

TILSHUNOSLIKDA LINGVISTIK BAZALAR VA ULARNING AHAMIYATI

*Nizomiddinov Obidxon,
Termiz davlat universiteti magistranti*

Annotatsiya. Ushbu maqolada tilshunoslikda lingvistik bazalar tuzish hamda ularning ahamiyati xususida soʻz yuritiladi. Bugungi kunda yaratilayotgan korpuslar, dunyodagi mashhur lingvistik bazalar va korpuslar, lingvistik baza tuzish, oʻzbek tilida lingvistik bazalarning yuzaga kelishi va hozirgi kundagi taraqqiyoti toʻgʻrisidagi fikrlar tahlilga tortiladi.

Kalit soʻzlar: tilshunoslik, lingvistik baza, ahamiyat, dunyo tilshunosligi, korpus, mashhur baza va korpus, oʻzbek tilidagi lingvistik bazalar.

Abstract. This article talks about the creation of linguistic bases and their importance in linguistics. Today's created corpora, world-famous linguistic bases and corpora, creation of a linguistic base, emergence of linguistic bases in Uzbek language and their current development are analyzed.

Keywords: linguistics, linguistic base, significance, world linguistics, corpus, famous base and corpus, linguistic bases in Uzbek language.

Абстрактный. В данной статье говорится о создании языковых основ и их значении в языкознании. Анализируются созданные сегодня корпуса, всемирно известные лингвистические базы и корпуса, создание лингвистической базы, появление лингвистических основ в узбекском языке и их современное развитие.

Ключевые слова: языкознание, языковая база, значение, мировая лингвистика, корпус, известная база и корпус, языковые основы в узбекском языке.

Barchamizga ma'lumki, bugungi kunda tilshunoslikning kompyuter lingvistikasi tarmogʻi juda ham taraqqiy etib kelmoqda. Sababi, bugungi texnika va texnologiyalar asrida til sohasining ham kompyuter bazasidan oʻrin olishi davr talabiga aylanib ulgurdi. Tillarning kelgusi taraqqiyoti ham ushbu sohaga bevosita aloqador desak, aslo mubolagʻa boʻlmaydi. Kompyuter lingvistikasining tillar taqdiriga taʼsiri benihoya ekan, uning tarkibidagi sohalarning ham ahamiyati juda ham yuqoriligi ham aslo sir emas. Alalxusus, lingvistik baza va korpuslar shular jumlasidandir. Bizning ushbu maqolamizda lingvistik bazalar, korpuslar va ularning tilshunoslikdagi ahamiyati xususida fikr yuritiladi.

Lingvistik ma'lumotlar bazasi (LMB) ostida mos ravishda tizimlashtirilgan lingvistik ma'lumotlar to'plami tushuniladi hamda u kompyuter lingvistikasi fanining asosiy bazasi hisoblanadi. Korpus esa ana shu lingvistik bazaning yaratilishidagi muhim poydevor sanaladi. Korpus tilning barcha jihatlarini qamrab olsa, lingvistik baza esa korpus asosida yuzaga keladi. Dunyo miqyosida yaratilgan qator lingvistik bazalar mavjud boʻlib, ularning yaratilishi turli davrlarga borib taqaladi. Xususan, rus tilida mavjud asosiy lingvistik bazalarning yaratilishi XX

asr oxiri – XXI asr boshlariga to'g'ri kelgan hamda ularni qaysi mavzu asosida yaratishda leksikografik manba sifatida foydalanish amaliyoti hisobga olindi[3]. N.Y. Shvedova ishlab chiqilgan LMB ning mazmuni rus semantik lug'atining birinchi jildi misollari asosida ekanligi borasida o'z fikrlarini keltirgan edi. Tadqiqot uchun material sifatida ko'rsatilgan lug'atdan doimiy tanlab olish usuli bilan tanlangan shaxs va shaxslar guruhining nominatsiyasi ma'nosiga ega otlar va salbiy-baholash semantikasi bo'ldi. Material mafkuraviy, semantik va uslubiy tamoyillar asosida tanlangan[5].

Ma'lumotlar bazasi texnologiyasi turli sohalarda, jumladan, tilshunoslikda, masalan, "tipologik va qiyosiy, fonetik, leksik, grammatik tadqiqotlarda; leksikografiya; semantik tadqiqotlar, shu jumladan, tezauruslarni tuzish; matematik tilshunoslikda va amaliy lingvistik muammolarni hal qilish uchun (lingvistik didaktik, avtomatlashtirilgan tarjima, avtomatik tanib olish va nutq sintezi) foydalaniladi. Mishankina gumanitar fanlarda ma'lumotlar bazasi texnologiyasiga asoslangan axborot tizimlarini yaratish va ma'lumotlarni qayta ishlash bilan bog'liq muammolarni ko'rib chiqadi, xususan, ushbu texnologiyadan filologik (lingvistik) tadqiqotlar sohasida foydalanish bilan bog'liq ekanligi xususida fikr yuritadi[9]. Tadqiqotchi mavjud ma'lumotlar bazalari haqida tarixiy ma'lumot beradi va LMBni alohida toifaga ajratib, uni ikki turga bo'lishni taklif qiladi:

- I. To'liq matnli (to'liq matnlarni o'z ichiga olgan).
- II. Faktografik (turli til darajasidagi lingvistik birliklarni o'z ichiga olgan)[11].

To'liq matnli LMBga misol sifatida biz Tomsk davlat universiteti jamoasi (N.A. Mishankina, I.V. Tubalova, Y.A. Emer) tomonidan ishlab chiqilgan "O'rta Sibir folklori" ma'lumotlar bazasini keltirishimiz mumkin. Ushbu ma'lumotlar bazasini yaratishdan maqsad Sibir xalq madaniyatining ilgari mavjud bo'lmagan folklor materialini antropologiya, etnografiya, madaniyatshunoslik, tilshunoslik va Sibir mintaqasi tarixi sohasidagi turli tadqiqotlar uchun qulay va munosib formatda birlashtirish va taqdim etishdir. Bu ma'lumotlar bazasiga rus tilidagi turli janrdagi xalq og'zaki ijodi matnlari, masalan, qo'shiqlar, topishmoqlar, maqollar, olmoshlar, til o'girishlari kiritilgan[13].

Shuningdek, boshqa adabiyotlarda lingvistik bazaning boshqa turlari ham alohida qayd etib o'tiladi. Uning birinchi turi, shuningdek, elektron kutubxonalar, matn to'plamlari va matn korpusini o'z ichiga oladi. Ikkinchi turdagi lingvistik bazalar orasida ulardagi til tavsifi darajasiga ko'ra, mos ravishda fonetik, morfologik, leksik va sintaktik lingvistik bazalarni ajratib ko'rsatish mumkin[4].

Leksik birliklar taqdim etilgan LMBlar, ayniqsa, an'anaviy va elektron lug'atlarni tuzish uchun leksikografik amaliyotda faol qo'llaniladi. Zamonaviy kompyuter texnologiyalari leksikografik ma'lumotlarni to'plash va saqlash jarayonini soddalashtirishga imkon beradi, yozuvlari an'anaviy katalog kartasiga o'xshash oddiy fayl o'rniga ma'lumotlar bazasidan foydalanadi. Biroq, leksik LMBlar nafaqat yangi lug'atlarni yaratishda, balki turli xil lingvistik tadqiqotlarda ham qo'llaniladi, chunki LMB leksikografik ma'lumotlarni taqdim etishning

mustaqil formati sifatida lug'atlarga nisbatan bir qator quyidagi afzalliklarni taqdim etadi[8]:

- to'plash;
- saqlash;
- lingvistik ma'lumotlarni tizimlashtirish va uning ichida tezkor qidiruvni amalga oshirish;
- shuningdek, til tizimida doimiy ravishda sodir bo'ladigan o'zgarishlarni amalga oshirish uchun ochiq tuzilishga ega [11].

Dunyo tilshunosligida xilma-xil lingvistik bazalar yaratilgan bo'lib, ularga misol sifatida quyidagilarni keltirishimiz mumkin:

"Ovoz" lingvistik bazasi (N.A. Mishankina, Tomsk davlat universiteti) semantikasida "tovush" komponentini o'z ichiga olgan fe'llarni qamrab oladi, "ovoz" leksik-semantik maydonini aks ettiradi va turini ko'rsatadigan audio yoki vizual demomateriallarni to'ldirish imkoniyatini beradi. Tovush, uning predmeti va u bilan bog'liq harakatlarni jamlagan ma'lumotlarni o'zida aks ettiradi.

Metaforik terminologiya lingvistik bazasi ilmiy nutqning kaliti bo'lgan konseptual metaforalarni aks ettiruvchi universal metafora modellarini aniqlashga qaratilgan baza hisoblanadi. Mualliflar (N.A. Mishankina, E.A. Panasenko, Tomsk politexnika universiteti) ishlab chiqilgan ma'lumotlar bazasidan bir qator muayyan muammolarni hal qilishda, masalan, milliy tilning atama yaratishda eng mashhur leksik birliklarini aniqlashda foydalanish mumkinligini ta'kidlaydilar. Ushbu ma'lumotlar bazasini yaratish uchun materiallar turli xil ilmiy sohalarning terminologik lug'atlariga asoslangan edi: biologiya, geologiya, informatika, tilshunoslik, tibbiyot, psixologiya, sotsiologiya, falsafa, ekologiya. Tahlil qilingan manbalarning umumiy soni 65 tani, aniqlangan metaforik terminologik birliklar soni esa 2000 ga yaqinni tashkil etdi. Ushbu ma'lumotlar bazasining markaziy jadvali "Leksik birlik", yordamchi jadvallari "Metaforik model" va mavzu sohalari jadvallari muayyan predmet sohasining metaforik terminologiyasidan iboratdir [12].

Psixolingvistik lingvistik bazasi (Tomsk davlat universiteti loyihasi: Z.I.Rezanova, A.A.Miklashevskiy) lingvistik semantikaning obrazli-idrok komponentini tavsiflash uchun ishlatiladi va yarim tizimli semantik ma'lumotlarni tahlil qilish uchun statistik usullardan foydalanishga imkon beradi. Ushbu lingvistik bazada psixolingvistik eksperiment davomida olingan 506 aniq va mavhum otlar haqida ma'lumotlar, ya'ni ularning idrok etish usullari (ko'rish, eshitish, ta'm, hid va teginish hissi) bilan aloqasi va boshqa qo'shimcha ma'lumotlar, masalan, so'zni o'rganishning subyektiv ahamiyati xususidagi ma'lumotlar o'rin olgandir [13].

Pragmatik belgilangan lug'at lingvistik bazasi Novosibirsk davlat pedagogika universiteti jamoasi tomonidan ishlab chiqilmoqda: T.A. Tripolskaya, E.Y. Bulygina, V.D. Chernyak (Sankt-Peterburg), V.A. Efremov (Sankt-Peterburg), M.A. Lappo, E.G. Basalaeva, L.N. Xramtsova, I.I. Sazhenin. Taxminan 400 birlikni o'z ichiga olgan ushbu ma'lumotlar bazasining demo versiyasi Internetda

mavjud. [14, 15] Keltirilgan tavsifga ko'ra, ma'lumotlar bazasi tavsifi rasmiy va mazmunli ravishda ikki qismga bo'lingan. Birinchi qismda rus tilining quyidagi klassik va zamonaviy lug'atlarining lug'at yozuvlari keltirilgan: "Rus tilining izohli lug'ati". D.N. Ushakov (1935–1940), 17 jildli zamonaviy rus adabiy tilining lug'ati (1950,1965), 4 jildli rus tilining lug'ati, A.P. Evgenyeva (1981-1984), "Rus tilining izohli lug'ati" S.I. Ozhegova va N.Yu. Shvedova (1999), "XX asr oxirida rus tilining izohli lug'ati. G.N. Sklyarevskaya (1998), "Rus tilining katta tushuntirish lug'ati", S.A. Kuznetsova (1998). Ma'lumotlar bazasining ikkinchi qismi mualliflar tomonidan ishlab chiqilgan "Pragmatik belgilar lug'ati" lug'at yozuvlari asosida tuzilgan ushbu lingvistik baza bugungi kunda juda muhim ahamiyat kasb etadi.

Hozirgi kunda lingvistik baza borasida o'tkazilgan tahlillar ko'rib chiqilayotgan LMBlarning har biri ma'lum bir tadqiqot muammosini hal qilish uchun yaratilgan degan xulosaga kelishga imkon beradi va shuning uchun ular bir qator mezonlar bo'yicha farqlanadi, buning asosida LMB ning quyidagi kichik turlarini ajratish mumkin:

- leksik birliklarning turiga ko'ra, ya'ni ularning qismiga mansubligi bo'yicha: fe'llarni o'z ichiga olgan LMB ("Ovoz" LB), otlar (psixolingvistik LB, pragmatik belgilangan lug'at ma'lumotlar bazasi), nutqning bir nechta qismlari birliklari (lug'at LB); shuningdek, ulardan foydalanish sohasiga ko'ra: maxsus lug'atni o'z ichiga olgan LMB (metaforik terminologiya ma'lumotlar bazasi);

- leksik semantikasi tavsiflash jihati nuqtayi nazaridan: "tovush" (LB "Sound") semantik komponentini tavsiflovchi LB, metaforik semantika (metaforik terminologiya MB), obrazli-idrok komponenti (psixolingvistik LB), pragmatik makrokomponent, (MB pragmatik belgilangan lug'at), dinamik jihatdagi me'yoriy-stilistik komponent (lug'at bazasi);

- ma'lumotlar manbai bo'yicha: bir nechta turdagi ma'lumotlarni o'z ichiga olgan LMB ("Ovoz" ma'lumotlar bazasi), eksperimental ma'lumotlar (psixolingvistik ma'lumotlar bazasi), lug'at ma'lumotlari (metaforik terminologiya ma'lumotlar bazasi, pragmatik belgilangan lug'at ma'lumotlar bazasi, lug'at MB);

- qo'llanilish sohasi bo'yicha: turli tadqiqot maqsadlari uchun mo'ljallangan LB ("Ovoz" ma'lumotlar bazasi", pragmatik belgilangan lug'at ma'lumotlar bazasi), terminologik tadqiqotlar (metaforik ma'lumotlar bazasi) terminologiya), psixolingvistik tadqiqotlar (psixolingvistik ma'lumotlar bazasi), leksik va grafik amaliyotda foydalanish (lug'at bazasi).

Xulosa sifatida aytish mumkinki, amaliy tilshunoslik va zamonaviy leksikografiya kesishmasidagi samarali yo'nalishlardan biri bu LB ning rivojlanishi bo'lib, undan foydalanish bir qator yangi muammolarni hal qilish va tilshunoslik uchun an'anaviy muammolarni hal qilish samaradorligini oshirish imkonini beradi. Lingvistik (va, umuman, gumanitar) tadqiqotlarda ma'lumotlar bazasi texnologiyasidan foydalanish nafaqat materiallarni qulay taqdim etish va ularni yagona tuzilishga birlashtirishga, balki ular bilan ishlash samaradorligini oshirishga yordam beradi. Mashinani qayta ishlash vositalaridan foydalanish,

shuningdek, yaratilgan ma'lumotlar bazalari yoki ulardan foydalanish asosida keyingi tadqiqotlar uchun yangi istiqbollarni ochadi.

Foydalanilgan adabiyotlar

1. Abdurahmonova N. Mashina tarjimasining lingvistik asoslari. –Toshkent: Akadernashr, 2012.
2. Alexander Clark, Chris Fox, and Shalom Lappin. (Edited) The Handbook of Computational Linguistics and Natural Language Processing. 2010 Blackwell Publishing Ltd.
3. Bender E.M., Langendoen D.T. Computational linguistics in support of linguistic theory. (Linguistic Issues in Language Technology, Vol. 1, 2010)
4. Igor A. Bolshakov and Alexander Gelbukh COMPUTATIONAL LINGUISTICS Models, Resources, Applications. – Mexico, 2004
5. Jurafsky D., J. H. Martin. Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition. 2006 (elektron variant)
6. Mitkov R. The Oxford handbook of Computational linguistics. Oxford university press, 2003 (elektron variant)
7. Pulatov A.K., Muhamedova S. Kompyuter lingvistikasi (o'quv qo'llanma). – Toshkent, 2014.
8. Pulatov A. Kompyuter lingvistikasi. –Toshkent: Akadernashr, 2011.
9. Rahimov A. Kompyuter lingvistika asoslari. -Toshkent: Akadernashr, 2011.
10. Волкова И.А. Введение в компьютерную лингвистику. Практические аспекты создания лингвистических процессоров. –Москва, 2006.— 43 с.
11. Зубов А.В., Зубова И.И. Основы лингвистической информации. –Минск: МГПИИЯ, 1992.
12. Марчук Ю.Н. Компьютерная лингвистика: учеб. Пособие –Москва, Восток-Запад, 2007.
13. Шемакин Ю.И. Начала компьютерной лингвистики. –М.:Высшая школа, 1992