TURLI TIZIMLI TILLAR IJTIMOIY-SIYOSIY DISKURSI PARALLEL KORPUSIDA EVFEMIZATSIYANI SUN'IY INTELLEKT ASOSIDA AVTOMATIK ANIQLASH

Malikajon Fayziyeva
Buxoro davlat universiteti Ingliz adabiyotshunosligi va tarjimashunoslik kafedrasi o'qituvchisi,
Pedagogika fanlari bo'yicha falsafa doktori (PhD)
m.ch.fayziyeva@buxdu.uz

Annotatsiya

Mazkur tezis turli tizimli tillar ijtimoiy-siyosiy diskursining parallel korpusida evfemizatsiyani sun'iy intellekt texnologiyalari asosida avtomatik aniqlash masalasiga bag'ishlangan. Evfemizatsiya kommunikativ jihatdan noqulay, salbiy, bahsli yoki ijtimoiy jihatdan sezgir mazmunni bevosita ifodalash o'rniga yumshatilgan, bilvosita yoki neytrallashtirilgan shaklda berish strategiyasi sifatida qaraladi. Tadqiqotda evfemizatsiyaning struktur-semantik va pragmatik xususiyatlari, uning ijtimoiy-siyosiy diskursdagi kommunikativ funksiyalari hamda turli tizimli tillarda namoyon bo'lishi tahlil qilinadi. Parallel korpus, lingvistik annotatsiya va tabiiy tilni qayta ishlash (NLP) texnologiyalarining integratsiyasi evfemistik birliklarni avtomatik aniqlash uchun istiqbolli metodologik asos sifatida taklif etiladi. Tadqiqotda evfemizmni aniqlashda kontekstual ma'no, semantik siljish, baholovchi komponent, diskurs sohasi, kollokatsion xususiyat va pragmatik funksiya kabi belgilarni hisobga olish zarurligi asoslanadi. Zamonaviy tadqiqotlar evfemizmlarni avtomatik aniqlashda kontekstning hal qiluvchi ahamiyatga ega ekanligini va bir xil birlik turli kontekstlarda literal yoki evfemistik ma'noda qo'llanishi mumkinligini ko'rsatmoqda (Lee et al., 2022; Poh et al., 2026).

Kalit so'zlar: evfemizatsiya, evfemizm, ijtimoiy-siyosiy diskurs, parallel korpus, sun'iy intellekt, avtomatik aniqlash, tabiiy tilni qayta ishlash, kompyuter lingvistikasi, pragmatika.

Аннотация

Данная статья посвящена проблеме автоматического выявления эвфемизации в общественно-политическом дискурсе типологически различных языков на основе параллельного корпуса и технологий искусственного интеллекта. Эвфемизация рассматривается как коммуникативная стратегия, посредством которой социально чувствительная, негативная или спорная информация выражается косвенно, смягчённо или нейтрализованно. В исследовании рассматриваются структурно-семантические и прагматические особенности эвфемистических единиц, их коммуникативные функции в общественно-политическом дискурсе и особенности их реализации в типологически различных языках. Интеграция параллельного корпуса, лингвистической аннотации и методов обработки естественного языка рассматривается как перспективная основа для автоматического выявления эвфемизации. Особое внимание уделяется контекстуальному значению, семантическому сдвигу, оценочности, дискурсивной сфере, коллокационным характеристикам и прагматической функции эвфемистических единиц. Современные исследования подтверждают, что контекст является одним из ключевых факторов при автоматическом выявлении эвфемизмов, поскольку одна и та же единица может употребляться как буквально, так и эвфемистически в зависимости от контекста (Lee et al., 2022; Poh et al., 2026).

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

Ключевые слова: эвфемизация, эвфемизм, общественно-политический дискурс, параллельный корпус, искусственный интеллект, автоматическое выявление, компьютерная лингвистика, прагматика.

Abstract

This paper investigates the automatic detection of euphemization in the socio-political discourse of typologically different languages using parallel corpora and artificial intelligence technologies. Euphemization is viewed as a communicative strategy through which socially sensitive, negative, controversial, or potentially face-threatening meanings are expressed indirectly, mitigated, or neutralized. The study focuses on the structural-semantic and pragmatic characteristics of euphemistic expressions, their communicative functions in socio-political discourse, and their cross-linguistic realization. The integration of parallel corpora, linguistic annotation, and natural language processing methods is proposed as a promising methodological basis for the automatic detection of euphemization. Particular attention is paid to contextual meaning, semantic shift, evaluative meaning, discourse domain, collocational patterns, and pragmatic function. Recent computational studies demonstrate that contextual information is crucial for euphemism detection because the same expression may have literal, figurative, or euphemistic interpretations depending on its context (Lee et al., 2022; Poh et al., 2026).

Keywords: euphemization, euphemism, socio-political discourse, parallel corpus, artificial intelligence, automatic detection, natural language processing, computational linguistics, pragmatics.

KIRISH

Tilning ijtimoiy-siyosiy kommunikatsiyadagi vazifasi faqat axborotni yetkazish bilan chegaralanmaydi. Siyosiy nutq va ommaviy axborot diskursida til vositalari voqelikni talqin qilish, muayyan hodisaga baho berish, ijtimoiy guruhlar haqidagi tasavvurlarni shakllantirish va adresatning munosabatini boshqarish vositasi sifatida ham ishlatiladi. Shu nuqtai nazardan evfemizatsiya ijtimoiy-siyosiy kommunikatsiyaning muhim diskursiv strategiyalaridan biri hisoblanadi.

Allan va Burridge evfemizmni noqulay yoki taqiqlangan mazmunni yumshoqroq shaklda ifodalash bilan bog'laydi hamda evfemizm, disfemizm va ortofemizmni bir-biridan farqlash zarurligini ta'kidlaydi (Allan & Burridge, 2006).

Siyosiy diskursda evfemizmning vazifasi faqat "yumshatish"dan iborat emas. U siyosiy subyektning ijobiy imidjini saqlash, noqulay masalalarni bevosita nomlamaslik, mas'uliyatni kamaytirish yoki bahsli hodisani ijtimoiy jihatdan maqbulroq shaklda taqdim etish uchun qo'llanishi mumkin. Crespo-Fernández siyosiy diskursda evfemizm o'zini himoya qilish, ijobiy o'zini namoyon qilish va noqulay mavzularni yashirish kabi funksiyalarni bajarishini ko'rsatadi (Crespo-Fernández, 2018).

Shu sababli evfemizatsiyani kompyuter yordamida avtomatik aniqlash masalasi oddiy leksik qidiruvdan ancha murakkabdir. Masalan, restructuring, collateral damage, enhanced interrogation yoki downsizing kabi birliklar turli kontekstlarda turlicha semantik va pragmatik qiymatga ega bo'lishi mumkin. Demak, avtomatik model nafaqat birlikning shaklini, balki uning kontekstual ma'nosi va kommunikativ funksiyasini ham hisobga olishi kerak.

So'nggi yillarda evfemizmlarni avtomatik aniqlash bo'yicha kompyuter lingvistikasi tadqiqotlari rivojlanmoqda. Zhu va Bhat (2021) evfemistik frazalarni avtomatik aniqlash uchun phrase mining, word

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

embeddings va masked language model yondashuvlarini birlashtirgan. Mualliflar ko'p so'zli evfemistik birliklarni aniqlash masalasini alohida NLP vazifasi sifatida ko'rsatgan.

2022-yildagi Euphemism Detection Shared Task esa evfemistik birliklarning literal yoki euphemistic usage ekanligini kontekst asosida aniqlash masalasini standart NLP vazifalaridan biri sifatida shakllantirdi (Lee et al., 2022).

2026-yilda esa ko'p tilli evfemizm benchmarki ishlab chiqilib, bir nechta tillarda evfemizmni aniqlash uchun yagona baholash muhiti taklif qilindi. Tadqiqotda evfemistik birliklarning turli tillar va madaniyatlarda shakl va kontekst jihatidan farqlanishi alohida qayd etilgan (Poh et al., 2026).

Shu jihatdan o'zbek va ingliz tillari kabi tipologik jihatdan farqli tillarning ijtimoiy-siyosiy diskursini parallel korpus asosida o'rganish dolzarb ilmiy masala hisoblanadi.

METODOLOGIYA VA TADQIQOT

Tadqiqotda korpus lingvistikasi, qiyosiy-tipologik tahlil, struktur-semantik tahlil, pragmatik tahlil, kontekstual tahlil va NLP/AI metodlari kompleks tarzda qo'llanadi.

Tadqiqotning konseptual modelini quyidagicha ifodalash mumkin:

Parallel korpus → preprocessing → lingvistik annotatsiya → evfemistik birliklarni belgilash → xususiyatlarni ajratish → AI modeli → avtomatik aniqlash → pragmatik tasnif.

Tadqiqot uchun ingliz va o'zbek tillaridagi ijtimoiy-siyosiy matnlardan tashkil topgan parallel korpus shakllantiriladi. Korpusga siyosiy nutqlar, hukumat bayonotlari, rasmiy xabarlar, xalqaro munosabatlarga oid materiallar va ommaviy axborot vositalaridagi ijtimoiy-siyosiy matnlarni kiritish mumkin.

Parallel korpusdan foydalanish bir xil kommunikativ mazmunning ikki tilda qanday ifodalanishini aniqlash imkonini beradi. O'zbek va ingliz siyosiy matnlarida parallel korpus va korpus-lingvistik metodlarning qo'llanishi siyosiy leksikaning pragmatik va ideologik xususiyatlarini aniqlash uchun samarali ekanligi zamonaviy tadqiqotlarda ham ko'rsatilgan.

Korpusdagi evfemistik birliklarni avtomatik aniqlashdan oldin ularni inson ekspertlari tomonidan annotatsiyalash maqsadga muvofiq.

Annotatsiya quyidagi parametrlarni o'z ichiga olishi mumkin:

№	Annotatsiya parametri	Mazmuni	Misol
1	**Formal structure** (Struktur shakl)	Evfemistik birlikning grammatik-leksik shakli	so'z, so'z birikmasi, frazeologizm, perifraz
2	**Literal meaning** (Bevosita ma'no)	Birlikning lug'aviy yoki to'g'ridan-to'g'ri anglatadigan ma'nosi	damage — zarar
3	**Euphemistic meaning** (Evfemistik ma'no)	Kontekstda bevosita nomlanmayotgan salbiy yoki noqulay mazmunni yumshatib ifodalashi	collateral damage — tinch aholiga yetkazilgan zarar
4	**Semantic mechanism** (Semantik mexanizm)	Evfemizatsiya yuzaga kelishining semantik usuli	substitution, metaphor, metonymy, generalization, abstraction
5	**Discourse domain**	Evfemizm qo'llanayotgan	political, military,

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

№	Annotatsiya parametri	Mazmuni	Misol
	(Diskurs sohasi)	ijtimoiy-siyosiy soha	economic, social, diplomatic
6	**Pragmatic function** (Pragmatik funksiya)	Evfemizmning kommunikativ maqsadi	mitigation, concealment, legitimization, face-saving
7	**Evaluative meaning** (Baholovchi ma'no)	Birlikning kontekstda yuzaga keltiradigan bahosi	positive, neutral, negative
8	**Target meaning** (Nishon ma'no)	Evfemistik ifoda orqali aslida nazarda tutilayotgan tushuncha	downsizing → job losses
9	**Contextual features** (Kontekstual belgilar)	Evfemistik talqinni yuzaga keltiruvchi atrofdagi lingvistik birliklar	kollokatsiyalar, sintaktik muhit, kontekst
10	**Cross-linguistic equivalent** (Tillararo ekvivalent)	Parallel korpusda boshqa tildagi mos ifoda	English → Uzbek
11	**Translation transformation** (Tarjima transformatsiyasi)	Evfemistik birlikning boshqa tilga o'tish usuli	substitution, modulation, explicitation, paraphrase, neutralization
12	**Euphemization degree** (Evfemizatsiya darajasi)	Asl mazmunning qanchalik yumshatilgan yoki yashirilganligi	low, medium, high

Bunday ko'p darajali annotatsiya kelajakda oddiy "evfemizm bor/yo'q" klassifikatsiyasidan tashqari, evfemizatsiya turini va pragmatik funksiyasini ham avtomatik aniqlash imkonini beradi.

Evfemizmlarni avtomatik aniqlashda zamonaviy NLP modellaridan foydalanish mumkin. Jumladan, contextual embeddings va transformer-based modellar kontekstga bog'liq ma'nolarni aniqlash uchun muhim imkoniyatlarga ega.

Zhu va Bhat (2021) ko'p so'zli evfemistik frazalarni aniqlashda phrase mining, word embeddings va SpanBERT asosidagi masked language modelni birlashtirgan.

Bundan tashqari, Hu va boshqalar (2024) evfemizm identifikatsiyasida domain gap muammosini ko'rsatib, turli domenlardagi ma'lumotlar taqsimotining farqlanishi model samaradorligiga ta'sir qilishini aniqlagan. Ular feature alignment asosida domenlararo farqni kamaytiruvchi modelni taklif qilgan.

Bu natija sizning tadqiqotingiz uchun ayniqsa muhim: ijtimoiy-siyosiy diskurs uchun yaratilgan modelni boshqa domenlardan olingan evfemizm ma'lumotlariga to'g'ridan-to'g'ri qo'llash yetarli bo'lmasligi mumkin.

NATIJALAR

Tadqiqotning nazariy va metodologik natijalari asosida evfemizatsiyani avtomatik aniqlash uchun bir nechta muhim xususiyatlarni ajratish mumkin.

Birinchidan, kontekstual xususiyat muhim hisoblanadi. Bir xil birlik literal yoki evfemistik ma'noda ishlatilishi mumkin. Zamonaviy multilingual benchmark ham aynan shu muammoni qayd etadi: bir ifoda kontekstga qarab literal, figurativ yoki evfemistik ma'noga ega bo'lishi mumkin (Poh et al., 2026).

Ikkinchidan, semantik kontrast muhim belgi bo'la oladi. Misol uchun, collateral damage, kontekstga qarab oddiy "zarar" emas, balki harbiy amaliyot oqibatida tinch aholining zarar ko'rishi yoki halok bo'lishini yumshatilgan tarzda ifodalashi mumkin.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

Uchinchidan, pragmatik funksiya evfemizatsiyaning asosiy ko‘rsatkichlaridan biridir. Siyosiy diskursda evfemizmlar noqulay mavzularni yumshatish, siyosiy subyektning ijobiy qiyofasini saqlash yoki bahsli masalalarni kamroq salbiy ko‘rinishda taqdim etish uchun xizmat qilishi mumkin (Crespo-Fernández, 2018).

To‘rtinchidan, parallel korpusdagi tillararo munosabat alohida xususiyat sifatida qaraladi. Ingliz tilidagi evfemistik birlik o‘zbek tilida evfemistik ekvivalent;

neytral ekvivalent; to‘g‘ridan-to‘g‘ri nominatsiya; izohlovchi birlik; perifraz orqali tarjima qilinishi mumkin.

Shu asosda evfemizatsiyaning quyidagi modelini taklif qilish mumkin:

Lexical form → Context → Semantic shift → Euphemistic interpretation → Pragmatic function → Cross-linguistic equivalent.

Avtomatik tizimning samaradorligini baholashda Precision, Recall va F1-score asosiy ko‘rsatkichlar sifatida qo‘llanishi mumkin. F1-score evfemistik va noevfemistik birliklarni aniqlashdagi muvozanatni baholash imkonini beradi. 2022-yilgi Euphemism Detection Shared Task ham evfemizmni kontekst asosida avtomatik aniqlashni alohida baholash vazifasi sifatida shakllantirgan.

Bundan tashqari, zamonaviy tadqiqotlarda GPT turidagi yirik til modellari va boshqa transformer modellarining ko‘p tilli evfemizmni aniqlashdagi imkoniyatlari ham sinab ko‘rilmoqda. 2026-yilgi multilingual benchmarkda (ko‘p tilli baholash mezoni) zero-shot (oldindan berilgan namunalarsiz yondashuv) va few-shot (cheklangan miqdordagi namunalar asosidagi yondashuv) yondashuvlar hamda GPT-5-nano va Qwen3-4B kabi modellar bo‘yicha tajribalar taqdim etilgan.

Bu holat evfemizatsiyani avtomatik aniqlashni faqat klassik machine learning bilan emas, balki LLM + korpus + lingvistik annotatsiya integratsiyasi asosida ham tadqiq qilish imkonini beradi.

XULOSA

Turli tizimli tillar ijtimoiy-siyosiy diskursi parallel korpusida evfemizatsiyani sun’iy intellekt asosida avtomatik aniqlash zamonaviy kompyuter lingvistikasi, pragmatika va korpus lingvistikasi kesishmasida shakllanayotgan istiqbolli yo‘nalishlardan biridir.

Tadqiqot natijalari evfemizatsiyani avtomatik aniqlashda faqat leksik shaklga tayanish yetarli emasligini ko‘rsatadi. Evfemistik birlikning kontekstual ma’nosi, semantik siljishi, baholovchi xususiyati, diskurs sohasi va pragmatik funksiyasi kompleks ravishda hisobga olinishi lozim.

Parallel korpusning qo‘llanishi esa o‘zbek va ingliz tillarida evfemizatsiyaning umumiy hamda milliy-madaniy xususiyatlarini aniqlash, evfemistik birliklarning tarjima ekvivalentlari va transformatsiyalarini qiyoslash imkonini beradi. Zamonaviy NLP tadqiqotlari evfemizmni avtomatik aniqlashning real va alohida kompyuter-lingvistik vazifa sifatida shakllanganini ko‘rsatmoqda (Zhu & Bhat, 2021; Lee et al., 2022; Hu et al., 2024; Poh et al., 2026).

FOYDALANILGAN ADABIYOTLAR

1. Allan K., Burridge K. Forbidden Words: Taboo and the Censoring of Language. Cambridge: Cambridge University Press, 2006. 320 p.
2. Crespo-Fernández E. Euphemism as a discursive strategy in US local and state politics // Journal of Language and Politics. 2018. Vol. 17, No. 6. P. 789–811. DOI: 10.1075/jlp.17040.cre.
3. Felt C., Riloff E. Recognizing euphemisms and dysphemisms using sentiment analysis // Proceedings of the Second Workshop on Figurative Language Processing. 2020. P. 136–145.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

4. Zhu W., Bhat S. Euphemistic phrase detection by masked language model // Findings of the Association for Computational Linguistics: EMNLP 2021. 2021. P. 1636–1647.
5. Lee P., Feldman A., Peng J. A report on the Euphemism Detection Shared Task // Proceedings of the 3rd Workshop on Figurative Language Processing. 2022. P. 142–151.
6. Hu Y., Li J., Wu M., Huang Z., Chen G., Sha Y. Uncovering and mitigating the hidden chasm: A study on the text-text domain gap in euphemism identification // Proceedings of the AAAI Conference on Artificial Intelligence. 2024. Vol. 38, No. 16. P. 18270–18278. DOI: 10.1609/aaai.v38i16.29786.
7. Poh W., Sammartino J., Andrew J., Kieraś W., Zawadzka-Paluektau N., Dilai I., Barak L., Peng J., Feldman A. The multilingual euphemism benchmark: Datasets and baselines for pragmatic language understanding // Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026). 2026. P. 7307–7319. DOI: 10.63317/5im6zovfyg3t.
8. Botirova N. Ideological framing in political translation: A comparative analysis of Uzbek and English news discourse using parallel corpora // Philology Matters. 2025. No. 3. P. 76–91.
9. Abdurakhmonova N., Botirova N. Контекстуальный анализ параллельных текстов при формировании переводческого словаря политической сферы // Международный журнал «Искусство слова». 2026.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.