

O‘ZBEK TILI MILLIY KORPUSINI LOYIHALASH

Qarshiyev Abduvali Berkinovich

Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti

Samarqand filiali professori,

Fizika-matematika fanlari nomzodi,

Tursunov Muhammadsolih Sa’din o’g’li

muhammadsolih927@gmail.com

Samarqand davlat universiteti mustaqil izlanuvchisi,

Maxmidov Shirinboy Botirovich

Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti

Samarqand filiali magistranti

Annotatsiya: Hozirgi kunda dunyo mamlakatlari o‘z tilining milliy mavqeini oshirish, tilining asl holatini saqlash, uni kelajak avlodga butunligicha yetkazish dolzarb masalalardan biridir. Albatta, bunda axborot texnologiyalarining o‘rni beqiyos. Axborot texnologiyalari bilan integratsiyasi tilni raqamlashtirish, elektron shaklda saqlash va til bo‘yicha tadqiqotlar olib borishda samarali hisoblanadi. Ushbu ishda milliy korpusni loyihalashning dastlabki bosqichi keltirilgan.

Kalit so‘zlar: *o‘zbek tili milliy korpusi, dasturiy ta’minot, korpus, korpus xususiyati, model, grammatik ma’lumot.*

Zamonaviy korpuslar – bu ma’lum bir tilda, elektron shakldagi matnlar to‘plamiga asoslangan axborot-ma’lumot tizimidir. Har bir korpus filologik yondashuv bilan matnlar ustida ishlash uchun, albatta, lingvistik apparat va dasturiy ta’minot bilan ta’minlanishi lozim.

Bugungi kunda korpuslar lug‘atlar va grammatika kabi tilshunoslikning ajralmas qismiga aylandi. Korpus paydo bo‘lganidan so‘ng tilshunoslik fanlari o‘zgarib ketdi, aytish mumkinki, butun tilshunoslik korpus tilshunosligiga aylandi. Eng taniqli va tan olingan lingvistik korpuslarga namuna sifatida quyidagilarni keltirish mumkin: Rus milliy korpusi (<https://ruscorpora.ru/new/>), Britaniya milliy korpusi (<http://www.natcorp.ox.ac.uk/>, <https://www.english-corpora.org/bnc/>), Turk milliy korpusi (<https://www.tnc.org.tr/>), Amerika milliy korpusi (<http://www.anc.org/>) va boshqalar.

Tabiiyki, bunday korpuslardan juda ko‘plarini ko‘rishimiz mumkin. Shuni yodda tutish kerakki, ba’zi bir yirik korpuslar ma’lum bir vazifalarni bajaradigan, foydalanuvchilarning ma’lum bir doirasiga yo‘naltirilgan bir nechta maxsus kichik korpuslar yoki qisman korpuslarni o‘z ichiga olishi mumkin. Masalan, rus milliy korpusining asosiy tarkibida 10 ta qo‘shimcha maxsus qisman korpuslar bor.

Mavjud korpuslarni tahlil qilish natijasida ko‘rish mumkinki, korpuslarni qurishda barcha tillar uchun yagona bo‘lgan metodologiya mavjud emas. Buning



sababi shundaki turli tillarda turli xil qoidalar, texnologik jarayonlar amal qiladi. Morfologik xarakteristikalar to'plamlari turlicha bo'lishiga qaramasdan, yuqorida qarab chiqilgan korpuslarning barchasida morfologik (yoki grammatik) razmetka ta'minlangan. Sintaksis razmetka esa ba'zi korpuslarda bor, lekin turlicha yondoshuvlar asosida amalga oshirilgan bo'lsa, ba'zilarida umuman yo'q. Korpuslarni tilni qamrab olish darajasi bo'yicha solishtirganda, rus tili milliy korpusini ham xronologiya, ham janrlararo eng muvozanatlashgan va matnlar xilma-xilligi ta'minlangan korpus deb hisoblash mumkin. Lekin rus tili milliy korpusining bu yutug'i uning asosiy kamchiligini ham yuzaga keltirgan. Har qanday yirik korpusda to'liq va aniq razmetkalashning imkoniyati cheklangan bo'ladi. Noaniqliklar, asosan, avtomatik razmetkalashda omonim so'zlar tufayli paydo bo'ladi.

Korpus lingvistikasi bo'yicha tadqiqotlar shu ko'rsatadiki til tadqiqotlarini o'rganish uchun mos asos bo'lgan mashinada o'qiladigan matnlar to'plami bilan shug'ullanadi. Matnlar to'plami odatda har qanday oqilona vaqt oralig'ida faqat qo'l va ko'z bilan tahlil qilishga to'sqinlik qiladigan hajmga ega. Shu sababli, korpuslar doimo dasturiy ta'minotning qidirish vositalaridan foydalaniladi. Konkordans foydalanuvchilarga so'zlarni kontekstda ko'rish imkonini beradi. Boshqa vositalar chastota ma'lumotlarini ishlab chiqarishga imkon beradi, masalan, so'zlar chastotasi ro'yxati, unda korpusda paydo bo'ladigan barcha so'zlar ro'yxati va ularning har biri ushbu korpusda necha marta sodir bo'lishini belgilaydi. Muvofiqliklar va chastota ma'lumotlari korpus lingvistikasi uchun bir xil darajada muhim bo'lgan tahlilning ikkita shaklini, ya'ni sifat va miqdorni ko'rsatadi [1].

Mavjud korpuslar tahlilidan ko'rish mumkinki, milliy korpus tilning qonuniyat va xususiyatlarini iloji boricha ko'p taqdim eta olishi va tarkibi bo'yicha muvozanatlashgan bo'lishi kerak. Matnlarning bir turi tildagi jami matnlar ichida qanday hajmiy ulushga ega bo'lsa, ana shunday miqdor milliy korpusda ham saqlanib qolishi kerak.

Lingvistik korpuslar qo'llanilishiga quyidagi talablar qo'yilishi mumkin [2]:

1. Korpus o'quv va ma'lumot qo'llanmalar yozishda, lug'atlar tuzishda hamda lingvistik ekspertizala va tahlillarni ishlab chiqishda, til ma'lumotlari sifatida ishonchli va vakolatli manba hisoblanadi;

2. Korpus turli xil lingvistik fanlarni o'qitishda xuddi tilning ko'rgazmali (illyustrasiya) manbasi, hamda tahlillarni taqdimotlilik ko'rinishi sifatida ishlatish mumkin;

3. Korpus ilmiy tadqiqot ishlarida, shu o'rinda ilmiy farazlarni tekshiruvchi sifatida qo'llanilishi mumkin;

4. Korpus til faoliyatining universal ma'lumotlar qo'llanmasi sifatida bo'lishi mumkin.

Korpus yaratishda quyidagi savollarga asoslangan holda asosiy talablarni belgilab olish muhim:

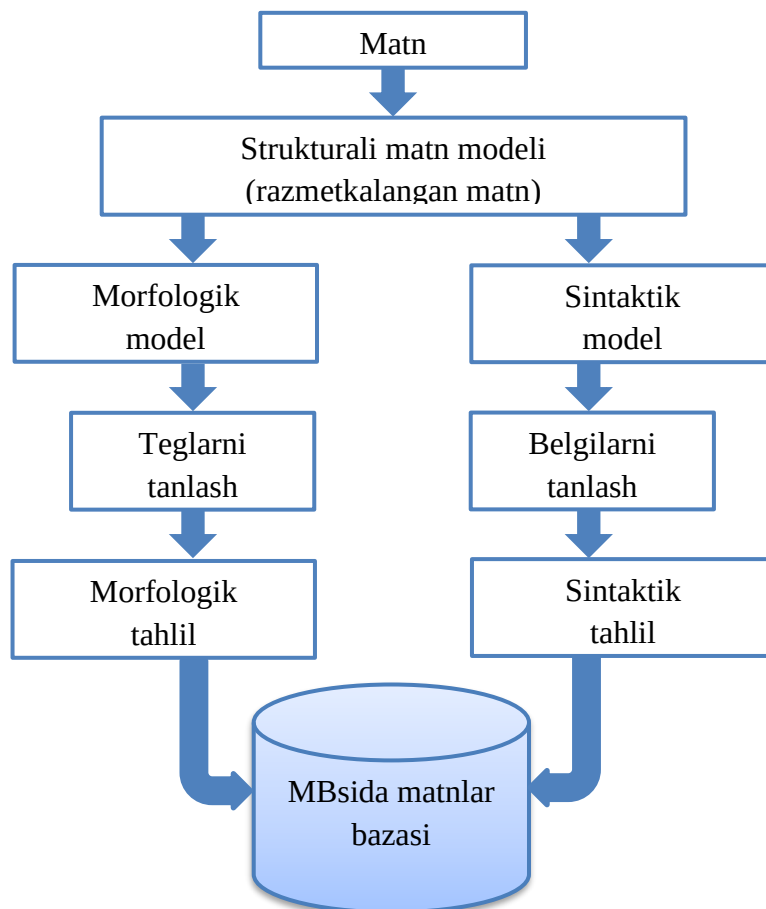


1. Korpus foydalanuvchisi kim? (Individual, guruh, lingvistik jamiyat).
2. Korpusning mantiqiy g'oyasi nimada?
3. Korpusni qurishda qancha hajmdagi ma'lumotlar bilan ishlamoqchisiz?
4. Nima ishlatiladi: matn parchalari, to'liq matnlar yoki ikkalasi ham?
5. Korpusdagi matnlarni tanlash tartibi qanday?

Korpus yaratishda turli xil razmetkaning bo'lish yoki bo'lmasligini va agar bo'lsa, uning aniqlik darajasini hisob olish maqsadga muvofiq. Korpus dasturiy ta'minoti tarkibida avtomatik razmetkalash dasturlari albatta bo'lishi kerak. Razmetkaning yuqori aniqlikda bo'lishini faqat filolog mutaxassis ta'minlashi mumkin. Filolog mutaxassis har bir matnning razmetkalanishini so'zma-so'z sinchiklab tekshirishi, avtomatik razmetka xatolarini tuzatishi va omonim holatlarni bartaraf qilishi kerak. Ishning asosiy qismi qo'l mehnati orqali bajarilishi mumkin. Shuning uchun filolog mutaxassisning kompyuterdagi ishiga mo'ljallangan maxsus razmetkalash dasturlari yaratilishi zarur. Bu dasturlar quyidagi sifatlarga ega bo'lishi kerak:

- foydalanuvchi so'zlar yoki gaplarning grammatik yoki sintaksis xarakteristikalarini yetarli darajada to'liq aniqlash va o'zgartirish imkoniyatiga ega bo'lsin;
- tahlil maksimal darajada qulay va sodda bo'lib, tez bajarilsin va bu tahlilning to'liqligini kamaytirmasin.

Korpus modelining sxematik shakli 1-rasmda tasvirlangan.



1-rasm. Korpus modelining sxematik tasviri

Foydalanilgan adabiyotlar:

4. McEnery T., Hardie A. Corpus linguistics: method, theory and practice. Cambridge University Press, 2012.
5. Баранов А. Н. Корпусная лингвистика // Баранов А. Н. Введение в прикладную лингвистику. Москва, 2001.