



*Nilufar ABDURAXMONOVA,*

*O'zMU Kompyuter lingvistikasi va amaliy tilshunoslik kafedrasini mudiri, f.f.d., professor*

*E-mail: n.abduraxmonova@nuu.uz*

*Karomat RAJABOVA,*

*O'zMU Kompyuter lingvistikasi mutaxassisligi magistranti*

*Zilola RO'ZIYEVA,*

*O'zMU magistratura bo'limi uslubchisi*

*O'zMU professori Z.Taxirov taqrizi asosida*

### KORPUS QIDIRUV MENEJERI SIFATIDA

Annotatsiya

Ushbu maqola konkordans terminining tarixiy rivojlanishi va lingvistik korpus tahlili uchun yaratilgan dasturiy instrumentariylar haqida ma'lumot beradi. Konkordansning dastlabki qo'l mehnatidan avtomatlashtirilgan versiyalarga o'tish jarayoni, Roberto Busa va Birmingham universiteti tomonidan ishlab chiqilgan dastlabki instrumentlardan boshlab, hozirgi zamonaviy konkordans instrumentlarigacha bo'lgan rivojlanish bosqichlari muhokama qilinadi. Korpus lingvistikasida keng qo'llaniladigan dasturlar va ularning funksiyalari haqida batafsil tahlil keltiriladi.

**Kalit so'zlar:** Konkordans, korpus menejeri, korpus, lingvistik instrumentariy.

### AS CORPS SEARCH MANAGER

Annotation

This article describes the historical development of the term concordance and the software tools developed for linguistic corpus analysis. The progression of the concordance from the original handwork to automated versions is discussed, from the early instruments developed by Roberto Busa and the University of Birmingham to the modern concordance instruments of today. A detailed analysis of commonly used programs in corpus linguistics and their functions is provided.

**Key words:** Concordance, corpus manager, corpus, linguistic instrument.

### МЕНЕДЖЕР ПО КОРПУСНОМУ ПОИСКУ

Аннотация

В данной статье описывается историческое развитие термина конкорданс и программные средства, разработанные для анализа лингвистического корпуса. Обсуждается развитие симфонии от оригинальной ручной работы к автоматизированным версиям, от ранних инструментов, разработанных Роберто Бузой и Бирмингемским университетом, до современных инструментов конкорданса. Приведен подробный анализ часто используемых программ в корпусной лингвистике и их функций.

**Ключевые слова:** Конкорданс, корпус-менеджер, корпус, лингвистический инструмент.

**Kirish.** Ilmiy manbalarda konkordans termini bilan bog'liq qiziq asoslar uchraydi. Chunonchi, 1230-yilda lotin vulgat Bibliyasi uchun ilk bor qo'l mehnati orqali konkordans yaratilganligi qayd etiladi. Korpus tahlili uchun dasturiy instrumentariysini yaratish Roberto Busa tomonidan taklif etilgan bo'lib, dastlabki konkordanser (1951) instrumenti shu olim nomi bilan tilga olinadi. Biroq bu texnologiya ham to'liq kompyuterlashtirilmagan edi. Dastlabki avtomatik ravishda ishlovchi konkordanser Birmingham universitetida CLOC korpusi uchun yaratilgan (1978, Reed).

Konkordansning 2-avlodi dunyo bo'ylab bir turdagi IBMning shaxsiy kompyuterlar ommaviylashgandan keyin foydalanish imkoniyati yaratildi [Tony McErery, Andrew Hardie, 2010]. Bu kabi kompyuterlarga konkordanser yozish imkoniyati kengaydi.

Korpus lingvistikasi 1980-yillarning oxirida juda tezlik bilan rivojlana boshladi. Natijada konkordans instrumentlari sifatida Kaye konkordanser (Kaye 1990), the Longman Mini-Konkordanser (Chandler 1989) va Micro-OCF (Hockey 1988) kabilar yaratildi. 1-avlodga nisbatan 2-avlod konkordans turlari avvalgi turlarning kamchiliklarini bartaraf etib, qidirilayotgan so'zni o'z va chap tomondan alifbo tarzda

tartiblash hamda olingna natijani statistic ma'lumot sifatida hisoblash imkoniyati yaratildi.

3-avlod konkordansi WordSmith (Scott 1996), MonoConc (Barlow 2000), AntConc (Anthony 2005) va Xaira kabilar. XML format kabi standart formatlarni konkordansga moslashtirilganligi va UNICODE kodlash tizimi joriy etilgani ijobiy ta'sir ko'rsatdi.

Bu avlod konkordanslar turli jihatlari bilan bir-biriga o'xshash bo'lib, har birida konkordans, chastotali lug'at, kollokatsiya hamda kalit so'zlar tahlili asosi instrumentariy hisoblanadi. Ushbu instrumentariylar korpusdagi matnlarni muayyan so'rovga ko'ra hisoblaydi (so'zlar chastotasi ro'yxati) hamda qidirilayotgan terminning barcha namunalarini (konkordans) ko'rsatadi. Yuqorida nomlari qayd etilgan instrumentariylar bir-biridan muayyan funksional imkoniyatlariga ko'ra farqlanadi. Masalan, Xaira konkordans WordSmith va AntConcga nisbatan solishtirganda kalit so'zlarni hisoblash uchun instrumentariy mavjud emas, lekin XML elementlari (annotatsiya, metama'lumot) bo'yicha qidiruvni amalga oshira oladi. Korpusdagi so'zlarni (n-gram) ketma-ketlikda berishda bir-biridan farq qiladi [Tony McErery, Andrew Hardie, 2010].

N	Concordance	File	Date	%
1	bosom of the air. O Romeo, Romeo! wherefore art thou Romeo? Deny thy	Romeo And Juli	1594	27%
2	it not a sin. Why, how now, kinsman! wherefore storm you so? Uncle, this	Romeo And Juli	1594	21%
3	How cam'st thou hither, tell me, and wherefore? The orchard walls are high	Romeo And Juli	1594	28%
4	three-hours wife, have mangled it? But, wherefore, villain, didst thou kill my	Romeo And Juli	1594	56%
5	slain my husband: All this is comfort, wherefore weep I then? Some word	Romeo And Juli	1594	56%

1-rasm

“Romeo va Juliyetta” asaridagi wherefore so‘zining konkordans bo‘yicha statistik ma‘lumoti [https://lexically.net/wordsmith/step\_by\_step8/English/overviewofconcord.html] Endilikda konkordans funksiyasiga ega bo‘lgan lingvistik korpus instrumentlarni soni miqdor jihatdan oshgan [https://corpus-analysis.com/tag/concordancer]: aConCorde (ingliz va arab tillari uchun konkordans), AntConc, AntPConc (parallel korpuslar uchun konkordans), BFSUParaConc (parallel konkordans), BFSUPowerConc, BNCWeb (Britaniya milliy korpusidan olingan grammatik va tekstual, leksik ma‘lumotlarni qayta ishlash va qidirish uchun webga asoslangan kliyent dasturi), buzz (pitonga asoslangan lingvistik tahlil instrumentariysi), CasualConc (MacOS uchun), CliC, Collocate, Concordance Randomizer, Concordancer, CorpKit, Corpus presenter, gwic, HeidelGram Web-based tools, IMS Corpus Workbench, KAT tool, KWords, Lestutor web concordancers, MLCT (korpus yaratish

va uni qayta ishlash uchun instrumentariyi), MonoConcEsy, OpenConc, ParaConc (bilingual/multilingual konkordanser), PhraseContext, Praaline, PyXMLConc, Shinyconc, Simple Concordance program, Sketch engine, Text Analysis Computing Tools (TACT), Textanz, TextSTAT, The Prime Machine, The simple Corpus tool, The SPAADIA concordance, The text feature analyzer, WordWander, WordCruncher.

Korpusga asoslangan tadqiqotlarning asosiy lingvistik instrumentariysi bu konkordans hisoblanadi. Konkordans foydalanuvchiga korpusdagi kerakli ma‘lumotni muayyan uzunlikda so‘z, so‘z qismi (morfem bo‘lak) yoki fraza shaklida qidirish imkonini beradi. Olingan natija bir qator trazida berilib, keyin topilgan natija uning tagidan, ya‘ni vertikal shaklda ekranda namoyon bo‘ladi <http://uzbekcorpus.uz/> elektron korpusida konkordans bo‘yicha qidirish:

USLUB DAVR TOMON GRAMM English Русский O'z

bola

IZLASH TOZALASH

46 ta hujjatda 936 ta so‘z uchradi.

Gafur Gulom «Shum bola» 1966	boyvachchalar“, degani hali ham qulog‘imdan Biz yalangoyoq, bo‘z ko‘ylak- ishtonli, kir-chir	bolalar	to‘tiqa – To‘ti, – To‘ti, to‘ti... – deb qichqirar edik. Asra kal...ko‘proq...
Gafur Gulom «Shum bola» 1966	To‘ti bo‘lsa –Buvingni... – deb Bozorda sang‘ib yurgan biz daydi	bolalar	uchun quvonchli ermaklardan biri bozor, mahalla, ko‘cha-ko‘y jinnilari edi. ...ko‘proq...
Gafur Gulom «Shum bola» 1966	ko‘payib ketganidan keyin bo‘z o‘tmay qolib, bozori kasodga uchrab,	bola-chaqasini	boqolmay, jinni bo‘lib qolibdi, deb Eshon oyi degan xotin ham ...ko‘proq...
Gafur Gulom «Shum bola» 1966	ko‘chaning o‘ng, chap tomonidagi pastqam, tor ko‘chalarda o‘g‘il va qiz	bolalar	to‘planib, har xil o‘yinlar o‘ynaymiz. Kurash, “botmon-botmon”, “oq terakmi-ko‘k terak”, ...ko‘proq...

2-rasm

Korpus menejeri bo‘yicha tatar tilida amalga oshirilgan ishlar yuzasidan ma‘lumotlar tahlil qilsak. Tatar tili korpus [http://tugantel.tatar] menejeri uchun MVC (Model-View-Controller) konsepsiyasi qo‘llangan. Tizim faoliyati asosiy boshqaruvchi (MainControl) orqali so‘rov (Request)dan boshlanadi. Dastlab, asosiy boshqaruvchi ShieldModel komponenti yordamida xavfsizlikni ta‘minlaydi, so‘ng obyekt tartibi (Queue)dan foydalaniladi. Agar Shield Model so‘rov xavfsiz deb hisoblasa, keying navbatni so‘ralayotgan vazifa bilan bog‘liq ravishda maxsus boshqaruvchiga o‘tkazadi. Maxsus tuzilmada 6 turdagi maxsus boshqaruv tizimi mavjud: statistik sahifadagi xulosa uchun SinglePageControl, qidiruv so‘rovini qayta ishlash uchun Search Control, kontekst kengaytmasi uchun Context Control, statistik sahifalarni boshqarish uchun SinglePageEditControl, statistik tizimlarni ko‘rish uchun StatisticsControl, korpus ma‘lumotlarni boshqarish uchun DataManagementControl, xavfsizlik sozlamasini boshqarish uchun SecurityControl. Dastlabki uch boshqaruv tuzilmasi foydalanuvchilar uchun bo‘lganlar AdminControl tomonidan boshqariladi. PageControl va AdminControl birgalikda yordamchi RequestControl boshqaruvchidan foydalanadi.

Manbalarda [Pavel Rychly Manatee/Bonito,2007] korpusning ideal menejeri uchun quyidagilardan iborat bo‘lishi lozimligini qayd etishadi:

Matn tayyorlash – turli formatlardan konvertlash, kodlash;

Metama‘lumot boshqaruvi – berilganlar manbasi, muallifi, mavzusi, janri haqida ma‘lumotlarni integratsiya qilish;

Tokenizatsiya – alohida birlik, so‘zni aniqlash;

Korpus annotatsiyasi – morfologik, sintaktik, semantik va pragmatik darajada qo‘l mehnati yoki avtomatik usulda tglash;

Yetarli korpus bazasi – barcha ma‘lumotlar bazasini tez qidirishi uchun tuzilmasini loyihalash;

Konkordanslash – foydalanuvchi so‘roviga binoan matndagi tegishli so‘z o‘ramlarni qayta ishlash;

Statistikani hisoblash – topilgan statistik ma‘lumotni turli belgilarga ko‘ra maxsus maza so‘rovi uchun qidirish.

Korpus manajment tizimi Manatee [wide Petr Sojka, Aleš Horák,2007] katta korpus ma‘lumotlaridan leksik statistikani hisoblash uchun yaratilgan platforma hisoblanadi. Ushbu tizim modulyar yondashuv asosida shakllantirilgan.

Indekslarni olish, yaratish va siqish uchun maxsus indekslar kutubxonasi, analizator so'rovi, so'rovlar bilan turli operatsiyalar uchun sinflar bilan baholash so'rovlari, sintaktik analiz uchun sintaktik daraxt kabi modullardan tashkil topgan. Manatee tizimi FinLib - matnli indekslash kutubxonasi so'zlarni qisish, invertlashgan indekslar asosidagi korpus ma'lumotlari uchun qidiruv protsedurasi va strukturlarni taqdim etadi. Tizim tashqi tokenizator yoki oddiy ichki tahlil asosida ishlaydi. Har bir token uchun lingvistik ma'lumotlar (POS yoki so'zning bosh shakli)ni hamda matndagi struktur annotatsiyani (gap chegarasi, satr boshi yoki hujjat, ot yoki fe'l) taqdim qiladi. Funktsiya har bir token uchun dinamik atributlarning ma'nosini qanday hisoblashni aniqlaydi. Dinamik atributlardan foydalanishning uch turi mavjud:

1) teg atributlarini o'tkazish (teglar kodining to'liq tavsifini ifodalash);

2) atributlarning ayrim ma'lumotlarini saralash;

3) tashqi manbalar ma'lumotlarini bog'lash (masalan, tezaurus ma'lumotlar bazasi yoki morfologik analizator).

Tanlab olingan atributlar bir ma'noli bo'lmasa, atribut ko'p ma'noga ega bo'ladi. Foydalanuvchi shu kabi tokenni topish uchun mos keluvchi ma'nodan birini olish mumkin bo'ladi.

Ushbu tizim xalqaro tillarga standartlashtirilgan bo'lib, UNICODE UTF-8 kodiga tegishli turli belgilardan iborat matn ma'lumotlarni o'qiy oladi.

Bonito klient/server arxitekturasidan foydalanadi. Klient internet orqali (TCP/IP) serverga bog'laydi. Agar

internetsiz bo'lmasa, shaxsiy kompyuter lokal diskiga korpus ma'lumotlari yuklanadi. Korpusning so'rov natijalari berilgan so'rovga mos ravishda korpusdagi barcha holatlarni konkordans ro'yxati bo'lib hisoblanadi. Konkordans ro'yxati KWIK (Key Word(s) In Context) formatda ko'rinadi. Qidirilayotgan so'zlar ustma-ust shaklda ekranda namoyon bo'ladi. Konkordans ro'yxati ba'zida konkordans bo'lib qisqartiriladi. Bonitoning markaziy obyekti bu konkordansdir.

S.Y. Toldova quyidagi lingvistik masalalarni hal qilishda konkordanslardan foydalanish mumkinligini qayd etadi :[<http://www.lomonosovfund.ru/>]

u yoki bu so'zning turli foydalanish imkoniyatlari solishtirish;

kalit so'zlarni analiz qilish;

so'z va so'z birikmalarining chastotatsini o'rganish;

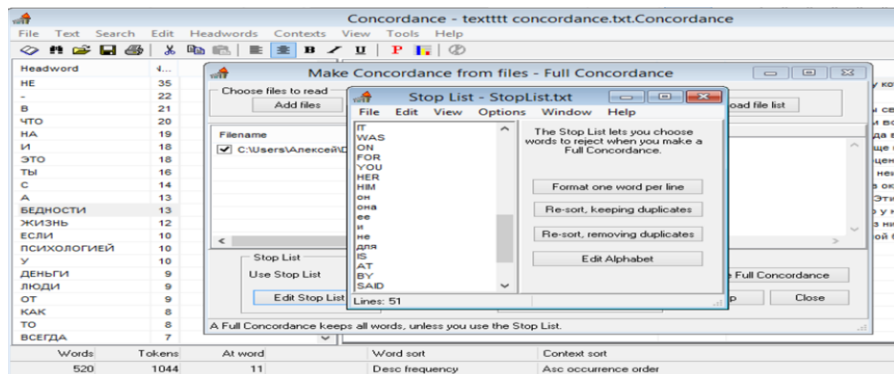
iboralarni qidirish va tadqiq qilish;

terminologiyani tarjima qilish;

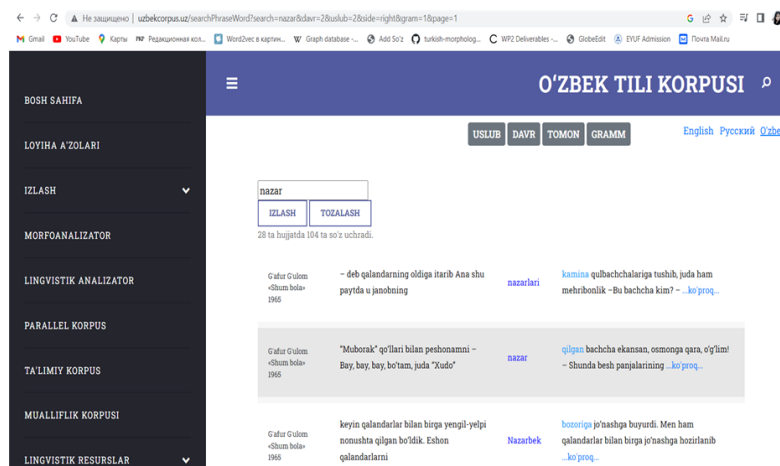
so'zlar ro'yxatini yaratish

Konkordanslardan foydalanish tilni xorijiy til sifatida o'qitishni avtomatlashtirishda (CALL – Computer Assisted Language Learning) ham faol qo'llanadi. Konkordans orqali u yoki bu so'zning o'zidan oldingi va keyingi so'zlar qurshovi orqali semantik munosabatlarini tahlil qilish imkoni yaratiladi.

CONCORDANCE.dasturiy ilovasi mustaqil ma'no anglatmaydigan so'zlarni (StopList) dastur ta'minotiga kiritish va tahrirlash mumkin.

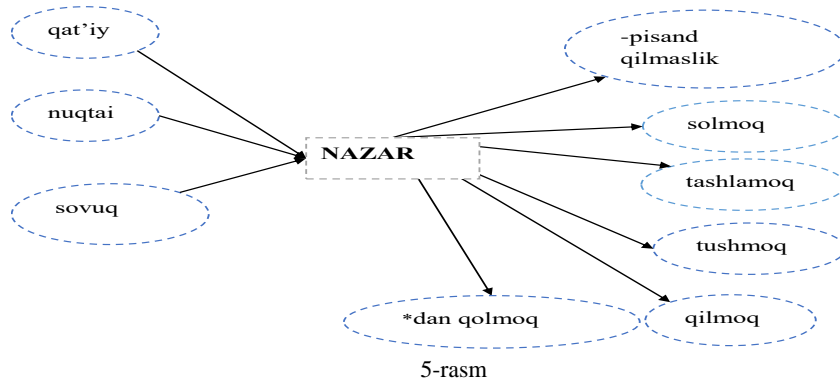


3-rasm



4-rasm

O'zbek tili elektron korpusi [<http://uzbekcorpus.uz/>] orqali konkordans instrumenti orqali u yoki bu so'zning n-gram bo'yicha so'z qurshovini aniqlash mumkin:



**Xulosa.** Xulosa qilib aytganda, konkordans tahlili korpusdagi matnlarni so'zlar, so'z qismlari yoki frazalar shaklida qidirish imkonini beradi va natijani vertikal shaklda taqdim etadi. Korpus menejerlari, masalan, Tatar tili korpusi uchun MVC konsepsiyasida ishlab chiqilgan. Korpus menejment tizimi, Manatee, lingvistik statistikani hisoblash uchun yaratilgan va xalqaro tillarga moslashtirilgan.

Konkordanslar tilni o'rganishda, kalit so'zlarni analiz qilishda, iboralarni qidirish va tadqiq qilishda keng qo'llaniladi. O'zbek tili elektron korpusida konkordans instrumenti orqali so'z qurshovini aniqlash mumkin. Konkordanslar tilni xorijiy til sifatida o'qitishda ham faol qo'llaniladi.

#### ADABIYOTLAR

1. Rychlý, P.: Corpus managers and their effective implementation. Ph.D. thesis, Faculty of Informatics, Masaryk University (2000).
2. Yergeau, F.: RFC2279: UTF-8, a transformation format of ISO 10646. Internet RFCs (1998).
3. Schulze, B.M., Christ, O.: The CQP User's Manual. (1996).
3. Kilgarriff, A., Rychlý, P., Smrž, P., Tugwell, D.: The Sketch Engine. Proceedings of Euralex (2004) 105–116.
4. Tony McEnery and Richard Xiao. What Corpora Offer in Language Teaching and Learning. 2010. – 364-373.
5. Eric D. Brill. A Corpus-Based Approach to Language Learning". 1993. – B. 1, 29, 133.
6. Abdurakhmonova N., Urdishev Q. Corpus based teaching Uzbek as a foreign language// Journal of foreign language teaching and applied linguistics, 2019
7. Abdurakhmonova N., Hayrullayeva G., Urdishev Q. Using A Corpus For Teaching Synonymy // In Uzbek Электронная письменность народов российской федерации: опыт, проблемы и перспективы Материалы II Международной научной конференции (Уфа, 11–12 декабря 2019 г.) Russian, Ufa