

Handwritten signature

O'ZBEKISTON RESPUBLIKASI
OLIY TA'LIM, FAN VA INNOVATSIYALAR VAZIRLIGI
BUXORO DAVLAT UNIVERSITETI



O'ZBEK TILSHUNOSLIGI VA JURNALISTIKA KAFEDRASI

“KORPUS LINGVISTIKASI”

O'QUV DASTURI

Bilim sohasi:	200 000 – San'at va gumanitar fanlar
Ta'lim sohasi:	230 000 – Tillar
Mutaxassislik:	70230401 – Kompyuter lingvistikasi
Ta'lim shakli:	kunduzgi

Mazkur fan dasturi xalqaro tan olingan reytinglarda birinchi 300 talik ro'yxatga kiruvchi **Australian National University (ANU), Australia (QS 32, THE 73)** ta'lim dasturi asosida tayyorlandi. (https://faculty.wcas.northwestern.edu/robvoigt/courses/2021_spring/ling334/)

Buxoro -2025

Fan/modul kodi KL 1-2 10	Fan/modul turi majburiy	Ta'lim tili o'zbek	Ishlab chiqilgan o'quv yili 2025/2026			
Semestr	Har bir semestrda fan/modulning nomi	ECTS - Kreditlar	Haftalik dars soatlari	Auditoriya mashg'ulatlari (soat)	Mustaqil ta'lim (soat)	Jami yuklama (soat)
1	Korpus lingvistikasi	5	4	60	90	150
2	Korpus lingvistikasi	5	4	60	90	150
	Jami:	10	12	120	180	300

2.	<p style="text-align: center;">I. Fanning mazmuni</p> <p>Fanni o'qitishdan maqsad – magistr'larga til korpusini yaratishning asosiy konsept va metodlarini tanishtirish hamda korpus uchun qidiruv tizimidagi til texnologiyalari va instrumentlarini o'rgatish va ularda Milliy korpus yaratish malakasini shakllantirish.</p> <p>Fanning vazifasi – korpus, korpus lingvistikasi, uning shakllanishi va taraqqiyoti, korpus lingvistikasining bugungi holatini, shuningdek, korpus tuzishning umumiy tamoyillarini tavsiflash, korpus tuzishning lingvistik asoslarini sharhlash, korpuslarning mushtarak hamda o'ziga xos jihatlari aniqlash va dalillash, korpus tuzish tamoyillarini ishlab chiqishni o'rgatish, korpus yaratishda kerakli bo'lgan lingvistik annotatsiyalash bo'yicha nazariy va amaliy bilimlarni hosil qilish, korpuslardan maqsadli foydalanish bo'yicha malaka hosil qilish.</p> <p style="text-align: center;">II. Asosiy nazariy qism (ma'ruza mashg'ulotlari)</p> <p style="text-align: center;">II.1. Fan tarkibiga quyidagi mavzular kiradi:</p> <p style="text-align: center;">1-mavzu. Korpus va korpus lingvistikasi</p> <p>Korpus tushunchasi va Milliy korpus, korpus lingvistikasi predmeti va uning shakllanish omili, korpus lingvistikasi bo'yicha asosiy tushunchalar, lingvistik tadqiqotlarda korpusning ahamiyati. Korpus lingvistikasining yondosh sohalari rivojida o'rni va ahamiyati.</p> <p style="text-align: center;">2-mavzu. Korpus lingvistikasining shakllanish taraqqiyoti</p> <p>Dunyo tillari korpuslarining paydo bo'lishi tarixi, taraqqiyot bosqichlari, korpuslarning tarkibiy tuzilishi, Braun korpusining yaratishi va rivojlanishi.</p> <p style="text-align: center;">3-mavzu. Korpus toksonomiyasi¹</p> <p>Maqsad va vazifasiga ko'ra korpus klassifikatsiyasi: diaxron korpus va zamonaviy korpus, milliy korpus, monolingval va bilingval (parallel),</p>
----	---

¹ Mazkur fan dasturi xalqaro tan olingan reytinglarda birinchi 300 talik ro'yxatga kiruvchi Australian National University (ANU), Australia (QS 32, THE 73) ta'lim dasturi asosida tayyorlandi. (https://faculty.wcas.northwestern.edu/robvoigt/courses/2021_spring/ling334/)

ko'ptilli korpuslar, ochiq va yopiq turdagi korpuslar, ta'limiy korpuslar, lingvistik annotatsiyalangan korpuslar, janr va matn hajmiga ko'ra tasnifiy korpuslar, piramidali korpuslar, muvozanatlashgan korpuslar.

Classification of corpora according to their purpose and objectives: diachronic corpus and contemporary corpus, national corpus, monolingual and bilingual (parallel) corpora, multilingual corpora, open and closed corpora, educational corpora, linguistically annotated corpora, genre- and text-size-based classified corpora, pyramidal corpora, balanced corpora.

4-mavzu. Korpus turlari

Korpusni tasniflash mezonlari. Korpuslar tipologiyasi: tili, foydalanish usuli, xronologik belgisiga ko'ra turlari. Reprezentativ korpus janrlari. Korpus tarkibidagi materialning turi va xususiyati. Korpuslarning reprezentativligi. Korpusning nutq shakliga ko'ra turlari: yozma, og'zaki, aralash. Og'zaki korpuslar ahamiyati va imkoniyatlari. Og'zaki nutqning kompyuter korpuslarini yaratish tamoyillari. Mualliflik korpus. Gazeta matnlari korpusi. Milliy korpus.

5-mavzu. Korpusning lingvistik va ta'limiy ahamiyati

Lingvistik tadqiqotlarda ta'limiy korpuslarning yaratilishi va qo'llanilishi. Korpus yirik leksikografik manba sifatida. Tilshunoslikning leksikologiya sohasini o'rganishda korpusning ahamiyati. Korpus va lingvodidaktika.

6-mavzu. Korpus yaratish texnologiyasi. Ma'lumot to'plash (ma'lumotlarni saqlash, konkordans dasturiy ta'minoti yordamida ma'lumotlarni olish, onlayn interfeyslar)

Korpus tuzishning umumiy tamoyillari. Korpus yaratishni loyihalash va uning texnologik jarayoni. Korpus lingvistikasida standartlashtirish.

7-mavzu. Korpus qidiruv tizim sifatida

Konkordanser va korpus menejeri. Internet tizimida mashina fondi bilan ishlash. Korpus menejerida til so'rovini yaratish va chiqish interfeysi. Korpuslar va korpus menejerining qiyosiy analizi.

8-mavzu. Korpus uchun kompyuter instrumentlari

Web-interfeyslar (COCA, NKRYA). AntConc ilovasi. Korpus menejerlari. Xaira. CQP. DDC. Manatee/Bonito. GATE. Sketch Engine.

9-mavzu. Korpusda lingvistik annotatsiyalash tamoyillari

Lingvistik annotatsiya. Korpus annotatsiyasi ekstralingvislik ma'lumot sifatida. Annotatsiya korpus ahamiyatini belgilovchi asosiy omil. Korpus annotatsiyasi bilan tanishish. Metama'lumot tushunchasi. Ekstralingvistik razmetka: bibliografik, topologik, tematik, sotsiologik, formal grammatik va texnologik. Lingvistik korpusda razmetkalash masalasi. Korpus razmetkalari, metatekst va lingvistik razmetkalash turlari. Lingvistik razmetkaning turlari. Sintaktik razmetkaning lingvistik asoslari.

10-mavzu. Korpusda matn tahlili texnologiyalari. Tilni modellashtirish (n-gramm, maksimal ehtimollikni baholash asosida)

Tabiiy tilni qayta ishlashning asosiy texnologiyalari: tokenizatsiya, lemmatizatsiya, stemming, parsing.

11-mavzu. Mualliflik korpusi

Mualliflik korpusi turlari va uni tuzishning lingvistik asoslari. Mualliflik korpusining tuzilishi va standart belgilari. Jahon korpus lingvistikasida yaratilgan mualliflik korpuslar va ularning umumiy va farqli jihatlari.

12-mavzu. Korpus - ilmiy tadqiqot materiali bazasi

Korpus bilan bog'liq leksikografik tadqiqotlar. Kvantativ lingvistika, mashina tarjimasini, lingvodidaktika va matn lingvistikasiga doir tadqiqotlarda korpus lingvistikasining o'rni va ahamiyati. Rus korpus lingvistikasida olib borilgan tadqiqotlar. O'zbek korpus lingvistikasi tadqiqotlar tahlili.

13-mavzu. Og'zaki nutq kompyuter korpuslari

Og'zaki nutqning kompyuter korpuslarini yaratish tamoyillari. Og'zaki nutq korpusining turlari va ularni yaratish texnologiyasi.

14-mavzu. Ko'p tili va parallel korpus texnologiyasi

Maxsus va parallel korpuslar, ularning ahamiyati. Mashina tarjimasida parallel korpuslar o'rganish obyekti sifatida. Lingvistik tarjima va kompyuter tarjimasini. Tarjima dasturlarini yaratishda korpusning o'rni. Parallel korpuslarda tarjima masalasi. Tillarni o'rganishda korpusning ahamiyati. Parallel korpus yaratish texnologiyasi.

15-mavzu. Korpus lingvistikasi va tarjima

Parallel va ko'p tili korpuslar. Tarjima texnologiyasida parallel korpuslar platformasidan foydalanish. Parallel korpus yaratish texnologiyasi

16-mavzu. Lingvistikada korpusga asoslangan metodlar

Kompyuter yo'nalishlari: statistik tahlil metodi, mashina tarjimasini, matn lingvistikasi va lingvodidaktika sohasidagi tadqiqotlarda lingvistik korpusning ahamiyati. Psixolingvistikada, sotsiolingvistikada, amaliy tarjimashunoslikda, dialektologiyada, diskurs tahlilda

17-mavzu. Korpus lingvistikasining xalqaro standartlari

Korpus lingvistikasini standartlashtirish tamoyillari va usullari, terminologik standartlarni ishlab chiqish. Til resurslarini ishlab chiqish, kompyuter leksikografyasi sohasida standart tamoyillar ishlab chiqish.

18-mavzu. Xorijiy til korpuslari tajribasidan

Milliy til korpuslari: Amerika, Rossiya, Turkiya, Buyuk Britaniya misolida

19-mavzu. O'zbek tili elektron korpusi va uning imkoniyatlari

O'zbek tili elektron korpusi yaratish tendensiyasi. O'zbek tili elektron korpusining qidiruv menejeri. O'zbek tili elektron korpusida sub korpuslar yaratish masalalari.

20-mavzu. O'zbek tilining ta'limiy korpusi

O'zbek tilining ta'limiy korpusini yaratish bosqichlari. Ta'limiy korpusi kontenti bilan ishlash. O'zbek tilining ta'limiy korpusida akademik lug'at yaratish masalalari.

III. Amaliy mashg'ulotlari bo'yicha ko'rsatma va tavsiyalar

Amaliy mashg'ulotlar uchun quyidagi mavzular tavsiya etiladi:

1. Korpus lingvistikasiga kirish
2. Korpus evolyutsiyasi tadqiqi
3. Korpus taksonomiyasi

4. Korpuslar tipologiyasi: tili, foydalanish usuli, xronologik belgisiga ko'ra turlari

5. Korpus yirik leksikografik manba sifatida
6. Korpus tuzishning umumiy tamoyillari
7. Korpus qidiruv tizim sifatida
8. Web-interfeyslar (COCA, NKRYA). AntConc ilovasi va boshqalar
9. Korpus annotatsiyasi bilan tanishish
10. Lingvistik korpusda razmetkalash masalasi
11. Tabiiy tilni qayta ishlashning asosiy texnologiyalari
12. Mualliflik korpusi
13. Korpus - ilmiy tadqiqot materiali bazasi
14. Og'zaki nutq kompyuter korpuslari
15. Parallel korpus texnologiyasi
16. Korpus lingvistikasi va tarjima
17. Lingvistikada korpusga asoslangan metodlar
18. Korpus lingvistikasining xalqaro standartlari
19. Xorijiy til korpuslari tajribasidan
20. O'zbek tili elektron korpusi va uning imkoniyatlari

Amaliy mashg'ulotlari multimedia qurilmalari bilan jihozlangan auditoriyada bir akademik guruhga bir professor-o'qituvchi tomonidan o'tkazilishi zarur. Mashg'ulotlar faol va interaktiv usullar yordamida o'tilishi, mos ravishda munosib pedagogik va axborot texnologiyalar qo'llanilishi maqsadga muvofiq.

IV. Mustaqil ta'lim va mustaqil ishlar

Mustaqil o'zlashtiriladigan mavzular bo'yicha magstrlar tomonidan referatlar tayyorlash va uni taqdimot qilish tavsiya etiladi.

Mustaqil ta'lim uchun tavsiya etiladigan mavzular:

1. Korpus lingvistikasining tarixi: kartotekadan korpusgacha
2. Korpus lingvistikasining rivojlanishi
3. Korpus menejeri
4. Lingvistik tadqiqotlarda korpusdan foydalanish usullari.
5. Leksikografiyada korpusdan foydalanish usullarini o'rganish.
6. Korpus lingvistikasi uchun veb-sayt yaratish,
7. Har xil turdagi mavjud bo'lgan korpuslarning ko'rinishi
8. Korpus qidiruv tizim sifatida
9. Korpusga asoslangan lingvistik tadqiqotlar
10. Ma'lumotlarni grafik tahlil qilish
11. WebCorp interfeysining tahlili va tavsifi
12. Tilshunoslik tadqiqotida korpuslardan foydalanish usullari
13. Leksikografiyada korpuslardan foydalanish usullari
14. Korpus lingvistikasida standartlashtirish
15. Annotatsiya korpus ahamiyatini belgilovchi asosiy omil

	<p>16. Mashina tarjimasida parallel korpuslar o'rganish obyekti sifatida</p> <p>17. Parallel korpuslarda tarjima masalasi</p> <p>18. O'zbek tili elektron korpusining qidiruv menejeri</p> <p>19. O'zbek tili elektron korpusida sub korpuslar yaratish masalalari</p> <p>20. O'zbek tilining ta'limiy korpusida akademik lug'at yaratish masalalari</p> <p>21. Korpusga asoslangan diplomatik diskurs tahlili</p> <p>22. Korpusga asoslangan tezaurus</p> <p>23. Korpusga asoslangan ta'limiy lug'atlar</p>
3.	<p>V. Fan o'qitilishining natijalari (shakllanadigan kompetensiyalar)</p> <p>Fanni o'zlashtirish natijasida magistr:</p> <ul style="list-style-type: none"> • magistrilar til korpuslari bo'yicha tayanch tushunchalarni, korpus terminologiyasi va vazifalarini bilishi hamda til korpuslarining tarkibi va tuzilishi haqida <i>bilimga ega bo'lishi</i>, til korpuslari, til korpusining lingvistik, lingvodidaktik, amaliy va ijtimoiy ahamiyati haqida <i>tasavvurga ega bo'lishi</i>; • xorijiy tillarni o'rganish, ilmiy tadqiqotlarni amalga oshirish jarayonida korpus materiallaridan foydalanish, fanga oid maxsus topshiriqlarni bajara olishi, shuningdek, xorijiy korpuslardan umumli va o'zbek tili milliy korpusiga doir tadqiqotlardan amalda to'g'ri foydalanish <i>ko'nikmalariga ega bo'lishi</i>; • magistrilar korpus menejerining formal-funksional modellari asosida korpus interfeysini shakllantirishga oid olgan bilimlarini faoliyatida qo'llay olish <i>malakalariga ega bo'lishi kerak</i>.
4.	<p>VI. Ta'lim texnologiyalari va metodlari:</p> <ul style="list-style-type: none"> • ma'ruzalar; • interfaol keys-stadilar; • mantiqiy fiklash, tezkor savol-javoblar; • guruhlarda ishlash; • taqdimotlar (PPT)ni qilish; • individual loyihalar; • jamoa bo'lib ishlash va himoya qilish uchun loyihalar.
5.	<p>VII. Kreditlarni olish uchun talablar</p> <p>Fanga oid nazariy va uslubiy tushunchalarni to'liq o'zlashtirish, tahlil natijalarini to'g'ri aks ettira olish, o'rganilayotgan jarayonlar haqida mustaqil mushohada yuritish va joriy, oraliq nazorat shakllarida berilgan vazifa va topshiriqlarni bajarish, yakuniy nazorat bo'yicha yozma ishni topshirish.</p>
6.	<p>Normativ-huquqiy hujjatlar</p> <p>1. O'zbekiston Respublikasi Prezidentining "O'zbek tilining davlat tili sifatidagi nufuzi va mavqei tubdan oshirish chora-tadbirlari to'g'risida"gi PF-5850-son Farmoni, Toshkent shahri, 2019-yil 21-oktabr</p>

	<p>2. O'zbekiston Respublikasi Prezidentining "Mamlakatimizda o'zbek tilini yanada rivojlantirish va til siyosatini takomillashtirish chora-tadbirlari to'g'risida"gi PF-6084-son Farmoni, 2020-yil 20-oktabr</p> <p>Asosiy adabiyotlar</p> <p>1. Abduraxmonova N. va Toirova G. Korpus lingvistikasi (darslik), 2025. – 385 b.</p> <p>2. Zaxarov V., Mengliyev B., Hamroyeva Sh. Korpus lingvistikasi. O'quv qo'llanma. GlobeEdit, 2021. – 198 b.</p> <p>3. Po'latov A. Kompyuter lingvistikasi. – Toshkent, 2011. – 520 b.</p> <p>4. Rahimov A. Kompyuter lingvistikasi asoslari. – Toshkent, 2011. – 160 b.</p> <p>5. Jurafsky D., J. H. Martin. Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition. 2023. – 620 p. (University of Southern California)</p> <p>Qo'shimcha adabiyotlar</p> <p>6. Захаров В.П., Богданова С.Ю. Корпусная лингвистика. Учебник. – Иркутск, 2011. – 161 с</p> <p>7. Хамроева Ш. Корпус лингвистика атамаларининг кискача атамалар лугати. Тошкент, 2018</p> <p>8. Горина О.Г. Инструменты корпусного анализа в обучении иностранному языку. Вестник Томского государственного университета. 2018. № 435. С. 187–194.</p> <p>9. Козлова Н.В. Лингвистические корпуса: определение основных понятий и типология / Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2013. Том 11, Выпуск 1. – С. 79- 88</p> <p>10. Evert S. et al. Corpora and collocations //Corpus linguistics. An international handbook. – 2008. – Т. 2. – С. 1212-1248. (Georgetown University)</p> <p>11. Gries, Stefan. On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement. Corpus Linguistics and Linguistic Theory. 2020. 16.3, 617-47. (The University of Manchester)</p> <p>12. Jurafsky D., J. H. Martin. Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition. 2023. – 620 p. (Australian National University (ANU))</p> <p>13. Jacob Eisenstein (NLP Notes) Natural Language Processing. 2018. P. 490 (Australian National University (ANU))</p> <p>14. Lindquist Hans, Levin Magnus. Corpus linguistics and the description of English. Second edition., Edinburgh, Edinburgh University Press, 2018. (Uppsala universitet)</p> <p>15. Meyer, C. English Corpus Linguistics. Cambridge: Cambridge University Press. 2002. (City University of Hong Kong)</p>
--	---

Axborot manbalari	
16.	http://nuu.uz/ – O‘zbekiston Milliy universiteti
17.	http://lex.uz
18.	https://uzbekcorpus.uz/
19.	http://corpus.byu.edu/coca
20.	http://tugantel.tatar/search/
21.	http://www.turklang.net
22.	https://mlcommons.org/en/multilingual-spoken-words/
23.	https://ruscorpora.ru/new/
24.	https://www.manchester.ac.uk/ – The University of Manchester
25.	https://www.uu.se/ – Uppsala University, Sweden
26.	https://www.cityu.edu.hk/ – City University of Hong Kong
27.	https://www.georgetown.edu/ – Georgetown University, Washington
28.	https://www.anu.edu.au/ – Australian National University (ANU)
7.	Fanning o‘quv dasturi Mirzo Ulug‘bek nomidagi O‘zbekiston Milliy universiteti tomonidan ishlab chiqilgan. O‘zMU kengashining 202__-yil “__”__dagi __-sonli bayonnomasi bilan ma‘qullangan. O‘zbekiston Respublikasi Oliy ta‘lim, fan va innovatsiyalar vazirligining 202__-yil “__”__dagi __-sonli buyrug‘i bilan ma‘qullangan fan dasturlarini tayanch oliy ta‘lim muassasasi tomonidan tasdiqlashga rozilik berilgan.
8.	Fan/modul uchun mas‘ul: G.I.Toirova – BuxDU O‘zbek tilshunosligi va jurnalistika kafedrası professori, filologiya fanlari doktori (DSc)
9.	Taqrizchilar: 1. B.Elov - ToshDO‘TAU dotsenti, texnika fanlari nomzodi; 2. N.Abduraxmonova – O‘zMU Kompyuter lingvistikasi va amaliy tilshunoslik kafedrası mudiri, filologiya fanlari doktori, professor

O‘quv-uslubiy boshqarma boshlig‘i
Kafedra mudiri
Tuzuvchi



S.Davlatov
B.Qilichev
G.Toirova

Mazkur o‘quv dasturi dunyoning e‘tirof etilgan xalqaro QS va THE reytinglarida nufuzli TOP-300 talikka kirgan quyidagi oliy ta‘lim tashkilotlarining ta‘lim dasturlari asosida ishlab chiqilgan.

No	OTM nomi	QS	THE	Top-300 ta‘lim dasturi asosida kiritilgan qo‘shimchalar	Mazkur dasturdagi mavzu nomi	Havolalar:
1.	The University of Manchester, Manchester, United Kingdom	35-o‘rin	53-o‘rin	Data collection 1 (data storage, retrieving data with concordance software, online interfaces) (<i>Ma‘lumot to‘plash (ma‘lumotlarni saqlash, konkordans dasturiy ta‘minoti yordamida ma‘lumotlarni olish, onlayn interfeyslar)</i>)	6-mavzu. Korpus yaratish texnologiyasi. Ma‘lumot to‘plash (ma‘lumotlarni saqlash, konkordans dasturiy ta‘minoti yordamida ma‘lumotlarni olish, onlayn interfeyslar) Eslatma: Mavzuga qo‘shimcha kiritildi	https://www.manchester.ac.uk/study/masters/courses/list/21198/msc-computational-and-corpus-linguistics/course-details/ELA60112#course-unit-details
2.	Uppsala University, Sweden	93-o‘rin	130-o‘rin	use a variety of corpus tools to perform searches of spoken and written language data (<i>og‘zaki va yozma matnlarni qidirishda korpus instrumentlaridan foydalanish</i>)	8-mavzu. Korpus uchun kompyuter instrumentlari Eslatma: Mavzu bir xil	https://www.uu.se/en/study/syllabus?query=32182
		63-o‘rin	78-o‘rin	Corpus annotation:	9-mavzu. Korpusda	https://www.cityu.edu.hk/ca

3.	City University of Hong Kong, Hong Kong, SAR			textual, grammatical, syntactic (<i>Korpus annotatsiyasi: morfologik, sintaktik, semantik</i>)	lingvistik annotatsiyalash tamoyillari Eslatma: Mavzu bir xil	talogue/pg/202122/course/L15421.pdf
4.	Australian National University (ANU), Australia	32-o'rin	73-o'rin	Language Modeling (n-grams, maximum likelihood estimation) (<i>Tilni modellashtirish (n-gramm, maksimal ehtimollikni baholash asosida)</i>)	10-mavzu. Korpusda matn tahlili texnologiyalari. Tilni modellashtirish (n-gramm, maksimal ehtimollikni baholash asosida) Eslatma: Mavzuga qo'shimcha kiritildi.	https://faculty.wcas.northwestern.edu/robvoigt/courses/2021_spring/ling334/
5.	Georgetown University, Washington D.C., United States	285-o'rin	201-250-o'rin	Multilingual and Parallel Corpora (<i>Ko'ptilli va parallel korpus</i>)	14-mavzu. Ko'p tilli va parallel korpus texnologiyasi Eslatma: Mavzu bir xil	https://gucorpling.org/amir/pdf/syllabi/4427_Corpus_Zeldes_F23.pdf