

O‘ZBEK TILI KORPUSINING DASTURIY TA’MINOTI IMKONIYATLARI

Abjalova Manzura Abdureshetovna

abjalova@navoiy-uni.uz

ToshDO‘TAU Kompyuter lingvistikasi
va raqamli texnologiyalar kafedrası dots.v.b.,
filologiya fanlari bo‘yicha PhD

Annotatsiya. Til korpuslari til ta’limi – lingvodidaktika, lug‘atlarni yaratish, shuningdek, turli ilmiy tadqiqotlarni amalga oshirish, tilni diaxron va sinxron aspektda o‘rganish, nutqiy kompetensiyani rivojlantirish, so‘z boyligini oshirish, gap modellarini shakllantirishda o‘ta muhim ish quroli hisoblanadi. Shu kabi mislsiz imkoniyatlarni berishga mo‘ljallangan til korpuslarini yaratish dolzarb masala hisoblanadi. Mazkur maqolada o‘zbek tilining ta’limiy korpusining ayni damdagi imkoniyatlari haqida so‘z yuritildi. Shuningdek, o‘zbek tilining ta’limiy korpusini yaratish omillari, ma’lumotlar bazasi, axborot tizimi va imkoniyatlari ko‘rsatildi.

Kalit so‘zlar: *ta’limiy korpus, leksikografiya, chastotali lug‘at, terminologik lug‘at, ma’lumotlar bazasi, model, axborot tizimi.*

Abstract. The linguistic corpus is an important tool for language education linguodidactics, the creation of dictionaries, as well as various studies, diachronic and synchronous language learning, the development of speech competence, vocabulary, speech patterns.

The article examines the factors, principles, databases, information systems and opportunities for creating an educational corpus of the Uzbek language.

Keywords: *educational corpus, lexicography, frequency dictionary, terminological dictionary, database, model, information system.*

Bugun kunda ko‘plab til korpuslari mavjud bo‘lib, turli maqsadlar doirasida yaratilganligi bois til korpuslarining maxsus turlari va turli imkoniyatlarini namoyon

etadi. Til korpuslari bo'yicha tadqiqotlar qamrovi keng bo'lib, ular, asosan, turli maqsaddagi til korpuslarining lingvistik va dasturiy ta'minotlarini yaratishga, mavjud korpuslarni rivojlantirishga hissa qo'shadi.

Til korpuslari, ayniqsa, zamonaviy elektron lug'atlar tuzish va undan foydalanish madaniyatini shakllantirish til imkoniyatini egallashda samarador ekanligi o'z isbotini topgan. Aynan til ta'limi va elektron lug'atlarni yaratishda dunyo miqiyosida yaratilishiga ehtiyoj nihoyatda yuqori bo'lgan til korpuslarining o'rni beqiyos.

Ma'lumki, kompyuter texnologiyalarining korpus tilshunosligida qo'llanilishi natijasida til korpuslari asosida so'zning qo'llanish davri va chastotasini aniqlash, terminologiya sohasini rivojlantirish, gap qurilishini o'rganish va tahlil qilish, tarjima dasturlari uchun n-tilli baza yaratish, uslubiyatni o'rganish, elektron lingvodidaktikani rivojlantirish, ayniqsa, turli lug'atlarni yaratish imkoniyatlarining kartotekadan avtomat jarayoniga o'tildi va bemiyl qulayliklarni yuzaga keltirdi [1]. Lug'atlarni tuzish masalasida til korpuslarining foydaliligi, ayniqsa, chastotali lug'atlarni yaratish, muayyan soha terminologiyasini shakllantirishda muhim sanaladi. Shuningdek, ikki yoki ko'p tilli parallel korpuslarning tarjima bazasi yordamida tarjimon lug'atlari [Boliang Zhang, 2020], avtomatik tarjima dasturlarining lingvistik bazalarini shakllantirishda yirik resurs vazifasini bajaradi.

Umuman, ko'p maqsadli va polifunksional til korpusi – muayyan tilning Milliy korpusi hisoblanadi.

Milliy korpus quyidagi talablarga javob berishi zarur:

- 1) tilning barcha uslublariga mansub o'ta katta hajmdagi matnlarning bo'lishi;
- 2) matnlarning lingvistik, morfologik va sintaktik jihatdan teglanishi;



3) korpusning meta-ma'lumotlari mavjud bo'lishi zaruriy talablardan hisoblanadi.

Mazkur talablar Milliy korpusning zaruriy lingvistik va lingvodidaktik texnologik vositaga aylanishini ta'minlaydi. Ta'limiy korpus Milliy korpusning xususiy turi hisoblanib, asosan, ta'lim jarayonida til o'rgatish [Nina Vyatkina, 2020], ta'lim jarayonida turli amaliy mashg'ulotlarni bajarish, tadqiqot jarayonida matn elementlaridan foydalanish, ta'limga oid fan sohalari terminologiyasini shakllantirish, ta'lim tezaurusini shakllantirish kabi masalalarda asosiy vosita hisoblanadi.

Ta'limiy korpus – maktab darsliklari va o'quv lug'atlariga asoslangan til korpusi bo'lib, o'zbek tili Milliy korpusining xususiy (tarkibiy) turi hisoblanadi. O'zbek tilining ta'limiy korpusi AM-FZ-201908172 raqamli “O'zbek tili ta'limiy korpusini yaratish” mavzusidagi amaliy loyiha doirasida yaratildi, 2021-yilning 23-aprelida dastlabki taqdimoti o'tkazildi, hozirda uning bazasi boyitililmoqda va takomillashtirilmoqda. O'zbek tilining ta'limiy korpusi 5 rukndan iborat:

1) Bosh sahifa; 2) Ta'limiy korpus (haqida ma'lumotlar); 3) Izlash; 4) Lug'atlar; 5) Kirish.

To'rtinchi – “*Lug'atlar*” rukni o'zbek tili ta'limiy korpusining leksikografik bazasi hisoblanib, SQL-SERVER2019 ma'lumotlar bazasi boshqaruv tizimidan foydalanilgan holda shakllantirilgan, ayni damda bazada sakkiz turdagi lingvistik lug'at birlashtirilgan (1-rasm).

Sayt test rejimida ishlamoqda!



Bosh sahifa

Ta'limiy korpus

Izlash

Lug'atlar

Kirish

So'zni kiriting

Izlash

- ☒ Izoh(lar)i
- ☒ Omonim(lar)i
- ☒ Sinonim(lar)i
- ☒ Antonim(lar)i
- ☒ Paronimi
- ☒ Qomus
- ☒ Ibora(lar)
- ☒ Darajalanish qatori
- ☐ Uslubiy xoslanishi

1-rasm. O'zbek tili ta'limiy korpusining lug'atlar rukni (8 ta lug'at)

“Lug'atlar” ruknida izohli lug'at, omonim, sinonim, paronim, antonim, iboralar va darajalanish lug'atlari bazalariga murojaat yordamida foydalanuvchiga lingvistik ma'lumotlar chiqarib beriladi.

Ta'limiy korpus integrallashgan veb-dastur tarzda shakllantirilib, foydalanuvchilarga ish jarayonida qulaylik yaratish, faoliyat jarayonlarini samarali amalga oshirish, vaqtdan unumli foydalanish, bir vaqtda muayyan leksik birlik bo'yicha turli ma'lumotlarni olish, leksik birlikning qo'llanilish ko'lamini aniqlash, misollar massivini topish, kontekstual ma'no qirralarini aniqlash kabi dolzarb masalalar yuzasidan keng ko'lamli imkoniyatlarni beradi.

Ta'limiy korpus foydalanuvchilari veb brauzer yordamida internet yoki lokal tarmoq orqali murojaat qiladi. Bunday holatda xavfsizlikni ta'minlash uchun veb brauzerdan veb servergacha kelgan trafikning maxfiyligini ta'minlash, foydalanuvchilarni autentifikatsiya qilish, veb serverni autentifikatsiyalashni ta'minlash kerak. Trafik himoyasini ta'minlash va veb serverni autentifikatsiya qilish uchun SSL sertifikatlardan foydalaniladi. SSL sertifikatlar foydalanuvchi brauzeri va server orasida ma'lumotlarni shifrlangan holda almashishni ta'minlaydi.

SSL sertifikatlar Registratsiya markazlaridan (VeriSign, Inc.; The Go Daddy Group, Inc.; Thawte va boshq.) [Zhao Shengxue, 2020] sotib olinishi mumkin. Internet yoki boshqa tarmoqlardan amalga oshirilayotgan hujumlarni aniqlash va bartaraf etish uchun IDS/IPS tizimlari, kirish va chiqish trafiginı filtrlash uchun tarmoqlararo ekrandan foydalaniladi.

Ta’limiy korpus imkoniyatlari. Korpus bo’yicha qidiruv foydalanuvchiga quyidagilarni aniqlash imkonini beradi:

- 1) muayyan soʻzning barcha shakllarini misollar massivi bilan aniqlash;

Korpus

O'zbek tilining ta'limiy korpusi

So'z yoki so'z birikmasi

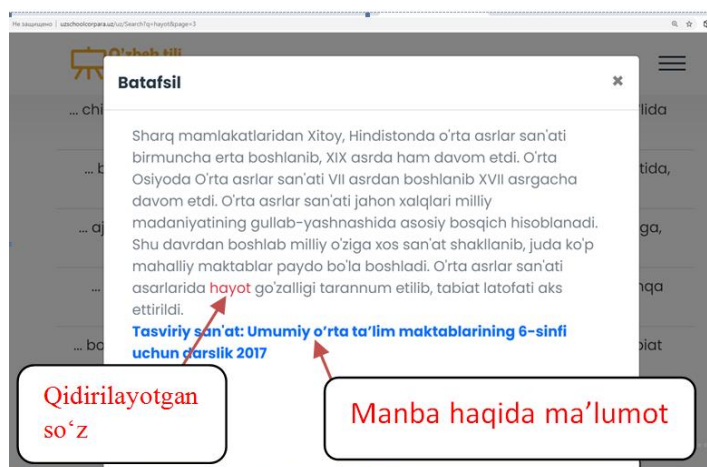
osmon

☐ Parametrlar
 ☐ CQL
 Uslubiy xoslanishi: uslubi betaraf.
 [Bigramma](#)

Izlash

2-rasm. “osmon” soʻzi qatnashgan misollar massivi

- 2) misollarning manbasini aniqlash;



3-rasm. “hayot” soʻzi ishtirok etgan abzats va uning manbasi nomi

3) manba haqida metamaʼlumot olish; Masalan:

Oʻzbek tili
711 milliy korpusi

Bosh sahifa Taʼlimiy korpus Izlash Lugʻatlar Kirish

osmonga qarab oʻtirar ekan. turdajon tika tuta qotai. toknadaagi deshinchi chiroq paxilab yonar, uning atrofida katta bir parvona aylanar ekan.

Oldingi Keyingi

Manba:

Nomi	Anor
Muallif(lar)i	Abdulla Qahhor 1907
Yaratilgan vaqti	27.02.1936
Nashr yili	2015
Nashr parametri	-
Nashriyoti	-
Qoʻllanish sohasi	adabiyotshunoslik
Adabiy turi	Dramatik
Janri	adabiy maktub
Vaqea vaqti va joyi	Oʻtmish voqeligi
Matn tipi	tasviriy
Uslubi	Badliiy
Auditoriya yoshi	18 yoshgacha
Auditoriyaning salohiyat darajasi	adabiyotshunoslik
Ichki korpus turi	davriy nashrlar korpusi
Soʻz(shakl) miqdori	0
Teglovchi	D.B. Axmedova

4-rasm. Abdulla Qahhorning “Anor” hikoyasiga berilgan metalingvistik maʼlumotlar

- 4) soʻzning izohi bilan tanishish;
- 5) soʻzning morfologik tahlili (morfoanalizator);
- 6) boʻgʻinlarga ajralishini oʻrganish;

7) qidiruvga berilgan so‘z bilan uning o‘ng va chap tomonidan birikish imkoniyatiga ega so‘zlar ro‘yxatini olish (N-gramma texnologiyasi);

8) so‘zning foydalanish chastotasi yoki statistikasini bilish;

9) so‘zning shakl va ma‘no munosabatlarini aniqlash (omonim, paronim, antonim, sinonim shakllari);

10) qidiruvga berilgan so‘z qatnashgan iboralarni taqdim etish;

11) o‘sha so‘zning ma‘no bo‘yoqdorligi, joylashuv o‘rni, qo‘llanilish ko‘lami kabilar bo‘yicha darajalanish qator(lar)ini taqdim etish;

12) “Ona tili qomusiy lug‘ati” bazasi asosida qidiruvga berilgan so‘z bilan bog‘liq ma‘lumotlarni taqdim etish imkoniyatlariga ega.

Mazkur imkoniyatlar korpus ruknlari doirasida namoyon bo‘ladi. Misol tariqasida korpusning “Lug‘atlar” ruknidagi qidiruv satriga “*falak*” so‘zini yozib natijani ko‘ramiz:

- 1) “so‘z tarkibi” – avtomatik morfologik tahlil natijasida “*falak*” asos ekanligi ma‘lum bo‘ladi;
- 2) “So‘z bo‘g‘inlari” *fa-lak* tarzida ajratib ko‘rsatildi;
- 3) *falak* so‘zining izohi taqdim etildi;
- 4) zid ma‘nosi yer so‘zi ekanligi ma‘lum bo‘ldi. Hatto shu o‘rinda yer so‘zining lisoniy ma‘lumotlarini ko‘rish uchun giperhavola mavjud;
- 5) *falak* lemmasining ma‘nodoshi taqdim etildi: *osmon*;
- 6) talaffuzdoshi (paronimi) *palak* ekanligi aniqlandi.

falak [Izlash](#)

So'z tarkibi: [falak]
So'z bo'g'inlari: fa-lak
Izoh(lar)i: falak [fa-lak]
1. Yerdan yuqoriga qaraganda gumbaz, qubba shaklida bo'lib ko'rinadigan fazo bo'shlig'i, osmon gumbazi; osmon, ko'k. *Yana chaqmoq chaqdi, yana falak gumburladi. H. G'ulom, Zamin yulduzlari*
Antonim(lar)i: yer
Sinonim(lar)i: osmon [f-t]
Omonimlari: mavjud emas.
Paronimi: palak – Qovun, tarvuz, bodringning yerda yoyilib o'sadigan tana, shox va barglari; so'zana.
Qomus: mavjud emas.

Ibora:(2 ta) [To'liq ko'rish](#)

1. Fig'oni falakka chiqdi: ruhan ortiq darajada bo'g'ilib, oh-voh qildi
2. Parvoyi falak: beparvo, beg'am

Darajalanish qatori:(1 ta) [To'liq ko'rish](#)

1. Osmon – ko'k – samo – falak – gardun
Bo'yog'iga ko'ra
(belgisining bo'yoqli-bo'yoqsizligiga ko'ra)

☒ Izoh(lar)i
☒ Omonim(lar)i
☒ Sinonim(lar)i
☒ Antonim(lar)i
☒ Paronimi
☒ Qomus
☒ Ibora(lar)
☒ Darajalanish qatori
☐ Uslubiy xoslanishi

5-rasm. O'zbek tili ta'limiy korpusi “Lug'atlar” ruknida “falak” so'zshakli qidiruvi natijasi

Mazkur misoldan anglashilganidek, ta'limiy korpus tarkibida sinonimayzer dasturi mavjud bo'lib, boshqa resurslarga bog'liq bo'lmagan holda sinonimlar bazasidan nomorfologik, ya'ni so'zning lemmasiga tayanib uning ma'nodoshi taqdim etiladi.

Ta'limiy korpus hajmi. Ayni damda o'zbek tili ta'limiy korpusi lingvistik ma'lumotlari hajmi quyidagi ko'rsatkichga ega:

№	Baza birligining nomi	soni
1	kitoblar	184
2	Internet matnlari	4 78 908
3	so'z turkumlari teglangan so'zlar	36 897
4	gaplar	5 600 015
5	kontekstlar	5 562 202
6	izohlangan so'zlar soni	4575
7	sinonim so'zlar qatori	993
8	antonim juftliklar	870



9	paronim juftliklar	558
10	omonim soʻzlar	1638
11	iboralar	1384
12	Darajalanish qatori	413

Xulosa tarzida aytish mumkinki, til korpusi milliy til birliklarining xususiyatlarini aniqlash maqsadida elektron qidiruv imkoniyati mavjud tizim, tabiiy tilning raqamlashgan yozma va ogʻzaki matnlari jamlanmasi hisoblanadi. Undan lingvistlar, leksikograflar, kompyuter lingvistlari, dasturchilar, muharrirlar, tarjimonlar, jurnalistlar, noshirlar, olimlar, oʻqituvchilar, taʼlim oluvchilar va boshqa har qanday soha mutaxassisi keng foydalanish imkoniyatiga ega boʻladi. Shu bois til taʼlimida yuqori samaradorlikka erishish uchun muntazam ravishda matnlar bazasi yangilanib boriladigan korpusga tayanish maqsadga muvofiq boʻladi.

Foydalanilgan adabiyotlar:

1. Oʻzbek tili milliy korpusi – muhim madaniy voqelik (davra suhbat) / “Maʼrifat”, 11.08.2021. № 32 (9357).
2. Boliang Zhang, Ajay Nagesh, Kevin Knight. 2020. Parallel Corpus Filtering via Pre-trained Language Models. *10.18653/v1/2020.acl-main.756*
3. Nina Vyatkina. 2020. Corpora as open educational resources for language teaching. *Foreign Language Annals, 10.1111/flan.12464*
4. Zhao Shengxue. 2020. English corpus translation system based on FPGA and machine learning. *Microprocessors and Microsystems, <https://doi.org/10.1016/j.micpro.2020.103464>*
5. <http://uzschoolcorpara.uz/> – Oʻzbek tilining taʼlimiy korpusi