EKSTREMISTIK MATNLARDA KATEGORIZATSIYA VA MUALLIF IDENTIFIKATSIYASI: O‘ZBEK TILIDAGI ONLAYN DISKURSGA SUD LINGVISTIK YONDASHUV

Abror Kurbonmuratov
Termiz davlat universiteti o‘qituvchisi

Annotatsiya

Ushbu tadqiqot sud lingvistikasi doirasida ekstremistik matnlarning kategorizatsiyasi va muallif identifikatsiyasini o‘rganadi. 2020–2024 yillar oralig‘ida Telegram kanallaridan to‘plangan 214 ta o‘zbek tilidagi matndan iborat korpus asosida uchta funksional matn kategoriyasi aniqlandi: Diniy esktremistik matn, siyosiy ekstremistik matn hamda ijtimoiy ekstremistik matn kabi kategoriyalarga ajratdik. Har bir kategoriya o‘ziga xos leksik, morfologik, sintaktik va pragmatik xususiyatlarga ega. O‘zbek tilining agglyutinativ tuzilishiga moslashtirilgan morfema darajasidagi stilometrik tahlil tasdiqlangan test to‘plamida 78,4% aniqlikka erishdi. Natijalar O‘zbekiston va kengroq Markaziy Osiyo uchun tergov amaliyoti, sud ishlari va media savodxonligi ta‘limi nuqtai nazaridan muhim ahamiyatga ega.

Kalit so‘zlar: *sud lingvistikasi, ekstremistik diskurs, muallif identifikatsiyasi, matn kategorizatsiyasi, o‘zbek tili, radikallashuv, Markaziy Osiyo.*

KIRISH

Raqamli platformalar orqali ekstremistik kontentning tarqalishi so‘nggi o‘n yil ichida Markaziy Osiyo jamiyatlari uchun eng dolzarb muammolardan biriga aylandi. Tarixiy muhim hududda joylashgan O‘zbekiston bu hodisani o‘ziga xos tarzda boshdan kechirmoqda. Viloyat shaharlarida, ayniqsa yosh erkaklar orasida smartfonlar va mobil internetning tez yoyilishi avlodlar uchun mutlaqo mavjud bo‘lmagan muloqot kanallarini ochib berdi. Bugungi vaziyatni tahliliy jihatdan ajratib turadigan narsa bu faoliyatning miqyosi emas, balki uning lingvistik xususiyati deb o`ylaymiz. o‘zbek tilidagi auditoriyaga yo‘lnaltirilgan ekstremistik kontentning katta qismi o‘zbek tilida yaratiladi, ammo ushbu kontentning lingvistik xususiyatlari deyarli tizimli ilmiy e‘tiborga olinmaydi.

Ushbu maqola aynan shu muammolarni yechishga qaratilgan. Lingvistik tahlilni huquqiy va tergov masalalariga tatbiq etuvchi fan hisoblangan sud lingvistikasiga tayanib, ikkita o‘zaro bog‘liq maqsad ko‘zlanmoqda: birinchisi o‘zbek tilidagi ekstremistik matnlarni kommunikativ funksiyasi bo‘yicha tasniflash uchun puxta asos yaratish; ikkinchisi bunday matnlarni stilometrik dalillar asosida mualliflariga nisbat berishning maqsadga muvofiqligini baholash. Ikki maqsad ham aniq amaliy ahamiyatga ega, shu bilan birga til va ideologiya o‘rtasidagi munosabat haqida nazariy savollarni ko‘taradi. Sud lingvistikasi huquqiy til talqinidan tortib muallif identifikatsiyasiga qadar keng ko‘lamli tahlil vazifalarini o‘z ichiga oladi.[1] Ushbu tadqiqot asosan ikki an‘anaga tayanadi: matnlar qanday qilib mafkuraviy pozitsiyalarni qurishini tushunishga yordam beradigan tanqidiy diskurs tahlili [2] va individual yozuvchilarni bir-biridan ajratib turadigan lingvistik odatlarni o‘lchash uchun miqdoriy usullar beruvchi stilometriya [3] sanaladi.

Ekstremistik diskursga oid tadqiqotlar, asosan arab va ingliz tilidagi kontentga qaratilgan holda, so‘nggi ikki o‘n yil davomida sezilarli darajada o‘sdi. Turkiy tillar, xususan o‘zbek tilidagi materiallarga oid qiyosiy ishlar ancha kamroq. Shu bilan birga, sobiq sovet davlatlarida onlayn radikallashuv bo‘yicha kichik bir taniqli adabiyot mavjud . O‘zbekchaning agglyutinativ morfologik tuzilishi — bitta so‘z butun gap ma‘nosini o‘z ichiga olishi mumkin bo‘lgan til — stilometrik tahlil uchun o‘ziga xos metodologik qiyinchiliklar tug‘diradi.

Korpus 2020–2024 yillar oralig‘ida ekstremistik yoki pro-ekstremistik kontent tarqatayotgan o‘n birta o‘zbek tilidagi Telegram kanalidan to‘plangan 214 ta matndan iborat bo‘lib, jami taxminan 87 000 so‘zni tashkil etadi. Kanallarni tanlashda o‘zbek fuqarolik jamiyati tashkilotlarining monitoring hisobotlari, nashr etilgan surishtiruv jurnalistikasi va bevosita kuzatuvdan foydalanildi. Barcha identifikatsiya ma‘lumotlari anonimlashtiriladi. Korpusga tegishli tadqiqot etikasi qoidalariga va O‘zbekistonning ekstremistik materiallar bilan ishlashni tartibga soluvchi huquqiy bazasiga muvofiq to‘plandi.

Har bir matn to‘rt darajada tahlil qilindi. Leksik darajada chastota tahlili o‘ziga xos so‘z boyligini, jumladan arabcha o‘zlashmalarni, diniy terminologiyani va kodlashtirilgan evfemizmlarni aniqladi. Morfologik darajada nominalizatsiya ko‘rsatkichlari, agent ot konstruksiyalari va fe‘l aspekti tarqalishi o‘rganildi. Sintaktik tahlil gap murakkabligi va buyruq konstruksiyalari zichligini kuzatildi. Pragmatik tahlil nutq akti turlariga, presuppozitsion mazmun va olmoshlar orqali auditoriya qurishga e‘tibor qaratdi.[4]

Muallif identifikatsiyasi uchun Burrouzning Delta usulining morfema darajasidagi varianti ishlab chiqildi:[5] butun so‘z shakllarini tahlil birligi sifatida ko‘rish o‘rniga, model chastota taqsimotlarini hisoblashdan avval har bir tokenni asos va qo‘shimcha komponentlarga ajratadi. Har bir matn uchun qirq yettita stilometrik xususiyat hisoblandi. Atribut tajribalari 70/30 o‘qitish-sinov bo‘linishi yordamida tergov jurnalistikasi orqali mustaqil ravishda muallifligi aniqlangan 31 ta matn to‘plamida tasdiqlandi.

Tahlildan uchta funksional kategoriya ajralib chiqdi. To‘g‘ridan-to‘g‘ri chaqiriq matnlari sintaktik jihatdan eng shaffof: ular ikkinchi shaxs buyruq konstruksiyalariga, diniy majburiyat ramkasiga va arab tilidan kelib chiqqan terminologiyadan keng foydalangan holda zo‘ravonlik lug‘atining zich qatlamiga tayanadi. Bu yerda lingvistik strategiya sodda: harakat talab qilinadi va diniy majburiyat o‘quvchidan har qanday shaxsiy ixtiyorni olib qo‘yish uchun jalb etiladi.[6]

Mafkuraviy ahamiyatiga ega bo`lgan matnlari farqli bo`ladi. Ular uzunroq, sintaktik jihatdan murakkab va o‘z ishontiruvchi ishlarini asosan presuppozitsiya orqali amalga oshiradi: davlatning qonuniy emasligiga, dunyoviy institutlarning korruptsiyasiga, qarshilik ko‘rsatish zaruriyatiga oid bahsli siyosiy da‘volarni bahs ochiq pozitsiyalar sifatida emas, balkim,asosli va tahlili insoniy hikoyalar orqali amalga oshiriladi. Bu presuppozitsion strategiya bunday matnlarni aniq qo‘zg‘atishga yo‘lnaltirilgan huquqiy doiralarda tasniflashni qiyinlashtiradi.

Yollash narrativlari eng murakkab auditoriya dizaynini ko‘rsatdi. Ular ochiq diniy lug‘atni minimallashtirib, umumiy etnik va ijtimoiy identifikatsiya, jamoaviy shikoyat

va tegishlilik hamda maqsad hissini ta‘kidladi. Dastlabki davat etish matnlarida aniq diniy ramkaning izchil bostirishi identifikatsiya murojaatigaidan mafkuraviy mazmunni bosqichma-bosqich kiritishga yo‘lnaltirilgan ataylab strategiyani ko‘rsatadi bu topilma onlayn yollanishga qarshi media savodxonligi dasturlari uchun bevosita ahamiyatga ega.[7]

Atribut tajribalari umumiy 78,4% aniqlikni, 81,2% aniqlik va 74,9% to‘liqlikni ko‘rsatdi. Morfema darajasidagi yondashuv so‘z darajasidagi bazaviy ko‘rsatkichlarni izchil ravishda oshib o‘tdi, ayniqsa so‘z-chastota taqsimotlari kamroq barqaror bo‘lgan qisqa matnlar uchun. Eng diskriminatsion xususiyatlar funksiya morfema chastotalari — xususan, kelishik qo‘shimchalari, ko‘makchilar va fe‘l kelishuvini bildiruvchi qo‘shimchalarning taqsimoti — keyin esa tinish belgilari odatlari va gap boshiga to‘g‘ri keladigan morfologik murakkablik bo‘ldi.

Korpusdagi ikkita matn birgalikda muallif identifikatsiyasi bilan mos keladigan aralash stilometrik profillarni ko‘rsatdi.[8] Ikkala holda ham bular uzunroq hujjatlar bo‘lib, tahlil ikki yozuvchining ishtirokini ko‘rsatdi — biri asosan mafkuraviy ramka bo‘limlari, ikkinchisi narrativ passajlar uchun mas‘ul. Bu topilma muallif identifikatsiyasining ehtimollik sifatida ko‘rilishi kerakligini va hal qiluvchi sifatida emas, boshqa tergov dalillari bilan birlashtirilishi zarurligini ta‘kidlaydi.

Matn kategoriyasi va muallif profili o‘rtasidagi korrelyatsiya eng qiziqarli natijalar qatoriga kirdi.[9] Mafkuraviy ishlov matnlari mualliflari mavzu yoki kanaldan qat‘iy nazar to‘g‘ridan-to‘g‘ri qo‘zg‘atish matnlari mualliflariga qaraganda sezilarli darajada yuqori morfologik murakkablikni ko‘rsatdi. Bu 4-bo‘limda aniqlangan funksional kategoriyalar nafaqat ritorik konventsiyalar ekanligini, balki ularni yaratayotgan shaxslarning ta‘lim va lingvistik fondidagi haqiqiy farqlarni aks ettirishini ko‘rsatadi.

Ushbu ishning amaliy oqibatlari uch sohaga taalluqlidir.[10]

Birinchidan, huquq-tartibot va sud amaliyoti uchun: bu yerda ishlab chiqilgan uch toifali tasnif tizimi ekstremizm bilan bog‘liq jinoyatlar turlari o‘rtasidagi mavjud huquqiy farqlarga maqbul darajada mos keladi va stilometrik usullar tegishli tasdiqlash va huquqiy standartlarga muvofiq daliliy vosita sifatida ko‘rib chiqilishiga yetarli aniqlikni namoyish etadi. Bir qancha yurisdiktsiyalar sudda sud lingvistik dalillarini qabul qilish uchun allaqachon protokollar ishlab chiqqan. O‘zbekiston huquq tizimi ham shunga o‘xshash doiralar ishlab chiqishdan foyda ko‘radi. Ikkinchidan, ekstremizmga qarshi kommunikatsiya uchun: chaqiriq matnlari aniq diniy ramkani tizimli ravishda bostirishi haqidagi topilma radikallashuvning dastlabki belgilari ommaviy ta‘limda qanday tasvirlanishini shakllantirishi kerak. Ekstremistik kontentning markeri sifatida faqat diniy lug‘atga qaratilgan xabardorlik kampaniyalari haqiqiy yollash materialining katta qismini o‘tkazib yuboradi.

Uchinchidan, kelajakdagi tadqiqotlar uchun: ushbu tadqiqotning eng muhim metodologik hissasi sud stilometriyasining o‘zbek tilining morfologik tuzilishiga muvaffaqiyatli moslashishi bo‘lishi mumkin. Buni tasdiqlangan, maqsadli vositaga aylantirish izohli korpuslar, avtomatlashtirilgan morfologik segmentatsiya va sud tasdiqlangan atribut algoritmlari bilan - tadqiqot ustuvorligi bo‘lishi kerak. O‘zbek tili

umumiy hisob-kitob lingvistik resurslar bilan ta‘minlanmagan va sud sohasi bu kamchilikka amaliy bahosi ayniqsa yuqori bo‘lgan kontekst hisoblanadi.

ADABIYOTLAR

1. Awan, I. (2017). Cyber-extremism: Isis and the power of social media. Society, 54(2), 138–149.

2. Burnap, P., & Williams, M. L. (2015). Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making. Policy & Internet, 7(2), 223–242.

3. Burrows, J. (2002). ‘Delta’: A measure of stylistic difference and a guide to likely authorship. Literary and Linguistic Computing, 17(3), 267–287.

4. Cheterian, V. (2012). The spread of radical Islam in post-Soviet states. Survival, 54(3), 1–28.

5. Coulthard, M., & Johnson, A. (2010). The Routledge Handbook of Forensic Linguistics. Routledge.

6. Grant, T. (2012). Txt 4n6: Idiolect free authorship analysis? Language and Law / Linguagem e Direito, 1(1), 1–15.

7. Koppel, M., Schler, J., & Argamon, S. (2009). Computational methods in authorship attribution. Journal of the American Society for Information Science and Technology, 60(1), 9–26.

8. Lemon, E. (2020). Manipulated inclusion: Authoritarian strategies in the post-Soviet space. Post-Soviet Affairs, 36(5–6), 422–440.

9. van Dijk, T. A. (1993). Principles of critical discourse analysis. Discourse & Society, 4(2), 249–283.

10. Wodak, R. (2001). The discourse-historical approach. In R. Wodak & M. Meyer (Eds.), Methods of Critical Discourse Analysis (pp. 63–94). SAGE.