

# 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 işlənəməkdədir.

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 qovuşduğu koqnitologiya adlanan yeni elm sahəsi yaranmışdır. Koqnitologiya dilçilik elminin də tədqiqat 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 məqsəda uyğun 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-işgüzar 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-danışıq 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 göstəriləndiyi kimi, süni intellekt – elə bir texnologiya, daha dəqiq desək, müasir elmi istiqamətdir ki, insan kimi düşünən 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 fonoloqik 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](http://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](http://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](http://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](http://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...>](http://studmed.ru>smolin-dv-vvedenie-v-is...>)
- [15] Хант Э. Искусственный интеллект /Artificial intelligence / Под ред. В. Л. Стефанюка. — М.: Мир, 1978. — 562 с. — [log-in.ru/books/khant-e-iskusstvennyiy-intellekt-khant-1-informatsiya](http://log-in.ru/books/khant-e-iskusstvennyiy-intellekt-khant-1-informatsiya)
- [16] Turing Alan (October 1950). Computing Machinery and Intelligence. *Mind* T. LIX( 236): 433-460. ISSN 0026-4423, doi:10. 1093mind/LIX.236.433.
- [17] Azərbaycan dilinin informasiya bazası <http://azerbaycandili.az>