



INGLIZ VA O'ZBEK TILLARIDA KOMPYUTER LINGVISTIKASINING LEKSIKOGRAFIK MUAMMOLARI

Bahriddinova Yulduzxon Bahriddin qizi

TerDU stajyor o'qituvchisi

yulduz.bahriddinqizi@gmail.com

<https://www.doi.org/10.37547/ejsspc-v03-i02-p1-26>

ARTICLE INFO

Received: 31th January 2023

Accepted: 09th February 2023

Online: 10th February 2023

KEY WORDS

Kompyuter lingvistikasi,
kompyuter leksikografiyasi,
raqamlashtirish inqilobi, tabiiy
til jarayoni (NLP).

ABSTRACT

Ushbu maqolada ingliz va o'zbek tilshunosligi sohalaridan biri kompyuter lingvistikasi va uning leksikografiyasi haqida, ushbu sohaning tarixi, rivojlanishidagi avzalliklar, hamda yuzaga keladigan muammolar haqida mulohaza yuritilgan. Bundan tashqari raqamli texnologiyaning kompyuter leksikografiyasi paydo bo'lishiga turtki bo'luvchi zarurlarning paydo bo'lishi va ularning yechimlari haqida qisqacha to'xtalib o'tilgan.

So'nggi yillarda axborot texnologiyalar va shunga oid bo'lgan raqamlashtirish jamiyatning har bir jabhasiga faol kirib boryapti, va albatta, ilm-fan sohalari ham bundan mustasno emas. Xususan uning tilshunoslikka ta'siri va shu asosida unda yangi sohalarning paydo bo'lishi uning hozirgi kundagi dolzarbligini yana bir bor oshiradi. Tilshunoslikning kompyuter lingvistikasi yo'nalishi buning yaqqol isbotidir.

Kompyuter lingvistikasi matematik lingvistikaning mantiqiy davomi bo'lib, u amaliy tilshunoslikning eng muhim qismini tashkil etadi. Kompyuter lingvistikasi 1954 yil AQSHda Jorjtaun universitetida mashina tarjimasini bo'yicha dunyoda o'tkazilgan birinchi tajriba asnosida yo'nalish sifatida shakllana boshladi, 1960- yilga kelib mustaqil fan sifatida shakllandi. Kompyuter lingvistikasi inglizcha "computational linguistics" so'zining kalkasidir. XX asrning 80-yillariga qadar bu fan turlicha nomlar bilan atalgan: hisoblash lingvistikasi, matematik lingvistika, kvantitativ lingvistika, injener lingvistikasi kabi. Bu fanning asosiy maqsadi lingvistik masalalarni yechishning kompyuter dasturlarini ishlab chiqish, inson va mashina (kompyuter) muloqotini optimallashtirish, tabiiy tilni qayta ishlash (Natural Language Processing) dir. [1] Ma'lumki, tevarak-atrofdagi ma'lumot, axborot ko'p. Inson uchun bu ma'lumotlarning barchasini to'plashi va o'z xotirasida saqlashi juda mushkul ish. Shunga ko'ra inson o'ziga kerakli bo'lgan axborotlarni yig'ish va ularga ishlov berishning yo'llarini kidirishga intiladi. Kompyuter qurilmalari bu sohada insonga katta yordam bermoqda. Hozirgi vaqtda kompyuter yordamida tilshunos olimlar so'zlarni bir joyga to'plash, ularga qayta ishlov berish, shu asosda turli xil lug'atlar, tezauruslar, rubrikatorlar, glossariylar, so'zlashgichlar yaratish bilan shug'ullanmoqdalar. Umuman, atamalarni to'plash va ularga qayta ishlov berish jarayonida kelib chiqqan katta amaliy ehtiyoj tilshunoslikda kompyuter texnologiyasidan kengroq foydalanish imkoniyatlarini vujudga keltirdi. Hozirgi kunga kelib fan-texnika inqilobi



natijasida tinmay yog'ilib turgan materiallar, axborotlar oqimini faqat kompyuter texnologiyasi yordamidagina boshqarish mumkin bo'lib qoldi.

Kompyuter lingvistikasining eng dolzarb yo'nalishlaridan biri bu kompyuter leksikografiyasidir. Kompyuter leksikografiyasi kompyuter orqali lug'at tuzish, lug'at tuzish tamoyillarini, lug'atshunoslik ishlarini avtomatlashtirish va ixchamlashtirish, bibliografik ma'lumotlarni avtomatik yo'l bilan qidirish va ularga qayta ishlov berish kabi masalalarni o'rganadi.[2] Zamonaviy tilshunoslikda kompyuter leksikografiyasini alohida soha sifatida ajratib o'rganish tavsiya etiladi, bu esa elektron lug'atlar tayyorlashda kompyuter texnologiyalaridan foydalanish usullarini o'rganishga xizmat qiladi. Aytish joizki, axborot texnologiyalari tilshunoslikda XX asrning 60-70-yillaridan boshlab qo'llanila boshlandi: kompyuterlar turli lug'atlar tayyorlash, shuningdek, eng oddiy statistik masalalarni yechish va ish materiallarini saralash uchun foydalanilgan. [3]. Ingliz tilshunosligida leksikografiya shakllanishi uchta katta revolyutsiyani o'z ichiga oladi:

1. Tasvirlovchi misollar revolyutsiyasi (The descriptive Paradigm)
2. Korpus revolyutsiyasi (The Corpus Revolution)
3. Raqamli revolyutsiya (The Digital Revolution)

Leksikografiyani raqamlashtirishning birinchi loyihasi 1980-yilda Shvetsiya va Birlashgan Qirollik davlatlarida amalga oshirildi [4]. Agar korpus inqilobining ta'siri asosan leksikografiyaning ichki tuzilishiga ta'sir ko'rsatadigan bo'lsa, raqamli inqilob nafaqat lug'at foydalanuvchilariga, balki har birimizning kundalik hayotimizga ta'sir qilibgina qolmay, lug'atlarni yaratish uslubimizni ham o'zgartirdi.

Qog'oz lug'atlar davrida leksikograflar tomonidan yozilgan va tahrirlangan tayyor manba o'rtasida yaqin aloqa mavjud edi. Bosma asar asosiy o'rinda bo'lib, leksikograf yaratgan qo'lyozma qog'oz shaklda, matnni qayta ishlash dasturida yoki ma'lumotlar bazasida yozilgan bo'lishidan qat'i nazar, jarayonning boshlang'ich bosqichi bo'lgan. Biz ma'lumotlar bazasi ishning markaziy elementi ekanligini va ma'lumotlar bazasi tuzilmasi turli ommaviy axborot vositalarida va turli platformalarda nashr etish uchun yaxshi tashkil etilgan va yetarlicha moslashuvchan bo'lishi kerakligidan ko'proq xabardormiz. Raqamli inqilob lug'atlar va lug'at yaratishni turli yo'llar bilan o'zgartirdi: biznes modeli, takomillashtirilgan qidiruv imkoniyatlari va til o'rganuvchilarga, ishonchsiz imlochilarga, foydalanuvchilar-ning ishtiroki, omma manbalari va hokazolar. So'nggi yillarda leksikografiya va tabiiy til jarayoni (NLP) hamjamiyati bir-biriga ikkita sababga ko'ra yaqinlashdi:

- Internet va veb-saytlarni ulash uchun giperhavolalardan foydalanish raqamli rivojlanish zaruratini tug'dirdi;
- Internetda ma'lumot izlash foydalanuvchilar sabrsiz bo'lishlariga sabab bo'ldi. Ular o'z savollariga tezda javob olishni xohlashadi. Lug'atlar endi nashriyot kompaniyalari uchun muhimlik darajasini yo'qotdi.

Bir tomondan, mavjud lug'at provayderlar ma'lumotlar tuzilmasi va qidiruv vositalarini yaxshilagan holda bosma nashrdan ekranga o'tishdi; Ehtimol, foydalanuvchilar foydalanadi degan ma'noda audio va videokliplar ko'rinishidagi yangi ma'lumotlar qo'shilgan yangi manbalar ham yaratilayapti. Boshqa tomondan, ularning harakatlari cheklangan muvaffaqiyatga erishganga o'xshaydi chunki ba'zi odamlar leksikografiyaning bunday o'zgarishlarini o'z baholashlarida salbiy holatda baholaydi: leksikografiyaning eng katta



muammosi leksikografik manbalarni endi foydalanuvchilar yetarli darajada idrok etmasdan shunchaki o'ziga kerak bo'lgan joyini elektron lug'atdan qidiruv berib ko'radi va bu so'z yoki jumlaning boshqa ma'nolari bilan umuman qiziqmaydi. Aksariyat foydalanuvchilar esa, aslida, lug'atlardan umuman foydalanishga ehtiyoj topmaydi va foydalanmaydilar, uning o'rniga ular muloqot qilish yoki ma'lumot qidirishda yordam kerak bo'lsa, shunchaki internetdan foydalanadilar.

O'zbek tilshunosligida ham elektron lug'atlarni yaratish bo'yicha bir qancha ishlar olib borilgan, jumladan, 1998-2001 yillar davomida mamlakatimizda bir guruh tilshunolar "Hozirgi o'zbek tili faol so'zlarining izohli lug'ati"ni tuzish jarayonida kompyuter texnologiyasidan unumli foydalandilar. Ular adabiy tilimiz materiallari ko'proq ijtimoiy-siyosiy matnlarda ifodalanishini inobatga olib, "Yoshlik", "Tafakkur", "Guliston", "Sharq yulduzi" kabi jurnallar va qator gazetalar sahifalarida qo'llangan barcha so'zlar maxsus dastur yordamida 94 ta fayl asosida sahifalab chiqildi. Ana shu fayllarda ajratib olingan so'zlar jami 8168 sahifani tashkil etdi.[5]

Zamonaviy axborot texnologiyalari barcha leksikografik kompyuter dasturlarini shartli ravishda guruhlariga bo'lish imkonini beradi: 1) leksikografik ma'lumotlarni yaratish va ularning ishlashini ta'minlash uchun zarur bo'lgan dasturlar (ular leksikografik ma'lumotlarni saqlash dasturlari deb ataladi); 2) elektron lug'atlar va leksikografik ma'lumotlar bazalari. Kompyuterda yaratilgan barcha lug'atlarni ikki guruhga bo'lish mumkin:

- 1) bosma lug'atlar elektron versiyalar ko'rinishida taqdim etiladi;
- 2) elektron format uchun ishlab chiqilgan lug'atlar bo'lib, ular kompyuter dasturi tomonidan taqdim etilgan qo'shimcha funksiyalar bilan tavsiflanadi.[3,4]

Tadqiqotchilar elektron va an'anaviy, ya'ni bosma materiallar o'rtasidagi gibridlanish tamoyillarini aniqlaydilar:

- 1) o'zgarmas lug'at materiallari va katta lug'aviy bazasi bo'lgan ko'p jildli qog'oz lug'atlardan foydalanishda noqulayliklar tug'diradi (masalan, ular harakatchan emas va axborotni tez olishda qiyinchilik tug'diradi);
- 2) lug'atga bosma formatda o'zgartirishlar kiritish va uni o'zgartirish jarayoni ko'p vaqtni va ish kuchini talab qiladi.[3,8]

Kompyuter leksikografiyasi amaliy tilshunoslik bo'limining zamon talablari asosida yangi va o'ziga xos yo'nalishi sifatid a bugungi kun tilshunoslari tomonidan turli elektron lug'atlarni yaratish va ulardan foydalanish tamoyillarini ishlab chiqishda alohida ahamiyatga ega.

References:

1. A.Rahimov. Kompyuter lingvistikasi asoslari fanidan o'quv qo'llanma. – T, 2011. - B 7.
2. <https://docs.yandex.ru/docs/view?tm=1675875969&tld=ru&name=rahimov%20a.kompyuterlingvistikasiasoslariarahimovpdf>.
3. D.B.O'rinbayeva. Kompyuter lingvistikasidan universal qo'llanma. – Samarqand, 2016. – B 8.
4. Sherqo'zi Qulbulov. Kompyuter leksikografiyasi tilshunoslik sohasi sifatida. https://uza.uz/ru/posts/kompyuter-leksikografiyasi-tilshunoslik-bolimi-sifatida_389299



5. Lars Trap-Jensen. Lexicography between NLP and Linguistics: Aspects of Theory and Practice. Lexicography in Global Contexts. – D,2018. –P 31.
<https://euralex.org/wpcontent/themes/euralex/proceedings/Euralex%202018/118-4-2949-1-10-20180820.pdf>
6. <https://elziyouz.blogspot.com/p/ozbekiston-respublikasi-oliy-va-orta.html>