2004, NIST Special Publication, pp.1–7.
16. Sparck J.K. A Probabilistic Model of Information Retrieval: Development and Comparative Experiments // Information Processing and Management, 2000, № 36(6), pp.809–840.
17. Некрестьянов И.С., Пантелева Н. Системы текстового поиска для Веб // Программирование, 2002, №4, с.33–57.
18. Indexing by Latent Semantic Analysis / S. Deerwester [and others] // Journal of the American Society for Information Science, 1990, №41(6), pp.391–407.
19. Мамедова М.Г., Скороходько Э.Ф. Автоматизированная система анализа терминологической лексики, г. Москва, ВНИИТИ, Научно-техническая информация, 1981, серия 2, №1, с.14–18.
20. Azərbaycan Respublikasının Nazirlər Kabineti yanında Terminologiya Komissiyası. www.terminologiya.az

УДК 004.82 (004.650:06.72:80)

Гурбанова Афруз М.

Институт Информационных Технологий НАНА, Баку, Азербайджан

afruz1961@gmail.com

Анализ методов автоматизации терминотворческой деятельности

В статье рассматриваются вопросы терминотворчества, являющегося результатом сознательного влияния общества на развитие языка. Было отмечено, что словарный состав любого языка обновляется за счет как внутренних, так и внешних ресурсов, а жизненный цикл слова зависит от активности его использования. Были определены принципы гармонизации терминологии. Был изучен процесс детерминологизации понятий, определены причины этого процесса. Отмечена роль специальных и общеупотребительных слов лексикона в процессе детерминологизации. Исследованы и проанализированы методы расчета частоты встречаемости и весовых коэффициентов терминов в текстах.

Ключевые слова: языковая политика, терминотворчество, гармонизация терминов, детерминологизация, весовой коэффициент термина.

Afruz M. Gurbanova

Institute of Information Technology of ANAS, Baku, Azerbaijan

afruz1961@gmail.com

Analysis of automation methods of the term creation activity

The article deals with the issues of terminology, which is a consequence of conscious influence of society on the development of language. It is highlighted that vocabulary of any language is updated due to both internal and external resources, and a life cycle of a word depends on its usage. The principles of harmonization of terminology are defined. The process of determinologization of concepts is studied and the reasons for this process are determined. The role of special and common words of the lexicon in the determinologization process is noted. The methods of calculating the frequency and weight coefficients of terms in texts are studied and analyzed.

Keywords: language policy, terminology creation, harmonization of terms, determinologization, weight factor of terms.