

**O'ZBEK KOMPYUTER LINGVISTIKASINING DOLZARB
MASALALARI VA ISTIQBOLLARI**

Siddiqova Mahliyo O'tkir qizi

*Oqdaryo tumani 40-umumta'lim maktabining ona tili va adabiyot
o'qituvchisi*

Annotatsiya

Ushbu tezisdan muallif tomonidan o'zbek kompyuter lingvistikasi sohasining bugungi kundagi dolzarb masalalari va mazkur sohaning rivojlanish istiqbollari haqida, Kompyuter lingvistikasining O'zbekistonda rivojlanishining dastlabki bosqichi 1960-1980-yillarga to'g'ri keladi. B u davrda Toshkentdagi ilmiy markazlarda dastlabki avtomatlashtirilgan lug'atlar va matnlarni kodlash ishlari olib borilgan. Asosiy maqsad esa kompyuterda o'zbek tilini kirill yozuvida ifodalash, matnlarni saqlash va qayta ishlash edi. Mustaqillikka erishganimizdan so'ng esa lotin yozuviga o'tish munosabati bilan o'zbek tilining lotin grafikasini kompyuterda standartlashtirish bo'yicha tadqiqotlar boshlandi

Kalit so'zlar: korpus, lingvistika, morfologik tahlil, qo'shma resurs, lingvistik resurs, kompyuter leksikografiyasi, qayta ishlangan resurs.

Bugungi kunda yoshlarimizni har tomonlama iqtidorli, yetuk mutaxassis qilib voyaga yetkazishda, avvalo, til haqidagi bilim va nutqiy ko'nikmalarga ega ekanligi bilan belgilanadi. "Yoshlarimizni mustaqil fikrlaydigan, yuksak intellektual va ma'naviy salohiyatga ega bo'lib, dunyo miqyosida o'z tengdoshlariga hech qaysi sohada bo'sh kelmaydigan insonlar bo'lib kamol topishi, baxtli bo'lish uchun davlatimiz va jamiyatimizning bor kuch imkoniyatlarini safarbar etamiz", - degan so'zlari biz uchun dasturulamal vazifasini o'tamoqda¹. Bugungi kunda o'zbek tilshunosligida jadal rivojlanib borayotgan yo'nalishlaridan biri, shubhasiz, kompyuter lingvistikasidir. Korpus lingvistikasining dastlabki bosqichlari Noam Chomskiy tomonidan ilgari surgan

¹ Mirziyoyev Sh.M. Erkin va farovon demokratik O'zbekiston davlatini birgalikda barpo etamiz. -

T:O'zbekiston, 2016. - B.14.

<https://scientific-jl.com/>

generativ mulohazalarida namoyon bo'la boshlagan bo'lsa-da, matnlarni amaliy jihatdan o'qiy oladigan shaklda yig'ib uni kompyuter tahlil qila oladigan darajaga darajadagi bosqich kompyuter lingvistikasi uchun muhim davr bo'ldi. Korpus lingvistikasi atamasidan oldin kompyuter korpusi termini qo'llanganligi haqidagi ma'lumotlar uchraydi. Grame Kennedi kompyuter lingvistikasining alohida bir soha bo'lib shakllanishida LOB korpusi, Brown korpuslarini kompyuter lingvistikasi shakllanishi uchun dastlabki avlod deya e'tirof etgan. Amerikalik olim Georg Zif matnda qo'llangan so'zlar chastotasi bilan matn uzunligi o'rtasidagi munosabatlarga asoslangan Zif qonunini yaratgan. Mazkur yo'nalishda o'zbek tilini kompyuter orqali qayta ishlash, tilshunoslik bilimlarini axborot texnologiyalari bilan uyg'unlashtirishdan iboratdir. Mazkur yo'nalishning asosiy maqsadi esa o'zbek tilida matnlarni avtomatik tahlil qilish, tarjima qilish, matn yaratish va nutqni tanish tizimlarni ishlab chiqishdir. Mazkur soha yurtimizda rivojlanayotgan yosh soha hisoblanadi, ammo jadal ravishda o'sib borayotgan yo'nalishdir. Kompyuter lingvistikasining O'zbekistonda rivojlanishining dastlabki bosqichi 1960-1980-yillarga to'g'ri keladi. Bu davrda Toshkentdagi ilmiy markazlarda dastlabki avtomatlashtirilgan lug'atlar va matnlarni kodlash ishlari olib borilgan. Asosiy maqsad esa kompyuterda o'zbek tilini kirill yozuvida ifodalash, matnlarni saqlash va qayta ishlash edi. Mustaqillika erishganimizdan so'ng esa lotin yozuviga o'tish munosabati bilan o'zbek tilining lotin grafikasini kompyuterda standartlashtirish bo'yicha tadqiqotlar boshlandi. Bu davrda birinchi elektron lug'atlar yartildi, o'zbek tilida dasturlash uchun klaviatura va kodlash standartlari ishlab chiqildi. 2010-2020-yillar oralig'ida sun'iy intellect yordamida imlo tekshirish, plagiat aniqlash dasturlari paydo bo'ldi. Google translate va Yandex translateda o'zbek tili paydo bo'ldi, o'zbek tilining milliy korpusi ustida tadqiqotlar yaratila boshlandi. O'zbek tilini qo'llab –quvvatlaydiga chatbootlar paydo bo'ldi, o'quv platformalar va tilni o'rgatuvchi ilovalar ko'paymoqda. Bugungi kunda O'zbekiston Fanlar Akademiyasi, Toshkent Axborot Texnologiyalari Universiteti va boshqa bir qator oily o'quv yurtlarida mazkur sohada izlanishlarn olib borilmoqda. O'zbek tilining electron korpuslari

yaratilmoqda, Matnlarni avtomatik tahlil qiluvchi dasturlar, elektron lug'atlar, ta'limiy platformalar rivojlanmoqda.

Kompyuter lingvistikasida korpus muhim ahamiyatga ega. Xususan, elektron lug'atlar bazasi, shuningdek, grammatik qoidalarni ishlab chiqishda korpusdan foydalaniladi. Korpus lingvistikasining mashina tarjimai, sentiment tahlil, matn tahlili va boshqa bir qator yo'nqalishlar uchun muhim sanaladi. Lingvistik ma'lumotlarni tahlil qilish hamda ularni boshqarish maqsadida bir qator darajadan iborat tizim yaratiladi. Resurslar 3 turga bo'linadi:

1. Qo'shma resurslar: bunday resurslar kirish metodi, chiqish metodi, standart ifodalarni o'z ichiga oladi.

2. Lingvistik resurslar: tilga qarashli bo'lgan ma'lumotlar bazasidan tashkil topgan bo'lib korpuslar, lug'at, lingvistik analizatorlar kiradi. Lingvistik resurslar omlin shaklda yoki avtonom ko'rinishda bo'lishi mumkin.

3. Qayta ishlangan resurslar: bu resurslar nutqni o'quvchi tizim, grammatikani tekshiruvchi vositalar, so'zlarni oldindan topish dasturi, qidiruv tizimi, imloni tekshiruvchi vositalar, sintaktik analizatorlardan iborat bo'ladi.

90- yillarda morfologik tahlil generativ model, paradigmatic model, ikki tarkibli morfologik model ² singari metodlarga asoslanadi. Turkiy tillar borasida morfologik analizator yaratish masalash XX asrning 60-yillaridan boshlandi. O'sha vaqtdan boshlab bir qator tadqiqotlar yaratildi. Masalan, tatar tilining morfologik analizatori uch modelga asoslanadi: generativ, paradigmatic va ikki bosqichli model. Morfologik tahlil sintaktik tahlil uchun muhim ahamiyatga ega. Chunki avtomat usulda matnni komponentlarga bo'lishda u yoki bu so'zning qaysi so'z turkumida ekanligini aniqlash lozim. Dastlabki morfologik (So'z turkumlari bo'yicha- Part of speech so'zidan olingan) annotatsiyalash Britaniya milliy korpusida foydalanilgan.

² Mahlov C., Piotrowski M. JSLM – COMPUTATION Morphology in the framework of the SLIM Theory of language / State of the art Computation Morphology. Zurich, -P.15

<https://scientific-jl.com/>

Shuningdek, Penn sintaktik korpusi, nemis tilining Negra korpusidagi teggerlar shular jumlasiga kiradi. Morfologik annotatsiya uchun ushbu korpuslarda standart teglardan foydalanilgan. Biroq har bir til o'zining imkoniyatidan kelib chiqib, ixtiyoriy teglarning to'plamlaridan foydalanadi.

O'zbek tilshunosligida o'zbek tilining morfologik annotatsiyasini yaratish maqsadida chekli avtomat transyuteridan foydalanildi va uning avtomatik morfotaktik bazasi hosil qilindi. Ikkinchi bosqichda esa morfotaktik ma'lumotlar tahrirlandi, uchinchi bosqichda yig'ilgan korpus asosida amalga oshirilgan morfologik analizator tekshirildi. Morfologik tahlilda bir qator fonetik hodisaga uchragan so'zlarimiz mavjudki, bunday so'zlarni tahlil qilishda ham bir qator yechim va algoritmlar mavjud. Shu bilan bir qatorda o'zbek tili leksikografiyasida ham bir qator yechimini kutayotgan masalalar mavjud. Xususan ma'lumot bazasining yetarli emasligi, standartlashtirish muammosi, ko'p tillilik muammosi, ya'ni o'zbekcha-ruscha, o'zbekcha-inglizcha lug'atlar mavjud, ammo boshqa tillar bilan lug'at borasida muammolar bor. Dasturiy interfeyslar ko'p hollarda foydalanuvchi uchun u qadar qulay shaklda emas. Yaqin yillar oralig'ida kompyuter lingvistikasi sohasi oldida turgan bbir qator vazifalar mavjud. Avvalo milliy til korpusini yaratmoq lozim. Zamonaviy morfologik va sintaktik analizatorlar yaratmoq kerak.

So'z turkumlarini teglashning dastlabki shakli Lankasteruniversitetining tadqiqot markazida o'tgan asrning 80-yillarida Jeorey Rich rahbarligida ishlab chiqilgan. Ingliz tili uchun CLAWS avtoamat tegi ishlab chiqilgan. Korpusning morfologik annotatsiyalashdan tashqari fonetik, semantik, pragmatic, diskurs, uslubiy, leksik kabi turlari ham mavjud. Korpusning nechog'lik lingvistik jihatdan annotatsiyalanganligi til texnologiyasida matn bilan bog'liq boshqa lingvistik instrumentlarni yaratishda qo'l keladi.

Xulosa qilib aytadigan bo'lsak o'zbek kompyuter lingvistikasining oldida bir qator yechimini kutayotgan masalalar mavjud bo'lib: morfologik va

sintaktik tahlilning to'liq avtomatlashtirilmaganligi, katta hajmdagi annotatsiyalangan matn bazalarining yetishmasligi kabi masalalardir.

Foydalanilgan adabiyotlar ro'yxati

1. Abduraxmonova N. Zamonaviy lingvistik korpuslarning kompyuter modellari. "Foreign Languages in Uzbekistan " , 2020, No 1(30), 50-58 <https://journal.fledu.uz> ISSN: 2181-8215(online)

2. A. Eshmuminov. O'zbek tili milliy korpusining sinonim so'zlar bazasi. Filol.fan.bo'yicha falsafa doktori (Phd)... dis.avtoref..- Toshkent, 2019

3. Пулатов , А.Қ. Компьютер лингвистикаси. – Тошкент: Akademnashr, 2011.-500 б;

4. Пулатов, А. Қ. Компьютер лингвистикаси / Масъул муҳаррирлар; А.А. Абдуазизов, М.М. Орипов. – Т.: Akademnashr, 2011. – 520 б. – Б.