



**ÜMUMMİLLİ LİDER HEYDƏR ƏLİYEVİN
100 İLLİK YUBİLEYİNƏ VƏ BEYNƏLXALQ
ANA DİLİ GÜNÜNƏ HƏSR OLUNMUŞ
“İKT-NİN AZƏRBAYCAN DİLİ PROBLEMLƏRİ,
AZƏRBAYCAN DİLİNİN İKT PROBLEMLƏRİ”
RESPUBLİKA ELMİ-PRAKTİKİ KONFRANSI**

Ə S Ə R L Ə R

21-22 fevral 2023-cü il

PROQRAM KOMİTƏSİ

SƏDR

akademik

İsa Həbibbəyli

HƏMSƏDRLƏR

akademik

Rasim Əliquliyev

akademik

Əli Abbasov

professor

Nadir Məmmədli

PROQRAM KOMİTƏSİNİN ÜZVLƏRİ

Tex.f.d. **Rəşid Ələkbərov**

AMEA-nın m.ü. **Ramiz Alıquliyev**

AMEA-nın m.ü. **Əminəğa Sadıqov**

Fil.e.d. **Məsud Mahmudov**

Fil.f.d. **Mehman Həsənli**

Fil.f.d. **Baba Məhərrəmli**

Fil.f.d. **Hikmət Quliyev**

AMEA-nın m.ü. **Kamil Aydazadə**

AMEA-nın m.ü. **Məsümə Məmmədova**

Tex.f.d. **Abzətdin Adamov**

Tex.e.d. **Zərifə Cəbraylova**

İ.e.d. **Əlövsət Əliyev**

Tex.f.d. **Samir Rüstəmov**

TƏŞKİLAT KOMİTƏSİ

Şakir Mehdiyev

Nərmin Adıgözəlova

Tex.f.d. **Gülzarə Nəbibəyova**

Əfruz Qurbanova

Tex.f.d. **Babək Nəbiyev**

Rasim Mahmudov

Anar Səmidov

Dinara Zeynalova

Könül Daşdəmirova

Gülzar Paşayeva

ISBN 978-9952-530-25-4

MÜNDƏRİCAT

1.	Rasim Əliquliyev, Ramiz Alıquliyev, Fərhad Yusifov. Elektron dövlət mühitində Azərbaycan dili ekosisteminin formalaşdırılmasının konseptual əsasları.....	5
2.	Rəşid Ələkbərov, Məmməd Həşimov. Bulud sistemləri əsasında Azərbaycan dili xidmətlərinin təşkili və intellektual analizi məsələləri.....	16
3.	Kamil Aida-zadə, Rustam Azimov. Analysis of texts features for the author recognition system.....	22
4.	Ümid Süleymanov, Samir Rüstəmov, Abzətdin Adamov. Azərbaycan dili üçün söz vektorları və böyük dil modelləri	26
5.	Rasim Alquliyev, Lyudmila Sukhostat. Issues of speech technologies application in the Azerbaijani language.....	31
6.	Ələvsət Əliyev, Roza Şahverdiyeva. Rəqəmsal iqtisadiyyat və transformasiyalar üzrə termin və anlayışlar bazasının formalaşdırılması problemləri.....	36
7.	Məsud Mahmudov. Azərbaycan dilinin milli korpusu süni intellekt problemləri kontekstində.....	44
8.	Ramiz Şıxəliyev, Fərqanə Abdullayeva. Kibertəhlükəsizlik problemlərinin həllində NLP effektiv vasitə kimi.....	50
9.	İsmayıl Sadıqov. Azərbaycan dilində mətnlərin oxunabilirlik səviyyəsinin qiymətləndirilməsi metodları.....	55
10.	Alakbar Valizada. Development of a real-time speech recognition system for the Azerbaijani language.....	61
11.	Hikmət Quliyev. Texnoloji proseslərin təsiri ilə dildə transformasiyalar: sosiopsixoloji faktorlar, folklorik effektlər və linqvisik təzahürlər.....	69
12.	Gulnara Nabibayova. Issues of linguistic demography formation on the national e-government platform.....	77
13.	Nəzakət Qazıyeva. Nitq texnologiyalarının eksperimental-fonetik aspektləri.....	82
14.	Yedgar Cəfərov. Milli onomastika sisteminin formalaşması aspektləri haqqında.....	89
15.	Rəna Mahmudova. Milli dil korpusunda Azərbaycan dilinin tədqiqi məsələləri.....	95
16.	Əfruz Qurbanova, Mehriban Bağirova. Dialektometriya və Azərbaycan dili: problemlər, həlli yolları və perspektivlər.....	99
17.	Rəsmiyyə Mahmudova, Kəmalə Qurbanova. Azərbaycanda jest dili və onun avtomatik tanınması problemləri.....	105

18.	Fərqanə Abdullayeva, Sabirə Ocaqverdiyeva. İnternet mühitində Azərbaycan dilində vulqar sözlər və onlarla mübarizə texnologiyaları.....	113
19.	Rasim Mahmudov. Dil sənayesi biliklər iqtisadiyyatının tərkib hissəsi kimi.....	119
20.	Günay İskəndərli. Sentiment analiz texnologiyaları və Azərbaycan dili.....	128
21.	Nigar Vəlizadə. Bədi mətnin linqvostatistik təhlili.....	134
22.	İradə Ələkbərova. Viki texnologiyaların milli dilin qorunmasında rolu və vikipediyanın Azərbaycan dili seqmentində problemlər.....	137
23.	Kəmalə Haşımova. İnternet mühitində azərbaycandilli reklam resurslarının formalaşması xüsusiyyətləri haqqında.....	145
24.	Səbinə Məmməd zadə. Milli transliterasiya sisteminin formalaşması məsələləri.....	151
25.	Maya Heydərova. Linqvistik verilənlər bazalarının quruluş xüsusiyyətləri.....	156
26.	Tamilla Bayramova. Proqram təminatı vasitələrinin lokallaşması problemləri haqqında.....	163
27.	Rafig İsmayılov, Kəmalə Qurbanova. Orfoqrafiya lüğətinin internet resurslarına inteqrasiyası və mətnlərin avtomatik düzəlişi məsələləri.....	167
28.	Sayal Quliyeva. Beynəlxalq təcrübə əsasında Azərbaycan dilinin tədrisi və qiymətləndirilməsi məsələləri.....	172
29.	Könül Həbibova. Həştəqlər xüsusi linqvistik hadisə kimi.....	180
30.	Bahar İsmayılova. Azərbaycan dilinin tədrisi prosesində intellektual oyunların tətbiqi məsələləri.....	186
31.	Aytən Qurbanova. Ümumtürk ədəbiyyatının təbliği və öyrənilməsində İKT-nin rolu.....	191
32.	Səyyar Abdullayev, Südabə Abasova. İKT terminlərinin azərbaycan dilinə tərcüməsi problemləri.....	194
33.	Nərmin Əliyeva, Rasim Heydarov. Korpus dilçiliyinin təşkilində leksikoqrafiyanın rolu və vəzifələri.....	200
34.	Sünbül Zalova. Jurnalistikada mətnlərin intellektual linqvistik təminatı problemləri.....	206
35.	Mehdi Rasul. Comparison of LR, NB and SVM Techniques on Sentiment Analysis of Azerbaijani Texts.....	211

Elektron dövlət mühitində Azərbaycan dili ekosisteminin formalaşdırılmasının konseptual əsasları

Rasim Əliquliyev¹, Ramiz Alıquliyev², Fərhad Yusifov³

^{1,2,3}İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

¹r.alguliev@gmail.com ²r.aliguliyev@gmail.com, ³farhadysifov@gmail.com

Xülasə — Məqalədə elektron dövlət mühitində Azərbaycan dili ekosisteminin formalaşdırılmasının konseptual əsasları tədqiq olunur. İnformasiya texnologiyalarının təsiri ilə sürətlə qloballaşan müasir dünyada dil müxtəlifliyinin qorunması aktual məsələlərdən birinə çevrilmişdir. Bu baxımdan e-dövlət platformasında Azərbaycan dilinin qorunması məsələləri mühüm tədqiqat istiqamətlərindən biridir. Milli dillərin e-dövlətin tərkib hissəsinə çevrilməsi və bu platformanın bütün seqmentləri üzrə təbii inkişafının təmin olunması üçün intellektual metodlar, alqoritmlər işlənməli və tətbiqlər hazırlanmalıdır. Məqalədə qloballaşmanın linqvomədəni fəzaya və dil proseslərinə təsirinə dair yanaşmalar araşdırılmışdır. Rəqəmsal əsrdə dünya xalqlarının dillərinin vəziyyəti və eləcə də, onlayn mühitdə Azərbaycan dilinin problemləri analiz olunmuş, təkliflər verilmişdir. E-dövlətdə dil müxtəlifliyinin qorunması və linqvistik texnologiyaların tətbiqi imkanları tədqiq olunmuşdur. Onlayn platformada Azərbaycan dili ekosisteminin yaradılmasının konseptual modeli təklif olunmuşdur. E-dövlət platformasında Azərbaycan dili ekosisteminin formalaşdırılmasına dair yanaşmalar verilmiş və linqvistik texnologiyaların tətbiqi

ilə həlli nəzərdə tutulan kompleks məsələlər göstərilmişdir.

Açar sözlər — e-dövlət; dil müxtəlifliyi; linqvistik texnologiyalar; dilin qorunması.

I. GİRİŞ

İnformasiya texnologiyalarının təsiri ilə sürətlə qloballaşan müasir dünyada dil müxtəlifliyinin qorunması aktual məsələlərdən birinə çevrilmişdir [1, 2]. Linqvistik texnologiyaların inkişafı sahəsindəki nailiyyətlər getdikcə daha çox insanlara öz dilində məlumat və xidmətlərə universal giriş imkanı yaradacaqdır. Qloballaşma prosesi təkcə texnoloji və iqtisadi sahədə deyil, həm də ictimai, siyasi və mədəni sahələrə də ciddi təsir göstərməkdədir. Hazırkı tendensiya onu göstərir ki, qloballaşmanın əsasını cəmiyyətin ingilis-amerikan modeli və onun iqtisadiyyatı, siyasəti, mədəniyyəti təşkil edir. Aydınır ki, bu model dünya dili roluna iddialı olan ingilis dili ilə sıx bağlıdır. Qloballaşma prosesi, o cümlədən dil sferasında baş verən dəyişikliklər elektron mühitdə Azərbaycan dilinin qorunması və inkişafı istiqamətində ciddi tədbirlərin görülməsini zəruri edir [2, 3].

İnformasiya Cəmiyyətinin formalaşması

istiqlamətində səylərin artırılması fonunda əksər ölkələr informasiya kommunikasiya texnologiyalarının (İKT) tətbiqi ilə elektron dövlət (e-dövlət) həllərinin inkişaf etdirilməsinə çalışırlar. E-dövlət təşəbbüsünün həyata keçirilməsi dövlət idarəçiliyində səmərəliliyə nail olmaqla yanaşı, tolerantlığın, mədəni, dini və linqvistik müxtəlifliyin qorunması üçün platforma rolunu oynayır və milli, regional, lokal səviyyədə ciddi addımların atılmasını zəruri edir.

E-dövlət yeni informasiya mühitini formalaşdıraraq dövlət sektorunun fəaliyyətini gücləndirməklə göstərilən xidmətlərin səmərəliliyinin artırılması vasitəsi kimi ortaya çıxmışdır [4-6]. Bununla yanaşı, inkişaf etməkdə olan ölkələrin e-dövlətin üstünlüklərindən tam istifadə etmək imkanları hələ də çox aşağıdır və e-dövlət layihələrinin həyata keçirilməsində çox sayda problemlərlə üzləşirlər. Bu problemlər sırasında siyasi-inzibati, infrastruktur, demoqrafik, dil və s. sosial amillər e-dövlət layihələrinin həyata keçirilməsində mühüm rol oynayır. Vətəndaşların demoqrafik xüsusiyyətləri, yaş, cins, təhsil, dil müxtəlifliyi və s. istifadəçilərin davranışlarına və onların onlayn xidmətləri qəbul etməsinə, istifadəsinə əhəmiyyətli dərəcədə təsir edən amillərdir. Bu baxımdan, e-dövlət xidmətlərinin yaradılmasında xüsusilə dil müxtəlifliyinin qorunması aktual məsələlərdən biridir. İKT-nin sürətli inkişafı və ingilis dilinin tədricən dominant dilə çevrilməsi ölkədə yaşayan əhalinin, azsayılı xalqların, etnik qrupların e-xidmətlərdən istifadə imkanlarını məhdudlaşdırır [2, 7, 8]. Xüsusilə, rəsmi dövlət dili ilə yanaşı

çoxdilliliyin mövcud olduğu ölkələrdə, məsələn, Latın Amerikasına və Afrika ölkələrində, Hindistanda, Çin Xalq Respublikasında və s. e-dövlət saytlarının dominant dildə olması əhalinin əksər hissəsinin xidmətlərdən istifadə edə bilməməsinə səbəb olur [2, 7-10]. E-dövlət mühitində dominant dillərin meydana çıxması dil müxtəlifliyi və dilin qorunması problemlərini aktuallaşdırır.

II. QLOBALLAŞMANIN LİNGVOMƏDƏNİ FƏZAYA TƏSİRİ

İstənilən informasiya bu və ya digər dilə və mədəniyyətə aid olan işarələrin köməyi ilə ötürülür. Bu aspektdən yanaşdıqda vahid informasiya fəzasının ən mühüm komponenti linqvomədəni fəza hesab olunur. İngilis-amerikan linqvomədəni fəzasının arealı mövcud olduqları dövlətlərin sərhədlərindən çox uzaqlara yayılaraq çox böyük ingilisdilli informasiya fəzasını formalaşdırır [11-13]. Postsovet ölkələrində baş verən sosial-siyasi xarakterli dəyişikliklərlə yanaşı, linqvistik-mədəni informasiya mühitində də ciddi dəyişikliklər nəzərə çarpır. Bu baxımdan qloballaşmanın linqvomədəni sferaya təsirinin nəticələrindən danışarkən ümumilikdə linqvomədəni proseslərin informasiya texnologiyalarının tətbiqinin təsiri ilə daha çox keyfiyyət dəyişikliklərinə məruz qaldığını qeyd etmək olar.

Qloballaşma prosesində bu və ya digər dilə real təhdid heç şübhəsiz dominant və ya başqa sözlə, artıq dünya dilinə iddialı olan ingilis dili hesab olunur [14-16]. İnförmasiya fəzasında iqtisadi, siyasi və sosial-mədəni xarakterli səbəblərdən

ingilisdilli resursların sürətlə yayılması bütün digər dillərdə olan resursları qat-qat üstələyir. Əlbəttə, ingilis dilinin belə sürətlə yayılmasının əsas səbəblərindən biri güclü media vasitələrinin olmasıdır.

Son onillikdə Avropada baş verən siyasi, iqtisadi və texnoloji dəyişikliklər nəticəsində ingilis dilinin rolu əhəmiyyətli dərəcədə yüksəlmişdir. İngilis dili Avropa, eləcə də Qərb və Şərq ölkələri üçün mübadilə və əməkdaşlıq dilinə çevrilmişdir. Holland linqvisti Olqa Fişerin fikirlərinə görə, dilin beynəlxalq kommunikasiya vasitəsi rolunun özünün mənfi tərəfləri var, belə ki, çox da yüksək olmayan mənimsəmə (bilmə) səviyyəsi, mədəniyyət müxtəlifliyi, müxtəlif dillərdə struktur və məna fərqliliyi, həmçinin digər dillərin nümayəndələrinin öz milli orijinallığını itirmək qorxusu, ingilis dilində təqdim olunan yad mədəniyyətin hökmranlığı altına düşmək qorxusu vardır [17].

90-cı illərdən başlayaraq “lingvistik imperializmi”, “mədəniyyət imperializmi”, “media imperializmi”, “informasiya imperializmi” adlanan anlayışlar meydana çıxmış və geniş müzakirələrə səbəb olmuşdur [11, 13, 14]. İnformasiya cəmiyyəti şəraitində linqvomədəni təsirlərin daha çox media kanalları vasitəsilə həyata keçirildiyini nəzərə alsaq dominant ingilis dilinin vahid informasiya fəzasına təsiri, o cümlədən onun Azərbaycan segmentinə də təsiri özünü göstərməkdədir. Bu baxımdan vahid informasiya fəzasında Azərbaycan dilinin qorunması çox əhəmiyyətli məsələdir və qeyd olunan geolinqvistik, dillərarası və dildaxili səviyyələr üzrə

qiymətləndirmələrin aparılması və effektiv mexanizmlərin işlənilməsi günün tələbidir.

III. RƏQƏMSAL ƏSRDƏ DÜNYA XALQLARININ DİLLƏRİNİN VƏZİYYƏTİ

Rəqəmsal əsrdə ABŞ, Böyük Britaniya və s. kimi aparıcı ölkələrin güclü iqtisadiyyatına, global media resurslarına, zəngin məlumat bazasına söykənən ingilis dili çox sürətlə milli dillərin linqvomədəni fəzasına daxil olaraq onları sıxışdırır. Hazırda mövcud internet resurslarının təxminən 60,7%-i ingilis dilinin, 31,5% əsasən, rus, ispan, türk, fars, fransız, alman, yapon, vyetnam, çin, ərəb və portuqal kimi aparıcı dillərin, qalan 7,8 %-i digər dillərin payına düşür [18]. İngilis dilinin dünyada belə üstün mövqə tutmasının əsas səbəbi bu dildə elm, təhsil və texnologiyaların sürətli inkişafı, başqa sözlə, sivilizasiyanın müasir inkişafı mərhələsində bu dilin elmi-texniki tərəqqinin (internet, fundamental biliklər üzrə əldə edilmiş nailiyyətlər və s.) lokomotivi rolunda çıxış etməsidir. Planetin dil mənzərəsini obyektiv qiymətləndirməyə çalışan UNESCO ekspertlərinin proqnozlarına görə, XXI əsrin sonuna qədər dünyadakı dillərin 50%-dən 90%-ə qədəri tamamilə məhv olub yoxa çıxa bilər [18,19].

UNESCO tərəfindən 2001-ci ildə qəbul olunan “Mədəni müxtəlifliyin ümumi Bəyannaməsi” bioloji, mədəni və dil müxtəlifliyi arasında qarşılıqlı əlaqələri nəzərdə tutur. Bəyannamə iştirakçı dövlətlərə aşağıdakı istiqamətlər üzrə əməkdaşlığı tövsiyə edir [18]:

- Dil müxtəlifliklərinə dəstək verilməsi və çox sayda dillərdən istifadə imkanlarının genişləndirilməsi;
- Təhsilin bütün səviyyələrində dil müxtəlifliyinin genişləndirilməsi, erkən yaşlardan bir neçə dilin öyrənilməsinin təşviqi;
- Ənənəvi pedaqoji sistemə müvafiq şəkildə mədəniyyət üzrə tam ünsiyyət imkanları yaradan metodların və məlumat mübadiləsi üsullarının daxil edilməsi;
- Kiberməkanda dil müxtəlifliyi də daxil olmaqla dünya informasiya şəbəkəsinə ictimai girişin təşviq olunması və s.

Avropa Birliyi ölkələrinin dayanıqlı və inklüziv inkişaf istiqamətlərini müəyyən edən “Avropa-2020” strategiyasında bu ölkələrdə e-dövlət quruculuğu zamanı cəmiyyətin sosial- mədəni, dil müxtəlifliyinin nəzərə alınması və qorunmasına səy göstərilməsi əsas baza prinsiplərindən biri kimi öz əksini tapmışdır [20]. E-dövlət mühiti elə bir həllin tapılmasına imkan verir ki, onunla həm dövlət dilinin konstitusiyaya ilə təsbit edilmiş statusunu və təbii inkişafını, həm ölkədəki milli azlıqların öz dillərindən istifadə ilə bağlı hüquqlarını, həm də ölkə ərazisində müvəqqəti yaşayan əcnəbilərin e-xidmətlərdən maneəsiz istifadəsini təmin etmək mümkün olsun.

Vahid informasiya fəzasında Azərbaycan dilinin qorunması üçün milli, korporativ və fərdi dil korpuslarının işlənməsi zəruridir. Bununla yanaşı, Azərbaycan dilində dair çoxlu sayda resursların (lüğətlər, atlaslar və s.) rəqəmsallaşdırılmasına və veb-saytlarının yaradılmasına ehtiyac vardır. İKT-nin tətbiqi ilə Azərbaycan dilinin tədrisi və

qiymətləndirilməsi üçün yeni mexanizmlərin işlənməsi olduqca aktualdır. Mühüm məsələlərdən biri informasiya texnologiyaları, süni intellekt və təbii dilin emalı metodlarının digər dillərdə geniş tətbiq olunmasına baxmayaraq Azərbaycan dilinə tətbiqdə problemlərin mövcud olmasıdır. Buna səbəb Azərbaycan dilinin ontologiyası, tezaurusu, müxtəlif təyinatlı tezlik lüğətləri, sentiment və vulqar sözlər bazası, Azərbaycan dilinin milli dil korpusu və digər zəruri resursların (steminq, lemmatizasiya, tokenizasiya və s.) mövcud olmamasıdır.

Qeyd olunan problemlərin aradan qaldırılması vahid informasiya fəzasında aşağıda göstərilən məsələlərin həll olunmasına imkan verəcəkdir:

- Azərbaycan dilində mətnləri, nitqləri xülasələşdirmək və təsnif etmək (text mining, audio mining və s.);
- Azərbaycan dilində nitqləri kompüter vasitəsilə tanımaq və analiz etmək;
- Kompüterdə olan azərbaycandilli mətnlərdən süni nitq sintez etmək, insan-kompüter (robot, dron və s.) dialoq sistemlərini (çatbot və s.) işləmək;
- Mətnlərin, nitqlərin emosional tonallığını müəyyən etmək (opinion mining, sentiment analiz və s.);
- Azərbaycan dili ilə dünyanın digər dilləri arasında keyfiyyətli kompüter tərcüməsi proqramları yaratmaq;
- Onlayn kanallar vasitəsi ilə Azərbaycan dilinə daxil olan yad terminləri avtomatik aşkarlamağa imkan verən texnologiyalar işləmək;

- Virtual məkanda Azərbaycan dilinin ekoloji vəziyyətini fasiləsiz monitoring etmək, qiymətləndirmək və dilin inkişafı ilə əlaqədar obyektiv və səmərəli qərarlar qəbul etmək;
- İnternet mühitdə milli informasiya axtarış sistemi yaratmaq;
- Azərbaycandilli virtual məkanda informasiya təhlükəsizliyinin təmin olunması, uşaqların ziyanlı mətn və audio-video kontentlərdən qorunmasına, kibercinəyatkırlıqla effektiv mübarizə aparılmasına nail olmaq;
- E-dövlətin formalaşmasına və vətəndaşlara Azərbaycan dili ilə bağlı müxtəlif onlayn xidmətlər göstərmək;
- Azərbaycandilli mətnlər üçün milli antiplagiat proqramları işləmək;

Bütün bunlar rəqəmsal transformasiyalar dövründə Azərbaycan dili ilə bağlı çoxlu sayda problemləri həll etməklə onlayn mühitdə Azərbaycan dilinin qorunmasını, inkişafını və müxtəlif məqsədlər üçün geniş tətbiqini təmin etməyə imkan verəcəkdir.

Bu sahədə həlli mühüm məsələlərdən biri e-dövlət mühitində Azərbaycan dili ekosisteminin yaradılmasıdır. Belə bir sosiotexnoloji arxitektura hesabına milli dillərin e-dövlət platformasının tərkib hissəsinə çevrilməsi, bu texniki platformada özünə yer tutması qloballaşma dövründə dillərin qorunması və inkişafına böyük töhfə verəcəkdir.

IV. E-DÖVLƏTDƏ DİLİN ARXITEKTUR-TEXNOLOJİ ƏSASLARI

Milli dillərin e-dövlətin tərkib hissəsinə

çevrilməsi və bu platformanın bütün seqmentləri üzrə təbii inkişafının təmin olunmasından ötrü yeni metodlar, alqoritmlər, yanaşmalar işlənməli, bir sıra tətbiqlər hazırlanmalıdır. Bu məqsədə çatmaq üçün aşağıdakı məsələlərin həlli zəruridir [18]:

- E-dövlətin bütün vətəndaşlar və ölkədə müvəqqəti yaşayan əcnəbilər üçün əlçatanlığının təmin olunması məqsədilə lokallaşdırma (milli dildən istifadə) və qloballaşdırma (əsas xarici dillərdən istifadə) problemlərinin həlli;
- Ölkə ərazisində yaşayan milli azlıqların dillərinin sənədləşdirilməsi (language documentation) üçün e-dövlət platformasının imkanlarından istifadə;
- E-dövlət sistemində toplanmış böyük həcmdə verilənlərin təhlili ilə ölkədəki mövcud dil situasiyasının mütəmadi qiymətləndirilməsi;
- E-dövlət mühitində, kütləvi informasiya vasitələrində, sosial mediada və s. dildən istifadənin intellektual analizi metodlarının işlənməsi və əks-əlaqə mexanizmlərinin tətbiqi;
- E-dövlət sistemində dilin təmizliyinin təmin olunması, orfoqrafiya, orfoepiya, transliterasiya, terminologiya, dialektologiya məsələləri, adlar və soyadlar, vulqarizmlər, dil sənayesi xidmətləri (məşin tərcüməsi, nitqin tanınması, dil öyrənmə və dil biliyi səviyyəsini qiymətləndirmə və s.), Dilin Milli Korpusu və s. xidmətlərin formalaşdırılması;

- Milli domen adlarının tətbiqi və çoxdilli internetin dəstəklənməsi üçün texniki məsələlərin həlli;
- Qloballaşma şəraitində milli dillərin qorunması və inkişafının təmin olunması üçün təklif edilən modellərin tətbiqi.

Belə bir modelin tətbiqi milli e-dövlət bazasında dilin müxtəlif funksional modulları üzrə məlumatların toplanmasına, təhlilinə, qiymətləndirilməsinə və onlardan məqsədyönlü şəkildə istifadəyə şərait yaradacaqdır.

V. E-DÖVLƏT MÜHİTİNDƏ AZƏRBAYCAN DİLİ EKOSİSTEMİNİN FORMALAŞDIRILMASI

Dil müxtəlifliyi, ifadə azadlığı ölkələrin, eləcə də, Avropa Birliyinin əsas dəyəri hesab olunmaqla e-demokratiyanın formalaşdırılmasının baza prinsiplərindən hesab olunur. Avropanın çoxdilli olması fundamental mədəni və sosial dəyəərə malik olsa da, dil müxtəlifliyinin olması insanlar arasında ünsiyyətə əhəmiyyətli dərəcədə təsir edə biləcək maneə hesab oluna bilər [1, 21, 22]. Bu baxımdan “birliyi və müxtəlifliyi” qoruyub saxlamaq olduqca mürəkkəb məsələdir. Hazırda bu məsələ dominant dil hesab edilən ingilis dili ilə müqayisədə çox sayda dillər üçün də aktualdır və millətlərin, əzsaylı xalqların dillərinin qorunması üçün ciddi tədbirlərin görülməsini labüd edir.

Onlayn mühitdə linqvistik bölünmənin qarşısının alınması və ya dilin qorunması üçün bir sıra texnoloji həllər mövcuddur. Dil texnologiyaları sahəsindəki yeniliklər maşın tərcüməsi, mətn analitikası, semantik

analiz və nitqin tanınması texnologiyaları insanlar arasında dil baryerinin aradan qaldırılmasına, eyni zamanda dil müxtəlifliyinin qorunmasına və rəqəmsal dünyada çoxdilliliyin inkişaf etdirilməsinə imkan verir [3, 22, 23]. Dil texnologiyalarının inkişafı nəticəsində insanlar internet mühitində öz dillərində yazsa, oxuya və danışsa bilirlər, həmçinin başa düşdükleri dildə informasiyaya çıxış imkanı əldə edirlər. Onlayn mühitdə dil texnologiyalarının tətbiqi dil baryerini aradan qaldırmaqla yanaşı, dilin qorunması və e-dövlət mühitində çoxdilliliyi inkişaf etdirməklə ölkələr arasında vahid informasiya fəzasının formalaşdırılmasına imkan verəcəkdir.

Hazırda qloballaşma prosesinin sürətlə getdiyi bir zamanda İnternetin inkişafı, ağıllı dövlətə (*smart government*) keçid və vahid informasiya fəzasının yaranması, dominant dil hesab olunan ingilis dilinin dünya dilinə iddialı olması dillərin itməsi prosesini sürətləndirən əsas amillərdən biri hesab olunur.

E-dövlət mühitində dilin qorunması üçün təşəbbüslərin yerinə yetirilməsi üçün müvafiq proqramların, layihələrin icrası mərhələli yanaşma əsasında həyata keçirilməlidir.

Azərbaycanda mövcud vəziyyətin analizi göstərir ki, müxtəlif təşkilatlar, fərdlər dillə bağlı ayrı-ayrı proqram məhsulları hazırlayırlar. Amma bu məhsullar müxtəlif səbəblərdən (kommersiya və s.) geniş istifadə oluna bilmir. Bu sistemləri e-dövlət sistemlərinə inteqrasiya etmək və ümumiyyətlə, Azərbaycan dilinin ekosistemini formalaşdırmaq lazımdır.

Bunun da nəticəsində virtual məkanda cəmiyyətə, vətəndaşlara, hər kəsə ödənişsiz, qeyri-məhdud olaraq Azərbaycan dili xidmətləri göstərilməsi (orfoqrafiya, terminologiya, dialektologiya, tərcümə və s.) mümkün olacaqdır. Aydınadır ki, bu texnoloji innovasiyalar dövlətin dillə bağlı siyasətini, yəni rəsmi dövlət dili kimi istifadə edilməsini, qorunmasını və inkişaf etdirilməsini, həmçinin başqa dillərə tərcümə, transliterasiya münasibətlərinin qurulmasını təmin etməlidir. Bu səbəbdən dövlət tərəfindən bu sahədə icra olunan layihələrə maliyyə dəstəyinin verilməsi zəruridir. Şəkil 1-də Azərbaycan dili ekosisteminin konseptual modeli təklif edilmişdir. Göründüyü kimi elektron mühitdə Azərbaycan dili ekosisteminin formalaşdırılması üçün e-dövlət və onun funksional komponentləri Azərbaycan dilinin tələblərinə uyğun qurulmalıdır. Bu məqsədlə, Azərbaycan dilinin funksional komponentləri rəqəmsal platformada elə yaradılmalıdır ki, e-dövlət mühitində həm bir-biri ilə qarşılıqlı əlaqədə olmaqla yanaşı, həm də vətəndaşlara fasiləsiz onlayn dil xidmətləri (terminoloji, orfoqrafik, vulqar sözlər, tərcümə və s.) göstərmək imkanlarına malik olsun. Bu baxımdan, Azərbaycan dilinin tədrisi və qiymətləndirilməsi üçün beynəlxalq standartlardan istifadə olunması zəruridir. Vətəndaşların dil bilikləri milli standartlar əsasında yazma, oxuma, dinləmə və danışma üzrə qiymətləndirilməlidir. Dövlətin dil siyasətinə dəstək məqsədi ilə e-dövlət mühitinin tərkib hissəsi kimi linqvistik demoqrafiya sistemi qurulmalı və müxtəlif aspektlər üzrə sosiolinqvistik təhlillər aparılmalı və zəruri təkliflər

işlənilməlidir. Şəkil 2-də e-dövlət platformasına Azərbaycan dili ekosisteminin konseptual modeli verilmişdir. Göründüyü kimi e-dövlət konsepsiyasının formalaşdırılması üçün onlayn mühitdə Azərbaycan dili ekosisteminin yaradılması mühüm əhəmiyyət kəsb edir.

Beynəlxalq təcrübə nəzərə alınaraq dilin ekosistemləri açıq arxitekturaya malik olmalı və texnoloji çağırışlara adekvat olaraq daim təkmilləşdirilməlidir. Eləcə də, bu işləri həyata keçirərkən Azərbaycan dili ilə yanaşı digər etnik azlıqların da dillərinin tələbləri nəzərə alınmalı, dövlətin multikultural siyasəti çərçivəsində onlar üçün müvafiq onlayn dil ekosistemləri yaradılmalıdır.

Qeyd olunanları nəzərə alaraq e-dövlət mühitində Azərbaycan dili ekosisteminin formalaşdırılması üçün aşağıda göstərilən məsələlərin kompleks həll olunması nəzərdə tutulur [1, 18]:

- E-xidmətlərin vahid informasiya məkanına inteqrasiya olunması və e-Azərbaycan mühitinin formalaşdırılması;
- E-xidmətlərin beynəlxalq standartlar əsasında standartlaşdırılması və çoxdilliliyin dəstəklənməsi, milli mənsubiyyət, ünsiyyət dili, yaşayış yerlərindən və s. asılı olmayaraq göstərilən xidmətlərin bütün vətəndaşlar üçün əlyətərli olmasının təmin edilməsi;
- E-xidmətlərin dil müxtəlifliyi nəzərə alınmaqla əhalinin bütün kateqoriyaları üçün əlyətərli olması üçün mexanizmlərin işlənilməsi;

- Virtual dil tədrisi platformasının yaradılması;
- Milli axtarış sisteminin yaradılması;
- Linqvistik texnologiyaların (maşın tərcüməsi, mətn analitikası, semantik analiz, dilin sanitarizasiyası, nitqin tanınması) işlənilməsi;
- Transliterasiya sistemlərinin yaradılması;
- Dil sənayesinin inkişaf etdirilməsi üçün tədbirlərin görülməsi;
- Milli antiplagiat sisteminin yaradılması;
- Milli antivirus xidmətinin yaradılması;
- Dialektoloji bazanın yaradılması və onlayn mühitdə dil mədəniyyətinin formalaşdırılması və s.

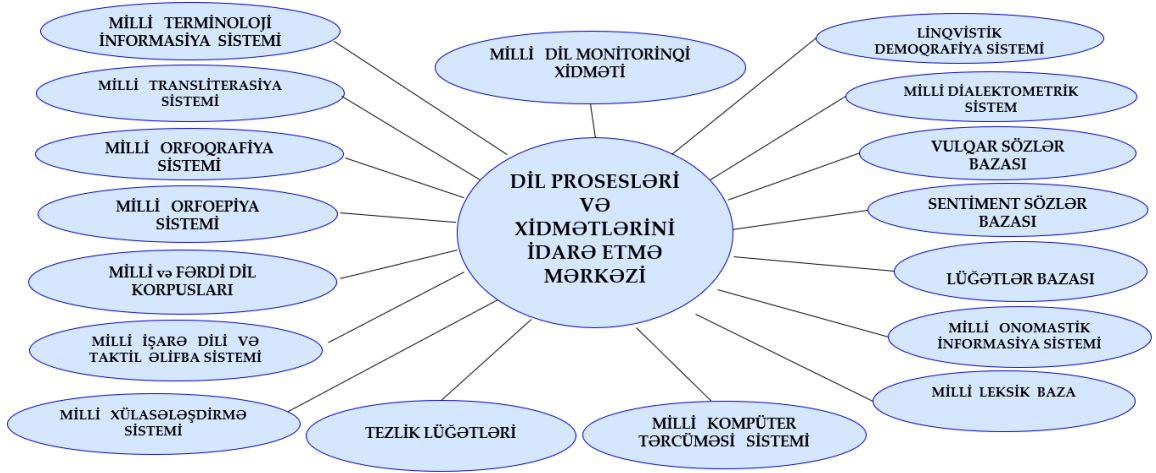
E-dövlət mühitində Azərbaycan dili ekosisteminin yaradılması dilin qorunması üçün yeni vasitələr, imkanlar təklif edir. Qlobal informasiya fəzasının e-Azərbaycan seqmentində Azərbaycan dilinin qorunması üçün bulud əsaslı linqvistik texnologiyaların tətbiq olunması son nəticədə istər geolinqvistik, istərsə də dillərarası və dildaxili səviyyələr üzrə qiymətləndirmələrin aparılmasına imkan verməklə, effektiv mexanizmlərin işlənilməsinə şərait yaradacaqdır.

NƏTİCƏ

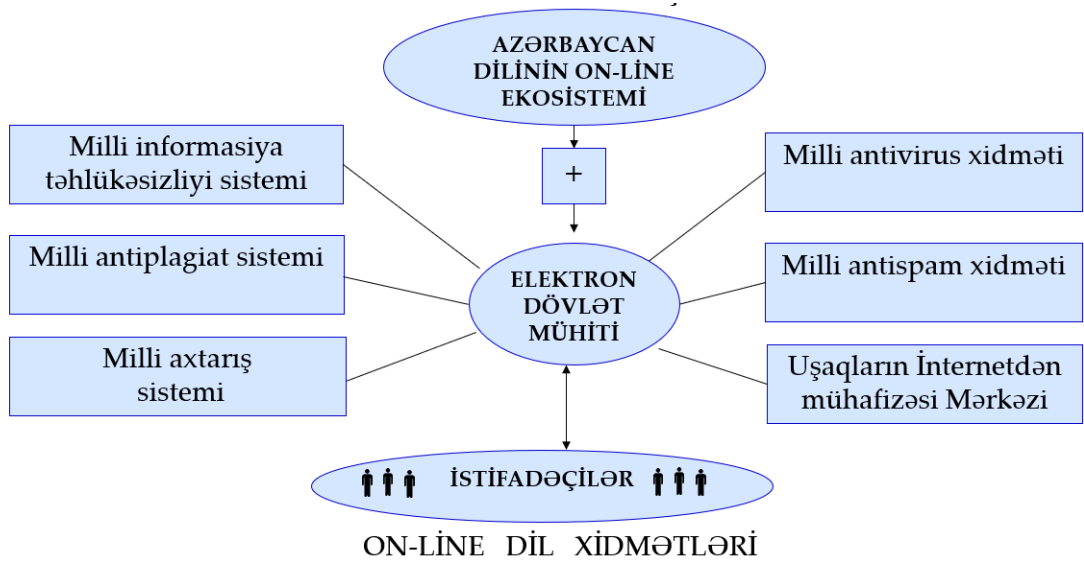
Beynəlxalq təcrübəyə nəzər saldıqda aydın olur ki, ümumiyyətlə, onlayn mühitində dilin qorunması vasitələri ilə bağlı müxtəlif ideyalar, yanaşmalar vardır. Qloballaşma prosesində dili təhlükə altında qoyan bir çox amillər mövcuddur və bu

amillərin hər birinin nəzarətdə saxlanması hesabına dilin qorunmasını təmin etmək mümkün deyil. Bu baxımdan İnternet, həm də dilin yox olması və dilin qorunması məsələləri barədə məlumatlılığın artırılması üçün istifadə edilə bilər.

Məqalədə onlayn platformada Azərbaycan dili ekosisteminin formalaşdırılmasının konseptual əsasları araşdırılmışdır. E-dövlət platformasında Azərbaycan dili ekosisteminin yaradılmasının konseptual modeli təklif olunmuşdur. Milli dillərin e-dövlətin tərkib hissəsinə çevrilməsi və bu platformanın bütün seqmentləri üzrə təbii inkişafının təmin olunmasından ötrü vahid ekosistemin yaradılmasının konseptual əsasları tədqiq olunmuşdur. Belə bir ekosistemin formalaşdırılması e-dövlət platformasında dilin müxtəlif funksional modulları üzrə məlumatların toplanmasına, təhlilinə, qiymətləndirilməsinə və onlardan məqsədyönlü şəkildə istifadəyə şərait yaradacaqdır. Bulud əsaslı linqvistik texnologiyaların tətbiqi İnternet üzərindən müxtəlif xidmətlərin (tərcümə, kataloqlaşdırma, informasiya resurslarının saxlanması, əlyetərlilik və s.) göstərilməsinə imkan verir, danışiq dilinin qorunması üçün də istifadə oluna bilər. Həmçinin, səs, nitqin tanınması üçün istifadə edilən texnologiyalar danışiq dilinin qorunması üçün istifadə oluna bilər. Vahid informasiya fəzasının e-Azərbaycan seqmentində də linqvistik texnologiyaların istifadəsi müxtəlif e-xidmətlərin göstərilməsinə imkan verməklə Azərbaycan dilinin qorunmasına və inkişafına xidmət edəcəkdir.



Şəkil 1. On-line platformada Azərbaycan dili ekosisteminin konseptual modeli



Şəkil 2. E-dövlət mühitində Azərbaycan dili ekosisteminin konseptual modeli

İSTİNADLAR

- [1] R. Alguliyev, F. Yusifov, A. Gurbanova, “Protection of Azerbaijani Language in e-government platform,” e-Journal of Linguistics, Udayana University, Indonesia, vol 15., No. 2, 2021, pp. 155-161.
- [2] Languages in a Globalising World, editors Maurais J. and Morris M.A., Cambridge, UK : Cambridge University Press, 2003, 345 p.
- [3] Ə.M. Qurbanova, “Virtual məkanda Azərbaycan dili: bəzi problemlər və həlli yolları,” İnformasiya cəmiyyəti problemləri jurnalı, №1, 2010, səh. 63-70.
- [4] Е.Г. Васильева, Д.В. Кононенко, “Современные интерпретации концепции электронного государства (электронного правительства),” Вестник Волгоградского государственного университета. Сер. 5, Юриспруд. № 1 (30), 2016, стр. 9-16.
- [5] M. Alshehri and S. Drew, “E-government principles: implementation, advantages and challenges,” International Journal Electronic Business, 9(3) pp. 255-270.
- [6] R.M. Əliquliyev, F.F. Yusifov, “Elektron dövlətin formalaşmasının bəzi aktual elmi-nəzəri problemlərinə həll perspektivləri,” İnformasiya Cəmiyyəti Problemləri elmi-praktiki jurnalı, 2014, s. 3-13.
- [7] P. Mittal, A. Kaur, “E-Governance - A challenge for India,” International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol. 2, Issue 3, 2013, pp.1196-1199.
- [8] W.K. Torgby and N.Y. Asabere, “Challenges of Implementing and Developing E-Government: A Case Study of the Local Government System in Ghana,” International Journal of Computer Science and Telecommunications, Vol. 5, Issue 8, 2014, pp. 39-48
- [9] S. Lata, S. Chandra, “Development of Linguistic Resources and Tools for providing multilingual Solutions in Indian Languages – A Report on National Initiative,” Proceedings of the International Conference on Language Resources and Evaluation, LREC’2010, 2010, Valletta, Malta, pp. 2851-2854.
- [10] G. Pérez-Salazar, A. Aguilar-Edwards, C.N. Mata-Martínez, “Linguistic diversity and accessibility in Mexican government web sites: executive branch,” Comparative Cultural Studies: European and Latin America Perspectives, No 2, 2016, pp.1-13
- [11] Т.Г. Добросклонская. “Языковые последствия информационной глобализации,” Языки в современном мире: материалы X международной конференции, Москва, 2012. стр. 59–71.
- [12] А.О. Лалетина, “Глобализация как объект лингвистических исследований,” Политическая Лингвистика, №3, 2011, стр.39-45.
- [13] D. Crystal, Language and the Internet. Cambridge, UK : Cambridge University Press, 2001, 272 p.
- [14] D. Crystal, English as a Global Language. Cambridge UK : Cambridge University Press, 2003, 160 p.
- [15] И.Э. Алекберова, “Английский язык в условиях глобализации,” Lingua mobilis, № 4 (37), 2012, стр.104-110
- [16] Е.С. Гриценко, “Язык и безопасность в контексте глобализации. Научный журнал Власть,” М., 2011, № 11, стр. 9-11.
- [17] O. Fischer, Morphosyntactic Change, Functional and Formal Perspectives, 2006, 398 p.
- [18] R. Əliquliyev, Y. Cəfərov, F. Yusifov, Ə. Qurbanova, “Kosmopolit e-dövlətdən milli e-dövlətə doğru: virtual mühitdə Azərbaycan dili xidmətlərinin yaradılması perspektivləri,” İnformasiya Cəmiyyəti Problemləri jurnalı, 2021, №2, səh. 3-25
- [19] Language Vitality and Endangerment. UNESCO Ad Hoc Expert Group on Endangered Languages, International Expert Meeting on UNESCO Programme Safeguarding of

- Endangered Languages, Paris, 10–12 March 2003, 27 p.
- [20] Europe 2020. A European strategy for smart, sustainable and inclusive growth. Brussels, 2010, 37 p.
- [21] The European Roadmap for Linguistic Diversity, Towards a new approach on languages as part of the European Agenda 2020, 2010, www.npld.eu/uploads/publications/313.pdf
- [22] The European Language Cloud, or how to enable a multilingual Europe, 2015, www.euractiv.com
- [23] R.M. Əliquliyev, R.Ş. Mahmudov, “Dil Sənayesi: imkanlar, perspektivlər və problemlər,” İnformasiya cəmiyyəti problemləri jurnalı, №1, 2018, səh. 3–26

Bulud sistemləri əsasında Azərbaycan dili xidmətlərinin təşkili və intellektual analizi məsələləri

Rəşid Ələkbərov¹, Məmməd Həşimov²

^{1,2}İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
t.direktor_muavini@iit.science.az, mamedhashimov@gmail.com

Xülasə — Məqalədə bulud sistemləri əsasında Azərbaycan dili servislərinin yaradılması və intellektual analizi məsələlərinə baxılmışdır. Virtual mühitdə yaradılan və yaradılması nəzərdə tutulan Azərbaycan dili servisləri sistemi analiz edilmiş və onların faydaları qeyd olunmuşdur. Azərbaycan dili servisləri sisteminin yaradılması və onların e-dövlət sisteminə inteqrasiyası zamanı meydana çıxan problemlər təhlil edilmişdir. Həmin problemlərin aradan qaldırılması üçün bulud sistemlərindən istifadənin üstünlükləri göstərilmişdir.

Açar sözlər — *e-dövlət; dil servisləri; verilənlərin emal mərkəzi; bulud texnologiyaları*

I. GİRİŞ

Dövlətlər, xüsusilə də insanlarla birbaşa ünsiyyətdə olan dövlət idarələri müvafiq texnologiyalar vasitəsilə gündəlik xidmətlərini həyata keçirirlər. Son texnologiyalardan istifadə, xidmətlərin hazırlanması üçün tələb olunan vaxtın azaldılması, yararlı və səmərəli xidmətin təqdim olunması vətəndaşlarla ünsiyyəti artırmaq üçün çox vacibdir. Dövlət orqanları arasında, dövlət və vətəndaşlar arasında, dövlət və biznes strukturları arasında

məlumat mübadiləsini həyata keçirən sistem “Elektron dövlət” sistemi adlanır. Elektron dövlət (e-dövlət) sistemi 7/24 İKT-nin (informasiya və telekommunikasiya texnologiyalarının) tətbiqi ilə dövlət qeydiyyatlarının rəqəmsallaşdırılması, vergi yığımlarının avtomatlaşdırılması, cəmiyyətdən müxtəlif tipli informasiyanın qəbul edilməsi, məlumatların toplanması və klassifikasiyası, seçkilərin keçirilməsi, istifadəçilərə yüksək səviyyədə elektron xidmətlərin göstərilməsinin təşkili və s. kimi idarəetmə proseslərini təmin edir [1, 2].

Hal-hazırda dünyanın müxtəlif ölkələrində e-dövlət konsepsiyasının tədqiqi və inkişaf etdirilməsi istiqamətində geniş işlər aparılır. E-dövlət müasir informasiya texnologiyalarından istifadə etməklə dövlət qurumları tərəfindən ölkə ərazisində yaşayan bütün vətəndaşlara, hüquqi və fiziki şəxslərə, xarici vətəndaşlara və vətəndaşlığı olmayan şəxslərə informasiya və elektron xidmətlərin göstərilməsinə şərait yaradır. Vətəndaşların dövlət orqanları ilə təmaslarının daha rahat və münasib şəkildə təşkili üçün dövlət orqanlarının göstərdiyi elektron xidmətlərin bir ünvanda cəmləşdirildiyi “elektron dövlət” portalı “bir pəncərə” prinsipi əsasında yaradılır. Ölkəmizdə də e-dövlətin formalaşdırılması beynəlxalq təcrübəyə əsaslanır və hal

hazırda, e-gov.az portalında müxtəlif nazirlik və komitələrə məxsus 454 xidmət növü fəaliyyət göstərir [3]. E-dövlət mühitində dövlətin dil siyasətini həyata keçirən xüsusi xidmətlərin yaradılması zəruridir. Bu servislər e-dövlət mühitində onlayn fəaliyyət göstərən, bütün vətəndaşlar üçün dilin müxtəlif xidmətlərini həyata keçirən funksional modullardan ibarət olmalıdır.

II. AZƏRBAYCAN DİLİ SERVLƏRİ

Virtual mühitdə yaradılan və yaradılması nəzərdə tutulan Azərbaycan dili servisləri aşağıda göstərilmişdir [4]:

Milli terminoloji informasiya sistemi. E-dövlət sistemində yeni terminlərin dilə daxil edilməsi, yeni terminlərin yaradılması prosesinə geniş ictimaiyyətin cəlb edilməsi, dilin terminoloji bazasınının yad sözlərdən qorunması, milli terminoloji bazanın davamlı inkişaf etdirilməsi məqsədilə terminologiya ilə bağlı xidmət, vahid onlayn terminoloji sistem yaratmağa ehtiyac vardır [5].

Milli orfoqrafiya və orfoepiya informasiya sistemi. E-dövlətdə orfoqrafiya və orfoepiya qaydaları ilə bağlı xidmətlər yönümlü elə bir sistem yaradılmalıdır ki, sosial şəbəkələrdə, kütləvi informasiya vasitələrində, sosial mediada ilk dəfə rast gəlinən sözü avtomatik olaraq xüsusi bazaya daxil etsin. Burada həmin sözlər elmi baxımdan təhlil olunsun, onların düzgün yazılışı və tələffüzü qaydaları təyin olunsun, dil üçün faydalılıq dərəcəsinin müəyyənləşdirilməsi üçün imkan yaransın.

Milli dialektoloji sistem. Ölkənin müxtəlif bölgələrində istifadə olunan dialekt sözlər elektron məlumat bazalarında toplanaraq linqvistik xəritə formasında virtuallaşdırılır. Bu bazalar bütün istifadəçilər üçün əlçatandır, belə ki, yeni dialekt sözlərin bazaya daxil edilməsi, sözlər haqqında əlavə izah və şərhlərin verilməsi imkanları yaradılır.

Milli transliterasiya sistemi. Rəqəmsal texnologiyalardan istifadəni asanlaşdırmaq, axtarış sistemlərindən daha asan və səmərəli istifadə etmək məqsədilə bir çox millətlər öz dillərini və ya hərflərini ingilis əlifbası ilə uyğunlaşdırmağa çalışırlar. Elektron xidmətlərdən istifadənin insanlar üçün daha rahat və əlçatan olması üçün latın qrafikalı əlifbaların oxunma və yazmasını asanlaşdıracaq vahid romanizasiya standartının tətbiq olunmasına ehtiyac vardır. Lakin bu tip dəyişikliklər düzgün siyasi-ideoloji, tarixi və elmi əsaslar üzərində aparılmasa, systemsiz şəkildə baş verərsə milli dillərin gələcəyi üçün müxtəlif təhlükələr yarana bilər [6].

Milli dil korpusu dilin bütün özəlliklərini, inkişaf mərhələlərini, üslublarını, fonetikasını, qrammatikasını, frazeologiyasını, lüğətlərini və s. özündə ehtiva edir. Milli korpusda müxtəlif istiqamətli və janrlı bədii mətnlər, publisistik, elmi, elmi-populyar, rəsmi-ışğuzar və hüquqi mətnlər, tədris mətnləri, texniki mətnlər, məişət mətnləri, multimedia materialları – videolar, səs yazıları və s. geniş əks olunur. Elektron dövlətdə milli dil korpusunun yaradılması vacib məsələlərdəndir.

Milli toponimik geoinformasiya sistemi. Ölkəmizin bütün regionlarının toponimik məlumatlarını bir-birinə bağlayan vahid, məlumat bazasının formalaşdırılması aktual məsələlərdəndir. Toponimlərin belə bir sistem əsasında pasportlaşdırılması, onların mənşəyinin zaman, məkan, aid olduqları xalq, şəxs, dövlət, relyef və s. üzrə toplanması və təsviri analitik qərar qəbul etmə məqsədilə vəziyyətin analizi, proqnozlaşdırma və s. üçün faydalı ola bilər [7].

Dillərin sənədləşdirilməsi dil materiallarının uzunmüddətli və çoxməqsədli qeydiyyatıdır. Dillərin sənədləşdirilməsinin başlıca ünsürü milli azlıqların dillərində olan materialların bir elektron mənbədə toplanmasıdır. Buraya əsasən, nağıllar, rəvayətlər, lətifələr, məişət dialoqları, laylalar, mahnılar və s. aiddir.

Dil sənayesi xidmətləri (tədris, tərcümə, lokallaşdırma və s.) qlobal informasiya cəmiyyətinin formalaşmasına və inkişafına xidmət edir ki, onun da əsasında bu sahədə İKT-nin uğurlu tətbiqi dayanır.

Vulqarizmlər bazası. Elektron mühitdə vulqarizmlərə qarşı uğurlu mübarizə məqsədilə e-dövlət platformasında qlobal vulqarizm bazasının yaradılması məqsədəuyğun hesab olunur.

Adlar-soyadlar sistemi. Elektron dövlət mühitində xüsusi adlarla (ad, ata adı və soyad) bağlı bir çox məsələlərin həlli xüsusi əhəmiyyət kəsb edir: vahid adlandırma standartının formalaşdırılması, adların modelləşdirilməsi, ad komponentlərinin müəyyənləşdirilməsi, adların orfoqrafiya qaydalarının və transliterasiyası

normalarının dəqiqləşdirilməsi, adların elektron bazasının yaradılması və onlardan istifadə mexanizmlərinin işlənilib hazırlanması və s.

Marka adları bazası. Dünyanın əksər ölkələrində adlandırma, adın analiz olunması, marka yaratmaq kimi xidmətlərdən ibarət sistemlər vardır. Ölkəmizdə də bazar araşdırmaları və marketingin əhəmiyyətli bir hissəsi olaraq belə bir sistemin yaradılması zəruriyyəti vardır.

Dilin təmizlənməsi. Virtual kommunikasiya vasitələrində, e-firlərdə və s. dil normalarının pozulması hallarının qarşısının alınması məqsədilə e-dövlət mühitində inteqrasiya edilmiş elektron mətnyaratma, korrektə, redaktə kimi proqramların tətbiqi nəzərdə tutulur.

Yuxarıda qeyd edilən dil servisləri əsasında e-dövlət sistemini formalaşdırmaq üçün yeni şəbəkə infrastrukturunun yaradılması tələb olunur. Şəbəkə infrastrukturunun yaradılması üçün isə kifayət qədər maddi vəsait və zaman tələb olunur. Hər hansı bir istifadəçi portala müraciət edib ona lazım olan xidmət növünə sorğu göndərdikdə, həmin sorğu aidiyyəti qurumun verilənlərin emalı mərkəzinə (VEM) göndərilir və orada emal edilir (Şəkil 1). Bu isə bir çox dövlət qurumlarının özünə məxsus VEM-in olmasını zəruri edir.

VEM-lər aşağıdakı əlamətlərə görə xarakterizə olunur:

- daim işçi vəziyyətində saxlanılması;
- müxtəlif platformalar üzərində qurulması;

- çoxsaylı əməliyyat sistemlərindən istifadə;
- mühəndis və proqramçı qruplarının olması;
- enerji sərfiyyatı;
- soyutma sistemlərindən istifadə.

VEM-lərin müxtəlif platformalardan və əməliyyat sistemlərindən istifadə etməsi onların bir-birinə inteqrasiya olunmasını çətinləşdirir. Eyni zamanda qeyd olunan dil servislərinin müxtəlif VEM-lərdə yerləşdirilməsi birgə analitik təhlillərin aparılmasında da müxtəlif problemlər yaradır.

Qeyd olunan problemlərin aradan qaldırılması və istifadəçilərə daha keyfiyyətli xidmətlərin göstərilməsi üçün respublikamızın dövlət qurumlarında ayrı-ayrı VEM-lərin yaradılması əvəzinə onların bulud texnologiyaları əsasında yaradılan vahid VEM-də cəmlənməsi təklif olunur.

III. BULUD SİSTEMLƏRİ ƏSASINDA AZƏRBAYCAN DİLİ SERVİSLƏRİNİN TƏŞKİLİ

Yüksək məhsuldarlıq, az texniki xidmət tələb edən, xərclərə qənaət üzrə geniş potensialı və bir sıra digər faydaları ilə seçilən bulud texnologiyaları gündən günə inkişaf edərək daha çox sahələrə nüfuz edir. Bulud texnologiyaları – istifadəçiyə xidmət şəklində internet və ya lokal şəbəkə vasitəsilə əlyətərli olan proqram-aparat təminatıdır. Bulud texnologiyaları müəyyən resurslara (hesablama resurslarına, proqram və məlumatlara) uzaq məsafədən müraciət

etmək üçün rahat interfeysdən istifadə etməyə imkan verir. İstifadəçi kompüteri bu halda, şəbəkəyə qoşulmuş adi terminal rolunu oynayır [8, 9]. Beləliklə, vahid VEM-in resursları seqmentlərə bölünərək dövlət qurumlarının istifadəsinə verilir və onlar istifadəçilər üçün lazım olan informasiya resurslarını həmin VEM-də yerləşdirirlər. Məqalədə təklif olunan yanaşma da müxtəlif təşkilatlar tərəfindən yaradılan dil servislərinin bulud əsaslı vahid mərkəzdə cəmlənməsidir. Bu isə yuxarıda qeyd olunan problemlərin aradan qaldırılmasına və xərclərin optimallaşdırılmasına imkan verir. Bulud texnologiyaları əsasında yaradılan dil servisləri sisteminin konseptual modeli şəkil 2.-də göstərilmişdir.

E-dövlətlə bağlı bulud texnologiyalarının tətbiqinin üstünlükləri [10]:

- xidmətlərin keyfiyyətinin yüksəldilməsi;
- aparat və proqram təminatına, xidmətə və elektrik enerjisinə olan xərclərin azaldılması;
- vaxta qənaət edilməsi, proqram və məlumatların asanlıqla əldə olunması;
- təşkilatlar arasında inteqrasiya və daha yaxşı əlaqənin qurulması;
- coğrafi, aparat və proqram təminatı baxımından məhdudiyətlərin aradan qaldırılması;
- “bulud”da saxlanılan sənədlərin istifadəçilər üçün istənilən zaman və istənilən yerdə əlçatan olması.
- istifadəçi sayının gündən-günə artması.

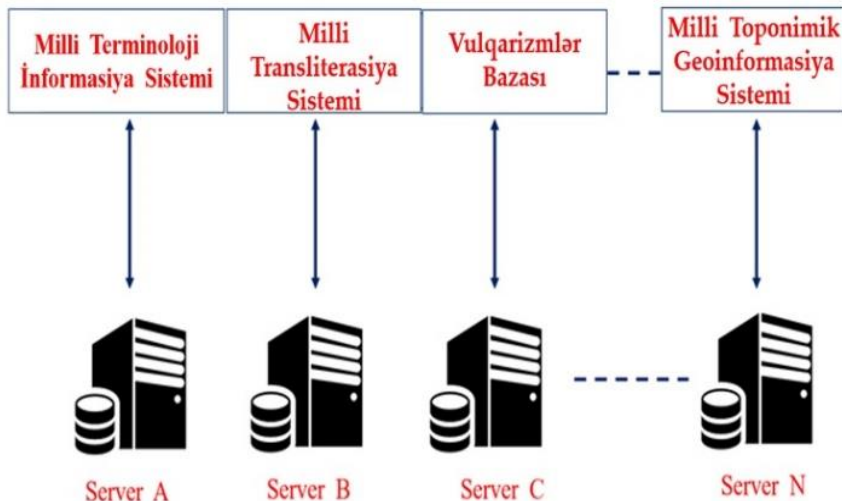
Digər tərəfdən e-dövlət sistemində bulud texnologiyaları əsasında Azərbaycan dili servisləri sistemlərindən toplanmış böyük həcmdə verilənlərin saxlanması və analitik təhlili ilə ölkədəki mövcud dil situasiyasının (dilənin müxtəlifliyinin qorunması, sosial şəbəkələrdə dil normalarının qorunması və s.) qiymətləndirilməsinə və onların məqsədyönlü şəkildə istifadəsinə nail olmaq olar.

NƏTİCƏ

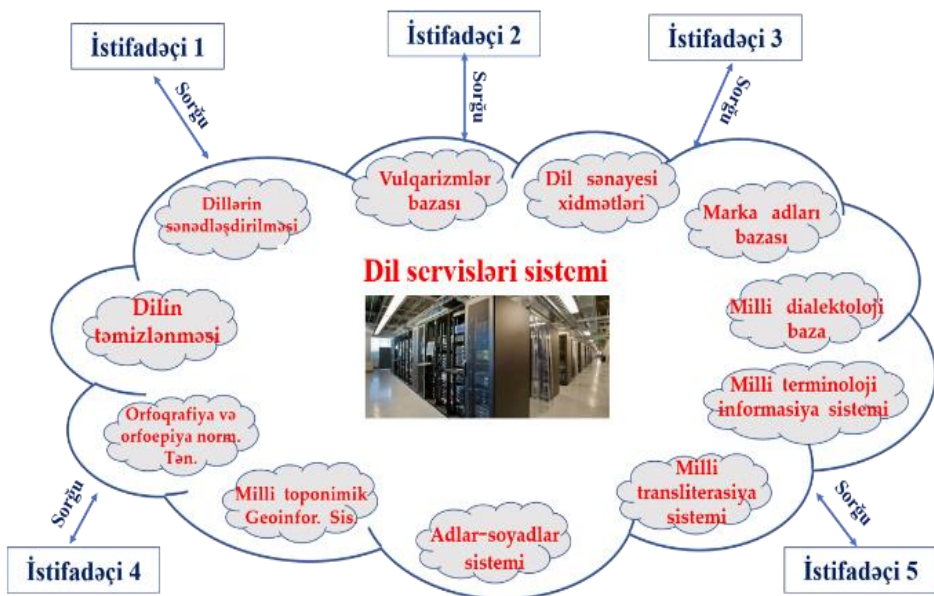
E-dövlət mühitində dil servislərinin yaradılması Azərbaycan dilinin qorunması və inkişafı üçün kifayət qədər vacib məsələlərdən biridir. Beynəlxalq təcrübə nəzərə alınmaqla, e-dövlət sistemində idarəetmənin səmərəliliyinin artırılması, xidmətlərin keyfiyyətinin yüksəldilməsi və xərclərin azaldılması məqsədilə bulud texnologiyalarından istifadə tövsiyə olunur. Bulud əsaslı sistemlərin rahatlığı, xərclərin səmərəliliyi, xidmətlərin inteqrasiyası, müvafiq təhlükəsizlik və etibarlılıq amilləri kimi faydalarına görə onun e-dövlətdə istifadəsi ən yaxşı alternativ ola bilər.

İSTİNADLAR

- [1] G. K. Behara, V. V. Varre and M. Rao, “Service Oriented Architecture for E-Governance,” 2009, www.bptrends.com
- [2] T. Riley, W. Sheridan, “Electronic Government in Developing and Developed Countries: The Basics By Thomas B. Riley and William Sheridan,” Commonwealth Centre for e-Governance, Ottawa, Canada, October 2006.
- [3] Elektron hökumət haqqında ümumi məlumat. <https://www.e-gov.az/az/content/read/2>
- [4] R. M. Əliquliyev, R. M. Alıquliyev, Y. M. Cəfərov, F. F. Yusifov, Ə. M. Qurbanova, “Kosmopolit E-dövlətdən Milli E-dövlətə Doğru: Virtual Mühitdə Azərbaycan Dili Xidmətlərinin Yaradılması Perspektivləri,” İnformasiya cəmiyyəti problemləri, №2, s. 3–25, 2021.
- [5] R. M. Alguliyev, A. M. Gurbanova, “The Conceptual Foundations of National Terminological Information System,” International Journal of Education and Management Engineering, no. 4, pp. 19-30, 2018.
- [6] S. V Məmmədzadə, “Verilənlərin Emalında Transliterasiya və Transkripsiyanın Rolü,” İnformasiya cəmiyyəti problemləri, №1, s. 115–122, 2018.
- [7] R. M. Alguliyev, Y. M. Jafarov, A. M. Gurbanova “Conceptual basis for creation of national toponymic geographic information system,” Problems of Information Society, vol. 13, no. 1, pp. 3–10, 2022.
- [8] R. Alguliyev, R. Alekperov. “Cloud Computing: Modern State, Problems and Prospects,” Telecommunications and Radio Engineering, vol. 73, no. 3. pp. 255-266. 2013.
- [9] T. Diaby, B. B. Rad, “Cloud Computing: A review of the Concepts and Deployment Models,” International Journal of Information Technology and Computer Science, vol. 9, no .6, pp. 50–58. 2017.
- [10] S. Hashemi, K. Monfaredi, M. Masdari, “Using Cloud Computing for E-Government: Challenges and Benefits,” World Academy of Science, Engineering and Technology, International Journal of Computer, Information, Systems and Control Engineering, vol. 7, no. 9, 2013.



Şəkil 1. Dil servisləri sisteminin infrastrukturunu



Şəkil 2. Bulud əsaslı dil servislərinin konseptual modeli

Analysis of texts features for the author recognition system

Kamil Aida-zade¹, Rustam Azimov²

^{1,2}Institute of Control Systems of the Ministry of Science and Education of the Republic of Azerbaijan, Baku, Azerbaijan

¹kamil_aydazade@rambler.ru, ²rustemazimov1999@gmail.com

Abstract — In the study a comparative analysis of the recognition effectiveness of the use of text features used in the author recognition computer system in combination with different methods and models of machine learning to recognize the authors of literary works in the Azerbaijani language was carried out.

Keywords — *authorship recognition of texts; authorship attribution; authorship identification; machine learning*

I. INTRODUCTION

Text authorship recognition has actual applications, such as the authorship of unattributed fictional works, as well as authorship identification of e-mails, texts on online forum pages, and blog posts.

Recognition of the authorship of texts is understood as selection of the author of a given text among certain author candidates based on the texts of those author candidates. For this, text samples known to be written by those author candidates are used [1]. Various methods and models of machine learning were used to establish the relationship between the features of these text samples and their authors.

After T.C. Mendenhall showed that the number of words of certain lengths (e.g.

number of 1-letter words, number of 2-letter words, etc.) remained invariant within the text of C.J.H. Dickens' Oliver Twist, he compared the features of the texts of Shakespeare's and Bacon's works [2], [3]. A.A. Markov used the frequency of occurrence of vowel-vowel, consonant-vowel, and so on in A.S. Pushkin's "Eugene Onegin" in his work devoted to an application of statistics [4]. Mosteller and Wallace used frequencies of selected words to identify the authors of twelve unknown authors among the Federalist papers [5].

In another study, recognition of the authorship of texts was carried out on 634 fictions by 55 authors [6]. However, as a result of the increase in Internet resources, in recent years there have been more works devoted to authorship identification of short, non-fiction texts such as articles in newspaper columns [7]-[9].

Statistical methods, machine learning methods and models were used to recognize the authorship of texts in the Azerbaijani language [10]-[13].

In the study based on the results of computer experiments, a comparative analysis of the effectiveness of a number of text features and several machine learning methods and models that can be used to recognize the authorship of texts in the

Azerbaijani language on an example of artistic works was conducted.

II. PROBLEM STATEMENT

In this study, the problem authorship recognition was considered on an example of the works of several famous Azerbaijani writers. A comparative analysis of the recognition effectiveness of the use of different feature sets with different methods and models of machine learning for use in a computer system that recognizes the authorship of texts in the Azerbaijani language was carried out.

A. Used Dataset

The dataset (dataset-0) consisting of 23 large-volume, 128 small-volume, total 151 works of 11 Azerbaijani writers was used in this work. Each of the large-volume works was divided into 10 parts (with almost equal number of characters), which were considered as separate texts, thus another dataset (dataset-1) consisting of $23 \times 10 + 128 = 403$ texts was prepared. The observations in this dataset-1 are split into training and test sets on 80-20 ratio (325 and 78 observations).

The recognition models were trained with the training set obtained from dataset-1, and their accuracy was evaluated on both the test set obtained from dataset-1 and dataset-0 itself. At the same time, accuracies were also evaluated on the test base obtained from dataset-1 with a certain confidence (recognition is rejected if a certain degree of confidence is not obtained, even if the recognition result is correct for that observation, it is considered wrong).

B. Used Text Features

Here, as text features, the frequencies of letter n-grams (a letter n-gram is a combination of n letters), the frequencies of words of a certain length (for example, the frequency of 5-letter words, the frequency of 6-letter words, etc.), the frequencies of sentences of a certain length (for example, the frequency of 5-word sentences, the frequency of 6-word sentences, etc.), letter n-gram variance features, which express how letter n-grams are distributed within a text (a given text is divided into parts, n-gram frequencies are calculated in each part, variance of them are found) were used. The feature sets consisting of different numbers of features selected from these features are described in Section III.

C. Used Machine Learning Methods and Models

Artificial Neural Network (ANN), Support Vector Method (SVM) and Random Forest (RF) were used in the study. Radial Basis Function was used as kernel in SVM. The characteristics of ANN architectures were defined according to the number of features. One set of letter 2-grams has also been used with a Convolutional Neural Network (CNN) in the form of a two-dimensional matrix.

III. DESCRIPTION OF USED FEATURE SETS AND INFORMATIVE FEATURES SELECTION ALGORITHM

A. Used Feature Sets

From the features of the frequencies of sentences of a certain length and the

frequencies of words of a certain length (let's call them sentence frequencies and word frequencies for short), two feature sets were made each. In one of the sentence frequency set, there are frequencies of 5-word, 6-word, ..., 14-word sentences, and in the other, there are frequencies of 5-word, 6-word, ..., 29-word sentences. One of the word frequency feature sets contains the frequencies of 3-letter, 4-letter, ..., 7-letter words, and the other has the frequencies of 3-letter, 4-letter, ..., 12-letter words. The frequencies of 5-word, 6-word, ..., 14-word sentences and the frequencies of 3-letter, 4-letter, ..., 12-letter words were also used.

Letter 1-grams and selected letter 2-grams (The selection of informative letter 2-grams is described in the next paragraph) feature groups were also used. Seven feature sets were obtained by choosing 5, 10, ..., 25, 50, 100 letter 2-grams from the possible letter 2-grams. Feature sets consisting of the variances of frequencies of letter 1-grams and selected 2-grams in separate parts of the given text were also created. Several mixed feature sets consisting of letter n-grams and their within-text variances were also used (4 feature sets with 10, 20, 30, 40 features using 5, 10, 15, 20 features from each of these feature classes).

B. Selection of Informative Letter bigrams

For the selection of informative letter 2-grams, all the texts in the training database were combined into one text, the frequencies of all possible letter 2-grams in this single text were calculated, and the high-frequency 2-grams were selected.

IV. RESULTS OF COMPUTER EXPERIMENTS

The maximum recognition accuracies obtained on dataset-0 when using different models and methods of machine learning with different feature sets are given in Table I.

TABLE I. RECOGNITION ACCURACIES OF FEATURE SETS

Feature classes	ANN	CNN	SVM	RF
n-gram	31.13%	44.3%	93.38%	96.69%
variation	31.13%	-	39.74%	68.21%
n-gram + variation	9.93%	-	58.94%	85.43%
sentence frequencies	42.38%	-	65.56%	78.81%
word frequencies	5.30%	-	56.95%	74.83%
sentence-word frequencies	15.89%	-	66.89%	86.75%

CONCLUSION

In this study based on the results of computer experiments, a comparative analysis of the effectiveness of different methods and models of machine learning to recognize the use of different feature sets, composed of different numbers of features belonging to different classes of features for use in a system that recognizes the authorship of texts on an example of the fictional works of several Azerbaijani authors, is given.

The feature sets consisting of letter n-grams outperformed other features in recognition accuracy. Although Random Forest gives the highest recognition accuracy among the models and methods,

more reliable results were obtained with Support Vector Machine.

REFERENCES

- [1] E. Stamatatos, "A survey of modern authorship attribution methods," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 3, pp. 538-556, 2009, doi: 10.1002/asi.21001.
- [2] T. C. Mendenhall, "The Characteristic Curves of Composition," *Science*, no. 214, pp. 237-246, 1887, doi: 10.1126/science.ns-9.214s.237.
- [3] Mendenhall, "A Mechanical Solution of a Literary Problem," *The Popular Science Monthly*, no. 60, pp. 97-105, 1901.
- [4] A. A. Markov, "An Example of Statistical Investigation of the Text *Eugene Onegin* Concerning the Connection of Samples in Chains," *Science in Context*, vol. 19, no. 4, pp. 591-600, 2006, doi: 10.1017/s0269889706001074.
- [5] F. Mosteller and D. L. Wallace, "Inference in an Authorship Problem," *Journal of the American Statistical Association*, vol. 58, no. 302, pp. 275-309, 1963, doi: 10.1080/01621459.1963.10500849.
- [6] Y. Zhao, J. Zobel, "Searching with style: Authorship attribution in classic literature," *In Proceedings of the thirtieth Australasian conference on Computer science*, vol. 62, pp. 59-68, January 2007.
- [7] J. Diederich, J. Kindermann, E. Leopold, G. Paass, "Authorship attribution with support vector machines," *Applied intelligence*, vol. 19, no. 1, pp. 109-123, 2003.
- [8] I. Erdoğan, M. Güllü, H. Polat, "Development of a Letter Recognition Application with Machine Learning Algorithms," *El-Jezeri*, vol. 9, no. 4, pp. 1303-1314, 2022.
- [9] V. Levent, B. Diri, "Author Recognition in Turkish Documents with Artificial Neural Networks," *Academic Informatics*, vol. 14, pp. 5-7, 2014.
- [10] K. R. Ayda-zade, S. G. Talibov, "Analysis of the Methods for the Authorship Identification of the Text in the Azerbaijani Language," *Problems of information technology*, vol. 8, no. 1, pp. 14-23, 2017.
- [11] R. B. Azimov, E. M. Mustafayev, "Comparison of SVM and ANN methods for recognition of authorship of texts". *Applied Mathematics and Fundamental Informatics: Proceedings of the XII Intern. youth scientific-practical. conf. with elements of science. schools*, pp. 60-61, May 2022.
- [12] K. R. Aida-zade, E. M. Mustafayev, R. B. Azimov, "Features analysis for application in a computer recognition systems of Azerbaijani texts authorship," *Second International Bilateral Workshop on Science Between Dokuz Eylül University and Azerbaijan National Academy of Sciences*, p.11, November 2022.
- [13] E. M. Mustafayev, R. B. Azimov, "Comparative analysis of different feature groups for use in a computer system for recognizing authors of texts in the Azerbaijani language," *II Republican scientific conference on "Fundamental problems of mathematics and application of intellectual technologies in education"*, pp. 34-39, December 2022.

Azərbaycan dili üçün söz vektorları və böyük dil modelləri

Ümid Süleymanov¹, Samir Rüstəmov², Abzətdin Adamov³

^{1,2,3}ADA Universiteti, Bakı, Azərbaycan

¹usuleymanov@ada.edu.az, ²srustamov@ada.edu.az, ³aadamov@ada.edu.az

Xülasə — Son illərdə söz vektorları və əvvəlcədən öyrədilmiş dil təmsil sistemlərinin uğuru təbii dilin işlənməsinə marağı xeyli artırmışdır. Onlar suallara cavab, tərcümə sistemləri, əhval-ruhiyyənin təhlili və s kimi Təbii Dilin Emalı tapşırıqlarında nəzərəçarpan irəliləmələrin əsasını təşkil edir. Söz vektorları üzərində tezis, metodların sürətlə artması nəzərə alınmaqla, az resurslu aqqlütinativ dil olan Azərbaycan dili üçün xüsusilə aktualdır. Bu tezis söz vektorlarının müxtəlif arxitektura yanaşmalarından istifadə edilməklə hazırlanmasını hədəfləyir və söz əlavələrinin effektivliyi hissələrin təhlili, mətn təsnifatı, semantik analogiya, sintaktik analogiya daxil olmaqla, müxtəlif təbii dil emal tapşırıqlarını ehtiva edən müxtəlif məlumat dəstləri üzərində empirik tədqiqatlar vasitəsilə araşdırılacaqdır.

Açar sözlər — söz vektorları; sentiment analizi; təbii dilin emalı; statik vektorlar

I. GİRİŞ

İnsan intellekti ilə təbii dil arasında əlaqənin mövcudluğuna inanılması, təbii dil emalı sahəsini tədqiqat üçün ən maraqlı sahələrindən birinə çevirir. Universal süni intellekt modellərini inkişaf etdirməzdən əvvəl, təbii dilin emalı ilə bağlı problemləri həll etmək çox vacibdir, çünki yalnız təbii dilin emalı ünsiyyət vasitəsilə, ağıllı

sistemlər özlərini ən effektiv şəkildə ifadə edə bilirlər. Hətta dahi alim və kompüter elmlərinin banilərindən biri olan Alan Turing özünün məşhur Turing testində [1] sorğu aparan şəxsin divarın arxasında maşınla, yoxsa insanla danışdığını müəyyən edə bilmirsə, maşının zəkaya malik olmasını təklif etmişdi. Bu, süni intellekt sahəsində təbii dilin emalı və generasiyasının əhəmiyyətinin daha bir sübutudur. Son zamanlardakı öncədən öyrədilmiş təbii dilin emalı modellərinin böyük uğurları da sahədəki tədqiqatı daha da zəruri edir.

Təbii dilin neyron təsvirində inqilab edən və onun müxtəlif sahələrində bir sıra irəliləyişlər yaradan söz vektorlarının yaradılması və istifadəsi bu tezisə əsas diqqət mərkəzində olacaqdır. Bu araşdırmanın məqsədi müxtəlif qabaqcıl metodlardan və maşın öyrənmə üsullarından istifadə edərək söz vektorları yaratmaqdır. Stabil söz vektorlarının yaradılması üçün publisistik, elmi və ədəbi üslubda yazılmış xəbər məqalələri, kitablar, elmi məqalələr, şeirlər və romanlar kimi müxtəlif mənbələrdən milyonlarla məlumat korpusu toplanmışdır. Söz vektorlarının effektivliyinin ölçülməsi üçün 2 yanaşma mövcuddur; intrinsik və ekstrinsik. Intrinsik yanaşmalarda [2] söz vektorlarının effektivliyi daxili analogiya tapşırığında ölçülür, ekstrinsik yanaşmada [3][4] isə

vektorlar hər hansı xarici təbii dilin emalı tapşırığında yoxlanılır.

II. ƏLAQƏLİ ƏDƏBİYYAT

Təbii dilin təsviri üsullarının müxtəlif aspektlərini əhatə edən zəngin ədəbiyyat mövcuddur. Collados və Pilehvar (2018) [5] müasir təbii dilin neyron şəbəkə üsullarından istifadə metodlarının xülasəsini təqdim edir və dil modelləşdirmə sistemlərinin sözün mənasını təmsil etməyə çalışarkən qarşılaşdığı çətinlikləri təqdim edir. Nəzarətsiz təlim vasitəsilə öyrənilən söz təmsilləri və nəzarətli təlimə əsaslanan təmsillər, texnikaları təmin etdikləri iki əsas kateqoriyadır. Collados və Pilehvar tərəfindən aparılan araşdırmada, həmçinin söz vektorlarının keyfiyyətinin ölçülməsi üçün qiymətləndirmə üsulları da müzakirə olunur (2018). Onlar tədqiqatı söz təmsillərinin istifadəsini təsvir etməklə və onları çoxlu faktorlar, o cümlədən mənanın kontekstə [6] uyğunluğu, şərh oluna bilməsi və s. müxtəlif kontekstlərdə tətbiqi baxımından qiymətləndirməklə yekunlaşdırırlar.

Mövcud dərin öyrənmə yanaşmaları və onların Təbii Dil Emalına təsirlərinin icmalı Otter və digərləri [7] tərəfindən özlərinin tədqiqat işlərində təqdim olunur. (2020). Təbii dilin emalının geniş icmalı və mövcud dərin öyrənmə arxitekturalarının araşdırılması işin birinci hissəsini təşkil edir. Təbii dilin emalı və dərin öyrənmə indi iki böyük mövzu olsa da, ayrı-ayrılıqda fərqli sahələr deyillər, onların kəsişməsini daha ətraflı araşdırmaq daha yaxşı olardı. Birinci hissədə dərin öyrənmə arxitekturasının izahı tez tamamlana bilərdi. Aşağıdakı bölmə dilin modelləşdirilməsi, morfologiyası,

təhlili və semantikasını daxil olmaqla [7] təbii dil emalının fundamental sahələrində dərin öyrənmənin istifadəsinə həsr edilmişdir (Otter və digərləri, 2020).

Peter və Pantel sözlərin semantik mənasının təmsil olunmasına diqqət yetirən əlavə bir araşdırma [8] aparmışlar (2010). Bu araşdırma ona görə fərqlənir ki, o, mövcud strukturları və metodları geniş təbii dil emal fərziyyələri ilə əlaqələndirməyə çalışır. Fərziyyələrin bir neçə nümunəsinə statistik semantik fərziyyə, sözlər paketi hipotezi, paylanma fərziyyəsi və başqaları daxildir. Nəticədə, sahəyə yeni daxil olan tədqiqatçılar bu araşdırmanı və onun araşdırdığı arxitekturaları başa düşməyi olduqca asan tapacaqlar. Tədqiqata linqvistik və riyazi baxımdan söz vektorlarının işlənməsinə dair ayrıca fəsillər də daxil edilmişdir.

III. METODOLOGİYA

Sözlərin yerləşdirilməsi üsulları son on il ərzində təbii dil emalı sahəsini kəskin şəkildə dəyişdirdi. Söz yerləşdirmələri müxtəlif maşın öyrənmə alqoritmlərini insanın təbii dilinin son dərəcə real təsvirini təmin edir. Tədqiqatçılar üçün əlçatan olan yaxşı formalaşmış korporasiyanın olmaması səbəbindən Azərbaycan dili aşağı resurs səviyyəsinə malikdir. Bundan əlavə, bu sintaktik və semantik cəhətdən mürəkkəb dilə yönəlmiş çox tədqiqat yoxdur. Azərbaycan dili aqlütinativ dildir, yəni sözlər və qrammatik cümlələr, adətən prefiks və postfiksədən istifadə etməklə əmələ gəlir. Təkcə bu fakt bu dil üçün söz yerləşdirmə konsepsiyasının effektivliyinin öyrənilməsini daha da münasib edir və bütün dünya üzrə təbii dil emalı

tədqiqatçıları və mütəxəssisləri üçün araşdırılması vacib bir mövzuya çevirir.

Sözlərin yerləşdirilməsi insan təbii dilinin sıx vektor təsvirlərini yaradır ki, bu da bir çox təbii dilin işlənməsi tapşırıqlarına, o cümlədən suallara cavab vermə, əhval-ruhiyyənin təhlili, mətnin təsnifatı və digərləri üçün uğurla tətbiq edilir və tez-tez, hətta bu tapşırıqlarda yeni ən müasir performansla nail olur. Bu, artıq danılmaz faktır. Bununla belə, tədqiqatların əksəriyyəti ingilis, fransız, alman və başqaları kimi dil ailələri üzərində cəmləşib və bu kontekstlərdə söz əlavələrinin tapşırıqlar üzərində effektivliyini nümayiş etdirir. Bu da öz nüvbəsində aşağıdakı tədqiqat məsələsini ortaya qoyur: “Sözün yerləşdirilməsi yanaşmaları Azərbaycan dili kimi aqqlütinativ dili uğurla təmsil edirmi?”

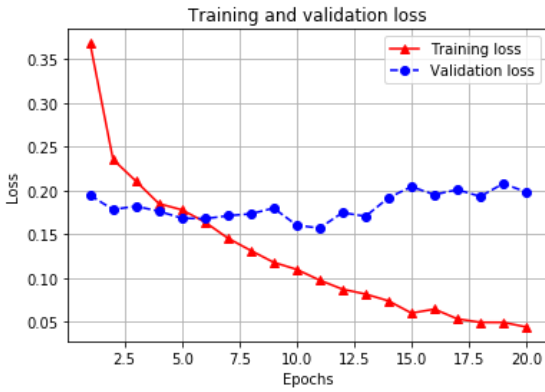
Biz bu tədqiqat problemini həll etmək üçün həm daxili, həm də xarici qiymətləndirmə tapşırığı parametrlərində söz daxiletmələrinin nəticələrini kəmiyyətləndirməyə və qiymətləndirməyə çalışacağıq. Bunun üçün söz daxiletmələrindən istifadə edərək bir sıra təcrübələr aparılacaqdır. *Word2vec* alqoritminin tərtibatçıları Mikolov və digərləri [9] tərəfindən məşhur şəkildə təklif olunan bu tapşırıqlardan söz daxiletmələrinin effektivliyini statistik qiymətləndirmək üçün istifadə etmək geniş şəkildə qəbul edilir. Nəzərə alın ki, bu tapşırıqların ayrıca Azərbaycan dilində versiyası hazırlanmışdır. Bu daxili qiymətləndirmə tapşırıqlarında həm semantik, həm də sintaktik bənzətmə tapşırıqlarından istifadə edilir və sözün

yerləşdirilməsinin düzgünlüyü qiymətləndirilir. Xarici qiymətləndirmə tapşırıqları üçün əhval-ruhiyyənin təhlili və mətn təsnifatı tapşırıqlarında sözlərin yerləşdirilməsinin effektivliyi göstəriləcək və dəqiqlik reytingləri göstəriləcək. Təcrübələrdən sonra bu aqqlütinativ dil üçün söz əlavələrinin effektivliyi müzakirə ediləcək və verilən tədqiqat sualına empirik nəticələr tapılmağa çalışılacaq.

FastText arxitekturası N-qram simvollarından istifadə edərək, altsöz məlumatlarının uçotunu təmin edir. Nəzərə alsaq ki, Azərbaycan dili aqqlütinativ dildir, bu, xüsusilə faydalıdır. Ən çox sintaktik qarşılıqlı əlaqənin altsöz səviyyəsində baş verdiyini nəzərə alsaq aqqlütinativ dildə altsöz strukturlarına müraciət daha da vacib olur. Nəticədə, bu arxitekturanın altsöz səviyyəli qatları, sintaktik düzgünlüyün daxili qiymətləndirilməsini tələb edən tapşırıqlarda daha yaxşı çıxış edəcəyini təxmin edə bilərik. Üç fərqli məlumat dəstində öyrədilmiş müxtəlif ölçülü söz vektor ölçüləri ilə biz aşağıdakı cədvəldə fastText arxitekturası üçün hərtərəfli dəqiqlik reytinglərini veririk. Ən yüksək ballar hər bir məlumat dəsti ölçüsü qruplaşması üçün qalın götürülür və altından xətt çəkilir. Modelin öyrəndiyi söz vektor ölçülərinin sayı vektor ölçüsü ilə təmsil olunur.

Söz yerləşdirmə arxitekturaları tərəfindən yaradılan söz vektorları hissələrin təhlili tapşırığına tətbiq edilir. Maşın öyrənmə arxitekturasının qiyməti söz vektorlarının keyfiyyətini əks etdirir. Sözün yerləşdirilməsi vektoru nə qədər yaxşı olarsa, model xarici qiymətləndirmə

tapşırığını bir o qədər yaxşı öyrənə bilər. Biz dondurulmuş yerləşdirmə vektorlarını sentiment təhlili tapşırığında yerləşdirmə təbəqəsi kimi istifadə etdik və maşın öyrənmə arxitekturasını öyrətdik. Maşın öyrənmə arxitekturası Yerləşdirmə Layeri, nizamlanma üçün Dropout qatı, 96 x 64 Convolution Layer, MaxPooling Layer, LSTM qatı və tam əlaqəli sıx təbəqədən ibarətdir. Arxitektura ümumilikdə 33 225 965 parametərə malikdir.



Şəkil 1. Öyrədilmə dəyərinin aşağı enməsi

Dövrələrin funksiyası kimi model arxitekturasının təlim və təsdiqləmə itkisi yuxarıdakı şəkildə göstərilmişdir. Göründüyü kimi, dövrlərin sayı artdıqca məşq itkisi azalır, lakin doğrulama itkisi 0,20 ətrafında statik nümunə nümayiş etdirir. Bu, modelin məlumat dəstindən kifayət qədər öyrəndiyi üçün 20 dövr sayının kifayət olduğunu göstərir. Beləliklə, daha çox dövrlər əlavə etmək həddindən artıq uyğunlaşma ilə nəticələnəcəkdir. Model təsdiqləmə məlumatlarını görmədiyini üçün təlim itkisindən daha böyük qiymətləndirmə itkisi adı haldır, buna görə

də doğrulama itkisi qaçılmaz olaraq təlim itkisindən daha yüksəkdir. Model təlim dəstindəki nümunəni uğurla öyrəndi, bunu təsdiqləmə və təlim itkisi üçün yaxın dəyərlər sübut edir.

İSTİNADLAR

- [1] Turing, A.M. (1950), 'Computing Machinery and Intelligence', *Mind* 59, pp. 433–460
- [2] Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT 2019*, pages 4171–4186.
- [3] Mikolov, T., Yih, W.-t., & Zweig, G. (2013d). Linguistic regularities in continuous space word representations. In *HLT-NAACL*, pp. 746–751.
- [4] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013c). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pp. 3111–3119.
- [5] Jose Camacho-Collados & Mohammad Taher Pilehvar. (2018). From word to sense embeddings: a survey on vector representations of meaning. *J. Artif. Int. Res.* 63, 1 (September 2018), 743–788. DOI:<https://doi.org/10.1613/jair.1.11259>
- [7] Joseph Turian, Lev Ratinov, and Yoshua Bengio. (2010). Word representations: a simple and general method for semi-supervised learning. In *Proceedings of ACL*, pages 384–394
- [6] Melamud, O., Goldberger, J., & Dagan, I. (2016). context2vec: Learning generic context embedding with bidirectional lstm. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pp. 51–61, Berlin, Germany
- [7] D. W. Otter, J. R. Medina and J. K. Kalita, (2020). "A Survey of the Usages of Deep Learning for Natural Language Processing," in *IEEE Transactions on Neural Networks and*

- Learning Systems, doi: 10.1109/TNNLS.2020.2979670
- [8] Turney Peter D. & Pantel Patrick (2010). From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, 37, 141- 188.
- [9] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.
- [10] Adams Wei Yu, David Dohan, Minh-Thang Luong, Rui Zhao, Kai Chen, Mohammad Norouzi, and Quoc V Le. (2018.) QANet: Combining local convolution with global self-attention for reading comprehension. In *ICLR*. 71
- [11] Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International Conference on Machine Learning*, pages 1188– 1196
- [12] Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of NAACL-HLT 2019*, pages 4171–4186.
- [13] Qing Cui, Bin Gao, Jiang Bian, Siyu Qiu, Hanjun Dai, and Tie-Yan Liu. (2015). KNET: A general framework for learning word embedding using morphological knowledge. *ACM Transactions on Information Systems*, 34(1):4:1–4:25.
- [14] Eric H. Huang, Richard Socher, Christopher D. Manning, and Andrew Y. Ng. (2012). Improving Word representations via global context and multiple word prototypes. In *Proceedings of Association for Computational Linguistics*, pages 873–882.
- [15] John Duchi, Elad Hazan, and Yoram Singer. (2011). Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *J. Mach. Learn. Res.* 12, (2/1/2011), 2121–2159
- [16] [13] Melamud, O., McClosky, D., Siddharth Patwardhan, and Mohit Bansal. 2016. The role of context types and dimensionality in learning word embeddings. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1030-1040.
- [17] Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. In *Proceedings of NAACL, New Orleans, LA, USA*.

Issues of speech technologies application in the Azerbaijani language

Rasim Alguliyev¹, Lyudmila Sukhostat²

^{1,2}Institute of Information Technology, Baku, Azerbaijan

¹r.alguliev@gmail.com, ²lsuhostat@hotmail.com

Abstract — With the development of state-of-the-art technologies, there is a need for convenient and natural communication between a person and a computer. The literature on automatic speech recognition has used many methods for various languages, including machine learning methods for speech analysis and classification. In terms of speech technologies, namely, speech synthesis and recognition, speaker identification, dialect recognition, and speaker accent, little research has been done on the Azerbaijani language. This paper analyses the issues of using speech technologies in the Azerbaijani language. The emergence of computers with artificial intelligence that can understand the Azerbaijani language will contribute to a new experience and a comprehensive intellectualization of human life.

Keywords — *speech signal; Azerbaijani language; automatic speech recognition*

I. INTRODUCTION

Speech is a complex signal resulting from several transformations occurring at various information levels: semantic, linguistic, articulatory, and acoustic. Speech is the easiest and most convenient way to

exchange information and the most natural way of interaction between people.

The automatic speech recognition system helps us with this. Such a system allows a computer to take an audio file or direct speech from a microphone as input and convert it to text, preferably in spoken language script [1].

The main speech technologies include speech synthesis, accent recognition, speech and language recognition, emotion recognition, gender detection, and speaker recognition (Fig. 1).

Recent advances in speech recognition create a dynamic environment as this technology allows to:

- access electronic services via voice without being distracted by the screen of the control device (keyboard, mouse, etc.);
- increase the security of personal data by using voice authentication;
- simplifies the input of search queries;
- make web resources accessible to people with disabilities.

Human speech, dialects, and accents have tremendous variety, and this variation in speech patterns is one of the biggest hurdles in building an autonomous speech recognition system.

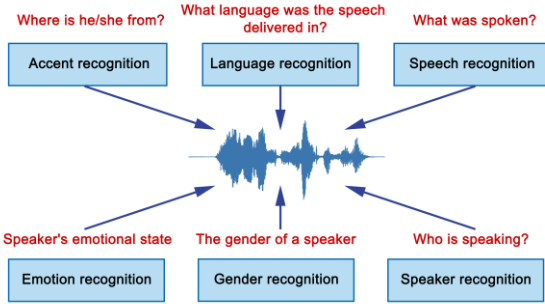


Fig. 1. The different speech tasks.

Furthermore, the problem arises when we add various factors like gender, style, and speed of speech. Another obstacle to creating an automatic speech recognition system is the presence of a large speech corpus, particularly for the Azerbaijani language. Along with the use of various methods and algorithms, it is necessary to consider the features of natural language in developing these systems. In this regard, at the moment, there are only some learning models for a small number of languages [2].

II. AUTOMATIC SPEECH RECOGNITION

As one of ICT, modern speech technologies have unique capabilities and are widely used in such areas as e-medicine, forensic phonetics, distance learning, and others.

Speech technologies using Big Data, the Internet of things, and cloud technologies

contribute to convenient and natural communication between people and a computer. Various methods have been proposed to solve the problems of automatic language identification based on machine learning methods, including deep neural networks [3].

The main stages of state-of-the-art automatic speech recognition systems are shown in Fig. 2.

A. Feature extraction

The feature extraction process is used to remove unnecessary information from the signal. A good feature extraction algorithm should be able to extract features in real time and contain as much information as possible. Feature extraction algorithms can also be classified based on speech features into temporal and spectral features. Time analysis methods analyse the audio signal in its original form in the time domain. The spectral analysis uses the spectral representation of a speech signal, i.e., the frequency domain. Some of the techniques used for feature extraction are MFCC (Mel-frequency cepstral coefficients), PLP (perceptual linear prediction), DWT (discrete wavelet transform), Relative Perceptual Spectral Linear Prediction (RASTA-PLP), and LPC (linear prediction coding).

B. Classification

After extracting the features, they are passed to the classifier's input. The task of the classifier is to study the relationship between the given sound input characteristics and the corresponding text. First, they are trained using a large enough

dataset to recognize specific patterns in the speech signal. The most commonly used classifiers are HMM (hidden Markov model), ANN (Artificial neural network), and SVM (support vector machine) [3].

C. Language model

An efficient language model is needed to create a real-time automatic speech recognition system. It uses the structural constraints of the language to predict the probability of a word occurring for a particular sequence of words. The difference between a classifier and a language model is that the classifier matches speech signals to the closest possible sequence of words. In contrast, the language model tests the probability of occurrence of the sequence of words generated by the classifier. Since in all world languages, some phrases sound the same but with different meanings. It can be fixed using a language model in conjunction with a classification model.

D. Acoustic model

Acoustic modeling analyses the training data in terms of relevant characteristics, for example, using various possibilities, expressing them as probabilities, and then combining these probabilities into a model, for example, HMM [4]. Various types of acoustic models have been used for automatic speech recognition, such as GMM (Gaussian mixture model), DNN (deep neural network), and LSTM (long short-term memory) in combination with HMM, and CNN (convolutional neural network) [5].

E. Decoding

Feature vectors are decoded into the linguistic units that make up speech using acoustic models derived from the recordings and their corresponding transcripts. Linguistic and pronunciation knowledge is often used to improve decoding performance.

III. PROBLEMS OF AUTOMATIC SPEECH RECOGNITION

Automatic speech recognition systems differ in the type of speaker, style of speech, and vocabulary size.

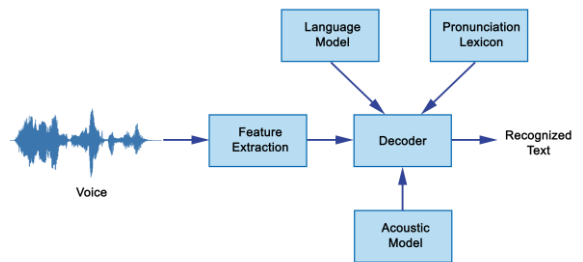


Fig. 2. Typical automatic speech recognition process.

They can be speaker-dependent, speaker-independent, and speaker-adaptive (Fig. 3). The efficiency of an automatic speech recognition system's efficiency depends on the dictionary's size.

The main problems of speech recognition are as follows:

- the same speech fragments have different characteristics and different duration;

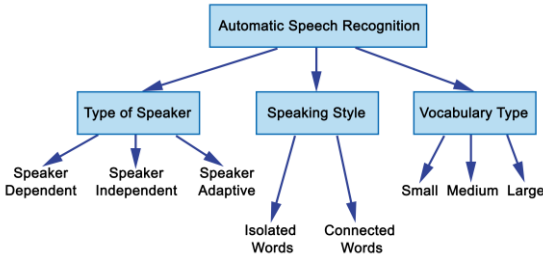


Fig. 3. Speech recognition systems classification.

- faulty and unclear pronunciation;
- poor diction of the speaker;
- high level of extraneous noise;
- insufficient or poor training of AI models;
- significant similarity of dictionary words;
- pronunciation with different intonation;
- the grammar used by the speaker and received by the system, the position of the microphone, and the user's speech speed;
- accent and dialect of the speaker.

IV. SPEECH TECHNOLOGIES IN THE AZERBAIJANI LANGUAGE

One of the main challenges for state-of-the-art research on speech production and speech technologies is the understanding and modeling of individual variations in spoken language. People have their speaking styles, depending on many factors. These differences demonstrate the difficulty of modeling large-scale speaker-independent systems designed to process information in a particular language. People

learn over the years, to some extent, to identify and interpret most of these speech aspects.

Over the past few decades, significant progress has been made in automatic speech recognition for various languages based on a given speech pattern [6, 7].

The Azerbaijani language belongs to the Turkic group of languages. This language group has more than 50 languages.

Azerbaijan is currently developing automatic speech recognition systems using the most modern methods [2, 8, 9].

Deep learning is fast becoming the method of choice over traditional automatic speech recognition methods. In addition, most research is developing toward multimodal and unsupervised speech recognition and Natural Language Processing (NLP) [9, 10].

CONCLUSION

Voice recognition of the language ensures its entry into all modern machine learning and artificial intelligence systems. Recognition, processing, and translation of speech require complex engineering and linguistic solutions. This paper considered the issues of using speech technologies in the Azerbaijani language.

Summing up, the following priority areas can be identified:

- Development of large speech corpora.
- Development of continuous speech recognition technologies, including ready-made acoustic-phonetic models.

- Creation of a program complex for restoring distorted and noisy speech messages in both a limited (thematic) dictionary and a mixed one.
- Creation of interpreters that correctly convey the meaning of a highly distorted speech message.
- Automated detection of linguistic information in texts and speech messages for biometric control.

REFERENCES

- [1] M. Malik, M. K. Malik, K. Mehmood, and I. Makhdoom, “Automatic speech recognition: A survey,” *Multimed. Tools Appl.*, vol. 80, pp. 9411-9457, 2021.
- [2] K. Aida-zade, S. Rustamov, E. Mustafayev, and N. Aliyeva, “Human-computer dialogue understanding hybrid system,” in *Proc. IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*, July 2012, pp. 1–5.
- [3] G. Singh, S. Sharma, V. Kumar, M. Kaur, M. Baz, and M. Masud, “Spoken language identification using deep learning,” *Comput. Intell. Neurosci.*, vol. 2021, Article ID 5123671, pp. 1-12, 2021.
- [4] L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proc. IEEE*, Vol. 77, pp. 257-286, 1989.
- [5] A. Pervaiz, F. Hussain, H. Israr, M. A. Tahir, F. R. Raja, N. K. Baloch, F. Ishmanov, and Y. B. Zikria, “Incorporating noise robustness in speech command recognition by noise augmentation of training data,” *Sensors*, Vol. 20, pp. 1-19, 2020.
- [6] M. Biswas, S. Rahaman, A. Ahmadian, K. Subari, and P. K. Singh, “Automatic spoken language identification using MFCC based time series features,” *Multimed. Tools Appl.*, pp. 1-31, 2022.
- [7] Y. Zhang, et al. “BigSSL: Exploring the frontier of large-scale semi-supervised learning for automatic speech recognition,” *IEEE J. Sel. Topics in Signal Proc.*, Vol. 16, pp. 1519-1532, 2022.
- [8] A. Abbasov, R. Fatullayev, and A. Fatullayev, “Speech technology market in Azerbaijan,” *American J. Manag.*, Vol. 21, pp. 95–101, 2021.
- [9] M. Mahmudov, R. Fatullayev, S. Abbasov, A. Fatullayev, and N. Abdullayev, “NLP systems for the Azerbaijani language and theoretical and practical tasks of creating the national corpuses,” *Turkology*, No. 4, pp. 15-28, 2016.
- [10] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, “Speech emotion recognition using deep learning techniques: A review,” *IEEE Access*, Vol. 7, pp. 117327–117345, 2019.

Rəqəmsal iqtisadiyyat və transformasiyalar üzrə termin və anlayışlar bazasının formalaşdırılması problemləri

Ələvsət Əliyev¹, Roza Şahverdiyeva²

^{1,2} İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

¹alovsat_qaraca@mail.ru, ²shahverdiyev@ gmail.com

Xülasə — Məqalə rəqəmsal iqtisadiyyat və transformasiyalar üzrə termin və anlayışlar bazasının formalaşdırılması problemlərinə həsr olunmuşdur. İqtisadiyyatın və cəmiyyətin rəqəmsal transformasiyası prosesində müxtəlif sahələrdə rəqəmsal iqtisadiyyat və transformasiya texnologiyalarının, Sənaye 4.0 platformasının tətbiqi əsasında Azərbaycan dilində formalaşdırılan termin və anlayışlar bazasının yaradılmasının aktuallığı əsaslandırılmışdır. Rəqəmsal iqtisadiyyatda milli kontentin inkişaf etdirilməsinin xüsusi əhəmiyyəti qeyd olunmuşdur. Rəqəmsal iqtisadiyyat və transformasiyanın (RİT) mahiyyəti, məqsədi və vəzifələri izah olunmuşdur. Rəqəmsal iqtisadiyyatın ÜDM-ə təsir aspektləri şərh olunmuş, rəqəmsal iqtisadiyyatda İKT infrastrukturunu və xidmətlərinin xüsusiyyətləri araşdırılmışdır. Azərbaycandilli proqram təminatlarının hazırlanması və geniş istifadəsinin təmin olunması istiqamətləri tədqiq olunmuşdur. RİT sferasında Azərbaycan dilindən global istifadə problemləri şərh olunmuşdur. Həmin sahədə izahlı lüğətlərin işlənilməsinin xüsusi məsələlərdən olduğu göstərilmişdir. RİT sferasında terminlərin və anlayışların çoxsahəliliyə, dinamikliyə və geniş əhatə dairəsinə malik olduğu göstərilmişdir. RİT sferası üzrə terminoloji vəziyyət araşdırılmış, onun Azərbaycan dilində termin və

anlayışlar bazasının formalaşdırılması mərhələləri göstərilmiş, terminlərin linqvistik xüsusiyyətləri üzrə araşdırmaların əsas istiqamətləri qeyd olunmuşdur. RİT sferasında terminoloji pərakəndəlik və uyğunsuzluq problemləri araşdırılmış, Azərbaycan dilində termin və anlayışların uyğunsuzluğu nümunələri verilmişdir. Məqalədə təhlil olunan problemlər üzrə müvafiq tövsiyələr verilmiş, həll mexanizmləri təklif olunmuşdur.

Açar sözlər — rəqəmsal iqtisadiyyat; rəqəmsal texnologiyalar; rəqəmsal transformasiya; termin və anlayışlar bazası; rəqəmsallaşma; innovasiyalaşma; rəqəmsal qurğu və avadanlıqlar; Sənaye 4.0 platforması

I. GİRİŞ

İqtisadiyyatın və cəmiyyətin sürətli rəqəmsal transformasiyası son illərdə ölkə qarşısında duran prioritet məsələlərdən birinə çevrilmişdir [1]. Ona görə də rəqəmsal transformasiyalar şəraitində, eləcə də Sənaye 4.0 platformasında rəqəmsal iqtisadiyyatın formalaşması prosesində həmin sahədə Azərbaycan dilində termin və anlayışlara ehtiyac kəskin şəkildə artmışdır. Bu mənada həmin termin və anlayışlar bazasının işlənilməsi müasir dövrün aktual məsələlərinə çevrilmişdir. İnformasiya, informatika, İKT, informasiyalaşma,

informasiya sistemləri, rəqəmsal texnologiyalar, yaşıl rəqəmsal texnologiyalar, rəqəmsal transformasiya, yaşıl texnologiyalar, yaşıl transformasiya, bilik iqtisadiyyatı, rəqəmsal iqtisadiyyat, sirkulyar iqtisadiyyat, ikinci əl İKT bazarı, rəqəmsal qurğu və avadanlıqlar (*digital devices*) bazarı, iqlim dəyişikliyi iqtisadiyyatı, Big Data, süni intellekt, Əşyaların İnterneti və s. terminlərin aid olduğu sferaların formalaşması və idarə olunması sahələrində ən yeni, müasir termin və anlayışlar (toplanması, seçilməsi, tərcümə edilməsi, təhlili, təsnifi) üzrə izahlı lüğətlərin işlənməsi məsələlərinə xüsusi zəruriyyət yaranır.

“Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsinə və ölkədə dilçiliyin inkişafına dair Dövlət Proqramı haqqında” [2]. 2012-ci il 23 may tarixli sərəncam ölkədə dil siyasətinin uğurla həyata keçirilməsində mühüm fəaliyyət sənədi olmuşdur. Müxtəlif illərdə bir çox dövlət sənədlərində milli kontentin inkişaf etdirilməsi öz aktuallığı ilə əhəmiyyət kəsb etmişdir [3]. Bütün bunlar hazırda Rəqəmsal iqtisadiyyat və transformasiya texnologiyalarının (RİTT) Azərbaycan dilində termin və anlayışlar bazasının formalaşdırılmasını zəruri hesab edir.

II. RƏQƏMSAL İQTİSADİYYAT VƏ RƏQƏMSAL TRANSFORMASIYANIN MAHIYYƏTİ, MƏQSƏDİ VƏ VƏZİFƏLƏRİ

Ölkənin İKT sahəsində yeni infrastruktur quruculuğu həyata keçirilir. Müasir İKT texnologiyaların tətbiqi və modernizasiyası

sahəsində mühüm işlər yerinə yetirilir. İnformasiya Cəmiyyətinin (İC) əsası olan rəqəmsal iqtisadiyyat iqtisadi proseslərin, strukturların İKT-nin tətbiqi nəticəsində yeniləşən, dəyişən, elektronlaşan hissəsidir. Rəqəmsal iqtisadiyyatın məqsədi cəmiyyətin iqtisadi həyatının bütün sahələrində rəqəmsal texnologiyaların sistemli tətbiqini və inkişafını təşkil etməkdir [4]. Rəqəmsal iqtisadiyyat elektron ticarəti, e-bankinqi, elektron pulları, e-maliyyəni, e-vergini, e-kömrüyü, e-sığortanı, e-informasiya mübadiləsini, kapitalın elektron hərəkətini və s. özündə birləşdirir. Rəqəmsal iqtisadiyyatın ən geniş yayılmış növü e-ticarətdir.

Rəqəmsal iqtisadiyyat e-kommersiya və e-biznesin, həmçinin e-məhsul və e-xidmətlərin sistemidir; İKT-dən istifadə olunan iqtisadi münasibətlər sistemidir; Virtual məkanda sosial və iqtisadi fəaliyyətin qlobal şəbəkəsidir; Bütün biznes proseslərinin elektron rejimdə həyata keçirilməsidir; Obyekt və proseslərin, tələb və təklifin, xidmətlərin, əmtəənin, müxtəlif idarəetmə modellərinin onlayn, elektron mühitə köçürülməsidir; Əsas resurs və başlıca kriteriyasının informasiya olduğu iqtisadiyyatdır.

Rəqəmsal transformasiya (digital transformation) – təşkilat tərəfindən əsas texnoloji proseslərin və idarəetmə sisteminin optimallaşdırılması ilə müşayiət olunan və ona gətirən rəqəmsal texnologiyaların tətbiqi prosesidir. *Digitalization* və ya rəqəmsal çevrilmə qlobal bir prosesə çevrilir, bu da həyatın müxtəlif sahələrində rəqəmsal texnologiyaların geniş yayılması deməkdir.

Rəqəmsal transformasiya istehsalın, xidmətin, satışların, biznesin və s. inkişafını sürətləndirmək üçündür. Rəqəmsal transformasiya – təşkilat tərəfindən rəqəmsal texnologiyanın adaptasiyasıdır. Rəqəmsal transformasiyanın həyata keçirilməsində əsas ümumi məqsəd səmərəliliyi, dəyəri və ya innovasiyanı təkmilləşdirməkdən ibarətdir. Rəqəmsal transformasiya qeyri-kommersiya təşkilatlarının (məsələn, universitetlər, digər təhsil müəssisələri, sosial obyektlər) səmərəliliyini artırmaq üçün nəzərdə tutulmuşdur. Təşkilatda rəqəmsal çevrilmənin dərəcəsini və müvəffəqiyyətini xarakterizə edən rəqəmsal inkişaf göstəricisi – təşkilatın rəqəmsal yetkinlik səviyyəsidir. Rəqəmsal transformasiya təkcə İKT-nin təkamülü deyil, bütün təşkilata təsir edən vahid biznes dəyişikliyi.

İqtisadiyyatda rəqəmsal texnologiyaların tətbiqi prosesi rəqəmsal transformasiyanın həyata keçirilməsinin ilkin şərtidir. Rəqəmsal transformasiya əsas tendensiyadır. İqtisadiyyatın müxtəlif sahələrinə və sosial sferaya xas olan, gündəlik prosesləri avtomatlaşdırmaq, bir çox proseslərin həyata keçirilməsi sürətini artırmaq və yeni imkanlar yaratmaqla iqtisadiyyatın səmərəliliyini əhəmiyyətli dərəcədə artırmaq olar. Rəqəmsal transformasiya, bir qayda olaraq, istehsalçılar və istehlakçılar arasında qarşılıqlı əlaqə üçün yeni zəmin yaradan, hər ikisinin ehtiyaclarını ödəməyə imkan verən innovasiyaların istifadəsi əsasında məhsulların, proseslərin, idarəetmənin (idarəetmənin) dəyişdirilməsinə yönəldilmişdir.

“Rəqəmsal transformasiya” termini xarici tədqiqatçılar tərəfindən birmənalı tərif almamışdır. Ən geniş yayılmış “rəqəmsal transformasiya” termininin şərhə – bu, rəqəmsal texnologiyaların istifadəsi ilə bağlı, təkcə iqtisadi deyil, cəmiyyətin bütün aspektlərindəki dəyişiklikdir [5].

Rəqəmsal transformasiya rəqəmsal olmayan məhsulların, xidmətlərin və ya əməliyyatların rəqəmsallaşdırılması üçün təşkilat tərəfindən rəqəmsal texnologiyanın qəbul edilməsidir.

Keyfiyyətli, təhlükəsiz və səmərəli rəqəmsal transformasiya, eləcə də bu sahədə resurslardan effektiv istifadə innovasiyaların inkişafında hərəkətverici qüvvə hesab olunur. Rəqəmsal xidmətlərin genişləndirilməsi və elektron hökumətin inkişafı vasitəsilə səmərəlilik və şəffaflıq artırılmaqdadır.

Rəqəmsal transformasiya sahəsində həyata keçirilən dövlət siyasəti üzrə idarəetmə, əlaqələndirmə və nəzarət mexanizmləri daimi təkmilləşdirilməkdədir. Rəqəmsal transformasiya sahəsində fəaliyyətin təşkili, əlaqələndirilməsi və həyata keçirilməsi dinamik xarakter daşıyır və ümumən ölkənin ÜDM-ə müsbət təsir göstərir [1].

Hazırda dünya dövlətlərinin ÜDM-nin təxminən 22%-i rəqəmsal iqtisadiyyat sayəsində formalaşır [6]. Bəzi iqtisadi şirkətlərin hesablamlarına görə iş yerlərinin avtomatlaşması hesabına dünya ÜDM-i 2030-cu ilə qədər 9 trilyon dollar arta bilər. 2030-cu ilə qədər dünya işçi qüvvəsinin təqribən 14 faizi öz peşəsini dəyişmək məcburiyyətində qalacaq. 2035-ci ildə dünya rəqəmsal iqtisadiyyatının

proqnozlaşdırılan həcmi 16 trln. dollar təşkil edəcəkdir. 2025-ci ildə Çində ÜDM artımının 22 %-ə qədər İnternet texnologiyası hesabına təmin ediləcəkdir. 2025-ci ildə ABŞ-da rəqəmsal texnologiyalar hesabına yaradılan dəyərin gözlənilən artımı 1,6-2,2 trln. dollar ətrafında olacaqdır. 2025-ci ilə qədər İnternet dünya iqtisadiyyatına ildə 4-11 trln. dollar gəlir gətirəcəkdir. 2025-ci ildə rəqəmsallaşma ÜDM-in artımının 19-34 %-ni təmin edəcəkdir. 2036-cı ildə iş proseslərinin 50%-ə qədər avtomatlaşdırılacaqdır [7-9].

III. RƏQƏMSAL İQTİSADİYYATDA İKT İNFRASTRUKTURUNUN, XİDMƏTLƏRİNİN VƏ MİLLİ KONTENTİN İNKİŞAF ETDİRİLMƏSİ

Rəqəmsal iqtisadiyyat ilk növbədə yüksək texnologiyalar sektorunun inkişafı hesabına formalaşır. Eyni zamanda aşağıdakılara da ciddi diqqət yetirilməlidir:

- Yüksək texnologiyaların inkişafı üzrə elmi-texniki potensialın gücləndirilməsi;
- Elektron hökumətin inkişafı;
- İKT cəmiyyətin inkişaf amili kimi;
- İKT üzrə kadr hazırlığı;
- İnformasiya təhlükəsizliyinin təmin edilməsi.

İnformasiya cəmiyyətinin inkişafı şəraitində milli kontentin genişləndirilməsinə, ölkənin milli-mədəni irsinin qorunub saxlanılmasına və inkişafına, əhalinin ondan müasir texnologiyalar vasitəsilə istifadəsinə və bilik

səviyyəsinin artırılmasına daim diqqət yetirilməkdədir. Milli kontentin zənginləşdirilməsinə xidmət edən İnternet resurslarının yaradılmasına dövlət dəstəyi mexanizmləri formalaşdırılmışdır.

Azərbaycandilli proqram təminatlarının hazırlanması və geniş istifadəsinin təmin olunması sahəsində: Gələcək nəsillər naminə mədəni irsin qorunması və rəqəmsallaşdırma prosesinin sürətləndirilməsi; Məkan məlumatları infrastrukturunun qurulması, o cümlədən müxtəlif təyinatlı coğrafi informasiya və naviqasiya resurslarının yaradılması; Onlayn medianın inkişaf etdirilməsi, yeni media vasitələrindən səmərəli istifadəyə dair tədbirlərin görülməsi və s. kimi işlər yerinə yetirilməkdədir. Bu və buna bənzər müvafiq sferalarda həm rəqəmsal iqtisadiyyat sektorları, həm də İKT infrastrukturunu və xidmətləri daimi olaraq yenilənməli və gücləndirilməlidir [10].

IV. RƏQƏMSAL İQTİSADİYYAT VƏ TRANSFORMASİYA SFERASINDA AZƏRBAYCAN DİLİNDƏN QLOBAL İSTİFADƏ PROBLEMLƏRİ VƏ İZAHLI LÜĞƏTLƏRİN İŞLƏNİLMƏSİ ZƏRURƏTİ

Bu sahədə ilk növbədə aşağıdakılara diqqət yetirilməlidir:

- Azərbaycan dilinin qloballaşma şəraitində istifadəsi;
- Ədəbi dilin normalarına ciddi surətdə əməl edilməsi;
- Müxtəlif sahələr üzrə yüksəkixtisaslı kadrların hazırlanması;

- Ölkədə dilçiliyin inkişafına dair Dövlət Proqramının icrası;

- Dil siyasətinin uğurla həyata keçirilməsində fəaliyyət sənədlərinin daimi təkmilləşdirilməsi və yenilənməsi.

Müasir qloballaşma dövründə rəqəmsal iqtisadiyyat və transformasiyalar sferasında Azərbaycan dilində termin və anlayışlar bazasının formalaşdırılması və həmin sferada izahlı lüğətlərin işlənməsi xüsusi məsələlərdən hesab olunur. Burada daha çox aşağıdakı sferalarda müvafiq məsələlərin həlli çox vacibdir.

- İnformasiya, informatika, İKT, informasiyalaşma, informasiya sistemləri;

- Yaşıl rəqəmsal texnologiyalar, yaşıl texnologiyalar, yaşıl transformasiya;

- Rəqəmsal texnologiyalar, rəqəmsal transformasiya;

- Bilik iqtisadiyyatı, rəqəmsal iqtisadiyyat, iqlim dəyişikliyi iqtisadiyyatı;

- Big Data, süni intellekt, Əşyaların İnterneti.

Göstərilən sferaların formalaşması və idarə olunması sahələrində ən yeni termin və anlayışlar (toplanması, seçilməsi, tərcümə edilməsi, təhlili, təsnifi) üzrə izahlı lüğətlərin işlənməsi məsələlərinə zərurət yaranmışdır. İKT-yə əsaslanan mövcud rəqəmsal iqtisadiyyatın Sənaye 4.0 platforması və innovativ texnologiyalar əsasında intellektuallaşdırılmasının prinsip, prioritet, xüsusiyyət, məqsəd və kriteriyalarının aydınlaşdırılmasına ciddi ehtiyac yaranmışdır. Bunu formalaşmış bilik

və texnologiya iqtisadiyyatının inkişafı probleminin müxtəlif aspektlərinin aktuallığı da təsdiq edir. Tədqiqatçılar konseptual aparata dair təkamül baxışların formalaşmasına, mövcud tendensiyaların öyrənilməsinə və biliklər iqtisadiyyatının inkişafının perspektivli istiqamətlərinin müəyyənləşdirilməsinə böyük diqqət yetirirlər. İqtisadi ədəbiyyatın təhlili göstərir ki, nə informasiya, nə də rəqəmsal innovasiya iqtisadiyyatı termininə vahid elmi yanaşma yoxdur [11]. Alimlər onların mənasını müəyyən etmək üçün bir çox müxtəlif terminlərdən istifadə edirlər.

Rəqəmsal iqtisadiyyat və transformasiya sferasında terminlər və anlayışlar çoxsahəliliyə, dinamikliyə və geniş əhatə dairəsinə malikdir.

Rəqəmsal iqtisadiyyat və transformasiya sferası üzrə terminoloji vəziyyətin araşdırılmasında aşağıdakı kimi məsələləri qeyd etmək olar:

- Rəqəmsal iqtisadiyyat, rəqəmsal transformasiya, İC və İKT sferası üzrə vahid terminoloji mühitin formalaşması;

- Rəqəmsal iqtisadiyyatın regional və sektorial inkişafının təminində terminologiyanın rolu, vəzifələri;

- Rəqəmsal iqtisadiyyat və rəqəmsal transformasiya sferasında terminlər və anlayışların çoxsahəliliyi, dinamikliyi və əhatə dairəsinin müxtəlifliyi.

Terminlər sisteminin yaranma prosesində əsas əhəmiyyətə malik amillərə bunları daxil etmək olar: 1) ekstralinqvistik – bilik və fəaliyyət dairəsi ilə bağlı anlayışların adlandırılmasının vacibliyini, həmçinin

fundamental elm sahələrinə dair anlayışların işlənməsi məsələlərini şərtləndirir; 2) linqvistik – terminlərlə ötürülən anlayış münasibətlərinin müxtəlifliyi, həmçinin terminlərdə komponentlərin birləşmə imkanları ilə bağlı olması [12].

Rəqəmsal iqtisadiyyatda İKT infrastrukturunu və xidmətlərinin həyata keçirilməsi məqsədi ilə İKT sahəsində işlədilən terminlərin ekstralingvistik və linqvistik təhlilinin aparılması əsas məsələlərdəndir. Göstərilən məqsədə nail olmaq üçün aşağıdakı məsələlərin həlli nəzərdə tutulmuşdur [12].

- Terminlər sistemi və biliklərin inteqrasiyası məsələləri;
- Terminlərinin yaranması və fənlərarası əlaqələr;
- İKT terminlərində interferensiya prosesi;
- Termin yaratmanın özəllikləri;
- Terminlərin çoxmənalılığı;
- Terminlərdə sinonimlik;
- Terminlər sistemində sahə anlayışı;
- Terminlərin beynəlmilliliyi;
- Terminlərin milli dillərə tərcüməsi problemləri;
- Kompüter proqramlarında Azərbaycan dilinə tərcümə olunmuş terminlər.

RİT sferasının Azərbaycan dilində termin və anlayışlar bazasının formalaşdırılması mərhələləri aşağıdakılardır:

- RİT sferasında Azərbaycan dilində işlənən termin və anlayışların toplanması, sistemləşdirilməsi, müqayisəli təhlili;
- RİT sferasında Azərbaycan dilində işlənən termin və anlayışlar bazasının rəsmi səviyyədə müzakirəsi və nəşri;
- RİT sferasında Azərbaycan dilində işlənən termin və anlayışlar bazası üzrə alt terminoloji sistemin proqram təminatının hazırlanması;
- RİT sferasında Azərbaycan dilində işlənən termin və anlayışlar bazasının onlayn mühitdə formalaşdırılması.

V. RİT SFERASINDA TERMINOLOJİ PƏRAKƏNDƏLİK VƏ UYĞUNSUZLUQ PROBLEMLƏRİ

RİT sferası üzrə bir çox termin və anlayışların təhlili göstərir ki, müvafiq sahədə pərakəndəlik mövcuddur və çoxlu sayda terminoloji uyğunsuzluqlar vardır. Bu uyğunsuzluqlar əsasən aşağıdakı sahələrdədir:

- İKT-nin sənaye, elm, xidmət, innovasiya sahəsində terminoloji problemləri;
- İKT üzrə milli standartlarda terminoloji uyğunsuzluqlar;
- RİT sferasında, İnformasiya Cəmiyyəti və İKT üzrə terminoloji pərakəndəlik və uyğunsuzluq problemləri.

RİT-in Azərbaycan dilində termin və anlayışlarının uyğunsuzluğu nümunələrinə aşağıdakıları misal göstərmək olar (cədvəl 1).

Cədvəl 1. RİT-in termin və anlayışlarının uyğunsuzluğu nümunələri

Sıra №	Termin azərbaycanca	Termin ingiliscə	Termin rusca	Kitab 1(A)	Kitab 2(C)	Kitab 3(M)
1.	Rəqəmsallaşma	digitization	Оцифровка	-	-	-
2.	Rəqəmsallaşdırma	digitization	Оцифровка	-	+	-
3.	Rəqəmsallaşdırılma	digitization	Оцифровка	-	-	-
4.	Kompüterləşdirmə	Computerization	Компьютеризация	+	-	-
5.	Kompüterləşdirilmə	Computerization	Компьютеризация	+	-	-
6.	Kompüterləşdirilmiş	Computerized	Компьютеризированный	+	-	-
7.	Kompüterizasiya	Computerization	Компьютеризация	-	-	+
8.	Kodlaşdırma	Coding	Кодирование	+	+	-
9.	Kodlaşdırılma	Encoding	Кодирование	-	+	-
10.	Kodlaşdırmaq	To encode	Кодировать	-	-	-
11.	Kodlaşdırılmış	Coded	Закодированный	+	-	+
12.	İnformasiyalaşdırma	Infomatization	Информатизация	-	+	-
13.	İnformasiyalaşdırılma	Infomatization	Информатизация	-	+	-
14.	İnformasiya sistemi	Information system	Информационная система	-	-	+
15.	İnformasiya texnologiyası	Information technology	Информационные технологии	-	-	-

1(A)* - Ə. Abbasov, S. Kərimov, N. Rəhimova, N. Vəliyev, T. İbrahimzadə, N. Cəfərov. İnformasiya-telekommunikasiya texnologiyaları üzrə ingiliscə-rusca-azərbaycanca izahlı lüğət, Bakı, ADNA, 2012.

2(C)* - İsmayıl Cəlalı. İnformatika terminlərinin izahlı lüğəti, Bakı - 2017.

3(M)* - Rafiq Məmmədli. İngiliscə-azərbaycanca izahlı kompüter lüğəti, Bakı - 2017.

Terminoloji pərakəndəlik və uyğunsuzluqlar, həmçinin aşağıdakı mənbələrdə də vardır:

- RİT üzrə ixtisaslarda, müasir rəqəmsal innovativ texnologiyalar üzrə keçirilən

tədris proqramlarında (bakalavr, magistr və fəlsəfə doktoru səviyyəsi üzrə);

- Termin və anlayışların ensiklopedik lüğətlərində, ISO standartlarında;
- Akademik institutların və ali təhsil müəssisələrinin saytlarında;
- Standartlaşdırma sferasında;
- İqtisadi strukturların saytlarında;
- Müvafiq elmi məqalələrdə və elmi jurnallarda;
- Media nümunələrində;
- İzahlı lüğətlərdə.

NƏTİCƏ

Azərbaycan dilçiliyində terminologiya-nın müxtəlif sahələrinə aid kifayət qədər araşdırmalar aparılsa da, RİT terminologiyası indiyə qədər hərtərəfli və sistemli şəkildə tədqiq olunmamışdır. İndiki qloballaşma şəraitində sürətlə inkişaf edən RİT, İC, İKT ilə əlaqəli terminlərin öyrənilməsi və düzgün tətbiqi olduqca vacibdir.

Bütün səviyyələrdə RİT, İC, İKT terminlərinin linqvistik mahiyyətinin aydınlaşdırılması, həmin terminlərin Azərbaycan dilində qarşılığının verilməsi, tərcüməsi və milliləşdirilməsi üsullarının işlənilməsi olduqca aktual məsələdir. RİT sferasında Azərbaycan dilində termin və anlayışlar bazasının rəsmi səviyyədə onlayn mühitdə formalaşdırılması aktual məsələdir və ona böyük tələbat vardır.

İSTİNADLAR

- [1] Rəqəmsal transformasiya sahəsində idarəetmənin təkmilləşdirilməsi haqqında Azərbaycan Respublikası Prezidentinin Fərmanı. Bakı, 27 aprel 2021. <https://president.az/articles/51299>
- [2] Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsinə və ölkədə dilçiliyin inkişafına dair Dövlət Proqramı. 09 aprel 2013. Azərbaycan Respublikası Prezidentinin 2013-cü il 9 aprel tarixli Sərəncamı. <https://president.az/az/articles/view/7744>
- [3] Azərbaycan Respublikasında İnformasiya Cəmiyyətinin inkişafına dair 2014-2020-ci illər üçün Milli Stratejiya. Azərbaycan Respublikası Prezidentinin Sərəncamı. Bakı 2 aprel 2014. <https://president.az/az/articles/view/11312>
- [4] A.G.Aliyev, R.O.Shahverdiyeva. Scientific and methodological bases of complex assessment of threats and damage to information systems of the digital economy. *International Journal Information Engineering and Electronic Business*, 2022, №2, pp.23-38.
- [5] А.И.Гниятуллина. Цифровая трансформация государственного управления: базовые термины и определения. В сборнике: Цифровые технологии в государственном и муниципальном управлении развитием территорий: новые концептуальные подходы. Научно-практической конференции. УФА, 2022, с.16-20.
- [6] İ.M.Abbasov. Rəqəmsal iqtisadiyyat və rəqəmsal təhsil. Rəqəmsal iqtisadiyyat: Azərbaycan yeni iqtisadi inkişaf mərhələsində beynəlxalq elmi-praktiki konfrans. Bakı, 25 noyabr, 2020, səh.5-9.
- [7] Digital Economy Report-2021. United Nations publication issued by the United Nations Conference on Trade and Development, 238 p.
- [8] А.И.Сарыгулов, В.Б.Рапоф. О некоторых подходах к оценке цифровой экономики. *Проблемы современной экономики*, 2020, №2(74), стр.46-50.
- [9] Digital Economy and Society Index-2020. <https://ec.europa.eu/digital-single-market/en/digital-economy-and-society-index-desi>
- [10] A.G.Aliyev, R.O.Shahverdiyeva. A conceptual approach to the formation of a digital innovation economy based on artificial intelligence technologies. *Artificial societies*, 2022, v.17, issue 4, pp.1-16.
- [11] А.С.Квилинский, Н.В.Трушкина, Н.С.Рынкевич. Концептуальные подходы к определению термина «информационная экономика. *Проблемы экономики*, 2019, №3(41), стр.147-155.
- [12] М.М.Мəmmədova. “İnformasiya-Kommunikasiya Texnologiyaları sahəsində işlədilən terminlərin ekstralingvistik və lingvistik xüsusiyyətləri.” *Filologiya üzrə fəlsəfə doktoru elmi dərəcəsi almaq üçün təqdim edilmiş dissertasiyanın Avtoreferatı*, Bakı, 2018.

Azərbaycan dilinin milli korpusu süni intellekt problemləri kontekstində

Məsud Mahmudov

AMEA Nəsimi adına Dilçilik İnstitutu, Bakı, Azərbaycan
mmasud@bk.ru

Xülasə — Süni intellekt elminin məqsədi insan aqlını, düşüncəsini, təfəkkürünü, insan beynini təqlid edə biləcək süni sistemlərin yaradılmasıdır. Süni intellektin bir elm kimi formalaşmasında riyaziyyat, fizika, biologiya, fiziologiya, biotexnika, kompüter elmləri, neyrofiziologiya, psixologiya, fəlsəfə, məntiq, dilçilik, iqtisadiyyat və s. elmlər mühüm rol oynayır. Süni intellektin ən əhəmiyyətli istiqamətlərdən biri təbii dilin emalı hesab olunur. Yeni linqvistik texnologiyaların yaradılması, təbii mətnin emalı, maşın tərcüməsi, məlumatların formal əlamətlərə görə axtarışı, tarixi abidələrin və klassiklərin dilinin riyazi-statistik metodlarla tədqiqi, statistik lüğətlərin tərtibi, elektron lüğətlərin hazırlanması sahəsində görülmüş işlər ənənəvi olaraq kompüter dilçiliyi və korpus dilçiliyinin tədqiqat obyektinə hesab olunurdu. Fikrimizcə, bu məsələləri süni intellekt sistemlərinin tədqiqat obyektinə kimi səciyyələndirmək daha düzgün olardı.

Açar sözlər — *süni intellekt; milli korpus; təbii mətnin emalı; statistik lüğətlər; linqvistik analizator*

I. GİRİŞ

Azərbaycan dilinin milli korpusunun yaradılması sahəsində aparılan işlər də süni intellekt sistemləri kontekstində

araşdırılmalıdır. Bu, dilçiliyin bir çox problemlərinə yeni baxış bucağından baxmaq, bu istiqamətdə aparılan elmi-tədqiqat işlərinin səmərəsini artırmaq imkanı yarada bilər. Təbii mətnin emalı, mətnin formal linqvistik təhlili, dilin riyazi-statistik metodlarla öyrənilməsi və bu səpkidə olan digər linqvistik texnologiyalar ayrı-ayrılıqda, bir-birindən təcrid olunmuş halda deyil, süni intellekt problemləri kontekstində tədqiq olunmalıdır. Belə olduqda, daha dəqiq və optimal nəticələr əldə etmək imkanı yaranır.

Korpus dilçiliyi hər hansı konkret dili təmsil edən mətnlərin qabaqcadan müəyyənləşdirilmiş qaydalar üzrə sistemli şəkildə elektron formatda toplanmasına istiqamətlənmiş məlumat axtarışı sistemlərinin yaradılması və onların tədqiqat prosesində istifadəsi məsələlərini öyrənir. Məlum olduğu kimi, “korpus” bir çox mənalarda işlənir: “hərbi birləşmə”, “tədris korpusu”, “gəminin korpusu”, “şrift dəsti” və s. Korpus dilçiliyində “korpus” elektron formatda toplanmış mətnlər toplusu, elektron fond mənasında işlənir.

II. MİLLİ DİL KORPUSLARI

Konkret dilin milli korpusunda müəyyən

zaman kəsiyində həmin dil tam şəkildə bütün janrları, üslubları, ədəbi dili, dialektləri, qrammatikası, lüğətləri və s. ilə təmsil olunur. Başqa sözlə, milli dil korpusu elektron daşıyıcıda saxlanan müxtəlif dil hadisələri və faktları ilə bağlı tədqiqatçılara operativ və dəqiq məlumat verən, müəyyən nizamla düzülmüş təbii mətnlər çoxluğudur. Burada “milli” sözü də müxtəlif anlamlarda izah olunur. Qeyd etməliyik ki, Britaniya milli korpusundan gələn ənənə kimi işlənməkdədir.

Süni intellekt dedikdə, ənənəvi olaraq insanın etməli olduğu, insana xas yaradıcı funksiyaları yerinə yetirə biləcək intellektual sistem, intellektual kompüter proqramlarının yaradılması elmi və texnologiyası başa düşülür.

İntellektual sistem dedikdə, həmin sistemin yaddaşında saxlanan, yaradıcı proses kimi səciyyələndirilən məsələləri həll etməyə qabil olan texnologiya və ya proqram sistemi başa düşülür.

III. TƏBİİ DİLİN EMALI

Təbii dilin emalı – NLP (Natural Language Processing) sistemləri dedikdə, maşın tərcüməsi, orfoqrafiyanın və qrammatikanın yoxlanması, nitqin yazıya (mətnə) çevrilməsi və başa düşülməsi (qavranması), sorğulara avtomatik cavab verilməsi, mətnin nitqə çevrilməsi (mətn üzrə insan nitqinin generasiyası) və s. sistemlər başa düşülür. Bu sistemlər insanla kompüter arasında daha təbii interfeys yaradılmasına, insanın kompüterlərdən daha səmərəli yararlanmasına və geniş istifadə edilməsinə şərait yaradır.

IV. ELEKTRON LÜĞƏTLƏR

Bəzən bu tipli lüğətlər kompüter lüğətləri də adlandırılır. Elektron formatda tərtib olunmuş və daha çox internetdə istifadə üçün nəzərdə tutulmuş elektron lüğətlər sistemidir. Lüğətlərin kitab (kağız) variantından fərqli olaraq daha geniş yayılma və sürətli istifadə imkanlarına malikdir. Hazırda kitab variantında olan qiymətli və nadir lüğətlərin də elektron variantlarının yaradılması sahəsində intensiv işlər aparılır. Konkordansları da elektron lüğətlərə aid etmək olar. Milli dil korpuslarının mühüm bloklarından biri də elektron lüğətlər altkorpusu hesab olunur.

Məlumdur ki, süni intellekt və milli dil korpuslarını yaradılması sahəsində bir çox dünya ölkələrində tədqiqatlar aparılmışdır. Tədqiqatın məqsədi bu sahədə aparılan işləri təhlil etməklə Azərbaycan dilinin materialında süni intellekt problemləri çərçivəsində tətbiqi və nəzəri linqvistik texnologiyaların hazırlanması, bu zəmində Azərbaycan dilinin milli korpusunun ayrı-ayrı bloklarının struktur və məzmununu müəyyənləşdirilməsidir. Məruzədə süni intellekt problemləri sahəsində böyük xidmətləri olan J.L.Loryer, H.Nilson, S.Rassel, E.Xant, A.Türinq, D.Smolın, V.Devyatkov, A.Jdanov və s. əsərlərindən [8-16], eləcə də internet resurslarından istifadə olunmuşdur. Həmin müəlliflərin əsərlərində süni intellektin yaranması, onun inkişaf istiqamətləri, ayrı-ayrı bölmələri, perspektivləri, nəzəri və praktik əhəmiyyəti, tətbiq sahələri barədə ətraflı məlumat verilir.

Yuxarıda qeyd edildiyi kimi, süni intellekt ənənəvi olaraq insanın etməli olduğu, insana xas yaradıcı funksiyaları

yerinə yetirə biləcək intellektual sistem, intellektual maşınların, xüsusilə intellektual kompüter proqramlarının yaradılması elmi və texnologiyasıdır.

Qeyd etmək lazımdır ki, mətnin formal təhlili, təbii mətnin emalı, mətnin səsə (nitqə), səsin (nitqin) mətnə çevrilməsi, maşın tərcüməsi, məlumatların avtomatik (formal əlamətlərinə görə) axtarışı, tarixi abidələrin və klassiklərin dilinin riyazi-statistik metodlarla tədqiqi, statistik lüğətlərin (tezlik, əks) tərtibi, elektron lüğətlərin hazırlanması sahəsində Azərbaycanda xeyli iş görülmüşdür [1-7].

Bu araşdırmaları və analogi istiqamətdə Azərbaycan Respublikasında aparılan digər tədqiqat işlərini süni intellekt sistemlərinə aid tədqiqatlar kimi səciyyələndirmək olar. Müasir Azərbaycan ədəbi dilinin ayrı-ayrı üslublarının, Azərbaycan klassiklərinin və yazılı tarixi dil abidələrin riyazi-statistik metodlarla öyrənilməsi sahəsində aparılmış tədqiqatlar da milli korpusun yaradılması prosesində böyük fayda verə bilər [3]. Məlum olduğu kimi, Azərbaycanda “Dilmanc” layihəsi çərçivəsində NLP – təbii dilin (mətnin) emalı sahəsində bir çox işlər görülmüşdür. Azərbaycan dilinin tezlik və əks lüğətlərinin tərtibi sahəsində də həmin yaradıcı kollektivin xidmətləri danılmazdır [1; 7].

Yuxarıda qeyd olunduğu kimi, təbii mətnin emalı, məlumatların formal əlamətə görə axtarışı, maşın tərcüməsi, elektron lüğətlərin tərtibi, dil materialının riyazi-statistik metodlarla tədqiqi məsələləri korpus dilçiliyinin, eləcə də süni intellekt sistemlərinin tədqiqat obyektinə hesab olunur. Yeni linqvistik texnologiyaların yaradılması

və tətbiqi süni intellekt sistemlərinin əsasını təşkil edir. Robot sənayesini bunlarsız təsəvvür etmək qeyri-mümkündür. Nitqi qavrayıb müəyyən suallara cavab verə bilən, şifahi və yazılı əməlləri icra etmək qabiliyyətinə malik robotlar təbii mətnin emalı texnologiyalarından istifadə etmiş olurlar.

Verilənlərdən biliyin alınması texnologiyası intellektual təhlilin əsas problemlərindən biridir. Bu problemin həlli ilə bağlı müxtəlif yanaşmalar mövcuddur. Bu yanaşmalardan biri də neyroşəbəkə texnologiyasına əsaslanır. Neyroşəbəkə – canlı orqanizmlərin sinir sistemlərinin quruluşu və funksiyalarını imitasiya edən riyazi modeldir.

Süni intellektlə neyrofiziologiya, epistemologiya və koqnitiv psixologiyanın qovuşuğunda koqnitologiya adlanan yeni elm sahəsi yaranmışdır. Koqnitologiya dilçilik elminin də tədqiq obyektlərindən biri kimi səciyyələndirilir.

Azərbaycan dilinin milli korpusunun yaradılması sahəsində artıq zəmin yaradıldığından hazırda daha geniş miqyasda işlərə başlamaq olar. Bunun üçün isə daha böyük mətnlər korpusuna ehtiyac duyulur. Belə ki, korpusa daxil edilən mətnlər Azərbaycan dilinin bütün üslublarını eyni səviyyədə və həcmdə ehtiva etməlidir. Məlum olduğu kimi, korpusun həcmi nə qədər böyük olsa, bir o qədər etibarlı və əhatəli hesab olunur. Azərbaycan dilinin milli korpusuna aşağıdakı mətnlərin daxil edilməsi məqsədəuyğun hesab edilir: Əsas korpus (təxminən 50 milyon söz-forma) – 1. Bədii mətnlər (nəsr və nəzm əsərləri, dramaturgiya, folklor mətnləri,

klassik və müasir Azərbaycan ədəbiyyatı nümunələri), publisist, rəsmi-ışgüzar və elmi üsluba aid mətnlər, dini və fəlsəfi mətnlər, dialekt mətnləri və s.; 2. Azərbaycan dilinin elektron lüğətlər altkorpusu (Azərbaycan dilinin izahlı, orfoqrafiya, terminoloji, omonimlər, sinonimlər, antonimlər, dialektoloji, tezlik, əks əlifba, ikidilli, çoxdilli, frazeologiya, orfoepiya, statistik və s. lüğətləri); 3. Şifahi mətn altkorpusları (məişət-danışqı dili, rəsmi çıxışlar, dialekt mətnləri və s.); 4. Paralel altkorpuslar (ikidilli – ingilis-Azərbaycan, Azərbaycan-ingilis, rus-Azərbaycan, Azərbaycan-rus, alman-Azərbaycan, Azərbaycan-alman və s., eləcə də analogi olaraq çoxdilli); 5. Çoxdilli və birdilli mətnlər; 6. Linqvistik analizatorlar — dilin müxtəlif səviyyələri (fonetik, morfoloji, sintaktik, sözdüzəltmə, semantik) üçün nəzərdə tutulmuş təhlil və axtarış proqramları; 7. Yazılı mətnin səsə (nitqə), səsin (nitqin) mətnə çevrilməsini təmin edən proqram sistemləri; 8. Tədris altkorpusları; 9. Konkordanslar; 10. Korpusun istifadəsi üçün mükəmməl proqram təminatı (axtarış proqramları).

Azərbaycan dilinin milli dil korpusunun yaradılması prosesində qabaqcıl dünya təcrübəsi nəzərə alınır. Burada korpuslar üçün çox önəmli olan mətnlərin formal meyarlara görə təhlili ön plana çəkilir. Hazırda “Azərbaycan dilinin tezlik lüğəti”nin birinci cildi çap edilmiş, “Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsinə və ölkədə dilçiliyin inkişafına dair” Dövlət Proqramına uyğun Azərbaycan dilinin ümumi informasiya bazası [17] yaradılmışdır. Elmin İnkişafı Fondunun

dəstəyi ilə hazırlanan “Elektron lüğətlər korpusu”nda isə bir sıra lüğətlərin (<http://korpus.azerbaycandili.az/>) elektron variantları qoyulmuşdur. Bütün bunlar Azərbaycan dilinin milli korpusunun yaradılması üçün ilkin baza rolunu oynaya bilər.

Yuxarıda qeyd edildiyi kimi, “Dilmanc” layihəsi çərçivəsində Azərbaycanda NLP sahəsində bir çox işlər görülmüşdür. Layihə çərçivəsində hazırlanmış, dilin bütün üslublarını əhatə edən ikidilli paralel mətn korpuslarını (ingilis-Azərbaycan ikidilli korpusu – 2 milyon cümlə, türk-Azərbaycan ikidilli korpusu – 277 min cümlə, rus-Azərbaycan ikidilli korpusu – 4,5 milyon cümlə) və konkret dilləri əhatə edən irihəcmli birdilli korpusları (Azərbaycan birdilli korpusu – 60 milyon cümlə, türk birdilli korpusu – 322 milyon cümlə) xüsusi olaraq qeyd etmək olar.

NƏTİCƏ

Korpus dilçiliyi çərçivəsində istifadə olunan linqvistik texnologiyalar, onların tətbiqi və istifadəsi məsələləri mahiyyətinə görə süni intellekt məsələləri ilə üzvi surətdə bağlıdır.

Korpus materialı süni intellekt məsələlərinin həlli prosesində əvəzsiz mənbə kimi uğurla tətbiq oluna bilər. Məhz ona görə də korpus dilçiliyi məsələlərinin və bu sahədə yaradılmış linqvistik texnologiyaların süni intellekt problemləri kontekstində araşdırılması, istifadəsi və dəyərləndirilməsi faydalı olardı.

Yuxarıda göstəriləyi kimi, süni intellekt – elə bir texnologiya, daha dəqiq desək, müasir elmi istiqamətdir ki, insan kimi düşünə bilən öyrədici kompüter,

robotlaşdırılmış texnika, analitik sistem yaratmaq üsullarını öyrənir. Beləliklə, süni intellekt müxtəlif mənbə və elm sahələrindən – informatika, riyaziyyat, dilçilik, psixologiya, biologiya, fiziologiya, biotexnologiya və s. məlumat toplamaqla öyrədici maşın texnologiyasının köməyi ilə verilənlər massivi əsasında insan intellektini təqlid etməyə qabil intellektual sistemlər yaratmağa cəhd göstərir. Bu prosesdə son zamanlar təbii dilin emalı sistemlərində uğurla tətbiq edilən neyron şəbəkələrdən, maşın tərcüməsi, mətnin səsə (nitqə), nitqin (səsin) mətnə çevrilməsi proqramlarından, təbii dili tam ehtiva edən korpus dilçiliyindən, statistik və elektron lüğətlərdən istifadə olunur. Bu sadalananlar robototexnikada, intellektual robotların yaradılmasında mühüm vasitələrdən hesab olunur. Fikrimizcə, əvvəllər riyazi dilçilik, hesablama dilçiliyi, statistik dilçilik, kompüter dilçiliyi, korpus dilçiliyi kimi dilçiliyin yeni sahələrinin tədqiqat obyektinə daxil olan bir çox məsələlər, eləcə də Azərbaycan dilinin milli korpusunun yaradılması sahəsində aparılan işlər süni intellekt sistemləri kontekstində araşdırılmalıdır. Bu, dilçiliyin bir çox problemlərinə yeni baxış bucağından baxmaq, bu istiqamətdə aparılan elmi-tədqiqat işlərinin səmərəsini artırmaq imkanı yarada bilər.

İSTİNADLAR

- [1] Azərbaycan dilinin tezlik lüğəti (söz kökləri). I cild (tərtibçilər – M. Mahmudov, Ə. Fətullayev və b.), Bakı, “Elm”, 2010, 464 s.
- [2] Azərbaycan dilində tabeli mürəkkəb cümlələrin formal meyarlara görə təsnifi və təhlili (M. Mahmudov, Ə.Fətullayev, R.Fətullayev, R. Dərvişov). Filologiya məsələləri. N 7, Bakı, “Elm və Təhsil”, 2015, s. 3-9.
- [3] Azərbaycan qəzet dilinin tezlik lüğəti (tərtibçilər – K. Vəliyeva, M. Mahmudov, V. Pines), Bakı, “Nurlan”, 2004, 264 s.
- [4] “Kitabi-Dədə Qorqud” dilinin statistik təhlili (tərtibçilər – K. Vəliyeva, M. Mahmudov, V. Pines, C. Rəhmanov V. Sultanov), Bakı, “Elm”, 1999, 248 s.
- [5] Mahmudov M., Fətullayev Ə. və b. Mətnin səsləndirilməsi sistemi və Azərbaycan dilinin fonemlərinin fonon-akustik bloku/Terminologiya məsələləri №1, Bakı, 2015, s. 33-49.
- [6] Mahmudov M. Ə. Mətnin formal təhlili sistemi. – Bakı: «Elm», 2002. – 242 s.
- [7] Mahmudov M.Ə., Fətullayev Ə.B. Azərbaycan dilinin əks əlifba lüğəti. – Bakı: «Nurlan», 2004. – 524 s.
- [8] Məhəmməd Füzulinin nəzm əsərlərinin əlifba – tezlik sözlüyü (tərtibçilər – K. A. Vəliyeva, M. Ə. Mahmudov, C. Ə. Rəhmanov, V. S. Sultanov), Bakı, “Elm”, 2004, 548 s.
- [9] Десятков В. В. Системы искусственного интеллекта / Гл. ред. И. Б. Фёдоров. — М.: Изд-во МГТУ им. Н. Э. Баумана, 2001. — 352 с. — (Информатика в техническом университете). — bookwork.ru/book/devyatkov_sii
- [10] Жданов А.А. Автономный искусственный интеллект. — М.: БИНОМ. 5-е изд. электрон. Лаборатория знаний, 2020. — 362 с. — avidreaders.ru/read-book/avtonominyu-iskusstvennyu-intellekt.html
- [11] Лорьер Ж.-Л. Системы искусственного интеллекта. — М.: Мир, 1991. — 568 с. — log-in.ru/books/sistemy-iskusstvennogo-intellekta-1-zh-1-loror-programming
- [12] Нильсон Н. Искусственный интеллект. — М.: Мир, 1973. — 274 с.- log-in.ru/books/nilson-iskusstvennyiy-intellekt-nilson-n-programming
- [13] Рассел С., Норвиг П. Искусственный интеллект: современный подход = Artificial Intelligence: a Modern Approach / Пер. с англ.

- и ред. К. А. Птицына. — 2-е изд. — М.: Вильямс, 2006. — 1408 с. — [studizba.com>files/show/djvu/3615-...>](http://studizba.com/files/show/djvu/3615-...>)
- [14] Смолин Д. В. Введение в искусственный интеллект: конспект лекций. — М.: ФИЗМАТЛИТ.2004. — 208 с. — studmed.ru>smolin-dv-vvedenie-v-is...>
- [15] Хант Э. Искусственный интеллект /Artificial intelligence / Под ред. В. Л. Стефанюка. — М.: Мир, 1978. — 562 с. — login.ru/books/khant-e-iskusstvennyiy-intellekt-khant-1-informatsiya
- [16] Turing Alan (October 1950). Computing Machinery and Intelligence. *Mind*. LIX(236): 433-460. ISSN 0026-4423, doi:10. 1093mind/LIX.236.433.
- [17] Azərbaycan dilinin informasiya bazası <http://azerbaycandili.az>

Kibertəhlükəsizlik problemlərinin həllində NLP effektiv vasitə kimi

Ramiz Şıxəliyev¹, Fərqanə Abdullayeva²

^{1,2}İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

¹shikhramiz61@gmail.com, ²a_farqana@mail.ru

Xülasə — Məqalədə NLP-nin kibertəhlükəsizlikdə tətbiq sahələri araşdırılıb, əsas tədqiqat istiqamətləri müəyyən edilib. Sosial media mənbələrində hücumçular tərəfindən yazılmış mətn tipli kibertəhdid xarakterli informasiyanın klassifikasiyası üçün NLP-yə əsaslanan yanaşma təklif edilib. Təklif olunan metod əvvəlcə kontentdə kibertəhdidlərlə bağlı sentimental izləri analiz edir, sonra isə proqnozlaşdırır. Yanaşma təşkilatların təhdidlərin prioritetləşdirilməsi, təhdidlərin avtomatik modelləşdirilməsi kimi proaktiv qərarlar qəbul etməsinə imkan yarada bilər.

Açar sözlər — Kiberhücum; CTI; NLP, Spam, Fişinq, Zərərli URL

I. GİRİŞ

Kibertəhlükəsizlik İnterneti və İnterneta qoşulmuş sistemləri (avadanlıq, proqram təminatı və verilənlər) kiberhücumlardan mühafizə etmək üçün resursların, texnologiyaların, proseslərin və idarəetmə elementlərinin təşkili təcrübəsidir [1]. Kibertəhlükəsizlik informasiyanın və sistemlərin kiberhücumlardan, verilənlərin sızmasından, şəxsi məlumatların oğurluğundan, zərərli proqramlardan, sosial mühəndislikdən, fişinqdən və digər kibertəhlükəsizlik təhdidlərindən mühafizəni əhatə edir [2].

Hər gün İnternetdə çox böyük həcmdə informasiya istehsal olunur və böyük əksəriyyəti strukturlaşdırılmamışdır. Yalnız strukturlaşdırılmış informasiyanın analizi əsasında kibertəhlükəsizlik problemlərini effektiv həll etmək mümkün deyil. Kibertəhlükəsizlik problemlərinin həlli üçün strukturlaşdırılmamış informasiyanın analiz edilməsinə ehtiyac vardır. Strukturlaşmamış informasiyanı analiz etməyə imkan verən bir sıra texnologiyalar mövcuddur, onlardan biri də Təbii Dilin Emalıdır (ing. Natural Language Processing, NLP). NLP kompüter elmləri, dilçilik və süni intellektin altsahəsidir. NLP mətn və ya nitq məlumatları şəklində təbii insan dilini emal etmək üçün hesablama linqvistikası, statistika, maşın təlimi və dərin təlim üsullarını birləşdirir. NLP qrammatika, semantika, praqmatika və morfologiyanın müxtəlif aspektlərini təhlil edərək təbii insan dilinin strukturunu və mənasını anlamağa imkan verir. Maşın tərcüməsi, məlumatların çıxarılması, ümumiləşdirmə, tibb, sorğu sistemləri və s. kimi müxtəlif sahələrdə istifadə olunur. NLP-nin kibertəhlükəsizlikdə tətbiq olunduğu ən maraqlı istiqamətlərdən biri kibertəhdidlərin intellektual analizidir (ing. Cyber Threat Intelligence, CTI). Bu istiqamət mətn tipli informasiya mənbələrindən təhdidlər

haqqında informasiyanın avtomatik çıxarılması və analizi ideyasına əsaslanır. Kibertəhdidlərin intellektual analizi sahəsində çox sayda üsullar işlənmişdir [3]. Kibertəhdidlər haqqında verilənlərin analizi mətn mənbələri üzərində həyata keçirilir. Belə mənbələrə kiberhücum strategiyalarının, prosedurlarının və alətlərinin şərh olunduğu təhdid hesabatlarını və onlayn bloq yazılarını, haker forumlarını, sosial media yazılarını, onlayn xəbər yazılarını misal göstərmək olar. Kibertəhdidlərin analizini aparmaq üçün supervizorlu və supervizorsuz maşın təlimi metodları təbii dilin emalı üsulları ilə birgə istifadə olunur. Təqdim olunan məqalədə NLP-nin kibertəhlükəsizlikdə tətbiqləri araşdırılıb, “Twitter” mənbələrində hücumçular tərəfindən yazılmış kibertəhdid xarakterli informasiyanın klassifikasiyası üçün yanaşma təklif edilib.

II. NLP-NİN KİBERTƏHLÜKƏSİZLİKDƏ TƏTBİQLƏRİ

Müdaxilələrin aşkarlanmasının ənənəvi üsulları nümunə uyğunluğundan, ya da qara siyahıdan istifadə edir. Hər iki halda naməlum və ya yeni zərərli hücumlar müəyyən edilə bilinmir. Davranışa əsaslanan üsul daha yüksək dəqiqliklə hücumları aşkar edir. Bu üsul görünməz zərərli trafik kimi bəzi tanınmış hücumları aşkar edə bilmir. NLP-nin istifadəsi aşkarlanma dəqiqliyini artırır, çünki NLP əsaslı aşkarlama mexanizmləri hücum metodlarından asılı olmur [4].

NLP-ni zərərli proqramların aşkarlaması üsulu kimi istifadə etmək mümkündür. Məsələn, NLP ilə icra edilə bilən faylların ASCII (American Standard Code for Information Interchange) sətirlərindən istifadə edərək zərərli proqramların aşkarlanması üsulu təklif edilmişdir [5]. ASCII sətirləri sözlərə bölür və sözlər təhlükəsiz və zərərli kimi təsnif edilir. NLP-dən istifadə edərək tez-tez istifadə olunan sözlərin korpusu xüsusiyyət vektoruna çevrilir. Bu zaman, söz tezliyi ilə müqayisədə NLP qeyri-adi sözlərə ehtiyac duymur.

Zərərli URL-lər (Uniform Resource Locator) bugünkü kiberhücumların və kiberfırıldaqaların əksəriyyətində istifadə edilir və e-poçt, mətn mesajları, pop-up-lar və ya reklamlar vasitəsilə istifadəçilərə göndərilir. Bu URL-lərin kliklənməsi və ya skanlanması oğurlanmış e-poçt hesablarına, fişinqə başlanmasına, zərərli proqramların və casus proqramların yüklənməsinə, həmçinin ciddi maliyyə itkilərinə səbəb ola bilər. E-poçtlarda zərərli URL-lərin aşkarlanması üçün maşın təliminə əsaslanan ansambl təsnifatı yanaşması təklif edilmişdir [6]. Bu yanaşmada strukturlaşdırılmamış URL sətirlərindən çıxarılan statik leksik xüsusiyyətlərdən istifadə edilir. Bu xüsusiyyətlərdən istifadə aşkarlama dəqiqliyini və sürətini artırır, çünki o, URL-lərin axtarışını və ya qara siyahıya salınmasını tələb etmir.

Spam filtrasiyası texnologiyaları e-poçt mesajlarında spam və ya fişinqi skan etmək üçün NLP mətn təsnifatından istifadə edir [7]. NLP həm spamerlərin şablonlarını və göndərdikləri mesajların növlərini müəyyən

etmək üçün, həm də e-poçtun daxili strukturunu başa düşmək üçün istifadə edilə bilər.

Fişinq kontekstində NLP, özünü insan kimi təqdim edən maşın tərəfindən göndərilən e-poçt mətnindəki botların və ya spamların davranışını anlamaq üçün istifadə edilə bilər [8]. Sosial mediadan əsas resurs kimi istifadə etdikdə fişinq daha da təsirli olur. NLP-dən istifadə sosial şəbəkələr və digər mənbələr vasitəsilə şəxsi məlumatların toplanması prosesini avtomatlaşdırır və bu, hücumun müəyyən edilməsini çətinləşdirir.

Adətən hücumun aşkarlanması və qarşısının alınması üzrə tədqiqatlar şəbəkənin özünün xüsusiyyətlərinə əsaslanan birbaşa sübutlara, məsələn, anormal trafik monitorinqinə yönəlir. Lakin dolayı sübutlarla hücumların mənbəyi müəyyən edilə bilər: məsələn, sosial media mətni vasitəsilə. Sübut kimi yalnız sosial şəbəkələrdən istifadə edərək DDoS (Distributed denial-of-service) hücumlarının aşkar edilməsi üçün NLP modellərinin istifadəsi təklif edilmişdir [9]. NLP modelləri hücumları aşkar etmək üçün istifadəçi mesajlarından kifayət qədər məlumat çıxarmağa imkan verir.

Şəxsi və konfidensial məlumatlar tez-tez sosial media mesajlarında yer alır. Bu, istifadəçilərin müəyyən sosial qruplarda qarşılıqlı əlaqələrinin sayını artırmaq istəyi, eləcə də məxfilik riskləri barədə zəif məlumatı ilə bağlıdır. Məlumatın məxfiliyini qiymətləndirə bilən texnologiyalar istifadəçini müəyyən məlumatların açıqlanması ilə bağlı risklər barədə xəbərdar etməklə məxfiliyin qorunmasını təmin edə bilər. İstifadəçilər

tərəfindən şəxsi məlumatları aşkar etmək üçün istifadə olunan təkrarlanan təbii dil nümunələri vasitəsilə məxfi məlumatların avtomatik tutulması üçün üsul təklif edilmişdir [10].

NLP-nin tətbiq olunduğu sahələrdən biri də Kiber Təhdidlərin Analizidir (Cyber Threat Analysis). Kiber Təhdidlərin Analizi termininə dünyanın aparıcı təşkilatları müxtəlif təriflər vermişdir. SANS İnstitutunun verdiyi tərif aşağıdakı kimidir:

CTI – qarşı tərəf, onun imkanları, tətbiq etdiyi texnikalar və taktikalar haqqında məlumatın toplanması və onların analizidir (SANS Institute).

CTI verilənləri – kibertəhdidləri aşkarlamağa, qiymətləndirməyə, izləməyə, cavab verməyə imkan verən istənilən informasiyadır. Bu informasiya aşağıdakıları əhatə edir:

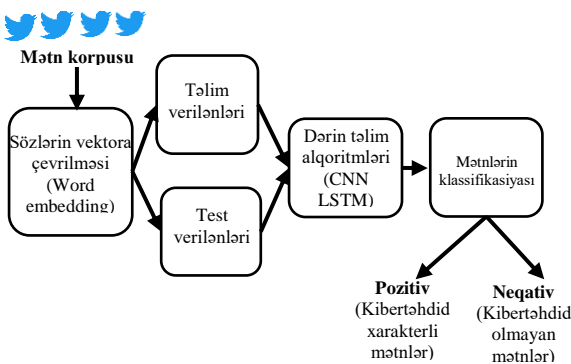
- komprementasiya indikatorları (file hash, IP address)
- təhdid aktorunun istifadə etdiyi taktikalar, texnikalar və proseduralar
- təhdidləri aşkarlamaq üçün təklif edilən əməliyyatlar
- hücumu həyata keçirmək üçün alətlərin adları
- hücum vektoru.

CTI məlumatlarının çıxarılması üçün aşağıdakı üsullar istifadə edilə bilər:

- CTI əsaslı mətnlərin klassifikasiyası;
- Hücum taktikalarının, texnikalarının, prosedurlarının çıxarılması;

- Kibertəhlükəsizliklə bağlı açar sözlərin çıxarılması;
- Kibertəhlükəsizliklə bağlı hadisələrin identifikasiyası;
- Kibertəhdid xəbərdarlıqlarının generasiyası;
- Haker resurslarının analizi;
- Proqram təminatı boşluqları haqqında informasiyanın çıxarılması;
- Təhdid hesabatlarının klassifikasiyası.

İşdə CTI mətnlərinin klassifikasiyası məsələsinə baxılmışdır. CTI mətnlərinin klassifikasiyası üçün təklif edilmiş *Word2Vec+CNN* modelinin arxitekturası Şəkil 1-də verilmişdir.



Şəkil 1. CTI mətnlərinin klassifikasiyası üçün təklif edilmiş *Word2Vec+CNN* modelinin arxitekturası

Haker forumları, təhdid hesabatları, sosial media yazıları, onlayn məqalələr CTI mətnlərinin analizi üçün mənbələrdir.

Təqdim olunan işdə sosial media saytı olan “Twitter” verilənləri kibertəhlükəsizlik hadisələrinin analizi məqsədi ilə istifadə edilmişdir. Təklif edilmiş yanaşmada sosial media saytlarından mətnlər götürülür, mətnlərin alqoritmlərə tanıtılması üçün onlar ilk növbədə ədədlərə çevrilməlidirlər. Bu prosesi həyata keçirmək üçün NLP-də vektorizasiya əməliyyatından istifadə edilir. İngilis ədəbiyyatlarında bunu “Word Embedding” adlandırırlar. Mətnlərin vektorizasiyasını həyata keçirmək üçün müxtəlif üsullar vardır. Bunlardan biri “Word2Vec” üsuludur. “Word2Vec” modelinin iki növü vardır. *Skip-gram* və *CBOW* (Continuous bag-of-words). Təqdim olunan işdə vektorizasiya skip-gram əsasında aparılmışdır. DDoS hücumunun proqnozlaşdırılması üçün kiberhücumla bağlı terminlər lüğətindən istifadə edilmişdir. İstifadə edilmiş lüğət *UFONet*, *Low Orbit Ion Cannon (LOIC)*, *Wannacry*, *Wannacrypt*, *Petya*, *Wcry*, *Petrwrap*, *Hakerlər*, *DDoS*, *DoS*, *Wikileaks*, *lulzsec* və s. kimi təhlükəsizliklə bağlı sözlərin adlarını ehtiva edir.

NƏTİCƏ

NLP üsulları strukturlaşdırılmamış mənbələrdən informasiyanın çıxarılmasını avtomatlaşdırma bilər. Texnologiya inkişaf etdikcə, NLP üsulları kibertəhlükəsizlikdə getdikcə daha çox istifadə olunur. NLP kibertəhlükəsizliyin çoxlu sayda müxtəlif məsələlərinin həllinə tətbiq edilə bilər. Nəzərdən keçirilən tətbiqlər kibertəhlükəsizlikdə NLP üsullarının əhəmiyyətini nümayiş etdirir, həmçinin kibertəhlükəsizlik sahəsində NLP-yə ehtiyacı təsdiqləyir.

Kiberhücumların sosial media verilənləri əsasında proqnozlaşdırılması istiqamətində geniş tədqiqatlar aparılmışdır. Mövcud işlər aşkarlama sisteminin sinifləndirilmiş nümunələr və qeyd olunmuş sayda əlamətlər əsasında öyrədilməsi ideyasına əsaslanır. Bu tip yanaşmalar dinamik təbiətli kiberhücumların aşkarlanmasını həyata keçirə bilmir. İşdə sosial media kontentində kibertəhdidlərin identifikasiyası üçün metod təklif edilmişdir. Bu məqsədlə təklif olunan metod əvvəlcə kontentdə kibertəhdidlərlə bağlı sentimental izləri analiz edir, sonra isə proqnozlaşdırır.

İSTİNADLAR

- [1] D. Craigen, N. Diakun-Thibault, and R. Purse, “Defining Cybersecurity,” *Technol. Innov. Manag. Rev.*, vol. 4, no. 10, pp. 13–21, 2014.
- [2] P.S. Seemaa, N. Sundaresan, M. Sowmiya, “Overview of Cyber Security,” *IJARCCCE*, vol. 7, pp. 125–128.
- [3] R.M. Alguliyev, R.M. Aliguliyev, F.J. Abdullayeva, “The improved LSTM and CNN models for DDoS attacks prediction in social media,” *International Journal of Cyber Warfare and Terrorism (IJCWT)*, vol. 9, no. 1, pp. 1-18, 2019.
- [4] S. Das, M. Ashrafuzzaman, F. T. Sheldon, and S. Shiva, “Network Intrusion Detection using Natural Language Processing and Ensemble Machine Learning,” in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, Dec. 2020, pp. 829–835
- [5] Ryo Ito and Mamoru Mimura, Detecting Unknown Malware from ASCII Strings with Natural Language Processing Techniques, 14th Asia Joint Conference on Information Security (AsiaJCIS), Kobe, Japan, 09 September 2019.
- [6] A. Joshi, L. Lloyd, P. Westin, and S. Seethapathy, “Using Lexical Features for Malicious URL Detection - A Machine Learning Approach,” 2019, p. 6.
- [7] S. Srinivasan, V. Ravi, M. Alazab, S. Ketha, A. M. Al-Zoubi, and S. Kotti Padannayil, “Spam Emails Detection Based on Distributed Word Embedding with Deep Learning,” in *Machine Intelligence and Big Data Analytics for Cybersecurity Applications*, Y. Maleh, M. Shojafar, M. Alazab, and Y. Baddi, Eds. Cham: Springer International Publishing, 2021, pp. 161–189.
- [8] R. Verma, N. Shashidhar, and N. Hossain, “Detecting Phishing Emails the Natural Language Way,” in *Computer Security – ESORICS 2012*, Berlin, Heidelberg, 2012, pp. 824–841.
- [9] N. Chambers, B. Fry, and J. McMasters, “Detecting Denial-of-Service Attacks from Social Media Text: Applying NLP to Computer Security,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, New Orleans, Louisiana, Jun. 2018, pp. 1626–1635.
- [10] G. Canfora, A. Di Sorbo, E. Emanuele, S. Forootani, and C. A. Visaggio, “A Nlp-based Solution to Prevent from Privacy Leaks in Social Network Posts,” in *Proceedings of the 13th International Conference on Availability, Reliability and Security*, New York, NY, USA, Aug. 2018, pp. 1–6.

Azərbaycan dilində mətnlərin oxunabilirlik səviyyəsinin qiymətləndirilməsi metodları

İsmayıl Sadıqov

İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
ismayil.sadigov@gmail.com

Xülasə — Məqalədə ingilis dilindəki mətnlər üçün oxunabilirlik düsturlarının Azərbaycan dilindəki mətnlərə tətbiqi probleminə toxunulur. Bu problemin həlli üçün düsturlardakı orta cümlə uzunluğu, orta söz uzunluğu kimi parametrlərə uyğun əmsalların korrektə edilməsindən, eləcə də “tanış sözlər” siyahısının hazırlanmasından danışılır. İngilis dilindəki mətnlər üçün dörd oxunabilirlik düsturunun Azərbaycan dilindəki mətnlər üçün uyğunlaşdırılmış variantı təqdim olunur.

Açar sözlər — *mətnin mürəkkəbliyi; oxunabilirlik düsturları; Fleşin oxuma asanlıığı düsturu; Fleş-Kinkeydin təhsil səviyyəsi düsturu; Deyl-Çellin yeni oxunabilirlik düsturu; Speşin oxunabilirlik düsturu.*

I. GİRİŞ

İnsanlar xüsusən məktəb illərində qazandıqları biliklərin böyük bir hissəsini dərslərdən alırlar. Tədqiqatlara görə şagirdlərdə oxu vərdişlərinin inkişafını ləngidən əsas amillərdən biri dərslərin mətnlərinin mürəkkəbliyi, daha dəqiqi, mətnlərin mürəkkəblik dərəcəsi ilə şagirdlərin yaş səviyyəsi arasında uyğunsuzluqdur. Həddən artıq mürəkkəb

mətnləri şagirdlər başa düşmədən əzbərləməyə məcbur olurlar.

Mətnin mürəkkəbliyi mətnin obyektiv xassəsi olub, onu oxuyan insandan asılı deyil. Mətndəki komponentlərin qiymətini, məsələn, cümlələrin uzunluğunu, mücərrəd sözlərin nisbətini və başqa parametrləri dəyişməklə onun mürəkkəbliyini də dəyişmək olar.

Mətnin mürəkkəbliyi onun başa düşülməsini çətinləşdirir. *Mətnin çətinliyi* təkcə onun mürəkkəbliyindən deyil, həm də oxucunun hazırlıq səviyyəsindən asılıdır. Bir (hazırlıqlı) oxucu üçün asan olan mətn, başqa bir (hazırlıqsız) oxucu üçün çətin ola bilər.

II. MƏTNİN MÜRƏKKƏBLİK FAKTORLARININ AŞKARLANMASI

Mətnin mürəkkəbliyinə təsir edən faktorları aşkarlamaq üçün belə bir sadə üsuldan istifadə olunur: Ekspert qiymətləndirməsi əsasında sadə və ya mürəkkəb hesab edilən müəyyən sayda mətn götürülür. Bu mətnlər ayrı-ayrı göstəricilər – məntiqi quruluş, mövzu, cümlələrin uzunluğu və başqa parametrlər üzrə müqayisə edilir. Əgər göstəricinin qiyməti asan mətnlərdən mürəkkəb mətnlərə keçdikcə dəyişilirsə, demək, bu

göstərici mətnin mürəkkəbliyi faktorlarından biridir. Məsələn, əgər sadə və mürəkkəb mətnlərin müqayisəsi zamanı mürəkkəb mətnlərdə daha çox naməlum sözlərin və uzun cümlələrin olması müəyyən olunursa, deməli, mətnin mürəkkəbliyi sözlərin tanışı olmasından və cümlələrin uzunluğundan asılıdır.

Tədqiqatlar nəticəsində mətnin mürəkkəbliyinə ən çox təsir edən faktorlar müəyyən edilmişdir. Onlardan bəziləri aşağıdakı cədvəldə göstərilmişdir (Cədvəl 1).

Cədvəl 1. Mətnlərin mürəkkəbliyinə təsir edən faktorlar

№	Parametrin (faktorun) adı
1	Orta cümlə uzunluğu (sözlərlə) – Average sentence length in words
2	Orta cümlə uzunluğu (hecalarla) – Average sentence length in syllables
3	Sadə cümlələrin sayı – Number of simple sentences
4	Bir abzasdakı cümlələrin sayı – Number of sentences per paragraph
5	Birhecalı sözlərin faizi – Percentage of monosyllables
6	Çoxhecalı sözlərin faizi – Percentage of polysyllables
7	Fərqli sözlərin faizi – Percentage of different words
8	Birinci, ikinci və üçüncü şəxs əvəzlərinin sayı – Number of first, second, and third-person pronouns

III. “OXUNABİLİRLİK” ANLAYIŞI VƏ OXUNABİLİRLİK DÜSTURLARI

Mətnin mürəkkəbliyi səviyyəsini göstərmək üçün “*oxunabilirlik*” (ingilis

dilində “*readability*”) terminindən istifadə olunur.

Hazırda hər hansı mətnin, xüsusən də şagirdlərin yaş xüsusiyyətlərini nəzərə almaqla tədris mətnlərinin oxunabilirliyini qiymətləndirmək üçün bir sıra riyazi modellər – *oxunabilirlik düsturları* mövcuddur. Elmi və populyar ədəbiyyatda *oxunabilirlik düsturları (readability formulas)* ifadəsi ilə yanaşı, *oxunabilirlik testləri (readability tests)*, *oxunabilirlik metrikası (readability metrics)* ifadələrindən də istifadə olunur.

Oxunabilirlik sahəsində ilk elmi tədqiqatlara 1920-ci illərdə ABŞ-da başlanılmış və ötən yüz ildə 100-lərlə oxunabilirlik düsturu işlənib hazırlanmışdır. Təbii ki, bu düsturların hamısı populyarlıq qazana bilməmişdir. Ancaq onlardan bəziləri bugün də geniş istifadə olunur:

- Fleşin oxuma asanlıığı düsturu (Flesch Reading Ease Formula);
- Fleş-Kinkeydin təhsil səviyyəsi düsturu (Flesch-Kincaid Grade Level Formula);
- Deyl-Çellin oxunabilirlik düsturu (Dale-Chall Readability Formula);
- Speşin oxunabilirlik düsturu (Spache Readability Formula);
- SMOG oxunabilirlik düsturu (SMOG Readability Formula);
- Frayn oxunabilirlik düsturu (Fry Readability Formula);
- Qanningin duman indeksi (Gunning Fog Index);
- FORCAST oxunabilirlik düsturu (FORCAST Readability Formula);

IV. ORTA CÜMLƏ UZUNLUĞU VƏ ORTA SÖZ UZUNLUĞUNA ƏSASLANAN OXUNABİLİRLİK DÜSTURLARI

Oxunabilirlik düsturları içərisində ən geniş istifadə olunanı *Fleşin oxuma asanlığı düsturudur*. Bu düstur iki dəyişənə (faktora) – mətndəki cümlələrin orta uzunluğuna və sözlərdəki hecaların orta sayına əsaslanır:

$$K_{en} = 206.835 - (1.015 \times S) - (84.6 \times W), \quad (1)$$

burada K_{en} – mətnin mürəkkəbliyinin (oxuma asanlığının) göstəricisi, S – orta cümlə uzunluğu (sözlərin sayının cümlələrin sayına nisbəti), W isə sözdəki hecaların orta sayıdır (hecaların sayının sözlərin sayına nisbəti). [1]

Düsturun asan tətbiqi üçün belə bir metodika təklif olunur: hər hansı mətndən 100 söz götürülür, cümlələrin sözlərlə orta uzunluğu (orta cümlə uzunluğu) və sözlərin hecalarla orta uzunluğu (orta heca sayı) hesablanır. Oxuma asanlığının qiyməti 100-dən 0-dək dəyişə bilər (Cədvəl 2). Ancaq hazırda çoxlu sayda proqram təminatı olduğundan mətnin ölçüsünün böyük götürülməsi hesablama işlərini çətinləşdirmir.

Cədvəl 2. Fleşin oxuma asanlığı qiymətləri

Oxuma asanlığının qiyməti (K_{en})	Təsviri
0 – 29	çox çətin mətn
30 – 49	çətin mətn
50 – 59	nisbətən çətin mətn
60 – 69	standart mətn
70 – 79	nisbətən asan mətn
80 – 89	asan mətn
90 – 100	çox asan mətn

Geniş istifadə olunan başqa bir oxunabilirlik düsturu – *Fleş-Kinkeydin təhsil səviyyəsi düsturudur*. Əslində, bu düstur Fleşin oxuma asanlığı düsturunun təkmilləşdirilmiş variantıdır. Bu düsturun tətbiqindən alınan qiymət mətnin hansı təhsil səviyyəsi üçün nəzərdə tutulduğunu göstərir. Düstur aşağıdakı kimidir:

$$GL_{en} = (0.39 \times S) + (11.8 \times W) - 15.59, \quad (2)$$

burada GL_{en} uyğun təhsil səviyyəsini göstərir. Məsələn, 8.3 qiyməti mətnin orta səviyyəli 8-ci sinif şagirdi (adətən, 12–14 yaş) üçün anlaşılacaq olacağını bildirir. Düsturdan görüldüyü kimi, cümlələrdəki sözlərin sayı və sözlərdəki hecaların sayı nə qədər az olarsa, alınan qiymət də bir o qədər kiçik olar, yəni mətni daha aşağı yaşlı oxucular başa düşə bilər.

V. MƏTNDƏKİ TANIŞ OLMAYAN SÖZLƏRİN SAYINA ƏSASLANAN OXUNABİLİRLİK DÜSTURLARI

1948-ci ildə ABŞ maarifçisi Edqar Deyl və psixoloq, yazıçı Janna Çell hazırda çox populyar və etibarlı hesab edilən oxunabilirlik düsturunu təklif etmişlər. [2] *Deyl-Çellin oxunabilirlik düsturunda* orta cümlə uzunluğu və mürəkkəb sözlərin faizindən istifadə olunmuşdur. Burada “mürəkkəb sözlər” dedikdə müəlliflər tərəfindən tərtib olunmuş və 80%-i 4-cü sinif səviyyəsində oxucuya məlum olan 3000 sadə sözdən ibarət siyahıya daxil olmayan sözlər nəzərdə tutulur. Əsas ideya ondan ibarətdir ki, oxucu tanış sözləri daha asan qəbul edir və başa düşür. Deyl-Çellin oxunabilirlik düsturu Fleşin oxuma asanlığı

düsturunun bəzi çatışmazlıqlarını aradan qaldırmaq məqsədilə işlənib hazırlanmışdır.

1995-ci ildə E.Deyl və J.Çell 3000 sözlük siyahını yenidən işləmiş və öz düsturlarının yeni versiyasını dərc etdirmişlər. [3] *Deyl-Çellin yeni oxunabilirlik düsturu* özündən əvvəlki düsturların üstün cəhətlərini nəzərə almış və mətnin oxunabilirliyini daha dəqiqliklə qiymətləndirir. Düstur aşağıdakı kimidir:

$$Raw\ Score = 0.1579 \times PDW + 0.496 \times ASL, \quad (3)$$

burada *Raw Score* – ilkin (xam) bal; *PDW* – mürəkkəb sözlərin (Deyl-Çellin siyahısında olmayan sözlərin) faizi; *ASL* – orta cümlə uzunluğudur.

Əgər *PDW* 5%-dən yüksək olarsa, yekun balı almaq üçün ilkin balın üzərinə 3.6365 əlavə olunur, yəni *Adjusted Score = Raw Score + 3.6365*, əks halda, *Adjusted Score = Raw Score*. Düsturun tətbiqi nəticəsində alınan qiymətin təhsil səviyyəsinə uyğunluğu aşağıdakı cədvəldə göstərilmişdir (Cədvəl 3).

Cədvəl 3. Deyl-Çell düsturundan alınan qiymətin təhsil səviyyəsinə uyğunluq cədvəli

Yekun bal	Təhsil səviyyəsi
4.9 və aşağı	orta səviyyəli 4-cü və ondan aşağı sinif şagirdi asanca başa düşür
5.0–5.9	orta səviyyəli 5-ci və ya 6-cı sinif şagirdi asanca başa düşür
6.0–6.9	orta səviyyəli 7-ci və ya 8-ci sinif şagirdi asanca başa düşür
7.0–7.9	orta səviyyəli 9-cu və ya 10-cu sinif şagirdi asanca başa düşür
8.0–8.9	orta səviyyəli 11-ci və ya 12-ci sinif şagirdi asanca başa düşür
9.0–9.9	orta səviyyəli kollec tələbəsi asanca başa düşür
10.0 və yuxarı	orta səviyyəli kollec məzunu asanca başa düşür

Deyl-Çellin yeni oxunabilirlik düsturuna bənzəyən başqa bir düstur 1953-cü ildə Corc Speş təklif etmişdir. Bu düsturda da məntdəki orta cümlə uzunluğu və “mürəkkəb” sözlərin faizi kimi parametrlərdən istifadə olunmuşdur. *Speşin oxunabilirlik düsturu* adlandırılan bu düstur ibtidai təhsil səviyyəsi (1–4-cü siniflər) üçün mətnlərin oxunabilirliyini qiymətləndirmək üçün nəzərdə tutulmuşdur. [4]

$$Index\ Spache = 0.141 \times ASL + 0.086 \times PDW + 0.839, \quad (4)$$

burada *PDW* – mürəkkəb (tanış olmayan) sözlərin faizi; *ASL* – orta cümlə uzunluğudur.

C.Speş 3000 sözlük siyahı əvəzinə, öz siyahısını təklif etmişdir: Speşin siyahısı ibtidai sinif şagirdlərinin bildiyi təxminən 1000 sözdən ibarətdir və bu siyahı ABŞ məktəblərinin 3-cü sinif şagirdləri ilə aparılan eksperiment nəticəsində müəyyən olunmuşdur: şagirdlərin 80%-nin qeyd etdiyi söz *tanış sözlər siyahısına* daxil edilmişdir.

VI. OXUNABİLİRLİK DÜSTURLARIN MODİFİKASIYA EDİLMƏSİ

Nəzərə almaq lazımdır ki, bu düsturların hamısı təcrübi yolla ingilis dilində olan mətnlər üçün alınmışdır. Hər bir dilin özünəməxsus sintaksisi, yəni söz və cümlə quruluşu vardır. Məsələn, ingilis dilində sözlərin orta uzunluğu Azərbaycan, yaxud rus dili ilə müqayisədə qısadır. Bundan başqa, müxtəlif dillərdə sözlərdə olan hecaların miqdarı fərqli qaydada hesablanır: dillərin əksəriyyətində sözdəki hecaların

sayı həmin sözdəki saiflərin miqdarı ilə müəyyən olunsada, bəzi dillərdə bu say tələffüzdən asılı olur. Məsələn, ingilis dilində “analyse” sözü üç hecadan ibarətdir, ancaq “analyses” sözü dörd hecalıdır. Ona görə də, bir dildəki mətnlər üçün hazırlanmış oxunabilirlik düsturunu başqa dildəki mətnlərə tətbiq etdikdə düzgün olmayan nəticələr alınabilir. Daha dəqiq desək, ingilis dilinin sintaksisinə uyğun hazırlanmış oxunabilirlik düsturlarını Azərbaycan dilində olan mətnlərə tətbiq etmək üçün həmin düsturlardakı əmsallar korrektə edilməlidir.

Tədqiqat göstərmişdir ki, ingilis dilindəki mətnlər Azərbaycan dili ilə müqayisədə orta hesabla 0.77 cümlə uzundur, ingilis dilindəki sözlər isə Azərbaycan dilindəki sözlərdən 1.91 heca qıtsadır. [5] Yəni Fleşin oxuma asanlığı düsturunu Azərbaycan dilindəki mətnlərə tətbiq etmək üçün orta cümlə uzunluğunun əmsalı (S) 0.77 dəfə, hecalarla orta söz uzunluğunun (W) əmsalı isə 1.91 dəfə korrektə edilməlidir. Nəticədə *Azərbaycan dilindəki mətnlər üçün Fleşin oxuma asanlığı düsturu* aşağıdakı şəkildə düşür:

$$K_{az} = 206.835 - (1.318 \times S) - (44.3 \times W). \quad (5)$$

Eyni qayda ilə Fleş-Kinkeyd düsturu da Azərbaycan dili mətnləri üçün uyğunlaşdırılmışdır. Beləliklə, *Azərbaycan dilindəki mətnlər üçün Fleş-Kinkeydin təhsil səviyyəsi düsturu* aşağıdakı şəkildə olmuşdur: [6]

$$GL_{az} = (0.51 \times S) + (6.18 \times W) - 15.59 \quad (6)$$

Azərbaycan dilindəki mətnlər üçün modifikasiya olunmuş növbəti *eyl-Çellin*

yeni oxunabilirlik düsturu (3) və *Speşin oxunabilirlik düsturudur* (4). Bu düsturlardakı orta cümlə uzunluğu (ASL) parametrisinin əmsalının korrektə edilməsi ilə bağlı problem yoxdur. Bu korrektəni etdikdən sonra *Azərbaycan dilindəki mətnlər üçün Deyl-Çellin yeni oxunabilirlik düsturu* aşağıdakı şəkildə düşəcəkdir:

$$\begin{aligned} \text{Əgər } PDW > 5\% \text{ isə, onda:} \\ \text{Adjusted Score} &= \text{Raw Score} + 3.6365, \quad (7) \\ \text{əks halda:} \\ \text{Adjusted Score} &= \text{Raw Score}. \end{aligned}$$

Eyni qayda ilə Speşin oxunabilirlik düsturu uyğunlaşdırıldıqdan sonra *Azərbaycan dilindəki mətnlər üçün Speşin oxunabilirlik düsturu* aşağıdakı şəkildə olacaqdır: [7]

$$GL = 0.183 \times ASL + 0.086 \times PDW + 0.839 \quad (8)$$

Əsas problem (7) və (8) düsturlarını Azərbaycan dilindəki mətnlərə tətbiq etmək üçün *PDW* kəmiyyətinin qiymətini, yəni mürəkkəb sözlərin faizini hesablamaqdadır. Çünki bunun üçün E.Deyl və J.Çell tərəfindən ingilis dili üçün hazırlanmış 3000 sözlük siyahının, eləcə də C.Speş tərəfindən təklif olunmuş ibtidai sinif şagirdlərinin bildiyi təxminən 1000 sözdən ibarət siyahının Azərbaycan dili üçün bənzərlərinin olması vacibdir. Ancaq Azərbaycan dili üçün indiyədək bənzər siyahılar olmadığından məqalə müəllifi tərəfindən belə siyahılar da hazırlanmışdır.

Azərbaycan dilindəki mətnlərin mürəkkəblilik səviyyəsini qiymətləndirmək üçün məqalə müəllifi tərəfindən iki proqram təminatı da işlənilib hazırlanmışdır.

Bunlardan biri “*Mətn analizi*” tətbiqi proqramı, digəri isə *www.oxunabilir.az* veb-proqramıdır. [8]

NƏTİCƏ

- Digər oxunabilirlik düsturlarının Azərbaycan dili üçün modifikasiya edilməsindən ötrü müvafiq proqram təminatının olması vacibdir.
- Belə proqram təminatının hazırlanmasına maddi dəstəyin olması lazımdır.
- Kompüterlərdə hazırlanmış və çap olunmuş bütün kitablardakı mətnlərin nəşriyyatlar tərəfindən “xam” şəkildə vahid mərkəzə təhvil verilməsi və tədqiqatçılar üçün həmin resursların əlçatanlığının təmin edilməsi vacibdir.

İSTİNADLAR

[1] Flesch R. Marks of a readable style. Columbia University contributions to education, 1943, no. 897. New York: Bureau of Publications, Teachers College, Columbia University.

- [2] Dale E., Chall J.S. A formula for predicting readability. Educational research bulletin. 1948, Jan. 21 and Feb. 17, 27: 1-20, 37-54.
- [3] Chall J.S., Dale E. Readability revisited, the new Dale-Chall readability formula. Cambridge, MA: Brookline Books, 1995.
- [4] Spache G. A new readability formula for primary-grade reading materials // Elementary school journal, 1953, 53: pp.410-413.
- [5] Sadıqov İ.C. Azərbaycan dili mətnlərinin mürəkkəbliyinin qiymətləndirilməsi üçün modifikasiya olunmuş Fleş düsturu // "İnformasiya texnologiyaları problemləri" jurnalı, 2018, №1, s.46-58.
- [6] Sadıqov İ.J. Mathematical and information models for evaluating readability of texts in Azerbaijani language // El-Cezeri Journal of Science and Engineering, V.5 – N.3, 2018, pp. 888-903.
- [7] Sadıqov İ.C. Azərbaycan türkçesindeki metinlər için kelime sıklıklarına dayalı okunabilirlik formülleri // “Uluslararası Mühendislik ve Doğa Bilimleri Çalışmaları Kongresi”, Ankara, 07-09 may 2021.
- [8] Sadıqov İ.C. Azərbaycan dili mətnlərinin mürəkkəbliyinin qiymətləndirilməsi üçün modifikasiya olunmuş Fleş düsturu // "İnformasiya texnologiyaları problemləri" jurnalı, 2018, №1, s.46-58.

Development of a real-time speech recognition system for the Azerbaijani language

Alakbar Valizada

Azerbaijan Technical University, Baku, Azerbaijan

alakbar.valizada@aztu.edu.az

Abstract — Real-time automatic speech recognition (ASR) systems have become increasingly important in today's fast-paced and interconnected world. This paper presents a real-time Automatic Speech Recognition (ASR) system for the Azerbaijani language, using a hybrid Hidden Markov Model/Deep Neural Network (HMM/DNN) model for acoustic modeling and a syllable-based n-gram model for language modeling. The ASR system is designed to handle streaming speech data transmitted via WebSockets, enabling efficient real-time recognition of speech in various applications, such as virtual assistants and transcription services. The proposed system leverages the Kaldi and SRILM toolkits for model training, and incorporates a WebSocket-based communication framework to facilitate seamless data transfer between the client and server. Experiments show that the syllable-based subword modeling approach is particularly effective for Azerbaijani, an agglutinative language, and the real-time ASR system delivers promising results in terms of accuracy and performance.

Keywords — ASR; HMM/DNN; n-gram; WebSocket; Kaldi; agglutinative languages; real-time ASR

I. INTRODUCTION

Automatic speech recognition (ASR) has become a critical technology in recent years, with applications in a wide range of fields, such as transcription services, virtual assistants, and human-computer interaction. As the demand for real-time speech recognition grows, it is essential to develop systems capable of processing and transcribing spoken language with minimal latency. This is especially crucial for under-resourced languages, where the availability of ASR systems is limited. One such language is Azerbaijani, an agglutinative language with a rich morphological structure, which presents unique challenges for ASR systems.

Previous research in Azerbaijani ASR has mainly focused on offline speech recognition systems for specific applications, such as emergency call centers [1] and taxi call service systems [2]. While these offline systems have made significant progress, the demand for real-time ASR systems is rapidly increasing, highlighting the need for efficient and accurate real-time speech recognition for Azerbaijani.

This paper introduces a real-time ASR system for the Azerbaijani language,

specifically designed to handle streaming speech data. The proposed system utilizes a WebSocket-based communication framework, which allows efficient transfer of speech data in real-time between clients and the recognition server. The system is built upon a hybrid HMM and time-delay neural network (TDNN) architecture for acoustic modeling, and employs a syllable-based n-gram model for language modeling. This combination allows the system to better handle the complexities of the Azerbaijani language, which often features a large number of unique word forms due to its agglutinative nature.

The implementation of the system relies on the Kaldi [3] and SRILM toolkits [4] for model training and evaluation. The paper presents the experimental setup, results, and performance metrics of the system, highlighting its effectiveness in real-time speech recognition for the Azerbaijani language.

The remainder of this paper is organized as follows: Section II provides an overview of related work in ASR and WebSocket-based communication. Section III details the experimental setup and results, including data preparation, model training, and evaluation metrics. Finally, Section IV concludes the paper and discusses future work.

II. SYSTEM OVERVIEW

A. Architecture

The proposed system consists of the following components:

- 1) *Client-side application*: Records user speech and streams audio data to the server using WebSocket.
- 2) *Server-side application*: Receives audio data, performs speech recognition using a hybrid HMM/DNN model, and sends recognition results back to the client.
- 3) *Acoustic model*: A hybrid HMM/DNN model trained using the Kaldi toolkit.
- 4) *Language model*: An n-gram language model created using the SRILM toolkit.

B. WebSocket Implementation

WebSocket is a protocol that enables bidirectional, low-latency communication between a client and a server over a single, long-lived connection (Fig. 1). This protocol is particularly suitable for real-time applications such as speech recognition, where data needs to be streamed continuously from the client to the server, and recognition results need to be returned promptly. In this section, we will discuss the server-side and client-side implementations of WebSocket for the proposed real-time speech recognition system for Azerbaijani.

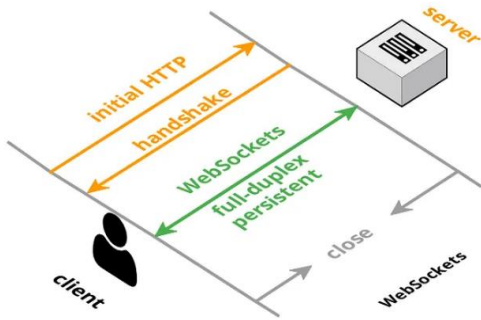


Fig.1. High-level overview of a WebSocket communication

The proposed system consists of the following components:

1) *Client-Side Implementation*

a) *Recording audio data:* The client application, which can be a web or mobile application, initiates a WebSocket connection to the server by providing the server's URL and the desired connection protocol. Upon establishing the connection, the client application maintains the connection throughout the user's interaction with the system.

The client-side application records the user's speech using the device's microphone and an appropriate audio recording API. The recorded audio data is typically stored in a buffer, which is a temporary storage area in memory.

To segment the audio data into smaller chunks, the client application sets up a timer or an event-based callback mechanism that triggers at regular intervals, say every T millisecond. When the timer or event is triggered, the client application retrieves the audio data from the buffer corresponding to the specified time interval.

For example, if the buffer size is set to 16,000 samples (corresponding to 1 second of audio at a 16 kHz sampling rate), and the timer interval is set to 250 milliseconds, the client application would retrieve 4,000 samples ($250 \text{ ms} * 16 \text{ samples/ms}$) from the buffer at each interval. These samples represent an audio chunk that is ready for transmission.

b) *Sending audio data:* Once the audio chunk has been prepared, the client application sends it to the server over the WebSocket connection. The audio data is typically encoded into a binary format, such as PCM, which preserves the original audio information and ensures efficient transmission.

Each WebSocket message sent to the server includes the binary-encoded audio chunk and any necessary metadata, such as the client's session ID, the current audio timestamp, or a sequence number to maintain the order of the chunks. The metadata can be sent as a separate message or included within the binary message, depending on the implementation.

c) *Receiving recognition results:* The client application listens for incoming messages from the server, which contain the recognition results. Upon receiving a message, the client application decodes the JSON object and updates the user interface to display the recognized speech in real-time.

2) *Server-Side Implementation:* On the server side, the application receives the incoming WebSocket messages containing the audio chunks and metadata. The server

is responsible for decoding the messages, extracting the audio data, and appending it to the appropriate audio buffer associated with the client's session ID.

As the audio chunks are received in the correct order (based on the sequence number or timestamp), the server application can concatenate them to reconstruct the original audio stream. The server maintains an audio buffer for each client session, ensuring that the audio data is correctly processed for each user.

When a sufficient amount of audio data has been accumulated in the buffer, the server processes the buffered audio using the ASR engine, which performs the speech recognition task. The length of the buffered audio can be determined by a fixed duration (e.g., 1 second), or adaptively based on the system's requirements and performance.

Once the ASR engine completes the recognition task, the server prepares the recognition results as a JSON object and sends it back to the corresponding client over the WebSocket connection, allowing the client to receive the results in real-time.

This process of segmenting audio data into chunks, transmitting them over WebSocket, and server-side collection and processing ensures that the real-time speech recognition system for Azerbaijani can efficiently and accurately recognize speech in a real-time scenario.

3) Acoustic model

The hybrid HMM/DNN architecture used in this study combines the strengths of HMMs and DNNs for acoustic modeling.

a) HMM Component

The HMM component is responsible for modeling the temporal structure of speech using a sequence of hidden states, $= q_1, q_2, \dots, q_T$, and observations, $O = o_1, o_2, \dots, o_T$. O observations represent the sequence of T acoustic feature vectors extracted from the speech signal at regular time intervals. Q hidden states represent the sequence of T underlying phonetic units or triphones corresponding to the observed acoustic features. The hidden states are not directly observable, and the goal of the HMM-based ASR system is to determine the most likely sequence of hidden states given the observed acoustic features [5].

The HMM is defined by the following parameters:

- State transition probabilities (A): $a_{ij} = P(q_t = j | q_{t-1} = i)$ represents the probability of transitioning from state i to state j .
- Emission probabilities (B): $b_j(o_t) = P(o_t | q_t = j)$ represents the probability of observing feature vector o_t at time t , given the state j .
- Initial state probabilities (π): $\pi_i = P(q_1 = i)$, represents the probability of starting in state i .

The likelihood of a given observation sequence, O , can be calculated using the forward algorithm, which computes the forward probability, $\alpha_t(i)$, as follows:

$$\alpha_t(i) = P(o_1, o_2, \dots, o_t, q_t = i | \lambda), \quad (1)$$

where $\lambda = (A, B, \pi)$ are the HMM parameters.

b) DNN Component

While HMMs are effective in modeling the temporal aspects of speech, they are limited in their ability to model the complex relationships between the acoustic features and the phonetic units. DNNs have proven to be highly effective in modeling these complex relationships due to their ability to learn hierarchical representations of the input features [6].

In the hybrid HMM/DNN architecture, the DNN is used to model the observation likelihoods $b_j(o_t)$ for each HMM state j . Specifically, the DNN is trained to predict the posterior probabilities $P(q_t=j / o_t)$ of the HMM states given the acoustic features o_t . Once trained, the DNN can be used to estimate the observation likelihoods for the HMM by applying Bayes' rule:

$$b_j(o_t) = P(o_t | q_t = j) = P(q_t = j | o_t) * P(o_t) / P(q_t = j), \quad (2)$$

where $P(o_t)$ and $P(q_t=j)$ are the priors for the observations and states, respectively.

By combining the strengths of HMMs and DNNs, the hybrid HMM/DNN architecture allows for more accurate modeling of the complex relationships between the acoustic features and phonetic units in speech signals, leading to improved recognition performance.

For the training of the AM, we used a time-delay neural network (TDNN) architecture [7] with maximum mutual information (MMI) sequence-level objective function [8]. TDNNs are a type of deep neural network (DNN) that have been widely adopted for speech recognition tasks due to their ability to model both short-term

and long-term temporal dependencies in the input data.

4) Language Modeling

The n-gram language model computes the probability of a sequence of words, $W = w_1, w_2, \dots, w_N$, based on the joint probability of the words. Using the chain rule of probability, we can write:

$$P(W) = P(w_1, w_2, \dots, w_N) = \prod_{i=1}^N P(w_i | w_{i-1}, w_{i-2}, \dots, w_1), \quad (3)$$

for an n-gram model, the probability of a word depends only on the previous n-1 words:

$$P(w_i | w_{i-1}, w_{i-2}, \dots, w_1) \approx P(w_i | w_{i-1}, w_{i-2}, \dots, w_{i-n+1}), \quad (4)$$

III. EXPERIMENTAL SETUP AND RESULTS

A. Data Preparation

1) Acoustic Model Data

For the acoustic model, audio data were collected from two main sources. The first source was a Telegram bot, which allowed users to contribute speech samples through a crowdsourcing approach. Users were prompted to record and submit phrases from a predefined list, ensuring the diversity and coverage of the collected data. The second source of audio data was from YouTube videos. Videos containing Azerbaijani speech were selected, and the audio tracks were extracted for further processing. To create a balanced dataset, both sources of data were combined, segmented, and annotated with corresponding transcriptions. The resulting dataset was then split into training, validation, and testing sets to enable the training and evaluation of the HMM/TDNN acoustic model.

2) *Language Model Data*

The language model was trained using text data collected from two main sources: Azerbaijani Wikipedia [9] and the AzerTag news portal (azertag.az) [10]. These sources were chosen due to their extensive coverage of various topics, which ensured the language model would be exposed to a wide range of vocabulary and linguistic structures. The collected text data were cleaned, tokenized, and processed to generate an n-gram language model using the SRILM toolkit [4].

A significant modification was made to the language modeling approach by adopting a syllable-based subword model instead of a traditional word-based model. This approach was chosen due to the agglutinative nature of the Azerbaijani language, which often results in a large number of unique word forms. By using a syllable-based subword model, the system can better handle out-of-vocabulary words and achieve more accurate recognition results [11].

B. *Acoustic and Language Model Training*

The experiments described in this paper were carried out using both the Kaldi and SRILM toolkits. The configuration of the Kaldi speech recognition system is based on the Kaldi chain recipe (s5_r2) of the TED-LIUM corpus, which integrates a hybrid DNN/HMM system. The "nnet3 chain" implementation was utilized for training the acoustic model (AM), employing a TDNN architecture with a maximum mutual information (MMI) sequence-level objective function. The context-dependent states,

acquired through the forced alignments of a GMM/DNN baseline system, served as targets for DNN training. Additionally, online i-vectors were used as input to the TDNN for improved speaker adaptation.

For the language model training, the SRILM toolkit was employed to train various language models (LMs). Based on our prior experience [1], modified Kneser-Ney smoothing has proven to be the most effective method for handling out-of-vocabulary (OOV) words and large text corpora.

C. *Evaluation Metrics*

The system was evaluated using the Word Error Rate (WER) and real-time factor (RTF) metrics. The WER measures (based on the Levenshtein distance?) the recognition accuracy, while the RTF indicates the efficiency and speed of a ASR system.

D. *Results*

The proposed system achieved a WER of 5.59%, which is competitive with state-of-the-art ASR systems for Azerbaijani. The system's performance can be further improved by incorporating additional data sources and refining the acoustic and language models.

The real-time factor (RTF) of the proposed system was found to be 0.165, indicating that the system processes the input speech efficiently, making it suitable for real-time applications. The use of WebSockets for streaming speech data and the optimized hybrid HMM/DNN model contribute to the system's efficiency.

IV. CONCLUSION AND FUTURE WORK

This paper presents a real-time speech recognition system for the Azerbaijani language, leveraging WebSockets for data streaming and a hybrid HMM/DNN model for ASR. The proposed system demonstrates promising results in terms of both accuracy and efficiency, making it a valuable contribution to the field of ASR for Azerbaijani. The Kaldi toolkit and SRILM toolkit were used for training the system, providing a solid foundation for further research and development.

Future work can focus on improving the system's performance by incorporating additional data sources, refining the acoustic and language models, and exploring alternative ASR architectures such as end-to-end neural models. Additionally, the system can be extended to support multilingual and code-switching scenarios, broadening its applicability and usefulness in real-world applications.

In conclusion, the proposed real-time speech recognition system for Azerbaijani offers a robust and efficient solution for ASR tasks, paving the way for a range of applications such as transcription services, voice assistants, and real-time translation.

ACKNOWLEDGMENT

I would like to thank Dr. Samir Rustamov for discussions and suggestions. This research experiments have been implemented in the Center for Data Analytics Research (CeDAR) at the ADA University/

REFERENCES

- [1] A. Valizada, N. Akhundova, S. Rustamov, “Development of Speech Recognition Systems in Emergency Call Centers.” *Symmetry* 2021, 13, 634, doi: 10.3390/sym13040634
- [2] S. Rustamov, N. Akhundova, A. Valizada. “Automatic Speech Recognition in Taxi Call Service Systems,” In *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering Emerging Technologies in Computing*; Springer: Cham, Switzerland, 2019; pp. 243–253.
- [3] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, J. Silovsky: “The Kaldi speech recognition toolkit,” In: *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding (No. EPFL-CONF-192584)*. IEEE Signal Processing Society (2011)
- [4] A. Stolcke (2002), SRILM - An Extensible Language Modeling Toolkit. *Proc. Intl. Conf. on Spoken Language Processing*, vol. 2, pp. 901-904, Denver.
- [5] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," in *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-286, Feb. 1989, doi: 10.1109/5.18626.
- [6] G. Hinton et al., "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," in *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82-97, Nov. 2012, doi: 10.1109/MSP.2012.2205597.
- [7] V. Peddinti, D. Povey, S. Khudanpur, “A time delay neural network architecture for efficient modeling of long temporal contexts,” in: *Sixteenth Annual Conference of the International Speech Communication Association* (2015)
- [8] D. Povey, V. Peddinti, D. Galvez, P. Ghahramani, V. Manohar, X. Na, Y. Wang, S. Khudanpur, “Purely sequence trained neural networks for asr based on lattice free mmi,”

- (author’s manuscript). Tech. rep., The Johns Hopkins University Baltimore United States (2016), doi: 10.21437/Interspeech.2016-595
- [9] “azwiki dump,” 2023. [Online]. Available: <https://dumps.wikimedia.org/azwiki/20230320/>
- [10] “The Azerbaijan State News Agency,” 2022. [Online]. Available: <https://azertag.az/>
- [11] A. Valizada, “Subword Speech Recognition for Agglutinative Languages,” 2021 IEEE 15th International Conference on Application of Information and Communication Technologies(AICT), Baku, Azerbaijan, 2021, pp. 1-6, doi: 10.1109/AICT52784.2021.9620466.

Texnoloji proseslərin təsiri ilə dildə transformasiyalar: sosiopsixoloji faktorlar, folklorik effektlər və linqvistik təzahürlər

Hikmət Quliyev
AMEA Folklor İnstitutu,
İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
quliyevh@mail.ru

Xülasə — Əsas funksiyası kommunikasiya olan dil, eyni zamanda mədəniyyətin ifadəsi üçün də mühüm vasitədir. Sosial mühitin formalaşmasında aktiv iştirak edən dil, bütövlükdə kommunikasiya vasitələri, texnoloji innovasiyaların təsiri ilə dəyişməkdədir. Kommunikativ dinamikanın spesifikliyi mədəni yaradıcılığın bütün səviyyələrində özünü göstərir. Bəşəriyyətin keçdiyi inkişaf yolu, epoxal dəyişikliklərə səbəb olan informasiya inqilablarının kökündə məhz texnoloji innovasiyalar dayanır. Bunu nəzərə alaraq, məqalədə yazının kəşfi, kitab çapı texnologiyasının ortaya çıxması, nəhayət, İnternetin – interaktiv ünsiyyət platformalarının ortaya çıxması kontekstində dildə meydana gələn dəyişikliklərin genezisində duran sosiopsixoloji faktorlar araşdırılmış, folklorik effektlər və linqvistik təzahürlər müəyyən olunmuşdur.

Açar sözlər — *texnoloji innovasiya; internet folkloru, linqvistik təzahürlər.*

I. GİRİŞ

Elmi ədəbiyyatlarda dilin kifayət qədər tərifi vardır. Lakin bu təriflər içərisində ən müştərək olanı dilin ilk növbədə

kommunikasiya vasitəsi kimi izahıdır. Dil, eyni zamanda mədəniyyət hadisəsidir. O mədəniyyətin, mədəni inkişafın ən mühüm göstəricilərindən biridir. Dilin vəziyyəti onu yaradan solumun intellektual-mədəni səviyyəsinin göstəcisidir. Orta əsrlərdə yazıya alınmış ən müxtəlif sahələr üzrə əlyazmalar – elm, təhsil, din, tibb, fəlsəfə və s. öz dövrünün həm dil xəzinəsi, həm də dil xəritəsidir. Bütövlükdə, dil – etnosun mədəni coğrafiyasının xəritəsidir. Bəşəriyyətin min illər ərzində qazandığı təcrübənin – mədəni yaradıcılığın istər diaxron, istərsə də sinxron planda ifadəsi, əsasən dil vasitəsilə ifadə olunmuşdur. Bu baxımdan dil sivilizasiyanın inkişafı, epoxal dəyişmələrə təkan verən elmi-texnoloji tərəqqi ilə sıx bağlıdır. Hər bir dövrün və yaxud mərhələnin mədəni tipologiyasının müəyyənlik şərtlərindən biri də kommunikasiya – ünsiyyət vasitələri və imkanları ilə şərtlənir desək, yanılmazdır. Beləliklə, bir tərəfdən dil kollektiv sosial mühitin yaradıcısı kimi çıxış edirsə, digər tərəfdən o, cəmiyyətin, mədəniyyətin, ideologiyanın vahid sistemdə təşkil olunmasına xidmət edir.

Əlbəttə ki, ilkin mərhələdə texnoloji innovasiya və dəyişiklik kimi diqqəti cəlb edən faktlar zamanla formal dəyişikliklərdən məzmun, mənə dəyişikliyinə doğru istiqamətlənməyə başlayır. Klassik ifadə ilə desək, kəmiyyət dəyişikliyi keyfiyyətə transformasiya olunur. Yəni informasiya almaq, xəbərləşmək, əlaqə yaratmaq, mobillik əldə etmək, əmək alətləri və prosesini təkmilləşdirmək və s. kimi zəruri ehtiyaclardan meydana çıxan texnoloji tərəqqi insana aid olan hər şeyi, ən sonda insanın özünü belə dəyişdirməyə doğru gedir. Homosapiensdən infosapiensə keçid nə qədər uzun yol hesabına başa gəlsə də, bu reallıqdır. Bu gün texnoloji proses, bütövlükdə sosial-mədəni sferanı, o cümlədən insan və onunla bağlı olan hər şeyi əhatə edir. Beləliklə, texnoloji innovasiya və dəyişmələr insanın *“homosapiens”dən “infosapiens” və “texnosapiens”ə* [1:176; 2: 84], *“human”dən “post human”ə* [2: 81-95], *“inforg”dan* [3: 43; 4: 128], *“cyborg”a* [2: 81-95] transformasiya olunmasına, mədəni və kommunikativ dinamika kontekstində yaradıcılığın məzmunu və mahiyyətinin dəyişməsinə gətirib çıxarmışdır. Bu prosesin məzmunu və mahiyyət hissəsi, məhz mədəniyyət, dil, folklor və s. ilə bağlıdır.

II. TEXNOLOJİ TƏRƏQQİ (İNFORMASIYA İNQİLABLARI) KONTEKSTİNDƏ KOMMUNİKASIYA: DİL, YAZI, MƏDƏNİYYƏT

Texnoloji tərəqqi bütün sferalarda dəyişikliklər yaradır. Bu dəyişikliklər bəzən epoxal proseslərə təkan verir. Məhz informasiya inqilabları bəşəriyyətin

inkışafında epoxal dəyişikliklərə səbəb olmuşdur. Xüsusən, informasiyanın əldə olunması, yadda saxlanması, emalı və ötürülməsi informasiya-kommunikasiya texnologiyalarının imkanlarına görə fərqli olmuşdur. Məsələn, yazının kəşfi ilə “... informasiya daşıyıcısı kimi təkci insanların yaddaşı deyil, həm də maddi daşıyıcılardan daş, heyvan dərisi, ağac və s. vasitələrdən istifadə edilməyə başlandı. Bəşər tarixində baş verən bu mühüm hadisə informasiyaların qeydiyyatı, uzun müddət yadda saxlanması, məkan və zamana görə yayılması üçün əvvəlki inqilaba nəzərən bir qədər geniş imkanlar yaratdı” [5: 9]. Beləliklə, yazının kəşfi aşağıdakılara şərait yaratdı:

- İnformasiyanın yadda saxlanması, tarixə, irsə çevrilməsi imkanı ortaya çıxdı;
- İnformasiyanın məsafə baxımından ötürülməsində məhdudiyət aradan qalxdı;
- İnformasiyanın çoxaldılması, uzunun köçürülməsi, daha da genişlənməsinə şərait yarandı;
- Nəhayət, yazı yazılı üslubu, yazılı mədəniyyəti yaratdı.

Yazının kəşfindən sonrakı mərhələdə isə kitab çapının meydana gəlməsi əldə olunmuş informasiya və biliyin daha geniş şəkildə yayılmasına, daha uzun müddətə qorunub saxlanılmasına təsir etdi. Təbii olaraq, bu proses bəşəriyyətin inkışafında, ideya və fəlsəfi düşüncələrin geniş kütlələrə çatmasında ciddi rol oynadı. Şübhəsiz ki, bütün bunlar öz növbəsində folklorun,

kollektiv mədəni yaddaşın qorunması, irs-
varis əlaqələrinin güclənməsi ilə nəticələndi.

Kitab çapından sonra elektrikin kəşfi, “elektron epoxa” adlandırılan dövrün meydana çıxması ilə xarakterizə olunur. Sözün texnologiləşməsi konsepsiyasının müəllifi olan Uolter J. Ong bu mərhələni “ikinci şifahi mədəniyyət epoxası” adlandıraraq yazır ki, “Telefon, radio və televiziya və s. tipli qurğular “ikinci şifahi mədəniyyət” epoxasını yaratdı” [6: 133].

Yazının, kitab çapının, elektrikin meydana gəlməsindən sonra bəşəriyyətin həyatında ən mühüm inqilablardan biri məhz İnternetin kəşfi olmuşdur. İnternet özünəqədərki inqilabi dəyişikliklərin heç biri ilə müqayisə oluna bilməyəcək proseslərə təkan verdi:

- Əlahiddə kommunikasiya imkanı yaratdı;
- Xalqlar, mədəniyyətlər arasında sərhədlər aradan qalxdı;
- Real sosial gerçəkliyin virtual əkizi ortaya çıxdı;
- Dil, ədəbiyyat, mədəniyyət, folklor və s. yaradıcılığı üçün yeni sosial mühit formalaşdı;
- Rəqəmsal texnoloji platforma, virtual ünsiyyət, elektron mədəniyyət anlayışları yarandı və s.

İnternetin meydana gəlməsindən sonra veb 2.0 – ikinci nəsil İnternet xidmətlərini özündə ehtiva etdirən texnologiyanın ortaya çıxması, onu məlumat, bilik arxivi funksiyasından bilik emalı platformasına transformasiya etdirdi. Burada əsas yenilik insanların iştirakçı sosial şəbəkələr hesabına

informasiya müzakirəsi və mübadiləsi prosesində aktivləşməsi, ortaq informasiya mühitinin formalaşmasıdır. Belə bir şərait, sosial medianı sosial-mədəni mühitə transformasiya etmişdir. Feysbuk, İntaqram, Yutub və s. kimi sosial şəbəkələr məhz iştirakçı sosial şəbəkələrdir. Əslində, bu mühit rəqəmsal folklor performansını üçün lazım olan bütün imkanları təqdim edir. Məhz bu imkanlar azərbaycandilli virtual sosial mühitin formalaşmasına, həmin dil seqmenti daxilində folklor resurslarının ortaya çıxmasına şərait yaratmışdır. Beləliklə, şifahi kommunikativ mühitdə meydana gələn, daha sonra yazıya transformasiya olunan, elektron epoxa daxilində audio informasiyanın bolluğu zəminində “ikinci şifahlıq” dövrünü yaşayan mədəniyyət İnternetin kəşfi ilə virtual platformaya transformasiya olundu. Heç şübhəsiz ki, hər bir kommunikasiya üsulu və kanalı özünəməxsus mədəniyyət tipinin ortaya çıxmasını şərtləndirdi. Hər bir mədəniyyət tipi isə ifadə olunduğu “dil”in özəlliklərini özündə əks etdirməkdədir.

III. TEXNOLOJİ İNNOVASİYALAR VƏ DİLDƏ TRANSFORMASIYALAR: SOSİOPSİXOLOJİ FAKTORLAR

Veb 2.0 texnologiyası: “prosumer” anlayışı

Veb 2.0 texnologiyasının hesabına interaktiv sosial media platformasının meydana çıxması İnternet istifadəçisini passiv dinləyicidən aktiv iştirakçıya çevirdi. Amerika folklorşünası S.J.Bronner A.Tofflerin “prosumer” (“producer” – istehsalçı və “consumer” – istehlakçı sözlərinin birləşməsindən əmələ gəlmiş)

anlayışına [7: 267] istinad etməklə belə bir sual irəli sürürdü: “istehsal edən və istehlak edən ortaq interaktiv mühitdə” birləşərsə və “istehsal edən-istehlak edən şəklində” ortaya çıxarsa, “nə baş verər?” [8: 24]. Beləliklə, virtual sosial şəbəkələrdə hər bir istifadəçinin həm informasiya qəbul edən, həm də istehsal edən statusunda olması, folklorşünaslıq termini ilə desək, İnternet istifadəçisinin həm söyləyici, həm də dinləyici kimi aktiv iştirakı virtual sosial mühiti ənənəvi, üz-üzə ünsiyyət mühitindən fərqləndirdi. Artıq informasiya, bilik, söz yaradıcılığında təkcə söyləyənlər, ötürənlər deyil, həm də dinləyənlər, qəbul edənlər aktiv duruma keçdilər.

Virtual sosial mühit, virtual sosial şərtlər və virtual insan

Texnoloji proseslər dərinləşdikcə formal dəyişikliklər məzmun dəyişikliyinə transformasiya olunur;

Virtual sosial mühit yeni sosial şərtlər və tələblər əsasında formalaşır;

Virtual insan/şəxsiyyət virtual mühitdə yaşam təcrübəsi qazanır və virtual ənənələr ortaya çıxır;

Virtual sosial mühitdə yeni davranış, yeni ənənə, yeni gerçəklik yaranır və bunlar dilə transformasiya olunur;

Beləliklə, texnoloji innovasiya və proseslər sosial mühitə təsir etməklə ilkin halda sosiotexnoloji platformanı formalaşdırmış, bunun davamı olaraq sosio-mədəni dinamika kontekstində yaradıcılığın məzmunu və mahiyyəti dəyişmişdir. Bütün bu faktorlar dil vasitəsilə ifadə olunduğundan, daha doğrusu, kommunikasiya zəminində baş verdiyindən

ənənəvi sosial mühitin “dili” ilə uzlaşmayan, yalnız “danışmağa”, “söhbət etməyə” deyil, həm də “göstərməyə”, “nümayiş etdirməyə” əsaslanan ünsiyyət zərurəti və imkanı ortaya çıxır. Bu isə, virtual ünsiyyət/elektron kommunikasiya/hibrid əlaqə və s. tipli dil işarəsi ilə şərtlənir.

IV. KOMMUNİKATİV DİNAMİKA KONTEKSTİNDƏ FOLKLOR VƏ DİL: METAFORİK-SİMVOLİK MƏNA

Yuxarıda qeyd etdiyimiz kimi, dil eyni zamanda simvol və işarələr sistemidir. İnsanın özünü və gerçəkliyi dəyərləndirməsi metaforik-simvolik mənalar çərçivəsində baş verir. Mədəniyyət, eləcə də folklor sosial gerçəkliyi metaforikləşdirməklə, simvol və işarələrə çevirməklə təcəssüm etdirir.

Folklor metaforik-simvolik sistem kimi folklorik kommunikasiya prosesində və kollektiv düşüncə modelində təcəssüm olunur;

Virtual folklorik faktlar məhz dilin metaforik-simvolik mənaları çərçivəsində yaradılır, qavranılır və dəyişdirilərək ötürülür;

İnternet folklorunda istifadəçi ilə folklorik fakt vahid semiotik proses tərkibində meydana çıxır: İŞARƏ – İŞARƏ EDƏN – İŞARƏ OLUNAN

İnternet “mem”ləri: metaforik-simvolik mənə (landırma)

İstifadəçidən istifadəçiyə “köçürmə” (“copy”) üsulu və təqlid vasitəsi ilə ötürülən, müəyyən bir fikir və yaxud davranışı əsas alaraq təcəssüm etdirən

mədəni vahidlər İnternet-memləri adlanır. İlk dəfə Riçard Doukins tərəfindən istifadə olunan bu anlayışın [9] hərfi mənası yunan dilindəki “mimeme” sözündən olub, “təqlid edilən bir şey” deməkdir. “... İnternet mühitində, sosial şəbəkələrdə insanlardan insanlara yayılan ideya, davranış üslubudur” [10: 98].

İnternet memləri folklorik yaradıcılıq nümunəsi kimi həm də metadil faktlarıdır. Memlərdə dilin universal imkanlarından – simvollaşdırmadan istifadə olunmaqla mənə yaradılır (Şəkil 1).



Şəkil 1. Memlərdə simvolların istifadəsi.

Üç müxtəlif əşyanın təsvirini birləşdirən fotoda, ilk iki əşya həqiqi tələ, üçüncü əşya – üzük isə nişanlanmaq, evlənmək mənasını ifadə edən, məcazi anlamda “tələ”dir. Buradakı folklorik mənə simvolik dillə - metadil vasitəsilə ifadə olunmuşdur.

V. SOSIAL MEDIANIN DİLİ: RƏQƏMSAL-TEXNOLOJİ

İnternetin əsas, aparıcı ünsiyyət vasitəsinə çevrilməsində sosial şəbəkələrin, bütövlükdə isə sosial medianın mühüm təsiri vardır. Əslində hazırkı çağda Azərbaycan dilində gedən dəyişikliklər real sosial mühitin davamı kimi ortaya çıxan virtual sosial mühitdə özünü göstərir. Lakin üz-üzə, səsli mükəllimədən fərqli olaraq, virtual kommunikativ mühitdə meydana

gələn ünsiyyət sosiotexnoloji şərtlərə uyğun olur. Odur ki, sosial medianın dilini müəyyənləşdirən faktorlardan biri rəqəmsal-texnoloji strukturdan irəli gələn tələblərdir. Buna misal olaraq, aşağıdakı bəzi məhdudiyətləri göstərə bilərik:

Twitter-də yazılı mətn, *YouTube*-də, əsasən video, *Instagram*-da foto və video resurslar üstünlük təşkil etdiyi halda, *Facebook* hər tip informasiya üçün uygundur;

Twitter-də mətn paylaşımaları 140 işarə ilə məhdudlaşdırılır. Bu isə fikrin ifadəsi üçün maksimal dərəcədə konkretliyi zəruri edir (şüar-mətn);

“TikTok”dakı qısa həcmli, video formatlı informasiyalar cəlbediciliklə bərabər, diqqət yayğınlığı və məzumsuz informasiya bolluğu yaradır.

VI. SOSIAL MEDIANIN DİLİ VƏ SOSİOPSİXOLOJİ DURUM: ANONİMLİK, SƏRBƏSTLİK, “ÇOXSƏSLİLİK”

Sosial medianın, virtual sosial şəbəkələrin istifadəçilər üçün yaratdığı ən əsas imkanlardan biri anonimlikdir. Hər bir istifadəçi ləqəbdən, saxta addan istifadə etməklə sosial şəbəkələrə üzv ola bilər.

Sosial medianın anonimlik imkanı:

- gizliliyi təmin edir;
- norma və qaydalara münasibətdə sərbəstlik yaradır;
- psixoloji baryerləri aradan qaldırır;
- realdan fərqli virtual şəxsiyyəti formalaşdırır;

Psixoloji olaraq bu durum bir sıra effektlər doğurur:

- antinorma;
- antistereotip;
- antiənənə;
- fərdilik;
- qeyri-rəsmilik;

Sosial mediada formalaşan mühit *çoxsəslili* mühit kimi xarakterizə olunur [11: 174]. Çoxsəslilik sosial medianın rəqəmsal-texnoloji strukturu, kommunikativ təbiəti, dinamik məzmunu ilə şərtlənməklə fərdlər çoxluğundan ibarət mühitin yaratdığı karnavalistik effektdir. Bu mühit isə dialoji, parodik və ironik xarakterə malikdir. O, antiierarxik təbiəti ilə avtoritar strukturu aradan qaldırır.

Beləliklə, sosial şəbəkələrin həm rəqəmsal-texnoloji bazisindən, həm buradakı sosial münasibətlərin ierarxik strukturundan, həm də bunların çulğalaşmasından ortaya çıxan mənə və məzmunundan – vernakular spesifiklikdən asılı olaraq sosial medianın dili müəyyənləşir. Bu dil isə Azərbaycan ədəbi dilinin rəsmi qayda və qanunlarına, norma və prinsiplərinə münasibətdə sərbəstdir. Daha doğrusu, sosial şəbəkələrdəki dil ədəbi dilin norma və prinsipləri baxımından imla xətası, orfoqrafik səhv, fonetik, leksik, qrammatik normaların pozulması ilə özünü göstərir.

VII. SOSİOTEXNOLOJİ GERÇƏKLİK: RƏQƏMSAL PLATFORMA VƏ ƏDƏBİ DİL

Əlbəttə ki, texnoloji tərəqqi informasiyanın əldə olunması, saxlanması, emalı və s. ilə bağlı kifayət qədər geniş

imkanlar yaratsa da, onun bəzi neqativ tərəfləri də vardır. Bütövlükdə, texnoloji tərəqqinin həm pozitiv, həm də neqativ tərəfləri vardır.

Pozitiv aspektlər:

- İnformasiyanı rahat şəkildə əldə etmək, yadda saxlamaq, ötürmək, dəyişdirmək və s.;
- Ünsiyyət və sosiallaşma ehtiyacının təmin edilməsi;
- Sürətli kommunikasiya imkanı və s.

Neqativ aspektlər:

- İnformasiyanın ifrat bolluğu;
- Konsentrasiya problemi;
- Kreativ, yaradıcı təfəkkürdən mexaniki, texniki vərdişlərə, hazır resurslara meyli;
- Dil qaydalarına riayət olunmaması;
- İmla səhvlərinin sürətlə populyarlaşması və s.

VIII. İNTERNETİN “DİLİ”: LİNGVİSTİK TƏZAHÜRLƏR

İnternetin dil tipologiyasını sosial media və şəbəkələrdəki kommunikasiya faktları təşkil edir. Burada diqqəti cəlb edən bir sıra məqamlar vardır:

- Xarici dilə məxsus sözlərdən istifadə etmək;
- Tələffüz olunduğu, deyildiyi kimi yazmaq;
- Dialekt faktlarından, məhəlli sözlərdən istifadə etmək;
- Qısaltma və abreviaturlardan istifadə etmək: OMG; LOL və s.
- Emojilərdən istifadə: 😊 😞
- Yeni sözlərin dilə daxil olması: “feyk”,

“çatlaşmaq”, “trollamaq”, “həştəq”

- Hərflər xəttinə yol verilməsi: ə-e, ö-o, ü-u, ş-sh, ğ-g
- Durğu işarələrindən səhv istifadə olunması və yaxud ümumiyyətlə istifadə olunmaması.

IX. İNTERNETİN DİLİ: ŞİFAHİ, YAZILI VƏ YAXUD HİBRİD

Bütün hallarda virtual ünsiyyət hər hansı bir dilin əlifbası əsasında baş verir. Virtual ünsiyyət klaviaturada hərflərlə qurulur. Lakin ünsiyyətin çevikliyi, danışıldığı kimi yazılması, emosional olaraq “danışmaq”, “söhbət etmək” xarakteri, durğu işarələrinə riayət olunmaması, emotikonlardan istifadə və s. bu ünsiyyəti şifahi edir. Beləliklə, virtual ünsiyyət formal olaraq yazılı (hərflərlə, işarələrlə) olsa da, psixoloji prototipində şifahilik emosiyası dayanır.

İnternetin dilinin şifahiliyi özünü bir sıra əlamətlərdə göstərir:

- İnternetin – sosial şəbəkələrin işlək dili gündəlik məişət üslubuna uyğun gəlir;
- Çevik kommunikasiya rahat, komfortlu ünsiyyəti – dil qaydalarına riayət olunmayan qeyri-rəsmi formatı zəruri edir;
- Virtual ünsiyyətdə dil qaydalarının, orfoqrafiya prinsiplərinin, üslubi normaların gözlənilməsi “soyuqluq”, “rəsmiyyət” yaratdığı üçün emosional olaraq “qəbul edilmir”;
- Bunun əksinə olaraq, dialekt faktlarından istifadə “səmimi mühit” yaradır;

- Qeyri-rəsmi ünsiyyət, gündəlik məişət üslubu söyüş, qeyri-etik ifadə və sairəni aktuallaşdırır.

Beləliklə, virtual sosial mühitdə ədəbi dil qaydalarının pozulması təkcə savadlılıq və yaxud savadsızlıqla deyil, onun genezisində dayanan sosiopsixoloji faktorlarla da bağlıdır.

NƏTİCƏ

Texnoloji innovasiya və dəyişikliklər bütün sferalarda olduğu kimi, sosial-mədəni mühitdə, onun təzahürü olan mədəni yaradıcılıqda, eləcə də dil faktlarında özünü göstərir. Hazırkı mərhələdə azərbaycandilli virtual sosial şəbəkələrdə meydana çıxan dil yaradıcılığı faktları Azərbaycan ədəbi dilinin qanun və qaydaları, norma və prinsipləri ilə müqayisədə uyğunsuzluq təşkil edir. Təbii ki, bunun özünəməxsus səbəbləri vardır.

Virtual sosial şəbəkələrdə kommunikasiya yazı vasitəsilə həyata keçirilsə də, bu dilin prototipində şifahilik emosiyası dayanır. Bununla əlaqədar olaraq, sosial şəbəkələrin dili Azərbaycan ədəbi dilinin orfoqrafiya qaydaları ilə deyil, orfoepiya prinsipləri ilə müəyyənləşir. Lakin bu prinsipdə də istisnalar mövcuddur. Belə ki, bəzən ədəbi dilin orfoepiya prinsipləri dialekt, məhəlli tələffüz ənənəsi ilə əvəz olunur. İnternet hər bir sosial qrupdan olan (savadlı, savadsız və s.) istifadəçilər üçün açıq olduğundan sosial şəbəkələrdəki dilin vəziyyəti də buna müvafiq şəkildə öz təsirini göstərir. Başqa sözlə desək, istifadəçilərin ədəbi dil qaydalarına müvafiq oxuma-yazma vərdişlərinin səviyyəsi sosial şəbəkələrdəki

dilin vəziyyəti ilə düz mütənasibdir. Lakin hesab edirik ki, virtual sosial şəbəkələrdə dil qaydalarına riayət olunmamasına səbəb, təkcə oxuma-yazma vərdişlərinin mənimsənilmə səviyyəsi ilə bağlı deyil, bu proses, həm də İnternet mühitinə xas olan sosiopsixoloji durumdan qaynaqlanır.

Araşdırma nəticəsində müəyyən olundu ki, İnternetin anonimlik, gizlilik imkanı psixoloji baryerlərin aradan qalxmasına, sərbəstliyə, qayda və normalara qarşı laqeydliyə əsas verir ki, bu da bütövlükdə həmin mühitin sosiopsixoloji durumunu müəyyənləşdirir. Bu gerçəklik ilk növbədə virtual sosial mühitin xarakterinin əks olunduğu dildə də öz təsirini göstərir.

Eyni zamanda sosial şəbəkələrin təsir imkanlarının genişliyi ədəbi dil baxımından norma pozuntusu, imla xətası olan faktların sürətlə populyarlaşmasına, yanlış şəkildə mənimsənilməsinə şərait yaradır.

Bundan başqa, sosial mediadakı ünsiyyət rəqəmsal-texnoloji platforma üzərində meydana çıxdığından buradakı dil bir sıra məhdudiyət və tələblərə uyğunlaşmalı olur. Bu proses ilkin halda formal dəyişikliklərə (mətnin qısaldılması, konkretləşdirmə, işarələrdən istifadə və s.), daha sonra üslubi spesifikaya təsir etməklə məzmun-mahiyyət planına da özünü göstərir.

İnternet mühitində, xüsusilə də virtual sosial şəbəkələrdəki ünsiyyət folklorik kommunikasiya rejimində meydana çıxdığından burada gündəlik danışmaq üslubuna məxsus olan şifahilik, canlılıq, emosional-ekspressivlik, məhəllilik və performativlik özünü açıq-aydın şəkildə göstərməkdədir.

İSTİNADLAR

- [1] Kızıldağ, H. “Arkaik devirlerden post-hümanizmə ölümsüzlük arayışındaki insan: ufuktaki transhümanizm ve kültür”. Sosyal Bilimle Işığında Doğa, İnsan, Toplum ve Kültür, s. 81-95. 2021.
- [2] Floridi, L. The Fourth Revolution: How the Infosphere is Reshaping Human Reality. Oxford: Oxford University Press, 272 p., 2014.
- [3] Gülüm, E. “Dijital iletişim teknolojileri aracılı bir folklorik deneyim alanı olarak sanal ortam”. Milli folklor, 119, s. 127-139, 2018.
- [4] Somerville, M. The Ethical Imagination Journeys of Human Spirit. Toronto: House of Anansi Press, 270 p. 2006).
- [5] Əliquliyev R.M., Valehov C.F., Mahmudov R.Ş. İnternet-jurnalistikanın formalaşmasının bəzi aspektləri. Ekspres-informasiya. İnformasiya cəmiyyəti seriyası. Bakı, “İnformasiya Texnologiyaları” nəşriyyatı, 32 səh., 2008
- [6] Ong, Walter J. Orality and Literacy The Technologizing of the Word. Routledge: Taylor & Francis e-Library, 203 p. 2005.
- [7] Toffler A. The Third Wave: the classic study of tomorrow. New York: Bantam, 1980
- [8] Bronner S J. “Digitizing and Virtualizing Folklore” İn Folklore and the Internet: Vernacular Expression in a Digital World. ed. Trevor J. Blank, Logan: Utah State University Press, p. 21-66, 2009.
- [9] Dawkins, R. The Selfish Gene / R.Dawkins. – Oxford; New York: Oxford University Press, 352 p., 1989.
- [10] Cəfərov Y. Sosial medianın dil-üslub xüsusiyyətləri. Journal of Baku Engineering University. Philology and Pedagogy, volume 1, number 1, s. 90-105, 2017.
- [11] Erdoğan, A. Teyfur ve Yumruksuz Ö. Sosyal medyanın doğası // Uluslararası Halkbilimi Araştırmaları Dergisi, sayı 6, s. 173-197, 2021.

Issues of linguistic demography formation on the national e-government platform

Gulnara Nabibayova

Institute of Information Technology, Baku, Azerbaijan

gulnarara@mail.ru

Abstract— The emergence of the field of electronic demography has become a prerequisite for developing an electronic demographic system on the e-government platform. Demography is a multidisciplinary field of science, and along with other areas, it also includes linguistic demography. In this regard, in the electronic demographic system, along with other segments, there is a segment of linguistic demography. Each area in the Data Warehouse (DW) has its own Data Mart (DM). This research emphasizes that when forming a DM of linguistic demography, it is necessary to solve some problems. These problems are related to the presence of different approaches to the definitions of the concepts of “language” and “dialect”, as well as to the establishment of a language indicator in the country, for which three options are possible - as a native language, as the language of everyday communication or as the language of use.

Keywords — *demography; Linguistic Demography; OLAP; electronic demographic system; Data Warehouse; Data Mart; native language; language of daily communication; language of use.*

I. INTRODUCTION

The implementation of a demographic policy aimed at improving the demographic

situation in the country is one of the important tasks facing each state. Demographic policy includes the solution of important tasks such as prolonging the life expectancy of the population, increasing the birth rate, reducing morbidity and mortality, regulating migration, etc. Demographic research plays an important role in solving these problems. The results of these studies constitute the basis of demographic policy.

Today, as a result of the impact of information technology on all spheres of human life and activity, along with the areas as electronic science, electronic medicine, etc. a new field of demography - electronic demography appeared on the e-government platform in Azerbaijan.

Demography (Greek δῆμος - people, Greek γράφω - I write) is a field of science that studies the patterns of population growth, its dependence on socio-economic conditions, natural factors, migration, population density and its distribution. Obviously, an important feature of demography is that it is an interdisciplinary science [1]. Demography is closely related to economics, political science, ethnography, medicine, geography, etc. In this regard, it is divided into a number of

specialized areas, each of which studies certain demographic processes.

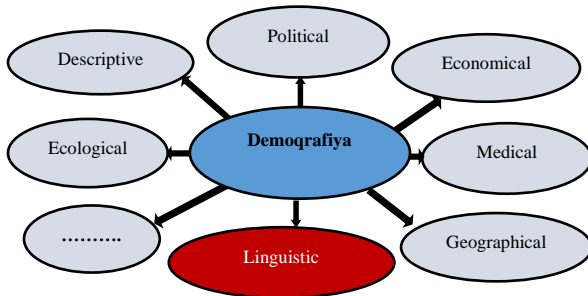


Fig. 1. Fields of demography

Fig. 1 shows the fields of demography, among which there is linguistic demography.

The emergence of electronic demography became a prerequisite for the need to develop an electronic demographic system [2]. One of the main unit in this system is “on-line analytical processing”, i.e. OLAP technology, which is a key element of the DW. OLAP technology is based on the representation of data in the form of OLAP cubes. They contain data used for analysis and decision making.

As noted above, demography is an interdisciplinary science, as a result of which the DW of the electronic demographic system is divided into several DM. Each DM corresponds to a distinct area of demographics. One of them deals with linguistic demography (Fig. 3).

II. FORMATION OF LINGUISTIC DEMOGRAPHY WITHIN THE ELECTRONIC DEMOGRAPHY SYSTEM

Linguistic demography is the statistical study of languages among all groups of populations. Fig. 2 illustrates the position of linguistic demography within the framework of national e-government.

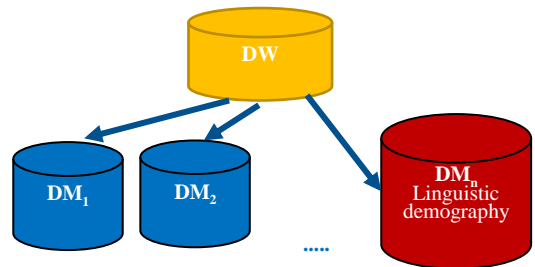


Fig. 3. DW structure

The data on the distribution and use of languages are included into the Linguistic Demographics DM. Specifically, these data include both the geography of the distribution of languages and the number of speakers of these languages.

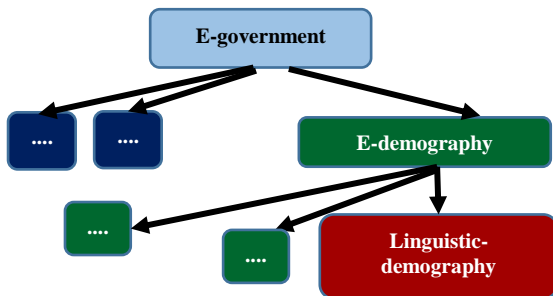


Fig. 2. The position of linguistic demography on the e-government platform in Azerbaijan

III. DIFFERENT APPROACHES TO THE DEFINITION OF THE CONCEPTS OF “LANGUAGE” AND “DIALECT”

There are currently over 7,000 languages in the world. Some languages have become the most widely used languages in the world. Fig. 4 displays the ranking of the 10 most spoken languages in the world [3]

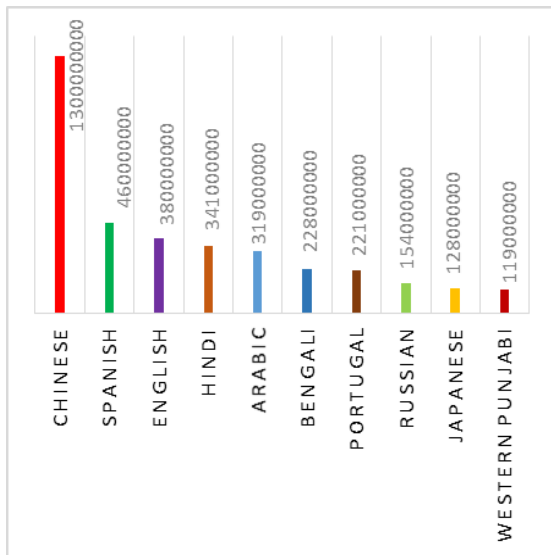


Fig.4. Top 10 most spoken languages in the world

It should be noted that the Azerbaijani language ranks 30th in the world in terms of the number of speakers. The area of distribution of the Azerbaijani language is Azerbaijan, Iran, Russia (Dagestan), Turkey, Georgia, and Iraq.

However, there are certain problems related to the estimation of the number of speakers of a particular language, and different estimates may vary significantly. The problems complicating the conduct of

research in this field primarily include: a) availability of different approaches to the definition of the concepts of “language” and “dialect”; b) availability of multilingualism among some individuals, which obscures the definition of the concept of “native language”; c) and the lack of sufficient population census data in many countries.

The availability of different approaches to the definitions of the concepts of “language” and “dialect” is directly related to the presence of different points of view on these concepts [4, 5]. If the concept of “language” is considered from a political point of view, then the language is the one that the people usually speak as a nation. For example, from 1850 to 1992 there was a Serbo-Croatian language, which included several dialects, including Serbian, Croatian and Bosnian. Nevertheless, following the breakup of Yugoslavia into several independent states in the mid-1990s, these dialects were recognized as separate languages. Another example is Spanish. The Spanish spoken in Spain significantly differs from its Latin American version. Nonetheless, despite the fact that it may well be considered a separate language, it is recognized as a dialect.

In order to carry out statistical studies in the field of linguistic demography, it is necessary to determine the current language indicator in the country, which can be set differently, namely as a native language, as a language of daily communication or as a language of use. Note that in order to make the results of these studies more reliable, demographers are recommended to ignore the language indicator of tourists, students

from other countries and workers on a temporary contract.

At the session of the International Statistical Congress of 1872, St. Petersburg, it was recommended to use “native language” as a language indicator. However, this recommendation was not followed in four censuses in multilingual Austria, where the “language of use” was adopted as a language indicator. In the Czech Republic, the recognition of “language of use” as a language indicator has been challenged. This was due to the fact that many citizens of the Czech Republic who were born in this country and speak Czech since childhood, but live in German-speaking cities in Austria, may consider German as their “native language”.

It should be noted that not only multilingualism, but also the presence of different approaches to the concept of “native language” complicate the definition of this concept.

The main approaches are as follows:

- Native language is a language acquired in childhood without teaching. A language that a person learns from childhood without special training, in a natural language environment.
- Native language is functionally the first language. The language in which they think automatically. In this language, a person easily expresses own opinion in oral and written form, and this is the most convenient form of expression for him/her.

- Native language is the language of ethnic identification. The language of a nation or ethnic group linking a person to previous generations.

Note that the native language may not coincide with the nationality of a person. Such cases may often be found in the context of global migration processes of the 20th and 21st centuries.

It is also noteworthy that recent studies in the field of linguistic demography have attracted more and more attention of demographers from the standpoint of sociolinguistics, which studies the relationship between the area of distribution of a language and the social conditions of its existence.

CONCLUSION

Due to the existing problems listed above, there are some difficulties when entering data on linguistic demography. For example: a) Should the dialect of the language be accepted as a new language? b) What does the term “native language” mean? and so forth.

All of the above problems facing world demographers are also important for Azerbaijani demographers.

In this regard, certain problems arise when entering data on linguistic demography into an electronic demographic system.

These problems may include:

- Whether the dialect of the language is accepted as another language or not

- Identification of native language of multilingual people
- Definition of the concept of “native language”.
- The need to conduct a census at least once every 5 years to determine the dynamics of the distribution of languages;
- What should be used as a language indicator - native language, everyday or spoken language.

These issues need to be addressed as soon as possible.

REFERENCES

- [1] <https://www.cambridge.org/core/books/essential-public-health/demography/CB1BF80C417CE4E5F980DF4D0A0D8D04>
- [2] Alguliyev R. M., Aliguliyev R. M., Yusifov F. F., Alakbarova I. Y. Formation of electronic demography as an effective tool for social research and monitoring of population data / Issues of state and municipal management. 2019. No. 4. pp. 61–86
- [3] <https://www.ethnologue.com/statistics/>
- [4] Soren Wichmann. How to Distinguish Languages and Dialects. Computational Linguistics, Volume 45, Issue 4 December 2019 <https://aclanthology.org/J19-4007/>
- [5] Koryakov Yu. The problem of “language or dialect” and an attempt of a lexical and statistical approach // Problems of Linguistics. 2017. No 6, pp. 79-101

Nitq texnologiyalarının eksperimental-fonetik aspektləri

Nəzakət Qaziyeva

AMEA Nəsimi adına Dilçilik İnstitutu, Bakı, Azərbaycan

n.qazi@inbox.ru

Xülasə — Məqalədə nitq texnologiyalarının inkişafında mühüm amillərdən olan eksperimental fonetik tədqiqatların əhəmiyyəti vurğulanır, nitq signalının akustik təhlilindən əldə olunmuş yeniliklərin bir çox müasir tədqiqat sahələri üçün vacib nəticələr verdiyi göstərilir. Təbii nitqə yaxın sintezin əldə olunması üçün nitq signalının fiziki-akustik parametrlərinin signalın mövqeyi, digər fonemlərlə əlaqəsi kimi amillər nəzərə alınmaqla əsaslandırılmasının verdiyi üstünlüklərdən bəhs olunur.

Açar sözlər — *nitq texnologiyaları; sintez; eksperimental fonetika; koartikulyasiya*

I. GİRİŞ

Müasir dövrdə informasiya texnologiyalarının nailiyyətlərindən bəhrələnməyə bir çox sahələr kimi, multidissiplinar istiqamət olan nitq texnologiyaları da sürətlə inkişaf etməkdədir. Dilin səs sistemi ilə əlaqəli biliklərin praktiki tətbiqi və insanın təbii nitq fəaliyyətini modelləşdirmə üsulları ilə bağlı olan nitq texnologiyalarının sürətli inkişafı müasir cəmiyyətin bu sahədəki bir sıra problemlərinin həllinə ehtiyacdən meydana çıxan tələblə əlaqədardır. Bu günə qədər dilçilik sahəsində təbii dillərin səs formasının təşkili və fəaliyyət prinsipləri haqqında xeyli məlumat toplanmışdır. Bu

məlumat nəzəri və eksperimental fonetik tədqiqatlar zamanı əldə edilmiş və əsasən təsviri xarakter daşımışdır. Müasir elmi baxışlar nitq texnologiyalarının fonetik məlumatlardan istifadə olunmaqla inkişafını aktuallaşdırır.

Nitq texnologiyalarına avtomatik nitqin tanınması və sintezi üçün sistemlərin inkişafı, nitq məlumatlarının korpusunun formalaşdırılması (akustik verilənlər bazası), nitq signalının avtomatik işlənməsi (əsas ton tezliyinin müəyyən edilməsi, formalı hərəkət trayektoriyalarının müəyyən edilməsi, avtomatik seqmentləşdirmə və s.), şəxsin səsə identifikasiyası və yoxlanılması, rabitə kanalları üzərindən ötürülməsi üçün nitq signalının sıxılması üsullarının işlənməsi və s. kimi məsələlər daxildir.

Nitq signalının rəqəmsal formada təqdim edilməsi onun təhlili və işlənməsi üçün geniş perspektivlər açır. Audio signalının səs analizi proqramları vasitəsilə təhlili zamanı səs signalının ossiloqramını və ya sonoqramını statik təsvir şəklində müşahidə etmək imkanı, signalı əvvəldən axıra və ya axırdan əvvələ doğru hərəkət etdirməyin mümkünlüyü, həmçinin həm bütün signalı, həm də onun hər hansı bir fraqmentini dəfələrlə dinləmək imkanı bir çox məsələlərin dəqiq təhlili və araşdırılmasına

şərait yaradır. Bu seçilmiş fraqmenti, nəinki dinləmək, həm də silmək, başqa yerə köçürmək və ya başqa səs signalına daxil etmək, həmçinin fiziki xüsusiyyətlərini dəyişdirmək mümkündür.

Eksperimental fonetikanın ənənəvi və aktual problemlərindən olan nitq axınının davamlılığı (qeyri-sabitliyi) ilə onun fonetik və fonoloji təsvirinin diskretliyi arasındakı ziddiyyət [1] bu məsələlərdə müəyyən problemlərin ortaya çıxmasına səbəb olsa da, nitq signalının akustik parametrlərinin eksperimental üsulla tədqiqinin müasir səviyyəsində bu tipli problemlərin həlli xüsusi çətinlik yaratmır. Bununla belə, müasir dövrdə bu sahə ilə bağlı yeni vəzifələr aktuallaşır. Bu problemləri həll etmək üçün artıq signalın spektral xüsusiyyətlərini dəqiq hesablamaq kifayət deyil, signalın qavrayış-perseptiv keyfiyyətini dəyişdirən amillər (əsas tezlik, formant mövqeyi və istiqaməti, fərdi tembr xüsusiyyətləri və s.) diqqət mərkəzində saxlanılmalıdır.

Nitq sintezi müəyyən informasiyanın çatdırılması tələb olunan istənilən sahədə tətbiq oluna bilər. Başlıca tətbiq sahələrindən isə informasiya-məlumat sistemləri, çağrı mərkəzləri, “Qaynar xətt” xidməti, sağlamlığı məhdud insanların, xüsusilə gözdən əlillərin cəmiyyətə inteqrasiyası məqsədilə “səsli” idarəetmə, nəqliyyat sektorunda – qatarların hərəkət qrafiki ilə bağlı, avtobus və metroda, eləcə də hava limanında müəyyən məlumatların verilməsi, eləcə də hərbi və aviakosmik sənaye qeyd oluna bilər.

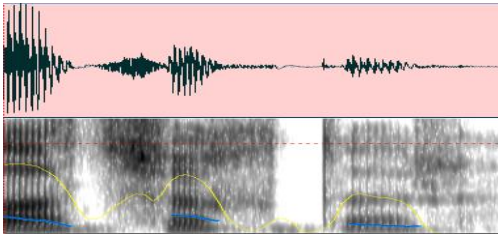
II. “KLONLAŞDIRMA” METODUNUN TƏTBİQİNDƏ NİTQİN FİZİKİ PARAMETRLƏRİNİN ROLU

Son zamanlar nitqin sintezi üçün verilənlər bazasında saxlanılan nitq signalının seqmentlərinin birləşdirilməsindən ibarət olan konkatenativ üsuldan geniş istifadə olunur. Nitq məlumat bazasının həcmi nə qədər böyükdürsə, onda konkret şəxsə aid səs, temporal və intonasiya dəyişkənliyi ilə bağlı material bir o qədər tam əks olunur. Nəticədə, sintez edilmiş nitq daha asan “tanınır” və ümumiyyətlə daha təbii səslənir. Bundan əlavə, nitq korpusu nə qədər böyükdürsə, onun elementlərinin əsas tonunun (ƏTT) müddətini və tezliyini dəyişdirmək lazım olduqda baş verən nitq seqmentlərinin təhrifi bir o qədər az olur. Başqa sözlə, nitq verilənlər bazası nə qədər böyükdürsə, onun tələb olunan kontekstdə, tələb olunan müddət və tezlik konturlu elementi ehtiva etmə ehtimalı bir o qədər yüksəkdir. Nitq bazasına daxil olan material universal deyil, həmişə fərdi olur və konkret natiqlərin tələffüz xüsusiyyətlərini əks etdirir. Əksər dillər üzrə istifadə olunan nitqin mətn üzrə eksperimental sintezatoru konkatenativ sintezatorlar üçün ənənəvi modullardan ibarətdir [2; 3].

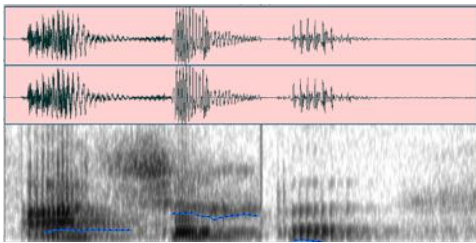
Adətən nitqin tanınması ilə bağlı tədqiqatlarda nitq signalının elementi kimi səs qəbul olunaraq fonemlərin bir-birinin ardınca – ardıcılıqla dərk olunduğu, tanındığı qeyd olunur. Nitqin qavranılması ilə bağlı tədqiqatlar, perseptiv araşdırmalar göstərir ki, fonemlər paralel şəkildə də qavranıla bilər və fonetik element kimi məhz heca qəbul olunmalıdır. Nitq sintezi

zamanı məhz hecanın xüsusi fonetik element olduğunu diqqət mərkəzində saxlamaq mühüm şərtidir. Bu zaman diqqət mərkəzində saxlanılmalı olan məsələlərdən biri də tənəffüs zamanı bir nəfəsvermədə tələffüz olunan hecaların sayı məsələsidir.

Nitq sintezi zamanı ritmik mənərə tonal vasitələrlə ifadə olunmalı, bunun üçün uyğun üsullar ilənib hazırlanmalıdır. Eksperiment zamanı prosodiya ilə bağlı problemləri aradan qaldırmaq məqsədi ilə səsini stasionar hissəsinin “klonlaşdırılması”ndan və pauzadan istifadə etdik. Vurgulu hecada saitin daha “qüvvətli” səslənməsinə nail olmaq üçün “klonlaşdırma”, eləcə də pauzadan istifadə sintezin keyfiyyətinə öz təsirini göstərir (Şəkil 1).



(açıqdır – təbii nitq)



(açıqdır - sintez)

Şəkil 1. Təbii nitqin sintezi

Qeyd edək ki, heca sərhədini müəyyən etmək məqsədilə samitin daxilində “klonlaşdırma” əməliyyatı aparılması da müsbət nəticə verir.

Sintezdə nailiyyət əldə etmək üçün nitqin fizioloji xüsusiyyətləri diqqət mərkəzində saxlanılmalıdır. Nəfəsalma, nəfəsvermə və onlar arasındakı pauzadan ibarət olan tənəffüs sikli 2,5-7 s. təşkil edir. Bir çox insanlarda nəfəsalmanın davamlılığı nəfəsvermədən daha azdır – bu mərhələ daha qısaadır. Pauzanın davamlılığı fərqlidir – hətta olmaya bilər.

Nəzərə almaq lazımdır ki, nitq əksər dünya dillərində nəfəsvermə zamanı baş verir. Nitq signalının seqmentlərə ayrıldığı parçada (bizim təcrübədə hecada seqmentləşmə samit və sait kimi yox, sait və sait olmayan) sait olmayan iki samit, samit və pauza, ya da pauza ola bilər. Prosodiyanın təmin olunması üçün sintezdə bu məqamlarda pauza xüsusi vasitə kimi əhəmiyyətli rola malikdir. Stasionar sahələrin davamlılığı (ölçüsü) keçid hissələrdən böyük olsaydı, həqiqətən də fonemlər ardıcıl dərk olunardı – eksperimental tədqiqat və təcrübə göstərir ki, adətən keçid hissələr daha “uzun”dur. Təbii nitq kəsiyində bəzən müəyyən hissədə müxtəlif fonemlərə məxsus xüsusiyyətlər “qarışıq” şəkildə müşahidə olunur – əvvəlki və sonrakı fonemin sərhədini dəqiq müəyyən etmək çətin olur. Bu hissələrdə samitə aid fiziki parametrlərin mövcudluğu onun samitə aid olunması fikrini doğursa da, bu hissə kəsildikdə saitin müəyyən olunması çətinləşir. Bu tipli eksperimentlərə əsasən,

sait və samitin ardıcıl yox, paralel şəkildə dərk olunması ehtimal olunur.

Saitlərin uzunluğu samitin növündən də asılıdır. Novlu samitdən əvvəl sait daha uzundur, nəinki kipləşən-partlayan samitdən. Nəzərə almaq lazımdır ki, sözdə hecalar artdıqca, nitq prosesi ilə bağlı olaraq fizioloji xüsusiyyətlərin təsiri ilə saitin ölçüsü qısalır. Kar və ya cingilti samitdən əvvəl olmasından asılı olaraq, saitin uzunluğu fərqli olur: kar samitdən əvvəl saitin uzunluğu cingiltidən əvvəl olanla müqayisədə uzunluğunun 1/2, 1/3-nə bərabərdir. Deməli, sintez zamanı samitin kar və ya cingilti olduğunu qabartmaq üçün ondan əvvəlki saitin uzunluğu əhəmiyyətli amil kimi istifadə oluna bilər. Sait səslərin uzunluğunun adi halda 150-350 ms. arası olduğunu nəzərə almaqla qeyd edə bilərik ki, sintez zamanı uzunluğu 50 ms. olan saitin 100 ms.yə çatdırılması onun daha yüksək səs kimi (təxminən 2-3 dB.) dərk olunmasına şərait yaradır. Maraqlıdır ki, 100-200 ms. arası dəyişmə bu effekti vermir.

Aparılan eksperimentlərin nəticəsinə görə, hecanın “davamlılığı” (uzunluğu - ölçüsü) saitin uzun tələffüzü hesabına təmin olunur – balans saithlərin ölçüsü vasitəsilə yaradılır. Bu eksperimentlərin nəticəsinə görə, vurğu da məhz sait səsin uzunluğu ilə təmin olunur. Nəzərə almaq lazımdır ki, hər dil üçün vurğunu müəyyən edən fiziki parametrlər eyni deyil. Bu kəmiyyət vurğusuna malik dillər üçün səciyyəvidir, yeri gəlmişkən, bu məsələ Azərbaycan dili üçün özünü doğrultmur. Belə ki, bir sıra

dillərdə başqalarından uzun heca vurğulu kimi dərk olunur. Bu vasitə ilə sintez zamanı cümlədə məntiqi vurğunun təmin olunmasına da nail olmaq olar. Qeyd edək ki, samitin uzunluğu sözün sərhədini bildirir və bu amil sintez zamanı təbii nitqə yaxın nəticə əldə etmək üçün xüsusi əhəmiyyət daşıyan amillərdən biri kimi istifadə oluna bilər.

Nitq mürəkkəb və çoxtərəfli prosesin nəticəsi olduğundan onun təhlili zamanı fizioloji, psixoloji, koqnitiv, akustik xüsusiyyətlərin nəzərə alınması və kompleks şəkildə təhlili vacibdir. Əks halda nitq aktının mexanizmi və nitq signalının parametrləri ilə bağlı dolğun təsəvvür əldə etmək mümkün deyil.

İ.M. Seçenovun da qeyd etdiyi kimi, biz səssiz düşünəndə belə, danışanda olduğu kimi, dilimizdə biotoklar güclənir. Bunu xüsusi qurğular vasitəsilə qeydə almaq mümkündür. Lakin bəzən müəyyən yazını xüsusi diqqətlə “ürəyimizdə” oxuduqda da bizdən asılı olmayaraq, nəinki dil, hətta dodaqlar belə tərpənir... “Daxili nitqin refleksləri funksional baxımdan dəyişilmiş adi nitq refleksləridir” [4]. Daxili nitqin mexanizmi sona qədər tədqiq olunmasa da, aparılan çoxsaylı eksperimentlərin nəticələri daxili nitqdə nitq analizatorlarının iştirakını sübut edir [5].

Alınan nəticələrə görə, çətin məsələləri düşünən zaman artikulyasiya hərəkətləri daha qabarıq şəkildə ifadə olunur. Daktil nitqi (lalların barmaq işarələri vasitəsilə həyata keçirdiyi nitq) mənimsəmiş sağlam insanlarda daxili nitq zamanı əvvəlcə dil,

daha sonra barmaq əzələlərində yüksək elektroaktivlik müşahidə olunur.

III. NİTQ SİNTEZİNDƏ NİTQİN PROSODİK VASİTƏLƏRİNDƏN İSTİFADƏNİN ÜSTÜNLÜYÜ

Nitq axınının fonetik (seqmental və suprasedgmental) təşkili haqqında məlumatlara kifayət qədər diqqət yetirilməməsi bəzən siqnalın təbiiliyinin pozulmasına səbəb olur. Bunu makrosintez nümunələrində təbii nitq siqnallarının kifayət qədər uzadılmış fraqmentlərində asanlıqla görmək olar. Makrosintez dəmir yolu vağzallarında elanlarda, danışan saatlarda və oyun proqramlarında istifadə olunur. Bu pozuntular sözlərin qovşağındakı samit birləşmələri, nitqin ritmik təşkilinin pozulması və ya qeyri-təbii intonasiyada ifadə olunur ki, bu da dil norması nöqtəyənəzərindən qəbul edilməz hesab olunur və sintezin keyfiyyətini müəyyən edən amillərdəndir.

Nitq siqnalının rəqəmsal şəkildə emalı əsas tonun tezliyinin və siqnal fraqmentlərinin müddətinin dəyişdirilməsi ilə sintezdə qarşıya çıxan müəyyən problemləri həll etməyə imkan verir. Səs mənbəyinin iştirakı olmadan həyata keçirilən nitq siqnalının fraqmentinin müddətini dəyişdirmək (məsələn, samitlərin parametrlərində dəyişiklik, fiziki fasilələr) kifayət qədər sadə üsullarla həll olunur. Kipləşən-partlayan samitlərdə pauzanın və ya küy hissəsinin müddətini dəyişdirmək – müddəti artırmaq üçün sıfır amplitudalı bir hissə əlavə etməklə və ya onu azaltmaq üçün küy və ya fasilənin bir hissəsini çıxarmaqla nəticə əldə etmək mümkündür. Afrikatların müddətini dəyişdirmək bir

qədər daha çətindir. Təcrübələr göstərir ki, yalnız pauza və ya yalnız səs-küyün müddətini dəyişdirmək orijinal fonetik kontekstdə səsin qavrayış keyfiyyətinin pozulmasına səbəb olur. Buna görə də, affrikatın müddətində dəyişiklik pauza və səs-küyün müddətinin mütənasib dəyişməsi ilə əldə edilə bilər. Təbii ki, kar samitlərin müddətini manipulyasiya edərkən, bu samitlərin partlayış sahəsinə və ətrafdakı səslərə bitişik ərazilərə müdaxilə etmək yolverilməzdir.

Kar samitlərə nisbətən sonorların və saitlərin müddətini dəyişdirmək daha çətindir. Belə ki, səs mənbəyinin iştirakı ilə yaranan səslərin müddəti üçün ölçü vahidi yalnız saniyələrlə (yaxud daha əlverişlisi millisaniyələrlə) ölçülən müddətlə deyil, həm də səs mənbəyinin dövrlərindəki müddəti ilə (bu cür səslərin daxili təşkili ilə) müəyyən edilən əsas tezlik hesablanır, dəyişə bilər. Bu baxımdan, belə səslərin müddətini dəyişdirmək yalnız əsas tezliyin dövrlərini manipulyasiya etməklə (lazımı sayda dövrləri silmək və ya köçürməklə) əldə edilə bilər.

Saiti əmələ gətirən dövrlərin seçilməsinə obyektiv olaraq əlavə məhdudiyətlər qoyulmaqla, onun müddətinin dəyişməsinə nail olmaq olar, lakin nəzərə almaq lazımdır ki, bu cür manipulyasiyalara ən dözümlü hissə saitin stasionar bölməsidir. Odur ki, manipulyasiya zamanı bitişik dövrlərdən – koartikulyasiya hissələrindən istifadə arzuolunmazdır, əməliyyata məruz qalan sahələr arasındakı məsafə ən azı bir dövr olmalıdır.

Saitlərdə və sonor səslərdə əsas tonun tezliyində xüsusi dəyişiklik baş verir ki, bu

da nitqin melodik tərtibatını müəyyənləşdirir. Bu baxımdan, konkret səsin müddətinin dəyişməsi nitqin melodik tərtibatına təsir etməməli və ya melodik tərtibatın korreksiyası (əsas ton tezliyinin modifikasiyası) ilə müşayiət olunmalıdır. Əks halda bu ona gətirib çıxarır ki, səslərin müddətinin dəyişməsi, bir qayda olaraq, nitqin və ya onun hissəsinin melodik tərtibatının dəyişdirilməsi ilə nəticələnir. Bununla belə, əsas tonun tezliyinin dəyişməsi həm də əsas tonun dövrlərinin müddətinin dəyişməsi deməkdir ki, bu da təbii olaraq siqnalın dəyişən fraqmentinin ümumi müddətinin dəyişməsinə gətirib çıxarır. Fonetik təcrübələrdə və yüksək keyfiyyətli avtomatik nitq sintezi sistemlərinin işlənilib hazırlanmasında əsas ton tezliyinin modifikasiyasından istifadənin əvəzsiz şərti onun qavrayış keyfiyyətinin və təbiiyyətinin mütləq şəkildə qorunub saxlanmasını təmin etməkdir. Problem ondadır ki, əsas tonun tezliyindəki təbii dəyişiklik, bir qayda olaraq, saitin linqvistik baxımdan tanınmasına cavabdeh olan formantların mövqeyinə təsir etmir, əksinə, bəzi harmonikalarda səsin fərdi tembri ilə bağlıdır və qalan harmonikaların yerini dəyişir. Odur ki, spektrin bütün digər komponentlərini eyni yerdə saxlamaqla yalnız əsas tezliyin dəyişdirilməsi qeyri-təbii siqnal təəssüratı yaradır.

Sintezin təbii nitqə daha yaxın olması üçün təbii nitqin komponentləri maksimal şəkildə sintezdə öz əksini tapmalı, həm seqment, həm də superseqment səviyyəyə aid elementlər korpusun təşkilində olduğu kimi, sintez prosesində də nəzərə

alınmalıdır. Adətən sintezin “süni” səslənməsi superseqment vasitələrin – prosodik xüsusiyyətlərin dəqiqliklə ifadə oluna bilməməsindən qaynaqlanır. Bir çox hallarda, hətta əhatəli korpusun mövcudluğu belə, sintezdə prosodiya məsələsini həll etmək üçün yetərli olmur.

İlk dəfə İ.Mətinqli qaydalar üzrə prosodik sintezatorun layihəsini təqdim etmiş [6], 2000-ci ildə isə Burkhardt [7] formant sintezindən istifadə edərək müxtəlif emosional kateqoriyalar üçün nitqin əsas akustik parametrlərini müəyyən etmişdir. Burkhardtın tədqiqatının məqsədi neytral şəkildə ifadə edilən nitqin akustik parametrlərini dəyişdirərək emosional qavrayışa təsir edən akustik xüsusiyyətləri müəyyən etmək idi. Onun təcrübələrinə görə, emosional nitqin formant sintezində əsas ton, onun orta qiyməti və dəyişmə diapazonu, nitqin sürəti, fonasiya, sait tərifinin dəqiqliyi kimi parametrlər həlledici əhəmiyyət kəsb edir.

Keyfiyyətli sintezin ilkin şərtlərindən biri bunun üçün qurulmuş korpusun mövcudluğudur. Müasir dövrdə geniş şəkildə istifadə olunan difonik sintez modeli səs bazasının formalaşması üçün müəyyən prinsipləri nəzərdə tutur [8]. Birinci mərhələdə, bütün mümkün birləşmələri sintez üçün istifadə olunan difonların sayını məhdudlaşdıran vahidlərin siyahısını müəyyən etmək, həm kəmiyyət, həm də keyfiyyət baxımından əsas allofonların siyahısının tərtib olunması, sait və samitlərin bəzi kombinator variantları, eləcə də saitlərin mövqe variantları nəzərə alınaraq korpus yaradılmalıdır. Bundan

sonra həm sözlün daxilində, həm də sözlərin qovşağında yaranan və nəzəri cəhətdən mümkün C + U, C + C, U + C, U + U birləşmələr nəzərə alınmaqla xüsusi difonik baza tərtib edilməlidir.

NƏTİCƏ

Nitq signalının işlənməsi zamanı fonemlərin koartikulyasiya, birləşmə və ritmik mövqe nəticəsində yaranan məcburi və potensial mümkün modifikasiyaları nəzərə alınmalıdır. Bu cür sistemləri öyrətmək üçün müəyyən bir dilin fonemlərinin bütün mümkün mövqe və kombinator variantlarını özündə birləşdirən ifadələr və ya ifadələr korpusunun olması əsas şərtidir. Bu prinsipləndən istifadə edən sistemlər müəyyən bir lüğətlə məhdudlaşmır, buna görə də əlavə təlim olmadan onlara istənilən sayda yeni söz əlavə edilə bilər. Konkatenativ üsulla sintezin uğurlu alınmasının ilkin şərti verilənlər bazasının böyüklüyü və əhatəliliyi olsa da, təbii nitqə yaxın sintezə nail olmaq üçün müasir eksperimental fonetik araşdırmalardan əldə olunan nəticələrin tətbiqi ilə “klonlaşdırma” üsulundan istifadə daha yaxşı nəticəyə nail olmaq üçün faydalı ola bilər.

İSTİNADLAR

- [1] Зиндер Л.Р. Общая фонетика. – м.,1979.
- [2] Krivnova O.F. Automatic synthesis of Russian speech // Proceedings of the XIV International Congress of Phonetic Sciences, Vol.1, pp. 507–510, San Francisco, 1999.

- [3] Лобанов Б.М., Карневская Е.Б., Левковская Т.В. Синтезатор речи по тексту как компьютерное средство «клонирования» персонального голоса // Тр. Международной конференции Диалог-2001 / М., 2001. С. 265-272.
- [4] Сеченов И. М. Рефлексы головного мозга. 1942.
- [5] Соколов А. Н. Внутренняя речь и мышление. М., Искусство, 1968.
- [6] Klatt D. H. Review of text-to-speech conversion for English // The Journal of the Acoustical Society of America. 1987. Vol. 82. No 3. P. 737–793.
- [7] Burkhardt F., Sendlmeier W. F. Verification of acoustical correlates of emotional speech using formant-synthesis // ISCA Tutorial and Research Workshop (ITRW) on speech and emotion. 2000.
- [8] Бондарко Л.В., Вербицкая Л.А., Гордина М.В. Основы общей фонетики. Л., 1983.

Milli onomastika sisteminin formalaşması aspektləri haqqında

Yedgar Cəfərov

İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
yedgar.jafarov@gmail.com

Xülasə — Məqalədə Milli Onomastika Sisteminin yaradılmasının ayrı-ayrı aspektlərinə diqqət yönəlib. Domen adları, adların modelləşdirilməsi, adların, o cümlədən toponimlərin, marka adlarının elektron bazalarının yaradılması və onlardan istifadə mexanizmlərinin işlənilməsi, onomastik vahidlərin transliterasiyası, adların orfoqrafiyası və s. məsələlər bu sistemin vacib komponentləri kimi nəzərdən keçirilib. Milli Onomastika Sisteminin formalaşdırılması üzrə müvafiq təkliflər irəli sürülüb.

Açar sözlər — *onomastika; xüsusi isim; domen adı; toponim; ad; soyad; insan coğrafiyası*

I. GİRİŞ

Dil, tarix və coğrafiya elmlərinin kəsişməsində yerləşən onomastika bu gün milli dil siyasətinin aparıcı istiqamətlərindən birinə çevrilmişdir. Qloballaşma dövründə onomastikanın elmi-praktik və mədəni-tarixi əhəmiyyəti ilə birgə, ictimai-siyasi mahiyyəti də getdikcə güclənməkdədir.

Azərbaycan dövlət müstəqilliyinə nail olaraq beynəlxalq münasibətlərin fəal subyektinə çevrildikdən sonra milli onomastika sistemi global məkana cəlb

olunmuşdur. Azərbaycan onomastik vahidlərindən bu gün siyasi, iqtisadi, sosial-mədəni məzmun daşıyan beynəlxalq rəsmi sənədlərdə geniş istifadə olunur və adlarımız funksionallığına görə artıq milli arealla sərhədlənmir. İndi tək-cə ünvan funksiyasını yerinə yetirməklə məhdudlaşmayan bu adlar həm də yüksək assosiasiyalı semiotik işarələr kimi özündə böyük milli potensial daşıyır. Eyni zamanda, beynəlxalq əmək bazarına, qlobal təhsil, elm, mədəniyyət mühitinə inteqrasiya Azərbaycanın xüsusi adlarının daha geniş sferada funksionallıq qazanmasına şərait yaradır. Adlarımız bütün dünyada milli nişanə kimi çıxış edir. İnternet cəmiyyətin həyatına dərinlən nüfuz etdikcə virtual mühitdə xüsusi adlardan, yəni onomastik vahidlərdən istifadə məsələlərinin də aktuallığı artır və xüsusi adların virtual mühitinə öyrənilməsi zərurətə çevrilir. “Virtual mühitdə xüsusi adların yerinə yetirdiyi funksiyaların çoxşaxəliliyini və bu sahədə həllini gözləyən elmi problemlərin əhatəsini nəzərə alaraq hazırda internetdə istifadə olunan xüsusi adların tədqiqi müasir dilçiliyin ən perspektivli istiqamətlərindən sayılır” [1].

Dilin virtual mühiti yeni bir linqvistik fenomeni – virtual onomastikanı, yaxud onomastik informatikanı formalaşdırır. Onomastik informatika elektron sferada, eləcə də internet məkanında adların istifadəsini öyrənən sahədir. Bu sahənin nəzəri və praktiki məsələlərinin həlli məqsədilə Azərbaycan Respublikasında onomastika məsələləri üzrə dövlət Konsepsiyasının hazırlanması və Milli Onomastika Sisteminin formalaşdırılması məqsədəuyğun olardı.

II. DOMEN ADLARI MİLLİ ONOMASTİKA SİSTEMİNİN TƏRKİB HİSSƏSİ KİMİ

Bu sistemin ən mühüm aspektlərindən biri domen adları ilə bağlıdır. Real dünyada coğrafi məkanı toponimlərin, insanı antroponimlərin, məhsulun çeşidini, növünü ktematonimlərin və s. köməyi ilə müəyyənləşdirdiyimiz kimi, virtual dünyada da bu vəzifəni domen adları yerinə yetirir. Domen adları yeni onomastik vahidlərdir. Onların qeydiyyatı, istifadəsi, alqı-satqısı, informasiya müharibəsində domen adlarından istifadə ilə bağlı məsələlərin tənzimlənməsi üçün bir sıra təxirəsalınmaz tədbirlərin həyata keçirilməsinə ehtiyac vardır [2].

Milli informasiya resurslarının genişləndirilərək təkmilləşdirilməsi, domen adlarının milli bazasının yaradılması, strateji əhəmiyyət daşıyan sözlərin domen adı kimi qeydiyyatdan keçirilməsinin təmin olunması zəruridir.

Beynəlxalq səviyyədə qeyri-latin domen adlarının tətbiqi genişləndikcə

Azərbaycanda da milli əlifba hərflərindən istifadə olunmaqla domenlərin yaradılması mümkün olacaq və bu, ölkəmizdə informasiyalı cəmiyyətin inkişafına hərtərəfli təkan verəcək. Ümumilikdə isə milli domen adlarının yaradılması ilə bağlı təşəbbüslərin türk dövlətlərini əhatə edən mədəni-coğrafi arealda həyata keçirilməsi, ortaq türk dili ideyasının məhz bu cür təşəbbüslərlə gerçəkləşdirilməsi daha məqsədəuyğun olardı. Türkdilli xalqlar arasında ümumi informasiya məkanının yaradılması üçün müştərək şəkildə müvafiq konvertasiya, tərcümə və çevirmə sistemlərinin işlənilməsi zəruridir. Bütün bu məsələlər Milli Onomastika Sisteminin əsas aspektlərindən birini təşkil etməlidir.

III. ADLARIN MODELLEŞDİRİLMƏSİ

Adların modelləşdirilməsi, vahid ad sisteminin yaradılması, ad və soyadlarda təhrif tiplərinin aradan qaldırılması, müxtəlif mənbələr əsasında adların elektron bazalarının yaradılması və onlardan istifadə mexanizmlərinin işlənilməsi, onomastik vahidlərin transliterasiyası normalarının dəqiqləşdirilməsi, adların orfoqrafiyası qaydalarının nizama salınması və s. elektron hökumət quruculuğunun mühüm istiqamətləridir. Bu problemlərin vaxtında öz həllini tapmaması elektron demokratiya modelinin cəmiyyətdə tam bərqərar olmasına, bu modelin tətbiq mexanizmlərinin təcrübədə özünü doğrultmasına müəyyən əngəllər yarada bilər. Bütün bunları nəzərə alaraq qeyd edilən məsələlər də Milli Onomastika

Sisteminin formalaşdırılması prosesində xüsusi diqqətdə saxlanmalıdır.

Məlum olduğu kimi, qlobal informasiya cəmiyyəti şəraitində Azərbaycan vətəndaşlarının adlandırılması məsələləri çox mühüm əhəmiyyət kəsb edir. Azərbaycanda ad və soyadlarla bağlı mövcud problemlərin həlli, bu sahədə texniki məsələlərin nizamlanması, qlobal müstəvidə xalqımızın milli identikliyinə təmin olunması, adların ekologiyası, təhriflərlə mübarizə, ad yaradıcılığı, yeni yaranan adlarla bağlı virtual referendumların təşkili, ad və soyadların milli-mənəvi dəyər kimi qorunaraq nəsillərdən-nəsillərə ötürülməsi, mültikulturalizm ənənələrinin inkişafı, dünya azərbaycanlılarının, eləcə də Azərbaycanda yaşayan əcnəbilərin ad və soyadlarının qeydiyyatı ilə bağlı çətinliklərinin həlli, bu sahədə onlayn məlumat mübadiləsi müasir dövrün aktuallaşdırdığı əsas məsələlərdəndir [3].

Əsrlər boyu milli ad sistemimizin formalaşmasında ciddi rol oynamış ərəb, fars və rus ad sisteminin təsirlərindən irəli gələrək bu gün Azərbaycanın ad sisteminə ciddi bir nizamsızlıq hökm sürür. Bu nizamsızlığa müxtəlif qurumlar tərəfindən hazırlanmış elektron məlumat bazalarının strukturunda da təsadüf edilir. Belə halların aradan qaldırılması üçün qeydiyyat prosesinin ad, ata adı və soyad, ata adının yazılması zəruri olmadıqda isə ad və soyad ardıcılığına əsaslanan dəqiq vahid standart üzrə həyata keçirilməsi təmin olunmalı, bu məsələlər də Milli Onomastika Sisteminin

son dərəcə vacib aspekti kimi nəzərə alınmalıdır.

Xüsusi adların elektron lüğətlərinin hazırlanması da günün tələbidir. Onomastik vahidlərin, o cümlədən şəxs adları və toponimlərin yerli dillərdən xarici, xarici dillərdən isə yerli dillərə düzgün transliterasiyası qloballaşma şəraitində müasir Azərbaycan dilçiliyinin ən aktual problemlərindən biridir. Uzun illər ərzində əcnəbi xüsusi adların ana dilimizə transliterasiyasının fonetik prinsip əsasında həyata keçirildiyini, bu ənənənin müsbət cəhətlərini nəzərə alaraq alınma sözlərin dilimizə transliterasiyası zamanı məhz fonetik prinsipə üstünlük verilməli və əcnəbi adlar dilimizdə tələffüz edildiyi kimi yazılmalı, zərurət olduqda mötərizədə adın orijinalı da göstərməli, adların transliterasiyasında türk dilinin təsirləri mümkün qədər neytrallaşdırılmalı, adların qarşılıqlı transliterasiyasını təmin edən texniki sistemlər hazırlanmalıdır. Eyni zamanda xüsusi adların lüğəti də tərtib edilməlidir.

Azərbaycan Respublikası Ədliyyə Nazirliyinin təqdim etdiyi ad və soyadlar siyahısı əsasında apardığımız təhlillər Azərbaycan ad və soyadlarında kifayət qədər təhriflərə yol verildiyini göstərir. Bu təhriflərin ortaya çıxmasında rus dili amilinin, məhəlli amillərin, insan amilinin, siyasi amillərin, etnik-mədəni amillərin, fonetik hadisələr amilinin, mühafizəkarlıq amilinin böyük rolu vardır. Sait və samitlərin əvəzlənməsi, ixtisarı, artımı, ahəng qanununun pozulması və yerdəyişmə, müxtəlifvariantlılıq, tire

işarəsindən istifadə, digər dillərin təsiri, adın xarakterik olmayan sait və samitlərlə başlaması, mürəkkəb sözlər əsasında düzələn adların komponentlərinin ayrı-ayrı yazılması, şəxs adları fəmiiliya düzəldən şəkilçi qəbul edərəkən sabitliyin pozulması və s. Azərbaycan ad və soyadlarında təsadüf edilən əsas təhrif tipləridir. Milli Onomastika Sisteminin yaradılması bu problemin həllinə kömək göstərə, daha nizamlı bir vəziyyət yarada bilər.

IV. MİLLİ TOPONİMİK GEOİNFORMASIYA SİSTEMİNİN PERSPEKTİVLƏRİ

Milli Onomastika Sisteminin ən mühüm komponentlərindən biri də Azərbaycanın bütün bölgələrinin toponimik məlumatlarını bir-birinə bağlayaraq vahid, qarşılıqlı əlaqəli və əlçatan məlumat bazasının mövcudluğuna xidmət edən Milli Toponimik Geoinformasiya Sistemi ola bilər [4].

Toponimlərin belə bir sistem əsasında pasportlaşdırılması, onların mənşəyinin zaman, məkan, aid olduqları xalq, şəxs, dövlət, relyef və s. üzrə toplanması, təsviri analitik qərar qəbul etmə məqsədilə vəziyyətin analizinə, proqnozlaşdırılmasına, müxtəlif tendensiyaların aşkarlanmasına böyük kömək göstərəcəkdir. Toponimlərin ərazidəki mikrotoponimlər fonunda statistik analizi, yer adlarının elektron məlumat bazalarının başqa ölkələrdə, xüsusilə də türkdilli ölkələrdə mövcud olan analoji bazalarla çarpazlaşdırılması, xalqların hərəkət trayektoriyalarının müəyyənləşdirilməsi, indiyə qədər qaranlıq

qalan bir çox tarixi həqiqətlərin üzə çıxarılması baxımından son dərəcə faydalı ola bilər” [5].

Dünyadakı qəbiristanlıqlar, ayrı-ayrı məzar yerləri haqqında data məlumatlarının, o cümlədən mətn qeydlərinin, fotosəkillərin və s. toplandığı *Findagrave.com* və *Billiongraves.com* kimi platformalar da belə bir sistemin tərkib hissəsi kimi böyük səmərə verəcəkdir.

Virtual mühitdə fiziki və hüquqi şəxslərin identifikasiyasının əsas elementlərindən olan adlar və soyadlar da elektron hökumətin vacib tərkib hissəsidir. Azərbaycan Respublikasının Elm və Təhsil Nazirliyinin İnformasiya Texnologiyaları İnstitutunun 2009-cu ildə irəli sürdüyü “Adlar və Soyadlar” Dövlət İnformasiya Sisteminin yaradılması təşəbbüsü isə bu sahədə mövcud olan problemlərin konseptual həllinə, cəmiyyətimizdə ad və soyadlarla bağlı həm indi, həm də gələcəkdə nizamlı bir vəziyyətin yaranmasına yönəlmiş mükəmməl bir platforma sayıla bilər [6]. Bu platforma da Milli Onomastika Sistemində əhəmiyyətli yer tutmalıdır.

Xüsusi adlar sosial informasiya daşıyıcılarıdır. İnsanların öz övladlarına qoyduqları adların statistik analizi ictimai şəurun inkişaf meyillərinin aydınlaşdırılmasından ötrü əsaslı istiqamətlər verir [7].

Böyük verilənlərin analizinin geniş imkanlar yaratdığı müasir dövrimizdə cəmiyyətin idarə olunmasında adlardan istifadə əmilindən irəli gələrək yeni bir elm sahəsi kimi insan coğrafiyasının (human geography) dünyanın müxtəlif elm

mərkəzləri üçün populyarlığı get-gedə artmaqdadır.

“İnsan coğrafiyası cəmiyyətləri, onların xarakterik xüsusiyyətlərini, fəaliyyətlərini, bu cəmiyyətə xas ictimai düşüncə tərzini onların bağlı olduqları coğrafi çevrə, milli məkan məlumatları ilə əlaqəli şəkildə öyrənən elmi istiqamətdir” [5].

Ərazinin tarixi-coğrafi xüsusiyyətləri ilə əhali tərkibi arasındakı münasibətlərin müəyyənləşdirilməsi, məskunlaşmaların bölgələr üzrə paylanmasına təsir göstərən təbii və bəşəri amillərin üzə çıxarılması, bu təsirlərin müasir dövrdə də izlənməsi zamanı onomastik vahidlər ciddi kriptografik məlumatların aşkarlanmasına imkan verir. Elə ona görə də İKT-nin tətbiqi ilə onomastikanın bir istiqaməti kimi geonomastika elminin meydana gəlməsi və inkişafı diqqəti cəlb edir.

Müxtəlif layihələrin icrası üçün geonomastika da Milli Onomastika Sistemi üçün son dərəcə faydalı bir metasahədir. Xüsusi adlar səviyyəsində aparılan statistik analizlər insan coğrafiyasının, dilin və mədəniyyətin öyrənilməsindən ötrü alternativ metodlar nümayiş etdirir. Adlandırmanın ideologiyalarla, milli, dini, siyasi və s. təmayüllərlə bağlılığını nəzərə alsaq, onların məzmununda ifadəsini tapan ictimai meyilləri də izləmək mümkündür.

V. MARKA ADLARI BAZASI

Milli Onomastika Sisteminin ən mühüm tərkib hissələrində biri də Marka Adları Bazası ola bilər. Məlum olduğu kimi, internet erasında xüsusi adlar

rəqəmsal marketingin vacib alətidir. “Xüsusi ad isə markanın vacib kommunikasiya vasitəsidir, marka və markalaşmanın əhəmiyyətli hissəsidir. O özünün fikir yükü ilə istehlakçıları cəlb edir” [8].

Müasir cəmiyyətdə marka adı və onun linqvo-funksional semantikasi dünyanın dil xəritəsinin mental vahidlərindən biri kimi də çıxış edir. Marka adları ilə bağlı nəzərə alınmalı olan çoxsaylı mühüm cəhətlər bir çox ölkələrdə olduğu kimi Azərbaycanda da marka adlarının elektron bazasının yaradılması zərurətini aktuallaşdırır. Adaptasiya, adların analizi, marka yaratma və s. bu kimi modulları özündə birləşdirən belə bir sistem əksər ölkələrdə olduğu kimi, Azərbaycanda da bazar araşdırmaları və marketing planının əhəmiyyətli hissəsinə çevrilə bilər. Bir çox xarici ölkələrdə ödənişli və yaxud ictimai əsaslarla marka adları təklif edən onlayn məlumat bazaları vardır. Bu bazalara daxil olmaqla minlərcə marka adı əldə etmək mümkündür. Bir sıra ölkələrdə bu cür bazalar marka adlarının qeydiyyatını aparan qurumlarda mövcuddur. Azərbaycanda da belə bir sistemin yaradıcısı və tənzimləyicisi qismində müvafiq elmi qurumlarla müştərək şəkildə aidiyyəti dövlət qurumu (Əqli Mülkiyyət Agentliyi) çıxış edə bilər. Bu sistemin yaradılması və marka adlarının analizinə dilçi-mütəxəssislərin cəlb olunması zəruridir. Adların analizi tələffüz, məna, qrafik həll, ad təkrarları, qlobalıq, vizyon, assosiasiya, marka arxitekturası və s. istiqamətlər üzrə aparılmalı və müştərilərə ən uyğun təkliflər verilməlidir.

Uğurlu marka adlarının tapılması üçün firmanın əməkdaşları arasında sorğu keçirilməsi, geniş ictimaiyyətin qatıldığı müsabiqələrin təşkili xarici ölkələrdə və Azərbaycanda səmərəli təcrübədir. Xüsusilə sosial şəbəkələr fikir mübadiləsi platforması kimi belə seçimlərin aparılması üçün ideal imkanlar yaradır.

NƏTİCƏ

Bütün bu qeyd edilənlər virtual məkanda, o cümlədən e-dövlət mühitində Azərbaycan dilinin istifadəsi, qorunması və inkişafı üçün Milli Onomastika Sisteminin də yaradılmasının vacibliyini göstərir.

Domen adları, adların modelləşdirilməsi ilə bağlı mexanizmlərin formalaşdırılması, xüsusi adların, o cümlədən toponimlərin, marka adlarının elektron bazalarının yaradılması və onlardan istifadə mexanizmlərinin işlənilməsi, onomastik vahidlərin transliterasiyası, orfoqrafiyası və s. məsələlər bu sistemin vacib komponentləri kimi çıxış edəcəkdir.

Milli Onomastika Sistemində Milli Toponimik Geoinformasiya Sisteminin də öz yerini tutması müxtəlif mövzularda araşdırmalara, nəsillər arasında rabitənin təmin olunmasına, qlobal məkanda Azərbaycanla bağlı tarixi həqiqətlərin təqdim olunmasına böyük kömək göstərəcəkdir.

Milli Onomastika Sisteminin yaradılması üzrə layihələr müxtəlif elm sahələrinin mütəxəssislərinin birgə əməkdaşlıq etdiyi multifunksional qruplar

tərəfindən yerinə yetirilməlidir.

İSTİNADLAR

- [1] Аникина, Т.В. Виртуальный антропоним в системе современной коммуникации // – Екатеринбург: Уральский филологический вестник Серия: Язык. Система. Личность: Лингвистика креатива, – 2012. № 2, – с. 58
- [2] Qasımova R.T., Cəfərov Y.M. İnternetdə Azərbaycan əlifbası ilə domen adlarının tərtibatındakı bəzi problemlər // İnformasiya cəmiyyəti problemləri, – 2020, № 1, s. 64-74
- [3] Əliquliyev R.M., Cəfərov Y.M., Qurbanova Ə.M. Qlobal informasiya cəmiyyəti şəraitində Azərbaycan vətəndaşlarının adlandırılması məsələləri // “Türkologiya” – 2020, № 1, – s. 31-42
- [4] Əliquliyev R.M., Cəfərov Y.M., Qurbanova Ə.M. Milli toponimik geoinformasiya sisteminin yaradılmasının konseptual əsasları // İnformasiya cəmiyyəti problemləri – 2022, № 1, s. 3-10
- [5] Cəfərov Y.M. Geonomastika və insan coğrafiyası məsələləri // Ümummilli lider Heydər Əliyevin anadan olmasının 93-cü ildönümünə həsr olunmuş Gənc Tədqiqatçıların IV Beynəlxalq Elmi Konfransının (Qafqaz Universiteti) materialları, – Bakı: – 29-30 Aprel, – 2016, – s. 158
- [6] “Adlar və Soyadlar” Dövlət İnformasiya Sisteminin yaradılması Konsepsiyası // AMEA-nın İnformasiya Texnologiyaları İnstitutu, Ekspres-İnformasiya, – Bakı: İnformasiya Texnologiyaları, – 2010. – 49 s.
- [7] Никонов В.А. Имя и общество / В.А.Никонов. – Москва: Наука, – 2007. – 278 с.
- [8] Будник А.В. Брендинг в управлении маркетинговой активностью / А.В.Будник, Н.К.Моисеева, М.Ю.Рюмин [и др.]. – Москва: Омега-Л, – 2003. – 410 с.

Milli dil korpusunda Azərbaycan dilinin tədqiqi məsələləri

Rəna Mahmudova

AMEA Nəsimi adına Dilçilik İnstitutu, Bakı, Azərbaycan

rena.memmedova.1991@inbox.ru

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

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

I. GİRİŞ

Korpus dilçiliyinin inkişafı, eləcə də korpusun qurulması müasir dilçiliyin aktual

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

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

II. MİLLİ DİL KORPUSLARI

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

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

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

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

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

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

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

III. AZƏRBAYCAN DİLİNİN MİLLİ KORPUSU

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

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

konkretləşdirilə və onlara lazım olan informasiyanı əldə etmək imkanı yaradılmalıdır.

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

IV. DİL KORPUSLARINDA LÜĞƏTLƏR BLOKU

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

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

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

terminlərin tədqiqat obyektini kompüter vasitəsi ilə tərtib olunan lüğətlərdir.

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

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

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

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

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

NƏTİCƏ

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

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

İSTİNADLAR

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

Dialektometriya və Azərbaycan dili: problemlər, həlli yolları və perspektivlər

Əfruz Qurbanova¹, Mehriban Bağirova²

^{1,2}İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

¹*afruz1961@gmail.com*, ²*mehriban.amea@gmail.com*

Xülasə — Dialektologiya dialektləri öyrənir, dialektometriya isə dialekt fərqlərinin, yəni linqvistik fərqlərin ölçülməsidir, onların paylanması ilk növbədə coğrafiya ilə müəyyən edilir. Hazırkı dövrdə yeni informasiya mənbələri və analitik proqram təminatları dialektometriyanın tətbiq dairəsini genişləndirir. Tədqiqat işində dialektometrik metodlar təhlil olunub. Azərbaycanda dialektometrik metodlardan istifadə imkanları araşdırılıb, tətbiqi istiqamətində tövsiyələr verilib.

Açar sözlər — *dialektologiya; dialektometriya; dialektoloji atlas; linqvistik coğrafiya*

I. GİRİŞ

Müasir dövrdə cəmiyyətdə baş verən dəyişikliklər, elm sahələrinin inteqrasiyasının güclənməsi dialektoloji tədqiqatlara yeni yanaşmanı zəruri edir.

Dialektlərin ən böyük tədqiq sahələri dialekt leksikoqrafiyası və linqvistik coğrafiyadır [1]. Bir dilin dialektlərinin linqvocoğrafi üsulla sistemli şəkildə öyrənilməsi Avropada XIX əsrin ikinci yarısında başlanmışdır. Bu üsul vahid sorğu anketi vasitəsilə çoxlu sayda yaşayış məntəqələrindən məlumat toplamaqdan ibarətdir. Anketin hər bir sualına verilən cavablar onların ərazi bölgüsünü əks etdirən dialekt xəritəsinin əsasını təşkil edir və eyni

yaşayış məntəqələri üçün tərtib edilmiş dialekt xəritələri toplusu dialektoloji atlasda birləşdirilir [2]. İlk belə atlas Q.Venkerin alman dialektləri atlası idi (1876-1881) [3]. J.Jilyeron (1902-1910) tərəfindən fransız dili atlası və digər Avropa dillərinin atlası nəşr edildi [4]. Hal-hazırda bütün əsas Avropa dillərində yüzlərlə və ya minlərlə xəritə daxil olmaqla dil atlasları var. Bir çox dildə məlumatların toplandığı layihə 1975-ci ildən nəşr olunan Avropanın Linqvistik Atlasıdır.

Son illər Azərbaycan dilinə yüksək dövlət qayğısının göstərilməsi dilçiliyimizin müxtəlif sahələrinin inkişafına geniş imkanlar açıb. AR Prezidentinin 9 aprel 2013-cü il tarixli Sərəncamı [5] ilə təsdiq edilmiş “Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsinə və ölkədə dilçiliyin inkişafına dair” Dövlət Proqramına “Azərbaycan dilinin müxtəlif dialekt və şivələrinin müasir dövrün tələblərinə uyğun olaraq öyrənilməsinin təmin edilməsi” məsələsinin daxil edilməsi Azərbaycan dili dialektlərinin müasir dilçiliyin nailiyyətləri səviyyəsində tədqiqini qarşıya mühüm vəzifə kimi qoyub. Proqramın icrası istiqamətində “Azərbaycan dilinin Naxçıvan dialektoloji atlası”, “Azərbaycan dilinin Qarabağ və Şərqi Zəngəzur dialektoloji atlası”, “Azərbaycan

dilinin Naxçıvan dialektoloji lüğəti”, “Azərbaycan dilinin Qarabağ dialektoloji atlası”, “Azərbaycan dilinin Qarabağ şivələri” monoqrafiyası və “Azərbaycan dilinin Qarabağ dialektoloji lüğəti” çap olunub.

Bu gün kütləvi informasiya vasitələrinin və yeni texnologiyaların sürətli inkişafı yerli dillərə və dialektlərə böyük təsir göstərərək onları yox olmaq təhlükəsi qarşısında qoyub. Yerli dillərin və dialektlərin qorunması istiqamətində dialektometrik tədqiqatların aparılması və dil variasiyalarının müəyyən edilməsi zəruridir.

Tədqiqat işində sürətlə qloballaşan hazırkı dövrdə riyazi metodlardan istifadə etməklə, xüsusən də dialektometriya üzrə tədqiqatlar aparmaqla Azərbaycan dilinin dialekt fərqlərinin daha dərinədən öyrənilməsinin zəruriliyi əsaslandırılmış, bu istiqamətdə təkliflər işlənmişdir.

II.DİALEKTOLOGİYA

Azərbaycan dili dialektləri xalqımızın tarixini, mədəniyyətini, etnoqrafiyasını və folklorunu ulu baba-nənələrimizin dilində yaşatmaqla yanaşı, dilimizin incəliklərini mühafizə edərək gələcək nəsillərə çatdırır.

Azərbaycan dili dialekt (ləhcə) və şivələrinin öyrənilməsi dilçiliyimizin ən önəmli məsələlərindəndir. Hər bir dialektin və ya şivənin materiallarının toplanması, tədqiqi xalqımızın dilinin və tarixinin araşdırılmasında mühüm rol oynayır.

Dialekt fərqləri dialektologiyanın əsas tədqiqat obyektidir. Dialektologiya dilçiliyin ərazi dialektlərini, dialekt fərqlərini və dialekt dilini hazırkı vəziyyətində və

tarixində öyrənən bölməsidir. Azərbaycan dilinin dialekt və şivələri coğrafi ərazi prinsipi üzrə 4 qrupa bölünür [6]:

- Azərbaycan dilinin şərq qrupu dialekt və şivələri - Quba, Bakı, Şamaxı dialektləri və Muğan, Lənkəran şivələri;
- Azərbaycan dilinin qərb qrupu dialekt və şivələri - Qazax, Qarabağ, Gəncə dialektləri və Ayrım şivələri;
- Azərbaycan dilinin şimal qrupu dialekt və şivələri - Şəki dialekti və Zaqatala-Qax şivələri;
- Azərbaycan dilinin cənub qrupu dialekt və şivələri - Naxçıvan, Ordubad, Təbriz dialektləri və İrəvan şivəsi.

Dialektlərin öyrənilməsində sosiolinqviistik üsullardan: uzunmüddətli panel müşahidəsi, nitq yazısının müxtəlif növləri, gündəliklər, anketlər, müsahibələr, sorğular, gizli testlər və s. istifadə edilir.

Dialektlər dilin yazılmamış formasıdır, dialektoloqlar onları öyrənmək üçün anket üsulundan və birbaşa müşahidə üsulundan istifadə edirlər.

Anket üsulu ilə dialektin linqviistik xüsusiyyətləri haqqında məlumatların toplanması xüsusi tərtib olunmuş sorğu vərəqəsinin suallarına linqviistik səriştəli şəxslərdən (müəllimlər, kənd ziyalıları və s.) yazılı cavablar almaqla həyata keçirilir.

Müasir dialektlərin öyrənilməsi üsullarından biri də tədqiqatçının canlı dialekt nitqini dinləyərək linqviistik xüsusiyyətləri müəyyənləşdirdiyi zaman birbaşa müşahidədir. Birbaşa müşahidə üsulu ilə dialektoloqlar əvvəlcədən tərtib edilmiş anket proqramı əsasında dialekt danışanlarının canlı nitqini qeydə alırlar.

Hansı nəsilin və danışanların daha çox dialekt olduğunu təyin etmək üçün dialektliyin ölçülməsi zəruridir. Xüsusi kompüter proqramlarından istifadə edərək nitq seqmentlərini dəfələrlə təkrarlamağa və təhlil etməyə imkan verən lent yazılarından daha çox məlumatlar əldə etmək olar. Belə yazıların fondu gələcək tədqiqatçılar üçün müasir dialekt nitqini qoruyub saxlayacaqdır [7].

III. DIALEKTOMETRİYA

20-ci əsrin ikinci yarısında çox sayda yığılmış dialektoloji material və informasiya texnologiyalarının inkişafı dialektlərin öyrənilməsinə yeni yanaşmanın - dialektometriyanın yaranmasına səbəb oldu. Dialektometriya statistik metodların istifadəsilə dil variasiyasını öyrənir. Dialektometrik metodlar dialektoloqlara dialekt fərqlərini kəmiyyətə qiymətləndirməyə və bunun əsasında dil dəyişikliyinə ölçməyə imkan verir [8, 9].

Dialektometrik metodlar 1970-1980-ci illərdə Jean Séguy (Fransa) və Hans Goebel (Avstriya) tərəfindən əvvəllər toplanmış atlas məlumatlarından istifadə edərək Roman dialektlərinin öyrənilməsinə statistik yanaşmalar tətbiq edildikdən sonra fəal şəkildə inkişaf etməyə başladı [10].

Dialektometrik metodlar yaşayış məntəqələri arasında “məsafə” anlayışına əsaslanır. Burada məsafə dedikdə, bir nöqtədə anket suallarına verilən cavabların digər nöqtədə suallara verilən cavablardan necə fərqləndiyini göstərən hər hansı iki yaşayış məntəqəsi (X və Y) üçün atlas məlumatlarına əsasən qiyməti hesablayan riyazi funksiya $d(X, Y)$ başa düşülür.

Praktikada müxtəlif funksiyalar tətbiq edilə bilər, lakin istənilən halda onlar ən azı üç tələbə cavab verməlidirlər [2]:

- Anketin bütün suallarının cavabları iki nöqtədə üst-üstə düşürsə, onların arasındakı məsafə 0-a bərabərdir $d(X, Y)=0$;
- Funksiya simmetrikdir, yəni $d(X, Y) = d(Y, X)$;
- X və Y anketləri böyük sayda suallarda nə qədər çox fərqlənirsə, onlar arasındakı məsafə bir o qədər çox olur.

Dialektometriya klassik dialektologiyadan yaranan yeni bir istiqamətdir ki, burada bir bölgədəki müxtəlif dialektlər arasındakı fərqlər statistik olaraq hesablanır, dialekt xəritələri və atlaslarından istifadə etməklə təqdim olunur.

Linqvistik xəritə bir dilin danışanlarının coğrafi paylanmasını və ya dil ailəsinin izoqlosslarını göstərən tematik xəritədir. Linqvistik atlaslar çoxlu sayda yerlərin dialekt profilini ətraflı sənədləşdirən empirik məlumat bazası kimi xidmət edir. Linqvistik atlaslarda olan çox sayda məlumatdan əsas nümunəni xülasələşdirmək və vizuallaşdırmaq üçün müxtəlif tanınmış ədədi təsnifat metodologiyalarından istifadə olunur.

Dialektoloji atlasın məlumatları əsasında hesablanmış qoşa nöqtələr matrisindən bir neçə üsulla istifadə etmək olar. Birincisi, bu bir çox dialekt xəritələrində eyni vaxtda variantların sərhədinin keçdiyi izoqlossların mühüm dəstələrini, yəni xətləri tapmağa imkan verir. İkincisi, avtomatik klasterləşmə metodunun köməyiylə dialekt və ləhcələrin

təsnifatını əldə etmək olar. Bu halda klasterlər (alt dialektlər) bir-birindən bir qədər fərqlənən, lakin qonşu klasterlərdən nəzərəcarpacaq dərəcədə fərqlənən nöqtələri birləşdirəcək [11].

Daha gənc bir sahə korpus dialektometriyasıdır, burada statistik metodlar dialekt xəritələrinə deyil, korpus məlumatlarına tətbiq edilir [12].

Dilin bir neçə dialekti arasında yaxınlığın qiymətləndirilməsi maraqlı, lakin mürəkkəb tədqiqat mövzudur. Bu cür qiymətləndirmə verilən dil məkanında dialektlərin nə qədər qarışdırılan və ya fərqləndirilən olduğunu göstərir. Beləliklə, əldə edilən verilənlər dialekt fərqlərini, izoqlossları, klasterləri və s. vizuallaşdırmaq üçün geniş imkanlar açır [13].

Dialektometrik analizin əsas məsələsi linqvistik deyimlərin bir-biri ilə müasir “oxşarlıq xəritəsi”nin əldə edilməsidir. Bu halda oxşarlıq təkcə idiomların əlaqəsi ilə deyil, həm də son miqrasiya və ya başqa bir dilin ümumi təsiri ilə yarana bilər. Qiymətləndirmə aşağıdakı bölmələr üzrə aparıla bilər [14]:

- nitqin fiziki səsləndirilməsindəki fərqlər;
- nitqin linqvistik ifadəsindəki fərqlər;
- qavramanın qiymətləndirilməsi əlamətləri;
- avtomatik nitq səsləndirmə sistemlərinin klassifikatorlarının fərqləri.

Dilləri, dialektləri və ya digər nitq növlərini müqayisə etməyin bir sıra yolları var. Dialektlər arasında tələffüz fərqlərini hesablamaq üçün Levenşteyn, Evklid və Manhattan məsafəsi kimi müxtəlif simli məsafələrdən istifadə olunur [15].

Levenşteyn məsafəsi iki ardıcılıq və ya sətirlər arasındakı fərqin ölçüsüdür. Belə ardıcılıqlar öyrənilən dilin sözləri ola bilər. Bir sözün iki tələffüz variantı arasındakı Levenşteyn məsafəsini ölçərkən, bir variantın digərinə çevrilməsi üçün keçməli olduğu əməliyyatların (simvolların daxil edilməsi, silinməsi və dəyişdirilməsi) minimal sayı hesablanır [16]. Çoxölçülü miqyaslama (ÇM) dialekt kontinuumlarını öyrənmək üçün istifadə edilən statistik üsuldur. ÇM mürəkkəb məsafə məlumatlarını şərh edilə bilən aşağı ölçülü təsvirlərə çevirir.

Klaster analiz alqoritmləri (məs. WPGMA, Weighted Pair Group Method using Arithmetic Averages) tədqiq olunan obyektlər çoxluğunu klassifikasiya etmək üçün həmin çoxluqda ən yaxın klasterləri müəyyən etməklə tətbiq olunur. Təhlil üçün mənbə materialı tədqiq olunan obyektlər arasındakı məsafələr matrisidir və alqoritmin işinin nəticəsi klasterlərin birləşmə ardıcılığını göstərən iyerarxik struktur kimi təqdim edilə bilər [17]. Vizual Dialektometriya (VDM) 1998-2000-ci illər arasında Salzburq Universitetində dialektometriya layihəsi kimi hazırlanmış və Dialektoloji Atlas məlumatlarının dialektometrik analizlərini dəstəkləyən alqoritmləri həyata keçirir. O, əvvəlcədən təsnif edilmiş Atlas məlumatlarının idarə edilməsi üçün funksionallıq, verilənlərin təhlilinə müxtəlif dialektometrik yanaşmalar və bu cür təhlillərin nəticələrini vizuallaşdırmaq üçün müxtəlif üsullar təklif edir (dendroqramlar, diaqramlar və ya xəritələr) [18, 19]. Gabmap dialektometriya və kartoqrafiya üçün veb proqramdır. Dialekt məlumatlarını müqayisə və statistik

təhlil etməyə imkan verir. Gabmap yalnız lüğətin və ya digər kateqoriyalı məlumatların müqayisəsini deyil, həm də redaktə məsafəsindən istifadə edərək tələffüzün müqayisəsini həyata keçirən qrafik istifadəçi interfeysidir. Gabmap dialektologiya üzrə tədqiqatçılara kompüter dəstəyi ilə kəşfiyyat və hesablamlar aparmağa imkan verir [20].

NƏTİCƏ

Tədqiqat işində dialektometrik metodlar təhlil olundu. Azərbaycanda dialektometrik metodların dərinlən öyrənilməsi və tətbiqinin vacibliyi müəyyənləşdirildi. Azərbaycan dilinin dialekt fərqlərinin hesablanması üçün dialektometrik metodlardan istifadə imkanları araşdırılaraq tövsiyə olunur:

- Azərbaycan dilinin dialektoloji atlasının məlumatlarının verilənlər bazasına daxil edilməsi və bu məlumatların hər kəs üçün əlverişliliyinin təmin edilməsi məqsədilə veb-portalin yaradılması zəruridir. Portal Azərbaycan dilinin vahid dialekt mühitinin yaranmasına, dialektlərin operativ axtarışına imkan verəcəkdir.
- Verilənlər bazasına daxil edilən məlumatların təhlili üçün müasir vasitələrdən, o cümlədən riyazi metodlardan, xüsusən də dialektometriya üzrə tədqiqatlarda geniş istifadə olunan alqoritmlərdən - klaster analizi və çoxölçülü miqyaslamadan istifadə etmək mümkündür.
- Azərbaycan dilinin vahid dialekt mühitinin yaranması bu dialektlərin daha

dərinlən öyrənilməsi, təsnifatı və təkamülü üçün texnoloji əsas yaradacaqdır.

İSTİNADLAR

- [1] Л.М.Захаровна, Лингвистическая география как основа ареальных исследований мордовских языков, Финно-угорский мир., 2016, №. 3 (28), pp.50-59
- [2] Т.А.Архангельский, Применение диалектометрического метода к классификации удмуртских диалектов, Урало-алтайские исследования, 2021, №2, с. 7-20
- [3] J.Herrgen, The digital wenker atlas (www.diwa.info): an online research tool for modern dialectology, Dialectologia: revista electrònica, 2010, pp. 89-95
- [4] P.Davoine et al. New approach to explore and to study cartographical heritage in dialectology: application to the Linguistic Atlas of France, International Cartographic Conference, 2005
- [5] “Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsinə və ölkədə dilçiliyin inkişafına dair Dövlət Proqramı”nın təsdiq edilməsi haqqında AR Prezidentinin Sərəncamı, <https://e-qanun.az/framework/25537>
- [6] M.Şirəliyev, Azərbaycan dialektologiyasının əsasları, Bakı, “Şərq-Qərb”, 2008, 416 səh.
- [7] Методы изучения диалектов, <https://studfile.net/preview/2378303/page:4/>
- [8] M.Wieling, J.Nerbonne, Advances in dialectometry, Annual Review of Linguistics, 2015, V. 1, №1, pp. 243-264
- [9] J.Nerbonne, W.Kretzschmar, Progress in dialectometry: toward explanation,
- [10] Literary and Linguistic Computing, 2006, V. 21, № 4, pp. 387-397
- [11] H.Goebel et al., Atlas, matrices et similarités: petit aperçu dialectométrique, Computers and the Humanities, 1982, T. 16, pp. 69-84.

- [12] V.Batagelj, T.Pisanski and D.Kerzic, Automatic Clustering of Languages, Computational Linguistics, 1992, V. 18, № 3, pp. 339-352
- [13] B.Szmrecsanyi, Corpus-based dialectometry a methodological sketch. Corpora, 2011, V. 6, №1, pp. 45-76
- [14] S.Pickl et al., Linguistic distances in dialectometric intensity estimation, Journal of Linguistic Geography, 2014, T. 2, №1, pp. 25-40
- [15] M.Mehrabani and J.H.Hansen, Automatic analysis of dialect/language sets. International Journal of Speech Technology, 2015, V. 18, pp. 277-286
- [16] H.Ruch, The role of acoustic distance and sociolinguistic knowledge in dialect identification, Frontiers in psychology, 2018, V. 9, p.818
- [17] W.Heeringa, Measuring dialect pronunciation differences using Levenshtein distance : дис. – University Library Groningen] [Host], 2004, p.324
- [18] S.Galdino, P.Maciel, Weight Pair Group Average Mean Clustering for Interval-valued Data, IEEE Latin American Conference on Computational Intelligence (LA-CCI), 2019. pp. 1-7
- [19] H.Goebl, Recent advances in Salzburg dialectometry, Literary and linguistic computing, 2006, V. 21, №. 4, pp. 411-435
- [20] “Dialektometrie Projekt” – Salzburg. URL: <http://www.dialectometry.com>
- [21] J.Nerbonne et al. Gabmap-a web application for dialectology, Dialectologia: revista electronica, 2011, pp. 65-89. <https://www.raco.cat/index.php/Dialectologia/article/view/245345>

Azərbaycanda jest dili və onun avtomatik tanınması problemləri

Rəsmiyyə Mahmudova¹, Kəmalə Qurbanova²

^{1,2} İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

¹rasmahmudova@gmail.com, ²kemalewamil@gmail.com

Xülasə — **İnformasiya-kommunikasiya texnologiyalarının (İKT) dinamik inkişafı müasir kompüter və robot sistemlərinin jestlərlə idarə olunması üçün əlverişli mühit yaratmışdır. İşdə jestlərin ünsiyyət prosesinin zəruri komponenti olduğu, mühüm əhəmiyyət kəsb etdiyi və müxtəlif peşələrdəki əvəzolunmaz rolu qeyd edilmişdir. Uzaq məsafədən jestlərlə idarə olunan texnologiyaların yaratdığı imkanlar haqqında məlumat verilmişdir. Jest dilinin avtomatik tanınması problemləri analiz edilmişdir. Bu istiqamətdə görülən işlər haqqında məlumat verilmiş və milli jest dili korpusunun yaradılmasının vacibliyi şərh olunmuşdur. Azərbaycan jest dilinin tanınmasında qarşıya çıxan problemlər və onların aradan qaldırılması üçün müəyyən tövsiyələr verilmişdir.**

Açar sözlər — jest dili; qeyri-verbal kommunikasiya; robototexnika; ağıllı ev; jestlərlə idarəetmə texnologiyaları

I. GİRİŞ

Müasir zamanda İKT-nin dinamik və sürətli inkişafı insanlarda informasiyanın daha sürətli və dolğun qəbul edilməsi vərdişləri yaratmışdır. İnformasiya cəmiyyətində əsas resursa çevrilmiş və

informasiya bolluğunun yaranmasına səbəb olmuşdur.

Sağlamlıq vəziyyətindən, maddi təminatından, bilik səviyyəsindən, dilindən, yaşından asılı olmayaraq cəmiyyətin hər bir üzvünün informasiya bolluğundan eyni dərəcədə bəhrələnməsi insanlar arasında sosial bərabərliyin əsasını təşkil edir. İnsanlar arasında sosial bərabərliyi təmin etmək ideyası bir sıra beynəlxalq sənədlərdə öz əksini tapmışdır: “Ümumdünya İnsan Hüquqları” Bəyannaməsi (1948), “Uşaq Hüquqları” Bəyannaməsi (1959), “Qadın hüquqlarının müdafiəsi haqqında” Konvensiya (1979), “Uşaq hüquqları haqqında” Konvensiya (1989), “Xüsusi tələbatlı insanların təhsili haqqında” Salamanka Bəyannaməsi (1994), “Əlillərin hüquqları haqqında” Konvensiya (2006) və s. Beynəlxalq hüquqi normativ bazalara əsaslanan bu sənədlərin qəbul edilməsində əsas məqsəd əlilliyi olan şəxslərin sağlam insanlarla bərabər cəmiyyətə inteqrasiya olunmasıdır [1].

Azərbaycan Respublikasında da əlilliyi olan şəxslərin sosial bərabərliyini təmin etmək istiqamətində müxtəlif hüquqi sənədlər qəbul olunmuşdur. Ölkəmizdə 2004-cü ildə “Azərbaycan Respublikasında xüsusi qayğıya ehtiyacı olan (sağlamlıq

imkanları məhdud) uşaqların təhsilinin təşkili ilə əlaqədar İnkişaf Proqramı (2005-2009-cu illər)” işlənib hazırlanmışdır. 2008-ci ildə ölkəmiz BMT tərəfindən qəbul edilmiş “Əlillərin hüquqları haqqında” Konvensiyaya qoşulmuşdur. Eyni zamanda 2017-ci ildə sağlamlıq imkanları məhdud şəxslərin təhsil hüququnu təhsilin bütün pillələri üzrə digər şəxslərlə bərabər səviyyədə təmin etmək və onların təhsil alması üçün maneəsiz mühit yaratmaq məqsədi ilə “2018-2024-cü illərdə Azərbaycan Respublikasında sağlamlıq imkanları məhdud şəxslər üçün inklüziv təhsilin inkişafı üzrə Dövlət Proqramı” qəbul edilmişdir. Dövlət Proqramında əlilliyi olan şəxslərin tələbatına cavab verən İKT infrastrukturunun yaradılması, istifadəçilərin psixofizioloji xüsusiyyətlərinə uyğun formatda rəqəmsal resursların hazırlanması, aparat-proqram təminatının xüsusi qayğıya ehtiyacı olan insanların tələbatına uyğun sazlanması və s. kimi vəzifələr qarşıya qoyulmuşdur [2].

Yer kürəsinin əhalisinin 5%-dən çoxu eşitmə və nitq məhdudluğu problemi yaşayır. Bu cür insanların yeganə ünsiyyət vasitəsi jest dilidir [3].

Jestlərin etibarlı və dəqiq tanınmasını təmin edən avtomatik tanınma sisteminin qurulması və mütəmadi olaraq təkmilləşdirilməsi vacib və aktual məsələdir.

Jestlərin istifadə dairəsi genişdir. İnformasiyanın jestlərlə ötürülməsi metodu müxtəlif sahələrdə uğurla istifadə olunur. Məsələn, Dövlət Yol Polisi İdarəsinin əməkdaşları yol hərəkətinin təşkilini (nəqliyyat vasitələrinin və ya piyadaların

təhlükəsizliyini qorumaq, yol hərəkətində nizam-intizam yaratmaq, yol-nəqliyyat hadisələrinin qarşısını almaq, fasiləsiz və rahat hərəkəti təmin etmək) təmin etmək məqsədi ilə sürücülərə xüsusi işarələrlə informasiya ötürür və ya idman yarışlarında hakim idmançılara xüsusi işarələr göstərərək oyununu idarə edir.

Müasir zamanda hərəkəti dəqiq aşkarlayaraq əks-əlaqə qura bilən universal qurğulara, smart televizorların, dronların və digər cihazların əl jestlərini idarə etməsinə, cərrahi əməliyyatlarda yüksək dəqiqlik və daha az risk vəd edən robotlara, “ağıllı evlər”ə nəzarətin avtomatlaşdırılmasına, bir sözlə, robototexnikaya tələbat artmışdır.

II. JESTLƏRİN TƏTBİQ SAHƏLƏRİ

Jestlər eşitmə və nitq məhdudluğu olan insanların yeganə ünsiyyət vasitəsidir. Tarixən jestlərlə ünsiyyət şifahi nitqdən əvvəl meydana gəlmişdir. Fikrini şifahi ifadə edən insanlar da jestlərdən mütəmadi olaraq istifadə edirlər. Jestlər və mimika danışan insanın fikrini daha aydın izah edir. Hətta insan danışmadıqda belə üz mimikası və ya hər hansı bir jest ilə bütöv bir fikri qarşı tərəfə ötürə bilir. Göründüyü kimi jestlər sağlamlıq imkanlarından asılı olmayaraq bütün insanların ünsiyyət vasitəsidir. İnformasiyanın sözsüz, jestlərlə ötürülməsi üsulu elmin müxtəlif sahələrindən olan mütəxəssislərin marağına səbəb olmuşdur və belə nəticə əldə olunmuşdur ki, jestlər ünsiyyətin zəruri komponentidir. Ünsiyyət prosesində jestlərin rolunu aşağıdakı kimi təsnifatlandırmaq olar [4]:

• Jestlər dinləyən insana dinlədiyi mövzu ilə yanaşı, eyni zamanda danışan insan haqqında da məlumat ötürür: danışan insanın aktivliyi, emosionallığı, təbiəti, özgüvəni və s. kimi şəxsi xüsusiyyətləri və keyfiyyətləri;

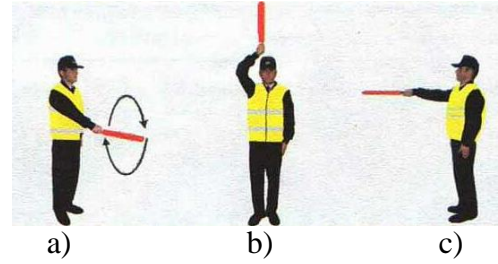
• Ünsiyyətdə olan insanların bir-birinə olan münasibəti haqqında məlumat verir: ünsiyyətin səviyyəsi – müzakirə və ya mübahisə olunan mövzuya münasibəti bildirərkən fikrini ifadə etmə forması;

• Ünsiyyətin gedişinin tənzimlənməsinə imkan verən məlumat: ünsiyyət müddətinin uzanmasına görə məmnunluq və ya əksinə narahatçılıq; üz ifadəsi ilə müzakirə olunan mövzuya maraq və ya etinasızlıq.

Müxtəlif peşələrdə vəzifənin icrası zamanı da jestlərin əvəzolunmaz rolu vardır. İdmançılar, dənizçilər, xüsusi təyinatlı hərbiçilər, yol patrul xidmətinin nizamlayıcıları və s. peşə sahibləri uzaq məsafəyə informasiya ötürərkən xüsusi jestlərdən istifadə edirlər.

Yol patrul xidmətinin nizamlayıcı inspektoru tərəfindən həyata keçirilən jestlər sürücü və piyadaların hərəkətinə yardım etməsi və ya mövcud vəziyyətə uyğun olaraq lazımı xəbərdarlığın verilməsi ilə bağlıdır. Nəqliyyat vasitəsi sürücülərinə nizamlayıcı inspektorun ötürdüyü siqnallara aid misal göstərək: “sürəti artırın” - sol qol dirsəkdən bükülərək sinə önündə çiyin səviyyəsində qaldırılır və sol əlin içi aşağı yönəlməklə saxlanılır, sağ qol dirsəkdən bükülərək bel səviyyəsində yana açılır və jezl biləkdən dairəvi hərəkət etdirilir; “diqqət” - sağ qol jezlə ön tərəfdə yuxarıya qaldırılır və bu vəziyyətdə tutulur; “sola

dönmə” - sağ qol ön tərəfdən irəli uzadılır (Şək. 1.) [5]:



Şəkil 1. a) sürəti artırın, b) diqqət, c) sola dönmə işarələri

Jestlər hər b zamanı kəşfiyyatçılar və xüsusi təyinatlı hərbi qüvvələr tərəfindən uzaq məsafəyə informasiyanın ötürülməsi üçün istifadə edilir (Şək. 2.) [6]:



Şəkil 2. Kəşfiyyatçıların istifadə etdiyi işarələr və onların mənası

Jestlərdən incəsənət sahəsində də geniş istifadə olunur. Buna rəqsi, baleti, pantomima teatrını və s. misal göstərmək olar. Hərflərdən sözlər, sözlərdən ifadə və cümlələr yarandığı kimi rəqqas da müxtəlif kombinasiyalar qurub rəqsin kompozisiyasını hazırlayır. Tamaşaçı məhz bu kombinasiyaları izləyərək rəqqasın “monoloq” və “dialoglar”ını “oxuya” bilir.

Göründüyü kimi, jestlər eşitmə və nitq məhdudiyətli insanların ünsiyyət vasitəsi olmaqla yanaşı, eyni zamanda müxtəlif sahələrdə informasiyanın ötürülməsində əvəzolunmaz funksiyaya malikdir.

III. JESTLƏRLƏ İDARƏ OLUNAN TEKNOLOGİYALAR

Jestlərlə idarəetmə texnologiyalarına əsaslanan sistemlərin tətbiq dairəsi fərdi kompüterlərdən tutmuş cərrahi robotlara qədər olduqca genişdir. Tək bir jest bir sıra klaviatura və siçan düymələrini əvəz edə bilir və işin sürətini kəskin artırır. Simsiz sistemlərin idarə edilməsi texnologiyaları uzaq məsafədən cihazı idarə etməyə imkan verir. Aşağıda göstərilən misallarla bir neçə cihazların iş prinsipini nəzərdən keçirək.

iMotion kontrolleri. iMotion hərəkəti dəqiq aşkarlayaraq əks əlaqə qura bilən universal bir qurğudur. iMotion klaviatura və siçan ilə idarə oluna bilən istənilən kompüterdə istifadə oluna bilər. Qurğu fərdi kompüterlərdə *Windows*, *Mac*, *Linux* əməliyyat sistemlərini, *iOS*, *Android* mobil telefonlarının mövcud proqramlarını, *Oculus Rift* virtual reallıq eynəklərini və oyunları dəstəkləyir. iMotion televizorun pultunu jest idarəetməsi ilə əvəz edərək köhnə televizorları Smart TV-yə çevirə bilir [7].

Kai kontrolleri. Kai cihazlara Bluetooth vasitəsilə qoşulur və onları əl jestləri ilə idarə etməyə imkan verir. Cihaz baş barmaq istisna olmaqla istifadəçinin ovcuna taxılır. Manipulyatorun gövdəsi 13×77×23 mm ölçüsünə malikdir. Cihaz yığcamdır və gündəlik istifadə üçün uygundur. Kai, həmçinin smart televizorları, dronları və

digər cihazları idarə etməyə imkan verir. Cihaz biləkdəki əyilmələri, ovucun dönməsini təsbit edə bilər. İstifadəçi əlini yuxarı və aşağı hərəkət etdirərsə və ya barmaqlarını əyərək sensoru örtərsə, cihaz həmin məlumatı aşkar edə bilər. O, həmçinin 3 ox ətrafında fırlanmaları hiss etmək və aşkar etmək qabiliyyətinə malikdir [8].

Gestonurse cihazı. Gestonurse qurğusu cərrahi əməliyyat zamanı tibb bacısı funksiyasını daşıyır. Ənənəvi cərrahi əməliyyatlarda insan faktoruna əsaslanaraq demək mümkündür ki, tibb bacısının səhv etmək riski vardır. Lakin Gestonurse yüksək dəqiqlik və daha az risk vəd edir. Gestonurse cərrahi alətləri ötürməklə əsas cərraha kömək edir. Belə bir robot sistemi lazımsız ünsiyyəti azaltmaq və kadr çatışmazlığını kompensasiya etmək potensialına malikdir. Robot əl jestlərini 97%-dən çox tanınma dəqiqliyi ilə başa düşür, əmrləri sürətli və dəqiq icra edir. Gestonurse cərrahi əməliyyatın effektivliyini artırır və nəticələri potensial olaraq yaxşılaşdırır [9].

Gest kontrolleri. Gest cihazı naqillərlə ovuc içi və barmaqların hər birinə bərkidilir. Qurğu əsasən “ağıllı ev” idarəetməsində istifadə olunur. Cihaz *Bluetooth* vasitəsilə kompüterə qoşulur. Bu texnologiya istifadəçi ilə fiziki olaraq eyni məkanda olmadan izlənilə və ya hətta idarə oluna bilər [10].

Arrm.io startapı. Arrm.io 2018-ci ildə Belarusiyada hazırlanmış işarə dilinin avtomatik tərcüməsini təmin edən startapdır. Jestlərin tanınması üçün RGB (Red (qırmızı), Green (yaşıl), Blue (mavi) rəng sistemi) kamerası olan fərdi kompüter və ya

noutbuk tələb olunur. Tanınma prosesi *Deep Sign Language* (dərin işarə dili) və neyron şəbəkə texnologiyasının tətbiqi ilə reallaşır. Cihaz eşitmə məhdudiyyətli insanların şifahi danışan insanlarla ünsiyyəti zamanı ən çox istifadə etdiyi (xəstəxanada qəbula yazılmaq, bankla, sığorta şirkəti ilə əlaqə saxlamaq və s.) jestlərin tanınması üçün nəzərdə tutulmuşdur. Məhsul uğurla istifadə olunmaqdadır [11].

Google MediaPipe modeli süni intellekt texnologiyalarına əsaslanaraq real zaman rejimində kompüter və ya mobil telefon kamerası ilə jestləri izləyə və tanıya bilir. Modelin real vaxt rejimində əl/ovuc aşkarlama dəqiqliyi 95,7% təşkil edir. MediaPipe çərçivəsinin üstünlüyü işarələrdən ibarət verilənlər bazasının zəngin olmasıdır və buna görə də müxtəlif işarə dillərinə şamil oluna bilir. Bu model jestlərlə idarə olunan texnologiyalar üçün yeni imkanlar yaradır. Google MediaPipe modeli “ağıllı evlər”ə nəzarətin avtomatlaşdırılması, robototexnika və video oyunlar sahəsində uğurla istifadə olunur [12].

Miyelofon. İşarə dilini tərcümə edən bu məhsulu rus məktəblisi Daniil Kazantsev hazırlamışdır. Məhsul əzələlərin elektromexaniki fəaliyyətini tanıyan elektromiyografiya texnologiyasına əsaslanır. Proqram məhsulu sınaqdan keçirilmiş və sistem jestləri 92,6% dəqiqliklə tanıyaraq uğurlu nəticə göstərmişdir [13].

Kompüter və robot sistemləri üçün jest idarəetmə texnologiyaları böyük potensiala malikdir və aşağıdakı problemlərin həllinə müsbət təsiri vardır [14]:

- naqilli texnologiyaların jestlərlə idarə olunan texnologiyalarla əvəz olunmasına;
- qeyri-standart mühitdə sistemlə istifadəçi arasında qarşılıqlı əlaqənin mümkün olmasına.

Jestlərin avtomatik tanınmasını reallaşdıran İKT projelərinin əsas məqsədi aparat və proqram təminatı vasitəsilə ünsiyyətə maksimum imkan yaratmaqdır. Müasir zamanda jestlərin tanınması üçün mürəkkəb və bahalı avadanlıqlara ehtiyac yoxdur, yalnız müasir mobil telefon və ya kompüter kamerası kifayət edir.

IV. JESTLƏRLƏ ÜNSİYYƏT VƏ ONLARIN AVTOMATİK TANINMASINDA QARŞIYA ÇIXAN PROBLEMLƏR

Jestlər bütün canlılar arasında ünsiyyət vasitəsidir, emosional siqnalların təbii vasitələrlə ötürülməsi üsuludur. Tarixən jest dili müstəqil yaranmışdır və hər millətin özünəməxsus jest dili formalaşmışdır. Jest dili şifahi danışq dilindən asılı olmayaraq inkişaf etmişdir [15,16].

Eşitmə və nitq məhdudiyyətli insanlar jest dilindən yeganə ünsiyyət vasitəsi kimi mütəmadi olaraq istifadə edirlər. Bütün cəmiyyətin diqqətini eşitmə məhdudiyyətli insanların ünsiyyət problemlərinə cəlb etmək məqsədi ilə BMT Baş Assambleyası 2018-ci ildən etibarən hər ilin sentyabr ayının 23-nü “Beynəlxalq Jest Dilləri” günü kimi qeyd olunmasını elan etmişdir.

İndiki dövrdə, robototexnikanın, kompüterlərin və İnternetin dinamik inkişaf etdiyi bir zamanda jest dilinin tanınması və bu dilin istifadəçiləri üçün ünsiyyətin asanlaşması problemi ön plana çıxır.

[17]-də göstərilir ki, əl ilə göstərilən jestlərin tanınması prosesi 4 mərhələdə aşağıdakı ardıcılıqla icra olunur: jesti göstərən əl şəkil və ya video kadr ardıcılığı formasında toplanılır; məlumatlar toplusu müəyyən tanınma metodlarının köməyi ilə analiz edilir, yəni təsvirdən və ya video ardıcılığın hər kadrından əl obyektini tapılaraq seqmentləşdirilir; seqmentləşdirilmiş hissə emal olunur və əsas xüsusiyyətləri aşkar edilir; alınmış nəticələr verilənlər bazasındakı şablonlarla müqayisə olunur və jest tanınır [17].

Təsvirdə jesti göstərən əl obyektinin seqmentləşdirilməsi işində müəyyən problemlər tanınma dəqiqliyinin kəskin şəkildə aşağı düşməsinə səbəb ola bilər [18]: 1) jesti göstərən əlin arxa fonu bir rəngdə olmayanda və ya jesti göstərən əl insanın açıq dəri hissəsi ilə kəsişdikdə seqmentləşdirmə işi çətinləşir; 2) jesti göstərən əl sensor müstəvisinə perpendikulyar olmazsa təsvirdən ovucu və barmaqları aşkar etmək çətinləşir; 3) jesti göstərən əldə əlcək, üzük və s. kimi aksesuarların olması tanınma prosesinə mənfi təsir göstərir; ünsiyyət zamanı jestlər ardıcıl olaraq göstərildiyindən ayrı-ayrılıqda jestlərin başlanğıcını və sonunu dəqiq müəyyən etmədikdə tanınma prosesində xəta baş verir.

Hal-hazırda jest dilinin elementlərini tanıyaraq mətnə, səsə və ya əksinə çevirən sistemlər mövcuddur. İnformasiyanın jestlərlə ötürülməsi üsuluna gözlərin baxışı, mimika, baş, bədən və əl hərəkətlərinin də daxil olması tanınma işini mürəkkəbləşdirir [19]: jestlərin oxşarlığı; hərəkət sürətinin müxtəlifliyi; əlin ölçüsünün və dərinin rənginin fərqliliyi; jesti göstərən insanın

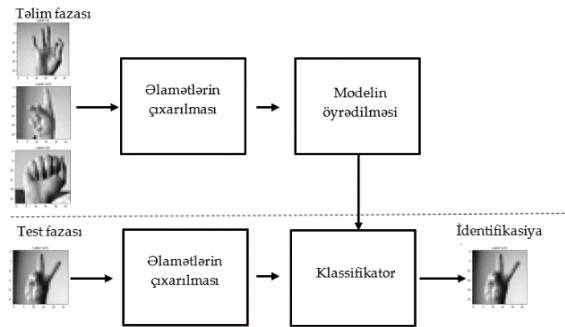
özünəməxsus jestikulyasiya stilinin olması; jest dilində müxtəlif dialektlərin mövcudluğu; və s.

Günbəgün inkişaf edən İKT-nin əldə etdiyi nailiyyətlər sayəsində jestlərin tanınması məsələsində qarşıya çıxan problemləri həll etmək mümkündür.

Azərbaycanda eşitmə və nitq məhdudiyətli insanların ünsiyyətini asanlaşdırmaq məqsədi ilə müvafiq tədbirlər həyata keçirilir. Jest dilinin elementlərinin avtomatik tanınması istiqamətində ölkəmizdə tədqiqatlar aparılmış və müəyyən nəticələr əldə edilmişdir.

İnformasiya Texnologiyaları İnstitutunun əməkdaşları tərəfindən əl jestlərinin tanınması üçün Neyron şəbəkə modelinə əsaslanan yeni yanaşma təklif edilmiş və eksperiment aparılmışdır [20].

Eksperiment təlim və test fazasından ibarətdir. Təlim fazasında əl jestləri təsvirlərinin verilənləri emal olunur və əlamətlər vektoru çıxarılır. Test fazasında əlamət vektorları təlim verilənləri ilə müqayisə olunaraq identifikasiya olunurlar. Şəkil 3-də jestlərin tanınması sisteminin strukturu təsvir olunmuşdur.



Şəkil 3. Jestlərin tanınması sisteminin strukturu

Eksperimentlər Python proqram mühitində yerinə yetirilmişdir. Təklif edilmiş model əlifbanın əksər hərflərini yüksək dəqiqliklə tanıya bilmişdir. Jestlərin tanınması üçün təklif edilən arxitektura jestlərin yüksək dəqiqliklə identifikasiyasını həyata keçirmək üçün müxtəlif maşın təlimi üsulları işlənib. Neyron şəbəkələri əsasında yaradılmış klassifikasiya üsulunun strukturu Cədvəl 1-də təsvir edilmişdir.

Cədvəl 1. Neyron şəbəkələri əsasında yaradılmış klassifikasiya üsulunun strukturu

Layer (type)	Output Shape	Param #
dense_11 (Dense)	(None, 100)	78500
dense_12 (Dense)	(None, 100)	10100
dense_13 (Dense)	(None, 24)	2424

Neyron şəbəkənin gizli laylarında 100 neyron, sonuncu çıxış layında verilənlər bazasındakı siniflərə uyğun olaraq 24 neyron istifadə edilmişdir. Modelin optimallaşma funksiyası rmsprop, itki funksiyası categorical_crossentropy funksiyalarıdır. Modelin öyrədilməsi zamanı batch_size=50, epochs=30 götürülmüşdür. Azərbaycanda milli jest dili korpusu olmadığına görə təklif edilmiş modelin parametrləri “American Sign Language letter database” verilənlər bazasının üzərində aparılan eksperimentlərin nəticəsində alınmışdır.

NƏTİCƏ

İKT-nin dinamik inkişafı jestlərin avtomatik tanınması probleminin həllinə yeni imkanlar yaradır. Milli jest dilimizin elementlərinin tanınması üçün ən başlıca

vəzifə milli jest dili korpusunun yaradılmasıdır. Milli jest dili korpusu ölkəmizdə yaşayan eşitmə və nitq məhdudiyətli insanların ünsiyyət vasitəsi olan jestlərinin əlamətlərini və hərəkət trayektoriyasını özündə saxlayan verilənlər bazasından ibarət olmalıdır. Milli jest dili korpusunun əsasında jestləri səsə, mətnə çevirən yeni metod və sistemlər işlənməlidir. Jestlərin avtomatik tanınması eşitmə və nitq məhdudiyətli insanların təhsil, tibb, ünsiyyət və s. kimi problemlərinin həllini asanlaşdırır və onların cəmiyyətə adaptasiyasını təmin edir. Eşitmə və nitq məhdudiyətli insanların cəmiyyətə inteqrasiya olunması üçün nitqin əsas rol oynamadığı fəaliyyət sahələrində onlardan peşəkar mütəxəssislər yetişdirmək olar və nəticə etibarilə onların mənəvi inkişafı, maddi təminatı yaranar.

İSTİNADLAR

- [1] Г.Н. Пенин, “Инклюзивное образование как новая парадигма государственной политики”, Образование для всех, №9(83), с.41–47, 2010, <https://cyberleninka.ru/article/n/inklyuzivnoe-obrazovanie-kak-novaya-paradigma-gosudarstvennoy-politiki>
- [2] R.Ş. Mahmudova, K.Ş. Qurbanova, “Eşitmə və nitq əngəlli insanların cəmiyyətə adaptasiyası problemləri” Azərbaycanda sosial müdafiə sisteminin inkişafı, dünən, bu gün və sabah mövzusunda beynəlxalq elmi praktik konfransı, Bakı, s. 88-91, 2019.
- [3] U.Farooq, M. S.Mohd Rahim, A.Abid “A multi-stack RNN-based neural machine translation model for English to Pakistan sign language translation”, Neural Computing and Applications. pp. 1-14. 2023.
- [4] Т.А.Гридина, Н.И.Коновалова, Н.А.Воробьева, “Невербальная

- коммуникация: практикум”, Екатеринбург, 106 с., 2015.
- [5] Yol-Patrol Xidməti Haqqında Təlimat, Azərbaycan Respublikasının Daxili İşlər Nazirliyinin Kollegiyası, Qərar № Q25-001-17, Bakı, “12” may 2017.
- [6] Язык жестов спецназа. Тактические жесты. Общение жестами в страйкболе, <http://swd.kiev.ua>
- [7] В.А.Лукашевич, А.П.Сиваков, “Сравнительная оценка степени трансформации двигательного стереотипа в условиях использования технологии "двухслойного экзоскелета" и среды виртуальной реальности”, Инновационные технологии в медицине, №. 2(3), с.80-88. 2015.
- [8] D.A. Booker, “The future of fashion & human gesture control: exploration of a wearable communication device for sign language speakers”, Massachusetts Institute of Technology, pp. 44-46, 2020.
- [9] M.Kangasniemi, K.Suyen, N.Colley, A.Voutilainen, “The use of robots and other automated devices in nurses' work: An integrative review”, International journal of nursing practice, №. 4 (25), pp. 12739, 2019.
- [10] J.O. Aguirre, “IntelihOgarT: A Smart Platform to Contribute Comfort in Intelligent Home Environments by Using Internet of Things Paradigm and Machine Learning”, Technologies and Innovation: 6th International Conference, Guayaquil, Ecuador, November 30–December 3, pp. 139, 2020.
- [11] M.D.Artemov, L.I.Voronova, A.G., “Vovik Designing Soft-Hardware Complex for Gesture Language Recognition using Neural Network Methods”, International Conference on Engineering Management of Communication and Technology (EMCTECH), pp. 1-6. 2020.
- [12] L.A. Domènech “Sing Language Recognition-ASL Recognition with MediaPipe and Recurrent Neural Networks”, дис. Universitat Politècnica de Catalunya, pp.13-14, 2020.
- [13] M.D. Artemov, L.I. Voronova, A.G. Vovik, “Designing Soft-Hardware Complex for Gesture Language Recognition using Neural Network Methods”, 2020 International Conference on Engineering Management of Communication and Technology (EMCTECH), pp. 1-6, 2020.
- [14] Б.С. Овчарова, Е.Н.. Барашко, "Технологии использования жестов для управления компьютерными и робототехническими системами" Форум молодых ученых, № (28), pp. 578-584, 2018.
- [15] Е.В. Прозорова, “Российский жестовый язык как предмет лингвистического исследования” опросы языкознания, №1, 44–61, 2007.
- [16] K.Sh. Gurbanova, “Gesture language: history, development stages and current state”, Problems of Information Society, №1, 94–99, 2018.
- [17] A.V. Bauer, “Real-time gesture recognition implementation. Microsoft Technologies in the Theory and Practice of Programming: Proceedings of the”, 12th All-Russian Scientific and Practical Conference of Students, Postgraduates and Young Scientist, Tomsk, March 25-26, pp. 56-58. 2015.
- [18] A.V. Bauer, “Dynamic Hand Gesture Recognition Based on Hidden Markov Models. Modern Techniques and Technologies”, Proceedings of the 10th International Scientific and Practical Conference of Students, Postgraduate Students and Young Scientists, Tomsk, April 14-18, pp. 145-146, 2014.
- [19] R.Sh. Mahmudova, K.Sh. Gurbanova, “The Opportunities of Technologies Application of Automatic Recognition of Gestures Shown by the Movement of Hands”, Problems of Information Technology, №2, pp.102–110, 2020.
- [20] F.J.Abdullayeva, , K.Sh. Gurbanova, “Sign Language Hand Gesture Recognition Method based on Machine Learning”, The 16th IEEE International Conference on Application of Information and Communication Technologies, Washington DC, 12-14 october, pp. 1-5. 2022.

İnternet mühitində Azərbaycan dilində vulqar sözlər və onlarla mübarizə texnologiyaları

Fərqanə Abdullayeva¹, Sabirə Ocaqverdiyeva²

^{1,2}İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

¹a_farqana@mail.ru, ²allahverdiyevabasira@gmail.com

Xülasə — Onlayn mühitdə qeyri-etik söhbətlərin, jarqon ifadələrin və bir sözlə vulqarizmlərin sayı sürətlə artır. İnternetdə senzura olmadığından istifadəçilərin belə kontentlərlə rastlaşma ehtimalı yüksək olur. Məqalədə vulqarizmlərlə qarşılaşma zamanı yaranmış problemlər şərh edilir. Vulqarizmlərə məhdudiyətlərin tətbiq edilməsinə dair rəsmi sənədlər araşdırılır. Azərbaycandilli mətnlərdə vulqarizmlərin aşkarlanması üçün maşın təliminə əsaslanan metod təklif edilir.

Açar sözlər — vulqarizmlər; uşaq təhlükəsizliyi; qeyri-etik danışq.

I. GİRİŞ

Rəqəmsal texnologiyaların inkişafı, İnternet xidmətlərinin sayının artması insanların sosial münasibətlərində yeni mərhələnin başlanmasına səbəb oldu. İnternet mühitinin imkanları istifadəçilərə fikir və rəylərini deməkdə sərbəstlik yaratmışdır. Artıq məkanından asılı olmayaraq insanlar sosial media platformalarından istifadə etməklə qarşılıqlı ünsiyyət qurur və müxtəlif paylaşımlar edirlər [1]. Son 10 il ərzində onlayn forum və sosial şəbəkə istifadəçilərinin sayının sürətlə artdığı nəzərə çarpır. Belə ki, *Facebook* sosial şəbəkəsi istifadəçiləri

tərəfindən 1 dəqiqə ərzində 510000 şərh generasiya edilir, 293000 status yeniləməsi və 136000 şəkil yüklənir. *Twitter* sosial şəbəkəsində isə həmin zaman müddətində yaradılan tvitlərin sayı 350.000 çatır. İnternet mühitində forumlarda və ya sosial şəbəkələrdə əlaqə quran insanların fikir ayrılıqlarına səbəb olan bir neçə faktoru qeyd etmək olar [2].

- Müxtəlif mədəniyyətlərə malik olması;
- Təhsil səviyyələrinin fərqlənməsi;
- Dini ayrılıqlar;
- İrqiçilik;
- İnsanların düşüncə tərzindəki fərqlilik;
- Adət-ənənələrin müxtəlif olması;
- Veb üzərindən senzurasız söz azadlığı.

Bu faktorlar onu göstərir ki, sosial şəbəkələrdə, forumlarda və digər rəy mənbələrində istifadəçilər tərəfindən daxil edilən nifrət dolu sözlərin, söyüş ifadələrinin, terror, qəddarlıq və s. məzmunlu fikirlərin artım dinamikası yüksəlməkdədir. Ekspertlərin fikrincə irqçilik, homofobiyaya meyilli şəxslər fikirlərini virtual məkanda yayır və öz qurbanlarını zorakılığa sövq etmək üçün boşluqlar tapırlar.

Onlayn mühitdə istifadə olunan nifrət, etik normadan kənar danışq, leksikonu

korlayan, vulqar sözlərdən ibarət təhqiramiz mətnlərin olması istifadəçilərə, xüsusilə də uşaq auditoriyasına öz mənfi təsirini göstərir. İnternetdə qeyri-etik ifadələrlə rastlaşma uşaqların informasiya cəmiyyətinin nümunəvi üzvü kimi yetişməsində düzgün tərbiyə olunmasında əhəmiyyət kəsb edir.

Bu tədqiqatda İnternet mühitində vulqarizmlərlə qarşılaşma zamanı yaranmış problemlər, beynəlxalq aləmdə qəbul edilmiş rəsmi sənədlər şərh edilir. Uşaqların İnternetdən istifadəsi zamanı informasiya təhlükəsizliyi baxımından kifayət qədər mühüm əhəmiyyət kəsb etdiyi və vulqarizmlə mübarizə yolları göstərilir. Azərbaycan dilində olan vulqarizmlərin məndən çıxarılmasını həyata keçirən metod təklif edilir.

II. VULQARİZMLƏRİN YARATDIĞI PROBLEMLƏR

“*Vulqarizm*” latın sözü olub “vulgaris”, yəni, “sadə xalq” mənası verən, ədəbi dil normalarına zidd olan, qaba (qeyri-ədəbi) söz, yaxud ifadələri bildirmək üçün istifadə olunan beynəlxalq termdir [3,4].

Vulqarizmlər ifadə formalarına görə fərqlənilir. Onlar mətn, video, audeo, simvollar və jestlər kimi təsvir edilir. İnsanların bir sıra mənfi xüsusiyyətlərini ifadə edən, şəxsiyyəti alçaldan ifadələr (söyüş, təhqir, qarğış, nifrət dolu ifadələr, jarqon) *vulqar leksikaya* aid edilir [3,4].

İnternetdə yayılan məlumatlar tənzimlənmədiyindən vulqar leksikadan istifadə sosial media platformalarında müxtəlif problemlərin yaranmasına səbəb olur. İnternetdə anonimliyin olması və

senzuranın olmaması insanlara özlərini daha çox olduqları kimi ifadə etməyə imkan yaradır. Çox vaxt istifadəçilərin bir-birinə olan qeyri-etik davranışı cəmiyyət, dövlət, vətəndaşlar üçün ideoloji və psixoloji problemlər yaradır. Fikir və ifadə azadlığı, ayrıseçkilik və ayrıseçkiliyin təbliği və ya təhrik edilməsi, düşmənçilik və ya zorakılıq kimi bir neçə hüquqi problemi ehtiva edir. İnsanların psixologiyasını və əhval-ruhiyyəsini korlayır və hətta depressiyaya səbəb olan halların baş verməsinə gətirib çıxarır.

Vulqarlıq çox vaxt təhqiramiz şərhələrlə əlaqəli olduğundan, sosial media məzmununda vulqarizmlərə daha çox rast gəlmək olur. Sosial media platformaları insanlara qarşı kiberzorakılıq edən şəxslərə öz hərəkətlərini ictimailəşdirməkdə geniş imkanlar verir və onların “qurbanları” aşağıdakı risklərə daha həssasdırlar:

- İstifadəçilər arasında kiberqarşıdurmalar yaranır;
- İnsanların özünə qarşı olan hörməti aşağı düşür;
- Psixoloji xəstəlikləri olan insanlara mənfi təsir edir;
- Virtual mühitdə işlədilən nifrət nitqi, söyüş və s. ifadələri real həyatda insidentlər yaradır;
- İntihar hallarına səbəb olur.

Dünyada bir çox dillərin (alban, ərəb, benqal, Çin, rus, ingilis, eston, fransız, alman, yunan və s.) dil korpusunda vulqarizmlərdən ibarət söz bazaları mövcuddur. Bu bazaların özləri də müəyyən qruplara: aqressiv məzmun, rasizm, söyüş, təhqir və s. bölünür [5].

III. VULQARİZMLƏRƏ MƏHDUDİYYƏTLƏR QOYAN RƏSMİ SƏNƏDLƏR

İnternet mühitində vulqar ifadələr işlənməsinə dair rəsmi sənədlər hazırlanmışdır və onlardan bəziləri haqqında aşağıda məlumat verilir.

Birləşmiş Millətlər Təşkilatı və bir çox digər aktorlar nifrət nitqinə qarşı mübarizə yollarını araşdırırlar. Bunlara ifadə azadlığı hüququnu təmin etməklə yanaşı, onlayn istifadəçilər arasında daha çox media və informasiya savadlılığının təşviq edilməsi təşəbbüsləri daxildir [6].

2005-ci ildə Tunisdə keçirilmiş informasiya cəmiyyəti üzrə Ümumdünya Sammitinin Fəaliyyət Planının 10-cu bəndində informasiya cəmiyyətinin etik aspektləri barədə qeyd edilir [7].

Avropa Şurasının “Uşaqların zərərli kontentdən qorunması” hesabatında uşaqlar üçün nəzərdə tutulmuş ziyanlı informasiyanın növlərindən biri kimi “Senzuradan kənar ifadələr (söyüş, jarqon, təhqiramiz ifadələr və s.)” göstərilir [8].

Avropa Şurasının Nazirlər Komitəsinin 97 (2) sayılı tövsiyəsində isə aşağıdakı kimi qeyd edilir: “...təcavüzkar millətçilik və etnosentrizmlə ifadə edilən dözümsüzlük, ayrışdırıcı yayan, təhrik edən, təşviq edən və ya haqq qazandıran bütün ifadə formalarını və azlıqlara, miqrantlara və immiqrant mənşəli insanlara qarşı düşmənçilik kimi başa düşülür [9].

Avropa İttifaqına daxil olan dövlətlər və qurumlar bu yeni realıqla mübarizəyə qoşulur. Məsələn, 2017-ci ildə Almaniyada təhqir, təhrik və dini ayrılıqların təbliği kimi qeyri-qanuni onlayn çıxışlara qarşı mübarizəni nəzərdə tutan Şəbəkə Mühafizə

Aktı (NetzDG) qəbul edilmişdir. Bu akta əsasən minimum 2 milyon istifadəçisi olan sosial media platformaları tərəfindən paylaşılan qeyri-qanuni məzmunun, o cümlədən nifrəti və dini təhqirləri ifadə edən söz və ifadələrin 24 saat ərzində silinməsinə və ya 50 milyon avroya qədər yüksək cərimənin ödənilməsinə tələb edir. Avropa Birliyi Komissiyası sosial media şəbəkələrinin qarşısında sərt tələblər qoyur və onları bu kodeksi imzalamağa məcbur edir [10].

Beynəlxalq aləmdə qəbul edilmiş, yuxarıda qeyd olunan bu və ya digər sənədlərin, cərimə və hüquqi tənzimləmələrin olmasına baxmayaraq vulqar fikir və ifadələrin virtual mühitdə yaratdığı problemlər çoxsaylıdır. Belə hallar həssas auditoriyaya təsir etməklə xoşagəlməz izlər buraxır və yeni texnologiyaların köməyi ilə qarşılıqlı tədbirlərinin görülməsini motivasiya edir. Bunu nəzərə alaraq növbəti bölmədə vulqar sözlərin aşkarlama üsulları haqqında məlumat verilir.

IV. MƏTNLƏRDƏ VULQARİZMLƏRİN AVTOMATİK AŞKARLANMASI TEXNİKASI

Mətn təhlilində, eləcə də vulqar sözlərin mətnlərdən avtomatik olaraq aşkarlanmasında müxtəlif üsullardan istifadə edilir. Aşağıda bunlardan bəziləri haqqında məlumat verilir [11]:

- *Lüğətə əsaslanan metod (dictionary-based methods)* – adından da göründüyü kimi, mətn təhlilində bu metoddan istifadə lüğətlərə, söz xəzinələrinə və leksik bilik bazasına əsaslanır.
- *Söz çantası (ing. Bag-of-words, BOW)*. – söz emalında istifadə olunan modellərdən

biri olmaqla, mətnə daxil olan nizamlanmamış sözlər toplusudur və sözlərin təbii dildən kompüterə uyğun olaraq rəqəm dilinə çevrilməsini həyata keçirir.

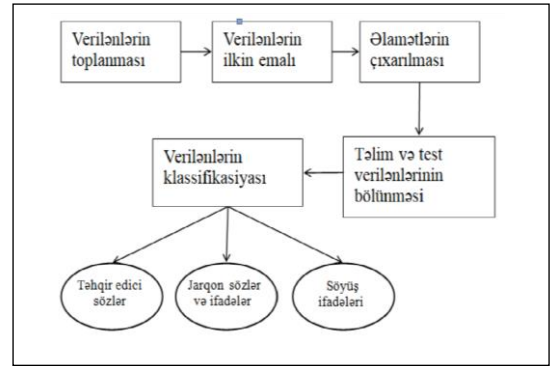
- *N-gram* - təbii dilin emalında yalnız sözlərin mənaları deyil, eyni zamanda sözün işlənmə konteksti də əhəmiyyətlidir. Mətnə və ya danışmada sözlərin mənasını aydın şəkildə anlamaq üçün dili kontekst baxımından başa düşməyin yollarından biridir.
- *Tf-idf* – Söz sırasında və ya söz korpusundakı bir sözün mətnə nə qədər uyğun olduğunu müəyyən edir.
- *Təbii dilin emalı (Natural Language Processing – NLP)* – kompüter vasitəsilə təbii dildə şərh edilən məzmunu oxumaq, anlamaq və anamlandırmaq xüsusiyyətlərinə malik olan süni intellekt sahəsidir.
- *Word2vec* – müəyyən söz korpusunda sözlərin paylanmış təsvirlərini təmsil etmək üçün istifadə edilən modellərin birləşməsidir. *Word2Vec (W2V)* mətn korpusunu giriş kimi qəbul edərək, hər bir sözün vektor təsvirini çıxaran alqoritmdir. Böyük ölçülü verilənlərin emalında daha effektiv nəticə göstərir.

Müsiq metodlara əsaslanan avtomatlaşdırılmış aşkarlama vasitələrindən istifadə etməklə, *Twitter*, *Facebook*, *Instagram* və *YouTube* kimi platformalarda getdikcə daha çox işarələnmiş və ya silinmiş məzmun barədə məlumat verilir. Qeyd edilir ki, 2021-ci ilin yanvar-mart ayları ərzində *YouTube* nifrət nitqi siyasətini pozan 85 247 videonu, *Facebook*, ümumilikdə 25,2 milyon, *Instagram* isə 6,3 milyon məzmunun işləndiyini bildirdi. *Twitter*-in

hesabatında isə 2020-ci ilin iyul və dekabr ayları ərzində nifrət nitqi siyasətinin pozulduğu 1628281 məzmunun silindişi qeyd edilir [12].

V. TƏKLİF OLUNAN METOD

Məqalədə uşaqların İnternetdən istifadəsi zamanı, onların qeyri-etik söz və ifadələrlə rastlaşmasının qarşısının alınması üçün metod təklif edilir. Maşın təlimi əsasında təklif olunmuş bu metod Azərbaycan dilində olan vulqar sözlərin mətnə çıxarılmasını həyata keçirir. Şəkil 1.-də təklif edilmiş arxitektura təsvir edilir [11].



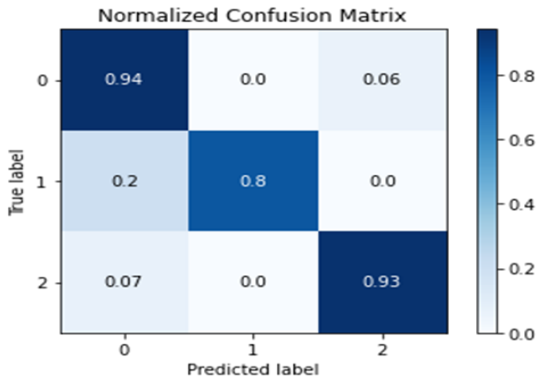
Şəkil 1. Azərbaycan dilində vulqarizmlərin aşkarlanması sisteminin arxitekturası

Azərbaycan dilində vulqar söz bazası olmadığı səbəbindən ideyanın effektivliyinin yoxlanması üçün eksperimentlərin aparılmasında müəlliflər tərəfindən “Vulqarizmlər_Az” kimi adlandırılan lüğət tərtib edilmişdir. Tərkibi 249 sözdən ibarət olmaqla və 3 sinifdə (təhqir edici sözlər, jarqon sözlər və ifadələr, söyüş ifadələri) təsnifatlaşdırılır. Mətnlərin emalı zamanı ilk növbədə *tokenləşdirmə* və *stemming* işləri yerinə yetirilir.

Tokenləşdirmə – strukturlaşdırılmamış məlumatları və təbii dildə olan mətni diskret elementlər kimi qəbul etməklə sözlərə və ya hissələrə bölür. Bu zaman paraqrafları cümlələrə, cümlələri sözlərə və sözləri simvollara bölmək olar [11].

Stemming – sözlərin kök halına gətirilməsi, yaxud da normallaşdırılmasıdır. Bu axtarışa qənaət etmək üçün cümlədəki sözlər toplusunun ardıcılığa çevrildiyi bir üsuldur. [13].

Eksperimənlər Python proqramlaşdırma mühitində və N-gram+TF-IDF əlamətləri əsasında aparılır. Klassifikasiyanın effektivliyi Accuracy, Precision, Recall, F-Measure metrikaları əsasında qiymətləndirilmişdir. GNB alqoritmi unigrams əlaməti ilə müqayisədə daha üstün nəticə göstərmişdir. Şəkil 2.-də təsvir edilən xətalər matrisinə əsasən demək olar ki, eksperimentlər effektiv nəticə göstərmişdir və bizim təklif etdiyimiz metod N-gram+TF-IDF əlamətlərinə görə hər üç sinfdən olan vulqarizmləri 0.94 dəqiqliklə tanıyırdır [11].



Şəkil 2. N-gram+TF-IDF əlamətlərinə görə GNB alqoritminin klassifikasiyası nəticələrinin xətalər matrisi

NƏTİCƏ

Onlayn mühitdə dil üslubu qeyri-rəsmidir, istifadəçilər də bundan “yararlanırlar”, vulqarizmlərdən isə sərbəst şəkildə istifadə edilir. Demək olar ki, onlayn mühitdə vulqar sözlərin istifadəsi mahiyyətə oflayn mühitdən fərqlənir və onlar baş verdiyi qarşılıqlı təsirlərin xarakterinə, daha sürətlə paylaşımalar edildiyinə görə və digər səbəblərdən az bir zamanda daha çox auditoriyaya yayıla bilər.

Bütün dillərdə vulqar söz və ifadələrdən istifadə edilir. Hər bir dilin özünəməxsus xüsusiyyətləri olduğundan bir dildə olan vulqarizmlərin digərləri ilə ciddi mənə fərqinin olduğu nəzərə çarpır. Azərbaycan dilində də vulqarizmlər bazasının yaradılması vacibdir və müasir dövrün tələbidir. Belə bir bazanın yaradılmasında məlum reallıqlar nəzərə alınmalıdır. Problemlərin həllində müəyyən beynəlxalq səviyyədə təşəbbüs irəli sürülməlidir. Bu verilənlərin digər dillərdə olan vulqar sözlər bazası ilə qarşılıqlı inteqrasiyası təmin edilməlidir. Həmçinin əlçatanlığı təmin edilməlidir ki, istifadəçilər, sadə vətəndaşlar və xüsusilə tədqiqatçılar üçün asan olsun. Vulqarizmlərlə mübarizə məqsədilə işlənən intellektual texnologiyalar milli tələbləri nəzərə alınmalıdır.

İSTİNADLAR

- [1] Alguliyev R.M., Ojagverdieva S.S. Conceptual Model of National Intellectual System for Children Safety in Internet Environmen // International Journal of Computer Network and Information Security, 2019, vol.11, no.3, pp. 40–47.
- [2] Wild and Interesting Facebook Statistics and Facts (2023), <https://kinsta.com/blog/facebook-statistics/>

- [3] Farajov R.A. İzahlı dilçilik lughati. Bakı, “Maarif”, 1969, 143 s. https://ec.europa.eu/commission/presscorner/detail/en/IP_16_1937
- [4] Aliguliyev R.M., Jafarov Y. Global vulgarizm bazası / İnformasiya təhlükəsizliyi problemləri üzrə I respublika elmi-praktiki konfransı, 2013, s. 60–62.
- [5] Aarts J., & Meijs, W. (2022). Corpus linguistics: recent developments in the use of computer corpora in English language research.
- [6] Understanding hate speech, <https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech>
- [7] World Summit on the Information Society (WSIS), <http://www.itu.int/wsis/docs/geneva/official/poa.html#c10>
- [8] European Commission, “Protection of personal data”, November 24, 2016. <http://ec.europa.eu/justice/data-protection/>
- [9] European Commission and IT Companies announce Code of Conduct on illegal online hate speech.
- [10] Germany: Network Enforcement Act Amended to Better Fight Online Hate Speech, <https://www.loc.gov/item/global-legal-monitor/2021-07-06/germany-network-enforcement-act-amended-to-better-fight-online-hate-speech/>
- [11] Abdullayeva F.C., Ocaqverdiyeva S.S. Vulqarizmlərin maşın təlimi əsasında aşkarlanmasına bir yanaşma// İnformasiya Texnologiyaları Problemləri, 2021, N: 2. s. 89–98.
- [12] Corporate author: United Nations Office on Genocide Prevention and the Responsibility to Protect, Addressing hate speech on social media: contemporary challenges, 2021, 10 pp.
- [13] Rizki A. S., Tjahyanto, A., Trialih, R. Comparison of stemming algorithms on Indonesian text processing. Telecommunication Computing Electronics and Control, 2019, 17(1), pp. 95-10.

Dil sənayesi biliklər iqtisadiyyatının tərkib hissəsi kimi

Rasim Mahmudov

İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

rasimmahmudov@gmail.com

Xülasə — Tədqiqat işində dil sənayesi ilə biliklər iqtisadiyyatının mahiyyəti, qarşılıqlı əlaqəsi, oxşar xüsusiyyətləri analiz edilir. Dil sənayesinin əhatə dairəsi müəyyən edilir, əsas seqmentləri təsnifatlandırılır. Dil sənayesinin biliklər iqtisadiyyatının bir sahəsi olduğu əsaslandırılır. Dil sənayesində İKT-nin tətbiqinin xüsusiyyətləri göstərilir. Eləcə də, dil sənayesinin gələcək perspektivlərinə dair fikirlər irəli sürülür.

Açar sözlər — *biliklər iqtisadiyyatı; dil sənayesi; lokallaşdırma; maşın tərcüməsi.*

I. GİRİŞ

Sürətlə qloballaşan müasir dünyada insanlar arasında ünsiyyət qurmaq, zəruri bilik və informasiyanı əldə etmək və ötürmək üçün linqvistik vasitəçilərə böyük ehtiyac vardır. Dünyanın müxtəlif ölkələrindən olan gənclər qabaqcıl ali məktəblərdə təhsil almaq, bacarıqlı mütəxəssis kimi yetişmək, əmək bazarında rəqabətə davamlı olmaq, uğurlu karyera qurmaq istəyirlər.

Transmilli şirkətlər öz bazarlarını daim genişləndirmək, çoxdilli bazarda fəaliyyət göstərmək, bütün dil seqmentlərini əhatə etmək iddiasındadırlar. Dünyada turizm sektoru sürətlə inkişaf edir, qlobal nəqliyyat-kommunikasiya infrastrukturunun inkişafı, insanların gəlirlərinin artması bu sahənin inkişafını sürətləndirir. İnternetin

köməyi ilə insanlar ən müxtəlif bilik və informasiya mənbələrinə daxil olaraq öz tələbatlarını dolğun şəkildə ödəmək niyyətindədirlər.

Qeyd edilən zərurət ünsiyyət vasitəçiliyinə - tərcüməçilik fəaliyyətinə, həmçinin dünyanın aparıcı dillərinin öyrənilməsinə tələbatı getdikcə daha da artırır. İnsanların bu tələbatını ödəmək üçün dünyada yeni intellektual xidmət seqmenti yaranır. Bu intellektual fəaliyyət sahəsi biliklər iqtisadiyyatının perspektivli istiqaməti olan dil sənayesini formalaşdırır.

Hazırda dil sənayesinin sürətli inkişafının əsasında həm də bu sahədə İKT-nin uğurlu tətbiqi dayanır. Bu sferada tədris, tərcümə, lokallaşdırma və digər xidmətlərin göstərilməsi, eyni zamanda, qlobal informasiya cəmiyyətinin formalaşmasına və inkişafına xidmət edir.

II. BİLİKLƏR İQTİSADIYYATININ MAHİYYƏTİ VƏ XÜSUSİYYƏTLƏRİ

“Biliklər iqtisadiyyatı” termini elmi leksikona 1962-ci ildə amerikalı alim F.Maxlup tərəfindən daxil edilib [1]. O, özünün “ABŞ-da biliklərin istehsalı və yayılması” əsərində ilk dəfə bu termindən istifadə edib. F.Maxlupa görə, bilik istehsalı elə bir prosesdir ki, onun vasitəsilə kimsə ona məlum olmayan nəyisə öyrənir. Onun

fikrincə, biliklərin istehsalında və yayılmasında resept yazan həkim də, məsləhət verən hüquqşünas da iştirak edir.

İqtisadi Əməkdaşlıq və İnkişaf Təşkilatının (Organisation for Economic Cooperation and Development, OECD) verdiyi tərifə görə, biliklər iqtisadiyyatı – iqtisadi və sosial inkişafın daha effektiv şəkildə sürətlənməsi üçün kodlaşdırılmış və qeyri-aşkar biliklərin əldə edilməsi, yaradılması, yayılması və istifadəsidir [2].

Ümumiyyətlə, biliklər iqtisadiyyatı iqtisadiyyatın elə bir sahəsidir ki, burada bilik və insan kapitalı əsas rola malikdir. İqtisadiyyatın bu sahəsinin inkişafı insan kapitalının keyfiyyətinin, həyat səviyyəsinin artırılması, elmtutumlu məhsulların, yüksək texnologiyaların istehsalının, innovasiyaların genişləndirilməsi ilə bağlıdır.

Başqa sözlə, biliklər iqtisadiyyatı – qeyri-maddi aktivlərin, insan kapitalının, elm və təhsilin, İKT-nin, intellektual xidmətlərin xüsusi çəkisinin yüksək olduğu sahədir. Bu sahənin əsas istehsal məhsulları bilik və informasiyadır.

Mütəxəssislərin fikrincə, biliklər iqtisadiyyatının mahiyyəti aşağıdakı qarşılıqlı əlaqədə olan proseslərlə müəyyən edilir [3-5]:

- informasiya və biliklər sosial-iqtisadi, texnoloji və mədəni inkişafın mühüm resursu və hərəkətverici qüvvəsinə çevrilir;

- təbii resurslar, əmək və kapital ilə birlikdə informasiya və biliklər də istehsal amillərinə çevrilir;
- informasiya və biliklərin yaradılması, ötürülməsi və istifadə edilməsini təmin edən sahələrin xüsusi çəkisi sürətlə artır;
- inkişaf etmiş informasiya infrastrukturunu beynəlxalq rəqabət qabiliyyətini təmin edən amilə çevrilir.

Biliklər iqtisadiyyatı, həmçinin insan kapitalının maddi kapitalla müqayisədə xüsusi çəkisinin daha artıq olduğu iqtisadi sahələrin məcmusu kimi xarakterizə olunur. Həmin sahələrdə biliklərin texnoloji reallaşdırılması həlledici rol oynayır, biliklərin istehsalı isə iqtisadi inkişafın mənbəyi kimi çıxış edir. Göstərilən yanaşmada biliklər iqtisadiyyatı elmi fəaliyyəti, təhsili, İKT-ni, innovasiyaları, intellektual xidmətləri əhatə edir.

Müvafiq baxışları, təsnifatları ümumiləşdirsək, biliklər iqtisadiyyatının əsas seqmentlərini bu cür müəyyən etmək olar [6-8]: elmi-tədqiqat; təhsil; İKT; KİV; informasiya texnologiyaları; nəşriyyat və reklam işi; bədii yaradıcılıq; şou-biznes.

III. DİL SƏNAYESİNİN BİLİKLƏR İQTİSADİYYATINA XAS OLAN XÜSUSİYYƏTLƏRİ

Biliklər iqtisadiyyatı ilə bağlı yuxarıda qeyd edilənləri nəzərə alsaq, dil xidmətlərini də intellektual bir fəaliyyət istiqaməti olaraq informasiya və bilik istehsalı sahəsinə aid etmək olar [9]. Belə

ki, dil xidmətlərinin aparıcı sahəsi olan tərcümə xidmətləri sayəsində insanlar hansısa kitabı, məqaləni, xəbəri oxuyaraq yeni informasiya və bilik əldə etmək imkanı qazanır. Yaxud dublyaj edilmiş filmləri izləyən insan dünyagörüşünü, biliklərini artırır. Həmçinin müvafiq tədris prosesini keçən yeni bir dil öyrənir, həmin dildə yazılan mətnlərlə, səsləndirilən audio-video materiallarla tanış ola bilir, bununla da, yeni biliklər əldə edir. Eləcə də, lokallaşdırma xidmətləri hesabına insanlar İnternet resurslarından səmərəli istifadə edə bilirlər.

Bir sıra iqtisadçı alimlər hesab edirlər ki, biliklər iqtisadiyyatı qloballaşma xarakterinə malikdir [9]. Onlar bunun üçün kompüterlərin, İnternetin, mobil telefonların meydana gəlməsi ilə iqtisadi sistemin qloballaşması proseslərinin sürətlənməsini əsas gətirirlər.

Bu aspektdən yanaşsaq, dil xidmətləri də məhz iqtisadi və mədəni sahələrdə qloballaşmaya təkan verir. Belə ki, istənilən sahədə qloballaşmanı əngəlləyən səbəblərdən biri də dil baryerinin olmasıdır. Məhz dil xidmətləri sahəsində transmilli iqtisadi subyektlər dil baryerlərini aşaraq qlobal bazarın milli seqmentlərinə nüfuz edə bilirlər. Bu, həm lokallaşdırma və tərcümə xidmətləri, həm də müvafiq mütəxəssislərin zəruri dil biliklərinə yiyələnməsi hesabına mümkün olur [8].

Qeyd etmək lazımdır ki, biliklər iqtisadiyyatı, ilk növbədə, İKT-yə əsaslanan innovativ iqtisadiyyatdır [9]. İKT yeni biliklərin yaradılması və qlobal miqyasda

istifadə imkanlarını əhəmiyyətli dərəcədə artırır. Bu aspektdən də yanaşsaq, dil sənayesinin inkişafı və effektivliyinin artması da məhz İKT-nin tətbiqi əsasında baş verir. Belə ki, hazırda dünya üzrə tərcümə xidmətlərinin əsas hissəsini İKT-nin nailiyyətlərinə əsaslanan maşın tərcüməsi təşkil edir. Həmçinin xarici dilin tədrisində İKT-nin imkanları uğurla tətbiq edilir.

Məlum olduğu kimi, biliklər iqtisadiyyatının aparıcı sahələrindən biri təhsil xidmətləridir. Bu gün dünyada təhsil sahəsində ən çox tələbat duyulan sahə müxtəlif dil biliklərinin əldə edilməsilə bağlıdır. İnsanlar digər sahələr üzrə bilikləri, əsasən, orta və ali təhsil sistemlərində əldə edirlər. Lakin dil biliklərinin əldə edilməsi üçün dünyanın istənilən ölkəsində müvafiq təhsil sistemlərindən kənarında xüsusi tədris mərkəzləri fəaliyyət göstərir. Həmin xidmətlər milli və beynəlxalq səviyyələrdə təhsil xidmətləri bazarının əsas hissəsini təşkil edir.

Qeyd etdiyimiz kimi, biliklər iqtisadiyyatının bir istiqaməti də nəşriyyat fəaliyyətidir. Belə ki, nəşriyyat xidmətləri informasiya və bilik daşıyıcıları olan müxtəlif təyinatlı qəzet, jurnal, kitab və s. çap edilməsidir. Nəşriyyat fəaliyyətinin əsas işlərindən biri məhz dil xidmətləri sırasına daxil olan redaktə və korrektə işləridir. Xüsusilə, elmi nəşriyyatlar dil, tərcümə və üslub redaktəsi işlərindən böyük həcmdə gəlir əldə edirlər. Nəzərə almaq lazımdır ki, müxtəlif elm sahələrində

çalışan alimlər dilçilik üzrə mütəxəssis deyillər. Bir çox hallarda onların yazdıqları elmi əsərlərin peşəkarlar tərəfindən tərcüməsinə və redaktəsinə ehtiyac duyulur.

Məsələn, dünyanın ən nüfuzlu elmi nəşrlərindən olan *Elsevier*-də mətnin həcmindən asılı olaraq məqalələrin redaktə işlərinin qiyməti 100-500 dollar arasında dəyişir [10]. Təqdim olunan məqalələrin bir çoxu dil və üslub tələblərinə cavab vermədiyinə görə ekspertizaya təqdim edilmək üçün, ilk növbədə, onların üzərində ödənişli əsaslarla müvafiq redaktə işləri aparılır.

Apardığımız müqayisəli analizlər göstərir ki, biliklər iqtisadiyyatının xarakterik xüsusiyyətlərini, eyni zamanda, dil sənayesinə də aid etmək olar. Bütün qeyd edilənləri nəzərə alaraq, dil xidmətlərini biliklər iqtisadiyyatının bir seqmenti hesab etmək olar.

IV. DİL SƏNAYESİNİN ƏSAS SEQMENTLƏRİ

Dil sənayesi bir-biri ilə bu və ya digər dərəcədə bağlı olan müxtəlif intellektual xidmət sahələrini özündə birləşdirir. Bu xidmətlərin hər birinin spesifik xüsusiyyətləri olsa da, onları birləşdirən bir əsas ümumi cəhət var ki, bu da insan amilinin həlledici rola malik olmasıdır. Yəni bu sahələrin hər birində göstərilən xidmətlərin keyfiyyəti, ilk növbədə, hansısa dili mükəmməl bilən mütəxəssislərdən asılıdır [9].

Eyni zamanda, bu xidmət seqmentləri arasında qarşılıqlı əlaqə və asılılıq mövcuddur. Belə ki, biri digərinin

fəaliyyətini təmin edir, yaxud stimullaşdırır. Məsələn, dil sənayesinin təhsil seqmenti digər seqmentlər üçün kadr hazırlığını həyata keçirir. Təhsil seqmentinin keyfiyyəti isə dil sənayesi üzrə elmi-tədqiqatların səviyyəsindən asılıdır. Tərcümə fəaliyyəti isə elmi-tədqiqat və təhsil seqmentləri üçün zəruri kontentin əlyətərliliyini təmin edir.

Bütövlükdə, dil sənayesinin aşağıdakı seqmentlərini fərqləndirmək olar (Şəkil 4) [11-13]:

- elmi-tədqiqat - dil sənayesini elmi-nəzəri və metodoloji biliklərlə təmin edir;
- təhsil - dil sənayesi üçün insan resursları (pedaqoqlar, poliqlotlar, tərcüməçilər, proqramçılar və s.) hazırlanır;
- tərcümə fəaliyyəti - müxtəlif xarakterli (elmi, pedaqoji, bədii, hüquqi, texniki, iqtisadi və s.) mətnlərin yazılı, müxtəlif toplanılarda, görüşlərdə çıxışların sinxron tərcüməsi həyata keçirilir;
- linqvistik informatika – maşın tərcüməsi, kompüter linqvistikası, terminoloji bazaların yaradılması, bulud xidmətləri və s. sahələri əhatə edir;
- lokallaşdırma fəaliyyəti - məhsul və xidmətlərin çoxdillilə bazarada istifadə olunması üçün onların yerli tələblərə adaptasiyası həyata keçirilir, yəni müşayiətedici kontent (o cümlədən veb-saytlar) hədəf auditoriyasına uyğun tərcümə edilir;

- nəşriyyat fəaliyyəti – redaktə işləri (qrammatika, üslub, orfoqrafiya, terminlərin unifikasiyası, oxunaqlıq və s.) yerinə yetirilir;
- turizm – bələdçi-tərcüməçi və informasiya xidmətləri göstərilir;
- kino, televiziya, əyləncə sənayesi – kinofilmlərin, televiziya verilişlərinin, oyun proqramlarının, digər audio-video materialların tərcüməsi və səsləndirilməsi həyata keçirilir.

V. DİL SƏNAYESİNDƏ İKT-NİN TƏTBİQİNİN PERSPEKTİVLƏRİ

Məlum olduğu kimi, biliklər iqtisadiyyatının aparıcı sahələrindən biri məhz İKT sektorudur. İKT sektoru yüksək səviyyəli intellektual və innovativ biliklər tələb edən bir sahədir. Həmin sektor məhz bu cür biliklərə malik olan kadrların hesabına sürətlə inkişaf edir və iqtisadiyyatın ən mühüm istiqamətlərindən birinə çevrilir.

Bütün sahələrdə olduğu kimi, dil sənayesində də İKT-nin imkanlarından geniş istifadə edilir. Bu gün dünyada İKT-nin xüsusi sahəsi olan kompüter linqvistikası müvafiq elmi-tədqiqat mərkəzlərinin əsas fəaliyyət istiqamətlərindən biridir. Bu sahədə intensiv elmi tədqiqat fəaliyyətini stimullaşdıran əsas faktor məhz kompüter linqvistikasının biznes perspektivləridir. Kompüter linqvistikasının əsas məhsulları olan maşın tərcümə sistemlərinə dil xidmətləri bazarında böyük tələbat vardır.

Hətta dil xidmətləri üzrə ixtisaslaşan dünyanın bir sıra transmilli şirkətləri müvafiq proqram təminatlarının istehsalı və tətbiqi ilə məşğul olurlar. Bütövlükdə, kompüter linqvistikası dil sənayesinin tərkib hissəsi kimi güclü iqtisadi potensiala malikdir.

Statistikaya görə, dünyada İnternet istifadəçilərinin sayı artıq 5,4 milyard nəfərə yaxındır [14]. Telekommunikasiya infrastrukturunun getdikcə genişlənməsi, smartfonların və digər mobil qurğuların yayılması İnternet istifadəçilərinin sürətlə artmasını stimullaşdırır ki, bu da, öz növbəsində, maşın tərcüməsinə olan tələbatın artmasına gətirib çıxarır.

Hazırda dil sənayesinin ən gəlirli və ən sürətlə inkişaf edən seqmenti tərcümə xidmətləridir [9]. Tədqiqatlar göstərir ki, tərcümə xidmətlərinin çoxu kiçik firmalar və frilanserlər tərəfindən həyata keçirilir. Tərcümə xidmətlərinin yazılı tərcümə, şifahi tərcümə, maşın tərcüməsi, lokallaşdırma, dublyaj və s. kimi növləri mövcuddur. Bu xidmətlərin bir çoxu qabaqcıl linqvistik texnologiyalar və proqram təminatları vasitəsilə dəstəklənir.

Dünyada İnternet istifadəçilərinin əksəriyyəti qlobal şəbəkədə mətnləri yalnız öz doğma dillərində oxuyub anlaya bilirlər. Bu da dünyanın müxtəlif regionlarında dil baryerinin aradan qaldırılması üçün effektiv texnoloji imkanların yaradılması işini stimullaşdırır.

Microsoft Bing Translator və *Google Translate* kimi ödənişsiz tərcümə proqramlarının əlyətərliliyi bu sahədə

rəqabət mühitini formalaşdırır, daha keyfiyyətli kommersiya xarakterli proqramların işlənilməsini stimullaşdırır [9]. Bu sahədə əsas problem tərcümənin keyfiyyətinin effektiv qiymətləndirmə mexanizmlərinin olmamasıdır.

Bütövlükdə, maşın tərcüməsinin əsas üstünlükləri kimi aşağıdakıları göstərmək olar [15, 16]:

- maşın tərcüməsi daha sürətli, asan və qənaətcildir;
- müasir tərcümə texnologiyaları mətnləri minimal səhvlərlə çevirməyə imkan verir;
- proqnozlara görə, SaaS (proqram təminatı xidməti kimi) bulud xidmətlərindən istifadə maşın tərcüməsinə tələbatın artmasına, müvafiq bazarın inkişafına ciddi təkan verəcək;
- Genişzolaqlı İnternetin və smartfonların geniş yayılması da maşın tərcüməsinin bazar perspektivlərini artırır;
- Ödənişsiz maşın tərcümə xidmətləri müvafiq kommersiya şirkətləri üçün ciddi rəqabət mühiti yaradır.

Şirkətlər tərəfindən kontentin müxtəlif dillərə tərcümə edilməsinə olan tələbat müştərilər üçün istifadə təlimatlarının hazırlanması, marketinq və kommersiya kataloqlarının tərtib edilməsi, malların təqdimatı üçün press-revizlərin hazırlanması və veb-kontentin yaradılması zərurətindən irəli gəlir.

Hazırda maşın tərcüməsinin əsas tətbiq sahələri bunlardır: hərbi sənaye, avtomobil sənayesi, elektronika, İKT və tibb. Maşın tərcüməsində hərbi sənaye sahəsinin payı daha çoxdur – təxminən 33%-dir [17]. Bu daha çox hərbi sənayedə böyük həcmli kontentin dəqiq və sürətli tərcümə edilməsinə olan tələbatdan irəli gəlir.

Dil sənayesinin vacib istiqamətlərindən biri də maşın tərcüməsi üçün yaddaş bazasının, terminoloji bazanın yaradılması, kontentin intellektual analizinin aparılması, mətnlərin tanınması və s. kimi işlərdir.

VI. DİL SƏNAYESİNİN GƏLƏCƏK PERSPEKTİVLƏRİ

Qeyd etdiyimiz kimi, dil sənayesinin inkişafına əsas tələbatı ingilis dili yaradır [9]. Bu dilin beynəlxalq arenada siyasi, iqtisadi, elmi, təhsil və mədəni sahələrdə, eləcə də, şəxsi müstəvidə əsas ünsiyyət vasitəsi kimi statusu dil sənayesinin bazar imkanlarını və perspektivlərini müəyyən edən əsas faktordur. Ona görə də qlobal dil sənayesinin perspektivləri ingilis dilinin gələcək vəziyyətindən əhəmiyyətli dərəcədə asılıdır. Burada paradoksal hal ondan ibarətdir ki, dil sənayesi, əsasən, ingilis dilini bilməyən insanların, ingilis dilinin geniş yayılmadığı ölkələrin, regionların tələbatı hesabına fəaliyyət göstərir.

Eyni zamanda, dil xidmətləri hesabına dünyada ingilisdilli insanların sayı artır, ingilisdilli ölkələrin, regionların dairəsi genişlənir. Yəni dil sənayesinin fəaliyyəti öz bazar və müştəri dairəsinin məhdudlaşmasına xidmət edir. Qlobal mühitdə ingilis dilinin gələcəyi, inkişaf

perspektivləri ilə bağlı müxtəlif yanaşmalar mövcuddur [18, 19]. Bəzi yanaşmalara görə, yaxın illərdə bu dildə sərbəst danışanların sayı 3 milyarda çatacaq. Bütün fəaliyyət sahələrində ingilis dili hakim mövqeyə yüksələcək.

Digər yanaşmalarda isə, ingilis dilinin müxtəlif variantlarının mövcud olduğuna diqqət çəkilir [20, 21]. Hesab edilir ki, yaxın gələcəkdə bu variantlar arasında fərqlər daha da artacaq, əsas rəqabət isə Britaniya və Amerika variantları arasında gedəcək. ABŞ-ın dünyadakı hərbi-siyasi və iqtisadi-texnoloji liderliyi praktiki cəhətdən ikinci variantı daha perspektivli edir.

Daha bir tendensiya ondan ibarətdir ki, ABŞ və Böyük Britaniyada əhalinin sürətlə qocalması, işçi qüvvəsinə (o cümlədən intellektual sahələrdə) getdikcə artan tələbat fonunda yaxın gələcəkdə bu ölkələrə güclü miqrant axını hesabına ingilis dili əvvəlki mövqeyini itirərək ikinci dilə çevrilə bilər.

Artıq ingilis dilinin Hinglish (hindi+ingilis), Japlish (yapon+ingilis), Chinglish (çin+ingilis), Spanglish (ispan+ingilis) və s. kimi versiyaları meydana çıxıb. Bu isə o deməkdir ki, ingilis dilinin ayrı-ayrı variantları gələcəkdə müstəqil dillərə çevrilə bilər. Məsələyə tarixi aspektdən yanaşanlar isə hesab edirlər ki, vaxtilə yunan, ərəb, rus, fransız və ispan dilləri dünyada lider mövqələrdə olduqları kimi, ingilis dili də yalnız sivilizasiyanın müasir mərhələsinin hakim dilidir. Gələcəkdə bu estafet Çin, yaxud yenidən ərəb dilinə verilə bilər.

ABŞ analitikləri hesab edir ki, ölkə vətəndaşlarının xarici dilləri öyrənmələri milli prioritetə çevrilməlidir [22]. Bu mövqə onunla əsaslandırılır ki, artıq 1975-ci ildən etibarən qlobal iqtisadiyyatda ingilisdilli dünyanın mövqələri zəifləməkdə davam edir və 2030-cu ildən Çinin iqtisadiyyatı ABŞ iqtisadiyyatından önə keçəcək. Həmçinin Latın Amerikasına (ispan və portuqal dilləri) və Cənubi Asiyaya (hindi və urdu dilləri) regionlarının iqtisadiyyatı da xeyli güclənəcək. Hesab olunur ki, ABŞ iqtisadiyyatının gələcək inkişafı daha çox ölkədə istehsal və xidmətlərin xarici istehlakçılara satılması imkanlarından asılı olacaq. Xarici bazarlarda uğur qazanmaq üçün isə müvafiq dilləri bilən mütəxəssislər indidən hazırlanmalıdır.

Hazırda dil xidmətləri göstərən şirkətlərin inkişaf konsepsiyaları məhz mövcud qlobal dil palitrasının xüsusiyyətlərinə, o cümlədən ingilis dilinin imkanlarına əsaslanır [9]. Yeni texnoloji platformaların, proqram təminatlarının hazırlanması və reallaşdırılmasında, kadr təminatının həyata keçirilməsində bu tendensiyalar nəzərə alınır. Gələcək perspektivdə dil sənayesinin daha çox tələbat duyulan dil segmenti də məhz qlobal mühitdə baş verəcək siyasi-iqtisadi proseslərdən asılı olacaq.

NƏTİCƏ

Apardığımız araşdırma göstərir ki, sürətlə qloballaşan müasir dünyada dil sənayesinin spesifik iqtisadi fəaliyyət sahəsi kimi rolu və əhəmiyyəti artmaqdadır. Maraqlıdır ki, müvafiq elmi yanaşmalarda,

adətən, qloballaşmanın əsas təkanverici qüvvəsi kimi İKT-nin sürətli inkişafı göstərilir. Burada dil faktoru, demək olar ki, nəzərə alınmır. Əslində isə, İKT qlobal informasiya və kommunikasiya mühiti üçün daha çox texnoloji infrastruktur yaradır. Əgər insanlar dil baryerinə görə ünsiyyət qura bilməsələr, informasiyadan, məqalədən, kitabdan istifadə edə bilməsələr, əcnəbi filmləri seyr edə bilməsələr İnternetin, mobil telefonların və s. texnoloji vasitələrin yaratdığı kommunikasiya mühitinin əhəmiyyəti olmazdı. Bu baxımdan, hesab etmək olar ki, informasiya və kommunikasiya mühitinin formalaşmasında dil faktoru həlledici əhəmiyyətə malikdir.

Doğrudur, indiyə kimi iqtisadiyyatda kommunikasiyaların rolu kifayət qədər araşdırılıb, lakin kommunikasiyanın effektivliyinin dil amilindən asılı olduğuna o qədər də diqqət yetirilməyib. Qlobal informasiya və kommunikasiya mühitinin formalaşması isə elm, təhsil, ticarət, turizm, KİV və s. fəaliyyət sahələri üçün yeni perspektivlər vəd edir. Dil sənayesinin inkişafını stimullaşdıran əsas amillərdən biri budur.

Bu onu göstərir ki, dil həm də mühüm iqtisadi inkişaf amilidir. Bir sıra iqtisadi fəaliyyət sahələrində uğur qazanmağın əsas vasitələrindən biri də dil imkanlarından səmərəli istifadə etmək bacarığıdır.

Dil xidmətlərinə olan yüksək tələbat həm də linqvistik informatikanın sürətli inkişafını stimullaşdırır. Bunun nəticəsidir ki, hazırda maşın tərcüməsi, terminoloji

informasiya sistemləri, xarici dil sahəsində öyrədici kompüter proqramları, dil sənayesinə xidmət edən bulud texnologiyaları sahəsində güclü rəqabət müşahidə olunur. Eyni zamanda, ənənəvi dil xidmətləri sektoru ilə texnoloji dil xidmətləri sektoru arasında da rəqabət artmaqdadır. Statistik analizlər göstərir ki, bu sahədə baş verən tendensiyalar ikincilərin xeyrinədir.

Burada paradoksal hal ondan ibarətdir ki, mobil sinxron tərcümə texnologiyalarının meydana gəlməsi və tədricən təkmilləşdirilməsi insanların xarici dil öyrənmək üçün motivasiyasını zəiflədir. Dil sənayesi spesifik intellektual fəaliyyət sahəsi kimi çıxış edir. Burada insan kapitalı və linqvistik informatika ən vacib resurslar hesab olunur.

Bütövlükdə, dil sənayesi öz mahiyyətinə görə biliklər iqtisadiyyatının xarakterinə uyğun gəlir. Ona görə də biliklər iqtisadiyyatının strukturu ilə bağlı elmi təsnifatlarda dil sənayesinə də yer ayrılması məqsədəuyğundur.

İSTİNADLAR

- [1] H.A. Cader, “The Evolution of the Knowledge Economy”, JRAP, 2008, No 38(2), pp. 117-129.
- [2] E.J.Stoilkova, S.Selimi, Z.B.Uraguchi, “Role of Knowledge Management in Knowledge Economies”, 31 October 2022, <https://www.helvetas.org>
- [3] M.G.Colombo, L. Grilli, L. Piscitello, “Science and Innovation Policy for the New KnowledgeEconomy”, Edward Elgar Publishing, 2011, 208 p.
- [4] V.Burduli, “Essence of Knowledge Economy and the Degree of Its Interoperability with Innovative Economy”, International Journal of

- New Economics and Social Sciences, June 2020, Vol.11(1), pp.61-82.
- [5] R.Zelazny, “Information society and knowledge economy - essence and key relationships”, Journal of Economics & Management, 2015, Vol.20(2), pp.5-22.
- [6] Sh. Hadad, “Knowledge Economy: Characteristics and Dimensions”, Management Dynamics in the Knowledge Economy, 2017, Vol.5, no.2, pp. 203-225.
- [7] R.M.Unger, “The Knowledge Economy: A Critique of the Dominant View”, American Affairs, Fall 2020, Volume IV, Number 3, pp. 48-76.
- [8] R.M.Unger, “The Knowledge Economy”, Verso Books, 2022, 297 p.
- [9] R.M.Algulyev, R.Sh.Mahmudov, “Language industry: opportunities, perspectives and problems”, Problems of Information Society, 2018, Vol.1, pp.3-26.
- [10] Elsevier, Language Services, <http://webshop.elsevier.com/language-services>
- [11] The Globalization and Localization Association, <https://www.gala-global.org>
- [12] Lionbridge, <https://www.lionbridge.com/language-cloud>
- [13] TransPerfect, <https://www.transperfect.com>
- [14] Internet World Stats, <https://www.internetworldstats.com/stats.htm>
- [15] Ch. Sin-wai The Future of Translation Technology: Towards a World Without Babel, Taylor&Francis, 2016, 316 p.
- [16] D.Kenny, “Machine translation for everyone: Empowering users in the age of artificial intelligence”, Language Science Press, 2022, 210 p.
- [17] Grand View Research, Machine Translation Market Size & Share, MT Industry Report, 2022, <http://www.grandviewresearch.com/industry-analysis/machine-translation-market>
- [18] H.Rose, “The future of English in global higher education: Shifting trends from teaching English to teaching through English”, 2019, <http://web.aou.edu.lb/online-journals/calr-issue9>
- [19] G.Diamantis, I.Karras, V.Zorbas, “Prospects for Global English: Which languages may rival English as a world lingua franca in the 21st century?”, Cultural Diversity Equity and Inclusion Intercultural Education in 21st century and beyond, Ioannina, 29.06-03.07 2015.
- [20] E.W.Schneider, “Hybrid Englishes: An exploratory survey”, World Englishes, Vol.35, Is.3, p.339-354.
- [21] M.Elisa, “A corpus-based analysis of new English blends”, Lexis - Journal in English Lexicology, 2019, <https://journals.openedition.org/lexis/3660>
- [22] Kassteen J. Global Trends in Foreign Language Demand and Proficiency, <https://studenttravelplanningguide.com/global-trends-in-foreign-language-demand-and-proficiency>

Sentiment analiz texnologiyaları və Azərbaycan dili

Günay İskəndərli

İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

gunayniftali@gmail.com

Xülasə — Sentiment analiz mətnin sentiment polyarlığının çıxarılması məqsədi daşıyır. Müxtəlif yanaşmalarda sözlərin sentiment polyarlığını müəyyən etmək üçün müxtəlif resurslardan istifadə olunur. İngilis dili bu cür resurslara sahib olmaq baxımından ən zəngin dil olsa da, Azərbaycan dili də daxil olmaqla, digər dillərin əksəriyyətində bu cür resurslar yoxdur. Bunu nəzərə alaraq, məqalədə sentiment analizin ingilis dilindən fərqli dillərdə mövcud problemləri araşdırılmış və Azərbaycan dilində mövcud vəziyyət nəzərdən keçirilərək bu istiqamətdə görülməli işlər müəyyən olunmuşdur.

Açar sözlər — *sentiment analiz; dil; sözlər şəbəkəsi.*

I. GİRİŞ

Bu gün internet şəbəkələrindən və sosial mediadan istifadə insanlara digər insanlar və ya qurumlar haqqında fikirlərini asanlıqla və aydın şəkildə ifadə etmək və şərh etmək imkanı verir. Bu isə öz növbəsində internetdə məlumatların kəskin artmasına yol açır. Bunun nəticəsidir ki, internetdən əldə edilən böyük verilənlərin təhlili, onların faydalı informasiyaya çevrilməsi və istifadəsi son illərdə intensiv şəkildə araşdırılan sahələrdən biridir. Sentiment analiz isə böyük verilənlərdə subyektiv məlumatların sistemli təhlilinə imkan verən təbii dilin emalı üsuludur. Mətnin

öyrənilməsi və kompüter linqvistikası alətlərindən istifadə edərək, tədqiqatçılara mətndəki emosiyaları müəyyən etməyə, onları emosiya kateqoriyalarına görə təsnif etməyə və kateqoriyaların emosional qütblərini tapmağa kömək edir.

Sentiment analiz mətnin sistemlik şəkildə tədqiqi, onun emosiya kateqoriyasının və qütblüyünün müəyyən edilməsi prosesidir. O cümlədən sentiment analiz vətəndaşların müəyyən bir mövzuya münasibətinin müsbət, mənfi və ya neytral olduğunu müəyyən etmək üçün mətnə ifadə olunan fikirlərin hesablaması yolu ilə müəyyən edilməsi və kateqoriyalara bölünməsi prosesidir [1].

Sentiment analiz dilçilik, maliyyə bazarları, marketing və sosial media təhlili kimi bir çox sahələrdə tez-tez istifadə olunur. Sentiment analiz üzrə aparılan tədqiqatlar populyar olsa da, hazırlanmış alqoritmlər və kitabxanalar nəzərə alındıqda, ingilis dili üzrə tədqiqatlar ön plandadır və müxtəlif dillərin təhlilinə imkan verən metod və kitabxanaların daha da inkişaf etdirilməsinə ehtiyac var. İngilis dili dünyada danışılan ümumi dil olduğundan, ədəbiyyatda ingiliscə mətnlər üzərində sentiment analiz üzrə çoxlu tədqiqatlar mövcuddur [2]. Bununla belə, digər dillərdə olan mətnlər üçün də sentiment analiz üzrə bir sıra işlər

görülmüşdür. Lakin Azərbaycan dilində olan mətnlərdə sentiment analiz hələ də açıq bir araşdırma sahəsi olaraq qalır.

Bu tədqiqatda deyilənləri nəzərə alaraq, sentiment analiz texnologiyaları nəzərdən keçirilmiş, digər dillərdə sentiment analiz üzrə mövcud problemlər araşdırılmış və Azərbaycan dili üzrə bu texnologiyaların tətbiqində qarşıya çıxan problemlər müəyyən edilmişdir.

II. SENTİMENT ANALİZ VƏ TƏTBİQ SAHƏLƏRİ

Sentiment analiz rəqəmsal məlumatların bol olduğu bu dövrdə biznes, hökumət və istehlakçı baxımından bir çox fayda və imkanlara malikdir. Bundan əlavə, Blokçeyn, Əşyaların İnterneti, Bulud Hesablama və Big Data kimi texnoloji yeniliklər sentiment analizin tətbiq dairəsini genişləndirərək, ondan praktiki olaraq istənilən sahədə istifadə etməyə imkan verir.

Sentiment analiz sosial mediada biznes, tibb, əyləncə və s. kimi sahələr üzrə yazılan rəylərin analizində uğurla tətbiq olunur. Aşağıda bu sahələrdə sentiment analizin tətbiqi imkanları nəzərdən keçirilmişdir [3]:

Biznes sahəsi: İstehlakçı firmalar məhsullarını təkmilləşdirmək, müştərilərin rəylərini araşdırmaq və innovativ marketing strategiyası hazırlamaq üçün sentiment analiz texnologiyalarından istifadə edə bilərlər.

E-ticarət inkişaf etdikcə satılan məhsulların sayı və müştərilərin rəyləri də artır. Sentiment analiz müştərilərə daha yaxşı məhsullar seçməyə kömək edə bilər. Sentiment analizin məqsədi hansı rəqiblərin

meydana çıxdığını və marketing kampaniyalarının necə müqayisə olunduğunu müəyyən etməkdir. Bir markanın və onun rəqiblərinin istehlakçı bazası haqqında tam təsəvvür əldə etmək üçün istifadə edilə bilər. Sentiment analiz bir neçə platformadan (*Twitter, Facebook, bloqlar*) məlumat toplayaraq nəzərə çarpacaq nəticələr verə və biznes analizindəki çətinlikləri aradan qaldıra bilər.

Səhiyyə və tibb sahəsi son dövrlərdə sentiment analizinin istifadə olunduğu sahələrdən biridir. Məlumatlar sorğular, *Twitter*, bloqlar, xəbər məqalələri, rəylər və s. kimi müxtəlif mənbələrdən toplanır. Bu məlumatlar daha sonra müxtəlif məqsədlər üçün təhlil edilə bilər. Məsələn, sentiment analiz standartların qiymətləndirilməsi və tibb sahəsindəki yeniliklərin təhlili məqsədlə istifadə oluna bilər. Sentiment analiz səhiyyə xidmətlərini yaxşılaşdırmaq üçün xəstələrin əhval-ruhiyyəsini, epidemiyaları, dərmanların mənfi reaksiyalarını və xəstəlikləri toplamaqda və qiymətləndirməkdə kömək edə bilər.

Əyləncə sahəsi: Sentiment analiz əyləncə sahəsində geniş istifadə olunur. Tamaşaçının reaksiyasını müəyyən etmək üçün film, şou və qısa filmlərin rəyləri təhlil edilə bilər. Bu, nəinki izləyicilərin daha yaxşı seçim etməsinə, həm də yaxşı məzmunun populyarlaşmasına kömək edə bilər.

Otellər və restoranlara dair rəylərin sentiment analizi müştərilərə daha yaxşı seçim etməyə, həmçinin sahibkarlara işlərini təkmilləşdirməsinə kömək edə bilər. Sentiment analiz vasitəsilə xidmət

təminatçıları daha çox mənfəət əldə edə bilirlər. Belə ki, onlar ən çox mənfə rəy alan xidmətləri aşkarlaya və onları təkmilləşdirə bilər.

Fond bazarı: Sentiment analizinin tətbiqlərindən biri səhm qiymətlərinin proqnozlaşdırılmasıdır. Bu, birja ilə bağlı bütün xəbərləri təhlil edərək və birja qiymətlərinin trendlərini proqnozlaşdırmaqla edilə bilər. Məlumatlar *Twitter*, xəbər məqalələri, bloqlar və s. kimi müxtəlif mənbələrdən toplanır. [4]-də aparılan tədqiqatda sentiment analiz fond bazarında tendensiyanın yüksələcəyini və ya azalacağını müəyyən etmək üçün istifadə olunur. Müsbət xəbərlər yüksəliş trendinə, mənfə xəbərlər isə aşağıya doğru meyl kimi qiymətləndirilir.

Siyasət: Dövlət qurumları sentiment analizdən istifadə edərək vətəndaşların ən çox hansı problemlərdən əziyyət çəkdiyini, hansı xidmətlərdən razı və narazı olduğunu müəyyən edə bilirlər. Sentiment analiz tendensiyaları, təkrarlanan problemləri və narahatlıqları aşkar etmək üçün məlumatları təsnif və təşkil etməyə imkan verə bilər.

Göründüyü kimi, sentiment analizinin tətbiq sahəsi olduqca genişdir və bu sahələrin hər birində həllini gözləyən bir sıra məsələlər mövcuddur.

III. SENTİMENT ANALİZİN ÜSTÜNLÜKLƏRİ VƏ ÇATIŞMAZLIQLARI

Opinion mining kimi də tanınan sentiment analiz, bir sıra sözlərin arxasındakı emosional çaların müəyyən edilməsi prosesidir və son illərdə

tədqiqatçıların daha çox diqqət mərkəzindədir. Bu sahədə həll edilməli olan bir-biri ilə əlaqəli bir çox altproblemlər var ki, bu da bu sahəni daha da çətinləşdirir. Bununla bağlı sentiment analizinin bir sıra üstünlükləri və çatışmazlıqları mövcuddur. Sentiment analizinin əsas üstünlükləri bunlardır:

Miqyaslılıq (genişlənmə qabiliyyəti) – sentiment analiz məlumatları səmərəli və sərfəli şəkildə emal etməyə imkan verir.

Real vaxt analizi – sentiment analiz vəziyyətləri dərhal müəyyən etməyə və hərəkətə keçməyə kömək edir.

Ardıcıl kriteriyalar – mərkəzləşdirilmiş sentiment analiz sistemindən istifadə edərək, şirkətlər bütün məlumatlarına eyni meyarları tətbiq edə bilirlər. Bu, səhvləri azaltmağa və məlumatların ardıcılığını yaxşılaşdırmağa kömək edə bilər.

Sentiment analizinin əsas çatışmazlıqları bunlardır:

- Mətnə istehzanın müəyyən olunması;
- İnkarnın idarə edilməsi;
- Cümlədə çoxsaylı fikirlərin olması;
- Çoxdillilik məlumatların mövcudluğu;
- Audio-video məlumatların olması və s.

Bu problemlərin aradan qaldırılması üçün tədqiqatçılar tərəfindən yeni metod və alqoritmlər təklif olunmasına baxmayaraq, hələ də həllini gözləyən problemlər qalmaqdadır.

IV. SENTİMENT ANALİZ VƏ DİLLƏR

Yuxarıda qeyd etdiyimiz kimi, sentiment analizin problemlərindən biri də dildir. İngilis dili, şübhəsiz ki, sentiment analizin əsas tədqiqat obyektidir. Bununla belə, ingilis dili fikirlərin onlayn ifadə olduğu yeganə dil deyil. Belə ki, internet məlumatlarının əksəriyyəti ingilis dilindən fərqli dillərdədir. Lakin zaman keçdikcə internetdə daha çox mənbə əldə olunmasına baxmayaraq, mənbələrin əksəriyyəti ilk növbədə ingilis dili üçün mövcuddur və bu isə müxtəlif dillərdə sentiment analizin tətbiqini çətinləşdirir. Bu isə tədqiqatçılar qarşısında belə bir mühüm sual çıxarır: niyə fərqli dillərdə sentiment analizin tətbiqi çətinləşir? Bunun bir neçə səbəbi var. Bunlardan bir neçəsi aşağıda nəzərdən keçirilmişdir [5]:

A. Fərqli kodlaşdırma ilə ilkin emal

Mətnin ilkin emalı istənilən təbii dilin emalı prosesinin tətbiqi üçün vacib bir addımdır. Kompüterlərin mətni emal etməsi üçün bugünkü sənaye standartı *Unicode*-dir. Qeyri-latin əlifbalarından və ya köhnə mətn məlumat mənbələrindən istifadə edən dillər kiril əlifbası üçün istifadə edilən KOI8 kimi fərqli kodlaşdırma üslubuna malik ola bilər. Bu isə həmin dillər üçün təbii dilin emalı prosesini ən başlanğıcda dayandırır.

B. Ümumişlək sözlər və mətn klassifikatoru

Təbii dilin emalı paketləri mənası olmayan sözləri (ümumişlək sözlər) təmizləmək üçün daxili söz siyahılarından kömək alır. Məsələn, ingilis dilindəki mətni emal üçün mənalı sözlərlə məhdudlaşdırmaq üçün mətndən “the, a, an və s.” kimi sözləri çıxarmaq lazımdır.

Ümumişlək sözlər müxtəlif dillərdə mövcuddur, lakin onların kompüterdə emal olunması üçün əl ilə işlənməlidir.

C. Dillər arasında struktur fərqləri

Sentiment analizin əsas addımlarından biri mətnin kiçik mənalı vahidlərə ayrılması prosesidir. Stanford Universitetində bir araşdırma qrupu tərəfindən müəyyən olunmuşdur ki, bu proses bir çox dillər üçün fərqlidir. Çünki müxtəlif dillərin qrammatik quruluşları fərqlidir.

Müxtəlif dillərdə sentiment analizin tətbiqi üçün əsas məsələlərdən biri sözlər şəbəkəsinin (*WordNet*) olmasıdır.

Hazırda 200-dən çox dil üçün nəzərdə tutulmuş “*WordNet*” şəbəkəsi mövcuddur. Qlobal *WordNet* Assosiasiyası dünyadakı bütün dillərə görə söz şəbəkələrini müzakirə etmək, paylaşmaq və birləşdirmək üçün platforma təmin edən açıq, ictimai və qeyri-kommersiya təşkilatıdır [6]. Bu platformadan istifadə edərək mövcud *WordNet* şəbəkələri ilə tanış olmaq olar.

Bir çox dillər üçün ingilis dilində yaradılmış *WordNet*-ə analoji olaraq belə söz şəbəkələrinin yaradılmasına baxmayaraq, bu şəbəkələr hələ də istənilən səviyyədə deyil. Belə ki, bu yaradılmış şəbəkələrin bir qismi açıq sistem olsa da, bir qismi açıq deyil. O cümlədən bu şəbəkələrdə semantik şəbəkələr istənilən səviyyədə deyil. Məsələn, ingilis dilində *WordNet*-də təxminən 117.000 semantik şəbəkə (*synset*) olmasına baxmayaraq, türkcə yaradılmış *WordNet*-də çox az – təxminən 15000 semantik şəbəkə (*synset*) var [7].

Bütün bunlar göstərir ki, müxtəlif dillərdə sentiment analizin tətbiq olunması üçün tədqiqatların daha da genişləndirilməsinə ehtiyac var.

V. AZƏRBAYCAN DİLİ VƏ ONUN SENTİMENT ANALİZ PROBLEMLƏRİ

Azərbaycan dili Altay dillərinin türk ailəsinin üzvüdür. Azərbaycan dilinin özünəməxsus xüsusiyyətləri bu dil üçün təbii dil emalını və sentiment analizin tətbiqini çətinləşdirir. Morfoloji cəhətdən Azərbaycan dili aqlütinativ bir dildir. Sözlər kökə çoxlu şəkilçilərin birləşməsi yolu ilə əmələ gəlir. Azərbaycan dili üçün sentiment analiz texnologiyalarının tətbiqinin əsas çətinlikləri bunlardır:

A. Azərbaycan dilinin morfoloqiyası

Azərbaycan dili aqlütinativ bir dildir. Bu törəmə və fleksiya şəkilçiləri sözün nitq hissəsini və semantik oriyentasiyasını dəyişə bilər (məsələn, “bəyəndim” (I liked it), “bəyənmədim” (I didn’t like it)).

B. Azərbaycan əlifbası

Azərbaycan dilində ingilis əlifbasında olmayan bir neçə simvol var: “ç”, “ğ”, “ı”, “ö”, “s”, “ü”, “ə”. Qeyri-rəsmi yazıda insanlar bu hərflərin bəziləri üçün ən yaxın ASCII simvollarını əvəz etməyə meyllidirlər (məsələn, “ç” “c” kimi yazılır). Bu sözlər kontekstdə insanlar tərəfindən oxuna bilsələr də, cümlədəki sözlərin dəqiq müəyyən edilməsində çətinliklər yaranır.

C. İnkarnın mürəkkəbliyi

Azərbaycan dilində bir sözü inkar etməyin bir neçə yolu var ki, bu da öz

növbaşında ifadə olunan hissi dəyişir. Bunlar aşağıdakılardır:

- ma/mə şəkilçiləri feli inkar edir (məsələn, “bəyənmədim” (I did not like it)).
- siz/sız/suz/süz şəkilçiləri, məsələn, “uğursuz” (without success – unsuccessful).
- “deyil” sözü (is/are not) , məsələn, “gözəl deyil” (is not beautiful).
- “Yox” sözü yoxluğu və ya əlçatmazlığı göstərir. (məsələn, “mövzu yoxdur” (does not have a topic))

NƏTİCƏ

Sentiment analiz mətnin verilmiş seqmentinin sentiment polyarlığını aşkarlamaq məqsədi daşıyır. Sentiment analizi ingilis dilindən fərqli dillərdə tətbiq etməyin əsas çətinliklərindən biri məhdud polyarlıq resurslarının mövcudluğudur. Belə ki, ingilis dili bu cür resurslara sahib olmaq baxımından ən zəngin dil olsa da, Azərbaycan dili də daxil olmaqla, digər dillərin əksəriyyətində bu cür resurslar yoxdur.

Yuxarıda qeyd etdiyimiz kimi, sentiment analizi Azərbaycan dilinə tətbiq etməyin əsas şərtlərindən biri sentiment analiz üzrə resursların mövcud olmasıdır. Bunun üçün isə ingilis dilində yaradılmış *WordNet* şəbəkəsinə analoji olaraq sözlər şəbəkəsi və SentiAzerNet şəbəkəsi yaradılmalıdır. Bu məqsədlə informasiya texnologiyaları və dilçilik üzrə tədqiqatçılar birgə araşdırma aparmalı və əməkdaşlıq etməlidirlər.

İSTİNADLAR

- [1] R.Dezhkarghani, Y.Saygin, B.Yanikoglu and K.Oflazer, “SentiTurkNet: a Turkish polarity lexicon for sentiment analysis,” Language Resources & Evaluation, Vol. 50, No.3, pp.667-685, 2016.
- [2] S.Deshmukh, N.Patil, S.Rotiwari and J.Nunes, “SENTIMENT ANALYSIS OF MARATHI LANGUAGE”, International Journal of Research Publications in Engineering and Technology, Vol.3, No.6, pp.93-97, 2017.
- [3] M.Wankhade, A.Ch.S.Rao and Ch.Kulkarni, “A survey on sentiment analysis methods, applications, and challenges”, Artificial Intelligence Review, Vol.55, No.7, pp.5731-5780, 2022.
- [4] F.Z. Xing, E. Cambria, R.Welsch, “Natural Language based financial forecasting: a survey”, Artificial Intelligence Review, Vol. 50, No.1, pp. 49–73, 2018
- [5] C.Dilmegani, Challenges and methods for Multilingual sentiment analysis in 2023. <https://research.aimultiple.com/multilingual-sentiment-analysis/>
- [6] <http://globalwordnet.org/>
- [7] G.Geziçi, B.A.Yanikoglu, “Sentiment Analysis in Turkish”, Turkish Natural Language Processing, pp.1-19, 2018.

Bədii mətnin linqvostatistik təhlili

Nigar Vəlizadə

AMEA Nəsimi adına Dilçilik İnstitutu, Bakı, Azərbaycan

n.valizade@mail.ru

Xülasə — Son illərdə meydana çıxmış, statistika ilə dilçiliyin qovuşmasının məhsulu olan statistik dilçilik elmi böyük maraq doğurur. Linqvostatistika son zamanlar sürətlə inkişaf etmiş və müstəqil bir elm sahəsi kimi formalaşmağa başlamışdır. Nəzərə almaq lazımdır ki, statistik dilçilik dilçiliyin bir sahəsi kimi tanınır. Bu termin bir çox dilçilər tərəfindən “linqvostatistika” kimi qəbul edilmiş və işlənməkdədir. “Linqvostatistika” dil hadisələrinin kəmiyyət səciyyələrinin öyrənilməsi və təsnifi üçün statistik üsulların tətbiqidir.

Açar sözlər — linqvostatistik təhlil; bədii mətn; müəllif üslubu, mətnin təhlili, tezlik siyahıları

I. GİRİŞ

Linqvostatistik material dedikdə, statistik məlumatlarla təchiz olunmuş dil materialı nəzərdə tutulur. Bu qəbildən olan materiallar linqvistik araşdırmalarda müxtəlif məqsədlərlə istifadə oluna bilər. Yəni, sözlərin, söz birləşmələrinin, söz-formalarının, köməkçi nitq hissələrinin, sadə və mürəkkəb cümlə konstruksiyalarının və s. işlənmə tezliyi, müxtəlif mətnlərdə rastgəlmə xüsusiyyətləri tədqiqatçı maraqlandıra bilər. Ayır-ayrı yazıçıların dilində bəzən müxtəlif cümlə konstruksiyalarından istifadə olunur ki, bu da müəllif üslubunun linqvostatistik

parametrlərini müəyyənləşdirmək baxımından əhəmiyyətlidir. Digər tərəfdən linqvostatistik parametrlərə görə anonim əsərlərin müəllifinin müəyyənləşdirilməsi də mümkündür.

Ümumiyyətlə, linqvostatistik təhlil zamanı mətnin əsas mənası, forması, quruluşu və s. oxuculara lazımı şəkildə çatdırılır. Mətnin linqvostatistik təhlili müəyyən elmi bazası və bu məsələlərlə bağlı elmi səriştəsi olmayan tədqiqatçılar üçün çətinlik törədir. Buna görə də, təhlilə başlamazdan əvvəl bəzi zəruri anlayışlarla və əsas terminlərlə tanış olmaq vacibdir. Tədqiqat materialı nə qədər çox olarsa bədii mətnin linqvostatistik təhlili ilə bağlı əldə olunmuş nəticələr bir o qədər etibarlı olar. Təhlil zamanı sözün fonetikasi və morfemik tərkibi də ətraflı şəkildə və statistik olaraq araşdırılır.

Dilçilikdə dil hadisələrinin linqvostatistik öyrənilməsində müxtəlif metodlardan istifadə edilir:

- 1) leksikoqrafik statistika metodları;
- 2) üslubi statistikasının metodları;
- 3) mətnin ölçülməsi üsulları.

Statistik metodlar dilçiliyin ən müxtəlif sahələrinə getdikcə daha dərinlən nüfuz etməkdədir. Dilçilik elmində geniş istifadə olunan bu metodların məcmusu

(lingvostatistika) nitqdə və məndə hansı vahidlərin daha tez-tez və ya çox az rast gələ bilməsi hallarını ehtimallarını) müəyyən etməyə imkan verir. Dil vahidlərinin işləkliyinin təhlili nitq ifadələrinin əmələgəlmə prosesləri üzrə müəyyən qanunauyğunluqların aşkara çıxarılmasını təmin edir [3].

II. TEZLİK LÜĞƏTLƏRİ

Dilçilikdə həm tezlik lüğətlərinin, həm də digər statistik lüğətlərin tərtibi sahəsində müəyyən təcrübə əldə edilmişdir. Ən tədqirleyiq nümunə isə 2004-cü ildə çap olunan “Azərbaycan dilinin tezlik lüğəti”dir. Lüğətin 3 cilddən ibarət olması nəzərdə tutulur. I cild söz köklərinin tezlik və əlifba-tezlik siyahılarını əhatə edir. II cildə əsasların statistik göstəriciləri əhatə olunacaqdır. III cildin isə söz-formaları əsasında tərtibi nəzərdə tutulur.

Azərbaycan dilinin tezlik lüğətlərinin tərtibi üçün bütün üslublardan istifadə olunaraq təxminən 50 milyona yaxın söz-forma lingvostatistik baxımdan nəzərdən keçirilmiş və nəticə olaraq 300 minə yaxın söz-forma yerdə qalmışdır. Mətnin statistik təhlili ana dilinin və xarici dillərin tədrisi ilə əlaqədar səmərəli metodların işlənməsi məsələlərini də əhatə edir. Lüğət-minimumların, danışiq kitabçalarının hazırlanması, tədris materiallarının seçilməsi prosesində tezlik lüğətlərindən faydalanmaq olar. Tezlik lüğətləri dilin sözmələgətirmə tiplərinin, sözdəyişdirmə imkanlarının araşdırılması, orfoqrafiya və qrafikanın təkmilləşdirilməsi baxımından da böyük səmərə verə bilər [1].

III. BƏDİİ MƏTNLƏRİN TƏHLİLİ

Son zamanlar Azərbaycan yazılı abidələrinin və klassiklərin dilinin lingvostatistik təhlili ilə bağlı da araşdırmalar aparılmaqdadır.

“Kitabi-Dədə Qorqud” milli mənşəyimizi, tariximizi, adət-ənənələrimizi əks etdirən ən möhtəşəm tarixi abidə, həyat ensiklopediyası, Azərbaycan xalqının milli sərəvətidir. Bizim 1300 illik tariximizi özündə cəmləşdirən “Kitabi-Dədə Qorqud” dastanında dilin fonetik, leksik, qrammatik səviyyələri qorunub saxlanılır. “Kitabi-Dədə Qorqud” dilində, Azərbaycan dilinə məxsus üçüncü növ təyini söz birləşmələrinin hər iki tərəfinin formal əlaməti işlənməmişdir. Bunlardan birinci tərəf həmişə ziyəlik hal, ikinci tərəf isə mənsubiyyət hal şəkilçisi qəbul edir. Nümunəyə diqqət edək: Kəndimizin şöhrəti bütün dünyaya yayılmış xalçaları muzeyləri bəzəyir. Xanəndənin səsi uzaqlara yayılan zənguləsi hamını heyran etmişdir. Səhranın qumu hər tərəfə səpələyən küləyi dayanmaq bilmirdi.

Dastanda “y” samitinin geniş işlənməsi də dastan dilinin qədimliyini göstərir: üz əvəzinə yüz, ürək əvəzinə yürək, igid əvəzinə yığit, uca əvəzinə yuca və s. [2].

Dastanın leksik səviyyədə tədqiqində bir çox maraqlı hallarla rastlaşmaq mümkündür. Burada eyni sözlərin təkrarı diqqət çəkir: qızıl-qızıl, avaz-avaz, qanlu-qanlu, diyar-diyar, tək-tək, salqum-salqum, qara-qara, gilə-gilə və s. Bununla yanaşı bəzi sözlər var ki, müasir dövrümüz üçün arxaik sayılır. Bunlara misal olaraq damu (cəhənnəm), güz (payız), uçmaq (cənnət), bez (əsgü), ərən (igid), tanuq (şahid) və s. sadalamaq mümkündür. Dastan mətninin lingvostatistik baxımdan araşdırılması bir

çox statistik qanunauyğunluqların aşkarlanması imkanı yaratmış olur.

Fonetik və leksik səviyyələrdə olduğu kimi, morfoloji səviyyənin də tədqiqi üçün “Kitabi-Dədə Qorqud” dastanını misal gətirə bilərik. Dastanda özüm, özün kimi qayıdış əvəzlilikləri də şəxs əvəzlilikləri ilə verilir: Gördünmü mən mənə nə etdim. Dastanın dilində s1x-s1x işlənən –ıbani//übəni felii bağlama formaları da dastanın dilinin çox qədim olduğunu göstərir: Gülübəni qarşuladular, qara otağa qondurdular, qara keçə altuma döşədilər, qara qoyun yəxnisindən ögümə gətürdülər. Qalqubanı yerindən turı gəldin [2].

Dədə-Qorqud dilinin statistik təhlilində şəxs əvəzliyinin birinci şəxsinin təkinin həm “mən”, həm də “bən” formaları ilə işlənməsinin şahidi oluruq. Konkret desək, dastanda “mən” sözü 121 (şəkilçili formaları 246, “bən” sözü 16 (şəkilçili formaları 43) kərə işlənmişdir [4]. Dastanın dilində işlənmiş bəzi sözlərə müasir dilimizdə də ümumxalq danışq dilində və ədəbi dildə işlək leksik vahidlər kimi geniş rast gəlinir. Məsələn, ana, ata, baba, əmi, dayə, gəlin, nişanlı, dayı, qız, qayın, qaynana, qaynata və s.

Dastanda yüksək tezlikli fellər çoxluq təşkil edir. Məsələn, yet (mək), ol (maq), öl (mək), düş (mək), sal (maq), var (maq), yar (maq), ver (mək), de (mək), qıy (maq), söylə (mək), yan (maq), tut (maq), din (mək), sor (maq), dön (mək), al (maq), vur (maq), gir (mək), dər (mək), istə (mək) və s. sadalamaq mümkündür. Aztezlilikli sözlərə: ildız, yavırım, kir, kiş, güzə, gölgə, ordu, özin, ögit, rum, panbuq, sağdıç, parə, rukət, ayğır, ağa, ağı, ata, bayat, bayandır, vay, vaqəə, qaftan, qəhr, qıl, qeyri, min, derəm,

de, dil, dəstur, eş, əvət, il, əyər, zəval, kar, keşiş, kətan, kərvan və s. misal göstərilə bilər.

NƏTİCƏ

Müasir dövrümüzün tələbinə əsasən, bir çox tədqiqata cəlb olunan mövzular yenidən işlənməli, daha da təkmilləşdirilməli və geniş şəkildə izahı verilməlidir. Azərbaycan dili tarixinin və inkişaf qanunauyğunluqlarının öyrənilməsində linqvostatistik təhlil metodu böyük əhəmiyyət kəsb edir.

Tədqiqat materialı nə qədər çox olarsa bədii mətnin linqvostatistik təhlili ilə bağlı əldə olunmuş nəticələr bir o qədər etibarlı olar. Təhlil zamanı sözün fonetikasi və morfemik tərkibi də ətraflı şəkildə və statistik olaraq araşdırılır.

Qeyd etməliyik ki, abidələr üzərində linqvostatistik baxımdan aparılan təhlillər də böyük əhəmiyyət kəsb edir. Konkret faktlara əsaslanaraq dilin yaranma və formalaşma tarixini öyrənməyə imkan verir. Hal-hazırda linqvostatistikanın köməyi ilə yalnız müasir dövrü deyil, eləcə də qədim dövr yazılı abidələrini tədqiq etmək və bilinməyən bəzi məqamları ortaya çıxarmaq mümkündür.

İSTİNADLAR

- [1] Azərbaycan dilinin tezlik lüğəti (söz kökləri). I cild (tərtibçilər – M. Mahmudov, Ə. Fətullayev və b.), Bakı, “Elm”, 2010, 464 s.
- [2] Axundov A.A. Azərbaycan dili saıtlərinin statistik təhlili təcrübəsindən, “ADU-nun elmi əsərləri (Dil və ədəbiyyat)”, №5, 1963, 63-67s
- [3] Mahmudov M.Ə. Kompüter dilçiliyi. – Bakı: “Elm və təhsil” nəşriyyatı, 2013. 356 s.
- [4] “Kitabi-Dədə Qorqud”un statistik təhlili (tərtibçilər: K.A Vəliyeva, M.Ə Mahmudov, V. Pines, C.Ə.Rəhmanov, V.S.Sultanov), Bakı, “Elm”, 1999, 248 s.

Viki texnologiyaların milli dilin qorunmasında rolu və vikipediyanın Azərbaycan dili seqmentində problemlər

İradə Ələkbərova

İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

airada.09@gmail.com

Xülasə — Xalqın dili ilk növbədə o xalqın mədəniyyətini, mənəvi zənginliyini, varlığını, keçmişi və gələcəyini əks etdirir. Cəmiyyətin bir xalq, millət kimi tanınmasında, milli-mənəvi özünüifadəsində dil çox vacib rol oynayır. Məqalədə viki texnologiyalarının milli dilin qorunmasında rolu göstərilmiş, Vikipediya virtual ensiklopediyasının mahiyyəti və Azərbaycan dili seqmentində vəziyyəti, dillə əlaqədar problemləri müəyyən edilmişdir.

Açar sözlər — viki texnologiyaları; Vikipediya; e-dövlət; informasiya cəmiyyəti; ölmüş dillər; dilin bərpası

I. GİRİŞ

Elm və texnologiyaların inkişafı, informasiya cəmiyyətinin genişlənməsi dilin fasiləsiz olaraq zənginləşməsinə, inkişafına və dəyişməsinə təsir göstərir. Dilin lüğət tərkibində yeni sözlərin yaranması ilk növbədə cəmiyyətdə gedən proseslərlə bağlıdır. Dilin qorunmasında elm, mədəniyyət, tədris, media, ictimaiyyət, ən əsası isə dövlətin yüksək dəstəyi vacibdir. Dilin dövlət dili kimi vacib bir statusa yüksəlməsi tarixi hadisə, millətin və dövlətin dünyada təsdiqidir. Dilin bu statusa malik olması millətin öz taleyinə

sahibliyini, dövlət qurmaq, bu dövləti idarə etmək və qorumaq qabiliyyətinə malik olmasını sübut edir.

İnformasiya resurslarının çoxalması, informasiya-kommunikasiya texnologiyalarının (İKT) inkişafından asılı olaraq insanlar arasında ünsiyyət formaları dildə yeni anlayışların, terminlərin yaranmasına gətirmişdir. Qloballaşmanın müasir prosesləri insanlar arasında ünsiyyətdə, təhsildə, informasiyanın əldə edilməsində və paylanmasıda kiçik və minoritar dillərin rolunun azalmasına səbəb olmuşdur. İKT vasitələrinin, əsasən də İnternetin insan həyatının bütün sahələrinə təsiri bu prosesin inkişafını daha aydın göstərməkdədir. İnformasiyanın həcminin durmadan artması insanlar arasındakı informasiya mübadiləsində yalnız bir-neçə dildən istifadə etməyə zərurət yaradır. 2001-ci ildə kompüter dünyasında yeni layihə – viki-texnologiyaları vasitəsilə yaradılan və idarə olunan Vikipediya (Wikipedia) virtual ensiklopediyası fəaliyyətə başladı. Bu layihə dillərin və mədəniyyətlərin qorunmasına xidmət etməklə yanaşı, harada yaşamasından, hansı dildə danışmasından asılı olmayaraq

insanların informasiya əldə etməsini asanlaşdırdı [1].

II. DÜNYADA DİL İLƏ BAĞLI VƏZİYYƏT

Ölkələrin informasiya cəmiyyətinə keçidi və İnternetin genişlənməsi dilə süni və ya təbii üsullarla yeni terminlər və anlayışların daxil olunmasına gətirir. Bu proses dilin saflığına çox ciddi zərbdır. Müşahidələr göstərir ki, qloballaşma dərinləşdikcə bəzi milli dillərin güclü dillər tərəfindən sıxışdırılması artır və dilin yox olmaq təhlükəsi yaranır [2].

Pensilvaniya ştatındakı Suortmor kollecinin professoru amerikalı linqvist Devid Xarrison “Dillər ölən zaman” (When Languages Die) kitabında bildirir ki, hazırkı vəziyyətdə tarix üçün dillərin indiki sürətlə yox olması qəbuledilməzdir. Kitabda deyilir: “Dillərin ölümü prosesi təbiətdəki bitki və heyvanların ölümündən daha sürətlə gedir. Əgər dillər haqqında Qırmızı kitab yaransa, oraya dünyadakı dillərin 40%-i salınmalıdır. Təbiətdə isə 8% bitkilərin və 18% canlıların yox olmaq təhlükəsi var ki, dillərlə müqayisədə bu o qədər də çox deyil” [3].

Alimlərin araşdırmaları göstərir ki, dünyada təqribən, 7 mindən artıq dil mövcuddur. Bu dillərin çox hissəsi insanların sıx yaşadığı və ərazicə böyük dövlətlərin payına düşür. Bu dövlətlər İndoneziya, Meksika, Braziliya, Kamerun, Hindistan, Çin, ABŞ və Rusiyadır [4]. Baxmayaraq ki, Avropa dilləri dünya dillərinin 4%-ni təşkil edir, dünyada ən çox yayılmış dillər əsasən Avropa ölkələri

dillərini təşkil edir. Bununla belə, Avropada təqribən 50 dili yox olmaq və ciddi dəyişikliklərə məruz qalmaq təhlükəsi gözləyir ki, bunlardan da birinci yerdə laplan dilidir ki, bu dildə Skandinaviya ölkələrində və Rusiyanın şimalında danışılır.

Rusiyada ölməkdə olan dillərə misal olaraq xanti, mansi, komi, momari, nanay, neqidal, orok, oroç, udeqe, kerek, ulç, evenk, qaqauz, urum dillərini göstərmək olar. Təəssüf ki, türk dillərinin də bir qismi yox olmaq ərəfəsindədir, məsələn, yakut, tivins, xakos, dolğan, altay dilləri bu siyahıya daxildir.

Alimlərin araşdırmaları göstərir ki, tarix boyu 9 mindən artıq dil yox olmuşdur. Belə dillər “ölü dillər” adlanır. Çox təəssüf ki, ölü dillər siyahısında bir zamanlar milyonlarla insanların danışdığı, qonşu dövlətlərə meydan oxuyan qüdrətli imperiyaların da dilləri daxildir. Ölü dillər siyahısına daxil olan belə dillərdən latın, qədim yunan, assuriya, qədim fars, qədim türk və s. dilləri göstərmək olar [5].

Birləşmiş Millətlər Təşkilatının (BMT) baş qərargahında keçirilən brifinqdə BMT-nin baş katibinin mətbuat katibi Mişel Mantasin xəbər verdiyi kimi, dünyada hər iki həftə ərzində bir dil ölür [6]. Dünyadakı dillərin 90%-də cəmi 100 min adam danışır. Dilçi alimlərin araşdırmaları göstərir ki, dilin yaşaması, qorunub saxlanması üçün həmin dildə heç olmasa 100 min insan danışmalı, yazmalıdır. Dilin qorunmasında yazının olması xüsusi yer tutur. Yazı vasitəsilə biz tarixdə mövcud olmuş dillər haqqında bilik əldə edirik. Tədqiqatlardan

məlumdur ki, hazırda 400 dili yox olmaq təhlükəsi gözləyir [7, 8].

BMT-nin 2002-ci il üçün nəşr etdirdiyi “Yox olmaq təhlükəsi altında olan dünya dillərinin atlası” (Atlas of the World's Languages in Danger of Disappearing) kitabında deyilir: “Dillərin yox olması ilə birgə insanların əsrlər boyu topladığı biliklər və fikirlər də məhv olur” [9].

21 fevral – Beynəlxalq Ana Dili Günüdür

1999-cu ilin noyabrında, UNESCO-nun baş qərargahında keçirilən BMT-nin 30-cu sessiyasında Banqladeş səfirinin təşəbbüsü ilə fevralın 21-i Beynəlxalq Ana Dili Günü elan edilmişdir və 2000-ci ilin 21 fevralından etibarən bu gün hər il qeyd edilir.

Beynəlxalq Ana Dili Günü elan edilməsinin tarixi səbəbləri var: 1952-ci il 21-22 fevral tarixlərində Pakistanda benqal dilinin qadağan edilməsinə etiraz olaraq keçirilən aksiyada polis və silahlı qüvvələrin müdaxiləsi nəticəsində 3 nəfər qətlə yetirilmişdir. Banqladeş nümayəndələri tərəfindən fevralın 21-nin məhz həmin şəhidlərin xatirəsinə ehtiram əlaməti kimi Ana Dili Günü adlandırılması ilə əlaqədar BMT-yə müraciət edilmişdir [10, 11]. Bu günün qeyd edilməsində əsas məqsəd dil və mədəniyyət müxtəlifliyinin, çoxdilliliyin qorunmasıdır

III. AZƏRBAYCAN DİLİNİN DÖVLƏT DİLİ STATUSU ALMASI

Azərbaycan dili çox çətin tarixi bir yol keçmişdir. Əsrlər boyu müxtəlif dövlətlər tərəfindən işğallara məruz qalan, müstəqil xalq kimi inkişafına süni maneələr

yaradılan, doğma dilində danışmağa belə icazə verilməyən bir millət üçün dilini əsrlərin keşməkeşindən itkisiz keçirərək dövlət dili statusuna qədər yüksəltməsi böyük nailiyyətdir.

Azərbaycan dili zəngin ənənəyə malikdir və Azərbaycan xalqı kimi çox qədim tarixi köklərə bağlıdır. Azərbaycan dilinin xarici təsirlərdən mühafizəsi və yad sözlərdən təmizlənməsi birinci növbədə dilimizin zəngin xəzinəsi hesabına baş verir. Azərbaycan dilinin zənginliyinin, linqvistik gözəlliyinin zəmanəmizə kimi qorunub saxlanması bir daha azərbaycanlıların öz kökünə bağlı, adət-ənənələrinə sadıq bir xalq olduğunu sübut edir. Dilimizin dövlət statusu alması siyasi, mədəni həyatda əldə edilən uğurlardan ən mühümüdür və ölkəmizdə aparılan dil siyasətinin real nəticəsidir.

1978-ci ildə qəbul edilmiş Azərbaycan SSR Konstitusiyasında Azərbaycan dili Azərbaycanın dövlət dili elan edilmişdir və bu mühüm hadisədə xalqımızın ümummilli lideri Heydər Əliyevin xidməti misilsizdir [12]. Ulu öndərimizin başçılığı ilə müstəqil Azərbaycan Respublikasının ilk konstitusiyasında da Azərbaycan dilinin dövlət dili statusu təsbit olundu. Konstitusiyanın “Dövlət dili” adlı 21-ci maddəsində deyilir: “Azərbaycan Respublikasının Dövlət dili Azərbaycan dilidir. Azərbaycan Respublikası Azərbaycan dilinin inkişafını təmin edir”. Bu gün Azərbaycan dili ictimai, siyasi, mədəni inkişafı ilə bağlı zənginləşir, onun publisistik, bədii və elmi üslubları daha da inkişaf edir. Azərbaycan dilinin dövlət dili

statusuna yüksəlməsi azərbaycanlıların bir millət kimi öz taleyinə sahibliyini, bir xalq kimi inkişafını və dünya siyasətində rolunu, dövlət qurmaq qabiliyyətini sübut etmiş oldu.

Azərbaycanda müstəqillik əldə etdikdən sonra ana dilimizin qorunması və inkişafı üçün bir sıra addımlar atılmışdır. Ölkəmizdə Azərbaycan dilinin zənginləşdirilməsi, işlək dilə çevrilməsi, beynəlxalq münasibətlər sisteminə yol tapması dilimizin gözəl bilicisi, ümummilli lider Heydər Əliyevin dilin qorunmasına yönələn düşünülmüş siyasətinin nəticəsidir. “Dövlət dilinin tətbiqi işinin təkmilləşdirilməsi haqqında” 18 iyun 2001-ci il tarixli fərman Azərbaycan dilinin qorunması işində yeni mərhələ oldu. Bu sənəd Azərbaycan dilinə verilən ən yüksək qiymət idi. Fərmanda orta və ali məktəblərdə Azərbaycan dilinin müasir dünya standartlarına uyğun tədrisinin həyata keçirilməsi, dilimizin tarixi ənənələrə, mənəvi köklərə uyğun olaraq inkişaf etdirilməsi və qorunması göstərilir [12].

Azərbaycan dilinin qorunması istiqamətində aparılan işlərin davamı olaraq Azərbaycan Prezidenti İlham Əliyev tərəfindən 12 yanvar 2004-cü il tarixdə “Azərbaycan dilində latın qrafikası ilə kütləvi nəşrlərin həyata keçirilməsi haqqında” sərəncam imzalanmış, Milli Məclisdə “Dövlət dili haqqında” qanun qəbul edilmişdir [13]. Dövlət tərəfindən Azərbaycan dili ilə bağlı fərman və sərəncamlar milli-mənəvi dəyərlərimizə,

elmin və mədəniyyətin inkişafına böyük diqqət yetirdiyini bir daha sübut edir.

IV. DİLLƏRİN İNKİŞAFI VƏ QORUNMASINDA VİKİ-TEKNOLOGİYALARININ ROLU

1995-ci ildə amerikalı proqramçı Uord Kenninqhem (Ward Cunningham) tərəfindən təklif olunan viki konsepsiyası İnternetə olan münasibətin dəyişməsinə səbəb oldu. Belə ki, istənilən şəxs xərc qoymadan və birbaşa brauzerdən virtual məkanda veb-səhifə yaratmaq hüququ əldə etdi. Bu konsepsiya virtual məkanda toplanan kontentə yeni münasibətlərin gətirilməsinə təkan verməklə milli dilin qorunması və biliyin paylanmasında inqilab etdi. Viki texnologiyaları İnternet istifadəçilərinin əməkdaşlığı, şəbəkə vasitəsilə koordinasiyası və real həyatın sosial aspektlərini özündə əks etdirən infrastrukturunu xarakterizə edir [14].

Viki texnologiyaları İnternet istifadəçilərinə tam hüquqlu veb-proqramçı kimi virtual mühitdə kontent yaratmaqla yanaşı, fayl yükləmək, müzakirələrdə iştirak etmək, digər istifadəçilərin yaratdığı məqalələrdə dəyişiklik etməyə icazə verir ki, bu da informasiyanın daha dolğun və aktual olmasına səbəb olur [15].

Viki texnologiyaları ilə idarə olunan Vikipediya 318 dildə fəaliyyət göstərən sosial şəbəkədir. Bu layihə ən geniş istifadəçi auditoriyasını qazanan veb-resurslardan, ən populyar sosial şəbəkələrdən biri, biliklərin toplandığı nəhəng baza kimi yüksək reyting qazanmışdır [16, 17]. 60 milyondan artıq

ensiklopedik məqaləyə və 106 milyondan artıq istifadəçiyə malik [18] Vikipediyanın populyarlaşmasının əsas səbəbi onun dinamik və daima yenilənən olmasıdır. Vikipediya praktiki olaraq bütün vacib sahələr haqqında müəyyən dərəcədə informasiya vardır. Nəzərə alsaq ki, "Britannika" ensiklopediyasında məqalələrin sayı 120 min, Böyük Sovet Ensiklopediyasında məqalələrin sayı 95279, Brokqauz və Efron ensiklopedik lüğətdə məqalələrin sayı 121240-dır, o halda Vikipediyanın necə nəhəng sistem olduğunu təsəvvür etmək olar.

Vikipediyanın nəhəng məlumat bazası hər saniyə yeni məqalə, xəbər, kitab, foto, audio və video fayllarla zənginləşir. Vikipediya genişləndikcə dünyanın istənilən nöqtəsində yaşayan insanlarla virtual ünsiyyət qurmaq, məktublaşmaq, gündəlik hadisələr haqqında operativ məlumatlar almaq mümkündür.

Dilin bərpasında viki texnologiyalarının əvəzolunmaz rolu var. Ölü dil hesab olunan latın, qədim rus, qədim fars, qədim ingilis və s. dillərdə Vikipediya ensiklopediyasında seqmentlər mövcuddur. Xalqın psixologiyasını və milli xüsusiyyətlərini xarakterizə edən əsas mədəni mənbələrdən olan milli dili mühafizə edərək gələcək nəsillərə olduğu kimi çatdırmaq müasir cəmiyyətimizdə ən aktual problemlərdəndir. Linqvistikada canlı və ölü dil məfhumları məlumdur.

Canlı dil insanların gündəlik ünsiyyətində işlədilən dildir. Dil elə bir məfhumdur ki, onu digər mədəniyyət nümunələri kimi muzeylərdə saxlayaraq

qorumaq mümkün deyil. Hər hansı bir dildə çap olunan lüğətlər, əlifbalar, dərsliklər bu dilin gələcəkdə olduğu kimi qorunub saxlanmasına zəmanət vermir. Bu səbəbdən, ölməkdə və yaxud sürətlə dəyişməkdə olan dillərin qorunması üçün ilk növbədə həmin dildə danışan insanların olması vacibdir.

Frigiya, finikiya, şumer, parfiya, puni, qədim çin, akkad, tirren, hun, ket, çağatay, xəzər, murom, meşşer, bulqar, soqdi, prussiya, sak, midiya, qədim makedoniya, qədim fars, qədim yunan, venet, yan, qədim rus, qədim ingilis, praslavyan, qot, qədim sakson kimi dillər ölü dillərə misal ola bilər. Müəyyən tarixi hadisələrin və texniki inqilabların təsiri altında ölü dillər bərpa oluna bilər. Məsələn, İsrail dövlətinin yaradılması (1947) ölü dil hesab olunan ivrit dilinin bərpasına səbəb oldu.

Viki texnologiyaları ensiklopedik məqalələrin istənilən İnternet istifadəçiləri tərəfindən və istənilən dildə yaradılmasına icazə verməklə, dillərin qorunmasına, zənginləşməsinə və hətta, ölü dilin “dirilməsinə” bilavasitə yardım etmiş olur.

Vikipediya *Unicode* (UTF-8) kodlaşmasından istifadə edir ki, bu da ensiklopediyada eyni zamanda müxtəlif dillərdə məqalələr yaratmağa imkan verir. Vikipediya müxtəlif dillərə malik seqmentlərin olması, onun bu və ya digər dilin qorunması, bu dilin mədəniyyətinə xas olan spesifik reallıqlar barədə ensiklopedik biliklər bazasının yaradılması istiqamətində unikal layihəyə çevirmişdir [18].

Dillərin inkişafında və qorunmasında Vikipediyanın əhəmiyyəti ondadır ki, o, hər

hansı bir dil üzrə seqment yaratmaqla dünya ictimaiyyətini bu dilin mövcud olması barədə xəbərdar etmiş olur. Vikipediya hər hansı mövzuda məqalə ilə tanış olarkən bu məqalənin daha hansı dillərdə mövcud olması haqqında məlumat almaqla, doğma dildə bu məqalə mövcud deyilsə, onu digər dillərdən tərcümə edərək asanlıqla yaratmaq mümkündür. Viki texnologiyalarının imkanları hətta ölmüş dillərin bərpasını da mümkün edir. Əgər bir dildə danışan, müzakirələr aparan, ensiklopedik və elmi məqalələr, kitablar yazan insanlar varsa, deməli o dil yaşayır.

V. VİKİPEDIYANIN AZƏRBAYCAN DİLİ SEQMENTİNDƏ VƏZİYYƏT

Azərbaycan dilində Vikipediya 2004-cü ildə www.az.wikipedia.org ünvanı altında fəaliyyətə başlamışdır. Bu gün Azərbaycan dilində məqalələrin sayı 192 mindən çoxdur [19]. Azərbaycan Respublikasında 2013-cü ilin “İnformasiya – kommunikasiya texnologiyaları ili” elan edilməsi ilə bağlı tədbirlər planı çərçivəsində Prezident İlham Əliyev tərəfindən imzalanan 2013-cü il 28 mart tarixli sərəncamda elm və texnikanın inkişafı çərçivəsində Vikipediyanın da inkişaf etdirilməsi ilə bağlı xüsusi bənd verilmişdir [20]: “Azərbaycana dair kontentin və Vikipediya kimi qlobal informasiya resurslarında Azərbaycan Respublikası ilə əlaqədar məlumatların inkişaf etdirilməsi üzrə tədbirlərin görülməsi (Azərbaycan Milli Elmlər Akademiyası, Təhsil Nazirliyi, Rabitə və İnformasiya Texnologiyaları Nazirliyi, qeyri-hökumət təşkilatlarını cəlb etməklə)”.

Vikipediyanın Azərbaycan dilində seqmentinin zənginləşdirilməsində, Azərbaycan dilində informasiya resurslarının virtual məkanda geniş yayılmasına xidmət edərək Azərbaycan dilinin qorunmasına, Azərbaycanın maraqlarının dünyada təbliğinə xidmət edən alimlər, tələbələr, media işçiləri ilə yanaşı yeni İnternet texnologiyalarından səmərəli istifadə etməyi bacaran mütəxəssislərə böyük ehtiyac vardır.

Vikipediyanın Azərbaycan dili seqmentində əsas problemlər aşağıdakılardır:

- Qrammatik və orfoqrafik səhvlər çoxdur;
- Maşın tərcüməsindən istifadə olunduğuna görə bəzi məqalələrdə məlumat düzgün verilmir;
- Məqalələrdə tez-tez əcnəbi sözlərdən istifadə olunur;
- Bəzən coğrafi adlar və tarixi abidələrin adları düzgün verilmir və ya bilərəkdən təhrif olunur;
- Əksər məqalələrin məzmunu qənaətbəxş deyil;
- Alim və mütəxəssislərin biliyinə, təcrübəsinə ehtiyac vardır.

Azərbaycan dilinin qorunması və geniş yayılması üzrə xüsusi proqramların hazırlanması və tədbirlərin görülməsində viki-texnologiyalarından istifadə edilməsi Azərbaycanın maraqlarının dünyada qorunması, ölkədə informasiya təhlükəsizliyi ilə əlaqədar görülən tədbirlərin tərkib hissəsidir. Vikipediya Azərbaycan dilinin qorunması və təbliği,

Azərbaycanın mədəniyyətinin, tarixinin tanınması ilə bağlı elm, mədəniyyət və təhsil müəssisələrinin üzərinə bir sıra vəzifələr düşür. Bu vəzifələr aşağıda verilmişdir:

- Elm, mədəniyyət və təhsil müəssisələri Vikipediya ensiklopediyasının Azərbaycanla bağlı əhəmiyyətli məlumatlarla zənginləşdirilməsi, nəzarətdə saxlanması, qorunması, təbliğində fəal iştirak etməlidir. Bu məqsədlə viki-texnologiyalarını yaxşı bilən xüsusi viki-qruplar hazırlanmalıdır;
- Vikipediyada və ümumiyyətlə İnternetdə Azərbaycan maraqları ilə əlaqədar gedən proseslər öyrənilməli, bu barədə əhalinin maarifləndirilməsi və məlumatlandırılması işləri aparılmalıdır;
- Vətəndaşların, əsasən də gənclərin Vikipediya layihələrində fəaliyyətinin aktivləşdirilməsi üçün lazımi tədbirlər görülməli, şərait yaradılmalıdır;
- Elmi müəssisə, universitet və müxtəlif təşkilatların əməkdaşlarının Vikipediyada iştirakını səmərəli etmək üçün zəruri biliklərin tədrisi və praktiki bacarıqların artırılması məqsədi ilə xüsusi kurslar yaradılmalıdır;
- Vikipediya layihələrindəki mövcud problemlərin aşkarlanması və həllinə diqqət artırılmalı, bununla əlaqədar elmi konfranslar, görüşlər və seminarlar təşkil olunmalıdır;
- Azərbaycan dilində kitabların, jurnalların və digər çap materiallarının, həmçinin müxtəlif elektron resursların Vikipediya layihələrində işıqlandırılması üçün elm

və təhsil müəssisələrində, kitabxanalarda maddi-texniki baza yaradılmalı, viki-qrupların səmərəli işinin təşkili həyata keçirilməlidir.

NƏTİCƏ

Hər bir dilin qorunması, onun daha da zənginləşdirilməsi işində virtual ensiklopediyaların, ilk növbədə isə Vikipediyanın daha kütləvi hal alması vacib şərtlərdəndir.

Dünya klassik ədəbiyyatının, texniki, ümumiyyətlə sahə ədəbiyyatının dilimizə tərcümə olunaraq Vikipediya ensiklopediyasına daxil edilməsi Azərbaycan dilinin zənginləşdirilməsində əvəzolunmaz işdir. Belə ki, gənc nəsil – şagirdlər, tələbələr təhsil prosesi zamanı, eləcə də müəllimlər lazım olan hər hansı bir məlumatı əldə etmək istədikdə xarici dillərdə olan saytlara yox, elə öz doğma ana dilimizdə olan Vikipediya ensiklopediyasına müraciət edərək, daha dolğun və geniş informasiya əldə edə bilsinlər. Son zamanların müşahidələri göstərmişdir ki, Vikipediya ensiklopediyasına tələbə və gənclərlə yanaşı dilçilərin, tarixçilərin, siyasətçilərin maraqları artmaqdadır.

Bütün deyilənlərdən bu nəticəyə gəlmək olar ki, qloballaşma və inteqrasiya şəraitində milli azlıqların dillərinin, eləcə də az istifadə olunan dillərin qorunmasında və yayılmasında viki texnologiyaları çox uğurlu vasitədir.

İSTİNADLAR

- [1] Ələkbərova İ.Y. Vikipediya virtual ensiklopediyasının əsasları. Bakı: “İnformasiya texnologiyaları” nəşriyyatı, 2017, 205 s.
- [2] Казарин В., Современный мир: глобализация и проблемы языкового многообразия // 2002, <http://russedina.ru/frontend/heading/time?id=2502>;
- [3] Хarrison D., When Languages Die // ABS, Pensillvaniya, 2007, 304 p.
- [4] Мир теряет языки, 2007, <http://www.akzia.ru/news/26-09-2007/655.html>;
- [5] Гафарлы М, 90% языков мира под угрозой исчезновения // "Новые известия", 2004;
- [6] Когда умирают языки, 2007, <http://www.bigbook.ru/litnews/detail.php?ID=3315>;
- [7] Павленко Ю.В., Культурно-языковые противоречия глобализации // “Зеркало недели”, № 32, 2004;
- [8] Kristall D., Английский язык как глобальный // İngilis dilindən tərcümə, “Весь мир”, М., 2001;
- [9] Lewis, M. Paul, Ethnologue: Languages of the World, Sixteenth edition. Dallas // Tex.: SIL International, 2009, Online version: <http://www.ethnologue.com/>;
- [10] International Mother Language Day, 21 February 2009 // 2008, <http://www.unesco.org/>;
- [11] Rics Smits, Yox olmaqda olan dillərin müdafiəsində UNESCO-nun fəaliyyəti, // İngilis dilindən tərcümə, 2007, <http://lingsib.unesco.ru/>;
- [12] Bədəlov Ə. Azərbaycan dilinin dövlət dili kimi rəsmiləşdirilməsi və inkişafı ümummilli lider Heydər Əliyevin adı ilə bağlıdır, Xalq qəzeti, 2009, http://www.anl.az/down/meqale/xalqqazeti/xalq_qazeti_avqust2009/87735.htm
- [13] Azərbaycan dilində latın qrafikası ilə kütləvi nəşrlərin həyata keçirilməsi haqqında, 2004, <https://e-qanun.az/framework/5427>
- [14] Leuf B., Cunningham W. The Wiki Way: quick collaboration on the Web, PA: Addison-Wesley, 2001, 200 p
- [15] Sumi R., Yasseri T., Rung A., Kornai A., Kertész J., Edit wars in Wikipedia / Privacy, Security, Risk and Trust (PASSAT), Proceedings of the IEEE Third International Conference on Social Computing, 9-11 Oct. 2011, pp. 724–727.
- [16] <http://www.wikipedia.org>;
- [17] Гүриев В., Цифровая Вселенная Ларри Сэнгера // журнал “Компьютерра”, №7, 2006.
- [18] Lewoniewski W., Węcel K., Abramowicz W. Multilingual Ranking of Wikipedia Articles with Quality and Popularity Assessment in Different Topics // Computers, 8(3):60, pp. 1–31.
- [19] https://meta.wikimedia.org/wiki/List_of_Wikipedias
- [20] “Azərbaycan Respublikasında 2013-cü ilin “İnformasiya-kommunikasiya Texnologiyaları ili” elan edilməsi ilə bağlı Tədbirlər Planı”nın təsdiq olunması haqqında, 2013, <https://e-qanun.az/framework/25505>

İnternet mühitində azərbaycandilli reklam resurslarının formalaşması xüsusiyyətləri haqqında

Kəmalə Haşımova

İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
hashimovakama@gmail.com

Xülasə — Reklam dünyada yeni məhsul və xidmət haqqında məlumat əldə olunması üçün bir vasitədir. İstifadəçilərinin sayı ildən-ilə artan reklamın cəlbediciliyi dil üslubunun düzgün seçilməsindən asılıdır. Burada reklam və marketing fəaliyyətində ifadə ediləcək sözlərin düzgün seçilməsi üçün təkliflər işlənmişdir.

Açar sözlər — reklam; alıcı; dil; marketing, informasiya

I. GİRİŞ

İnformasiya Kommunikasiya Texnologiyalarının inkişafı bir çox sahələrlə yanaşı reklam sahəsinə də öz təsirini göstərmişdir. Müasir üsulla və süni intellektə əsaslanan yeni texnologiyalar iqtisadiyyatın inkişafında əsaslı rol oynayır. Reklamın inkişafında müasir texnologiyalardan istifadə etdikdə alıcı və satıcı arasında münasibət daha şəffaf karakter alır.

Reklamın bir çox növləri vardır: Məhsul-orientasiyalı reklam, müqayisəli reklamçılıq, inandırıcı imic, təbliğat, sms reklam, qəzet və jurnallarda yayılan reklam, televiziya reklam və s. Bundan başqa onlayn məkanda internet reklamın

çoxsaylı növləri: konteks, sms, rich media, spam xarakterli reklam, viral, infomercials (reklam çarxları) və s. vardır. Hələ daş dövründən bugünkü günə kimi təkamül dövrü keçmiş reklam daim inkişafdadır.

II. REKLAMDA YENİ TEKNOLOGİYALAR

Reklamlar yeni texnologiyanın tətbiqi ilə iqtisadiyyatın inkişafının sürətləndirilməsinə xidmət edir. Bunun üçün reklam baxımlı, gözəl tərtibatlı olması ilə yanaşı dili də aydın, səliqəli və anlayışlı olmalıdır. Reklam dilinin cəlbedici, inandırıcı olub-olmaması məsələsi daim diqqət mərkəzindədir. Reklamın uğuru onu necə başa düşülməsi və piar etdirilməsi ilə əlaqəlidir. Dünya üzrə bu mövzuda araşdırmalar davamlı olaraq aparılır. Reklam edənlər və marketoloqlar bu sahəyə İnformasiya Texnologiyaları sahəsindəki yenilikləri (süni intellekt, maşın öyrənməsi (machine learning), böyük verilənlər və s.) cəlb edirlər. Bununla yanaşı tədqiqatçılar bu mövzuda yeni metod və təkliflər irəli sürürlər. Reklam üzrə bir çox modellər mövcuddur ki, buna misal olaraq Aida,

Acca, 4P modellərini göstərə bilərik. Bu modellərdə də digərləri kimi birinci növbədə müştərinin diqqətini cəlb etmək durur. Bu mərhələdə dizaynla yanaşı dil heç də az rol oynamır.

Təqdim edilən mövzuda bir çox məqalələr çap edilmişdir: məsələn, Belarus Dövlət İnformatika və Radioelektronika Universitetində S.Musavinin müəllifi “Nümunə üzrə müasir “facebook” və “vkontakte” sosial şəbəkələrində reklam, internet reklamının dili” məqalə eləcə də, Sürix Universitetindən A.Yanoshkanın “Veb reklam: İnternetdə yeni ünsiyyət formaları” adlı kitabını və s. göstərmək olar [1, 2]. Məqalədə məşhur “Facebook” və “VKontakte” sosial şəbəkələrində kontekstli reklam mətnləri timsalında müasir internet reklam dilinin xüsusiyyətləri müzakirə olunur. Kontekstli reklam bannerlərinin mətninin ən çox istifadə olunan sintaktik fiqurları və şifahi ifadə vasitələrinə misallar verilir, həmçinin reklam mətnlərində müstəqil nitq hissələrinin qrammatik xüsusiyyətləri öyrənilir.

Kitabda İnternetin interaktiv imkanları, xüsusən də onlayn reklamlar və veb-reklamlar vasitəsilə yaranmış yeni ünsiyyət formalarını araşdırır. O, kütləvi kommunikasiya və şəxsiyyətlərarası ünsiyyəti birləşdirən yeni onlayn kommunikasiya modelini hazırlayır. İnteraktiv kütləvi kommunikasiya hipermətn məlumatlarının vahidləri baxımından daha yüksək mürəkkəblik dərəcəsi ilə üzləşən onlayn kommunikasiya

tərəfdaşlarının rollarını yenidən müəyyənləşdirir.

Veb reklamında interaktivliyin bu yeni aspekti linqvistik olaraq müxtəlif növ şəxsi ünvan formalarında, direktivlərdə və sövqedici sözlərində əks olunur. Bu tədqiqat həmçinin veb reklamların aktivləşdirilməsinə başlamağa cəhd etdiyi müxtəlif inandırma strategiyalarını təhlil edir.

Veb Reklamı hipermətn linqvistikası, reklam, kommunikasiyalar və media tədqiqatları sahəsində alimlər, tədqiqatçılar və tələbələr üçün veb reklamının dili haqqında mühüm məlumatlar təqdim edir.

III. AZƏRBAYCANDA REKLAM

Dünya ölkələri ilə yanaşı Azərbaycanda da reklamın özünə məxsus rolu vardır. Daim inkişaf edən bir ölkə kimi burada reklamın istər adi, istərsə də onlayn məkanda müasir texnologiya ilə ayaqlaşması hiss olunur. Bu sahədə ciddi qaydalar mövcuddur. 15 may 2015-ci ildə Azərbaycan Respublikasının Prezidenti İlham Əliyev tərəfindən imzalanmış: Reklamda “Dövlət dili haqqında” Azərbaycan Respublikası Qanununda deyildiyi kimi “Reklam qurğularında yerləşdirilən reklamın mətnində Azərbaycan Respublikasının dövlət dili ilə yanaşı digər dillərdən istifadə oluna bilər. Bu halda xarici dildə olan mətn dövlət dilində olan mətndən sonra yerləşdirilməli, dövlət dilində olan mətnin tutduğu sahədən böyük olmamalıdır. Mətnində digər dillərdə istifadə olunan əmtəə nişanı və coğrafi göstərici müvafiq icra hakimiyyəti orqanının müəyyən etdiyi orqan (qurum)

tərəfindən qeydə alındığı formada göstərməlidir [3]. Həmin qanunvericilikdə bu illərə kimi müxtəlif müddəalarda kiçik dəyişikliklər olması dövrün tələbi ilə əlaqələndirilir. Reklamda olan qaydalara riayət edilməsinə xüsusi qurumlar nəzarət edir.

IV. REKLAM İFADƏLƏRİNİN BAĞLILIĞI

Reklamın dəqiqliyi və effektivliyi reklam verən tərəfindən tərtib edilmiş sözlərdən yerində istifadə edilməsindən asılıdır. Məhsul və ya xidmət haqqında informasiyanı auditoriyaya dəqiq və canlı çatdırmaq üçün mənalı, lokanik sözlərdən istifadə olunur. Burada reklam dilini ədəbi baxımdan araşdırmaq lazımdır. Əsasən iki cür məna vardır:

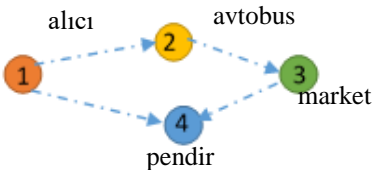
- denotativ
- konnotativ

Denotativ məna hərfi və ya adi lüğət mənasıdır, konnotativ məna isə qiymətləndirici, nəzərdə tutulan və ya emosional mənadır [4]. Eyni lüğət sözlərini işlədən insanlar üçün təyinatlar təxminən eyni olacaq, müxtəlif insanlar üçün isə fərqli mənalara mövcuddur. Burada dəyərlər istifadə olunduğu kontekstdən, göndərən, qəbul edən arasındakı əlaqədən və bir çox digər dəyişənlərdən asılıdır. Deməli, sözün denotasiyası açıq hərfi məna, konnotasiya isə onun bağlı olduğu mənadır. Sözlər çox vaxt güc, etibarlılıq, mükəmməllik, şöhrət və bu kimi mənalara ifadə edən keyfiyyətlərdir. Reklam dilinin konnotativ olduğunu deyə bilərik.

Semantika isə dilin mənalara sistemini öyrənir və yanaşmaları çox müxtəlifdir. İlk növbədə məna dillə xarici aləm (referensial və ya denotativ məna), sözlə onun işarə etdiyi anlayış arasındakı əlaqədir. İkincisi, bir sıra şəxsi, emosional tonlarda (affektiv və ya konnotativ məna) əks olunan danışanın psixi vəziyyəti ilə əlaqələndirilir. Semantika reklamı lazımi yüksəkliyə çatdıran linqvistik formalardan biridir. Sözlərin və assosiasiyaların seçiminin unikallığı dil reklamını ixtisaslaşdırılmış tədqiqat sahəsinə çevirir. Müəyyən edilmiş dilin kodu əsasında müxtəlif söz manipulyasiyaları, düzlülüşi, söz seçimi, söz birləşməsinin sırası, reklam məkanlarının denotativ, konnotativ mənalara və digər kənarlaşmalarla hərtərəfli yoxlanılır. İdeya dilin fakturasını, mənasını və anlayışını daha yaxşı başa düşmək, qiymətləndirməkdir. Reklamın nitq seçimi müstəsna bacarıq tələb edir. Bu, müəyyən vəziyyətə uyğun sözlərin düzgün uzlaşmasına kömək edəcək. Səhv leksik seçim ya ünsiyyətin məqsədini qeyri-dəqiq təhrif edəcək, ya da ən yaxşı halda nəzərdə tutulan fikri qeyri-müəyyən və ya qismən çatdıracaq [5].

Reklam olunan məhsul alıcıya aydın dildə izahat edilməli və ya tez-tez rast gəlməklə təsir edilməlidir. Eləcə də reklam mətni tərtib olunan zaman vulqar sözlərə yer verməməli və yaxud başqa dillərə tərcümə zamanı təhqiredici səslənməməlidir. Reklamın hansı məkanda olmasından asılı olmayaraq ilk növbədə ədəbi dil normaları gözlənilməlidir. Dil mütəxəssisləri reklam müəssisələri ilə

müntəzəm əlaqə qurmalı, reklam agentləri ilə məsləhətləşməlidirlər. Reklamda mətnlər obrazlı, aydın, lakonik olmalıdırlar. Mətnlərin strukturu, məna və üslubi keyfiyyətləri düzgün seçilməlidir. “Alıcı avtobusa əyləşib, digər qəsəbəyə getdi və oradakı mağazadan pendir aldı”. Bu cümlədə 4 obyekt vardır: alıcı (1), avtobus (2), digər qəsəbədəki mağaza (3) və pendir (4). Burada əsas məqsəd digər qəsəbədə marketin olduğu və ya qarşılaşdırılmış pendirə diqqətin yönəldilməsidir.



Şəkil 1. Reklam edilən obyektlər arası əlaqə

Baxdığımız mətni bu şəkildə də ifadə etmək olar: “Alıcı avtobusa əyləşdi və avtobusla digər qəsəbəyə getdi. Oradakı marketdə pendir vardı. Pendir xüsusi qabda idi. Alıcı pendir qabını götürdü”. Göstərdiyimiz nümunədə ikinci cümlə, yorucu olduğuna görə, müştəri cəlb etmə ehtimalı demək olar ki, yox dərəcəsindədir. Bu səbəbdən reklam oluna bilməz [6].

V. İNTERNET MARKETİNQ

Hal-hazırda bəşəriyyətdə inkişafın növbəti mərhələsi–informasiya cəmiyyətinin inkişaf mərhələləri həyata keçirilir. İnternetin meydana gəlməsi ilə ticarətin yeni məkanda sürətli inkişaf mərhələsinə keçməsi, müasir konsepsiyasında fərqli mərhələyə qədəm

qoydu. İnternetdə reklamın inkişaf etməsi İnternet marketing fəaliyyəti üçün böyük imkanlar yaratdı. İnternet-marketing artıq istənilən müasir şirkətin fəaliyyətinin ayrılmaz hissəsinə çevrildi. İnternet sayəsində istehsalçılar alıcıların tələbatları, istəkləri, zövqləri, imkanları barədə daha operativ və dəqiq informasiya əldə etməsini təmin edir. Marketing istehlakçı ilə istehsalçı arasında sıx kommunikasiya əlaqələri yaradaraq, təşkilat, şəxsiyyət arasında qarşılıqlı anlaşmanı və əməkdaşlığı dəstəkləməyə kömək etməkdir [7].

Müasir dövr texnika əsri olsa da öz təbələri vardır. Onlayn mühitdə alıcı ilə satıcı arasında rabitə yaradan vasitələrdən biri şirkətin veb – saytıdır. Saytın cəlbedici olması, sürətli işləməsi, rahatlığı və müştərini maraqlandıran məhsul haqqında informasiyanın ön planda yerləşməsi əsas amillərdən biridir. Şirkət haqqında qısa, məhsul haqqında isə geniş məlumat, qiymət cədvəli və əks əlaqə yerləşdirilir. Əgər yeni məhsul, onların faydaları, endirimlər haqqında mütəmadi məlumat saytda yerləşdirilərsə coxsaylı alıcı cəlb etmiş olur. Azərbaycanda bir çox informasiya saytları müqavilə əsasında reklam yerləşdirirlər. Lakin bu istifadəçiyə mane olmamalıdır.

Hər bir sahədə yeniliklər tətbiq olunur və bu barədə məlumat verilir. İcad edilmiş hər bir yeni məhsulun üzərində ondan istifadə qaydaları göstərilir. Nəzərə alınsa ki, bu məhsul digər ölkələrə ixrac edilir, deməli, istifadə qaydaları daha bir neçə dildə təqdim olunmalıdır. Bu zaman ən çox istifadə olunan dillər mütləq şəkildə bu

siyahıda yer alır. Məhsula tələbatı artırmaq məqsədi ilə həmin dildə məhsul haqqında reklam təqdim edilməlidir. Daha sonra marketinqin düzgün təşkili çox önəmli məsələlərdir. Reklam yerləşdirərkən istehlakçıda inam yaratmaq üçün informasiya asimetriyasını aradan qaldırması vacib şərtidir.

Məhsul haqqında reklam hazırlanarkən asimetriya ilə yanaşı, psixoloji təsir də önəmli yer tutur. Müasir dövrdə psixologiya və dilçilik arasında sıx əlaqə geniş yer tutmağa başlamışdır.

Reklam sahəsini, müştəri fikrini məhsulu almaq üçün kifayət qədər uzun müddətdə bu istiqamətə yönəldən bir elm kimi təqdim etmək olar [8]. Beləliklə, reklamın bugünkü iş dünyasında həlledici rol oynadığı aydın olur. Məsələ reklam dillərinin cəlbedici və inandırıcı olub-olmamasıdır. Beləliklə, Azərbaycan dilində reklamın dil xüsusiyyətlərinin hərtərəfli öyrənilməsi, uğurlu reklamın necə başa düşülməsini və inkişaf etdirilməsini yaxşılaşdırmağa kömək edir.

İnternet məkanında tez-tez reklamların nümayişi adətən izləyiciləri yorur. Belə halların qarşısını almaq üçün xüsusi proqramlar mövcuddur. Adblock Plus, AdBlocker Ultimate, AdGuard və s. nümunə olaraq bu kimi bloklayıcı proqramları göstərmək olar.

Nümayiş etdirilən reklamda Azərbaycan dilindən başqa digər dillərə olan sözlərdən də istifadəyə rast gəlinir. Problem olaraq qeyd etmək olar ki, alınma sözlər heç bir səbəb olmadan bəzən orijinalda olduğu kimi yazılır. Xüsusilə mərkəz şəhərlərdə

fırma, mağaza, restoran, və s. adları əcnəbi dillərdə yazılır, bununla da dil qanunları pozulur.

Araşdırmalar onu da göstərir ki, qrammatik qaydalara heç də həmişə riayət olunmur. Reklam və elanlarda Azərbaycan dilinin qrammatik qaydaları bəzən pozulur, milli xüsusiyyətə əməl edilmir. Azərbaycan dilində olan adlar “ə” hərfinin “e” hərfinə çevrilməsi ilə yazılır: “Mermer”, “Gözəllik salonu”, “Perde evi” və s. əcnəbi sözlərdən istifadə edilir. Televiziya reklamları adətən xarici dövlətlərin reklamlarının dublyajı ilə nümayiş etdirilir. Bu, heç də həmişə uğurlu alınmır. Bu kimi neqativ halların aradan qaldırılması üçün dövlət qurumları xüsusi tədbirlər həyata keçirirlər.

NƏTİCƏ

Son dövrlərdə rəqəmsal iqtisadiyyatın inkişafı reklam və marketinqin müasir mərhələyə qədəm qoyması ilə bərabər, bu sahədə yaranan çətinlikləri də aşkarlayıb aradan qaldırmağa yardım edir. Daim inkişaf edən ölkələr sırasında olan Azərbaycan öz dilini qoruyub saxlamaqdadır. Hər bir sahədə olduğu kimi reklam və marketinq sahəsində çalışan mütəxəssislər qaninvericilik tələblərinə əməl etməli, Azərbaycan dilində olan zənqin sözlərin düzgün ifadə edilməsinə, mədəniyyətimizin qorunub saxlanmasına diqqət etməlidirlər.

İSTİNADLAR

- [1] С.С. Мусави, “Язык современной интернет-рекламы на примере рекламы в социальных сетях “facebook” и “вконтакте””, 56-я Научная Конференция Аспирантов,

- Магистрантов и Студентов БГУИР, Минск, 2020, с.81-83
- [2] A. Janoschka, “Web Advertising: New forms of communication on the Internet”, 2004, pp.230
- [3] “Azərbaycan Respublikasında dövlət dili haqqında” Qanun <http://www.e-qanun.az/framework/1865>
- [4] A. Məmmədov, “Reklamlarımızın dili”, 2019
- [5] Н. Х. Уразаева, Язык современной интернет-рекламы на примере рекламы в социальных сетях Филология и лингвистика в современном обществе: материалы II Междунар. науч. конф. г. Москва, 2014.: <https://moluch.ru/conf/phil/archive/107/4157/>
- [6] Т. С. Мачушенко, Метафора в рекламных текстах, Молодой ученый 2017, – № 12 (146), <https://moluch.ru/archive/146/40984/>
- [7] К.К. Həşimova, İnternet mühitində reklam problemləri və onların həlli yolları, / Ekspress–İnformasiya, İnformasiya cəmiyyəti seriyası, Bakı, 2014, s.58.
- [8] <https://stm.az/ru/news/91/reklamlarımızın-dili>

Milli transliterasiya sisteminin formalaşması məsələləri

Səbinə Məmmədzadə

İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
sabinamammadzadeh@gmail.com

Xülasə — Transliterasiya əsasən sözlərin fonoloji quruluşunu qorumaq məqsədi daşıyır. Maşın tərcüməsinin keyfiyyətinin yüksəldilməsində ad və soyadların, o cümlədən toponim və qurum adlarının düzgün transliterasiyası çox mühüm rol oynayır. Təqdim edilmiş tədqiqat işində Azərbaycan dili üçün istifadə edilən latın qrafikasının digər dillərin qrafikaları ilə transliterasiyanı həyata keçirən proqram təminatının formalaşması məsələlərinə baxılmışdır. Milli Transliterasiya Sisteminin üstün və çatışmayan cəhətləri və gələcək inkişaf perspektivləri şərh edilmişdir.

Açar sözlər — transliterasiya; konvertasiya cədvəlləri; qrafem; fonem

I. GİRİŞ

Maşın tərcüməsi və xarici dildə məlumat axtarışının nəticəsi əsasən transliterasiyadan asılı olur. Transliterasiya prosesi bir dildə olan sözlərin həmin dilin təxmini fonetik və ya orfoqrafiya ekvivalentlərindən istifadə etməklə digər dildə təmsil edilməsini əhatə edir [1; 2].

Son illərdə bütün proseslərin virtual mühitə keçməsi ilə əlaqədar olaraq transliterasiyanın da avtomatlaşdırılması sahəsində geniş tədqiqatların aparılmasına zəmin yaranmışdır. Hazırda qrafik sistemlərinin əksəriyyətinin latın

əlifbasından istifadə etməsinə baxmayaraq, onlar müxtəlif müasir dillərdə, adətən, fərqli fonemləri ifadə edir. Bu səbəbdən də, maksimal dərəcədə transliterasiya dəqiqliyinə nail olmaq üçün müasir metodların yaradılması zəruridir. Bu baxımdan, Azərbaycan dili üçün istifadə edilən latın qrafikasının digər dillərin qrafikaları ilə transliterasiyası üçün metodların işlənməsinə ehtiyac yaranmışdır.

II. “MİLLİ TRANSLİTERASIYA SİSTEMİ”NİN YARADILMASININ NORMATİV-HÜQUQİ ƏSASLARI

Hazırda Azərbaycan dilində yazılan ad və soyadlar şəxsiyyət vəsiqələrində vahid formada transliterasiya olunsada bəzi qaydalara tam olaraq riayət edilmir və yaxud fonetik incəliklərə fikir verilmir. Belə ki, “ə” hərfi ingilis dilinə tərcümə zamanı yalnız “a” hərfi ilə əvəzlənir, “e” hərfi ilə əvəzlənmənin mümkünlüyü isə ümumiyyətlə nəzərə alınmır və s. Azərbaycan Respublikası Prezidentinin 09 aprel 2013-cü il tarixli 2837 nömrəli Sərəncamı ilə təsdiq edilmiş “Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsi və ölkədə dilçiliyin inkişafına dair Dövlət Proqramı”nın [3] 6.1.4 və 6.3.9. bəndləri

müvafiq olaraq “Milli transliterasiya standartlarının hazırlanması” və “Milli transliterasiya standartları əsasında Azərbaycan əlifbasından digər əlifbalara transliterasiyanı həyata keçirən proqram təminatının hazırlanması”-ni diktə edir. Hər iki bəndin icrası məqsədlə müəyyən sahənin mütəxəssisləri tərəfindən işlər aparılmışdır. Buna baxmayaraq, bir sıra məsələlər hələ də tam həllini tapmamışdır.

Azərbaycan dili qrafikasının transliterasiyası sahəsində mövcud vəziyyət

Hal-hazırda transliterasiya problemi qlobal virtual məkanın çoxdilli xarakteri ilə əlaqələndirilir. Bununla belə, dünyada itməkdə olan dillərin sayı artmaqdadır. Bu, əsasən, latın qrafikasının digər qrafikalarla müqayisədə üstünlük təşkil etməsi ilə bağlı ola bilər. Hətta latın qrafikasından istifadə etməyən dillər belə romanizasiyadan istifadəyə üstünlük verir. Romanizasiya prosesi isə müəyyən bir dilin qrafik və fonetik xüsusiyyətlərini geniş şəkildə əks etdirən avtomatlaşdırılmış dillərarası transliterasiya sistemi ilə təmin edilməlidir.

III. MİLLİ TRANSLİTERASİYA SİSTEMİNİN LİNGVİSTİK PROBLEMLƏRİ

Azərbaycan dili qrafikasının digər dillərin qrafikaları ilə avtomatik transliterasiyasını təmin etmək məqsədilə ilk növbədə Azərbaycan dilinin qrafik və fonetik xüsusiyyətləri araşdırılmalı, Azərbaycan dilinin digər dillərlə fonetik və qrafik uzlaşması öyrənilməlidir. Milli Transliterasiya Sisteminin lingvistik problemlərinə ilk növbədə müxtəlif dillərin

fərqli qrafik xüsusiyyətlərə malik olması aiddir:

- Müəyyən hərflərin əksər dillərdə tanınmaması. Latın qrafikasının istifadəsinə baxmayaraq, Azərbaycan dili qrafikası ingilis, fransız, alman və digər geniş yayılmış dillər üçün qəbul edilən latın qrafikalarından fərqlənir. Azərbaycan latın hərflərinin 7 hərfi (ç, ğ, ı (“I” üçün kiçik hərflər), İ “i” üçün böyük hərflər), ö, ş, ü dəyişdirilmiş və xüsusi olaraq bir hərflər (ə) əlavə edilmişdir.
- Bəzi hərflərin qarşılıqlı olmaması səbəbindən olduğu kimi saxlanması, məs., ü, ö və s.
- Bir neçə hərfin eyni səsi verməsi ilə əlaqədar olaraq yalnız birinin seçilməsi, məs., z səsinə verən fars dilində dörd hərflər mövcuddur, bu isə subyektiv fikirlərdən asılıdır.
- Bir hərfin bir neçə səsi bildirməsi, məs., fars dilindəki əlifbədə a, e, ə, bəzən u və i səslərini də verə bilər.
- Hərflərin bir neçə formada yazılması (fars dilində): əvvəldə, ortada, sonda və ayrılıqda. Bu isə əksinə transliterasiyanı mümkün edir.

Əslində, transliterasiya prosesində bu hərflər yalnız öz qrafik təsvirləri ilə deyil, həm də ingilisdilli auditoriya üçün tələffüz baxımından da çox mürəkkəb hesab edilir. Bu hərflər ž, j, ŷ, î, û, š, č, ê kimi qəbul edilmiş transliterasiya cədvəllərində Azərbaycan oxucusu üçün də anlaşılmaz diakritik işarələrlə təmsil olunduqda bu problem daha da mürəkkəbləşir.

Digər məsələ ad, soyad və toponimlərin transliterasiyası üçün sistemin etibarlı olmaması ilə əlaqəlidir. Buna bir neçə səbəb göstərmək olar:

- Ad, soyad və toponimlərin transliterasiyası üçün xüsusi standartların işlənməməsi;
- Ekspertlərin fikir ayrılığının olması;
- Bəzi hərflər üçün konkret strukturun olmaması, məs., rus dilinə transliterasiya zamanı ö və ü hərfləri üçün ümumi qəbul olunmuş qayda mövcuddur. Ö hərfi əvvəldə - O ilə, ortada isə Ę ilə əvəz olunur: Ömər – Омар, Gözəl - Гѐзьяль, Ə hərfi üçün isə bu qayda mövcud deyil: Ə hərfi a, e, я və э hərfləri ilə əvəzlənə bilər: Əsmər – Эсмер, Heydər – Гейдар
- Toponimlərin indiyədək bir neçə variantlarda yazılması (buna rus dilinin uzun illər göstərdiyi təsiri, sözün mənası və siyasi məqsədlər səbəb ola bilər). Rayon və şəhər adlarının əksəriyyəti əvvəl sovet dövründə rus dilinin fonoloji xüsusiyyətləri ilə uyğunlaşdırılmış və daha sonra xarici dillərə çevrilmişdir. Məsələn, Zaqatala – Zagatala, Zakataly, Balakən – Balakan, Belokany, Belokan, Şamaxı – Shamakhi, Shemakha və Naxçıvan– Nakhichevan, Nakhichivan. Bu toponimlərin qlobal informasiya məkanında təhrif edilməsi milli identikliyin pozulması ilə nəticələnir. Bu təhriflər, bəzi hallarda, yalnız tərcümə dilinin qrammatik xüsusiyyətləri ilə deyil, həm də siyasi-ideoloji amillərlə də əlaqəli ola bilər.

IV. MİLLİ TRANSLİTERASIYA SİSTEMİNİN TEXNİKİ PROBLEMLƏRİ

Milli Transliterasiya Sisteminin texniki problemlərinə aşağıdakılar aid ola bilər:

- Bu sahədə Azərbaycan dilinin transliterasiyası üçün kifayət qədər metod və alqoritmlərin işlənməməsi (bunun üçün yalnız əvəzlənmə, silinmə və əlavə etmə qaydalarından istifadə olunmuşdur);
- Ad və toponimlər üçün vacib olan dil korpusunun işlənməməsi;
- Fars-Azərbaycan transliterasiyasının hələ işlənməməsi;
- Saytın transliterasiyasında problemlərin olması;
- Bu sahədə hər dil üzrə bir neçə mütəxəssisin cəlb edilməməsi.

Bütün qeyd edilənlərlə yanaşı Milli Transliterasiya Sistemi bir sıra üstünlüklərə malikdir. Belə ki, Milli Transliterasiya Sistemi Azərbaycanca ilk çoxdilli transliterasiya sistemidir; bu sistemdə yeni qəbul edilmiş (milli) transliterasiya standartları tətbiq edilir; Transliterasiya zamanı fonetik xüsusiyyətlər nəzərə alınır; Yalnız romanizasiya deyil, ingilis, alman, fransız, ispan və italyan dillərinin qrafikası ilə transliterasiya tətbiq edilir; Mövcud qaydaların (alqoritmlərin) dəyişdirilməsi imkanı vardır; Yeni qrafikaların əlavə edilmə imkanının olması; Yalnız mətnin deyil, bütövlükdə veb-saytın transliterasiyası da mümkündür və s.

V. MİLLİ TRANSLİTERASIYA SİSTEMİNİN İNKİŞAF PERSPEKTİVLƏRİ

NƏTİCƏ

Ümumilikdə, qeyd edilən problemləri həll etməklə Milli Transliterasiya Sisteminin bir sıra inkişaf perspektivlərinə aşağıdakıları aid etmək olar:

- Konvertasiya cədvəllərinin, həmçinin bütövlükdə sistemin standartlaşdırılması;
- Ad, soyad və toponimlər üçün bazanın hazırlanması;
- Fars-Azərbaycan transliterasiyası üçün yeni metodların işlənməsi (keçid transliterasiyadan istifadə və ya təlim bazasının yaradılması);
- Ərəb qrafikası üçün transliterasiya cədvəllərinin (müvafiq sahə mütəxəssisləri tərəfindən artıq işlənilib hazırlanmışdır) əlavə edilməsi;
- Toponimlər üçün xüsusi cədvəllərin işlənilib tətbiq edilməsi (BGN/PCGN [4] (ABŞ Coğrafi Adlar üzrə Şura və Birləşmiş Krallıqda rəsmi istifadə üçün Coğrafi Adlar üzrə Daimi Komitə (1993) standartından istifadə). Qeyd etmək lazımdır ki, Azərbaycan dilçiliyi sahəsində, xüsusilə də transliterasiya sahəsində ən mühüm elmi-tədqiqat işlərinin biri kimi AMEA Nəsimi adına Dilçilik İnstitutunun Hind-Avropa dilləri şöbəsinin müdiri, filologiya üzrə elmlər doktoru, professor İlham Tahirov tərəfindən “Azərbaycan Respublikasının ərazi vahidlərinin adları ingilis dilində” adlı kitab hazırlanmışdır;
- İctimai rəylərdən yararlanmaqla sistemin daim yenilənməsi.

Azərbaycan Respublikası Prezidentinin 09 aprel 2013-cü il tarixli 2837 nömrəli Sərəncamı ilə təsdiq edilmiş “Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsi və ölkədə dilçiliyin inkişafına dair Dövlət Proqramı”nın 6.1.4 “Milli transliterasiya standartlarının hazırlanması” bəndinin icrası məqsədi ilə AZS 890:2020 “Milli transliterasiya standartları” dövlət standartı qəbul olunmuşdur [5]. Standart 7 dilin (rus, ingilis, alman, fransız, ispan, italyan, fars) qrafikaları ilə transliterasiyanı əhatə edir. Qeyd edilən Dövlət Proqramının 6.3.9. “Milli transliterasiya standartları əsasında Azərbaycan əlifbasından digər əlifbalara transliterasiyanı həyata keçirən proqram təminatının hazırlanması” bəndinin icrasının nəticəsi olaraq Milli Transliterasiya Sistemi yaradılmış və 18 sentyabr 2020-ci ildə Azərbaycan Milli Elmlər Akademiyasının Humanitar Elmlər Bölməsinin və Fizika Riyaziyyat və Texnika Elmləri Bölməsinin birgə təşkilatçılığı ilə Milli Transliterasiya Sisteminin və “Transliterasiya” veb-portalının onlayn təqdimatı keçirilmişdir [6].

Qəbul edilmiş standartın müvafiq dövlət qurumlarında, xüsusilə də Azərbaycan ad və soyadlarının sənədlərdə düzgün və vahid formada tərcüməsi (transliterasiyası) ilə məşğul olan Azərbaycan Respublikası Daxili İşlər Nazirliyinin Baş Pasport, Qeydiyyat və Miqrasiya İdarəsində tətbiqi təmin olunmalıdır. Bu, Azərbaycan dilində yazılan ad və soyadlarının əcnəbi dillərə transliterasiyasının vahid formada həyata keçirilməsinin təmin edilməsinə və çoxvariantlılığın qarşısının alınmasına şərait

yaradacaq və müasir dövrün ən mühüm problemlərindən biri olan milli identikliyin qorunması məsələsinin də həllinə imkan yaradacaqdır.

İSTİNADLAR

- [9] Regmi K., Naidoo J., Pilkington P. Understanding the Processes of Translation and Transliteration in Qualitative Research. *International Journal of Qualitative Methods* (2010), 9(1)
- [10] Wehmeier, S., McIntosh, C., Turnbull, J., & Ashby, M. (2005). *Oxford advanced learner's dictionary of current English*. Oxford: Oxford University Press
- [11] Azərbaycan Respublikası Prezidentinin 09 aprel 2013-cü il tarixli 2837 nömrəli Sərəncamı ilə təsdiq edilmiş “Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsi və ölkədə dilçiliyin inkişafına dair Dövlət Proqramı” <https://e-qanun.az/framework/23583>
- [12] United States Board on Geographic Names (BGN) and the Permanent Committee on Geographical Names for British Official Use (PCGN) romanization standards www.translitteration.com
- [13] AZS 890:2020 “Milli transliterasiya standartları” dövlət standartı <http://azstand.gov.az/az/news/102>
- [14] Mammadzada Sabina 7th International Conference on Turkic Languages Processing (TURKLANG2020) Development of the National Transliteration System of Azerbaijan. Russia, Bashkortostan, Ufa. 18-19 October 2020. pp235-241

Linqvistik verilənlər bazalarının quruluş xüsusiyyətləri

Maya Heydərova

AMEA Nəsimi adına Dilçilik İnstitutu, Bakı, Azərbaycan

maya-heyderova@rambler.ru

Xülasə — Dilçilikdəki tədqiqatların nəticələrinin cəmiyyətdə aktiv istifadəsi bu sahədə rəşional şəkildə təşkil edilmiş informasiya resursları ilə birbaşa bağlıdır. Bu baxımdan da tədqiqatçıların lazımi məlumatlara girişini təmin etmək üçün elmdə və xüsusən dilçilikdə daha səmərəli informasiya sistemlərinin yaradılmasının zəruriliyi artır. Bu zaman informasiyanın daha effektiv saxlanması və emalı üçün verilənlər bazası (VB) texnologiyalarından istifadə etmək ən optimal variantdır. Verilənlər bazası kompüter texnologiyalarından istifadə edərək yaddaşda saxlanılan və idarə oluna bilən strukturlaşdırılmış verilənlər bazası yaratmağa imkan verir. Məqələdə fonetik, leksikoqrafik və qrammatik səviyyələrdə hazırlanmış linqvistik verilənlər bazaları və onların quruluş xüsusiyyətləri müzakirə olunur.

Açar sözlər — kompüter dilçiliyi; verilənlər bazası; fonetika; leksika; qrammatika

I. GİRİŞ

Linqvistik tipologiyanın inkişafı dilçilik sahəsində linqvistik verilənlər bazasının yaradılmasına səbəb olmuşdur. Linqvistik tipologiya – aralarındakı genetik əlaqənin təbiətindən asılı olmayaraq dillərin struktur və funksional xüsusiyyətlərinin müqayisəli tədqiqədir [1]. Linqvistik verilənlər bazasında müxtəlif dillərin xüsusiyyətləri cəmlənir. Buna görə də müqayisəli və

tipoloji dilçilikdə tədqiqatlar aparmaq üçün daha səmərəli vasitə olan linqvistik verilənlər bazalarının hazırlanmasına ehtiyac yaranmışdır.

Verilənlər bazası texnologiyası həm ənənəvi, həm də elektron lüğətlərin yaradılması prosesində istifadə olunur. Hal-hazırda tətbiqi dilçiliyin müxtəlif sahələrində informasiyanın kompüterdə verilənlər bazası şəklində təsvirindən istifadə olunur:

- Müqayisəli və tipoloji dilçilikdə, fonetik, leksik, qrammatik tədqiqatlarda;
- Tezaurusların tərtib edilməsi daxil olmaqla semantik tədqiqatlarda;
- Kompüter dilçiliyində və tətbiqi dilçilik məsələlərinin həllində, xüsusilə də:
 - Linqvodidaktika sahəsində tədqiqatlar;
 - Psixolinqvistika sahəsində tədqiqatlar;
 - Sosiolinqvistika sahəsində tədqiqatlar;
 - Kompüter vasitələrinin köməyi ilə dilin tədrisi;
 - Avtomatik tərcümə sistemlərinin yaradılması;
 - Avtomatik tanıma və nitqin sintezi;

– Dialektologiya sahəsində tədqiqatlar [2].

Hazırda dil verilənlərinin strukturlaşdırılmasına və sistemləşdirilməsinə imkan verən, onlara rahat əlçatanlığı təmin edən və axtarış imkanı verən verilənlər bazası idarəetmə sistemi (VBİS) geniş istifadə olunan vasitədir. Lakin bütün dilçilər VBİS-lə işləmə vərdişlərinə malik deyildir və nəticədə onlar heç də həmişə öz verilənlərini VBİS cədvəllərində saxlaya bilmir. Bunun üçün dilçi tədqiqatçıların istifadə edə biləcəyi interfeysdə verilənlər bazasını həm iyerarxik, həm də relyasiya modelinə uyğun hazırlamışlar.

Bu bazalar əsasən dilçiliyin üç səviyyəsini əhatə edir:

- Fonetik;
- Leksikoqrafik;
- Qrammatik.

II. FONETİK VERİLƏNLƏR BAZASI

Fonetik “nitq korpusu” müəyyən bir dildə toplanmış nitq fraqmentləri toplusundan və onların əsasında yaradılmış proqram təminatı ilə təmin olunan verilənlər bazasıdır.

Fonetik verilənlər bazasının istifadəsi ilə həll olunan dilçilik məsələləri:

- Nitqin fonetik xüsusiyyətlərinin öyrənilməsi;
- Nitqin avtomatik tanınması və sintezi sistemlərinin işlənməsi;
- Səsinə görə şəxsiyyətin identifikasiyası sistemlərinin işlənməsi [3].

Fonetik verilənlər bazasına nümunə kimi aşağıdakıları göstərə bilərik:

TIMIT (Texas Instruments, Massachusetts Institute of Technology) akustik-fonetik korpusu geniş fonetik tədqiqatlar üçün, habelə ingilis dilinin Amerika versiyası çərçivəsində avtomatik fasiləsiz nitq tanıma sistemlərinin hazırlanması və sınaqdan keçirilməsi üçün nəzərdə tutulmuşdur. Korpusda səslərin qeyd olunmasında ABŞ-ın 8 regional dialekt zonasından 630 diktör iştirak edir [4].

RuSpeech (<http://russpeech.spbu.ru/>) müəyyən mətnə kəsilməz rus nitqinin fraqmentlərini, fonetik transkripsiyasını və informatorlar haqqında əlavə məlumatları özündə birləşdirən nitq verilənlər bazasıdır. Hazırda RuSpeech hər tələffüz edilən cümlənin fonetik işarələnməsi olan 50 mindən çox cümləni ehtiva edir. Korpusun hazırlanmasında hər biri orta hesabla 250 cümlə söyləyən 220 diktör dəvət edilmişdir [4].

III. LEKSİKOQRAFİK VERİLƏNLƏR BAZASI

Leksikoqrafik verilənlər bazasına müxtəlif tipli elektron lüğətlər daxildir.

Elektron lüğətlər - xüsusi kompüter formatında hazırlanmış və ya daha mürəkkəb kompüter proqramlarının (maşın tərcüməsi sistemləri) tərkib hissəsi kimi insanların istifadəsi üçün nəzərdə tutulmuş sözlərin toplusudur. Elektron lüğətlərin iki növü vardır: tətbiqi proqram şəklində olan offlayn lüğətlər və İnternetdə olan onlayn lüğətlər.

Tətbiqi proqram şəklində olan lüğətlərdən istifadə etmək istəyən istifadəçilər bu proqramları telefonuna, kompüterinə yükləyib İnternetə çıxışı

olmadan istifadə edir. Bu tip lüğətlər kommersiya məqsədilə geniş şəkildə istifadə olunur. Belə lüğətlərə misal olaraq, geniş istifadəçi auditoriyasına malik olan və özündə 150-dən çox lüğət bazasını birləşdirən “ABBYY Lingvo” lüğətini göstərə bilərik. Bu lüğət 1996-cı ildə istifadəyə verilmiş və qısa müddət ərzində istifadəçilər arasında geniş yayılmışdır. Son illər müxtəlif təşkilatlar tərəfindən hazırlanan ödənişli və ödənişsiz proqram versiyaları geniş vüsət almışdır. Azərbaycanda bu tipli proqramlara nümunə kimi, Poliqlot və Dilmanc avtomatik lüğətlərinin proqram versiyalarını göstərmək olar.

İnternet lüğətlərindən internetə birbaşa çıxışı olan hər bir istifadəçi yararlanı bilər. İnternet lüğətləri proqram lüğətləri ilə müqayisədə daha çox istifadə olunur. Bir çox axtarış sistemlərində (portallarda) onlayn tərcümə lüğətləri yerləşdirilmişdir. Buna misal olaraq, google.com, yandex.ru, rambler.ru, mail.ru axtarış sistemlərini göstərmək olar. *Google* və *Yandex* kimi tərcümə lüğətləri 100-dən çox dildə, o cümlədən, Azərbaycan dilindən digər dillərə və yaxud da digər dillərdən Azərbaycan dilinə tərcümə həyata keçirən pulsuz onlayn lüğətdir. Avtomatik tərcümə lüğətlərinin müəyyən çatışmazlıqları da vardır. Məsələn, *Google* tərcümə sistemi istifadəçiyə xarici dildə olan mətnin ümumi məzmununu anlamağa kömək edir, lakin mətni dəqiq tərcümə edə bilmir və nəşr standartlarına uyğun məzmun təqdim etmir, məsələn, bir çox hallarda sözləri mətndən ayırı, kontekst nəzərə alınmadan tərcümə edir və qrammatik qaydalara əməl olunmur.

Azərbaycan dilinin ikidilli onlayn lüğətlərindən olan Dilmanc hazırda Azərbaycanda ən çox istifadə edilən tərcümə lüğətidir. Bundan başqa ikidilli və izahlı lüğətlərin elektron versiyasını özündə ehtiva edən Azleks, Azərbaycan dilinin orfoqrafiya lüğəti əsasında hazırlanan Azərdict, Azərbaycan dilinin orfoqrafiya və izahlı lüğətlərinin, şəxs adları və ixtisarlar lüğəti əhatə edən “Azərbaycan dilinin elektron lüğətlər korpusu” onlayn versiyada hazırlanmışdır (korpus.azerbaycandili.az) [2].

IV. QRAMMATİK VERİLƏNLƏR BAZASI

Etimologiya (sözün mənşəyini öyrənən dilçilik sahəsi) sahəsində tədqiqatlar aparmaq üçün hazırlanan “Babil qülləsi” linqvistik verilənlər bazasını nümunə kimi göstərə bilərik [5]. Bu verilənlər bazası müqayisəli-tarixi dilçiliyə həsr edilmişdir və internet üzərində onlayn olaraq fəaliyyət göstərir. Bazada Avrasiyanın bütün əsas dil ailələrinin materiallarını birləşdirən çoxlu etimologiyalar var. “Babil qülləsi”ndə bütün etimoloji bazalara giriş eyni qaydada qurulur, saytın açılan səhifəsində bazalar sadalanır, bazaların yaradıcıları və yaradılma vaxtı göstərilir. İstifadəçinin hər birinin bazaya (onun tərtibçiləri, lüğət mənbələri, bazalar arasındakı əlaqələr) və bazadakı materiallara ardıcıl baxmaq imkanı vardır. İstifadəçi müvafiq formanı dolduraraq xüsusi bir sorğu yarada və bütün türk dilləri üçün məlumat ala bilər.

Tipoloji tədqiqatlar aparmaq üçün hazırlanan tipoloji verilənlər bazaları (TVB) müxtəlif təbii dillərin müəyyən olunmuş xarakteristikaları haqqında verilənləri strukturlaşdırılmış sistemli şəkildə özündə saxlayır. Tipoloji verilənlər bazalarında müxtəlif dillərin qrammatik xüsusiyyətləri və xarakteristikaları haqqında informasiyalar saxlanılır, məsələn, müxtəlif dillərdə müşahidə olunan qrammatik cinslərin sayı və s. Qrammatik cəhətdən cinslərə bölünməyən, iki cinsə (kişi, qadın) bölünən, üç cinsə (kişi, qadın, orta) və daha çox cinslərə bölünən dillər vardır. TVB-yə daxil olan verilənlər tədqiqatçılar tərəfindən müvafiq dillərin qrammatik xüsusiyyətlərinə uyğunlaşdırılmışdır. TVB-də nəzərdən keçirilən dillərin hər biri bütövlükdə və ya ayrı-ayrı dil hadisələri kəmiyyətinə və tərkibinə, təbiətinə, onlara daxil olan əlamətlərin tərkibi və quruluşuna görə xarakterizə olunub, fərqlənə bilər.

TVB-nin iki əsas növü vardır:

- xarakterik TVB – geniş spektrli dil hadisələrini təşkil edir;
- ixtisaslaşdırılmış TVB – ayrıca dil hadisələrinin öyrənilməsi üçün hazırlanmışdır [6].

Xarakterik tipoloji verilənlər bazaları sxeminin əsası nisbətən sadə “matris” strukturudur, verilmiş n sayda dillərin m sayda əlamətləri $n*m$ kimi göstərilir. Başqa sözlə desək, n sayda dillər m qədər əlamətlə xarakterizə olunur. Sxemin qurulmasında oxşar prinsiplər olan TVB-ni şərti olaraq xarakterik adlandırmaq olar. TVB-nin əsas fərqləndirici xüsusiyyətlərindən biri müqayisə edilən dillərin siyahısı və onları

təsvir etmək üçün əlamətlər dəstidir. Bu əlamətlər dəsti öyrənilən dil hadisəsinin miqyasına tədqiqatın məqsəd və vəzifəsinə uyğun olaraq müəyyən edir. Əgər dil hadisələrinin məlumatları qlobal miqyasda tədqiq olunursa, o zaman dillərdən nümunə götürülməsi adətən maksimum areal (dilin yayıldığı ərazi) və genetik təmsilçilik nəzərə alınmaqla aparılır, yəni tədqiqatçılar TVB-yə mümkün qədər fərqli dil qruplarının və regionların nümayəndələrini daxil etməyə çalışırlar. TVB nisbətən kiçik dil qruplarını da əhatə edə bilər, məsələn roman dil qrupu (cins prinsipinə görə) və ya Afrikan dil qrupu (coğrafi prinsipə görə).

Uzun illər ərzində işlənən, bir çox dil qruplarının xüsusiyyətlər massivini dəstəkləyən iddialı iki əsas TVB layihəsi vardır:

- TVB olan Dil Strukturlarının Dünya Atlası (World Atlas of Language Structures – WALS) maksimum 192 əlamətdə 2679 dil haqqında informasiyaları saxlayır [7]. Bu 192 əlamət hər bir dil üçün deyil, məsələn rus dili üçün bu əlamətlərin sayı 156-dır [8]. Bu baza 2005-ci ildə kitab şəklində, 2008-ci ildən isə onlayn versiyada istifadəçilərə təqdim olunmuşdur. Burada dillərin qrammatik, sintaktik, həmçinin leksik və fonetik xüsusiyyətləri əks olunur.
- Dillərin xüsusiyyətlərini özündə saxlayan ikinci böyük TVB Dünya dilləri (Языки Мира – ЯМ) 315 dilin 3821 əlamətini özündə saxlayır [9].

Struktur parametrik modellərə əsaslanan çoxdilli Linqvistik verilənlər bazalarının (LVB) səmərəliliyini artıran digər funksiya

coğrafi informasiya sistemləri ilə birləşməsidir. Belə sistemlərə nümunə kimi Picin və Kreol Dillərinin Strukturları Atlası (Atlas of Pidgin and Creole Language Structures – APiCS) çoxdilli linqvistik verilənlər bazasını göstərmək olar. ApiCS-də kreol və picin dilləri üçün xüsusi parametrlər vardır. Məsələn: bir çox dillərdə işlənən “uşaq” və ya “balaca” sözü portuqal dilində “pequeno”, yəni “kiçik” sözüə çevrilir [10]. Hal-hazırda bu LVB hər biri 130 -ə qədər fonetik, leksik və qrammatik parametrlər ilə xarakterizə olunan 76 dili təsvir edir və dil hadisələrinin coğrafi bölgüsünü nəzərdən keçirməyə imkan verən coğrafi informasiya sistemi ilə inteqrasiya olunmuşdur.

Bunlardan başqa, daha bir neçə linqvistik verilənlər bazasını misal göstərmək olar: Tipoloji Verilənlər Bazası Sistemi (Typological Database System – TDS), Fonoloji Segment İnteraktiv Bazası (Phonological Segment Inventory Database), Dünya Alınma Söz Bazası (The World Loanword Database – WOLD) və s. [13]

İxtisaslaşdırılmış tipoloji verilənlər bazası geniş profilli xarakterik TVB-lərdən fərqli olaraq, çox geniş və ya məhdud şəkildə dil materiallarını, yəni tədqiqatçıların müvafiq dilləri xarakterizə edib, təsnif etdikləri təhlilə əsasən bəzi dizaynları və ya mətn parçalarını əhatə edir. Konkret layihələr üzərində işləyən tədqiqatçıların məqsədlərinə və nəzəri əsaslara uyğun olaraq hazırlanmış bu TVB-lər daha çətin struktura malik olur. Məsələn, Şəxs və işarə əvəzlilikləri TVB-si ümumi dil məlumatları ilə yanaşı, özündə tipoloji fərqli 109 dilin dəstini saxlayır.

Eyni zamanda bu əvəzliliklərin paradigmalarını, həmçinin hər bir dil üçün əvəzliliklər sisteminin xülasəsini özündə ehtiva edir [11]. Bununla yanaşı hər bir əvəzlilik verilən dəstin morfoloji və sintaktik xüsusiyyətləri ilə təsnif olunur.

Bu cür TVB-lərin məqsədi, bir qayda olaraq, tipoloji tendensiyaları və tipoloji hadisələri aşkara çıxarmaqdır. Məsələn, Şəxs və işarə əvəzlilikləri TVB-nin təhlili nəticəsində müəyyən edilmişdir ki, onda olan məlumatlar morfoloji və sintaktik atributların universalələrinə uyğun gəlir. Bununla yanaşı göstərilir ki, məsələn, tay dilinin əvəzliyi digər dillərin əvəzliliklərindən fərqlənir. Bu fərqlilik o qədər ciddidir ki, bu əvəzliliklərin şəxs əvəzliliklərinə və yaxud isimlərə aid olması sual yaradır. Beləliklə, Şəxs və işarə əvəzlilikləri TVB-si ilə iş tədqiqatçılara belə bir fundamental sual yaradır, həqiqətən də şəxs əvəzlilikləri mövcuddurmu və bütün dillərdə bu əvəzliliklər varmı?

Dil hadisələrini və obyektlərini təsvir edən daha bir TVB Surrey Universitetinin (Böyük Britaniya) morfoloji qrupu tərəfindən hazırlanmışdır. Bu TVB müxtəlif dillərdə sinkretizm haqqındadır. Sinkretizm, adətən eyni formaya malik olan, morfoloji və sintaktik xüsusiyyətləri fərqli olan sözlər arasındakı əlaqəyə deyilir [12].

Bu TVB-də sinkretizmi təsvir edən on mümkün xarakteristika dəsti diqqət çəkir: cəm, hal, cins, müəyyənlik, şəxs, zaman, feil şəkilləri və növləri, tərz, inkarın olub, olmaması. Bu xarakteristikaların mənalari ayrıca cədvəldə cəm şəklində qruplaşdırılır, sinkretizm özü isə bu birləşmələrin cütü şəklində təsvir olunur. TVB müxtəlif dillərdə sinkretizm haqqında yalnız binar

(ikili) sinkretizmlərin qeyd edilməsini təmin edir: əgər ikidən çox birləşmənin ümumi forması varsa, onda belə sinkretizm bir neçə ikiliyə bölünür.

Surrey qrupu müxtəlif dillər üçün supletivizm üzrə TVB hazırlayıb. Bu tip dillərdə bir leksik vahidin iki flektiv formaya malik olmasının aydın fonoloji əlaqəsi yoxdur. Supletivizm hadisəsi odur ki, söz kökünün dəyişməsi yolu ilə yeni sözlər əmələ gəlir. İltisacı dil olan Azərbaycan dilində bu tip nümunələrə rast gəlinməsə də, flektiv dillərdə bu hadisəyə tez-tez rast gəlinir (məsələn, good – better – best, bad – worse – worst, человек – люди, итди – шел). Supletivizm nizamsız hadisə olduğundan ona nizamlı model qurmaq olmur. Müəlliflər bu TVB-nin təsviri üçün bir çox hallarda sinkretizmi təsvir edən TVB-nin sxemini təkrarlayan verilənlər sxemindən istifadə edir. Burada mənalari ayrıca cədvəldə cəm şəklində qruplaşdırılan dil xarakteristikalarını saxlayan on cədvəl vardır. Sonra daha yüksək səviyyəli cədvəldə leksik vahidlərin kökləri malik olduqları birləşmələrə bağlanır və daha sonra yuxarı səviyyəli bir cədvəldə xüsusi bir leksem supletivizimlə öz aralarında əlaqəli olan müxtəlif köklərə birləşdirilir. Beləliklə, TVB-də təsvir olunan hadisələr ağaca bənzər məntiqi struktura malik olur və verilənlər bazası relyasiya modelinin cədvəli şəklində təsvir olunur [12].

NƏTİCƏ

Nəticə olaraq qeyd etmək lazımdır ki, linqvistik verilənlər bazası ölkəmizin dilçilik tədqiqatlarında böyük əhəmiyyət kəsb etsə də, hələ də kifayət qədər öyrənilməyib. Bu səbəbdən də onun

hərtərəfli öyrənilməsi və Azərbaycan dilinin bu tip dialektoloji və etimoloji bazalarının yaradılması dilçilik baxımından mühüm məsələlərdəndir.

İSTİNADLAR

- [1] Р.Дж. Абдрахманова. Лингвистическая Типология: учебное пособие. Бишкек: Изд-во КPCY, 2014. 108 с.
- [2] М. Heydərova. Elektron məkanda dil materiallarının verilmə üsulları. Lənkəran Dövlət Universitetinin Elmi Xəbərləri. 2019 №1, s. 35-44
- [3] М. Heydarova. Compiling of phonetic database. Path of Science: International Electronic Scientific Journal. Vol 7, № 4, 2021, s. 4001-4006
- [4] Р. К. Потапова, В. В Потапов. Речевые базы данных как часть мультимодальных корпусов в интернете. Вестник МГЛУ. Гуманитарные науки, 2018. Вып. 6 (797). С. 99-116
- [5] The Tower of Babel [Elektron resurs]. URL: <http://starling.rinet.ru> [istinad tarixi: 01.02.2023].
- [6] Г.Кружков. Информационные ресурсы контрастивных лингвистических исследований: типологические базы данных, Системы и средства информ., 2015, том 25, выпуск 1, 198–212
- [7] М. Haspelmath. The typological database of the World Atlas of Language Structures. The use of databases in cross-linguistic studies. Empirical approaches to language typology. Berlin -New York: Mouton De Gruyter, 2009. Vol. 41. P. 283-300.
- [8] WALS: Language Russian. [Elektron resurs]. https://wals.info/languoid/lect/wals_code_rus [istinad tarixi: 01.02.2023].
- [9] В.Д.Совольев Типологические базы данных: перспективы и использования. Вопросы языкознания, 2010. № 1. с. 94-110.

- [10] Language Structures Online [Elektron resurs]. URL: <http://apics-online.info/> [İstinad tarixi: 01.02.2023].
- [11] H. Bliss, E. Ritter. 2009. A typological database of personal and demonstrative pronouns. The use of databases in cross-linguistic studies. Empirical approaches to language typology. Berlin -New York: Mouton De Gruyter. 41:77-116.
- [12] P. H. Matthews The Concise Oxford Dictionary of Linguistics. Oxford: Oxford University Press, 1997. 432 p.
- [13] A. Dimitriadis, M. Windhouwer, A. Saulwick, R. Goedemans and T. Bıró 2009. The Typological Database System. The use of databases in cross-linguistic studies. Empirical approaches to language typology. Berlin -New York: Mouton De Gruyter. 41:155-207.

Proqram təminatı vasitələrinin lokallaşması problemləri haqqında

Tamilla Bayramova

İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

tamilla@iit.science.az

Xülasə — Məqalədə proqram təminatının lokallaşdırılması problemindən, yəni onun başqa dil və mədəniyyət nümayəndələrinə uyğunlaşdırılması proseslərindən danışılır və proqram təminatının lokallaşdırılmasının linqvistik aspektləri təhlil edilir. Lokallaşma sürətlə inkişaf edən kompüter texnologiyaları ilə müasir tərcümə texnologiyalarının qovşağında yerləşir və proqram təminatı sənayesinin ən aktual problemlərindən biridir.

Açar sözlər — *lokallaşma; tərcümə; proqram təminatı; interfeys.*

I. GİRİŞ

“Düşüncələrdəki dərinliyi, hissələrdəki incəlikləri bütünlüklə ifadə etmək kamilliyinə yetişə bilməsi üçün hər hansı xalqın bir neçə minillik tarixi yaşaması lazım gəlir. Azərbaycan dilinin bugünkü inkişaf səviyyəsi göstərir ki, Azərbaycan xalqı dünyanın ən qədim xalqlarındandır” *H.Əliyev.*

İnformasiya texnologiyalarının inkişafı dilimizin qorunması sahəsində həlli vacib olan yeni, aktual məsələlərin yaranmasına gətirmişdir. Hal-hazırda Azərbaycan dilinin əsas problemlərindən biri də istifadə etdiyimiz əməliyyat sistemləri, proqram vasitələri, oyunlar və mobil əlavələrin başqa

dillərdə (əsasən ingilis, türk və rus dillərində) olmasıdır. İnternetdən müxtəlif dillərdə kompüter oyunları, öyrədici proqramlar, mobil proqram əlavələri və s. yükləmək və intuitiv şəkildə onun iş prinsipini öyrənmək olar. Lakin təcrübəsi olmayan digər insanlar, əsasən də uşaqlar və yaşlılar dili bilmədiklərinə görə bu proqram vasitələrindən istifadə edə bilmirlər. Ona görə də əsas proqram vasitələrinin lokallaşmış, yəni Azərbaycan dilinə uyğunlaşdırılmış versiyasının işlənməsi onun daha çox insanlar tərəfindən istifadə edilməsinə gətirir.

II. PROQRAM TƏMİNATININ LOKALLAŞMASI VƏ BEYNƏLMİLƏLLƏŞDİRİLMƏSİ

Proqram təminatının lokallaşdırılması (*software localization*) proqram təminatının müxtəlif ölkə və regionlarda istifadə edilə bilməsi üçün yerli dilə və mədəniyyətə uyğunlaşdırılmasıdır. Bu zaman interfeys elementləri, məlumatlandırıcı və dəstək faylları, sənədlər müvafiq dillərə tərcümə edilir [1].

Lokallaşma sürətlə inkişaf edən kompüter texnologiyaları, kompüter linqvistikası və müasir tərcümə texnologiyalarının (o cümlədən maşın tərcüməsi) kəşişməsində yerləşir. Proqram

sistemlərini işləyən şirkətlər öz məhsullarının eyni vaxtda bir çox dillərdə və dünyanın müxtəlif ölkələrində buraxılmasını istəyirlər. Bu cür problemlərin həlli proqram təminatının lokallaşdırılmasının köməyi ilə mümkün olur [2].

Əsasən aşağıda göstərilən proqram vasitələri lokallaşdırılır:

- Əməliyyat sistemləri, axtarış sistemləri, mühasibat proqramları;
- Kütləvi istifadə üçün nəzərdə tutulan proqram məhsulları (Microsoft Office, Adobe Photoshop və s.);
- Mobil proqram əlavələri, oyunlar;
- və s.

Lokallaşdırma sadəcə olaraq mətnin tərcüməsindən ibarət deyil, tərcümə ilə yanaşı aşağıda göstərilən köməkçi tədbirlərin də yerinə yetirilməsi vacibdir [3]:

- İnsan adları, tarix, zaman, çoxrəqəmli ədədlər və kəsr ədədlərin formatlarının milli formata uyğunlaşdırılması;
- Milli ölçü sistemi, telefon kodları, valyuta vahidlərinə keçid;
- Milli qanunvericiliyin xüsusiyyətlərinin nəzərə alınması:
 - Vergi sistemi;
 - Dövlət tərəfindən verilən sənədlərin formatı;
 - Proqram təminatına dövlət orqanları tərəfindən qoyulmuş məhdudiyyətlər;
 - İnterfeys elementlərinin düzgün yerləşdirilməsi (sözlərin uzunluğu digər dilə keçdikdə əhəmiyyətli dərəcədə fərqlənə bilər);

- Siyahıların əlifba sırası ilə düzgün çeşidlənməsi (milli əlifba siyahısının daxil edilməsi);
- Terminlərin düzgün tərcümə edilməsi;
- Lazım gəldikdə yazı sisteminin dəyişdirilməsi (məsələn, ərəb dilində sağdan sola).

Kompüter oyunlarının və öyrədici proqramların lokallaşdırılması bir qədər fərqlidir. İnteraktiv əyləncə proqram sənayesi son illərdə əhəmiyyətli dərəcədə inkişaf etmişdir. Bu kompüter oyunlarının yaxın gələcək üçün ən gəlirli və perspektivli biznes sahəsinə çevriləcəyini göstərir. Rəqabətli bazarda uğur qazanmaq üçün interaktiv proqram təminatının işlənilmə prosesində lokallaşma məsələləri həll edilməlidir [4].

Kompüter oyunlarının lokallaşdırılması zamanı müxtəlif məsələlərə baxılır:

- İstifadəçi təlimatının tərcümə edilməsi;
- Milli mentalitetin, milli yumorun (kompüter oyunlarında personajlara adların verilməsi, qrafikada) nəzərə alınması;
- Ssenarilər, dialoq və subtitrlər, personajların adları, əşyaların adları, interfeys və menyunun tərcümə edilməsi;
- Şəkillər üzərində mətn varsa qrafikanın yenidən çəkilməsi;
- Yol nişanları, poçt qutuları və s. bu kimi nişanların milli standartlara uyğunlaşdırılması;
- Səs fayllarının dublyaj edilməsi və yenidən yazılması;
- Lokallaşdırılmış materialların oyuna əlavə edilməsi, yeni versiyanın sınağı və həmin oyunun müşayiət edilməsi (yeni

versiyaların və dəyişikliklərin vaxtında tərcümə edilərək əlavə edilməsi).

Beləliklə, kompüter oyunlarının lokallaşdırılması digər proqram məhsullarının lokallaşdırılması ilə müqayisədə unikal və eyni zamanda mürəkkəb bir prosesdir.

Proqram təminatının beynəlmilləşməsi proqram məhsulunun və ya sənədlərin proqramın lokallaşdırılmasını mümkün qədər asanlaşdıracaq şəkildə layihələndirilməsini və işlənməsini əhatə edən ümumi bir anlayışdır. Başqa sözlə desək, beynəlmilləşdirmə kontenti əvvəlcədən lokallaşdırılmaya hazırlamaqdır və proqram təminatının yaradılmasının ilkin mərhələlərində nəzərə alınmalıdır. Əslində isə 100% beynəlmilləşdirilmiş proqram təminatı yoxdur. Beynəlmilləşmə və lokallaşma proqram məhsulunun qloballaşması prosesinin tərkib hissəsidir.

III. LOKALLAŞMANIN MƏRHƏLƏLƏRİ

Xarici proqram məhsullarını daxili bazarda satmaq və ya milli proqram məhsullarını xarici bazara çıxarmaq üçün onların lokallaşdırılması lazımdır. Proqram təminatının uğursuz tərcüməsi şirkətin gəlirini və əməyin effektivliyini aşağı salır və istifadəçilərin haqlı narazılığına səbəb olur. Oyunların xətalı tərcüməsi süjetin anlaşılmasında və məhsuldan narazılıqla nəticələnir. Bu problemləri aradan qaldırmaq məqsədilə proqram təminatı vasitələrini lokallaşdırmaq üçün mütəxəssislərdən ibarət qrup təşkil edilməlidir. Bu qrup əsasən tərcüməçilər, lokallaşma üzrə mühəndislər, keyfiyyətin

təmin edilməsi üzrə mühəndis və layihə rəhbərindən ibarət olur.

Lokallaşma prosesi bir neçə mərhələdə yerinə yetirilir:

- Proqram təminatının texniki imkanları analiz edilir. Bu mərhələdə lokallaşma zamanı yarana biləcək çətinliklər təhlil edilir.
- Proqram təminatının tərcüməsi və uyğunlaşdırılması. Bu mərhələyə daxildir:
 - Lokallaşdırma üçün qrupun formalaşdırılması;
 - Tərcümədə terminoloji uyğunluğu qorumaq üçün qlossarilərin tərtib edilməsi;
 - Mətnin tərcümə üçün əlverişli olan formata keçirilməsi və tərcümə edilməsi;
 - İnterfeysin və xidməti proqramların göndərişlərinin tərcümə edilməsi;
 - Kömək fayllarının və sənədləşmənin tərcümə edilməsi;
 - Linqvistik dəstək.
- Lokallaşdırılmış proqram təminatının sınağı və sazlanması.

Bu mərhələdə tərcüməçilər və texniki mütəxəssislər birlikdə iştirak edirlər. Tərcümədə qüsurlar, ayrı-ayrı mətnlərdən götürülmüş sözlərin uzlaşması, mətnin ekranda düzgün yerləşdirilməsi və s. yoxlanılır.

Proqram təminatının lokallaşdırılması **LISA (Localization Industry Standards Association)** tərəfindən müəyyən edilmiş standartlara uyğun həyata keçirilməlidir [5].

NƏTİCƏ

Azərbaycanda da proqram təminatı vasitələrinin lokallaşdırılması aktual problemlərdən biridir.

Türkiyə, Rusiya, Ukrayna kimi dövlətlərdə bu problemlər müəyyən qədər həll edilmişdir və bir sıra əməliyyat sistemləri, ofis proqramları, oyunlar və s. bu ölkələrə uyğunlaşdırılmışdır. Proqram təminatının lokallaşdırılması xidməti göstərən müxtəlif saytlar var. Eyni zamanda lokallaşma işlərini asanlaşdırmaq üçün avtomatlaşdırılmış tərcümə alətləri yaradılmışdır (Computer Aided Translation – CAT). Bu alətlər proqram təminatının əvvəlki versiyalarından tərcümə edilmiş fraqmentləri avtomatik şəkildə götürərək yeni versiyaya əlavə edə bilər [6].

Microsoft şirkəti öz məhsullarının dünyanın bir çox ölkələri üçün lokallaşdırılmasını təmin etmək üçün bir sıra fundamental tədbirlər görmüşdür [7]:

- lokallaşdırma üzrə 100-dən çox təlimat (məsələn, müəyyən dil üçün qayda və tələblərin məcmusu, lokallaşma üzrə tövsiyələr, regional standartlar və müəyyən dillərin spesifik xüsusiyyətləri haqda məlumatlar və s.);
- Microsoft şirkətinin məhsullarını əhatə edən terminlər bazası;
- istifadəçi interfeysinin sətir elementlərinin tərcümələri və qlossariləri;
- proqramların müxtəlif dillərdə lokallaşdırılması üçün tərcümə alətləri və istinad materialları.

İSTİNADLAR

- [1] K. Aida-zade, S. Rustamov, E. Mustafayev, and N. Aliyeva, “Human-computer dialogue understanding hybrid system,” in Proc. IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA), July 2012, pp. 1–5
- [2] N. Kano, Developing International Software, Microsoft Press, Redmond, Wash., 1995.
- [3] Craciunescu O. Machine Translation and Computer-Assisted Translation; a New Way of Translating? // Translation Journal. 2004. Vol. 8. No. 3. URL: www3.uji.es/~aferna/EA0921/7a.pdf
- [4] R. W. Collins, "Software localization for Internet software, issues and methods," in IEEE Software, vol. 19, no. 2, pp. 74-80, March-April 2002, doi: 10.1109/52.991367.
- [5] Морозов Михаил Дмитриевич Особенности процесса локализации компьютерных игр // Вестник АГТУ. 2019. №2 (68). URL: <https://cyberleninka.ru/article/n/osobennosti-protsesssa-lokalizatsii-kompyuternyh-igr> (дата обращения: 27.03.2023).
- [6] LISA (Localization Industry Standards Association). Mode of access: <http://www.lisa.org>
- [7] Baker M., Saldanha G. Routledge Encyclopedia of Translation Studies, 2nd edition. –Routledge, Abingdon: 2009. – 674 с.
- [8] Muntés, V. [et al.]. Context-aware machine translation for software localization. A: Conference of the European Association for Machine Translation. "Proceedings of the 16th Annual Conference of the European Association for Machine Translation: EAMT 2012: Trento, Italy, May 28th-30th 2012". Trento: 2012, p. 77-80.
- [8] Microsoft Language Portal. URL: <https://www.microsoft.com/en-us/language/Search?searchTerm=hands-free&langID=594&Source=true&productid=0>

Orfoqrafiya lüğətinin internet resurslarına inteqrasiyası və mətnlərin avtomatik düzəlişi məsələləri

Raqif İsmayılov¹, Kəmalə Qurbanova²

^{1,2} İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

¹raqif.ismayilov@mail.ru, ²kemalewamil@gmail.com

Xülasə — Müasir zamanda informasiya kommunikasiya texnologiyalarının dinamik inkişaf etdiyi hazırkı qloballaşma dövründə milli dil korpusunun onlayn mühitə tətbiqinə tələbat yaranmışdır. Milli dil korpusu zəngin məlumat bazasına və geniş axtarış imkanları olan proqram təminatlarına malik olmalıdır. İşdə elektron linqvistik korpusların yaranma tarixinə nəzər salınıb və dil korpusunun məlumat axtarış sistemi olduğu qeyd olunmuşdur. Azərbaycan milli dil korpusunun yaradılmasının və bu platformanın inkişafının təmin olunmasının zəruriliyi əsaslandırılmışdır.

Açar sözlər — milli dil korpusu; orfoqrafiya lüğəti; kompüter dilçiliyi; mətn korpusu; linqvistik korpus

I. GİRİŞ

İstənilən dilin orfoqrafiya lüğəti ədəbi dildə işlənən sözlərin düzgün yazılış formalarını özündə ehtiva edən mötəbər bir mənbədir. Zaman keçdikcə yeni sözlər yaranır ki, bu təbii proses də dilin orfoqrafiya lüğətini zənginləşdirir. Eyni zamanda hər millətin orfoqrafiya lüğəti onun dil vahidinin etimologiyasını da əks etdirir.

Yaşadığımız qloballaşma dövründə orfoqrafiya lüğətinin internet resurslara

inteqrasiyası hər millətin milli dil korpusunun yaranmasına zərurət yaratmışdır. Dil korpusu şifahi nitq və ya yazılı mətnlər şəklində müəyyən bir dilin təhlili üçün linqvistik məlumatların məcmusudur [1].

Azərbaycan dilinin lüğət tərkibində də ciddi dəyişikliklər özünü göstərmiş və ədəbi dilimiz zamanın tələbi ilə bir sıra yeni anlayışları ifadə edən söz və ifadələrlə xeyli zənginləşmişdir. Milli dilimizin orfoqrafiya lüğətinin internet resurslarına inteqrasiyası zamanın tələbidir.

II. AZƏRBAYCAN ORFOQRAFIYA LÜĞƏTİNİN İNKİŞAF MƏRHƏLƏLƏRİ

Azərbaycan dilinin ilk orfoqrafiya lüğəti 1929-cu ildə nəşr olunmuşdur. Linqvistika üzrə görkəmli alim Vəli Xulufunun latın qrafikalı əlifba əsasında tərtib etdiyi bu orfoqrafiya lüğətindəki sözlərin sayı cəmi 11 min olmuşdur. Orfoqrafiya lüğətinin sonrakı təkmilləşdirilmiş nəşrləri 1940 (20 min söz), 1960 (40 min söz), 1975-ci ildə (58 min söz) işıq üzü görmüşdür. 2013-cü ildə nəşr olunmuş sonuncu orfoqrafiya lüğəti isə həcminə görə daha böyük olmuşdur. Lüğətdəki sözlərin sayı 104 minə keçmişdir [2].

Azərbaycan dilinin orfoqrafiya lüğəti 2021-ci ildə Azərbaycan Milli Elmlər Akademiyasının həqiqi üzvü, filologiya elmləri doktoru, professor Möhsün Nağısoylunun rəhbərliyi ilə yenidən işlənmiş 7-ci nəşr Azərbaycan Respublikası Nazirlər Kabinetinin 16 aprel 2019-cu il tarixli qərarı ilə təsdiqlənmiş və 3 noyabr 2020-ci il tarixli 438 nömrəli qərarı ilə müəyyən dəyişikliklər edilmiş “Azərbaycan dilinin orfoqrafiya normaları” əsasında hazırlanmış ilk nəşrdir. Lüğətə 90 minə yaxın söz daxil edilmişdir [3].

Son dövrlərdə informasiya kommunikasiya texnologiyaları (İKT) dinamik inkişaf edərək bütün sahələrə öz töhfəsini vermiş və həmin sahənin sürətlə inkişafına səbəb olmuşdur. Digər sahələr ilə yanaşı dil korpusunun onlayn mühitə inteqrasiyası istiqamətində aparılan işlər İKT-nin imkanları sayəsində yeni səviyyəyə yüksəlmişdir. Respublikamızda milli dil siyasətinin inkişafı Azərbaycan xalqının ümummilli lideri Heydər Əliyevin adı ilə bağlıdır. Azərbaycan Respublikası Prezidentinin imzaladığı proqramlar çərçivəsində milli dilimizin istifadəsinə və tədqiqinə dövlət qayğısı artırılmış və ölkədə dilçilik elmi əsaslı surətdə inkişaf etmişdir. Proqramda milli dilimizin öyrənilməsi və təbliği sahəsinə İKT-nin tətbiqi alimlərin qarşısına vəzifə kimi qoyulmuşdur. Ana dilimizin qlobal mühitə inteqrasiyası üçün avtomatik tərcümə sistemlərinin işlənməsi, linqvistik texnologiyaların yaradılması və inkişaf etdirilməsi Dövlət Proqramının icrasına dair tədbirlər planında əks olunmuşdur [4].

Milli dilimizdə olan məlumatların onlayn mühitdə axtarış imkanlarını genişləndirmək üçün dil korpusunun hazırlanması zəruri və aktual məsələdir. Dil korpusu zəngin məlumat bazasına və bu baza üzərindən geniş imkanlara malik proqram təminatlarına malik olmalıdır.

III.DÜNYADA YARADILAN ELEKTRON DİL KORPUSLARI

Dillərin öyrənilməsi və öyrədilməsi prosesində dil emalının yeni mütərəqqi tətbiqlərdən, nitqin tanınması, axtarış sistemləri və avtomatik tərcümə korpuslarından istifadə daha səmərəli nəticə verir.

Araşdırmalara görə dünyada elektron linqvistik korpusların yaranma tarixi 1960-cı illərə təsadüf edir. Birinci nəsil elektron linqvistik korpuslara aşağıdakıları misal göstərmək olar [8]:

- *The Brown Corpus*. Bu linqvistik korpus Amerika Braun (Brown) Universitetində 1961-ci ildən 1964-cü ilə qədər, 4 il müddətinə tərtib edilmişdir. Bir milyon ifadədən ibarət olan bu korpusun dili ingilis (Amerika) dilidir və yazılı mətnlərdən ibarətdir. Mətnlərin strukturu və kateqoriyaların seçimi çox yaxşı düşünülmüş olduğundan, korpus tədqiqatı üçün yararlıdır.
- *Lancaster-Oslo/Bergen (LOB) Corpus*. Layihə 1970-1978 illərdə Lankaster, Oslo universitetlərinin və Bergendəki elmi mərkəzin birgə fəaliyyəti nəticəsində ərsəyə gəlmişdir. Korpus ingilis (Britaniya) dilində bir milyon ifadədən ibarət idi. Strukturu *Brown Corpus*-a bənzəyir. Alimlərin

fikrincə bir milyon söz istifadəsi dilin aşağı tezlikli elementlərini təhlil etmək üçün kifayət etmirdi. Lakin buna baxmayaraq, o dövrdə yüzlərlə yüksək keyfiyyətli və maraqlı tədqiqatlar *Brown* və *LOB* korpusuna əsaslanmışdı.

- *London-Lund Corpus (LLC)*. London Universiteti Kollecinin əməkdaşları tərəfindən 1975-ci ildə şifahi ingilis dilinə əsaslanan LLC korpusu yaradıldı. Korpus orfoqrafik transkripsiya, fonetik və prosodik (şifahi olmayan) işarələrdən ibarət idi və təxminən 500 min işarəni ehtiva edir. Əvvəlcə kağız üzərində hazırlanmış bu layihə İsveçin dilçi alimləri tərəfindən kompüter formasına çevrilmişdir.

- *The Cobuild Project / The Bank of English*. 1990-cı ildə Birminhem Universitetinin təşəbbüsü ilə ərsəyə gəlmişdir. Korpus daim yeniləndiyindən dəqiq ölçüyə malik deyil. 1997-ci ildə təxminən 300 milyon ifadə olan korpusda, 2005-ci ildə artıq 525 milyon ifadə var idi. Korpusun 25 faizi şifahi, 75 faizi yazılı ifadələrdir.

- *The International Corpus of English (ICE)*. Bu korpus ingilis dilinin rəsmi dil olduğu ölkələrin (Avstraliya, Kanada, Yeni Zelandiya və s.) bir neçə universitetlərinin birgə layihəsi ilə 1996-cı ildə meydana gəlmişdir. Bu korpusun təhlili üçün xüsusi olaraq mürəkkəb proqram təminatı hazırlanmışdır.

- *Gigaword corpora*. 2014-cü ildə yaranan bu nəhəng korpus Avropa İttifaqı tərəfindən maliyyələşdirilib və ingilis, ərəb, çin və digər dillərin linqvistik monitorinqini apara

bilir. Korpus 1 milyard söz bazasından ibarətdir.

Texnologiyanın dinamik inkişaf etdiyi son zamanlarda mətnlərin emalı, axtarışı və saxlanması istiqamətində əldə olunmuş yeniliklər yüz milyon və ya daha çox sözdən ibarət korpus yaratmağa imkan verir. Buna misal olaraq 2019-cu ildə görkəmli alim Mark Davies tərəfindən yaradılmış English-Corpora.org platformasını misal göstərmək olar. Bu layihə tərkibində ən çox istifadə olunan onlayn korpusları birləşdirir və tədqiqatçılar tərəfindən müxtəlif məqsədlər üçün uğurla istifadə olunur. English-Corpora.org platforması müxtəlif sahələrdə, xüsusən də texnologiya və dil öyrənmə sahəsində böyük müəssisələr tərəfindən istifadə edilmişdir (Şək. 1) [9].



The screenshot shows the English-Corpora.org website interface. At the top, there is a navigation bar with links for 'corpora', 'guides', 'related resources', 'users', 'my account', 'upgrade', and 'help'. Below the navigation bar, there is a table listing various corpora. The table has columns for 'Corpus (see tour)', 'Download', '# words', 'Dialect', 'Time period', and 'Genre(s)'. The table lists several corpora, including 'News on the Web (NOW)', 'Web: The Intelligent Web-based Corpus', 'Global Web-Based English (GloWBE)', 'Wikipedia Corpus', 'Coronavirus Corpus', 'Corpus of Contemporary American English (COCA)', 'Corpus of Historical American English (COHA)', 'The TV Corpus', 'The Movie Corpus', and 'Corpus of American Soap Operas'. Each row includes a download icon, the number of words, the dialect, the time period, and the genre(s).

Corpus (see tour)	Download	# words	Dialect	Time period	Genre(s)
News on the Web (NOW)	1	16.8 billion+	20 countries	2010-yesterday	Web: News
Web: The Intelligent Web-based Corpus	1	14 billion	6 countries	2017	Web
Global Web-Based English (GloWBE)	1	1.9 billion	20 countries	2012-13	Web (incl blogs)
Wikipedia Corpus	1	1.9 billion	(Various)	2014	Wikipedia
Coronavirus Corpus	1	1.5 billion	20 countries	Jan 2020-Dec 2022	Web: News
Corpus of Contemporary American English (COCA)	1	1.0 billion	American	1990-2019	Balanced
Corpus of Historical American English (COHA)	1	475 million	American	1820-2019	Balanced
The TV Corpus	1	325 million	6 countries	1950-2018	TV shows
The Movie Corpus	1	200 million	6 countries	1930-2018	Movies
Corpus of American Soap Operas	1	100 million	American	2001-2012	TV shows

Şəkil 1. English-Corpora.org platforması

IV. ORFOQRAFİYA LÜĞƏTİNİN İNTERNET RESURSLARINA İNTEQRASIYASI

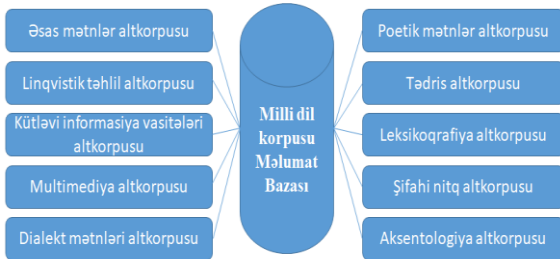
Orfoqrafiya lüğətinin internet resurslarına inteqrasiyası üçün Azərbaycan milli dil korpusunun yaradılması geniş imkanlar yarada bilər. Milli dil korpusu hər hansı bir konkret dildə mətnlərin elektron

formada toplanmasına istiqamətlənmiş məlumat axtarış sistemidir [5].

Dil korpusunda təmsil olunan dilin qrammatikası orfoqrafiyası və lüğəti ehtiva olunur. Eyni zamanda, dil korpusunda müxtəlif istiqamətli mətnlər (bədi, texniki və s.) də əks olunur. İstənilən dil korpusunda həmin dilə müraciət üsulları müəyyən qaydalar üzrə nizamlanır. İstifadəçi korpusdan maraq dairəsinə uyğun olan məlumatı asanlıqla əldə edə bilər [6, 7].

Azərbaycan milli dil korpusunun yaradılması üçün zəngin məlumat bazasının və bu baza üzərindən geniş imkanlara malik proqram təminatlarının hazırlanması zərurəti yaranır.

Milli dil korpusu ümumiləşmiş məlumat bazasından ibarət olmalı və bu baza alt korpuslardan təşkil olunmalıdır (Şək. 2.).

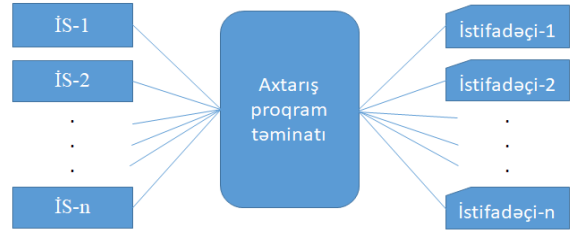


Şəkil 2. Azərbaycan milli dil korpusunun baza strukturu

Milli korpus ümumi və alt sistemləri əhatə etməklə yanaşı mövcud bazalar üzərindən analitik təhlillərin aparılması üçün geniş imkanlara malik olmalıdır, digər mühitlərdə və sistemlərdə istifadə olunacaq proqram servislərini özündə saxlamalıdır.

Azərbaycan milli dil korpusunun yaradılması üçün ikinci əsas məsələ olaraq

vəb-proqram təminatı hazırlanmalı, məlumat bazasının zənginləşdirilməsi məqsədi ilə veb-servis proqram təminatlarının tərtib edilməsi və istifadəçilər üçün rahat mühit təqdim edilməlidir (Şək. 3.).



Şəkil 3. İnternet mühitində mövcud olan zəngin informasiya sistemlərindən istifadə

V.MƏTNLƏRİN AVTOMATİK DÜZƏLİŞİ MƏSƏLƏLƏRİ

Azərbaycan milli dil korpusunun yaradılması mətnlərin yazılışında qrammatik və orfoqrafik səhvlərin avtomatik düzəlişi, durğu işarələrinin məntdəki yeri və söz seçimi məsələsini asanlaşdırma bilər.

Microsoft Word proqramında mətnlərdə qrammatik və orfoqrafik səhvləri göstərən *When correcting spelling and grammar in Word* funksiyası mövcuddur. Bu tip problemləri həll etməyə yardımçı olan digər proqram təminatlarını misal göstərək:

- *Key Switcher* – müxtəlif qrammatik və orfoqrafik səhvləri aşkar edən və düzgün variantı göstərən proqram vasitəsi;
- *Punto Switcher* – bir yazı sistemindəki hərflərin başqa yazı sisteminin hərfləri ilə

verilməsi əməliyyatını icra edən proqram vasitəsidir;

• Grammarly – məndə durğu işarələrinin mövqeyi və söz seçimi ilə bağlı yaranmış səhvləri aşkar etmək funksiyasına malikdir. Veb saytlarda müştəri ilə əlaqəni təkmilləşdirmək məqsədi ilə istifadə olunur. Süni intellekt texnologiyası istifadəçilərinə yazı üslubunu təkmilləşdirmək imkanı yaradır.

NƏTİCƏ

İnternetin yaratdığı qloballaşma dövründə informasiya texnologiyalarının və proqram təminatının dəstəyi ilə dilin qorunması və inkişafı üsulları ilə bağlı problemlər aktual məsələlərdən birinə çevrilmişdir. Müasir mühitdə linqvistik texnologiyaların davamlı şəkildə təkmilləşdirilməsi hesabına milli dilimizdə gedən prosesləri intellektual analiz etmək mümkündür.

Azərbaycan milli dil korpusu yaradılmalı və bu platformanın inkişafı təmin olunmalıdır. Belə bir korpusun yaradılması və veb-servis proqram təminatlarının tərtib edilməsi milli dilin təkmilləşdirilməsinə, məlumatların toplanmasına, analitik təhlilinə, qiymətləndirilməsinə və onlardan məqsədyönlü şəkildə istifadəyə şərait yaradacaqdır.

İSTİNADLAR

[1] K.H. “Волченкова Параллельный корпус как справочная база данных в работе

переводчика”, Проблемы и перспективы развития образования в России, №. 33, с. 32-35, 2015.

- [2] M.Nağısoylu, “Azərbaycan dilinin orfoqrafiya lüğəti”, Bakı, “Şərq-Qərb” nəşriyyat evi, s. 840, 2013.
- [3] M.Nağısoylu, “Azərbaycan dilinin orfoqrafiya lüğəti”, Bakı, “Elm” nəşriyyat evi, s. 756, 2021.
- [4] İ.Əliyev, “Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsinə və ölkədə dilçiliyin inkişafına dair Dövlət Proqramı”nın təsdiq edilməsi haqqında Azərbaycan Respublikası Prezidentinin Sərəncamı, Bakı, 9 aprel 2013-cü il, № 2837, <https://e-qanun.az/framework/25537>.
- [5] M.Mahmudov, “Kompüter dilçiliyi”, Bakı, “Elm və təhsil”, 356 s., 2013.
- [6] R. Məmmədova, R.Şahverdiyeva, L.Əkbərova, “Milli Dil Korpusunda Lüğətlər Blokunun Proqram Təminatının İlanılması Məsələləri”, Proqram mühəndisliyinin aktual elmi-praktiki problemlər I respublika konfransı, Bakı, 181-183, 17 may 2017.
- [7] R.M.Əliquliyev, R.M Alıquliyev, Y.M.Cəfərov, F.F. Yusifov, Ə.M. Qurbanova, “Kosmopolit e-dövlətdən milli e-dövlətə doğru: virtual mühitdə azərbaycan dili xidmətlərinin yaradılması perspektivləri” İnformasiya cəmiyyəti problemləri, №2, s. 3-25, 2021.
- [8] М. И. Солнышкина, Г. М. Гатиятуллина, “История развития корпусной лингвистики (на примере англоязычных корпусов)”, Вестник Томского государственного университета. Филология, №. 63, с. 132-160, 2020.
- [9] English-Corpora.org, [://www.english-corpora.org/byu-corpora.asp](http://www.english-corpora.org/byu-corpora.asp)

Beynəlxalq təcrübə əsasında Azərbaycan dilinin tədrisi və qiymətləndirilməsi məsələləri

Sayalı Quliyeva

İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
msayali@mail.ru

Xülasə — Tədqiqat işində Azərbaycan dilinin tədrisinin hazırkı vəziyyəti, onun mənimsənilməsi və qiymətləndirilməsi məsələləri araşdırılmışdır. Dil biliklərinin qiymətləndirilməsi sahəsində mövcud beynəlxalq standartlar təhlil edilmişdir. Azərbaycan dilinin tədrisi və qiymətləndirilməsi ilə bağlı müəyyən problemlər diqqətə çatdırılır. Azərbaycan dilinin tədrisinin hazırkı vəziyyəti, onun müasir tələblərə uyğunlaşdırılması və bu sahədə İKT-nin rolu göstərilmişdir. Beynəlxalq təcrübəyə və İKT-nin tətbiqinə əsaslanaraq Azərbaycan dilinin tədrisi və qiymətləndirilməsi metodologiyasının təkmilləşdirilməsi və inkişaf mexanizmlərinin işlənilməsi istiqamətində təkliflər verilmişdir.

Açar sözlər — Azərbaycan dilinin tədrisi; dil bacarıqları; qiymətləndirmə; CEFR; İKT

I. GİRİŞ

Müasir dövrdə Azərbaycan dilinin inkişafına xüsusi diqqətin ayrılması ölkədə dilçilik elminin daha da inkişafı üçün əlverişli zəmin yaratmışdır. Həmçinin informasiya-kommunikasiya texnologiyalarının (İKT) sürətlə inkişaf etdiyi qloballaşma dövrü Azərbaycan

dilinin zənginləşməsi və istifadə imkanlarının genişləndirilməsi istiqamətində aparılan işlərin yeni səviyyəyə yüksəldilməsini tələb edir. Qeyd edək ki, dillərin bir-birinə inteqrasiyası, eyni zamanda elm və texnologiyanın inkişafı dildə yeni sözlərin yaranmasına və dilin zənginləşməsinə təsir edir. Bu təsirlər özünü müsbət və mənfi yöndə göstərə bilər. Onlar bəzən dil normalarının pozulmasına səbəb olur. Dilin saflığının qorunması və onun inkişaf etmiş dünya dilləri sırasına gətirilməsi üçün dövlət səviyyəsində mühüm addımların atılması zərurəti yaranır [1].

Dil dövlət siyasətinin tərkib hissəsidir və dil quruculuğu, onun öyrədilməsi, düzgün istifadəsi və qiymətləndirilməsi bu gün qarşıda duran prioritet məsələ hesab olunur. Ölkədə konstitusion tələb olaraq həyata keçirilən dil siyasətinə dəstək məqsədilə Azərbaycan dilinin tədrisinin müasir vəziyyətini araşdırmaq, ədəbi dil qaydalarına, üslub və normalarına xələl gətirmədən onun tədrisinin, mənimsənilməsinin və qiymətləndirilməsinin müasir tələblərə və beynəlxalq standartlara uyğunlaşdırılması qarşıda duran əsas vəzifələrdən biridir. Bu

məqsədlə tədqiqatda ilkin olaraq dil biliklərinin öyrənilməsi və qiymətləndirilməsi sahəsində dünya təcrübəsi araşdırılmışdır. Dil biliklərinin öyrənilməsi və qiymətləndirilməsi üçün nəzərdə tutulmuş mexanizmlərin işlənilməsi zamanı İKT-nin rolu qiymətləndirilmişdir. Beynəlxalq təcrübə öyrənilərkən bir sıra standartlar araşdırılmış və onlar arasından “Ümumavropa dilləri üçün tövsiyələr toplusu” (Common European Framework of Reference – CEFR) ən münasib standart kimi seçilərək daha geniş təhlil olunmuşdur. Tədqiqatın əsas məqsədi beynəlxalq təcrübəyə və İKT-nin tətbiqinə əsaslanaraq Azərbaycan dilinin tədrisi və qiymətləndirilməsi üçün bəzi tövsiyələrin verilməsidir.

II. DİL BİLİKLƏRİNİN QIYMƏTLƏNDİRİLMƏSİNDƏ BEYNƏLXALQ TƏCRÜBƏ VƏ STANDARTLAR

Avropada xarici dillərin tədrisi metodikası, həmçinin dillərin mənimsənilmə və qiymətləndirmə mexanizmlərinə nəzər salsaq görərik ki, burada yalnız dilin tədrisinə deyil, eyni zamanda, dil biliklərinin qiymətləndirilməsinə də xüsusi diqqət yetirilir. Xüsusilə XX əsrin II yarısından başlayaraq bilinqualizm (ikidillilik) və plurinqualizm (çoxdillilik) kimi konsepsiyaların meydana gəldiyi bir şəraitdə dil biliklərinin əldə olunmasında yeni yanaşmalar və qiymətləndirmə metodları yaranmağa başladı.

Tədqiqat zamanı, dil biliklərinin öyrənilməsi, tədrisi və qiymətləndirilməsi

istiqlamətində bir sıra standartların olduğunu müşahidə etdik.

Təşkilatlararası Linqvistik Dəyirmi Masa Şkalası (*Interagency Language Roundtable Scale- ILR*) 1968-ci ildə ABŞ Dövlət Departamentinin Xarici Xidmət İnstitutu (Foreign Service Institute) tərəfindən hazırlanmış və bütün ABŞ dövlət qurumları üçün dil biliklərinin qiymətləndirilməsində standart kimi qəbul edilmişdir [2]. ILR tövsiyələr toplusu əsasında tərtib olunmuş testlər yalnız ABŞ dövlət işçilərinin dil bacarıqlarının ölçülməsi üçün nəzərdə tutulmuşdur. ILR tövsiyələr toplusunda test meyarlarının standartlaşdırılması, təşkilatlararası təlim, sınaq və tərcümə metodologiyası sahəsində əməkdaşlığın təmin edilməsi, dil ilə əlaqəli yeni məlumatların və təcrübələrin həm ictimai arenada, həm də xaricdə yayılması üçün şəraitin yaradılması əsas məqsəd hesab olunurdu.

ILR standartına görə bir insanın sahib olmalı olduğu minimum dil bacarıqlarına dinləmə, oxu, danışma, yazı, tərcümə və interpretasiya aiddir. İnsanların dil bacarıqları beş ballıq şkala ilə qiymətləndirilir. Dil bacarığı qiymətləndirilən şəxsin göstəriciləri bir bacarıq səviyyəsini əhəmiyyətli dərəcədə aşıb, növbəti səviyyəyə isə tam cavab vermədikdə onun balı “+” ilə təyin edilir (0+, 1+, 2+, 3+ və ya 4+ şəklində). Qiymətlər oxu, danışma, dinləmə, yazma, tərcümə, audio tərcümə, tərcümə və mədəniyyətlərarası ünsiyyət kimi müxtəlif bacarıqlar üçün ayrıca təyin edilə bilər [2].

Xarici Dillərin Tədrisi üzrə Amerika Şurası Standartı (*American Council on the Teaching of Foreign Languages - ACTFL*) - tədrisin bütün səviyyələrində dillərin öyrənilməsinə təkmilləşdirmək və qiymətləndirmək məqsədi daşıyan tövsiyələr toplusudur [3]. 1980-ci illərdə ILR tövsiyələr toplusu əsasında Müasir Dil Assosiasiyasının (MLA-Modern Language Association) kiçik bir qolu kimi yaradılmışdır. ACTFL-in kvalifikasiya testləri bütün dünyada akademik qurumlar, dövlət qurumları və özəl korporasiyalar tərəfindən istifadə olunur. ACTFL-də dil bacarıqlarının qiymətləndirilməsi üçün ümumilikdə on bir səviyyə var. Bunlardan beşi əsas səviyyə hesab olunur. Əsas səviyyələr isə üç əltsəviyyəyə bölünür [3].

Ümumavropa dilləri üçün Tövsiyələr Topusu (Common European Framework of References for languages - **CEFR**) - Avropa Şurasının və Kembric Universitetinin dəstəyi ilə dil biliklərinin tədrisi və qiymətləndirilməsi üçün qəbul olunmuş beynəlxalq standart hesab olunur [4]. 1991-ci ildə İsveçrədə keçirilən simpoziumda “Ümumavropa dilləri üçün tövsiyələr toplusu”nun tərtib edilməsi ideyası irəli sürülmüş və sənədin son rəsmi variantı 2001-ci ilin yanvarında nəşr olunmuşdur. “Ümumavropa dilləri üçün tövsiyələr toplusu” ilk dəfə iki dildə – ingilis və fransız dillərində çap olunmuş, daha sonralar isə dünyanın 40-a yaxın dilinə tərcümə olunmuşdur. Beləliklə, CEFR dil biliklərinin tədrisi və qiymətləndirilməsi sahəsində əsas beynəlxalq standartda çevrilmişdir [4].

“Ümumavropa dilləri üçün Tövsiyələr Topusu”nda dil biliklərinin qiymətləndirilməsi altı səviyyə (A1, A2, B1, B2, C1, C2) üzrə aparılır. Cədvəl 1-də dil biliklərinin səviyyələr üzrə bölgüsü təqdim edilir. Dil öyrənən şəxs dil bacarıqları üzrə fərqli göstəricilərə malik ola bilər. Belə ki, yazma və oxumada aşağı səviyyədə olduğu halda, dinləmə və danışmada orta səviyyəli nəticə göstərə bilər. Beləliklə, səviyyələr üzrə qiymətləndirmə aparılarkən dil öyrənənin dil bacarıqları ayrı ayrılıqda yoxlanılır və qiymətləndirmə ona uyğun aparılır [5].

Cədvəl 1. CEFR-da dil biliklərinin səviyyələr üzrə bölgüsü

Təcrübəli istifadəçi	C2	Qabaqcıl səviyyə
	C1	Üstün səviyyə
Müstəqil istifadəçi	B2	Orta səviyyə
	B1	Alt orta səviyyə
Yeni başlayan istifadəçi	A2	İbtidai səviyyə
	A1	Başlanğıc səviyyə

CEFR üzrə səviyyələrin ümumi xarakteristikasına nəzər salmaq:

- A1 - Günlük istifadə olunan sadə söz və ifadələrdən istifadə edərək ünsiyyət qura bilər.
- A2 - Tez-tez istifadə olunan ifadə və cümlələri başa düşə, ətrafda gördüklərini izah edə və həmin anda baş verən hadisələri müzakirə edə bilər.
- B1 - Standart və tanış mövzuları başa düşür, dilin olduğu mühitdə yarana biləcək istənilən situasiyada vəziyyəti idarə edə, şəxsi marağında olan

mövzularda mətnlər qura və öz fikirlərini nisbətən rahat ifadə edə bilir.

- B2 - Konkret və abstrakt mövzuda olan mətnlərin əsas ideyasını başa düşür, çətinlik çəkmədən yerli dil daşıyıcısı ilə qarşılıqlı əlaqə qura bilir.
- C1 - Mürəkkəb və qarışıq mövzular ətrafında müzakirə apara, dildən akademik və peşəkar məqsədlər üçün effektiv şəkildə istifadə edə bilir.
- C2 - Faktiki olaraq eşitdiyi və oxuduğu hər şeyi başa düşə, daha mürəkkəb situasiyalarda sərbəst və dəqiq şəkildə öz fikirlərini ifadə edə bilir.

CEFR-də dildən istifadə əsas şərtidir. Burada dillərin öyrənilməsi, tədrisi, qiymətləndirilməsi əsas götürülür. CEFR-in əsas məqsədi dil mütəxəssislərini ümumi baza ilə təmin etmək və fərqli təhsil sistemlərinin olması səbəbindən ünsiyyət zamanı yaranan maneələrin aradan qaldırılmasını təmin etməkdir. CEFR tövsiyələr toplusunda dil biliklərinin öyrədilməsi və qiymətləndirilməsində əsasən “bacarır” (“can do”) yanaşmasından istifadə olunur. Bu yanaşma öyrənənlərə dil bacarıqları üzrə hansı səviyyədə olduqlarını və hansı səviyyəyə keçməli olduqlarını müəyyən etməkdə köməklik göstərir [5].

CEFR-i digər standartlardan fərqləndirən üstün cəhət onun konkret bir dil üçün nəzərdə tutulmamasıdır. Bu tövsiyələr toplusundan istifadə edərkən hər bir dilin öz daxili linqvistik imkanları da nəzərə alınmalıdır.

Bu gün bir çox inkişaf etmiş və inkişaf etməkdə olan ölkələr CEFR tövsiyələr toplusundan istifadə edərək öz dillərinə uyğun imtahan test sistemlərini qurmuş və müvəffəqiyyətlə tətbiq edirlər, məsələn, TOEFL (ingilis), ESOL (ingilis), IELTS (ingilis dili), TestDaF (alman), TOPIK (koreya dili), JLPT (yapon dili), CCE (çex dili), TOCFL (Çin dili), TPKİ (rus), TYS (türk dili) və s. [6].

Hal-hazırda respublikamızda da qeyd olunan dil səviyyə imtahanlarından xarici dil biliklərimizin qiymətləndirilməsi üçün (TOEFL, IELTS, TOPIK və s.) istifadə olunur.

III. AZƏRBAYCAN DİLİNİN TƏDRİSİNİN MÖVCUD VƏZİYYƏTİ VƏ PROBLEMLƏRİ

XXI əsrin ilk onilliyi Azərbaycan təhsil sistemində köklü dəyişikliklərin baş verdiyi bir dövr kimi səciyyələnir. 2003-2013-cü illər dünya bankı ilə II kredit sazişi əsasında islahat proqramının həyata keçirilməsi dövrü hesab edilir [7]. Qeyd olunan təşəbbüslərin nəticəsi olaraq, Azərbaycan təhsil sistemində bir çox islahatlar həyata keçirilmişdir. Əldə olunan ən mühüm nəticə kimi “Milli Kurrikulum” konsepsiyasını göstərə bilərik [8]. “Milli Kurrikulum” konsepsiyası yeni standartlar əsasında formalaşdırılmışdır. Burada milli maraqlar və ana dilinin tədrisi dövlətçilik baxımından vacib məsələlərdən biri hesab edilir. Belə hesab olunur ki, dil şəxsin aktivliyinin əsas göstəricisidir, bu da öz növbəsində cəmiyyətdə müstəqil, yaradıcı və azad düşüncə qabiliyyətinə malik şəxsiyyətin formalaşmasını təmin edir. Dünya ölkələrində ana dilinin öyrədilməsi milli

təfəkkürün formalaşmasında mühüm vasitə olduğu kimi, “Milli Kurrikulum” konsepsiyasında da, ana dilinin tədrisi ön plana çəkilir [8]. Ana dilinin düzgün formada tədrisi şagirdlərin ünsiyyət imkanlarını genişləndirir, onlarda nitq fəaliyyətinin növləri ilə yanaşı, düşünmə, öyrənmə və danışma bacarıqlarını formalaşdırır. Təhsilin ümumi inkişaf səviyyəsini müəyyənləşdirmək üçün keçirilən beynəlxalq və milli qiymətləndirmələrdə ana dili üzrə standartların mənimsənilməsinə xüsusi diqqət yetirilir. Bütün bunlar bir fənn kimi ana dilinin şəxsiyyətin formalaşdırılmasındakı müstəsna rolu ilə bağlıdır [7].

Yuxarıda qeyd olunan dəyişikliklərə baxmayaraq, bu gün Azərbaycan dilinin tədrisində bəzi çatışmazlıqlar özünü göstərir. Belə ki, Azərbaycan dilinin tədrisi əsas dil bacarıqları olan dinləmə, oxu, yazı və danışmaq üzrə aparılır. “Milli Kurrikulum” konsepsiyasına qədər Azərbaycanda mövcud olan ənənəvi tədris sistemindəki çatışmazlıqlardan biri Azərbaycan dilinin tədrisinin nitq bacarıqları üzərində deyil, dilçilik üzərində qurulmasıdır. Belə ki, dil bacarıqlarının inkişaf etdirilməsi və qiymətləndirilməsi metodikası formalaşdırılmamışdır. Bu da öz növbəsində dil öyrənilməsi və dil biliklərinin qiymətləndirilməsində bir sıra çətinliklər yaradırdı. Həmçinin şifahi nitq bacarıqlarının, xüsusilə dinləmə və danışmanın diqqətdən kənar qalması, uşaqların və yeniyetmələrin əksəriyyətinin, hətta doğma dildə belə öz fikrini səlis ifadə etməkdə çətinlik çəkməsinə səbəb olmuşdur. Məhz qeyd olunan faktorlar

kurrikulum sisteminə tam keçidi təmin etməkdə çətinliklər yaradır.

Beləliklə, Azərbaycan dilinin inkişaf etdirilməsi və beynəlxalq aləmdə nüfuz qazanması üçün dilin tədrisi məsələlərinə yanaşmalarda köklü dəyişikliklərin edilməsi məqsədəuyğun hesab edilir. Bu məqsədlə dövlət səviyyəsində bir sıra fərman və sərəncamlar imzalanmışdır. “Azərbaycan dilinin saflığının qorunması və dövlət dilinin istifadəsinin daha da təkmilləşdirilməsi ilə bağlı tədbirlər haqqında” 1 noyabr 2018-ci il tarixli Fərmanda dilimizin imkanlarından yetərincə və düzgün istifadə edilməməsi, habelə televiziya və radio kanallarında, internet resurslarında, mətbu nəşrlərdə və reklam daşıyıcılarında ədəbi dilin normalarının kobud şəkildə pozulması, leksik və qrammatik qaydalara əməl edilməməsi, məişət danışığından istifadə olunması, əcnəbi söz və ifadələrin yersiz işlədilməsi və bu sahədə hüquqi tənzimləmənin və daimi nəzarətin tətbiq olunması tələbi öz əksini tapır [9]. “Azərbaycan dilinin elektron məkanda daha geniş istifadəsinin təmin edilməsi ilə bağlı bir sıra tədbirlər haqqında” 2018-ci il 17 iyul tarixli Sərəncamda Azərbaycan dilinin elektron məkanda daha geniş istifadəsinin, bu dillə maraqlanan şəxslər üçün əlçatanlığın təmin edilməsi və öyrədilməsinin asanlaşdırılması məsələləri öz əksini tapmışdır [10].

Son illər xarici ölkələrdən Azərbaycana əcnəbilərin miqrasiyasında xeyli artım müşahidə edilir. Hər il yüzlərlə xarici ölkə vətəndaşı müxtəlif məqsədlərlə (təhsil almaq, işləmək, ticari əlaqələr qurmaq və s. üçün) ölkəmizə miqrasiya

edirlər. Məhz bu səbəbdən, ölkə vətəndaşları ilə yanaşı, əcnəbilərin də Azərbaycan dili bilgilərinin vəziyyətinin qeyri-müəyyənliyi və onun aradan qaldırılması üçün zəruri tədbirlərin görülməsi aktual məsələlərdən biridir. Eyni zamanda respublika və xarici ölkə vətəndaşlarına Azərbaycan dili bilgilərini inkişaf etdirmələri üçün stimullaşdırıcı vasitələrin (sertifikatların verilməsi) və digər imkanların yaradılması zəruridir.

Ölkədə konstitusion tələb olaraq həyata keçirilən dil siyasətindən irəli gələn vəzifələrin reallaşdırılmasına dəstəyin verilməsi məqsədiylə atılan addımlardan biri də məhz beynəlxalq standartlar əsasında Azərbaycan dilində dil səviyyələrinin işlənilməsi və dil biliklərinin qiymətləndirilməsinin bu səviyyələr üzrə aparılmasıdır.

IV. AZƏRBAYCAN DİLİNİN TƏDRİSİ VƏ QIYMƏTLƏNDİRİLMƏSİNDƏ İKT-NİN ROLU

İKT-nin sürətlə inkişaf etdiyi müasir dövrdə yüksək texnologiyalı milli iqtisadiyyatın və İnformasiya cəmiyyətinin yaradılması ölkələrin dövlət siyasətinin prioritet istiqamətindən biri hesab olunur. İKT bu gün bir çox sahələrdə mühüm amilə çevrilmişdir və dövlətlərin siyasi, iqtisadi və sosial fəaliyyətinə aktiv təsir edərək iqtisadiyyatın və ictimai münasibətlərin qloballaşmasına da öz təsirini göstərməkdədir.

Təhsil sahəsində İKT-nin tətbiqi mühüm əhəmiyyət kəsb edir. Xüsusilə, müasir texnologiyaların verdiyi imkanlar sayəsində

dillərin öyrənilməsində daha yaxşı nəticələr əldə etmək mümkündür. Dil mütəxəssisləri tərəfindən aparılan araşdırmalar göstərir ki, dilin tədrisində İKT-nin tətbiqi öyrənmə keyfiyyətinin artırılmasına, mövcud informasiya və biliklərə əlyətərlik imkanının yaradılmasına, təhsil xərclərinin azaldılmasına, akademik və elmi mətnlərin keyfiyyətinin, dəqiqliyinin yaxşılaşdırılmasına, öyrənməyə marağın artmasına səbəb olur. Həmçinin İKT-nin tətbiqi müəllimlərə tələbələrin bilik və bacarıqlarını qiymətləndirmək imkanı yaradır. Eyni zamanda, yeni texnoloji vasitələr dil öyrənmələrə ünsiyyət qurmaq, mətni redaktə etmək, öz fikrini izah etmək və mətni tez və çevik şəkildə hazırlamaq imkanı verir [11].

Beynəlxalq təcrübə əsasında Azərbaycan dilinin tədrisi və qiymətləndirilməsi metodologiyasının təkmilləşdirilməsi və mexanizmlərinin işlənilməsi də İKT-nin tətbiqinə əsaslanmalıdır. Lakin Azərbaycan dilinin tədrisində İnternet resurslarının, elektron dərsliklərin az olması, elektron məkanda Azərbaycan dili ilə bağlı kifayət qədər məlumatların olmaması dilin qiymətləndirilməsi metodologiyasının və mexanizmlərinin işlənilməsində çətinliklər yarada bilər.

Bu istiqamətdə dövlət tərəfindən bir sıra mühüm addımlar atılmışdır. “Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsinə və ölkədə dilçiliyin inkişafına dair Dövlət Proqramı”nın təsdiq edilməsi haqqında Sərəncam bunun ən bariz nümunəsidir. Qəbul olunmuş dövlət proqramında

Azərbaycan dilinin lüğət tərkibinin və qrammatik normalarının sistemləşdirilməsi, Azərbaycan dilini öyrənən xarici vətəndaşlar və xarici dil öyrənmək istəyən azərbaycanlılar üçün tədris vəsaitlərinin daha intensiv hazırlanmasının təmin olunması, Azərbaycan dilinin düzgün işlənməsində çevikliyin təmin olunması üçün müvafiq internet resurslarının yaradılması, Azərbaycan dilində internet resursların, elektron və interaktiv dərsliklərin yaradılması sahəsində işlərin görülməsi və s. kimi məsələlər öz əksini tapır [12].

Azərbaycan dilinin tədrisi və qiymətləndirilməsi üçün Azərbaycan dilinin ümumi informasiya bazasının yaradılması, terminoloji lüğətlərin hazırlanması, Azərbaycan dilində internet resurslarının, elektron və interaktiv dərsliklərin yaradılması, məşin tərcüməsi sistemlərinin və digər müasir tətbiqi – linqvistik texnologiyaların yaradılması mühüm prioritet istiqamətlərdir.

NƏTİCƏ

Azərbaycan dil biliklərinin qiymətləndirilməsi sahəsində Avropa təcrübəsi əsasında yeni meyarların işlənilməsi zəruridir. Tədqiqat işində CEFR tövsiyələr toplusu əsasında Azərbaycan dilinin linqvistik xüsusiyyətlərini nəzərə almaqla “Azərbaycan dili üzrə səviyyə imtahanı”nın təşkili təklif olunur. Həmçinin müasir Azərbaycan dilinin tədrisinin və qiymətləndirilməsinin iki istiqamətdə aparılması (Azərbaycan Respublikasının vətəndaşları və əcnəbi vətəndaşlar üçün)

nəzərdə tutulur. Dil biliklərinin qiymətləndirilməsinin dörd dil bacarığı üzrə aparılması məqsədəuyğundur. Bu məqsədlə ayrıca tədris planı və bu tədris planına uyğun dərsliklər və dərs vəsaitlərinin hazırlanması zəruridir. Qeyd olunan təkliflərin həyata keçirilməsində İKT-nin dəstəyi ilə elektron bazanın yaradılması təmin edilməlidir. Azərbaycan dili üzrə səviyyə imtahanları müxtəlif sahələrdə müxtəlif məqsədlər (təhsil, işə qəbul, miqrantlarla iş və s.) üçün tətbiq oluna bilər.

İSTİNADLAR

- [1] A. Qurbanov. Müasir Azərbaycan ədəbi dili. I cild (2 cildə). Bakı, 2019. 464 səh.
- [2] Interagency Language Roundtable, <https://www.govtilr.org/Skills/IRL%20Scale%20History.htm>.
- [3] American Council on the Teaching of Foreign Languages, https://www.actfl.org/uploads/files/general/AC-TFLProficiencyGuidelines2012_Azerbaijani.pdf.
- [4] Common European Framework of Reference for Languages: Learning, Teaching, Assessment, Council of Europe, April 2020, Companion volume, 278 p.
- [5] Introductory Guide to the Common European Framework of Reference (CEFR) for English Language Teachers, Cambridge University Press, 2013, 12 p.
- [6] Wikipedia Free Encyclopedia: List of language proficiency test, https://en.wikipedia.org/wiki/List_of_language_proficiency_tests
- [7] Ş. Nəbiyeva, Ana Dili Tədrisi Metodikasının Aktual Problemləri (metodik vəsait), Bakı, 2010, 64 səh.
- [8] “Azərbaycan Respublikasında ümumi təhsilin konsepsiyasının (Milli Kurikulumun) təsdiq

- edilməsi haqqında” Azərbaycan Respublikası Nazirlər Kabinetinin 2006-cı il 30 oktyabr tarixli qərarı. <https://e-qanun.az/framework/19682>.
- [9] “Azərbaycan dilinin saflığının qorunması və dövlət dilinin istifadəsinin daha da təkmilləşdirilməsi ilə bağlı tədbirlər haqqında” Azərbaycan Respublikası Prezidentinin Fərmanı, 1 noyabr 2018-ci il. <https://president.az/az/articles/view/30563>
- [10] “Azərbaycan dilinin elektron məkanda daha geniş istifadəsinin təmin edilməsi ilə bağlı bir sıra tədbirlər haqqında” Azərbaycan Respublikası Prezidentinin Sərəncamı, 17 iyul 2018-ci il. <https://e-qanun.az/framework/39503>
- [11] S. Ammanni, U. Aparanjani, The Role of ICT in English Language Teaching and Learning, International Journal of Scientific & Engineering Research, Volume 7, Issue 7, July-2016, ISSN 2229-5518
- [12] “Azərbaycan dilinin qloballaşma şəraitində zamanın tələblərinə uyğun istifadəsinə və ölkədə dilçiliyin inkişafına dair Dövlət Proqramı”nın təsdiq edilməsi haqqında Azərbaycan Respublikasının Prezidenti Sərəncamı, 2013-cü il 9 aprel 2013-cü il. <https://president.az/az/articles/view/7744>
- [13] Lala N. Adiyeva, Rəsmiyyə Ş. Mahmudova, Analysis of international experience in the field of teaching and assessment of language proficiency, Problems of Information Society, 2023, vol.14, no.1, 85-92

Həştəqlər xüsusi linqvistik hadisə kimi

Könül Həbibova

Azərbaycan Milli Elmlər Akademiyası, İ.Nəsimi adına Dilçilik İnstitutu, Bakı,
Azərbaycan

Konul_habibova@mail.ru

Xülasə — Bu tezisdə virtual məkanda istifadə edilən həştəqlər xüsusi linqvistik hadisə kimi nəzərdən keçirilir. Rəqəmsal texnikanın daha geniş istifadə edildiyi müasir dövərdə həştəqlər qloballaşmış informasiya mübadiləsi prosesinin vacib elementi olduğundan onların dilçilik baxımından araşdırılması xüsusi aktualıq kəsb edir. Araşdırmanın əsas məqsədi həştəqlərin praktik-linqvistik özəlliklərinin araşdırılmasından ibarətdir.

Açar sözlər — *həştəq; pragmatika; dilçilik; virtual məkən; linqvistik hadisə*

I. GİRİŞ

Həştəqlərin linqvistik hadisə kimi öyrənilməsi onların virtual məkanda dil elementlərindən necə istifadə olunmasının öyrənilməsi istiqamətində araşdırılır. Həştəqlər virtual məkanda bu və ya digər informasiya yüklü yazının məzmununu təsvir etmək üçün istifadə edilən “#” işarəsi ilə başlayan açar sözlər və ya ifadələrdir.

Dil fenomeni kimi həştəqlər üzərində aparılan araşdırmalar, adətən, insanların virtual şəxsiyyət yaratmaq, emosiyaları ifadə etmək, digər istifadəçilərlə ünsiyyət qurmaq və qrup şəkilli müzakirələrdə iştirak etmək üçün həştəqlərdən necə istifadə etdiyinə diqqət yetirir. Linqvistik tədqiqatlar, həmçinin həştəqlərin

strukturunu, metaforik mənalarını, həştəq yaradarkən dialekt formalarından istifadəni və digər aspektləri təhlil edə bilər.

Müasir dövərdə virtual məkanda geniş istifadə edilən həştəqlər xüsusi linqvistik fenomen hesab olunur. Həştəqlər yalnız müxtəlifməzmunlu kontekstləri deyil, həm də sıxılmış formada həştəq vasitəsilə ötürülən əsas hadisə və ya hərəkət haqqında məlumatın virtual məkanda “tanınmasını”, onun kimin maraqlarına uyğun olmasını da təmin edir, həmçinin assosiativ bağlantıları və konnotasiyaları bərpa etməklə həştəqləri şərh etməyə imkan verir [1]. Hesab edirik ki, həştəqlər internet üzərində qurulan ünsiyyət prosesi zamanı böyük kommunikativ potensiala malikdir. Bir çox müasir ictimai hərəkətlərin iştirakçıları həştəqlərdən öz ideologiyalarının internet üzərindən daha sürətli şəkildə yayılması və geniş auditoriyaya çatdırılması məqsədi ilə istifadə edirlər.

Həştəqlərin linqvistik hadisə kimi tədqiqi aktual mövzudur, çünki insanların sosial media və internet üzərindən ünsiyyəti bütün dünyada çox geniş yayılmış və şəbəkələr insanlar üçün vacib amilə çevrilmişdir. Həştəqlər isə sosial şəbəkələrdəki ünsiyyətin əsas elementidir və müxtəlif fikir və emosiyaları ifadə etmək üçün geniş şəkildə istifadə edilir.

Həşteqlərin öyrənilməsi dəyərli elmi və praktik nəticələr verə bilər, çünki o, insanların onlayn mühitdə məna yaratmaq və sosial kimliklərini ifadə etmək üçün dildən necə istifadə etdiyini anlamağa kömək edə bilər. Bu, marketing, siyasi kampaniyalar, mədəniyyətlərarası ünsiyyət, həmçinin sosial medianın ünsiyyət təcrübəsinə və ümumilikdə mədəniyyətə necə təsir etdiyini anlamaq üçün vacibdir.

Üstəlik, həşteqlər daim dəyişən və inkişaf edən dinamik dil hadisəsidir, ona görə də onlar haqqında öyrənmək onlayn mühitdə dil dinamikasını və təkamülünü izləməyə kömək edə bilər.

II. TƏDQIQATIN METODLARI

Bu araşdırma zamanı həşteqlərin linqvistik hadisə kimi təhlili məqsədi ilə dilçiliyin təsviri, təsviri-müqayisəli metodlarından istifadə edilmişdir. Eləcə də dil materialları virtual şəbəkələr üzərindən seçildiyindən burada kontent-analiz metoduna da müraciət edilmişdir. Bu metodun tətbiqi ilə həşteqlərin istifadə tezliyi, onların mövzu əhatəsi, emosional çalarları və digər parametrlərini araşdırmaq mümkündür.

III. HƏŞTƏQLƏRİN LİNQVİSTİK MAHİYYƏTİ

Həşteqlər linqvistik hadisədir, çünki onlar müəyyən mövzuları, ideyaları, hadisələri və ya emosiyaları təsvir etmək üçün dildə istifadə olunan söz və ya ifadələrdən ibarətdir. Dil fenomeni olaraq həşteqlər aşağıdakı xüsusiyyətlərə malikdir:

Linqvistik struktur: həşteqlər “#” işarəsindən sonra heç bir boşluq

buraxılmadan söz və ya ifadələrin ardıcılığından ibarətdir.

Semantik məna: həşteqlərin spesifik semantikasi var, yəni konkret mövzuları, ideyaları, hadisələri və ya emosiyaları təsvir etmək üçün istifadə olunur.

Kontekstlə əlaqə: həşteqlər konkret kontekstdə istifadə olunur, yəni harada və necə istifadə olunduğundan asılı olaraq müxtəlif mənalar kəsb edir.

Ekspressivlik: həşteqlər gülüş, kədər, qıcıq və s. kimi emosiya və hissləri ifadə etmək üçün istifadə edilə bilər.

Metaforik məna: həşteqlərdə müəyyən olaylara bağlı metaforalar və eyhamlar ola bilər ki, bu da onlardan sosial kimliyi ifadə etmək vasitəsi kimi yararlanmağa imkan verir.

Dinamik: həşteqlər daim dəyişən mövzulara bağlı olaraq dəyişir və inkişaf edir.

Ümumiyyətlə, həşteqlərin linqvistik mahiyyəti ondan ibarətdir ki, onlar virtual məkanda və sosial şəbəkələrdə açar söz funksiyalı linqvistik ünsiyyət vasitəsi kimi çıxış edir.

Qeyd etmək lazımdır ki, “müasir zamanda ictimai cərəyanların əhəmiyyətli bir hissəsi məhz virtual məkanda yaranır və inkişaf etdirilir. Sonra bu hərəkətlərlə bağlı həşteqlər formalaşdırılır ki, məhz bunun sayəsində də hərəkət iştirakçıları internet vasitələri üzərindən bir araya gələ bilmək şansı qazanır” [1]. Məsələn, bu kimi hərəkətlərdən birinə COVID19 pandemiyası dövründə *#evdəqalsaglamqal* həşteqini misal göstərmək olar. Bu həşteq altında insanlara təhlükəli virusdan qorunma yolları və vasitələri haqqında ətraflı məlumat və məsləhətlərin verilməsi istiqamətində geniş

və tutumlu, eləcə də əhəmiyyətli məlumatın yayılması uğurla həyata keçirilmişdir. İnternet üzərindən qurulan ünsiyyət müasir insanların məişətinin ayrılmaz hissəsinə çevrilmişdir.

L.Y.Şipitsinanın fikrincə, həşteqlər “...kommunikasiya vasitələrinin inkişafının tarixi mərhələlərindən birini təmsil edir” [2] ki, onun xarakterik əlaməti kimi kompüter və informasiya texnologiyaları vasitələri bu prosesdə aparıcı vasitə rolunu oynayır. İnternet ictimai ünsiyyət sferasını kökündən dəyişdirərək, subyektlərinin sayını artırmaqla, onların öz fikirlərini ifadə etmək azadlığını, informasiya resurslarına və auditoriyaya çıxışını genişləndirmişdir. İnternet kommunikasiya bir sıra xüsusiyyətlərə malikdir, xüsusən də buraya dil resurslarından maksimal şəkildə lakonik olaraq istifadə edilməsi daxildir. “O, mətnin yığcamlığına, dinamikliyinə kömək edir, onu üstüörtülü şəkildə ifadə olunan mətnə mənaları ilə zənginləşdirir, mətnə məlumatın lazımsız təkrarlanmasından xilas edir, çünki onun sayəsində daha az informativ əhəmiyyət kəsb edən məlumatlar ya qismən buraxılır, ya da tam olaraq ixtisar edilir” [3]. İnternet istifadəçiləri kontentə uyğun davranarkən dolayı və ya birbaşa ictimai-siyasi proseslərə təsir edən ictimai davranışları formalaşdırırlar.

Eyni hal sosial şəbəkələr üzərindən paylaşılan məlumatlar, həşteqlər və ya məntiq vasitəsilə təsnif edilmiş məlumat obyektinə ilə əlaqəli etiketlərdən istifadə edərək birgə təsnif edildikdə baş verir [1]. Hazırda həşteq yalnız fərdi tvitləri və yazıları qeyd etmir, həm də onlardan əsas ideyanı və ya yaradılmış obrazı ön plana çıxarmaqla məlumatın pragmatik və

manipulyativ potensialını bir araya cəmləyir [4].

A. Həşteqlərin linqvistik funksiyaları

Həşteqlərin ilkin funksiyasından danışarkən qeyd etmək lazımdır ki, onların istifadəsinin rahatlığı informasiya axınının sıxlığının artması və informasiyanın həcmində idarə olunması, həmçinin məlumatın saf-çürük edilməsi zərurəti ilə bağlıdır. Qrafiklərin köməyi ilə əldə edilən məlumatlara əsasən, #qarabağazərbaycandır həşteqi ilə eyni kontekstdə ən çox rast gəlinən həşteqlər sistemləşdirilmiş, onların semantik təhlili aparılmışdır. Biz “həşteq buludunun əsas həşteq tərəfindən ifadə olunan konsepsiyaları qərəzsiz şəkildə müəyyən edən çoxqatlı və çoxvektorlu konstruksiya” [1] olması faktını qəbul edirik.

Həşteqlər müxtəlif İnternet istifadəçiləri tərəfindən hazırlanmış müxtəlif tutumlu informasiyaları bir məzmun ətrafında birləşdirir və müxtəlif həşteqlərin bir kontekstdə birləşməsi vahid məzmun sahəsini təşkil edir. Hər bir həşteq “insan şüurunun konseptual şəbəkəsinin davamlılığını əks etdirən qapalı şəbəkə yaradaraq bir çox keçidlərə uyğun gəlir” [5].

İnternet məkanda rastlanan həşteqlərin nominativliyinə nəzər saldıqda qeyd etmək lazımdır ki, bunlar arasında struktur baxımdan sadə nominativ quruluşa malik olanlar çoxluq təşkil edir ki, bunlarda da əsasən baş verən hadisələrin əsas iştirakçılarının adlarından, baş verən hadisələrin vaxət olduğu yerlərin, məkanların adlarından çağırış məzmunlu predikativ konstruksiyalardan istifadə edilir.

Virtual ünsiyyətdə əsas dil tendensiyası kimi etiketlərin qısalığı onlarla hadisə arasında güclü assosiativ əlaqələr yaratmağa və müvafiq tematik məlumatların axtarışını sadələşdirməyə imkan verir. Həştəqlər, həmçinin mürəkkəb emosional cavabları kontekst və dilin üslubi imkanları və fiqurları vasitəsilə ötürür. Onların yayılma sürəti və müzakirə olunan hadisənin qavranılmasının emosional vektorunu formalaşdırmaq qabiliyyəti diqqətəlayiqdir.

Danılmaz reallıqdır ki, bəşəriyyət dünyanın müasir inkişaf mərhələsində qloballaşma prosesinin təsiri altındadır, bunun nəticəsində ölkələrin iqtisadi, mədəni, siyasi və elm sahəsi daha da genişlənir. Getdikcə artan sayda yeni leksemlər meydana çıxır ki, onların arasında həm qrafik, həm də semantik səviyyədə linqvistik sıxılma prosesinə məruz qalan leksik vahidlər geniş yer tutur.

İnternetin müasir virtual məkanının mühüm hissəsini ingilisdilli saytlar tutur. Veb-səhifələrin və veb-saytların 50%-dən çoxu hələ də ingilis dilindən əsas dil kimi istifadə edir, lakin digər dillərdə təqdim olunan İnternet məlumatlarının müəyyən seqmentlərinin daha sürətli böyüməsi səbəbindən bu nisbət tədricən azalır. Tədqiqatların birində qeyd edilir ki, “W3Techs-də (World Wide Web Technology Surveys), 01.04.2011-ci ildən 01.11.2017-ci ilədək olan göstəricilərə əsasən, İnternetdəki saytların ingilis dilindən istifadə nisbəti 6% civarında azalaraq 57,6%-dən 51,2%-ə enmişdir” [6].

Virtual kommunikasiya zamanı meydana gələn bütün dil və nitq hadisələri bu və ya digər dərəcədə ingilis dilinin təsiri ilə formalaşır və ya bu dilin qrammatik qanunları, ya da leksik tərkibinə daxil olan elementlərin iştirakı ilə meydana gəlir. Bu baxımdan söyləmək lazımdır ki, həştəqlərdə də ingilis dili elementlərindən geniş istifadə edilir. Bir çox hallarda isə açması ingiliscə olan abreviaturlardan da istifadə edilir. “Bundan əlavə, dilin sintetik və ya analitik xüsusiyyətlərinin abreviasiya proseslərinə fərqli təsirini sintetik dillərdə daha az, ingilis dili kimi analitik dildə isə ən geniş şəkildə ifadə olunan semantik törəmənin inkişafını müşahidə etməklə tapmaq olar [1].

Bundan əlavə, müasir zamanımızın tanınmış dilçi alimlərindən olan Y.A.Dyujikovanın təbirincə desək, bizim dövrümüzdə ixtisarin sözyaradıcı proses kimi tətbiq tezliyi baxımından ingilis dili yeni leksik vahidlərin əmələ gəlməsinin əsas yollarına meyl edir və konversiya, affiksasiya və mürəkkəbləşmə ilə yanaşı öz yerini tutur. Bu, böyük ölçüdə həyatın kompüterləşdirilməsinin və kommunikanın internet resurslarından intensiv istifadəsinin nəticəsidir [7]. Semantik səviyyədə linqvistik kompressiya prosesi virtual ünsiyyətdə əsas siyasi, iqtisadi, sosial və digər hadisələri əks etdirən həştəqlər vasitəsilə çox intensiv şəkildə həyata keçirilir. Nəticə etibarilə, cəmiyyətdə baş verən dəyişiklikləri başa düşmək üçün bu linqvistik transformasiyaları öyrənmək aktual görünür.

Həştəqlərin virtual məkandakı funksiyalarını ümumiləşdirdikdə, onları aşağıdakı şəkildə qeyd etmək olar:

Həştəqlər identifikasiya funksiyası yerinə yetirir: xüsusi mövzuları, ideyaları, hadisələri və ya emosiyaları müəyyən etmək və təsvir etmək üçün istifadə olunur ki, bu da istifadəçilərə məlumatları tez tapıb təşkil etməyə kömək edir.

Həştəqlər kommunikativ funksiyaya malikdir: ümumi maraqları bölüşən istifadəçilər arasında əlaqə yaratmağa kömək edir, həmçinin nəşrlərin auditoriyasını genişləndirməyə kömək edir.

Həştəqlər modifikator funksiyası yerinə yetirir: istinad olunan sözün və ya ifadənin mənasını dəyişdirmək üçün istifadə edilə bilər. Məsələn, #training həştəqi fiziki məşqlərə istinad etmək üçün istifadə oluna bilər, lakin zehni məşqlərə və ya öz işinə istinad etmək üçün də istifadə edilə bilər.

Həştəqlərin yaradıcı funksiyası da mövcuddur: yeni terminlər və ifadələr yaratmaq üçün istifadə edilə bilər ki, bu da dil təcrübəsini genişləndirməyə və istifadəçilərin yaradıcılığını stimullaşdırmağa kömək edir.

Ümumiyyətlə, həştəqlər virtual kommunikasiya istifadəçilərinə effektiv ünsiyyət qurmağa, digər istifadəçilərlə əlaqə saxlamağa və sosial kimliklərini ifadə etməyə kömək edən bir çox linqvistik funksiyaları yerinə yetirir.

B. Həştəqlərin ideoloji potensialı

İnternet məkənində “#QarabağAzərbaycandır” həştəqinin geniş yayılması və Azərbaycan Respublikasının ərazi bütövlüyünün vahid ideyasına çevrilməsi Azərbaycan Respublikasının

Prezidenti cənab İ.Əliyevin Ermənistan Respublikasının Prezidenti N.Paşinyan arasındakı məşhur debatından sonra geniş vüsət almışdır. “#qarabağazərbaycandır” həştəqi Azərbaycan Respublikasının ərazi bütövlüyünə qarşı qeyri-dəqiq və ya təxribat xarakterli məlumatların yer aldığı kontentlərdə aktuallaşır və bu istiqamətdə hər hansı qeyri-dəqiq informasiyanın formalaşması zamanı trendə çevrilərək virtual məkanda bu ideya ətrafında bütün dünya azərbaycanlılarının, eləcə də dövlətimizin maraqlarını müdafiə edən qüvvələrin bir arada birləşməsinə təmin edir. Məsələn, internet üzərindən “#qarabağazərbaycandır” həştəqini axtarış panelində açar sözü kimi daxil etdikdə axtarışın nəticəsi kimi bununla əlaqədar olan aşağıdakı həştəqlərin də verildiyini müşahidə etmək mümkündür: #azerbaijan #baku #weki #pakistan #kewvet #türkazerikardeşliyi #azerbaycantürkiye #armenia #арменя #şəhidlərölməzvətənbölməz #QarabağAzərbaycandır #keşfet #qarabağbizimdir #qarabağazərbaycandır #azərbaycanordusugüclüdür və s.

#qarabağazərbaycandır həştəqinin virtual məkanda fərqli yazılış formalarına da rast gəlmək mümkündür. Məsələn, “#qarabağ_azərbaycandır” və ya “#QarabağAzərbaycandır”.

Həştəqlərdə öz əksini tapan “ad-etiketlər təyin olunmuş xarakter və ya fenomenin müəyyən çalarlı təsvirini yaratmaq vasitəsinə çevrilir. Belə bir funksiyada bu və ya digər həştəqin kütləvi istifadəsi media hadisəsinə çevrilir və media səhifələrinə

daxil olur” [8]. Həştəq şəklində şüarların yaradılması müxtəlif insanları və cərəyanları bir araya gətirməyə kömək edir. Beləliklə, bu dil vahidi xüsusi mədəni əhəmiyyət kəsb edir.

NƏTİCƏ

Bu araşdırma zamanı əldə edilən nəticələri aşağıdakı kimi müddəalar şəklində qeyd etmək olar:

- Həştəqlər sosial şəbəkələrdə ünsiyyət və təşkilatlanma üçün böyük əhəmiyyət kəsb edən mühüm linqvistik və sosial fenomendir;
- Həştəqlər mövzunu işarələmək, duyğuları və münasibətləri ifadə etmək, icmaları təşkil etmək və məlumat yaymaq kimi müxtəlif linqvistik funksiyaları yerinə yetirir;
- Həştəqlər ideoloji potensiala malik ola bilər və insanları müəyyən ideyalar, dəyərlər və təşəbbüslər ətrafında səfərbər etmək, həmçinin siyasi, sosial, mədəni və ekoloji baxış və mövqeləri ifadə etmək üçün istifadə edilə bilər.

Yekun olaraq vurğulamaq lazımdır ki, həştəqlərin dil hadisəsi kimi tədqiqi aktualdır və həştəqlərin linqvistik xüsusiyyətlərinin müəyyən edilməsi və təhlili, onların sosial və ideoloji potensialının öyrənilməsi, həmçinin sosial mediada məlumatların tədqiqi və təhlili metodlarının işlənilməsi hazırlanması kimi müxtəlif məqsədlərə xidmət edə bilər.

Həştəqlərin öyrənilməsi marketing, siyasət, sosiologiya, dilçilik və s. kimi

müxtəlif sahələr üçün praktik əhəmiyyət kəsb edə bilər.

İSTİNADLAR

- [1] Н.А. Рябченко, О.П. Малышева, «Управление политическим контентом: семантический анализ тегов и хэштегов» // VIII международная социологическая Грушинская конференция «Социолог 2.0: трансформация профессии». М., 2018, с. 362-366.
- [2] Л.Ю. Щипицина, «Компьютерно-опосредованная коммуникация: Лингвистический аспект анализа». М., 2010.
- [3] А.П. Атягина, «Языковая компрессия в Twitter» // Медиаскоп. 2012 №3, с. 4, URL: <http://www.mediascope.ru/node/1136>.
- [4] В. Емельяненко, «Слова за решёткой: Хештег в 2012 году стал новой нормой интернет-языка» // Русский мир.ru, 2013. № 2. URL: <https://rusmir.media/2013/02/01/reshetka>
- [5] А.А. Беловодская, «Об исследовательском потенциале изучения хэштега как медиадискурсивного феномена» // Медиалингвистика. 2019, № 6 (1), с. 60-74. DOI: 10.21638/spbu22.2019.105.
- [6] А.В. Алексеев, «Процесс сокращения хештегов в современной виртуальной коммуникации» // Научный диалог. 2019, № 10, с. 9-8.
- [7] Е.А. Дюжикова, О.В. Афанасьева, «Парадокс «идентичности» инициально-словных сокращений современного английского языка» // Вестник Московского государственного областного университета. Серия : Лингвистика. 2017, № 6, с. 50-58.
- [8] Ю.Е. Галямина, «Лингвистический анализ хештегов Твиттера» / В кн: Современный русский язык в интернете. М.: Просвещение, 2014.

Azərbaycan dilinin tədrisi prosesində intellektual oyunların tətbiqi məsələləri

Bahar İsmayılova

İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan

www.biaz.ucoz.net, bahar_ismayilova@yahoo.com

Xülasə — Məqalədə Azərbaycan dilinin tədrisi prosesində qarşıya çıxan bəzi problemlər araşdırılmış, mövzuların mənimsənilməsi, qiymətləndirilməsi, asudə vaxtın səmərəli təşkilində, Azərbaycan dilinin tədrisinin müasir tələblərə uyğunlaşdırılmasında İKT-nin rolunun və tətbiqinin üstünlükləri qeyd olunmuş, təlimin keyfiyyət göstəricilərinin yüksəldilməsi üçün müxtəlif intellektual oyunlardan, xüsusilə krosvordların hazırlanması və həlli, onlayn yarışlardan istifadənin zəruriliyi əsaslandırılmışdır.

Açar sözlər — *tədris yönümlü intellektual oyunlar; dayanıqlı daxili motivasiya, krosvord; onlayn yarışlar; təlimin fərdiləşdirilməsi; özünüqiymətləndirmə; özünütəhsil; özünütədqiqmə*

I. GİRİŞ

İnformasiya cəmiyyətində təhsilin modernləşdirilməsi üçün hansı texnologiyalardan necə istifadə edilməsi aktual məsələlərdən biridir. Bu gün dünya iqtisadiyyatı daha çox bilik yönümlü olduğundan ölkələrdə müasir tələblərə cavab verən təhsil sisteminin qurulması tələb olunur. Bu amil inkişaf etmiş ölkələrdə təhsilin informasiyalaşdırılması istiqamətində sistemli şəkildə bir neçə

mərhələdə islahatların keçirilməsinə səbəb olmuşdur. Neçə illərdir ki, respublikamızda ictimai və iqtisadi sahələrdə olduğu kimi təhsil sferasında da ciddi islahatlar aparılır [1]. “Təhsil millətin gələcəyidir” deyərək təhsilə yüksək qiymət verən və onun perspektivlərini dahiyənə şəkildə göstərən ulu öndər Heydər Əliyevin layiqli davamçısı Azərbaycan Respublikasının Prezidenti İlham Əliyev təhsili belə qiymətləndirir: “Təhsil bizim gələcəyimizdir, inkişaf perspektivlərimizi müəyyən edən ən başlıca sahədir”. Bu ideyalar əsasında bütün fənlərin, o cümlədən Azərbaycan dili fənni üzrə təhsilin yeni məzmunu, məzmun standartları, fənn kurikulumu, buna uyğun olaraq, dərslik və dərs vəsaitləri işlənib hazırlanmış, davamlı olaraq müxtəlif dövlət proqramları həyata keçirilir [2]. “Azərbaycan dilinin saflığının qorunması və dövlət dilindən istifadənin daha da təkmilləşdirilməsi ilə bağlı tədbirlər haqqında” 1 noyabr 2018-ci il tarixli fərmanda bu istiqamətdə ardıcıl dövlət siyasətinin aparılmasının nə qədər zəruri olduğu vurğulanır [3].

Kompüter texnologiyalarına dərindən yiyələnmək, informasiya-kommunikasiya texnologiyalarından (İKT) yüksək səviyyədə istifadə bacarıqlarına malik

olmaq müasir dövrün tələbidir. Ona görə də ümumtəhsil məktəblərində Azərbaycan dili fənninin tədrisi üsullarının daha da təkmilləşdirilməsi, İKT ilə əlaqəli yeni metodların işlənməsi və şagirdlərin yaş xüsusiyyətlərindən asılı olaraq müxtəlif vasitələrdən istifadə etməklə onların bu fənnə marağının artırılması olduqca zəruridir. Tədrisdə İKT-nin tətbiqinin bir çox üstünlükləri vardır:

- Qavrama imkanlarının, öyrənmə keyfiyyətinin artırılması;
- Öyrənməyə marağın yaradılması, artırılması;
- Tədrislə bağlı hər cür informasiyaya əlyətərlik imkanı;
- Bilik və bacarıqların qiymətləndirilməsi, özünüqiymətləndirmə, özünütəkmilləşdirmə və s.

Gündən-günə təkmilləşən, inkişaf edən İKT-nin imkanlarından Azərbaycan dilinin, eləcə də bütün fənlərin tədrisində tətbiq edilən müxtəlif məqsədlərlə istifadə etmək olar. Müxtəlif səviyyəli inteqrativ, interaktiv krosvordlardan, işgüzar oyunlardan istifadə gənclərin müxtəlif mənbələrlə işləməsinə, idraki fəallığının artmasına, özünüqiymətləndirmə və özünütəkmilləşdirməsinə şərait yarada bilər.

II. AZƏRBAYCAN DİLİ FƏNNİNİN TƏDRİSİNDƏ QARŞIYA ÇIXAN PROBLEMLƏR

Neçə illərdir ki, məktəblilərin, gənclərin söz ehtiyatının azlığı, bir çox hallarda informasiya texnologiyalarından əyləncə

məqsədlə istifadə etməsi müasir dövrün aktual problemlərindən biri olaraq müəllimləri, valideynləri çox narahat edir. Bu çətinliklərin səbəbləri ilə bağlı sorğunun nəticəsində aldığımız cavablar aşağıda təsvir olunmuşdur:

- Maddi-texniki bazanın lazımi səviyyədə olmaması;
- Təlimin fərdiləşdirilməsi, hər bir şagirdin müstəqil fəaliyyətinin təşkilində yaranan çətinliklər;
- Şagirdlərdə fənni öyrənmək üçün motivasiyanın qənaətbəxş olmaması.

Xüsusilə region məktəblərində maddi-texniki baza lazımi səviyyədə olmadığı üçün dərs prosesində aşağıdakı kimi problemlər yaranır:

- İnternet vasitəsilə araşdırma tipli tapşırıqların yerinə yetirilməsi çətin olur;
- Təlimin fərdiləşdirilməsi, praktiki işlərin təşkilində çətinliklər yaranır;
- Dayanıqlı öyrənmə motivasiyasının yaradılması (şagird özü istəməsə heç kəs onu içindən gələn istəklə öyrənməyə məcbur edə bilməz!) çətinləşir;
- Özünüqiymətləndirmə, özünütəkmilləşdirmə zamanı çətinliklər yaranır;
- Passiv şagirdi, tələbəni fəaliyyətə həvəsləndirmək, hər dərsdə ən azı 5-7 dəqiqə təhsilənlərin 100% fəallığını təmin etmək çətin olur.

Bütün fənlərin, xüsusilə Azərbaycan dilinin tədrisi prosesində hər bir şagirdin dərsi mənimsəmə səviyyəsini hər dərs

yoxlamaq, biliyində olan zəiflikləri ona anlatmaq, bilmədiyi hissələri öyrətmək vacibdir. Müəllimlər yaxşı bilir ki, bu işi reallaşdırmaq çox çətindir. Ancaq hər bir şagird, tələbə dərslər prosesində öz potensialına uyğun krosvord hazırlayarsa, həll edərsə, qısamüddətli onlayn yarışlarda iştirak edərsə, özünüqiymətləndirmə prosesini aparmış olur və bu fəaliyyət zamanı öz səhvlərini görür. Bir səhifə mətni öyrənmək əvəzinə həmin mövzu üzrə 10-20 interaktiv test işləmək, onlayn yarışlarda iştirak etmək, anında səhvinə, zəif bildiyi hissələri görüb, bilmədiklərini öyrəniyə yenidən yarışa qatılmaq şagird, tələbə üçün daha maraqlıdır və o, yarışlarda yer tutmaq üçün yorulmadan, əzmkarlıqla işləməyə çalışır.

Bəzi saytlardan [4] istifadə etməklə buraxılış, ali məktəbə qəbul imtahanları və s. üçün hazırlanmış testlərdən ibarət elektron baza əsasında keçirilən onlayn yarışlar təhsilalanlara müxtəlif istiqamətlərdə uğur qazanmağa kömək edir:

- Şagirdlərə IX və XI siniflərdə buraxılış imtahanlarını, ali məktəbə qəbul imtahanlarını müvəffəqiyyətlə verməyə yardımçı olur;
- Tələbələrə kollokvium və semestr imtahanlarında uğur qazanmağa kömək edir;
- Magistratura, işə qəbul imtahanlarında və s. istiqamətlərdə fəaliyyət zamanı yüksək nəticələr əldə etməyə əhəmiyyətli dərəcədə müsbət təsir edir.

Hər bir fənnin tədrisi zamanı müəllim ilk növbədə terminlərin, anlayışların

mənimsənilməsinə çalışır. Müxtəlif mövzular üzrə hazırlanan interaktiv krosvordlar, testlərlə iş zamanı təhsilalanlar terminləri daha yaxşı başa düşür, onları müqayisə edir, mövzulara aid bilikləri dərinləşir, praktik bacarıqları artır, təkmilləşir. Fikrimizcə, Azərbaycan dili fənninin tədrisi prosesində müxtəlif səviyyəli, integrativ, interaktiv krosvordlardan, onlayn yarışlardan istifadə özünüqiymətləndirmə, özünütəhsil, özünütəkmilləşdirmə bacarıqlarını, öyrənmə motivasiyasını artırır və nəticədə tədrisin keyfiyyəti yüksəlir.

Araşdırmaların nəticələrinə görə krosvordlar, tədrisyonümlü oyunlar oxuyub-anlama, dərkətmə, problemləri hissetmə və çevik, optimal həll etmə bacarığını artırır. [5].

Amerikada ali məktəblərə qəbul imtahanları zamanı məktəblilərin intellektual qabiliyyətlərini yoxlamaq üçün xüsusi ümumfederal testlərdən istifadə olunur [6]. Bu testlər vasitəsilə həm də şagirdlərin müxtəlif fənləri mənimsəmə səviyyəsi, həyatı bacarıqları da ölçülür. Milli Terminoloji İnformasiya Sisteminin veb-portalında [7] yerləşdirilən lüğətlərdən istifadə etməklə hazırlanan interaktiv, integrativ testlər, krosvordlar vasitəsilə təhsilalanların hər bir fənn üzrə nailiyyətlərini qiymətləndirmək mümkündür.

Hər hansı metod, üsul, vasitənin üstünlüyü aşağıdakı parametrlərlə xarakterizə olunur:

- Əlçatanlıq (müəllim və təhsilalan üçün);

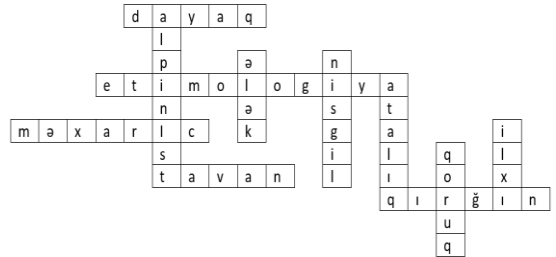
- Təlimi fərdiləşdirmə, özünüqiymətləndirmə, özü-nütəhsil, özünənəzarət, özünüidarəetmə imkanı ilə;
- Müstəqil fəaliyyət zamanı sürət, nəticələrin kəmiyyət, keyfiyyət göstəriciləri ilə;
- Həyati bacarıqların (sosiallaşma, özünüifadə, özünütəsdiqetmə və s.) artması ilə.
- Krosvordların hazırlanması və həlli, mərhələli oflayn və onlayn yarışlar bu baxımdan ideal fəaliyyətlərdən biri olmaqla hər bir təhsillənən üçün əlçatandır.

Krosvordların hazırlanması və həllinin aşağıdakı kimi əhəmiyyəti vardır:

- Təhsilənlərin müstəqil işləməsinə, özünü qiymətləndirməsinə, özünütəhsilə, özünü təkmilləşdirməsinə və reallaşdırmasına imkan yaranır;
- Fənnə, nəzəriyyəni öyrənməyə maraq, kitabla işləmə həvəsi artır;
- Diqqəti mərkəzləşdirmə bacarığı, yaddaş inkişaf edir, fənnə aid terminlər təhlil olunur, yadda saxlanılır, söz ehtiyatı artır, təhsilənlən savadlı yazılışı öyrənir;
- Yaradıcı təfəkkür inkişaf edir, hazırladığı krosvordlar başqa siniflərdə istifadə olunduqda öyrənənənin yaradıcı əməyə motivasiyası yüksəlir, öz broşurunun olmasını arzulayır;
- Başladığı işi axıra çatdırma, çevik düşünmə, lüğətlə işləmə bacarığı təkmilləşir, qazandığı uğurları gördükcə motivasiyası, özgüvən hissi artır;
- Əzmkarlığı artır və şagirdə yarışmaq, özünü təsdiq etmək arzusu yaranır, qətiyyət, məsuliyyət, inadkarlıq tərbiyə olunur;

- Rəqabət hissi şagirdləri daha çətin krosvordlarla işləməyə həvəsləndirir, könüllü olaraq dərscdənkənar saatlarda cütlərlə, qrup şəklində müxtəlif növ suallarla işləyir, idraki fəallıq, nəticədə təlimin keyfiyyəti yüksəlir;
- Şagird krosvordla iş zamanı daha çox düşünür, kreativ olur. Test iş zamanı əzbərcilik vardır, cavablar olduğu üçün şagird bəzən məntiqlə cavabı tapa bilər.

Krosvord hazırlanması və həlli, mərhələli yarışlar fikrimizcə ideal fəaliyyətdir, ən zəif təhsilənlənin inkişafına imkan yaradır (Şəkil 1.). Bir qələm, vərəq, broşur, sayt kifayətdir ki, şagird müstəqil işləsin, anında səhvini görsün, düzəltsin, təkmilləşsin.



Şəkil 1. İntegrativ krosvord nümunəsi

Krosvord hazırlamağın və həll etməyin aşağıdakı imkanları yaranır:

- Sərbəstlik, müstəqillik, seçim azadlığı;
- Məhdudiyətsiz, maraqlı məşqetmə;
- Mərhələlərlə özünüqiymətləndirmə, özünütəhsil, hədəflərə çatma, özünütəsdiq;
- Sağlamlığa müsbət təsir, əyləncə, istirahət etmə imkanı.

NƏTİCƏ

Aparılan tədqiqatların nəticələrinə əsaslanaraq aşağıdakı kimi təkliflər hazırlanmışdır:

- İnformatika, Azərbaycan dili və riyaziyyat fənlərinin integrativ tədrisi ilə bağlı Almaniya təcrübəsinin araşdırılması, qəzet redaksiyalarının əməkdaşları ilə görüşlərin, disputların, ekskursiyaların keçirilməsi təhsilçilərin integrativ biliklərinin dərinləməsinə, həyati bacarıqlarının artmasına çox böyük təsir edə bilər. [8].
- Müəllimlərdə intellektual oyunların tədris prosesinə tətbiqi ilə bağlı bacarıqların formalaşdırılması və təkmilləşdirilməsi üçün mərhələli təlimlərin keçirilməsi təhsil müəssisələrində təlimin keyfiyyət göstəricilərinin artmasına, qiymətləndirmədə olan bəzi problemlərin həllinə müsbət təsir edə bilər.
- Siniflər üzrə müxtəlif səviyyəli tədrisyönlü krosvordlara aid broşurlar hazırlanması, hazırlanmış interaktiv krosvordların disk şəklində broşura əlavə edilərək məktəblərdə yayılması Azərbaycan dilinin tədrisində keyfiyyət göstəricilərinin yüksəlməsinə effektiv təsir edə bilər;
- Azərbaycan dili və informatika fənləri üzrə integrativ dərslərin keçirilməsi bu istiqamətdə görülən işlərə təkan verə bilər.
- V-XI siniflərdə tədrisin məzmununa uyğun olaraq müxtəlif səviyyəli, integrativ, interaktiv krosvordlar, testlər, quizlər, müxtəlif mərhələli oflayn və onlayn yarış nəticələri, integrativ dərslər və s. daxil ediləcəyi saytın hazırlanması şagirdləri bu istiqamətdə aktiv fəaliyyətə stimullaşdırar, Azərbaycan dili fənninin tədrisi keyfiyyətinin yüksəlməsinə müsbət təsir edər, Azərbaycan dili fənninə İKT-nin tətbiqi imkanlarını genişləndirər.
- Multimedia qurğularından tez-tez istifadə monotonluq yaradır, aktiv fəaliyyətə marağı azaldır. Məqsədyönlü, mövzulara uyğun dəyişikliklər, rollu, işgüzar oyunların keçirilməsi (məsələn, informatika müəlliminin istiqamətləndirməsi ilə krosvordlardan ibarət broşurların, fəxri fərmanların, təbrik açıqcalarının, valideyn iclaslarına dəvətnamələrin və s. hazırlanması və yayılması ilə bağlı), esse yazma üzrə yarışların keçirilməsi və s. dayanıqlı motivasiyanı artırır, onun idarəedilməsini təmin edir.

İSTİNADLAR

- [1] Azərbaycan Respublikasında təhsilin inkişafı üzrə Dövlət Strategiyası, URL: <https://president.az/articles/9779>
- [2] Azərbaycan Respublikasının təhsil sahəsində İslahat Proqramı, 1999. URL: <http://www.e-qanun.az/framework/5363>.
- [3] <https://president.az/az/articles/view/30563>.
- [4] <https://quizizz.com>
- [5] Qurbanova Ə.M., İsmayılova B.İ., Tədrisyönlü oyunlardan, intellektual testlərdən istifadə etməklə informatika fənninin tədrisində olan bəzi problemlərin həlli yolları // Bakı:AMEA İnformasiya Texnologiyaları İnstitutu, “İnformasiya cəmiyyəti problemləri”
- [6] <http://institute-of-progressive-education-and-learning.org/elearning/elearning-educational-games/>
- [7] <http://www.azterm.az>
- [8] Michael Weigend. Обучение информатике в Германии - тенденции и проблемы (Перевод с нем. М.И.Юдовина) // Компьютерные инструменты в школе, №2, 2012, с. 1-9

Ümumtürk ədəbiyyatının təbliği və öyrənilməsində İKT-nin rolu

Aytən Qurbanova

AMEA Nizami Gəncəvi adına Ədəbiyyat İnstitutu, Azərbaycan Pedaqoji Universiteti
filoloq86@mail.ru

Xülasə — Müasir gəncliyin, ümumiyyətlə cəmiyyətin formalaşmasında rol oynayan mühüm məqamlardan biri də İKT-nin sürətli inkişafıdır. İKT-nin cəmiyyətin həyatına bu qədər daxil olduğu bir dövrdə, hətta mütaliənin də elektronlaşmasının özünəməxsus müsbət və mənfi tərəfləri vardır. Belə ki, xüsusən, mütaliə mövzularının nə dərəcədə uğurlu seçilməsi bu sahədə diqqət olunması vacib olan məsələlərdən biridir. Nəzərə alsaq ki, turançılıq və türkçülük ölkəmizdə və bütün dünyada geniş şəkildə yayılmış, eləcə də türk dövlətlərinin rəhbərləri tərəfindən son illərdə ən mühüm ideyalardan birinə çevirilmişdir. Bu səbəbdən də tələbə və şagirdlərdə bu istiqamətdə oxu vərdişlərini aşılamağa, o cümlədən ciddi elmi tədqiqatlar aparmağa ehtiyac vardır.

Açar sözlər — *ümumtürk ədəbiyyatı; muxtariyyətdə olan türk dövlətləri; informasiya texnologiyaları; türk dilləri; türkçülük; turançılıq*

I. GİRİŞ

“Mən fəxr edirəm ki, türk dillərinə mənsub olan Azərbaycan dili bu qədər bədii ifadələrlə zəngindir və biz həyatın bütün sahələrinə aid olan fikirlərimizi öz ana dilimizdə ifadə edə bilərik” [1]. Ümummilli lider Heydər Əliyevin dediyi bu fikir həm

bütövlükdə türk dillərinin, həm də türk dil ailəsinin bir qolu olan doğma Ana dilimizin şirinliyindən, zənginliyindən, poetikliyindən xəbər verir. Əsrlərlə təqiblərə məruz qalan türk dilləri bütün türkdilli dövlətləri birləşdirən əsas mənəvi dəyərdir. Məhz bu səbəbdən də, milli dilimiz daxil olmaqla, bütün digər türk dillərinin əsrlər boyunca qorunub saxlanması, qarşılıqlı inkişaf etdirilməsində, yayılmasında görkəmli türk liderlərinin, ziyalılarının rolu böyükdür.

II. ÜMUMTÜRK ƏDƏBİYYATININ TƏBLİĞİ VƏ ÖYRƏNİLMƏSİNDƏ İKT-NİN ROLU

“Parçala, hökmranlıq et!” siyasəti ilə türk soyundan olan millətlərin illərlə etnik tərkibinin dəyişdirilməsinə çalışan və təzyiq altında saxlayan siyasi rejim – Sovet hakimiyyəti parçalandıqdan sonra müstəqil Azərbaycan Respublikasında türkçülük, turançılıq ideyalarının inkişafı yolunda mühüm addımlar atıldı [2]. 1993-cü ildə Azərbaycanda hakimiyyətə gələn ulu öndər Heydər Əliyev yenidən ümumtürk mədəniyyətinin təkamülü və türk dövlətləri, topluları arasında dostluq, qardaşlıq münasibətlərinin inkişafı yolunda ilk addımı məhz həmin gün – 15 iyun 1993-cü ildəki çıxışı zamanı atdı.

O, türk xalqlarının birləşdirilməsi yolunda türkün düşmənlərinin anlamayacağı bir strategiya və taktika yeridərək, “Ey türklər, gəlin birləşək!” demədi, müştərək türk eposları “Kitabi-Dədə Qorqud”-un və “Manas”-ın müvafiq olaraq 1300 və 1000 illik yubileylərini ümumtürk hadisəsi kimi keçirdi. Dahi Azərbaycan şairi Məhəmməd Füzulinin 500 illik yubileyini türk bədii təfəkkürünün hadisəsi kimi qeyd edərək Türkiyə Böyük Millət Məclisinin tribunasından bütün dünyaya bəyan etdi [2]: “Füzuli bütün türkləri birləşdirən şəxsiyyətdir!”

TÜRKSÖY-un formalaşması, türk xalqlarının mədəni abidələrinin, şəxsiyyətlərinin, türkmən şairi Məhdimqulunun, qazax ədəbiyyatının klassiki Abayın, böyük özbək hökmdarı Əmir Teymurun, İbn-Sinanın, Nizaminin, Nəvainin, Yunis İmrənin, Füzulinin, Mahmud Qaşqarının, bir çox başqa dahilərimizin, “Kitabi-Dədə Qorqud”, “Alpamış”, qırğız eposu “Manas”-ın yubileyinin təşkil edilməsində Ümummilli liderin iştirakı, əslində, bu istiqamətdə çox önəmli addımlar idi. 1996-cı ilin noyabrında dahi şairimiz Məhəmməd Füzulinin 500 illiyi, 2000-ci ilin aprelində isə böyük türk eposu “Kitabi-Dədə Qorqud”-un 1300 illiyi ilə bağlı türk dünyasının təmsilçilərini, dövlət xadimlərinin, alim, şair və yazıçıların, mədəniyyət xadimlərinin Bakıya toplaşması ümumtürk ədəbiyyatının, mədəniyyətinin əlaqəli şəkildə inkişafına mühüm töhfə verdi. Ancaq nəzərə alsaq ki, bir neçə on il əvvəldən söhbət gedir, o zaman İKT-nin inkişaf səviyyəsi imkan vermirdi ki, bu tədbirlər, eyni zamanda,

ortaq ədəbiyyat, mədəniyyət geniş şəkildə türkdilli ölkələr arasında təbliğ oluna bilsin.

Fikrimizcə, hələ də yaxasını Rusiya və ÇXR kimi imperialist dövlətlərin pəncəsindən qurtara bilməyən və onların ərazisində muxtariyyətdə yaşayan 15 türksöylü xalqın ümumtürk mədəniyyətinin təşkili və təbliği məsələlərindən geri qalması problemi, əgər siyasi məsələlər olmasa idi, sadəcə, İKT-nin yüksək inkişafı ilə aradan qalxa bilərdi. Çünki İKT-nin sürətli inkişafı bütün dünyada zaman və məkan məhdudiyətlərini aradan qaldırır [3].

Azərbaycanın şair və yazıçıları hələ qədim dövr farsdilli ədəbiyyatımızdan başlamış, hətta sovet dövrü də daxil olmaqla, öz yaradıcılıqlarında türk, Turan sevgisini gizlətməmişlər [4]. Nizami Gəncəvinin poemalarında da türk sevgisini çox aydınlıqla sezə bilirik. Belə ki, müəllif farsın dilində türkü vəsf edən, farsın dilində türkün tarixini yaradan dahi sənətkar idi. Elə bu səbəbdən də klassiklərimizdən qalan türk ədəbiyyatını, türkcülük ideyasını müasir gəncliyimiz arasında təbliğ etməliyik.

Hazırda İKT-nin yüksək səviyyədə inkişafı və demək olar ki, yer kürəsinin çox geniş bir ərazisində İKT-nin əlçatanlığı imkan verir ki, ümumtürk ədəbiyyatını və mədəniyyətini ən ucqar ərazilərdə məskunlaşmış eyni kökdən olan qardaşlarımıza belə çatdırıla bilsin. Bu məsələdə aşağıdakı formatda tədbirlərin həyata keçirilməsinin artıq nə qədər məqsədəuyğun olması ciddi müzakirə mövzudur:

- müştərək ədəbi portalların təşkili və təhsil müəssisələrində reklam olunması;
- müştərək ədəbi dərnəklərin təşkili, bütün türk ölkələrinin bədii dəyərlərinin təbliğ olunmasına nəzarət etmək;
- bütün türk xalqlarının ədəbiyyatı ilə ədəbi əlaqələr fonunda simpozium və konfransların təşkili;
- müştərək kitab sərgilərinin təşkili;
- tələbələrədən təşkil olunmuş könüllü teatr truppalarının yaradılması;
- ədəbi-bədii gecələrin təşkil edilməsi;
- festivalların təşkil olunması;
- bədii əsərlərə çəkilən filmlərin kinostudiya şəklində izlədilməsi;
- türk xalqlarının yazıçı və şairlərlə görüşlərinin təşkil edilməsi.
- müştərək onlayn dərslərin tədrisi və s.

NƏTİCƏ

Son dövrlərdə İKT-nin yüksək səviyyədə inkişafı yuxarıda həyata keçirilməsi məqsəduyğun olan məsələlərin əksəriyyətinin məhz virtual platforma üzərindən reallaşdırılmasına imkan verir. Belə ki, müştərək ədəbi portalların hazırlanması şərait yaradır ki, məsələn,

Azərbaycanda hələ də tədqiqatında boşluqlar olan başqırd, çuvaş, tuva, xakas və s. türkdilli xalqların ədəbiyyatı haqqında kifayət qədər dolğun məlumatlar əldə edə bilək.

Türkdilli ölkələr arasında onlayn konfransların daha sıx təşkili həyata keçirilərsə, bu qardaş xalqlar arasındakı münasibətləri və dostluğu daha da genişləndirə bilər. Virtual şeir gecələrinin təşkilini həyata keçirməklə qədim ənənələrə söykənən ümumtürk ədəbiyyatını türk xalqlarının birlik və güc simvoluna çevirmək mümkündür.

Beləliklə, İKT-nin hazırda həyatımızın çox hissəsini zəbt etməsini biz əslində daha səmərəli şəkildə istifadə edərək dəyərlərimizin inkişafına və təbliğinə yönəldə bilərik.

İSTİNADLAR

- [1] <https://azerbaijan-news.az/az/posts/detail/heyder-eliyev-ve-dil-medeniyetimiz-1650490040N>.
- [2] N.Cəfərov. Türk dünyasının birliyi strategiyası: Heydər Əliyevdən İlham Əliyevə”, Xalq qəzeti.- 2009.- 18 iyul.- s. 3.
- [3] S.Abdullayev, S.Abasova. Təhsil prosesində İKT-nin tətbiqinə və keyfiyyətinə sistemli yanaşma prinsipləri haqqında. “Elektron dövlət quruculuğu problemləri” I Respublika elmi-praktiki konfransı, 4 dekabr, 2014 İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan.
- [4] Şəmsizadə “Turan birliyi: İdealdan həqiqətə”. 525-ci qəzet.- 2022.- 10 iyun.- s.13

İKT terminlərinin Azərbaycan dilinə tərcüməsi problemləri

Səyyar Abdullayev¹, Südabə Abasova²

^{1,2}İnformasiya Texnologiyaları İnstitutu, Bakı, Azərbaycan
¹sayyar.abdullayev@yandex.com, ²abasova.sudaba@gmail.com

Xülasə — Məqalədə İKT terminlərinin Azərbaycan dilinə tərcüməsi sahəsində vəziyyətin təhlili, dillərarası qarşılıqlı əlaqələrin inkişafı, əcnəbi dillərdən alınan söz və terminlərin tədqiqi, kalka üsulu ilə alınması problemləri, alınma sözlərin unifikasiya və standartlaşdırma formalarının kompleks şəkildə öyrənilməsi məsələlərinə baxılmışdır. Alınma sözlərin dilin sonrakı inkişaf mərhələlərində yaratdığı problemlər araşdırılmış və onların həlli istiqamətində tövsiyələr işlənmişdir.

Açar sözlər — *termin; milli termin; alınma sözlər; kalka üsulu; leksikoqrafiya, lüğət*

I. GİRİŞ

2023-cü ildə 100 illik yubileyini qeyd etdiyimiz Ulu öndər Heydər Əliyevin həyata keçirdiyi dil siyasəti xalqımızın milli-mənəvi dəyərlərinin qorunmasında, tariximizin düzgün istiqamətdə öyrənilməsində, ana dilimizin inkişafında böyük rol oynamışdır. Hal-hazırda da Azərbaycan Prezidenti İlham Əliyev Ulu öndər Heydər Əliyevin ana dilimizin sosial-siyasi, ideoloji problemlərinin həllində gördüyü işləri uğurla davam etdirir.

Elmlərin differensiyası nəticəsində yeni sahələrin yaranması prosesi baş vermişdir. Hal-hazırda elmlərin inteqrasiyası prosesi isə müasir elmi istiqamətlərin yaranmasında, formalaşmasında və

gələcəkdə bu istiqamətlərin müstəqil bir elm kimi inkişafında böyük rola malikdir. Cəmiyyətdə, ictimai həyatda, elmi-texniki tərəqqidə baş verən bütün elmi yeniliklər dildə öz əksini tapmaqla yanaşı, insanlar arasında ünsiyyət formalarının əsasını təşkil edir. Dil, həmçinin cəmiyyətin müxtəlif sosial qruplarının təsirinə məruz qalır və dilin lüğət tərkibi zənginləşir.

Bununla da dilin lüğət tərkibinin əsasını təşkil edən terminologiyanın formalaşması və daimi inkişaf prosesi baş verir. Hər bir dildə olduğu kimi Azərbaycan dilində də terminlərin yaranma prosesində iki əsas istiqamət nəzərdə tutulur [1,2]:

- dilin lüğət tərkibindən istifadə edilməsi
- elmi terminlərin dilə daxil olunması

II. TERMİNLƏRİN MİLLİLƏŞDİRİLMƏSİ AMİLLƏRİ

Terminlərin milliləşdirilməsi prosesində ən əsas amillərdən biri yeni terminlərin yaranması və onların dünya dilləri arasında yer almasıdır. Yeni terminlərin yaradılması prosesində daha çox beynəlxalq dillərdən istifadə edilir və bunun nəticəsində dilin lüğət tərkibi zənginləşir. Bu sahədə ingilis dilinin aparıcı mövqeyə malik olması danılmazdır. Dilin üstünlüyünü şərtləndirən

əsas cəhətlərdən biri də elmtutumlu informasiyaların dilə daxil olunması və inkişaf etdirilməsidir. Tədqiqatçılar yeni terminlərin yaradılmasında fərqli fikirlər yürüdürlər. Belə ki, bəziləri terminlərin öz ana dilində qarşılığını axtarıb tapmağa səy etmək, yəni dilin öz vasitələrindən istifadə etməklə yeni termin yaratmağa çalışmaq, digər qismi isə alınma terminlərə üstünlük verirlər. Hazırkı dövrdə elmi-texniki tərəqqinin, intensiv informasiya axınının, baş verən ictimai-siyasi və sosial dəyişikliklərin, beynəlxalq əlaqələrin inkişafı dillərin lüğət tərkibinin zənginləşməsinə səbəb olur. Ölkələrin dünyaya inteqrasiyası, beynəlxalq münasibətlərin sürətli inkişafı, xüsusən Qərbin aparıcı dövlətləri ilə bir çox sahələrdə qarşılıqlı əlaqələrin genişlənməsi dilə alınma söz və terminlərin daxilolma prosesini sürətləndirir. Bu prosesdə dilin öz daxili imkanları ilə yanaşı, başqa dillərdən alınma sözlər və terminlər hesabına da dildə zənginləşmə baş verir. Beləliklə, dilin lüğət tərkibində xüsusi bir qat təşkil edən alınma söz və termin sistemi yaranır.

III. TERMİNLƏRİN MİLLİLƏŞDİRİLMƏSİ SAHƏSİNDƏ VƏZİYYƏTİN TƏHLİLİ

Müxtəlif dillərdən bir-birinə keçən söz və terminlər qanunauyğun bir prosesdir. Belə ki, bu proses ilk növbədə, həmin dillərin daşıyıcılarının elmi, ədəbi, mədəni, iqtisadi, diplomatik və s. əlaqələri üçün böyük əhəmiyyət kəsb edir. Dillərarası qarşılıqlı əlaqələrin inkişaf etdiyi müasir dövrdə əcnəbi dillərdən alınan söz və terminlərin tədqiqi, müasir müstəvidə nəzərdən keçirilməsi və elmi təhlili, alınma

sözlərin dilin sonrakı inkişaf mərhələlərində yaranan problemlərinin aktuallığını şərtləndirən amillərdən biridir.

Hər hansı bir dildən söz və termin alınması tamamilə təbii və progressiv bir prosesdir. Bu prosesin nəticəsi olaraq söz və termin bir dildən digərinə keçir. Qohum və ya uzaq dillər özlərində olmayan sözləri digər dildən almaqla ya mövcud olan məfhumlarına uyğun söz və termin tapır, yaxud da əcnəbi dildən aldığı söz və termini dəyişmədən müvafiq şəkildə və ya kalka üsulu ilə saxlayır.

Qeyd etmək lazımdır ki, hər bir dilin lüğət tərkibi müxtəlif dillərdən alınma söz və terminlər vasitəsi ilə zənginləşir. Belə ki, dillərin hər birində sahələr üzrə terminologiyaların formalaşmasının əsası və terminlərin bir dildən başqa dillərə keçməsi prosesi elmlərin inkişafı və diferensiasiyası, təhsil prosesinin inkişafının sürətləndirilməsi, texniki nailiyyətlərin tətbiqi və istifadəsinin artması nəticəsində baş verir.

IV. TERMİNLƏRİN MİLLİLƏŞDİRİLMƏSİNİN MAHİYYƏTİ VƏ VƏZİFƏLƏRİ

Hazırda ingilis dilinin dünyada geniş yayılması, onun beynəlxalq dil statusu qazanması, bütün sahələrlə yanaşı İKT üzrə yeni anlayış, söz və terminlərin həmin dil əsasında formalaşmasını zəruri edir. Azərbaycan dilində söz və terminlərin uzun müddət vasitəçi dillərdən, hazırda ingilis dilindən birbaşa alınması, onların nizamasalma və unifikasiya prosesi həm leksik sistemdə baş vermiş dəyişmələrin xüsusiyyətləri, həm də alınma söz və

terminlər baxımından öyrənilməsi aktual məsələlərdəndir.

Dilin terminoloji sistemində xüsusi qat təşkil edən alınma söz və terminlərin unifikasiya (ümumi bir sistemə, qaydaya salınması) və standartlaşma formalarının kompleks şəkildə öyrənilməsi aşağıdakı vəzifələrin yerinə yetirilməsində nəzərdə tutulmuşdur:

- dilin leksik sisteminə dil əlaqələrinin təsirini və rolunu aydınlaşdırmaq;
- Azərbaycan dilində alınma söz və terminlərin assimilyasiya (uyğunlaşma) prosesini müəyyənləşdirmək;
- alınma söz və terminlərin dilə daxil olma yollarını və üsullarını aydınlaşdırmaq;
- alınma söz və terminlərin inkişaf qanunauyğunluqlarını və onun işləklilik qazanmasını dəqiqləşdirmək;
- alınma söz və terminlərin milli dillərdə paralelliyini yaradan səbəbləri müəyyənləşdirmək;
- beynəlmiləl söz və terminlərin Azərbaycan dilinə transliterasiya yollarını və ümumi xüsusiyyətlərini təhlil etmək;
- Azərbaycan dilində fonetik, qrammatik və semantik səviyyələrdə alınma söz və terminlərin mənimsənilməsini araşdırmaq;
- Azərbaycan dilində alınma söz və terminlərin qanunauyğunluqlarının mənimsənilməsini müəyyənləşdirmək;

- Azərbaycan dilində alınma söz və terminlərin sistemə salınması, unifikasiyası və standartlaşdırılması yollarını müəyyənləşdirmək [2].

TERMINLƏRİN MİLLİLƏŞDİRİLMƏSİ ÜZRƏ PROBLEMLƏR

Terminlərin yaradılması prosesində qarşıya çıxan problemlərdən biri də kalka üsulu ilə alınmadır. Belə ki, kalka üsulu ilə alınan sözlərin alınma söz və terminlərə aid edilib-edilməməsi məsələsi sual yaradır. Bir sıra dilçilərin bu problemlə bağlı müxtəlif yanaşmaları mövcuddur. R.A.Budaqov qeyd edir ki, “Kalka üsulu alınmanın xüsusi bir növüdür. Bu sözlər müvafiq əcnəbi sözlərin quruluşu kimi formalaşan, lakin onların maddi əsasını götürməyən sözlərdir”. M.Ş.Qasimov isə həmin yanaşmanı inkar edir və öz fikrini digər formada izah edir: “Əslində bu fikirlərin heç biri dəqiq deyildir. Kalka üsulunu nə alınma söz və terminlər nə də əsl mənada tərcümə hesab etmək olmaz” [3,4].

Söz və terminlərin yaradılması prosesində kalka üsulu:

- dildə olan söz və terminlər əcnəbi dillərə məxsus söz və terminlərin quruluşu kimi formalaşır;
- yeni söz və terminlər hərfi yolla tərcümə edilir;
- dildə digər dilin təsiri nəticəsində yeni söz və termin modeli əmələ gəlir.

Kalka üsulu ilə alınma söz və terminlərin yaradılması prosesi milli termin arasındakı keçidi xarakterizə edir. Belə ki, alınma söz və terminlərin məzmununun dəyişməməsi,

milli formanın isə olduğu kimi saxlanması nəzərdə tutulur. Azərbaycan dilində də, bütün dillərdə olduğu kimi kalka üsulu ilə digər dillərdən alınma söz və terminlər yaranır və zənginləşir. Alınma söz və terminlər həm də milli dilin daxili imkanları hesabına yaranır və milli dil vahidi ilə əvəz olunur. Beləliklə, milli söz və terminlərin yaradılması prosesi dilin daxili inkişaf qanunlarının əsasını təşkil edir.

İllər öncə dilimizin rus dili ilə əlaqələrinin daha geniş olması səbəbindən ingilis mənşəli söz və terminlər dilimizə rus dili vasitəsi ilə keçirdi. Hal-hazırda isə dövlətlər arasında yaranan əlaqələrin sayəsində söz və terminlərin dilimizə birbaşa keçməsinə imkan yaranır. Söz və terminlərin digər dillərdən alınması, həm ictimai hadisə, həm də dilin formalaşması prosesidir. Bu prosesdə bir dildən başqa dilə daha çox elmi, texniki tərəqqi və yeni texnologiyalar ilə bağlı terminlər daxil olur. Qeyd etmək lazımdır ki, müstəqillik illərindən bu günə qədər dilimizin həm ingilis, həm Avropa mənşəli söz və terminlərin alınması hesabına zənginləşməsinə imkan yaranmışdır [2].

V. İKT ÜZRƏ TERMİNLƏR LÜĞƏTİ

Azərbaycan dili bazasının inkişafına, yeni termin və anlayışların yaranmasına və respublikamızda elmin inkişafına təkan vermiş amillərdən biri də uzun illər ərzində elmin bir çox sahələrinin sistemli şəkildə araşdırılmasıdır.

Ana dilimizdə elmi əsərlərin və digər sənədlərin tərtibi və tərcüməsi zamanı yaranan çətinlikləri aradan qaldırmaq üçün elmin müxtəlif sahələrində çalışan

tədqiqatçılar qarşısında mühüm vəzifələr qoyulmuşdur. İnformasiya texnologiyaları və İKT üzrə ən müasir terminlər bazasını (termin, məfhum anlayışlarını əhatə edən) özündə əks etdirən lüğətlərin ana dilində işlənilib hazırlanması zərurəti məhz bu səbəbdən yaranmışdır [5].

Lüğətlərin yaradılması dilçiliyin xüsusi bir sahəsi kimi leksikoqrafiyanın vəzifələrindən biridir.

Leksikoqrafiya (yunanca “*leksikon*” – “söz” və “*qrafo*” – “yazıram”, yəni “söz yazıram”) dilçiliyin lüğətlərin yazılması və tədqiqi ilə məşğul olan xüsusi bir sahəsidir. *Lüğət* adətən, əlifba sırası ilə verilmiş söz, söz birləşmələri və terminlər toplusundan ibarətdir. Lüğətdə söz, söz birləşmələri və terminlərin izahı, şərh və ya başqa dilə (dillərə) tərcüməsi verilir. Son zamanlar informatika, İKT üzrə müxtəlif lüğətlər yaradılmışdır [6-11].

Müasir dövrdə İKT sferasının dinamik inkişafı informasiya sistemləri, blokçeyn texnologiyaları, bulud texnologiyaları, rəqəmsal transformasiya, yaşıl texnologiyalar, Big data, süni intellekt, Əşyaların İnterneti və s. kimi yeni texnologiyaların yaranmasına səbəb olmuşdur. Bunun da nəticəsində İKT sferasında kifayət qədər yeni söz, termin və anlayışların yaradılması, onların milliləşdirilməsi zəruriyyəti yaranır. İnformasiya Texnologiyaları İnstitutunda ən yeni termin və anlayışlar üzrə izahlı lüğətlərin işlənməsi məsələlərinə baxılmış, tədqiqat işləri aparılmışdır, 5 kitab (ekspres-informasiya, informasiya texnologiyaları seriyası) olmaqla

informasiya texnologiyaları üzrə müasir terminlər lüğəti yaradılmışdır [12-17].

Bütün qeyd olunanlar nəzərə alınaraq, təqdim olunan lüğətlərdə informasiya texnologiyalarının müxtəlif istiqamətləri – Big data, blokçeyn texnologiyaları, alqoritmləşdirmə, bulud texnologiyaları, GPS texnologiyalar, 3D texnologiyalar, informasiya sistemləri, informasiya təhlükəsizliyi, proqram mühəndisliyi, vikipediya, internet, sosial şəbəkələr və s. üzrə azərbaycanca-ingiliscə-rusca ən müasir terminlərin hər üç dildə izahı verilmişdir.

Son illərdə bir sıra alınma terminlərin dilimizə daxil olması, yeni terminlərin qarşılığının işlənilməsi cəmiyyətimizin, xüsusilə də elmimizin beynəlxalq aləmə birbaşa inteqrasiyası nəticəsində baş vermişdir. Beləliklə də, lüğətdə terminlərin, termin-söz birləşmələrinin lüğətçilik işində müəyyənləşdirilmiş meyar və prinsiplər əsasında verilməsi əsas məsələlərdən biri hesab olunur. Belə ki, lüğətdə aşağıdakı qanunauyğunluqların gözlənilməsi vacibdir:

- Azərbaycan dilində terminlərin dəqiq, konkret izahı verilməli;
- Termin beynəlmiləl mahiyyətlidirsə, olduğu kimi saxlanılmalı və izahı verilməlidir.

Azərbaycan dilində informasiya texnologiyaları üzrə ən müasir terminlərlə bağlı terminoloji sistemin formalaşması bu lüğətlərdə nəzərə alınmışdır. Burada Avropa dillərindən keçən bəzi terminlərin dilimizdə mövcud olan terminlərlə ifadə oluna bilməməsi səbəbindən onlar mənbədə olduğu kimi verilmişdir. Lüğətdə terminlərin böyük bir qisminin ana

dilimizdə işlənən sözlərlə verilməsi müsbət hal kimi qiymətləndirilir.

NƏTİCƏ

Demək olar ki, ən çox işlənən terminlər müəyyən müddətdən sonra adi, ümumişlək terminlərə, yəni sözə çevrilir. Ona görə də ən müasir terminlərin yaradılması və milliləşdirilməsi zəruriyyəti yarandığından, bu istiqamətdə tədqiqatların aparılması və izahlı lüğətlərin işlənilməsi vacib məsələlərdən biridir.

Lüğətlərin tərtibi zamanı yeni anlayışların lüğətə daxil edilməsi üçün alınma sözlərdən çox sözseçmə prinsiplərinə riayət olunmalıdır. Dilin lüğət tərkibində işlənən, onların qarşılığını işlətmək yolu ilə yaradılan terminlərə üstünlük verilməlidir. Lüğət tərtib olunarkən bu sahədə mövcud olan əsas prinsiplərə riayət olunmalıdır. Beynəlxalq və alınma terminlərin Azərbaycan dilinin qayda-qanunlarına uyğunlaşdırılması, onların izahının verilməsi də lüğətlərdə nəzərə alınmalıdır. Terminlərin izahı elə qurulmalıdır ki, təsnifat nəticəsində onlar arasında yaranmış əlaqə itməsin. Terminlərin əlifba sırası ilə tərtib olunması lüğətçilik ənənəsinə uyğun olaraq yaradılmalıdır.

İSTİNADLAR

- [1] N. Bayramlı, Yeni Azərbaycan, 2013, 2 mart, S.4., http://www.anl.az/down/meqale/yeni_az/2013/mart/297038.htm
- [2] G. Abdullayev S.H., Kazımov T.H., Abasova S.E. Qloballaşma şəraitində dilin inkişaf tendensiyaları və təhlükəsizliyi, https://ict.az/uploads/konfrans/info_sec_2018/R

- S30_DEVELOPMENT_TRENDS_AND_SAFETY_OF_LANGUAGE.pdf
- [3] Alməmmədova S.M. Müasir Azərbaycan dilində alınma terminlərin unifikasiyası və standartlaşdırma formaları, Bakı - 2018,
- [4] Budaqov R.A. Dilçiliyə dair oçerklər. Bakı, Azərnəşr, 1956, səh 88
- [5] Qasimov M.Ş. Azərbaycan dilinin terminologiyasının əsasları. Bakı, Elm, 1973, səh.124.
- [6] Məmmədova M.M. İnformasiya-telekommunikasiya texnologiyaları sahəsində işlədilən terminlərin ekstralinqvistik və linqvistik xüsusiyyətləri,
- [7] What is lexicography, <https://www.engati.com/glossary/lexi-cography>
- [8] Ə.Abbasov, S.Kərimov, N.Rəhimova, N.Vəliyev, T.İbrahim-zadə, N.Cəfərov İnformasiya-telekommunikasiya texnologiyaları üzrə ingiliscə-rusca-azərbaycanca izahlı lüğət, Bakı, ADNA, 2012.
- [9] Abbasov Ə.M., Əliyev F.Ə., Əliyev Ə.Ə., Əhmədov F.B. İnformatika telekommunikasiya radioelektronika üzrə ingiliscə-rusca-azərbaycanca terminoloji lüğət, (2-ci nəşr), Bakı, “Elm”, 2004.
- [10] Sadıqov İ.C. İnformatika terminlərinin izahlı lüğəti, Bakı, 2017
- [11] Məmmədli R. H. ingiliscə-azərbaycanca izahlı kompüter lüğəti, Bakı, 2014 və s.
- [12] Abdullayev S.H., Abasova S.E. İnformasiya texnologiyaları üzrə müasir terminlər lüğəti, Bakı, 2015
- [13] Abdullayev S.H., Abasova S.E. İnformasiya texnologiyaları üzrə müasir terminlər lüğəti, Bakı, 2016
- [14] Abdullayev S.H., Abasova S.E. İnformasiya texnologiyaları üzrə müasir terminlər lüğəti, Bakı, 2017
- [15] Abdullayev S.H., Abasova S.E. İnformasiya texnologiyaları üzrə müasir terminlər lüğəti, Bakı, 2018
- [16] Abdullayev S.H., Abasova S.E. İnformasiya texnologiyaları üzrə müasir terminlər lüğəti, Bakı, 2020
- [17] Abdullayev S.H., Abasova S.E. İnformasiya texnologiyaları üzrə müasir terminlər lüğəti, Bakı, 2021

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 unutmalarına 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, danışiq, 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 axtarıqlarının (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 sözün 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, anlaşılma 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 olunsun 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 mənfə 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

Jurnalistikada mətnlərin intellektual linqvistik təminatı problemləri

Sünbül Zalova

İnformasiya Texnologiyaları İnstitutu,
Bakı Dövlət Universiteti, Bakı, Azərbaycan
sunbulzalova@gmail.com

Xülasə — Tezisdə linqvistik xüsusiyyətlərin jurnalistikada rolu, jurnalist mətnlərində informasiya qəbulunu asanlaşdırmaq üçün mətnlərin keyfiyyətinin yaxşılaşdırılması, süni intellekt texnologiyalarının jurnalistikaya tətbiqinin yaratdığı yeni istiqamətlər haqqında məlumat verilmiş. linqvistik təminat vasitələri, vətəndaş jurnalistikası ilə peşəkar jurnalistikanın dil-üslub fərqləri ilə bağlı tədqiqatlar araşdırılmış, müasir jurnalistikada yazı işlərinin keyfiyyətinin artırılması üçün informasiya kommunikasiya texnologiyalarının imkanlarından istifadənin zəruriliyi əsaslandırılmışdır.

Açar sözlər — linqvistik təminat vasitələri; vətəndaş jurnalistikası; süni intellekt

I. GİRİŞ

Jurnalistika sahəsi cəmiyyətin problemlərinin araşdırılması, hadisələrin düzgün təqdim edilməsi kimi məsələlərin həllində vacib rol oynayır. Burada yazı dili və onun xüsusiyyəti çox önəmlidir. Jurnalistikanın hər bir janrında dilin incəliklərinə və imkanlarına yaxşı bələd olmaq və onlardan məharətlə istifadə etmək tələb olunur. Jurnalistlər hadisələri, faktları, problemləri söz və cümlələrin düzgün

seçilməsi ilə müxtəlif bucaqlardan təqdim edə bilirlər [1]. Burada hər hansı bir məsələnin neqativ və ya pozitiv təqdim olunması bir çox hallarda jurnalistin seçdiyi ifadələrdən asılıdır.

II. İNTELLEKTUAL LİNGVİSTİK TƏMİNAT VASİTƏLƏRİ

Mətn axtarış sistemlərində tammətli sənədlərlə iş prosesində təbii dilin emalı vasitələrindən geniş istifadə olunur. Bu vasitələr mətn axtarış sistemlərinin mürəkkəb komponentləridir. Təbii dilin emalı vasitələri mətn sənədlərindəki, sorğulardakı terminləri, sözləri, söz birləşmələrini və s. təhlil edir. Təbii dilin emalı vasitələri mətnlərdəki məzmununu aydınlaşdırır, terminlər arasındakı münasibətləri aşkar edir və səmərəli axtarışı təmin edir. Mətnlərin keyfiyyətinin yaxşılaşdırılması məqsədi ilə mətni axtarış sistemlərində linqvistik təminat vasitələri kompleksindən istifadə olunur. Bunlara müxtəlif lüğətlər, tezauruslar, sistemin predmet sahəsinin ontoloji xüsusiyyətlərini əks etdirən materiallar daxildir [3].

Sistem lüğətləri dedikdə təbii dilin emalı ilə məşğul olan mətn axtarış sistemlərində ümumdil leksikası, və predmet sahəsi leksikası üzrə lüğətlərdən istifadə edilir [3].

Bu lüğətlər mətni morfoloji təhlil və axtarış edən zaman müxtəlif qrammatik formalardakı sözləri tanımağa imkan verir.

Tezaurus – dilin əsas anlayışları lüğətidir [3]. Tezaurları xüsusi lüğətlər kimi də adlandırmaq olar. Əsasən mətn tipli sənədləri təqdim etmək və sorğunu genişləndirmək üçün tezaurus adlanan xüsusi lüğətlərdən istifadə olunur.

Təbii dildə təqdim olunan mətn sənədlərinin və axtarış sorğularının adekvat dərki üçün predmet sahəsinin əsas anlayışları və onların arasındakı müxtəlif semantik əlaqələri təyin edən vasitələrə *ontoloji materiallar* deyilir. Bura predmet sahəsinin tezaurusu və birinci dərəcəli məntiq dillərini (*Birinci dərəcəli məntiq dedikdə predikat məntiqi, kəmiyyət məntiqi və birinci dərəcəli predikat hesabı nəzərdə tutulur ki, bu da riyaziyyat, fəlsəfə, dilçilik və kompüter elmlərində istifadə olunan formal sistemlərin məcmusudur*) və s. aid etmək olar [4].

İnformasiya kommunikasiya texnologiyalarının (İKT) sürətli inkişafı jurnalist yazılarında mətlərin oxunaqlığının yoxlanmasına, mətlərdə ölkədə qəbul edilməmiş terminlərin aşkarlanmasına, vulqar sözlərin avtomatik redaktəsinə şərait yarada bilər. İKT-nin köməyi ilə, əsasən də, süni intellekt vasitəsilə bu gün mətlərin keyfiyyəti avtomatik yoxlanılır və yaxşılaşdırılır.

III. SÜNİ İNTELLEKTİN JURNALİSTİKA SAHƏSİNDƏ YARATDIĞI İMKANLAR

Struktur linqvistik xüsusiyyətlər yazılı ünsiyyətin mühüm aspektidir. Araşdırmalar göstərir ki, linqvistik mürəkkəblilik insanların

informasiyanı emal etməsində mühüm rol oynayır. Tədqiqatçılar avtomatlaşdırılmış məzmun təhlili metodlarından istifadə edərək müxtəlif vətəndaş və peşəkar jurnalist təşkilatları arasında dil mürəkkəbliyindəki fərqləri araşdırıb. Təhlil nəticəsində müxtəlif kütləvi informasiya vasitələrində siyasi informasiyanın təqdim edilməsi nümunələrinin fərqli olduğu göstərilmişdir. Araşdırmada jurnalistikada mətn mürəkkəbliyinin rolu və əhəmiyyəti, tabloid qəzetlərinin və vətəndaş jurnalistikası məqalələrinin mətnin mürəkkəbliyi və ölçüləri baxımından necə fərqləndiyi barədə dəyərli fikirlər təqdim olunmuşdur [2]. Bu araşdırmada üç müxtəlif media növündən olan siyasi jurnalist mətnlərinin struktur xüsusiyyətləri, xüsusən də peşəkar qəzet məqalələri, tabloid qəzet məqalələri və siyasi veb-bloq məqalələri müqayisə edilmişdir. Tədqiqatın məqsədi mətn mürəkkəbliyinin bu medialar arasında fərqləndirilməsini, xüsusən də vətəndaş jurnalistikası mətnlərinin digər iki peşəkar mətn növündən fərqləndirilməsini araşdırmaq idi. Dilçilikdən tutmuş sosial elmlərə qədər bir sıra sahələrdən mətn mürəkkəbliyi ölçüləri seçilmiş və ~930.000 məqaləyə tətbiq edilmişdir. Müəyyən edilmişdir ki, bu ölçülər faktiki olaraq mətn mürəkkəbliyinin üç müxtəlif ölçüsünü ölçür: sintaktik, semantik və məzmun mürəkkəbliyi.

Maşın tərcüməsi, avtomatik referatlaşma, mətnin mövzuya uyğun təsnifatlaşdırılması süni intellektin jurnalistika sahəsində yaratdığı yeni imkanlardır. Süni intellektin jurnalistikaya verdiyi töhfələri aşağıdakı kimi təsnif etmək olar:

Mətnlərin oxunaqlığının və mürəkkəbliyin səviyyəsinin qiymətləndirilməsi dedikdə, mətnlərin mürəkkəbliyinin qiymətləndirilməsi üçün riyazi modellərin qurulmasının zəruriliyinin əsaslandırılması nəzərdə tutulur. Həmçinin mövcud qiymətləndirmə düsturları Azərbaycan dilinin özəllikləri nəzərə alınmaqla uyğunlaşdırılmalıdır. Bu prosesə Azərbaycan dilindəki mətnlərin mürəkkəbliyin səviyyəsinin qiymətləndirilməsi üçün avtomatlaşdırılmış texnologiyaların işlənilməsi ilə yanaşı bəzi dünya ədəbiyyatı nümunələrinin orijinalının Azərbaycan dilinə tərcümələrinin oxunaqlığının müqayisəsi və hazırlanmış metodika əsasında məktəb dərslərinin mürəkkəbliyinin şagirdlərin yaş səviyyəsinə uyğun olub-olmamasının qiymətləndirilməsi kimi məsələlər də daxildir [5].

Mətnlərdə istifadə olunan terminlərin müqayisəli analizi dedikdə, mətnlərdə ölkədə qəbul edilməmiş terminlərin aşkarlanması (məsələn: “Gəncə Destinasiya Mərkəzi” tipli adlara qarşı terminaloji filtrasiyadan istifadə) və Milli Terminoloji İnformasiya Sistemi ilə müqayisəsi nəzərdə tutulur. Müasir dövrdə bir çox elm sahələrində də (leksikoqrafiya, terminşünaslıq, informasiya axtarışı və s.) emal olunmalı informasiyanın həcminin artım dinamikası termin və açar sözlərin avtomatik seçilməsi məsələsini aktual edir. Hazırda təbii dilin emalı sahəsində qoyulan qaydaları təqdim edən mətnlərdən terminlərin avtomatik çıxarılması üçün bir çox fərqli yanaşma və sistem işlənmişdir [6].

Mətnlərdə vulqar sözlərin aşkarlanması İnternet mühiti şifahi nitqdən fərqlidir və burada müəyyən texniki əsaslar çərçivəsində vulqarizmlərə qarşı mübarizə aparmaq mümkündür. Milli dillərdə yaradılan vulqarizm bazalarının inteqrasiyası ilə qlobal vulqarizm bazasının yaradılması və UNESCO-nun koordinatorluğunun bu baza üzrə AntiVulgar sisteminin tərtibi İnternet mühitində vulqarizmlərə qarşı uğurlu mübarizə platforması kimi nəzərdən keçirilməlidir.

Mətnlərdə müxtəlif növ sirr daşıyan məlumatların müəyyənəşdirilməsi. Müasir dövrimizdə texnologiyanın sürətli inkişafı ilə əlaqədar olaraq informasiyanın toplanması, saxlanması, emalı və ötürülməsi prosesləri müasir İKT vasitələrindən istifadə etməklə həyata keçirilir. Bu isə konfidensial informasiyanın asanlıqla başqalarının əlinə keçməsi təhlükəsini artırır. İnformasiya təhlükəsizliyinin təmin edilməsinin mühüm şərtlərindən biri informasiyanın konfidensiallığı məsələsidir. İnformasiya resursları ilə işləyən insanlar ayrı-ayrı vətəndaşlara və ya təşkilatlara aid olan informasiyanın konfidensiallığını qorumaq üçün hüquqi məsuliyyətlərini dərk etməli və informasiya təhlükəsizliyini təmin etməlidirlər.

Müxtəlif növ sirr daşıyan məlumatların olub-olmamasının aşkarlanması müvafiq qanunvericilik aktları ilə tənzimlənə bilər.

Mətnlərdə plagiatlığın aşkarlanması. Plagiatlıq dedikdə başqa müəllifə məxsus fikir və ideyaların öz adından təqdim olunması nəzərdə tutulur [7]. Bu hadisəni oğurluq və piratçılıqlada sıx əlaqələndirmək

olar. Plagiatlıq həmçinin kompüter virusları, spamlar, fişinq və s. kimi kibercinayətkarlığın bir növü hesab olunmalıdır. Plagiatlıq İnternetdən, kitabdan, jurnaldan və s. müxtəlif mənbələrdən birbaşa mətnin köçürülməsi və yaxud az sayda dəyişiklik etməklə baş verir.

Plagiatlıq avtomatik aşkarlama kompüter proqramları vasitəsilə həyata keçirilir və hazırda plagiatlıq aşkarlayan çoxlu sayda proqram təminatları mövcuddur: PlagAware, PlagScan, Check for Plagiarism, iThenticate, PlagiarismDetection, Academic Plagiarism, The Plagiarism Checker, Urkund, Docoloc və s. [7].

Maşın tərcüməsi prosesi sözlərin, cümlələrin və ya böyük həcmli mətnin tərcüməsi olan kompüter sistemi vasitəsilə təbii dilin başqa təbii dilə avtomatik çevrilməsidir [8]. Tərcümə proqramlarının özləri də iki kateqoriyaya bölünür: bütün mətni tərcümə edən və hər bir sözün ayrı-ayrılıqda mənasına uyğun sözlər siyahısını təqdim edən proqramlar.

Avtomatik referatlaşdırma mətndən zəruri məlumatların çıxarılması və bu məlumatın əsas məzmununu saxlamaqla yığcam formada istifadəçiyə təqdim edilməsi prosesidir. Mətnlərin avtomatik referatlaşdırılması üçün klasterləşmə və optimallaşdırma metodlarına əsaslanan iki mərhələli cümlə seçimi modeli təklif olunub [9].

Mətnin təsnifatlandırılması prosesi göstərilən təsnifatlandırma sistemi çərçivəsində məzmunu uyğun olaraq mətnin hansı kateqoriyaya aid olmasını avtomatik müəyyən edir [10]. Lakin cümlələr oxşar

olduqda bu tip məsələlərin həlli çətinləşir, qeyri-səlis klasterləşmə tələb olunur.

Yazı üslubunu analiz etməklə, müəllifin müəyyənləşməsi. Hər bir müəlif öz dəsti-xətti, üslubu və araşdırması ilə fərqlənir. Müasir İKT vasitələrinin köməyi ilə mətnlərdə müəllifin dəsti-xəttinin avtomatik olaraq aşkarlanması mümkündür.

NƏTİCƏ

Nəticə olaraq demək olar ki, İKT-nin köməyi ilə mətnləri daha anlaşıqlı formada təqdim etmək, tanınmayan terminlərdən, vulqar sözlərdən təmizləmək mümkündür. Həmçinin plagiatlığın yoxlanılması və referatlaşma, mətnlərin təsnifatlandırılması və xarici dillərdən avtomatik tərcüməsi mətnlərin keyfiyyətinin yaxşılaşdırılmasında texnologiyanın necə vacib rol oynadığının əyani sübutudur. Müasir İKT vasitələrinin tətbiqi ilə jurnalistikada da yazı işlərinin keyfiyyəti artırıla bilər.

İSTİNADLAR

- [1] John Richardson “Language and journalism: An expanding research agenda”, *Critical Discourse Studies* 2015, vol 12, pp 370-373
- [2] Petro Tolochko, Hajo G. Boomgaarden, *Journalism Studies*, 2018, vol 19, pp 1786-1803
- [3] İsa Musayev, Mətləb Əlizadə, *İnformatikanın əsaları (dərs vəsaiti)*, 2014
- [4] https://en.wikipedia.org/wiki/First-order_logic
- [5] İ.Sadıgov, "Mathematical and Information Models for Evaluating Readability of Texts in Azerbaijani Language", 2018, *El-Cezeri*, vol. 5, no. 3, pp. 888-903, Sep.
- [6] Qurbanova Ə.M, Mətnlərdən terminlərin avtomatik çıxarılması metodlarının araşdırılması və müqayisəli analizi, *İnformasiya cəmiyyəti problemləri*, 2021, №1, səh. 55–69

- [7] R.S. Hasanova, F.Ş.Əsgərov, Plagiatlıq və onunla mübarizə vasitələri: antiplagiat sistemləri, İnformasiya cəmiyyəti problemləri, 2019, №2, səh. 94–101
- [8] Ramiz M. Aliguliyev, Makrufa Sh. Hajirahimova “An optimization model for automatic text summarization”, Problems of information technology, 2015, №2, pp 84–90
- [9] Rasim M. Alguliyev, Ramiz M. Aliguliyev, Nijat R. Isazade, Asad Abdi, Norisma Idris, COSUM: Text summarization based on clustering and optimization, “Expert Systems”, 2019, vol 36
- [10] Алекперова И.Я., Анализ тональности текста как эффективный подход к анализу поведения // Информационные технологии, 2022, том 28, № 11, стр. 590–596.

Comparison of LR, NB and SVM Techniques on Sentiment Analysis of Azerbaijani Texts

Mehdi Rasul

Landau School, Baku, Azerbaijan

mehdi@rasul.az

Abstract—The prediction of sentiment in text has been a challenging problem across domains. This paper analyzes movie reviews in English and Azerbaijani using machine learning techniques like Logistic Regression, Naive Bayes, and Support Vector Machine (SVM). The results indicate the importance of language corpus and compare performance between languages.

Keywords — *sentiment analysis, machine learning, Logistic Regression, Naive Bayes, Support Vector Machine*

I. INTRODUCTION

Sentiment analysis, also known as opinion mining, identifies and extracts subjective information such as opinions, emotions, and attitudes from text data using natural language processing (NLP) and computational linguistics. It has applications across business, politics, and social media monitoring. Companies use it to evaluate customer feedback, monitor reputation, forecast behavior, etc.

Recent research shows machine learning algorithms, particularly supervised learning and deep learning, produce good results in automating textual sentiment analysis. Studies show that combining computer science and math can improve accuracy in sentiment analysis automation.

Well-defined English language corpus aids in building powerful and accurate models for English text analysis. The abundant NLP libraries and tools in English, like word correction, grammar checking, tokenization, etc. make accurate sentiment models easier to create. However, research on under-resourced languages like Azerbaijani is limited despite its linguistic distinctness and approximately 30 million speakers worldwide. The lack of labeled data and sophisticated corpus in Azerbaijani makes sentiment analysis uniquely challenging.

This paper compares several machine learning techniques on an English and Azerbaijani movie review dataset including Logistic Regression (LR), Naive Bayes (NB) and Support Vector Machine (SVM). The results address the importance of language corpus and differences in techniques for Azerbaijani versus English sentiment analysis.

II. REVIEW OF LITERATURE

There have been several studies that have applied machine learning techniques for sentiment analysis across different domains.

Neethu and Rajasree performed sentiment analysis on Twitter data using classifiers like SVM, NB, Maximum Entropy and Ensemble methods [1]. They achieved up to 90% accuracy. Chandra and Jana (2020) also used machine learning and deep learning models on Twitter data, finding that LSTM gave 97% accuracy while LR, NB and SVM gave around 82% accuracy [2].

For political sentiment analysis, Heredia et al. used deep convolutional neural networks to analyze tweets and predict the 2016 US elections with 84% accuracy [3]. Additionally, in Indonesia, tweets of @jokowi, account of current President of the Republic of Indonesia, have been scraped and fitted into different machine learning models to find the sentiment, specifically positive and negative labels. The results indicated that Sequential Minimal Optimization (SMO) gave 82.7% of accuracy, precision and recall [4].

For the Azerbaijani language, Rustamov and Hasanli classified 12,000 Azerbaijani tweets into positive and negative sentiment using LR, NB and SVM with accuracy up to 94% using bag-of-words features [5]. Suleymanov et al. (2019) did text classification on Azerbaijani news articles, finding SVM with TF-IDF features gave 93% accuracy, outperforming NB at 56.53% accuracy [6].

Overall, prior research indicates machine learning approaches can effectively classify sentiment for both Azerbaijani and English texts. Algorithms like SVM, NB and LR have been most widely used.

III. EXPERIMENT

In the project, Python has been utilized in data cleaning and modeling stages. Specifically, Python built-in library *sk-learn* was useful by simplifying the use of various ML algorithms. The trained dataset is about movie reviews in English, and it was translated into Azerbaijani. There are 2000 reviews, half of which was in Azerbaijani, and the other half was in English language. The process for both Azerbaijani and English languages has been outlined below.

A. Azerbaijani Language Processing Approach

Data Cleaning

In the data cleaning phase, several essential steps were undertaken to prepare the Azerbaijani dataset for sentiment analysis. Words with fewer than 3 characters were removed, and non-alphabetical characters, such as digits and punctuations, were eliminated because they do not contribute to the sentiment of the text. To maintain dataset relevance, a comprehensive list of Azerbaijani stop-words was manually created and applied to the text. To ensure consistency, all text entries were converted to lowercase, and sentence tokenization was employed.

Modeling

To enhance result quality, experiments were conducted using various machine learning models, specifically LR, NB and SVM. Different feature extraction techniques, including TF-IDF and *Countvectorizer*, were explored.

Given the relatively small dataset, the k-fold cross-validation method was employed to ensure robust model training and evaluation.

In the initial stage of the pipeline, text data was tokenized and countvectorized using a (1,2) n-gram with a word analyzer. The data was divided into an 80-20 train-test split. Among the models tested, LR, with BOW, achieved an impressive 84.25% accuracy and 86% precision (Table II). However, the models exhibited susceptibility to overfitting, with training accuracy values ranging from 95% to 99%. To mitigate this, regularization techniques were applied.

Table II. Accuracy Scores with Default Parameters

Algorithm	TF-IDF	BOW
LR	86%	84.25%
NB	74%	79%
SVM	85.25%	77.5%

Later, regularization was applied to minimize the possibility of overfitting. Overfitting is the phenomenon in machine learning when the model adapts to the training data too much, and, consequently, the model performs exceptionally well on training data, but poorly on testing data. Therefore, it is preferred to build a model that achieves similar performance with both training and testing data and over the predefined thresholds. After L1 regularization was applied, most models still showed either no or very little improvement in accuracy score. The only model which saw significant increase in accuracy score from L1 regularization was

NB with BOW. The results show that NB algorithm with BOW produced 86% accuracy, which is a 7% increase from the result with default parameters.

In a separate experiment, the k-fold cross-validation approach was adopted, dividing the dataset into equally sized folds and iteratively using one-fold for testing while training on the remaining k-1 folds. This method allowed assessment of the models' performance to be more comprehensive. The accuracy scores for the three selected algorithms are as follows:

Table III. Accuracy Scores using K-fold Approach

Algorithm	TF-IDF	BOW
LR	81%	84.6%
NB	78%	79%
SVM	81%	82%

B. Language Processing on English Dataset

An English version of the same dataset was utilized in the modeling process for predicting the general sentiment of reviews. Python, enriched with built-in libraries for English language modeling, particularly through nltk, facilitated tasks like word tokenization, stemming, and feature extraction without requiring custom implementations. Text files were read in Python and structured using the Pandas library.

In subsequent stages, common stop-words like "and," "or," and "is" were removed from the texts using a list obtained from the nltk corpus library. Additionally, each text was tokenized using the word tokenizer function. The Porter Stemming

method was applied to maintain the root of various word versions in the dataset, ensuring consistency across words like "computers," "computation," and "computed," which were all transformed to "comput" in the texts.

Further preprocessing involved retaining words with a minimum length of 4 characters while removing those with 3 or fewer characters.

In the modeling stage, various machine learning algorithms were employed. Initially, models were trained with default parameter values. Subsequently, regularization was applied to address issues such as overfitting. Data was divided into train and test datasets, with the k-fold approach also implemented to assess model performance and minimize overfitting risks.

Based on default parameter values, LR with TF-IDF feature extraction yielded an accuracy of 87%, while SVM and Multinomial NB with TF-IDF achieved 85% accuracy (Table IV).

Table IV. Accuracy Scores for Sentiment Analysis for Dataset in English

Algorithm	TF-IDF	BOW
LR	87%	78%
NB	85%	81%
SVM	85%	82%

IV. CONCLUSION AND FUTURE WORK

In this research, various ML algorithms have been used to predict the sentiment of the movie reviews in both English and Azerbaijani. The study has indicated the achievements attained in building models with various techniques. TF-IDF and BOW

(or Count Vectorizer) have been implemented for feature extraction methods from the texts and tested in different models, namely LR, NB, and SVM.

For Azerbaijani version of the dataset, combined with TF-IDF feature extraction, LR and SVM algorithms produced the best result (81% accuracy, measured as the mean of 10 folds' results). When BOW feature extraction was used, LR produced accuracy of 84%, while SVM reached 82% accuracy score.

The same dataset in English has also been modeled to compare the results with different preprocessing techniques that were not applied in Azerbaijani version, such as stemming, stopwords list, etc. Overfitting was much less of an issue for models based on English dataset. The highest score was attained with BOW feature extraction combined with LR method. Additionally, SVM and NB algorithms performed well with TF-IDF feature extraction method and achieved 85% accuracy score.

The research indicates that the results are similar for both language models. However, it should be noted that the dataset was relatively small for sentiment analysis. Therefore, the amount of data in the dataset should be increased so that a better comparison can be made.

Python has enriched libraries for building language models in English. These include as list of stopwords, word tokenization, stemming and lemmatization techniques. Particularly, English language corpus is well developed, which helps achieve higher results. However, Azerbaijani language lacks the language

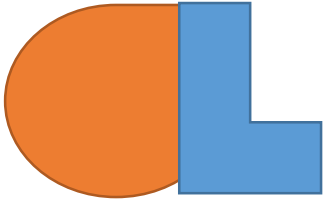
corpus which makes it harder to build generalized models. Also, there are not any libraries like NLTK and BERT for Azerbaijani language, which makes creation of accurate sentiment analysis models a lot harder. Therefore, a lot of resources should be dedicated to research and creation of language corpus for Azerbaijani Language, which would help getting higher accuracy scores.

ACKNOWLEDGEMENT

I am extremely grateful to Mr. Mammadzada for his guidance and technical support throughout this research experiment. In addition, I am thankful to Dr. Rustamov for providing the essential dataset for the research project.

REFERENCES

- [1] M.S. Neethu & R. Rajasree (2013). “Sentiment analysis in twitter using machine learning techniques”. Proceedings of the 2013 Fourth International Conference on Computing, Communications and Networking Technologies, 2013, pp.1-5. <https://doi.org/10.1109/ICCCNT.2013.6726818>
- [2] Y. Chandra & A. Jana (2020) “Sentiment analysis using machine learning and deep learning”. Proceedings of the 2020 7th International Conference on Computing for Sustainable Global Development, New Delhi, India, pp.1-4. <https://doi.org/10.23919/INDIACom49435.2020.9083703>
- [3] B. Heredia, J.D. Prusa, & T.M. Khoshgoftaar (2018). “Location-based twitter sentiment analysis for predicting the U.S. 2016 Presidential Election”. Proceedings of the 31st International Florida Artificial Intelligence Research Society Conference, vol. 2009, pp.265–270.
- [4] F.A. Wenando, R. Hayami, Bakaruddin & A.Y. Novermahakim (2020). “Tweet sentiment analysis for 2019 Indonesia presidential election results using various classification algorithms”. Proceedings of the 2020 1st International Conference on Information Technology, Advanced Mechanical and Electrical Engineering, Yogyakarta, Indonesia, pp.279-282. <https://doi.org/10.1109/ICITAMEE50454.2020.9398513>
- [5] H. Hasanli & S. Rustamov (2019). “Sentiment analysis of Azerbaijani tweets using logistic regression, Naïve Bayes and SVM”. Proceedings of the 2019 IEEE 13th International Conference on Application of Information and Communication Technologies, Baku, Azerbaijan, pp.1-7. <https://doi.org/10.1109/AICT47866.2019.8981793>
- [6] U. Suleymanov, B.K. Kalejahi, E. Amrahov, & R. Badirkhanli (2019). “Text classification for Azerbaijani language using machine learning and embedding”. arXiv:1912.13362 <https://doi.org/10.48550/arXiv.1912.133>



ÇAP POLİQRAFİYA

Mətbəə və poliqrafiya xidmətləri

Texniki redaktorlar: **Salman Əhmədov**

Zülfüqar Hüseynli

Kompüter dizaynı: **Fərqanə Manafova**

Kompüter yığımı: **Sürayyə Zərifova**

Korrektor: **Şəbnəm Sultanlı**

Çapa imzalanmışdır 01.06.2023 format 60x84 1/8

Sifariş №24. Sayı 200 ədəd
