

TILSHUNOSLIKDA LINGVISTIK BAZALAR VA ULARNING AHAMIYATI

Asatullayev Furqat
TerDU 2-kurs magistranti

Lingvistik ma'lumotlar bazasi obyekt va ularning munosabatlari to'g'risida boshlang'ich, oraliq va yakuniy ma'lumotlarni saqlash va ulardan samarali foydalanishga xizmat qiladigan maxsus ma'lumotlar tuzilishidir. Lingvistik baza tuzilma bo'lib, uning uchun samarali optimallashtirish algoritmi mavjud bo'lib, ma'lumot qo'shish, o'zgartirish va qidirish muammolarini hal qilishni osonlashtiradi.

Korpus (lotincha: corpus – tana) – 1) inshoot, mexanizm, asboblar, apparatlar, qurilmalar va ba'zi transport mashinalarining qismi; 2) yaxlit maydondagi bir necha binodan bittasi yoki yirik inshoot ning alohida qismi; 3) kegeli (o'lchami) 10 punktga (3.76mm) ga teng bo'lgan bosma shrift; 4) biron maxsus ish bilan mashg'ul bo'lgan mutaxassislar hay'ati (mas, diplomatic korpus, ofitserlar korpusi va boshqalar).

Korpus lingvistikasi amaliy tilshunoslikning ma'lum bir qismi bo'lib, kompyuter qurilmalari yoramida korpusni yaratish va uning uchun muhim hisoblanadigan lingvistik ma'lumotlar bazasini tuzish lozim bo'ladi.

Korpus – bu ma'lum bir tilning o'ziga xos xususiyatlari va variantlarini aks ettiruvchi bir necha belgi asosida tanlab olingan electron shakldagi matn parchalari, lingvistik tadqiqotlar uchun asos vazifasini bajaradi. Korpus lingvistikasi haqida ma'lumotlarni yanada kengaytiradigan bo'lsak, korpus

Lingvistik baza yoki lingvistik ma'lumotlar bazasi tillarning katta hajmli (mashina tarjimasini, tahrirlovchi dasturlar, sinonim lug'atlar, omonim va antonym lug'atlar) vazifalarni bajarishga yuklatilgan katta ma'lumotlar ombori hisoblanadi. Lingvistik baza bu korpus uchun yaratilayotgan ma'lumotlar bazasi hisoblanib, korpusning eng asosiy jihati sanaladi. Lingvistik ma'lumotlar bazasida ishlanayotgan muayyan bir tilning hamma jabhalariga oid manbalar yig'iladi. Hozirgi kunda tilshunosligimizda lingvistik bazalarni yaratish bo'yicha amalda ilk ishlar bajarilmoqda.

Lingvistik ta'monot, ya'ni lingvistik ma'lumotlar bazasi muayyan tilning grammatik qonuniyatlari, qolaversa, semantic imkoniyatlari asosida yaratiladi. Bizga ma'lumki, har qanday korpus lingvistikasiga doir dasturlarni yoki lug'atlarni ularning lingvistik bazasiz taassavur qilish qiyin. Shuningdek ular lingvistik bazaga ehtiyot ham sezadi. O'zbek tilida nafaqat ot so'z turkumi grammatik shakllarini, balki boshqa turkumlarni ham bira yo'la lingvistik ma'lumotlar bazasini yaratish mumkin emas.

Hozirgi vaqtda ma'lumotlar bazalari ko'rinishida ma'lumotlarni kompyuterda tasvirlash tilshunoslik va amaliy tilshunoslikning turli sohalarida qo'llaniladi:

• tipologik va qiyosiy tilshunoslikda keng ma'lum bo'lgan va kam o'rganilgan tillarning tavsifi;

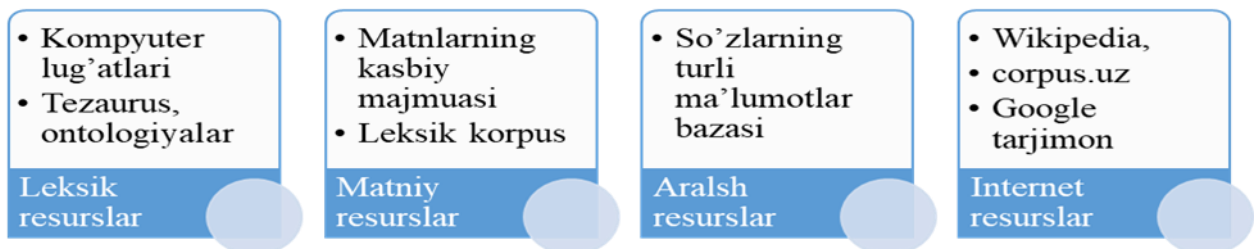
- fonetik tadqiqotlar;
- leksikografiya, shu jumladan lug'at tuzish;
- grammatikani o'rganish, grammatik qoliplarni aniqlash va baholash;
- swmantik tadqiqotlar, jumladan tezauriyalarni tuzish;
- hisoblash tilshunosligida va amaliy lingvistik muammolarni hal qilish uchun, xususan:

1. kompyuter vositalaridan foydalangan holda tillarni o'rgatish;
2. avtomatik tarjima tizimlarini yaratish;
3. nutq texnologiyalarining lingvistik jihatlarini ta'minlash (avtomatik tanib olish va nutq sintezi);
4. psixolingvistika sohasidagi tadqiqotlar;
5. sotsiolingvistika bo'yicha qilingan ma'lumotlar;
6. dialektologiya sohasidagi tadqiqotlar.

Lingvistik bazasi hozirgi kunda korpus lingvistikasining asosiy ta'minoti hisolanib, ko'pgina ma'lumotlar ushbu baza orqali korpusga uzatiladi. Ya'ni ma'lum bir ma'lumot bazaga joylashtirilgan keyin, korpus qidiruv tizimida uni ko'rsata oladi.

Lingvistik bilimlar bazasi haqida Nilufar Abduraxmonova quyidagicha izoh beradi: "Lingvistik bilimlar bazasi mashina uchun matnlarni qayta ishlashda zarur bo'lgan barcha ma'lumotlarni o'z ichiga qamrab oladi".

Lingvistik bazalarga quyidagilarni misol qilib keltirishimiz mumkin:



Yuqorida ko'pgina ta'kidlaganimizdek lingvistik baza korpus uchun ta'minot hisoblanib, sanab o'tilgan lingvistik bazalar haqida ma'lumotlar berib o'tamiz.

Kompyuter lug'atlari – bu ma'lum bir soha yoki kasbga tegishli bo'lgan so'zlar yoki iboralar lug'ati bo'lib, ma'lum bir buyruq asosida ishlash qobiliyatini o'zida mujassam etgan dastur hisoblanadi. Bunga misollar sifatida ko'pgina lug'atlarni sanash mumkin. Masalan ismlar ma'nosi lug'ati. Bu lug'atda ma'lum bir miqdorda ismlar joylangan bo'lib, ularning har birining ma'nosi dasturning ta'minnot bazasi, ya'ni lingvistik bazasini tashkil etadi.

Tezaurus – muayyan tildagi barcha so'zlarniqamrab oladigan, ularning matnda qo'llanish holatlarini to'la-to'kis aks ettiradigan lug'at.

Korpus uchun lingvistik baza yoki lingvistik ma'lumot deyilganda, og'zaki hamda yozma manbalar anglandi.

Korpus bo'yicha jahonda ko'plab ishlar qilingan bo'lib, rus tili misolida oladigan bo'lsak, N.A.Mishanka, V.P.Zaxarov, Y.V.Nedoshivina, V.V.Rikov, V.Plungyanlarning korpus bo'yicha birdirgan fikr, taklif va mulohazalari korpus bo'yicha qilingan ilk tadqiqotlar sifatida aytib o'tish mumkin. Masalan rus tili korpusi bo'yicha ilk tadqiqotlar olib borgan Mishankin lingvistik ma'lumotlar bazasi haqida quyidagi fikrlarni berib o'tgan: "Ma'lumotlar bazasi texnologiyasi ham an'anaviy, ham elektron lug'atlarni yaratish jarayonida qo'llaniladi. Maxsus va terminologik lug'atning bazalari faol ishlangan".

Xulosa qilib shuni ta'kidlash joizki, korpus lingvistikasi va lingvistik bazalar mukammallashuvi uchun grammatik shakllar bazasini yaratilishi kerak. Ularning lingvistik ma'lumotlar ta'minoti shakllantirilishi lozim. Bu vazifani amalga oshirish uchun dastlab so'z turkumlari doirasida grammatik baza va ularning lingvistik ma'lumotlar ta'minoti yaratilishi maqsadga muvofiqdir.

Foydalanilgan adabiyotlar ro'yxati

1. Захаров. Компьютер лингвистики и лингвистик базы данных. Конференция. 2002 г
2. Саодат Муҳамедова "Компьютер лингвистикаси. Ўқув қўланма. Тошкент-2007. Электрон адабиёт. 20-бет
3. Баранов А.Н. Введение в прикладную лингвистику. – М., 2001.
4. N.Abduraxmonova Kompyuter lingvistikasi. Darslik. "Nodirabegim" nashriyoti. Toshkent – 2021.
5. Френсис Н., Кучера Г. Вычислительный анализ современного американского варианта английского языка – М., 1967.; Синклер Дж. Предисловие к книге «Как использовать корпуса в преподавании иностранного языка» (Электронный ресурс) URL https://studiorum-ruscorpora.ru/wp-content/themes/bootstrapbasic4child/texts/sinclair_introduction.pdf
6. Н.А.Мишанка. Базы данных в лингвистических исследованиях. УДК 81`374