

e-ISSN 3060-5059

DOI: 10.67227



IJTIMOYIY-GUMANITAR SOHADA ILMIY-INNOVATSION TADQIQOTLAR

ILMIY METODIK JURNAL

Scientific and Innovative Research in
Social and Humanitarian Studies

Vol. 3 • Issue 8

2026



Founded in 2024



publishscience.uz



publish_science_journal@mail.ru



PEER REVIEWED • OPEN ACCESS

SO‘Z TURKUMLARINING KORPUS TADQIQOTIDAGI OPERATSIONAL TASNIFI: MILLIY AN‘ANA VA UNIVERSAL STANDART UYG‘UNLIGI

Abdullayev Ikramjon Xashimdjanoich

Andijon davlat chet tillari instituti, tayanch doktorant.

e-mail: annora1976@gmail.com ORCID: <https://orcid.org/0009-0007-9353-1036>

Annotatsiya.

Maqolada o‘zbek tili so‘z turkumlarining milliy tasnifi universal teg standarti bilan qay darajada mos kelishi 12 048 tokenli qo‘lda teglangan mikrokorpus materialida o‘lchandi. Ikki qatlam 1 480 tokenida ajralishi va nomuvofiqlik beshta turkumda to‘planishi aniqlandi.

Kalit so‘zlar: so‘z turkumlari, operatsional tasnif, korpus lingvistikasi, teglash, universal teg standarti, milliy tasnif, mikrokorpus, atoqli ot, modal so‘z.

ОПЕРАЦИОНАЛЬНАЯ КЛАССИФИКАЦИЯ ЧАСТЕЙ РЕЧИ В КОРПУСНОМ ИССЛЕДОВАНИИ: СОГЛАСОВАНИЕ НАЦИОНАЛЬНОЙ ТРАДИЦИИ И УНИВЕРСАЛЬНОГО СТАНДАРТА

Абдуллаев Икрамжон Хашимджанович

Андижанский государственный институт иностранных языков, базовый докторант.

Аннотация.

В статье на материале размеченного вручную микрокорпуса объемом 12 048 токенов измерена степень соответствия национальной классификации частей речи узбекского языка универсальному стандарту тегов. Установлено, что два слоя расходятся в 1 480 токенах, а расхождения сосредоточены в пяти категориях.

Ключевые слова: части речи, операциональная классификация, корпусная лингвистика, разметка, универсальный стандарт тегов, национальная классификация, микрокорпус, имя собственное, модальное слово.

OPERATIONAL CLASSIFICATION OF PARTS OF SPEECH IN CORPUS RESEARCH: LIGNING THE NATIONAL TRADITION WITH THE UNIVERSAL STANDARD

Abdullayev Ikramjon Xashimdjanoich

Andijan State Institute of Foreign Languages, Basic Doctoral Student.

Abstract.

The article measures how far the national classification of Uzbek parts of speech corresponds to the universal tagset, drawing on a manually annotated microcorpus of 12,048 tokens. The two layers diverge in 1,480 tokens, and the divergence is concentrated in five categories.

Keywords: parts of speech, operational classification, corpus linguistics, tagging, universal tagset, national classification, microcorpus, proper noun, modal word.

Kirish. So‘z turkumlarini ajratishda semantik, morfologik va sintaktik mezonlar har doim ham bir-biriga mos kelavermaydi. Nazariy tavsifda bu ziddiyat izohlanadi va shu bilan cheklanadi. Korpus tadqiqotida esa undan qochib bo‘lmaydi: annotator har bir tokeniga bitta teg qo‘yishi shart, tokenlar soni esa o‘n minglab bo‘ladi. Shu bois teglashga mo‘ljallangan tasnif nazariy tavsifdan farqli talabga javob berishi, ya‘ni har bir holatda bir qiymatli va qayta tekshiriladigan qarorni ta‘minlashi lozim. Ushbu tadqiqotda bunday tasnif **operatsional tasnif** deb ataladi.

O‘zbek tili uchun bu masala ikki yo‘nalishda talqin qilinadi. Bir tomondan, milliy an‘anada o‘n ikki turkumli tasnif shakllangan bo‘lib, u maktab va oliy ta‘lim amaliyotida mustahkam o‘rin egallagan. Ikkinchi tomondan, korpus resurslari xalqaro qiyoslanuvchanlikni talab qiladi va bu Universal Dependencies loyihasining teglar to‘plamiga tayanishni taqozo etadi [11]. Ikki tizim bitta korpusda qo‘llanilganda quyidagi savol yuzaga keladi: ular bir-birini takrorlaydimi yoki har biri alohida axborot beradimi?

Mazkur masalaning amaliy jihati ham mavjud. Davlat tilining nufuzini oshirishga qaratilgan normativ hujjatlarda o‘zbek tilining raqamli resurslarini yaratish ustuvor vazifalardan biri sifatida belgilangan [1], sun‘iy intellekt texnologiyalarini rivojlantirish strategiyasida esa til texnologiyalari uchun sifatli ma‘lumotlar bazasini yaratish zarurligi alohida ta‘kidlangan [2]. Teglangan korpus ana shunday bazaning eng muhim tarkibiy qismlaridan biridir. Undagi teglar tizimining qanday qurilgani esa keyinchalik ushbu korpusga tayanuvchi morfologik analizatorlar, elektron lug‘atlar va boshqa lingvistik texnologiyalarning imkoniyatlarini belgilaydi.

Mazkur maqolaning maqsadi ushbu savolga miqdoriy tahlil asosida javob berishdan iborat. Aniqrog‘i, milliy va universal teg qatlamlarining qaysi hajmda va qaysi turkumlarda ajralishi o‘lchab chiqildi. Olingan natijalar amaliy ahamiyatga ega: agar ikki qatlam deyarli mos kelsa, ikkinchi qatlamni yuritish ortiqcha mehnat bo‘ladi; aksincha, ular

tizimli ravishda farqlansa, ikki qatlamli teglash zaruratga aylanadi.

Adabiyotlar tahlili va metodologiya.

O'zbek tilshunosligida so'z turkumlari tasnifi uch mezonning uyg'unligiga tayanadi. Sh. Shoabdurahmonov va hammualliflari turkumni ma'no, shakl va vazifa birligida ta'riflaydi [3; 12–14-b.]. A. Nurmonov va hammualliflarining nazariy grammatikasida bu uyg'unlik tizim tushunchasi orqali aniqlashtiriladi: turkum yakka birlik emas, balki munosabatlar to'plami sifatida qaraladi [4; 9–11-b.]. A. Nurmonovning tilshunoslik tarixiga bag'ishlangan ishida bu tasnifning shakllanish bosqichlari va uning arab hamda rus grammatika an'analari bilan aloqasi ko'rsatilgan [5; 87–90-b.]. Z. Isaqov esa turkumlarni belgilarni ifodalash darajasiga ko'ra o'zaro ziddiyatda turuvchi guruhlar deb talqin qiladi [6; 16-b.] va shu bilan chegaraviy holatlar muammosini ilgari suradi. J. Eltarovning turkumlararo ko'chishga bag'ishlangan tadqiqoti bu muammoni alohida obyektga aylantirgan [7; 11-b.].

Sanab o'tilgan ishlarning umumiy jihati shundaki, ularda tasnif tavsif vositasi sifatida quriladi. Teglash talabi esa boshqacha: unda tasnif qaror qabul qilish tartibiga aylanishi kerak. Farqni bir misolda ko'rish mumkin. Nazariy tavsif "harakat nomi fe'l va ot xususiyatlarini birlashtiradi" deb yozishi mumkin va bu to'g'ri bo'ladi. Annotator esa shu birlikka NOUN yoki VERB tegidan bittasini qo'yishga majbur. Demak, tavsif darajasidagi to'g'ri javob teglash darajasida yetarli emas.

Xalqaro amaliyotda bu vazifa ikki qatlamli teg tizimi orqali hal etiladi. Universal Dependencies standartida har bir token uchun universal teg (UPOS) va tilga xos teg (XPOS) alohida maydonlarda saqlanadi [11; 258–260-p.]. Standart mualliflari universal qatlamni tillararo qiyoslash uchun, tilga xos qatlamni esa milliy an'anani saqlash uchun mo'ljallanganini ta'kidlaydi. Nazariy jihatdan bu yechim mos ko'rinadi, ammo uning o'zbek materialidagi natijasi hozirgacha o'lchanmagan.

O'zbek tili uchun korpus resurslari so'nggi yillarda faollashdi. N. Abduraxmonovning qo'llanmasida korpus tuzilishi va annotatsiya qatlamlari tavsiflangan [8; 41–45-b.]. B. Elov va hammualliflari milliy korpusning maqsad va vazifalarini bayon qilgan [9; 27-b.]. M. Sharipov va hammualliflari ko'p sohali teglangan tanlanmani e'lon qildi [12], L. Bobojonova va hammualliflari esa neyron model asosidagi teglagichni taqdim etdi [13]. Sh. Zarmasovning ot turkumini modellashtirishga bag'ishlangan tadqiqotida turkumni rasmiy model shaklida berish tajribasi keltirilgan [10; 21–23-b.]; unda modellashtirish bitta turkum doirasida olib borilgan va ikki teg qatlamli masalasi ko'tarilmagan. Bu ishlar teglash amaliyotini rivojlantirdi, biroq ularda milliy tasnif bilan universal to'plam o'rtasidagi moslik alohida o'lchov predmeti sifatida qaralmagan. Mazkur maqola shu bo'shliqni to'ldirishga qaratilgan.

Tadqiqot materiali muvozanatlashgan mikrokorpusdir. Unga uch funksional uslubdan — badiiy, ilmiy va publitsistik — o'n manbadan taxminan teng hajmda parcha olingan: mos ravishda 4 006, 4 004 va 4 038 token, jami 12 048 token. Har bir manbadan asar yoki maqola boshidan uzluksiz parcha olindi, chunki tasodifiy gaplar to'plami qo'shni tokenlar munosabatini yo'qotadi.

Teglash tartibi ikki bosqichli: avtomatik dastlabki teglash va undan keyin har bir tokenni qo'lda tekshirish. Har bir token uchun milliy qatlam (XPOS) va universal qatlam (UPOS) mustaqil to'ldirildi. Ikki qatlam bir-biridan avtomatik hosil qilinmadi — bu muhim shart, chunki hosil qilingan qatlam farqni ko'rsata olmasdi. Annotatsiya ishonchligi ikki bosqichda o'lchandi va turkum tegi bo'yicha Koen kappa koeffitsiyenti 0,850 ni tashkil etdi.

O'lchov usuli oddiy: har bir token uchun XPOS va UPOS qiymatlari qiyoslanadi va farq qiladigan tokenlar sanaladi. Farq milliy turkumlar kesimida guruhlanadi, so'ng har bir guruhning tabiati kontekstual tahlil bilan aniqlanadi.

Izoh. Maqolani tayyorlash jarayonida matnni tahrirlash va kvantitativ hisob-kitoblarni qayta tekshirishda sun'iy intellekt vositalaridan foydalanildi. Tadqiqot materiali, teglash qarorlari, tahlil va xulosalar muallifga, ya'ni menga tegishli.

Muhokama.

Ikki qatlamli teglashning zarurati haqidagi bahs, bizningcha, ikki xil tasnif tamoyilining to'qnashuvidan kelib chiqadi. Milliy an'ana turkumni birinchi navbatda shakl paradigmasi va lug'aviy ma'noga tayantiradi: ot son, egalik va kelishik ko'rsatkichlarini oladi, fe'l tuslanadi. Universal standart esa sintaktik xulqni birlamchi deb hisoblaydi va shu bois bir milliy turkum ichidagi sintaktik jihatdan boshqacha yo'l tutadigan birliklarni ajratadi. Ikki tamoyil bir-birini inkor etmaydi, lekin ular bir xil natijani ham bermaydi.

Bu farqning nazariy asoslari xalqaro adabiyotda ancha oldin muhokama qilingan. R. Kvirik va hammualliflari turkumlarni ochiq va yopiq sinflarga ajratib, yopiq sinf a'zosining qiymati faqat shu sinfdagi qolgan a'zolarga nisbatan aniqlanishini ko'rsatgan edi [14; 67-p.]. V. Kroft esa bundan uzoqroqqa borib, kategoriyalarni tilga xos qurilmalar sifatida talqin qiladi va universal ro'yxatning o'zini shubha ostiga qo'yadi [15; 63–65-p.]. Ikki qarash orasidagi tanlov amaliy oqibatga ega: R. Kvirik yo'nalishi universal to'plamni istalgan tilga tatbiq etishga ruxsat bersa, V. Kroft yondashuvi har bir tilda o'z tasnifini talab qiladi.

Mazkur tadqiqotda ushbu bahsga qat'iy javob berilmaydi, chunki bitta til materiali tipologik xulosa uchun yetarli emas. Buning o'rniga biz masalani o'lchov darajasiga o'tkazdik: agar V. Kroft haq bo'lsa, milliy va universal qatlam korpusda tizimli ravishda ajralishi kerak; agar R. Kvirik talqini kuchliroq bo'lsa, ajralish tasodifiy va kam sonli bo'lib chiqadi. Shu tarzda nazariy bahs tekshiriladigan shaklga keltiriladi.

Ikki qatlamni saqlashning muqobil yo'li ham bor edi va uni ko'rib chiqish o'rinli. Birinchi muqobil — faqat universal qatlam bilan cheklanish. Bu yo'l teglashni soddalashtiradi, lekin o'zbek an'anasidagi modal so'z va taqlid so'z turkumlari yo'qoladi: universal to'plamda ular uchun alohida teg yo'q. Ikkinchi muqobil — faqat milliy qatlam. Unda korpus o'zbek tilshunosligi doirasida yopiladi va boshqa tillar korpuslari bilan qiyoslanmaydi. Uchinchi muqobil — bir qatlamni ikkinchisidan avtomatik hosil qilish. Bu yechim tejamkor ko'rinadi, ammo u farqni yashiradi, holbuki tadqiqotimizning predmeti aynan farqning o'zi.

Shu sabablarga ko'ra, ushbu tadqiqotda ikki qatlam mustaqil to'ldiriladigan parallel maydon sifatida qabul qilindi. Bunday qaror mehnat hajmini oshiradi, buning evaziga esa qatlamlararo munosabatni o'lchash imkonini beradi.

Natijalar.

Korpusning barcha tokenlari bo'yicha ikki qatlam qiyoslanganda milliy va universal teg 1 480 tokenda mos kelmadi. Bu — 12 048 tokenning 12,3 foizi, ya'ni har sakkizinchi token. Mazkur qiymat turkumlar kesimidagi farqlarning yig'indisi sifatida hisoblangan; uning tarkibi 1-jadvalda keltiriladi.

1-jadval. Milliy turkumlarning universal teglarga proyeksiyasi

Milliy turkum (XPOS)	Universal teg (UPOS)	Token	Farq
Ot (NOUN)	NOUN 5 151 · PROPN 606	5 757	606
Fe'l (VERB)	VERB 2 627 · AUX 54	2 681	54
Olmosh (PRON)	PRON 383 · DET 245	628	245
Bog'lovchi (CONJ)	CCONJ 433 · SCONJ 17	450	450
Modal so'z (MOD)	AUX 71 · PART 51	122	122
Taqlid so'z (IMIT)	INTJ 3	3	3
Sifat, ravish, son, ko'makchi, yuklama, undov	o'zgarishsiz	2 407	0
Jami		12 048	1 480

Jadval nomuvofiqlikning tasodifiy tarqalmaganini ko'rsatadi. Sakkiz turkum universal qatlamga o'zgarishsiz o'tadi va ular birgalikda 2 407 tokenni qamrab oladi. Farqning barchasi olti turkumda to'plangan.

Farqning tabiati esa bir xil emas va uni ikki turga ajratish mumkin. Birinchi tur — nomning almashishi. Bunda milliy turkumning universal to'plamda bir xil nomli muqobili umuman yo'q, shu bois turkumning barcha tokenlari farqli hisoblanadi. Bog'lovchi, modal so'z va taqlid so'z shu toifaga kiradi; ular birgalikda 575 tokenni tashkil etadi. Ikkinchi tur — ichki bo'linish. Bunda turkum, asosan, o'zini takrorlaydi, lekin uning ma'lum qismi boshqa tegga ajraladi: ot, fe'l va olmoshda shunday holat kuzatiladi va bu qism 905 tokenni tashkil etadi.

Ikki turning farqi metodologik jihatdan muhim. Nomning almashishi tarjima masalasi bo'lib, uni jadval orqali bir marta hal qilish mumkin. Ichki bo'linish esa har bir tokenda alohida qaror talab qiladi, chunki bir xil lemma bir kontekstda bir tegni, boshqa kontekstda esa ikkinchi tegni oladi.

Ichki bo'linishning eng katta ulushi otga to'g'ri keladi. Milliy tasnifda atoqli ot mustaqil turkum emas, otning ma'no turi hisoblanadi; universal to'plamda esa u alohida teg (PROP) sifatida ajratiladi. Korpusda 606 token shu holatga to'g'ri keladi:

(1) ...smetasi yeb bitirilmagan boshqa obyektga o'tadi-ketadi. Toshkentda nima ko'p — qurilish ko'p... (B1_Hoshimov, ID 112).

Bu yerda Toshkentda birligi milliy qatlamda NOUN, universal qatlamda esa PROP tegini oladi. Grammatik jihatdan u oddiy otdan farq qilmaydi: o'rin-payt kelishigi qo'shimchasini olgan va hol vazifasida kelgan. Farq faqat referensial xususiyatga ega: birlik yagona obyektni nomlaydi. Agar uni turdosh ot bilan almashtirsak, masalan, *shaharda*, gapning grammatik tuzilishi o'zgarmaydi, ammo atoqlilik yo'qoladi. Demak, bu bo'linish grammatik emas, referensial mezonga tayanadi va shu bois u milliy qatlamda emas, universal qatlamda qayd etiladi.

Olmoshdagi bo'linish boshqacha asosga ega. Universal standart o'rin bosuvchi olmoshni (PRON) aniqlovchi vazifasidagi olmoshdan (DET) ajratadi; milliy tasnifda esa ikkalasi ham olmosh sanaladi. Korpusda bunday holat 245 marta uchraydi:

(2) ...naryadni o'rtacha o'n so'mdan kam yopmasdim. Bu safar hammasiga o'sha A-1 armaturali panel sabab bo'ldi... (B1_Hoshimov, ID 153).

Bu birligi bu yerda mustaqil turmaydi — u safar otiga bog'lanib, uni aniqlaydi. Shuning uchun universal qatlamda DET tegini oladi. Agar shu birlik boshqa kontekstda o'rin bosuvchi sifatida kelsa (*Bu meni qiziqtirmaydi*), universal teg PRONGa o'zgaradi, milliy teg esa har ikki holatda ham o'zgarishsiz qoladi. Ya'ni bu yerda bo'linish kontekstga bog'liq va uni lemma darajasida hal qilib bo'lmaydi.

Modal so'z esa nomning almashishiga misol bo'ladi va u eng ko'p savol tug'diradigan holatdir. Milliy an'anada modal so'z yaxlit turkum, universal to'plamda esa u ikkiga bo'linadi:

(3) ...qimirlasa terisi shilini chiqadi. Unga bitta narsa kerak: qay yo'l bilan bo'lmasin, planni bajarish... (B1_Hoshimov, ID 271).

(4) ...chiqishingiz bilan orqangizdan yomonlaydi. Xullas, uch kun avval Moskvichida kelib qoldi... (B1_Hoshimov, ID 303).

Uchinchi misoldagi *kerak* predikativ vazifada keladi va universal qatlamda AUX tegini oladi; to'rtinchi misoldagi *xullas* esa kirish bo'lak bo'lib, PART tegiga tushadi. Milliy qatlamda ikkalasi ham modal so'z. Bizning kuzatishimizcha, aynan shu turkum ikki tasnif tamoyilining farqini eng aniq ko'rsatadi: milliy an'ana birlikning modallik ma'nosiga qaraydi, universal standart esa uning gapdagi sintaktik o'rniga. 122 tokenning 71 tasi birinchi, 51 tasi ikkinchi guruhga tushdi, ya'ni turkum ichida ikki xil sintaktik xulq mavjud va u milliy qatlamda ko'rinmaydi.

Bog'lovchi turkumi nomning almashishiga yana bir misol beradi, ammo undagi taqsimot notekis. Universal standart teng bog'lovchi (CCONJ) bilan ergashtiruvchi bog'lovchini (SCONJ) ajratadi; korpusda 450 tokenning 433 tasi birinchi, atigi 17 tasi ikkinchi guruhga tushdi:

(5) ...Mulla aka, yaxshi keldingizmi, deb qo'limga suv quyadi. Agar ikkinchi smenada bo'lmasa... (B1_Hoshimov, ID 303).

1293).

Agar birligi ergash gapni bosh gapga bog'laydi va shu bois SCONJ tegini oladi. Ergashtiruvchi bog'lovchining bunchalik kam uchrashi o'zbek tilining tipologik xususiyati bilan izohlanadi: ergashish munosabati, asosan, ravishdosh va sifatdosh shakllari, ya'ni fe'lining tuslanmaydigan shakllari orqali ifodalanadi va alohida bog'lovchi so'zga ehtiyoj kamayadi. Ushbu kuzatuv milliy qatlamda ko'rinmaydi, chunki unda ikki tur bir turkum ostida birlashadi.

Taqlid so'z esa eng kichik guruh — korpusda uch token. Milliy an'anada u mustaqil turkum, universal to'plamda esa undov (INTJ) tegiga qo'shiladi. Bu holatda son statistik xulosa uchun yetarli emas, ammo teglash qarori jihatidan u boshqa turkumlar bilan bir xil mexanizmga ega: milliy qatlam turkumini saqlaydi, universal qatlam esa uni yaqin tegga birlashtiradi.

Fe'ldagi bo'linish miqdoran eng kichik, ammo u bir xil mexanizmga ega:

(6) ...A-2 armaturali panelni yetkazib berganida olam guliston edi. Tag'in do'q urishiga o'laymi... (B1_Hoshimov, ID 193).

Edi bog'lamasi milliy qatlamda fe'l, universal qatlamda yordamchi fe'l (AUX) sanaladi. 2 681 fe'l tokenidan atigi 54 tasi shunday, ya'ni fe'l turkumi ikki qatlamda deyarli bir xil ko'rinadi.

Nomuvofiqlikning uslublar bo'yicha taqsimoti ham bir xil emas: publitsistik matnda 650, ilmiy matnda 425, badiiy matnda 405 token. Farqning publitsistik uslubda ko'proq to'planishi tasodif emas — bu uslubda tashkilot, joy va sana nomlari zich keladi, ular esa universal qatlamda PROPIN tegini oladi.

Xulosa. Milliy va universal teg qatlamlari o'zbek korpusida 1 480 tokenida, ya'ni materialning 12,3 foizida ajraladi. Bu miqdor ikki qatlamni saqlash zaruratini ko'rsatadi: agar farq bir-ikki foizni tashkil etganda, ikkinchi qatlamni yuritish qo'shimcha mehnat bo'lar edi. Nomuvofiqlik tasodifiy emas va u ikki mexanizmga bo'linadi. Bog'lovchi, modal so'z va taqlid so'zda milliy turkumning universal to'plamda muqobili yo'q — bunda 575 token qamrab olinadi va masala moslik jadvali bilan hal etiladi. Otda, olmoshda va fe'lda esa turkum ichida bo'linish yuz beradi va 905 token har bir holatda alohida qaror talab qiladi. Bizningcha, bu natijaning asosiy ma'nosi shundaki, farqning o'zi ma'lumot beradi. U milliy tasnif va universal standart bir xil narsani turlicha nomlashini emas, balki ular boshqa-boshqa mezonlarga tayanishini ko'rsatadi: birinchisi shakl paradigmasi va lug'aviy ma'noga, ikkinchisi esa sintaktik xulqqa. Modal so'z misolida bu ayniqsa aniq ko'rinadi — yaxlit milliy turkum universal qatlamda ikki sintaktik guruhga ajraladi. Mazkur yondashuvning cheklangan tomoni shundaki, o'lchov bitta mikrokorporus materialida olib borildi va tanlanma hajmi 12 048 token bilan chegaralangan. Kichik turkumlar, jumladan, taqlid so'z uch token bilan qatnashgani bois, ular bo'yicha xulosa faqat tendensiya darajasida qabul qilinishi mumkin. Korpus kengaytirilganda nisbatlar aniqlashadi, ammo bo'linishning tuzilmaviy tabiati saqlanishi kutiladi, chunki u tanlanma hajmiga emas, ikki tasnifning mezonlariga bog'liq.

FOYDALANILGAN ADABIYOTLAR

1. O'zbekiston Respublikasi Prezidentining ***O'zbek tilining davlat tili sifatidagi nufuzi va mavqeyini tubdan oshirish chora-tadbirlari to'g'risida***gi 2019-yil 21-oktabrdagi PF-5850-son Farmoni // Qonun hujjatlari ma'lumotlari milliy bazasi.
2. O'zbekiston Respublikasi Prezidentining ***Sun'iy intellekt texnologiyalarini 2030-yilga qadar rivojlantirish strategiyasini tasdiqlash to'g'risida***gi 2024-yil 14-oktabrdagi PQ-358-son qarori // Qonun hujjatlari ma'lumotlari milliy bazasi.
3. Shoabdurahmonov Sh., Asqarova M., Hojiyev A., Rasulov I., Doniyorov H. Hozirgi o'zbek adabiy tili. I qism. — Toshkent: O'qituvchi, 1980. — 446 b.
4. Nurmonov A., Shahobiddinova Sh., Iskandarova Sh., Nabiyeva D. O'zbek tilining nazariy grammatikasi. Morfologiya. — Toshkent: Yangi asr avlodi, 2001. — 80 b.
5. Nurmonov A. O'zbek tilshunosligi tarixi. — Toshkent: O'zbekiston, 2002. — 232 b.
6. Isaqov Z. O'zbek tilida so'z turkumlarining o'zaro munosabati va unda yuklamalarning o'rni: monografiya. — Qo'qon, 2011. — 83 b.
7. El'tazarov J. D. O'zbek tilida so'z turkumlarining o'zaro aloqasi va ko'chishi: Filol. fan. dokt. ... diss. — Samarqand, 2005. — 284 b.
8. Abduraxmonova N. Z. Korpus lingvistikasi. — Toshkent, 2023. — 357 b.
9. Elov B., Primova M., Amirkulov M. O'zbek tili korpusi: tarixiy rivojlanishi, maqsad va vazifalari // Kompyuter lingvistikasi: muammolar, yechim, istiqbollar: V xalqaro ilmiy-amaliy konferensiya materiallari. — Toshkent, 2025.
10. Zarmasov Sh. Sh. O'zbek tili so'z turkumlarini modellashirish tamoyillari (ot so'z turkumi misolida): Filol. fan. bo'yicha fals. dokt. (PhD) diss. avtoref. — Qo'qon, 2025. — 58 b.
11. de Marneffe M.-C., Manning C. D., Nivre J., Zeman D. Universal Dependencies // Computational Linguistics. — 2021. — Vol. 47, No. 2. — P. 255–308.
12. Sharipov M., Kuriyozov E., Vičič J. UzbekPOS: A Multi-domain Dataset for Uzbek Part-of-Speech Tagging // Data in Brief. — 2026. — Vol. 66. — Art. 112640.
13. Bobojonova L., Akhundjanova A., Ostheimer P. S., Fellenz S. BBPOS: BERT-based Part-of-Speech Tagging for Uzbek // Proceedings of the First Workshop on Language Models for Low-Resource Languages. — Abu Dhabi: Association for Computational Linguistics, 2025.
14. Quirk R., Greenbaum S., Leech G., Svartvik J. A Comprehensive Grammar of the English Language. — London–New York: Longman, 1985. — 1779 p.
15. Croft W. Radical Construction Grammar: Syntactic Theory in Typological Perspective. — Oxford: Oxford University Press, 2001. — 416 p.