



Gülçöhrə Qədir qızı QƏDİROVA

Azərbaycan Universiteti, Tərcümə və filologiya kafedrasının baş müəllimi, Bakı, Azərbaycan

E-mail: gulchohra.gadirova@au.edu.az

ORCID ID: 0009-0009-5866-5471

**KORPUS ƏSASLI DİALEKTOLOGİYADA SÜNİ İNTELLEKT
TEXNOLOGİYALARININ TƏTBİQİ: AZƏRBAYCAN VƏ İNGİLİS DİLLƏRİ ÜZRƏ
MÜQAYİSƏLİ YANAŞMA**

Xülasə

Tədqiqatın məqsədi Azərbaycan və ingilis dillərinin dialektoloji xüsusiyyətlərinin korpus əsaslı və süni intellekt dəstəklə metodlarla təhlil imkanlarını araşdırmaqdır. Tədqiqatda dialektal materialların toplanması, annotasiyası və avtomatik emalı üçün korpus dilçiliyi, maşın öyrənməsi və təbii dil emalı metodlarından istifadə olunmuşdur. Verilənlərin strukturlaşdırılması zamanı sosial-demografik göstəricilər və linqvistik variasiyalar nəzərə alınmışdır. Tədqiqatın nəticələri göstərir ki, korpus əsaslı yanaşmalar dialektal variativliyin daha dəqiq modelləşdirilməsinə və regional dil xüsusiyyətlərinin sistemli şəkildə təhlilinə imkan yaradır. Süni intellekt texnologiyaları isə dialektlərin avtomatik identifikasiyası və təsnifatında yüksək effektivlik nümayiş etdirir.

Açar sözlər: korpus dilçiliyi, dialektologiya, süni intellekt, maşın öyrənməsi, NLP.

UOT: 81'322:81'282:004.8

DOI: <https://doi.org/10.54414/LKID9957>

Giriş

Müasir dilçilikdə korpus əsaslı metodologiyaların tətbiqi dialektoloji materialların sistemləşdirilməsi, təhlili və rəqəmsal mühitdə qorunması baxımından mühüm elmi əhəmiyyət kəsb edir [3]. Xüsusilə Azərbaycan və İngilis dillərinin fonetik, leksik və semantik xüsusiyyətlərini əks etdirən dialektal vahidlərin rəqəmsal bazalarda emalı dilin tarixi inkişaf dinamikasının və regional xüsusiyyətlərinin daha dəqiq öyrənilməsinə imkan yaradır. Bu yanaşma dialektoloji materialların kompüter texnologiyaları və müasir metadata standartları vasitəsilə sistemləşdirilərək etibarlı elmi bazaya çevrilməsini təmin edir.

Rəqəmsal dialektologiya sahəsində aparılan tədqiqatlar göstərir ki, dialektal məlumatların çoxqatlı linqvistik annotasiya modelləri ilə zənginləşdirilməsi onların avtomatik emalı və linqvostatistik təhlili üçün geniş imkanlar yaradır [18]. Bu proses dialektlərin areallara görə təsnifatının dəqiqləşdirilməsi ilə yanaşı, regional ləhcələrin süni intellekt əsaslı nitq emalı sistemlərində tanınmasını və təhlilini də asanlaşdırır [18].

İngilis dili dialektlərinin korpus əsaslı tədqiqi tarixi və sosiolinqvistik dəyişikliklərin empirik şəkildə araşdırılmasına imkan verdiyi halda, Azərbaycan dilçiliyində bu istiqamətin inkişafı milli korpusun funksional imkanlarının genişləndirilməsi ilə sıx bağlıdır [13].

Son illərdə maşın öyrənməsi, böyük verilənlərin emalı və təbii dil emalı texnologiyalarının inkişafı dialektoloji tədqiqatların metodoloji bazasını əhəmiyyətli dərəcədə genişləndirmişdir [20]. Dialektal vahidlərin avtomatik tanınması, rəqəmsal xəritələndirilməsi və elektron leksik bazalara integrasiyası dilin rəqəmsal infrastrukturunun formalaşdırılmasında mühüm rol oynayır. Eyni zamanda, süni intellekt texnologiyalarının tətbiqi morfofonoloji formaların identifikasiyasını və regional semantik xüsusiyyətlərin kontekstual təhlilini sürətləndirir [15].

Müasir tədqiqatlarda neyron şəbəkələr və dərin öyrənmə metodologiyalarının dialektoloji korpuslara tətbiqi nitqin tanınması sistemlərinin dəqiqliyinin artırılmasına və avtomatlaşdırılmış tərcümə sistemlərinin

təkmilləşdirilməsinə şərait yaradır. Bu texnologiyalar dialektal nitqin emalında regional xüsusiyyətlərin nəzərə alınmasına imkan verərək dilin texnoloji adaptasiya imkanlarını genişləndirir. Bununla yanaşı, rəqəmsal platformalar dialektal materialların yalnız arxivləşdirilməsini deyil, həm də onların müasir elmi standartlara uyğun şəkildə sistemli təsnifatını təmin edir.

Dilçilik və süni intellektin integrasiyası nəticəsində formalaşan yeni metodoloji yanaşmalar dilin struktur və funksional xüsusiyyətlərinin daha dəqiq modelləşdirilməsinə imkan yaradır [2]. Bu istiqamətdə aparılan tədqiqatlar süni intellektin fonetik, leksik və semantik modelləri avtomatlaşdırılmış şəkildə emal etməklə dilçilik elminin nəzəri və praktik imkanlarını genişləndirdiyini göstərir. Eyni zamanda, rəqəmsal transformasiya şəraitində dilçi alimlərdən süni intellekt texnologiyaları və koqnitiv dilçilik prinsipləri üzrə yeni peşəkar kompetensiyaların formalaşdırılması tələb olunur [11].

Beləliklə, korpus əsaslı dialektologiya və süni intellekt texnologiyalarının integrasiyası regional dil müxtəlifliyinin qorunması, sistemləşdirilməsi və qlobal rəqəmsal mühitə adaptasiyası üçün mühüm elmi platforma formalaşdırır. Bu yanaşma yalnız akademik tədqiqatların inkişafına deyil, həm də dil irsinin gələcək nəsillər üçün qorunmasına və rəqəmsal mühitdə davamlılığının təmin edilməsinə xidmət edir.

Ədəbiyyat icmalı

Müasir dilçilikdə süni intellekt texnologiyalarının tətbiqi ilə bağlı aparılan tədqiqatlar həm nəzəri, həm də praktik baxımdan geniş tədqiqat istiqaməti formalaşdırmışdır. Xüsusilə təbii dilin emalı və maşın tərcüməsi sahəsində əldə olunan nailiyyətlər dilçilik tədqiqatlarının avtomatlaşdırılmış analiz sistemlərinə integrasiyasını sürətləndirmişdir. Bu texnologiyalar mətnlərin avtomatik identifikasiyası, semantik analiz və dil vahidlərinin emalı kimi mürəkkəb filoloji proseslərin daha yüksək dəqiqliklə həyata keçirilməsinə imkan yaradır.

Bir sıra tədqiqatçılar süni intellektin tədris prosesindəki rolunu xüsusi vurğulamışlar. Abbasova [1] və Hamidovna [12] adaptiv təlim sistemləri və virtual köməkçilərin tələbələrin dil bacarıqlarının inkişafına müsbət təsir göstərdiyini qeyd etmişlər. Turkanın [14] araşdırmalarında isə süni intellekt əsaslı tətbiqlərin söz ehtiyatı, qrammatika və tələffüz bacarıqlarının inkişafında interaktiv və fərdiləşdirilmiş öyrənmə mühiti yaratdığı göstərilmişdir. Bu yanaşmalar tədris prosesində motivasiyanın artırılması və fərdi öyrənmə ehtiyaclarının nəzərə alınması baxımından əhəmiyyətli hesab olunur.

Dialektoloji korpusların formalaşdırılması və rəqəmsal dil resurslarının yaradılması istiqamətində aparılan tədqiqatlar da mühüm elmi nəticələr ortaya qoymuşdur. Tədqiqatçılar süni intellektə əsaslanan proqram təminatlarının dialektal variativliyin dəqiq kodlaşdırılması və korpus bazalarının formalaşdırılmasında metodik üstünlüklərini qeyd etmişlər [14]. Eyni zamanda, rəqəmsal korpus texnologiyaları dialektoloji materialların sistemli şəkildə saxlanması və gələcək tədqiqatlar üçün etibarlı verilənlər bazasının yaradılması baxımından mühüm əhəmiyyət daşıyır.

Bununla yanaşı, bir sıra müəlliflər süni intellekt texnologiyalarının tətbiqi zamanı meydana çıxan etik və metodoloji problemlərə də diqqət yetirmişlər. Şimşək [17] məlumatların məxfiliyi, alqoritmik qərəzlilik və mədəni xüsusiyyətlərin qorunması məsələlərini əsas problemlərdən biri kimi qiymətləndirmişdir. Digər tədqiqatlarda isə akademik dürüstlüyün qorunması, plagiat hallarının qarşısının alınması və tədqiqatçıların rəqəmsal kompetensiyalarının inkişaf etdirilməsi zəruri istiqamətlər kimi göstərilmişdir. Bu baxımdan, süni intellekt alqoritmlərinin verdiyi nəticələrin filoloji və metodoloji baxımdan yoxlanılması mühüm elmi tələb hesab olunur.

Mövcud tədqiqatlarda süni intellektin dil tədrisi, korpus texnologiyaları və təbii dil emalı sahələrinə tətbiqi geniş şəkildə araşdırılsa da, Azərbaycan və İngilis dillərinin dialektoloji materiallarının müqayisəli korpus əsaslı və süni intellekt dəstəklı təhlili kifayət qədər öyrənilməmişdir. Bu baxımdan, dialektal vahidlərin rəqəmsal korpuslar və süni



intellekt texnologiyaları vasitəsilə sistemli təhlili müasir dilçilik üçün aktual tədqiqat istiqamətlərindən biri hesab olunur.

Metodologiya

Bu tədqiqatda Azərbaycan və İngilis dillərinin dialektoloji xüsusiyyətlərini əhatə edən korpusların yaradılması və təhlili üçün korpus dilçiliyi, maşın öyrənməsi və təbii dil emalı (NLP) yanaşmaları integrativ şəkildə tətbiq olunmuşdur. Tədqiqatın əsas məqsədi dialektal materialların rəqəmsal mühitdə sistemləşdirilməsi, avtomatik emalı və süni intellekt əsaslı analizi üçün etibarlı metodoloji çərçivənin formalaşdırılmasıdır.

Korpusların formalaşdırılması mərhələsində dialektal materiallar müxtəlif mənbələrdən toplanmış, onların fonetik, leksik və qrammatik xüsusiyyətləri nəzərə alınaraq strukturlaşdırılmışdır. Toplanmış verilənlərin emalında maşın öyrənmə modellərindən istifadə edilmiş, dialektal variantlar arasındakı leksik və qrammatik fərqlərin avtomatik aşkarlanması təmin olunmuşdur. Bu modellər regional dialektlərin təsnifatını apararaq dilin lokal və global dəyişmə dinamikasını sistemli şəkildə modelləşdirməyə imkan vermişdir.

Metodoloji çərçivəyə həmçinin avtomatlaşdırılmış geri əlaqə (feedback) mexanizmləri daxil edilmişdir. Bu mexanizmlər korpus məlumatlarının keyfiyyətinə nəzarət, annotasiyanın dəqiqliyinin artırılması və təsnifat səhvlərinin azaldılması üçün tətbiq olunmuşdur. Eyni zamanda, etik standartlara riayət olunması və akademik dürüstlüyün qorunması üçün nəzarət mexanizmləri inteqrasiya edilmişdir.

Korpusların strukturlaşdırılması zamanı metadata yanaşmasından istifadə olunmuş, respondentlərin sosial-demoqrafik göstəriciləri (yaş, cins, coğrafi məkan və s.) sistemli şəkildə kodlaşdırılmışdır [7]. Bu yanaşma dialektal materialların kontekstual təhlilində daha yüksək dəqiqlik təmin etmişdir.

Bundan əlavə, sosiolingvistik müəhibə metodları generativ süni intellekt alqoritmləri ilə birləşdirilmişdir. Bu inteqrasiya dialektal məlumatların daha geniş və operativ şəkildə toplanmasına imkan yaratmış, açıq sual

modelləri vasitəsilə dil müxtəlifliklərinin daha dərinədən üzə çıxarılmasını təmin etmişdir.

Müzakirə

Dialektal tədqiqatlarda korpus əsaslı yanaşmaların tətbiqi dilin qeyri-rəsmi və kontekstdən asılı formalarının, o cümlədən neologizmlərin və kod-deyişmə (code-switching) hallarının sistemli şəkildə sənədləşdirilməsinə imkan yaradır [4]. Bu yanaşma ənənəvi dialektologiya ilə hesablama dilçiliyi arasında metodoloji körpü rolunu oynayaraq lingvistik variasiyaların avtomatik identifikasiyası və təsnifatını mümkün edir [19].

Avtomatik nitq tanıma texnologiyalarının tətbiqi böyük həcmli dialektal məlumatların toplanmasına şərait yaratsa da, bu sistemlərdə yaranan səhvlər tədqiqat nəticələrinin etibarlılığına təsir göstərə bilər. Buna görə də avtomatik nəticələrin insan ekspertlər tərəfindən yoxlanılması və audio materiallara zaman kodları vasitəsilə birbaşa çıxışın təmin edilməsi vacib metodoloji tələb kimi qalır [5]. Eyni zamanda, transkripsiya məlumatlarının normallaşdırılması və annotasiya səviyyəsində standartlaşdırılması fonoloji variasiyadan asılı olmayan daha sistemli analizlər aparmağa imkan verir.

Dialektlər sabit strukturlar deyil, müxtəlif sosial və situativ kontekstlərdə aktivləşən dinamik lingvistik seçimlər sistemi kimi qəbul edilməlidir. Bu səbəbdən tezlik əsaslı analizlər dilin istifadəyə əsaslanan modellərində gizli qanunauyğunluqların üzə çıxarılmasında mühüm rol oynayır. Az resurslu dialekt mühitlərində korpusların balanslaşdırılması və çəkiləndirmə mexanizmlərinin tətbiqi model performansını artırmaqla yanaşı, potensial qərəzliliyin aşkarlanmasına da imkan verir.

Müasir dərin öyrənmə modelləri və böyük dil modelləri dialektoloji tədqiqatları ənənəvi statistik yanaşmalardan daha dərin səviyyəyə daşıyır. Bu sistemlər yalnız səthi leksik strukturları deyil, həm də asılılıq qrammatikası və diskurs səviyyəsindəki münasibətləri təhlil etməyə imkan yaradır. Bununla belə, bu texnologiyaların tətbiqində etik məsələlər xüsusi diqqət tələb edir. Dialektal xüsusiyyətlərin sosial-iqtisadi status

və ya demoqrafik göstəricilər üçün dolayı proksi kimi istifadəsi ciddi qərəzlilik riskləri yarada bilər [8].

Bu kontekstdə dil texnologiyalarının inkişafı yalnız texniki effektivlik deyil, həm də sosial ədalət və inklüzivlik prinsiplərinə əsaslanmalıdır. Sosiolingvistik müxtəlifliyin korpuslarda bərabər şəkildə təmsil olunması böyük dil modellərinin daha balanslı və etibarlı nəticələr verməsinə şərait yaradır. Eyni zamanda, birgə annotasiya prosesləri və sahəyə uyğunlaşdırılmış təlim strategiyaları lingvistik qərəzliliyin azaldılmasında effektiv vasitə kimi çıxış edir.

Azərbaycan və digər çoxdilli və post-sovet lingvistik mühitlərdə süni intellekt texnologiyalarının tətbiqi dil iyerarxiyalarının və sosiolingvistik güc münasibətlərinin necə yenidən formalaşdığını anlamaq baxımından xüsusi əhəmiyyət daşıyır. Bu baxımdan yerli icmaların tədqiqat prosesinə aktiv şəkildə cəlb edilməsi epistemik ədalətin təmin olunması və mədəni uyğunluğun artırılması üçün vacib şərtidir.

Tədqiqatın nəticələri göstərir ki, avtomatlaşdırılmış korpus əsaslı yanaşmalar dialektoloji variativliyin daha dəqiq, sistemli və miqyaslı şəkildə modelləşdirilməsinə imkan yaradır. Xüsusilə səs yazılarının rəqəmsal korpuslara integrasiyası dialektal məlumatların yanlış interpretasiya riskini azaldır və həm diatopik, həm də diastratik dəyişkənliyin kompleks təhlilini mümkün edir.

Maşın öyrənməsi və süni intellekt alqoritmlərinin tətbiqi dilin diaxron inkişafını izləməyə, leksik və qrammatik dəyişikliklərin sosial faktorlarla əlaqəsini müəyyən etməyə şərait yaradır [16]. Bundan əlavə, sentiment analizi və mövzu modelləşdirmə kimi metodlar regional dil davranışlarının və dil ideologiyalarının daha dərinlən təhlilinə imkan verir.

Tədqiqat həmçinin göstərir ki, avtomatlaşdırılmış sistemlər dialektlərin fonoloji və leksik xüsusiyyətlərinin sənədləşdirilməsində mühüm rol oynayır və ənənəvi metodlarla müəyyən olunması çətin olan gizli dialektal nümunələrin aşkar edilməsini təmin edir.

Bununla belə, süni intellekt modellərinin məhdudiyyətləri, xüsusilə verilənlərdəki

qərəzlilik və dialekt müxtəlifliyinin tam əhatə olunmaması kimi problemlər dialektoloji tədqiqatlarda diqqətlə nəzərə alınmalıdır [9]. Bu səbəbdən avtomatik transkripsiya və analiz proseslərində insan ekspert nəzarəti zəruri olaraq qalır [6].

Nəticələr göstərir ki, insan mütəxəssisləri və süni intellekt sistemləri arasında qurulan integrativ əməkdaşlıq dialektal materialların daha etibarlı və dəqiq təhlilinə imkan verir. Bu yanaşma həm də metodoloji etibarlılığı artıraraq, etik risklərin və analitik qərəzliliyin azaldılmasına töhfə verir [10].

Tədqiqat göstərir ki, korpus əsaslı yanaşmaların sosiolingvistik prinsiplərlə integrasiyası dialektal variasiyaların daha dəqiq və sistemli şəkildə modelləşdirilməsinə imkan verir. Eyni zamanda, bu integrasiya süni intellekt əsaslı dil modellərində mövcud olan epistemik və təmsilçilik bərabərsizliklərinin azaldılmasına töhfə verir.

Süni intellekt və korpus dilçiliyinin birgə tətbiqi dil modellərinin mədəni və regional müxtəlifliyi daha real şəkildə əks etdirməsinə şərait yaradır. Bu isə həm texnoloji performansın, həm də sosiolingvistik etibarlılığın yüksəldilməsinə xidmət edir.

Gələcək tədqiqatlar az resurslu dialektlərin daha yaxşı təmsil olunması, modellərdə qərəzliliyin azaldılması və sosiolingvistik faktorların daha dərinlən integrasiya olunması istiqamətində inkişaf etdirilməlidir. Bu yanaşma yalnız texnoloji sistemlərin təkmilləşdirilməsi deyil, həm də dil müxtəlifliyinin qorunması və rəqəmsal mühitdə davamlılığının təmin edilməsi baxımından strateji əhəmiyyət daşıyır.

Nəticə etibarilə, sosiolingvistik və texnoloji yanaşmaların integrasiyası dil resurslarının rəqəmsallaşdırılmasında daha inklüziv və dayanıqlı sistemlərin formalaşdırılmasına imkan yaradır. Bu isə dil texnologiyalarının həm akademik, həm də sosial baxımdan daha məsuliyyətli inkişafına xidmət edir.

ƏDƏBİYYAT SİYAHISI

1. Abbasova T. İngilis dilinin tədrisində süni intellektdən istifadə. Elmi Tədqiqat Beynəlxalq Elmi Jurnal, 2025;5(9):74–79. Doi: 10.36719/2789-6919/49/74-79



2. Abdurahmanova K. Müasir dilçiliyin inkişaf istiqamətləri və süni intellektin dil tədqiqatlarına təsiri. *Filologiya məsələləri*, 2025;(4):278–285. Doi: 10.62837/2025.4.278
3. Abdurahmanova S.B. Principles of linguistic annotation in language corpora. *International Multidisciplinary Journal for Research*, 2026;13(3):515–519. Doi: 10.5281/zenodo.19125219
4. Alekseev A., Turatali T. KyrgyzNLP: Challenges, progress, and future. Analysis of images, social networks and texts: 12th International Conference, AIST 2024, Bishkek, Kyrgyzstan, October 17–19, 2024, p. 3–39, Doi: 10.1007/978-3-031-88036-0_1
5. Coats S. Dialect corpora from YouTube. In Busse B., Dumrukic N., Kleiber I. (Eds.), *Language and Linguistics in a Complex World*. Berlin, Boston: De Gruyter, 2023: 79–102. Doi: 10.1515/9783111017433-005
6. Curry N., Baker P., Brookes G. Generative AI for corpus studies. *Applied Corpus Linguistics*, 2024;4(1):1–8. Doi: 10.1016/j.acorp.2023.100082
7. Çelikkol M., Körber L., Zhao W. Exploring Diachronic and Diatopic Changes in Dialect Continua: Tasks, Datasets and Challenges. In *Proceedings of the 5th Workshop on Computational Approaches to Historical Language Change*, pages 12–22, Bangkok, Thailand. Association for Computational Linguistics. 2024. Doi: 10.18653/v1/2024.lchange-1.2
8. Demszky D., Sharma D., Clark J., Prabhakaran V., Eisenstein J. Learning to recognize dialect features. *Proceedings of NAACL-HLT*, 2021:2315–2338. Doi: 10.18653/v1/2021.naacl-main.184
9. Dwiputri A. AI and sociolinguistics. *LIER*, 2025;1(2):36–45. Doi: 10.71435/688505
10. Farooq M.Z., Hafiz U., Hussain M.A. Speech recognition and phonetic variation: understanding the impact of accents and dialects on AI-based speech systems. *Qualitative Research Journal for Social Studies*, 2025;2(3):296–311. Doi: 10.63878/qrjs295
11. Ganiyeva L., Iskandarova M. Digital transformation in linguistics. *International conference on modern development of pedagogy and linguistics*, 2026;3(2):71–76. Doi: 10.5281/zenodo.18746602
12. Hamidovna H.M., Worldly Knowledge Publishing Centre. Sun'iy intellekt asosida ingliz tilini o'qitishning zamonaviy yondashuvlar. In *Ilm Fan Xabarnomasi*, 2026;11(02):1103–1107. Doi: 10.5281/zenodo.18267003
13. Həbibova K.K. Korpus dilçiliyində təməl yanaşmalar və Azərbaycan dilçiliyində bu sahə ilə bağlı perspektivlər. *Terminologiya məsələləri*, 2025;(1):84–93. Doi: 10.5281/zenodo.18370856
14. Quluzadə T.Ə. İngilis dilinin tədrisində süni intellektə əsaslanan yeniliklər. *İnnovasiya və davamlı inkişaf: yeni ideyalar və həllər I Beynəlxalq Elmi Konfransı*. Naxçıvan, 2025. Doi: 10.5281/zenodo.17688498
15. Mahmudov M.Ə. Dilçiliyin süni intellektlə bağlı yeni sahələri. Müasir incəsənət məkanında süni intellekt: problemlər və perspektivlər, 2025:257–259. Doi: 10.25045/asucaai.2025.45
16. Nourin N., Muniroopesh R., Shoaib A.S. M., Basharat A. Machine learning and the study of language change: a review of methodologies and application. *International Journal of Management Information Systems and Data Science*, 2024;1(2):48–57. Doi: 10.62304/ijmids.v1i2.144
17. Şimşek Ö. Yapay Zeka ve dil öğrenme süreçlerinde kullanımına genel bakış. *Temel Eğitim Araştırmaları Dergisi*, 2025; 5(1):118–130. Doi: 10.55008/te-ad.1567106
18. Vosiljonov A.B. Theoretical and methodological foundations for creating a linguistic base of the Fergana Valley dialects and integrating them into the national corpus of the Uzbek language. 2026. Doi: 10.5281/zenodo.18766020

19. Williams J.N., FitzMorris M., Aka O., Laszlo S. DrawL: understanding the effects of non-mainstream dialects in prompted image generation. arXiv preprint arXiv:2405.05382, 2024. DOI: 10.48550/arXiv.2405.05382.
20. Vintoniak O., Hnatyuk M., Minailo R.V., Turyшева O., Kotvytska V.

Dialectology in modern linguistic research: theoretical approaches and methods. AD ALTA: Journal of Interdisciplinary Research, 2024;14(1):231–236. Doi: 10.33543/1401393944

Gulchohra Gadir GADIROVA

Senior Lecturer, Department of Translation and Philology, Azerbaijan University, Baku, Azerbaijan

CORPUS-BASED DIALECTOLOGY AND THE APPLICATION OF ARTIFICIAL INTELLIGENCE TECHNOLOGIES: A COMPARATIVE STUDY OF AZERBAIJANI AND ENGLISH LANGUAGES

Summary

The aim of this study is to investigate the potential of corpus-based and artificial intelligence-supported methods for analyzing dialectal features of Azerbaijani and English languages. The study employs corpus linguistics, machine learning, and natural language processing techniques for the collection, annotation, and automatic processing of dialectal data, considering socio-demographic and linguistic variation factors. The results demonstrate that corpus-based approaches enable more accurate modeling of dialectal variation, while artificial intelligence techniques provide effective tools for automatic dialect identification and classification.

Keywords: corpus linguistics, dialectology, artificial intelligence, machine learning, natural language processing.

Гюльчохра Гадир ГАДИРОВА

Старший преподаватель кафедры Перевода и Филологии Азербайджанского университета, Баку, Азербайджан

КОРПУСНО-ОРИЕНТИРОВАННАЯ ДИАЛЕКТОЛОГИЯ И ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: СРАВНИТЕЛЬНЫЙ АНАЛИЗ АЗЕРБАЙДЖАНСКОГО И АНГЛИЙСКОГО ЯЗЫКОВ

Резюме

Цель исследования заключается в анализе диалектологических особенностей азербайджанского и английского языков с использованием корпусных методов и технологий искусственного интеллекта. В исследовании применяются методы корпусной лингвистики, машинного обучения и обработки естественного языка для сбора, аннотирования и автоматической обработки диалектных данных. Метод исследования включает использование корпусных данных, алгоритмов машинного обучения и лингвистической разметки с учетом социодемографических параметров. Результаты исследования показывают, что корпусный подход обеспечивает более точное моделирование диалектной вариативности, а технологии искусственного интеллекта эффективно применяются для автоматической идентификации и классификации диалектов.

Ключевые слова: корпусная лингвистика, диалектология, искусственный интеллект, машинное обучение, обработка естественного языка.

Daxil olub: 02.06.2026