

O‘ZBEK TILIDAGI HARBIY TERMINLARNI MILLIY TIL KORPUSIDA TEGLASH MASALALARI

Ilhom Xolmurotov

Nizomiy nomidagi O‘zbekiston milliy pedagogika universiteti o‘qituvchisi

Annotatsiya: Ushbu maqolada o‘zbek tilidagi harbiy terminlarning lingvistik xususiyatlari, ularni Milliy til korpusida teglashning ilmiy-nazariy asoslari hamda amaliy yechimlari tahlil qilinadi. Harbiy terminlar tizimi o‘zbek tili terminologik qatlamining muhim tarkibiy qismi sifatida ko‘riladi. Maqolada harbiy terminlarning semantik, morfologik va funksional xususiyatlari, ularning etimologik manbalari, shuningdek, korpus lingvistikasida ularni avtomatik aniqlash va teglash masalalari yoritilgan. O‘zbek tili uchun yaratilayotgan Milliy til korpusida harbiy terminlarni alohida teglash zarurati asoslanadi hamda “tagset” modeli uchun takliflar ilgari suriladi. Tadqiqot natijasida o‘zbek tilining harbiy leksikasini elektron korpusga kiritish jarayonida ilmiy-texnik va metodologik yondashuvlar ishlab chiqilgan.

Kalit so‘zlar: harbiy terminologiya, Milliy til korpusi, teglash, annotatsiya, korpus lingvistika, lingvistik model, tagset

1. Kirish

XXI asrda tilshunoslikning eng faol rivojlanayotgan yo‘nalishlaridan biri korpus lingvistikasidir. Bu soha tildagi birliklarni raqamli shaklda jamlash, ularni morfologik, sintaktik va semantik jihatdan belgilash orqali til tizimini chuqur tahlil qilish imkonini

beradi. Til resurslarining raqamli shaklda to‘planishi, ularning avtomatik qayta ishlanishi, mashina o‘qitish (machine learning) texnologiyalari bilan uyg‘unlashuvi bugungi kunda lingvistik tadqiqotlarning asosiy yo‘nalishlaridan biriga aylangan.

O‘zbek tili uchun Milliy til korpusi (O‘zMK) yaratilayotgani tilshunoslikda yangi bosqichni boshlab berdi. Mazkur korpus o‘zbek tilining leksik, morfologik, sintaktik va uslubiy qatlamlarini to‘liq qamrab olishni maqsad qilgan. Shu jihatdan, korpusda alohida tematik qatlamlarni, xususan, **harbiy terminologiyani** aniqlash va teg bilan belgilash muhim vazifa hisoblanadi.

Harbiy terminlar har qanday millatning tarixiy taraqqiyoti, davlat mustaqilligi, harbiy madaniyati va texnologik taraqqiyot darajasini ifoda etadi. O‘zbek tilida ham bu qatlam boy va murakkab tizimni tashkil etadi: u qadimgi turkiy terminlardan tortib, rus va g‘arb tillaridan kirib kelgan zamonaviy harbiy atamalargacha bo‘lgan so‘zlarni o‘z ichiga oladi. Masalan, “lashkar”, “sarkarda”, “jangovar”, “artilleriya”, “poligon”, “tankchi”, “brigada” kabi terminlar semantik, morfologik va funksional jihatdan turlicha xususiyatlarga ega.

Shu boisdan, o‘zbek tilidagi harbiy terminlarni Milliy til korpusida alohida teglash ularni lingvistik, terminologik va texnik jihatdan tahlil qilishni osonlashtiradi. Bu nafaqat ilmiy maqsadlarda, balki harbiy tarjima tizimlari, axborot xavfsizligi, avtomatik tahlil va sun‘iy intellekt asosidagi dasturlarda ham muhim ahamiyat kasb etadi.

2. Adabiyotlar sharhi

O‘zbek tilshunosligida terminologiya masalalari ko‘plab olimlar tomonidan tadqiq etilgan. Xususan, A. Qodirov (2020) o‘zbek tilida harbiy terminlarning shakllanish tarixini o‘rganib, ularning asosiy qatlamlarini qadimgi turkiy va rus tilidan kirgan birliklar tashkil etishini qayd etgan. B. Alimov (2021) o‘zbek terminologiyasining tizimli tuzilishini, ularning xalqaro terminologik standartlar bilan bog‘liqligini yoritgan. N. Abdullayeva va

boshqalar (2023) esa O‘zbek tilining Milliy korpusini yaratish jarayonidagi leksik va morfologik muammolarni ko‘rsatib, ularni avtomatlashtirish masalalarini tahlil qilganlar.

Chet tillardagi tajribalar ham bu yo‘nalishda muhim ahamiyatga ega. Masalan, rus tilidagi Milliy korpus (RNC), ingliz tilidagi British National Corpus (BNC) yoki Xitoy Harbiy Korpusi (Chinese Military Corpus) harbiy terminlarni alohida tematik qatlam sifatida ajratgan. Bu tajribalar o‘zbek tili uchun moslashtirilgan model ishlab chiqishda amaliy asos bo‘la oladi.

O‘zbek tilida korpus lingvistikasi sohasidagi dastlabki ishlar M. Sharipov (2023), M. Abjalova (2022), O. Umirov (2021) tomonidan olib borilgan. Ularning tadqiqotlarida morfologik teglash, tokenizatsiya, lemmatizatsiya va korpus tahlilining avtomatik modellari ishlab chiqilgan. Shu bilan birga, maxsus terminologik qatlamlar, jumladan, harbiy terminlarni teglash masalasi hali keng qamrovda o‘rganilmagan.

Harbiy terminologiyaning o‘ziga xos xususiyatlari haqida o‘zbek tilshunosligida ham qator asarlar mavjud. Masalan, Qodirov (2020) “O‘zbek tilida harbiy terminlarning shakllanishi va rivoji” nomli tadqiqotida ularning ijtimoiy-siyosiy omillar bilan bog‘liqligini ko‘rsatgan. Shuningdek, T. Yusupov (2018) harbiy leksikaning sinonimik qatlamini, S. Jo‘rayev (2019) esa semantik differensiasiyasini o‘rgangan.

Demak, mavjud adabiyotlar tahlili shuni ko‘rsatadiki, o‘zbek tilida harbiy terminlarni **korpus asosida teglash** masalasi hali ilmiy jihatdan to‘liq ishlab chiqilmagan. Aynan shu jihat mazkur maqolaning ilmiy yangiligini belgilaydi.

3. O‘zbek tilidagi harbiy terminlarning shakllanish manbalari

O‘zbek tilida harbiy terminologiya shakllanishining tarixiy jarayoni uch asosiy manbadan oziqlanadi:

1. Qadimgi turkiy harbiy leksika,

2. Arabcha va forsha harbiy terminlar,

3. Rus va g'arbiy Yevropa tillaridan kirgan zamonaviy terminlar.

3.1. Qadimgi turkiy qatlam

Qadimgi turkiy yozma yodgorliklarda — *Orxon-Enasoy bitiklari*, *Qutadg'u bilig*, *Devonu lug'otit-turk* asarlarida — ko'plab harbiy terminlar uchraydi. “Lashkar”, “sarkarda”, “qal'a”, “sipoh”, “jang”, “otliq”, “yov”, “qilich”, “nayza”, “qalqon” kabi birliklar o'zbek harbiy leksikasining ildizini tashkil etadi. Ularning ko'pi semantik jihatdan umumiy harbiy tushunchalarni ifodalaydi va bugungi kungacha faol qo'llanadi.

Bu qatlamdagi terminlar asosan o'zbek tili fonetikasi va morfologiyasiga to'liq moslashgan, shuningdek, so'z yasash imkoniyatlari keng. Masalan, “jang” so'zidan “jangchi”, “jangovar”, “jangarilik” kabi hosilalar yasalgan.

3.2. Arabcha va forsha qatlam

IX–XV asrlar oralig'ida o'zbek tili (turkiy tillar)ga islom madaniyati orqali kirib kelgan arabcha va forsha harbiy terminlar ham muhim qatlamni tashkil etadi. “Amir”, “sipoh”, “sarboz”, “himoya”, “mudofaa”, “dushman”, “muhozira”, “g'alaba” kabi so'zlar arabcha yoki forsha manbalarga ega.

Bu terminlar asosan diniy-harbiy ma'nolarni ifodalagan va o'zbek tilining adabiy leksikasiga chuqur singib ketgan. Ularning semantik maydoni kengayib, bugungi kunda ham rasmiy-idoraviy til uslubida faol ishlatiladi.

3.3. Rus va Yevropa qatlam

XIX asr oxiridan boshlab harbiy terminologiyada rus tili orqali g'arb tillaridan kirgan terminlar faol qo'llanila boshladi. “Armiya”, “batalyon”, “diviziya”, “artilleriya”, “tank”, “raketa”, “general”, “polkovnik”, “aviatsiya”, “strategiya”, “taktika” kabi terminlar

xalqaro xarakterga ega bo‘lib, o‘zbek tili fonetik tizimiga nisbatan moslashgan holda ishlatiladi.

Bu qatlam hozirgi o‘zbek harbiy terminologiyasining asosiy qismini tashkil etadi va xalqaro standartlarga mos ravishda yangi terminlar bilan boyib bormoqda (masalan, “dron”, “robototexnika”, “kiberxavfsizlik”, “logistika” va boshqalar).

4. Harbiy terminlarning morfologik va semantik tahlili

Harbiy terminlar, boshqa terminologik birliklar singari, tildagi umumiy grammatik tizimga bo‘ysunadi. Ammo ularning **morfologik shakllanishi va semantik turlanishi** o‘ziga xos xususiyatlarga ega.

4.1. Morfologik xususiyatlar

Harbiy terminlar so‘z turkumlari nuqtayi nazaridan asosan otlar (armiya, askar, jangchi, general), fe‘llar (jang qilmoq, himoya qilmoq, saf tortmoq), sifatlar (jangovar, mudofaa, strategik) shaklida ifodalanadi. So‘z yasashda esa ot + affiks (jang+chi), ot + ot (harbiy xizmat), sifat + ot (strategik reja), yoki qisqartmalar (O‘zDJT — O‘zbekiston Davlat Jangovar Tayyorlov) kabi modellar ustuvor hisoblanadi.

Affikslarning termin hosil qilishdagi roli alohida e‘tiborga loyiq:

- **-chi** (jangchi, artilleriyachi, razvedkachi) — kasb yoki faoliyat sohibini bildiradi;
- **-lik** (harbiylik, zobitlik) — holat yoki xususiyat;
- **-kor** (jangkor, g‘alabakor) — harakatga moyillikni bildiradi.

4.2. Semantik turlanish

Semantik jihatdan harbiy terminlar bir necha tematik maydonlarga ajratiladi:

1. Harbiy unvon va lavozimlar: general, polkovnik, mayor, serjant;

2. **Harbiy qismlar va tuzilmalar:** batalyon, brigada, shtab, polk;
3. **Qurol-aslaha va texnika:** tank, artilleriya, samolyot, dron;
4. **Harbiy harakatlar:** hujum, mudofaa, razvedka, operatsiya;
5. **Tashkiliy tushunchalar:** strategiya, taktika, logistika;
6. **Harbiy joy nomlari va ob'ektlar:** poligon, garnizon, front, baza.

Bu tasnif korpus teglash jarayonida semantik segmentatsiyani aniq belgilashga yordam beradi. Har bir maydon uchun alohida teg kategoriyasini yaratish korpusdagi semantik tahlil sifatini oshiradi.

5. Harbiy terminlarni teglashning nazariy asoslari

Korpus lingvistikasi terminlarni teglash (tagging) jarayonida lingvistik birlikning ma'nosi, sintaktik funksiyasi va kontekstual roli asosida belgilanadi. Harbiy terminlarni teglashda quyidagi **nazariy tamoyillar** muhimdir:

1. **Kontekstual aniqlik** – termin faqat harbiy ma'noda ishlatilganida teg qo'llanadi (*“brigada” so'zi tibbiyot brigadasi emas, harbiy brigada sifatida*).
2. **Morfologik moslik** – terminning affiksial shakllari (*jangchi, jangchilar, jangchilik*) bir lemma sifatida ko'riladi.
3. **Ko'p ma'nolilikni hisobga olish** – homonimik so'zlar (*“pozitsiya” – harbiy joy yoki nutqdagi pozitsiya*) uchun disambiguatsiya zarur.
4. **Semantik darajalilik** – harbiy terminologiya ichida yuqori (*strategik*) va past (*taktik*) darajadagi birliklar ajratiladi.

Bu tamoyillar asosida yaratilgan tagset korpusda terminlarni avtomatik tarzda tanib olish imkonini beradi.

6. Milliy til korpusi va harbiy terminlar uchun tagset modeli

O‘zbek tilining Milliy korpusi (O‘zMK) bugungi kunda 100 milliondan ortiq so‘z shaklini o‘z ichiga olgan yirik lingvistik platformadir. Ammo bu korpusda tematik teglash, jumladan, harbiy qatlamning alohida belgilanmaganligi tadqiqotchilar uchun cheklov yaratadi.

6.1. Tagset modeli

Harbiy terminlar uchun maxsus **“MIL-Tagset”** modeli taklif etiladi. U quyidagi asosiy teglarni o‘z ichiga oladi:

Teg nomi	Tavsifi	Misol
MIL_TERM	Umumiy harbiy termin	<i>harbiy, jangovar</i>
MIL_RANK	Harbiy unvon	<i>serjant, kapitan</i>
MIL_UNIT	Harbiy qism yoki tuzilma	<i>batalyon, brigada</i>
MIL_EQUIP	Qurol-aslaha, texnika	<i>tank, raketa, artilneriya</i>
MIL_ACTION	Harbiy harakat yoki operatsiya	<i>hujum, mudofaa, razvedka</i>
MIL_LOC	Harbiy joy nomi	<i>front, baza, poligon</i>
MIL_ORG	Harbiy tashkilot yoki muassasa	<i>shtab, garnizon, akademiya</i>
MIL_LEGAL	Harbiy huquq va tartib terminlari	<i>nizom, intizom, buyruq</i>
MIL_MISC	Boshqa harbiy terminlar	<i>askarlik, zobitlik, xizmat</i>

Mazkur model XML yoki CoNLL-U formatida kodlash imkonini beradi. Masalan:

<hw lemma="jangchi" pos="NOUN" tag="MIL_TERM"/>

<hw lemma="kapitan" pos="NOUN" tag="MIL_RANK"/>

6.2. Annotatsiya vositalari

Amaliy teglash jarayonida **WebAnno**, **INCEpTION** yoki **BRAT** kabi vositalar qo'llanishi mumkin. Bu dasturlar ko'p foydalanuvchili annotatsiyani, mashina yordamida avtomatik teglashni va inson tomonidan nazoratni uyg'unlashtirish imkonini beradi.

Bunda "semi-automatic tagging" — ya'ni dastlab mashina annotatsiyasi, so'ng inson tomonidan tuzatish — eng maqbul yondashuv deb hisoblanadi.

7. Annotatsiya jarayoni: inson va mashina uyg'unligi

Korpus lingvistikasi amaliyotida annotatsiya — bu til birliklarini ma'lum belgilarga (morfologik, sintaktik, semantik, pragmatik) ko'ra tavsiflash jarayonidir. Harbiy terminlarni annotatsiyalash jarayoni ham xuddi shu tamoyillarga asoslanadi, biroq u o'ziga xos jihatlarni talab etadi. Chunki harbiy leksika o'zining maxsus, rasmiy va ko'p qatlamli xususiyatlari bilan ajralib turadi.

7.1. Annotatsiya bosqichlari

Harbiy terminlarni Milliy til korpusida belgilash jarayoni quyidagi bosqichlarda amalga oshiriladi:

1. **Material tanlash** — harbiy matnlar, hujjatlar, darsliklar, farmonlar va maqolalardan tanlab olinadi.
2. **Termin lug'atini yaratish** — harbiy atamalar ro'yxati yaratiladi (1000 dan ortiq birlik).
3. **Avtomatik teglash** — mashina o'qitish asosida dastlabki teglash amalga oshiriladi.

4. Inson annotatsiyasi — ekspert tilshunoslar tomonidan avtomatik teglar tekshiriladi va tuzatiladi.

5. Sifat nazorati — inter-annotator reliability (annotatorlar o‘rtasidagi kelishuv darajasi) ko‘rsatkichi hisoblanadi.

7.2. Mashina va inson annotatsiyasining uyg‘unligi

Amaliyotda mashina annotatsiyasi jarayonni tezlashtiradi, lekin semantik aniqlikni inson nazoratisiz ta‘minlay olmaydi. Shu sababli, “hybrid annotation” modeli qo‘llaniladi — dastlab mashina (NER modeli yoki termin lug‘atiga asoslangan classifier) teg qo‘yadi, so‘ng annotator (tilshunos) tekshiradi.

Masalan, *INCEpTION* platformasi mashina tomonidan kiritilgan teglar uchun ishonchlilik ko‘rsatkichini chiqaradi. Annotator esa noto‘g‘ri joylarni tuzatadi. O‘zbek tilida bu jarayonni avtomatlashtirish uchun morfologik analizator (UzMorph, 2023) bilan integratsiya muhim ahamiyatga ega.

8. Harbiy terminlarni teglashda amaliy yechimlar (O‘zbek tili misolida)

Harbiy terminlarni avtomatik tanib olish uchun bir necha yondashuv sinovdan o‘tkazilgan. Quyida ularning afzallik va kamchiliklari keltiriladi:

Yondashuv	Tavsif	Afzallik	Cheklov
Lug‘atga asoslangan	Tayyor terminlar ro‘yxati bilan solishtirish	Aniqlik yuqori (Precision)	Yangi yoki kontekstual terminlar aniqlanmaydi
Qoidaga asoslangan (rule-based)	Morfologik va sintaktik shablonlar asosida	Grammatik tahlil aniq	Korpusdagi xilma-xillikka moslashishi qiyin

Yondashuv	Tavsif	Afzallik	Cheklov
Mashina o‘qitish (ML)	Belgilangan terminlardan o‘rganish (SVM, CRF)	Moslashuvchanlik yuqori	Katta belgilanadigan ma’lumot kerak
Chuqur o‘qitish (DL)	NER model (BERT, RoBERTa, UzBERT) asosida	Semantik aniqlik yuqori	Hisoblash resurslari ko‘p talab etiladi
Hybrid model	Lug‘at + ML integratsiyasi	Har ikki afzallikni birlashtiradi	Dastlabki sozlash murakkab

Tadqiqot doirasida **UzBERT-NER** modeli asosida harbiy terminlarni tanib olish uchun tajriba o‘tkazildi. Model 15 ming teglangan token asosida o‘qitilib, sinov korpusida quyidagi natijalarni ko‘rsatdi:

- Aniqlik (Precision) – 0.91
- Qamrov (Recall) – 0.87
- Umumiy F1 – 0.89

Bu natijalar o‘zbek tili uchun harbiy terminlarni avtomatik teglash imkoniyatining yuqori darajada ekanini ko‘rsatadi.

9. Natijalar tahlili

O‘tkazilgan tadqiqotlar quyidagi natijalarni berdi:

- 1. Termin bazasi yaratildi:** 7200 dan ortiq o‘zbekcha harbiy terminlar aniqlanib, lemmasi, manbasi, so‘z turkumi va teg turi bilan belgilandi.

2. **Tagset modeli ishlab chiqildi:** 9 asosiy teg (MIL_TERM, MIL_RANK, MIL_UNIT, MIL_EQUIP, MIL_ACTION, MIL_LOC, MIL_ORG, MIL_LEGAL, MIL_MISC) bo'yicha teglash amalga oshirildi.
3. **Annotatsiya tizimi sinovdan o'tkazildi:** 100 000 tokenli matnlar korpusida mashina va inson annotatsiyasi uyg'unligi 92% aniqlik bilan ta'minlandi.
4. **Tilshunoslik amaliyotida yangi metodika yaratildi:** O'zbek tilidagi terminologik qatlamlarni korpusda semantik jihatdan belgilashning tizimli yondashuvi ishlab chiqildi.
5. **Texnik integratsiya amalga oshirildi:** UzMorph morfologik analizatori bilan INCEpTION annotatsiya tizimi o'zaro integratsiya qilindi.

Bu natijalar nafaqat o'zbek tilining raqamli resurslarini boyitadi, balki xalqaro darajada o'zbek korpus lingvistikasi rivojiga hissa qo'shadi.

10. Xulosa

O'zbek tilidagi harbiy terminlarni Milliy til korpusida teglash masalasi filologik, lingvistik va texnologik jihatdan murakkab, ammo dolzarb yo'nalish hisoblanadi. Tadqiqot natijalaridan ko'rinadiki:

- Harbiy terminlar o'zbek tilining tarixiy, madaniy va ilmiy qatlamida mustahkam o'rin tutadi;
- Ularni semantik va morfologik jihatdan aniqlash, maxsus tagset asosida teglash — korpus lingvistikasida terminologik qatlamni chuqur o'rganish imkonini beradi;
- Mashina o'qitish va inson annotatsiyasini uyg'unlashtirgan hybrid model eng samarali yechim sifatida tavsiya etiladi;
- O'zbek tili uchun ishlab chiqilgan "MIL-Tagset" modeli til resurslarini boyitish va avtomatik til texnologiyalarini rivojlantirishda muhim ahamiyat kasb etadi.

Mazkur tadqiqot natijalari keyinchalik **sohaviy korpuslar** (harbiy, texnik, tibbiy)ni shakllantirishda, hamda o‘zbek tilini sun’iy intellekt tizimlariga moslashtirishda metodologik asos bo‘lib xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Abdullayeva, N. va boshq. (2023). *O‘zbek tilining Milliy korpusi: muammolar va yechimlar*. — Toshkent: TDSHU nashriyoti.
2. Abjalova, M. (2022). *Korpus birliklarini teglash nazariyasi*. — Toshkent: Fan.
3. Alimov, B. (2021). *O‘zbek terminologiyasi va xalqaro atamalar tizimi*. — Toshkent: Fan nashriyoti.
4. Qodirov, A. (2020). *O‘zbek tilida harbiy terminlarning shakllanishi va rivoji*. — O‘zbekiston FA Tilshunoslik instituti nashri.
5. Sharipov, M. (2023). *O‘zbek tilida morfologik teglash modellari*. — “Filologiya masalalari” jurnali, №3.
6. Jo‘rayev, S. (2019). *Harbiy terminlarning semantik differensiasiyasi*. — Toshkent: TDSHI.
7. Yusupov, T. (2018). *Sinonimik qatlam va harbiy terminlar*. — Toshkent: Fan.
8. Karimov, A. (2022). *O‘zbek tilida avtomatik lemmatizatsiya muammolari*. — TDYuU ilmiy to‘plami.
9. Qodirova, M. (2021). *Korpus lingvistikasining o‘zbek tilshunosligidagi istiqbollari*. — “Til va madaniyat”, №1.
10. Hasanov, D. (2024). *UzMorph morfologik analizatori texnik tavsifi*. — “Kompyuter lingvistika” jurnali, №2.
11. Umirov, O. (2021). *O‘zbek tilida tokenizatsiya va teglash texnologiyalari*. — Toshkent: IT Park.

12. Xo‘jayev, A. (2020). *Tilshunoslikda korpus va sun’iy intellekt*. — Toshkent: Innovatsiya nashriyoti.
13. Rasulov, B. (2022). *Korpus tahlilida terminologik birliklarni segmentatsiya qilish*. — “Filologiya masalalari”, №4.
14. Turaev, S. (2023). *Korpus lingvistikasida annotatsiya usullari*. — Toshkent: TDShU.
15. Xudoyberdiyev, R. (2021). *O‘zbek tili va texnik terminlar*. — Toshkent: Texnika nashriyoti.