

KOMPYUTER LINGVISTIKASI YO'NALISHLARI USTIDA OLIB BORILAYOTGAN IZLANISHLAR

Mohichehra Abramatoeva Abdurazzoq qizi

Termiz davlat universiteti o'zbek filologiyasi fakulteti

abramatovamohichehra@gmail.com

Annotatsiya

Hozirgi vaqtda chet el tilshunosligida korpus lingvistika deb nomlanuvchi yo'nalish paydo bo'ldi. Korpus tilshunosligi elektron shaklda oldindan qayta ishlangan, belgilangan va tizimlashtirilgan katta hajmli matnlar (korpora) materiallari bo'yicha nutq shakllarini o'rganishga qaratilgan. Bunday matn korpusining oddiy elektron kutubxonalardan farqi shundaki, maxsus dasturlar yordamida matnlarning kerakli bo'laklarini ko'rsatilgan parametrlar bo'yicha izlash, so'ngra olingan ma'lumotlarni umumlashtirish va tahlil qilish mumkin. Bunday holda, tadqiqotning o'zi soniyalarda o'lchanadigan vaqtni oladi.

Kalit so'zlar

Kompyuter, lingvistika, korpus.

Korpus lingvistikasining an'anaviy tilshunoslikdan farqi nafaqat kompyuter texnikasidan foydalanishda, balki o'rganilayotgan hodisalarga umumiy yondashuvda hamdir. Korpus lingvistikasi muayyan tillar matnlarida mujassamlangan nutqni o'rganishga qaratilgan. U nutq ma'lumotlaridan xulosalarga o'tadi, ya'ni. miqdoriy tadqiqotlarga tayangan holda induktiv empirik usuldan foydalanadi [1]. Tadqiqotchilarning korpus lingvistikasiga qiziqishi tadqiqotchilarga axborotni olish va qayta ishlash uchun alohida imkoniyatlar yaratgan kompyuter texnologiyalari va dasturiy ta'minotining keng rivojlanishi bilan bog'liq. Ilgari tadqiqotchi bir necha oylar va yillar davomida kutubxonalarda o'tirib, kerakli ma'lumotlarni qo'lda yozgan ma'lumotlar va ma'lumotlar miqdori endi tom ma'noda "barmoq uchida" bo'lib chiqdi va ularni bir necha daqiqada olish va qayta ishlash mumkin. Biroq, axborot portlashi nafaqat foyda, balki muammolarni ham keltirib chiqaradi. Ajablanarlisi shundaki, ularning eng muhimi - ma'lumotlarning ko'pligi.

Jamiyat bu muammoga XX asrning ikkinchi yarmida, axborot texnologiyalari hali bunday rivojlanishga erishmagan paytda duch keldi. 60-yillarda. O'tgan asrning sanoati rivojlangan mamlakatlarida shunday qoida mavjud edi: agar mavjud bo'lgan 50 000 dollarlik loyihani amalga oshirish zarur bo'lsa, lekin bu haqda ma'lumot topish uchun vaqt kerak bo'lsa, shunga o'xshash loyihani qayta ishlab chiqishdan ko'ra arzonroq edi. u haqida ma'lumot qidiring. Internetning rivojlanishi bilan ma'lumot olish jarayoni beqiyos oson, tez va arzonlashdi, ammo yangi muammo paydo bo'ldi: axborotning katta oqimida kerakli ma'lumotni qanday tanlash kerak. Masalan, "Korpus lingvistikasi" so'rovi uchun o'zbek qidiruv tizimi Yandex bu mavzu bo'yicha 2033 sahifa va 373 ta saytni, "Korpus lingvistikasi" so'rovi uchun ingliz tilidagi Yahoo tizimida esa 938 000 ta iborani zikr qilgan. Internet va Yahoo-da qidirish vaqti 0, 14 soniyani oladi.

Agar biz olingan ma'lumotnomalarning butun hajmini hisobga olsak, unda barcha ma'lumotlarni ko'rish uchun sarflangan vaqt XX asrda kutubxonada o'tkazgan vaqtga nisbatan nomutanosib ravishda ko'proq bo'ladi. Shu bilan birga, bog'lanishlarning aksariyati o'zaro bog'langan, ya'ni. mualliflar bir-biriga murojaat qiladilar yoki so'rov so'zlarini eslatib o'tish nashrlar yoki maxsus

avtoreferatlar va dissertatsiyalar reklamasini nazarda tutadi, shuningdek, seminarlar va konferentsiyalarning e'lonidir.

Boshqacha qilib aytganda, havolalarning aksariyati ilmiy, lingvistik ma'lumotlarni o'z ichiga olmaydi va filtrlash kerak bo'lgan "axborot shlaki" bo'lib, bu ham vaqt talab etadi. Bunday vaziyatda Muersning birinchi qonuni kuchga kiradi: "iste'molchi ma'lumotsiz qidirishdan ko'ra ma'lumot izlash qiyinroq bo'lsa, qidiruv tizimidan foydalanishdan qochadi" [3]. Sanoat uy bekalari uchun uy ishlarini osonlashtirish uchun tobora ko'proq maishiy texnika ishlab chiqarganidek, dasturchilar ham ma'lumotni muntazam qidirishni osonlashtirish va foydalanuvchini keyingi ilmiy izlanishlarga undash uchun mo'ljallangan ko'proq dasturiy mahsulotlarni yaratish ustida ishlamoqda.

Ammo "kerakli ma'lumotlarni aniq, ilmiy asoslangan izlash uchun maxsus semiotik (belgi) tizimini qurish kerak" [2]. Ushbu kontseptual tarkib ko'pincha foydalanuvchining ma'lumot uchun og'zaki so'rovida ifodalanmaydi, bu taqdim etilgan ma'lumotlarning katta miqdorini yaratadi, bu esa u bilan keyingi ishlashni bloklaydi.

Qadim zamonlardan boshlab tilshunoslar mazmun (tushunchalar) va uning lisoniy ifodasi (so'zlari) o'rtasida bog'lanish o'rnatishga, mazmun va shakl o'rtasidagi qonuniyatlar va munosabatlarni aniqlashga intilganlar. Biroq, bunday tadqiqotlar boshi berk ko'chaga kirib qoldi, chunki tushunchalar falsafa va psixologiya sohasiga, so'zlar esa tilshunoslik sohasiga tegishli. Xuddi shu tushuncha turli so'zlar bilan ifodalanishi, bir so'zda turli mazmun singdirilishi mumkin. Humpty Dumpty "Ko'zoynak orqali" kitobida aytganidek, "Men bir so'zni qabul qilsam, bu men xohlagan narsani anglatadi, ko'proq va kam emas". Ammo so'zlar ortidagi tushunchalarning tafovutiga qaramay, tadqiqotchilar shakl mazmunini ifodalash naqshlarini o'rnatishga urinishlaridan voz kechmaydilar. Masalan, Muers har bir tushunchani maxsus leksik birlik - deskriptor orqali tasvirlash mumkin, deb hisoblaydi [4].

U so'z bilan bog'liq bo'lgan odamlarni "fikir ko'r" deb hisoblaydi va bu turdagi odamlar uchun ma'lumotni qidirish imkonsiz deb hisoblaydi, shu bilan birga eng mavhum g'oyalar ham majburiy ravishda ifodalanishi kerakligini unutadi. ko'proq qochdi. Leksik so'rovdan kontseptual so'rovga o'tish shakli bo'lgan va kompyuterda qidirish uchun mos bo'lgan deskriptorlar usuli R.P.Futrel va S.Gauch [5] ishlarida bayon etilgan. Bootstrapping ("o'z-o'zini ta'minlash") so'zini o'z ichiga olgan asarning nomi, so'rovning og'zaki tavsifidan kontseptualga o'tish bir xil tildagi so'zlar asosida sodir bo'lishini ko'rsatadi, faqat cheklangan. ularning tarkibida va turlicha tashkil etilgan.

Ular o'tish jarayonini so'rovni kengaytirish deb atashadi. So'rov kengaytiriladi va kerakli lug'at (onlayn ma'lumotlar bazasi) va ixtisoslashtirilgan "o'xshashlik matritsasi" ning operatsion korpusiga qo'ng'iroqlar yordamida "kontseptual" bo'ladi. Shu tariqa, qidiruv jarayonida ma'nosiz leksik qiyoslash (so'zlarni moslashtirish) bartaraf qilinadi va kontseptual qidiruvga (kontseptual qidiruv) o'tish amalga oshiriladi. Ularning tadqiqotlari 200 000 so'zdan iborat korpus materiallari bo'yicha olib boriladi, ular dastlabki belgilar yordamida ierarxik tarzda joylashtirilgan uyalar (klasterlar) tashkil etuvchi sinflarga va o'xshashlik guruhlariga (simsetlarga) bo'linadi.

Kompyuter tadqiqotlari ma'lumotlari Roget tezausining materialida tasdiqlangan. Shunday qilib, kompyuter tahlilining natijalari Rogetning empirik tahlili natijalaridan unchalik farq qilmadi, lekin kompyuterni o'rganishning o'zi an'anaviyga qaraganda beqiyos kamroq vaqt talab qildi. Axborotni qidirish samaradorligini oshirish qidiruv so'rovi va korpus matnlarida semantik tuzilmalarni taqqoslash orqali amalga oshiriladi. G'arb tilshunosligida til korpusi muammosi

ancha oldin muvaffaqiyatli hal etilgan. 1967 yilda G. Kuchera va N. Frensis 1 million so'zdan iborat zamonaviy Amerika versiyasi matnlaridan iborat Braun korpusida amalga oshirilgan "Zamonaviy Amerika ingliz tilining hisoblash tahlili" asarini nashr etdilar. 1960 yilda nashr etilgan Amerika merosi lug'ati matn korpusidan tuzilgan birinchi lug'at edi.

Undan keyin Bank of English korpusiga asoslangan Collins' COBUILD ingliz lug'ati paydo bo'ldi. Undan keyin LOB (1960-yillar Britaniya ingliz tili), Kolxapur (Hind ingliz tili), Vellington (Yangi Zelandiya ingliz), ACE (Avstraliya ingliz), Frown Corpus (90-yillarning boshi Amerika ingliz), FLOB Corpus (1990-yillarning boshidagi Britaniya ingliz). 90-yillarda. noshirlar konsorsiumi, Oksford va Lankaster universitetlari, Britaniya kutubxonasi, 100 million so'zdan iborat Britaniya Milliy korpusi ishtirokida tuzilgan. Lotin matnlari korpusi "Perseus", Charlz universitetining Chexiya korpusi va boshqalar ham ishlab chiqilgan. million so'z. O'zbek tiliga kelsak, mahalliy tilshunoslik bu borada chet elshunoslikdan orqada.

Ular o'tish jarayonini so'rovni kengaytirish deb atashadi. So'rov kengaytiriladi va kerakli lug'at (onlayn ma'lumotlar bazasi) va ixtisoslashtirilgan "o'xshashlik matritsasi" ning operatsion korpusiga qo'ng'iroqlar yordamida "kontseptual" bo'ladi. Shu tariqa, qidiruv jarayonida ma'nosiz leksik qiyoslash (so'zlarni moslashtirish) bartaraf qilinadi va konseptual qidiruvga (kontseptual qidiruv) o'tish amalga oshiriladi. Ularning tadqiqotlari 200 000 so'zdan iborat korpus materiallari bo'yicha olib boriladi, ular dastlabki belgilar yordamida ierarxik tarzda joylashtirilgan uyalar (klasterlar) tashkil etuvchi sinflarga va o'xshashlik guruhlariga (simsetlarga) bo'linadi.

Kompyuter tadqiqotlari ma'lumotlari Roget tezausining materialida tasdiqlangan. Shunday qilib, kompyuter tahlilining natijalari Rogetning empirik tahlili natijalaridan unchalik farq qilmadi, lekin kompyuterni o'rganishning o'zi an'anaviyga qaraganda beqiyos kamroq vaqt talab qildi. Axborotni qidirish samaradorligini oshirish qidiruv so'rovi va korpus matnlarida semantik tuzilmalarni taqqoslash orqali amalga oshiriladi. G'arb tilshunosligida til korpusi muammosi ancha oldin muvaffaqiyatli hal etilgan. 1967 yilda G. Kuchera va N. Frensis 1 million so'zdan iborat zamonaviy Amerika versiyasi matnlaridan iborat Braun korpusida amalga oshirilgan "Zamonaviy Amerika ingliz tilining hisoblash tahlili" asarini nashr etdilar. 1960 yilda nashr etilgan Amerika merosi lug'ati matn korpusidan tuzilgan birinchi lug'at edi.

Undan keyin Bank of English korpusiga asoslangan Collins' COBUILD ingliz lug'ati paydo bo'ldi. Undan keyin LOB (1960-yillar Britaniya ingliz tili), Kolxapur (Hind ingliz tili), Vellington (Yangi Zelandiya ingliz), ACE (Avstraliya ingliz), Frown Corpus (90-yillarning boshi Amerika ingliz), FLOB Corpus (1990-yillarning boshidagi Britaniya ingliz). 90-yillarda. noshirlar konsorsiumi, Oksford va Lankaster universitetlari, Britaniya kutubxonasi, 100 million so'zdan iborat Britaniya Milliy korpusi ishtirokida tuzilgan. Lotin matnlari korpusi "Perseus", Charlz universitetining Chexiya korpusi va boshqalar ham ishlab chiqilgan. million so'z. O'zbek tiliga kelsak, mahalliy tilshunoslik bu borada chet elshunoslikdan orqada....

Unga ko'plab ilmiy muassasalar va universitetlarning tilshunos olimlari jalb etilgan. Milliy korpus - bu ma'lum bir tilni mavjudligining ma'lum bir bosqichida ifodalovchi, ma'lum bir tilni barcha janrlar, uslublar, ijtimoiy va hududiy dialektlarda va hokazolarda aks ettiruvchi elektron shakldagi matnlar to'plami. Korpus vakili bo'lishi kerak, ya'ni. iloji bo'lsa, tilda ifodalangan yozma va og'zaki matnlarning barcha turlarini o'z ichiga oladi va bu matnlarning barchasi, iloji bo'lsa, ularning tegishli davr tilidagi ulushiga mutanosib ravishda korpus tarkibiga kiritilishi kerak. Bu faqat korpusning katta hajmi (yuz millionlab so'zlardan foydalanish) bilan mumkin.

Tuzuvchilar tomonidan rejalashtirilgan O'zbek tili Milliy korpusining hajmi 200 million so'zni tashkil etadi. Belgilash ishning asosiy xususiyati hisoblanadi. U korpusni bugungi kunda Internetda to'plangan matnlar to'plamidan (yoki "kutubxonalar") ajratib turadi. Belgilash qanchalik boy va xilma-xil bo'lsa, korpusning ilmiy va ma'rifiy ahamiyati shunchalik yuqori bo'ladi. O'zbek tili Milliy korpusida hozirda to'rt xil belgi qo'llaniladi: metamatn, morfologik, urg'u va semantik; Yaqin kelajakda sintaktik belgilarni joriy etish rejalashtirilgan. Belgilash tizimi doimiy ravishda takomillashtirilmoqda. Milliy korpusning yaratilishi lingvistik tadqiqotlarning barcha sohalar uchun katta imkoniyatlar yaratadi.

Matnlarni ommaviy statistik qayta ishlash imkoniyati gipotezalarni matematik jihatdan tasdiqlash yoki rad etish, grammatika va lug'atlarni tuzish imkonini beradi. Misol uchun, hozirda Britaniya Milliy Korpusiga tayanmasdan bitta inglizcha izohli lug'at tuzilmagan. Milliy korporatsiyaning asosiy foydalanuvchilari birinchi navbatda tilshunoslardir. Shu bilan birga, ayrim til davrlariga oid statistik ma'lumotlar adabiyotshunos va tarixchilar, oliy o'quv yurtlari o'qituvchilari va talabalari uchun ham qiziqarli bo'lishi mumkin. Darslik va lug'atlarni jamlash korpusga tayanmasdan ham mumkin emas. Ishlab chiqilayotgan o'zbek tilining Milliy korpusi birinchi navbatda 19-asr boshidan XXI asr boshlarigacha bo'lgan davrni qamrab oladi.

Bu davr oldingi davrlar tilini ham, zamonaviy tilni ham turli sotsiolingvistik variantlarda - adabiy, so'zlashuv, qisman dialektal tilda ifodalaydi. Korpus tarkibiga madaniy ahamiyatga ega bo'lgan va til nuqtai nazaridan ham qiziqish uyg'otadigan original (tarjima qilinmagan) badiiy adabiyot (nasr va dramaturgiya, bundan keyin she'riyat) kiradi. Badiiy matnlardan tashqari, korpusda yozma (va zamonaviy bosqich uchun - og'zaki) tilning ko'plab boshqa namunalari mavjud: xotiralar, insholar, jurnalistika, ilmiy-ommabop va ilmiy adabiyotlar, ommaviy nutqlar, shaxsiy yozishmalar, kundaliklar, hujjatlar, va hokazo n Milliy korpora yordamida tadqiqot tadqiqotchilar beradi.

Ko'p imkoniyatlar va yo'nalishlar. So'z shakllarining chastotasini aniqlashdan tashqari, korpus muvofiqlik haqida fikr beradi. Muvofiqlik - matnda topilgan so'z shakllarining alifbo tartibida joylashtirilgan ro'yxati. Lug'atdan farqli o'laroq, bu erda so'z og'zaki muhiti bilan berilgan. Shunga o'xshash, ammo bir xil bo'lmagan, tasodifiy moslashuvlar foizidan oshadigan so'zlar yoki atamalar ketma-ketligini o'rganish. Kollokatsiya tushunchasi so'z birikmalaridan foydalanish qoidalarini, masalan, ma'lum old qo'shimchalar va fe'llar yoki fe'llar va otlarning birikmalarini anglatadi.

Shunday qilib, korpus tadqiqotlari nutq birliklarini tabiiy muhitdan ajratmasdan kontekstda o'rganish imkonini beradi. Matn korpusi bilan ishlash distributiv texnikani yangi bosqichga ko'tarish imkonini beradi, bu quyidagilarning umumiylikni aniqlashga qaratilgan: 1) berilgan til birligining barcha chiziqli muhitlari; 2) o'rganilayotgan til birligining barcha birikmalari. Distribyutsiya usuli semantikani kontekstli o'rganishni ham o'z ichiga olishi mumkin, lekin birinchi navbatda u nutqdagi (matndagi) ba'zi elementlarni boshqalarning bilimlari asosida bashorat qilish imkoniyatidan iborat. Masalan, frazeologik jihatdan bog'liq bo'lgan birlik yoki ko'p so'zlarning grammatik shakli ko'pincha boshqa so'z yoki grammatik shaklni to'liq taxmin qilish imkonini beradi.

So'zlarni taqsimlashning boshqa holatlarida prognozlash ehtimoliy naqshlarga bo'ysunadi, ya'ni. so'z yoki grammatik shaklni faqat ma'lum bir ehtimollik darajasi bilan bashorat qilish mumkin. Korpus lingvistikasi bu ehtimollikni aniq raqamlarda ifodalash imkonini beradi. Ikki yoki undan ortiq matn korpusini taqqoslash tarjima qilish uchun namunalarni tezda topish uchun tarjimonga ikki tildagi ekvivalent matn qismlarini maxsus ma'lumotlar bazasida to'plash imkonini beruvchi

tarjima diskleri deb ataladigan kompyuter dasturlarini, interfaol vositalarni yaratishga imkon beradi. kelajakda yangi matnlar. So'zlar, iboralar va butun iboralar bo'laklar vazifasini bajarishi mumkin. Janr va mavzu bo'yicha bir-biriga yaqin bo'lgan matnlar ustida ishlashda bunday vositalar ma'lumotlar bazasi o'sishi bilan tarjimani soddalashtiradi va tezlashtiradi.

Dastur tarjima qilinadigan matnni segmentlarga ajratadi (iboralarni yoki ajratilgan iboralarni ajratib ko'rsatadi) va natijada olingan kombinatsiyalarni ma'lumotlar bazasi elementlari bilan taqqoslaydi, agar mos kelsa, tarjimondan fragmentning tugagan tarjimasini yakuniy matnga almashtirishni taklif qiladi (yoki uni avtomatik ravishda almashtirish). Eng ilg'or dasturlarda bir nechta tillar uchun o'rnatilgan morfologik modullar va imlo tekshirgichlari, shuningdek, ikki tilda juftlangan fragmentlarni avtomatik ravishda shakllantirish uchun nisbatan aqlli parallel matn moslashtirish vositalari mavjud. Turli dasturlarda moslashish tartib-qoidalari har xil, lekin, qoida tariqasida, ular dialog elementlarini o'z ichiga oladi va ba'zan parallel matnlardan birini o'zgartirishni talab qiladi. Biroq, matn korpusini yaratish tadqiqotchilar oldida turgan barcha muammolarni hal qilmaydi. Ishning muhim qismi - bu korpuslar bilan samarali ishlash imkonini beruvchi dasturlarni yaratish. Avvalo, bular korpus menejerlari deb ataladigan ma'lum bir korpusga asoslangan qidiruv tizimlari. Misollardan biri, Sankt-Peterburg davlat universitetida ishlab chiqilgan "Ekspert lingvistik tizim" dasturi.

ELS dasturidan korpus matnlarining leksik tarkibini o'rganish, leksemalar kontekstini yoki korpusdagi so'z shakllarini izlash (xususan, wordnet tezaurusini yaratishda), barcha turdagi chastotali lug'atlarni (masalan, yozuvchilar) yaratishda foydalanish mumkin. til lug'atlari). ELS dasturining asosiy vazifasi korpus matnlaridagi leksemalarning chastotasini tahlil qilishdan iborat. Chiqarish ro'yxatida siz kontekstni barcha tokenlar uchun yoki foydalanuvchining iltimosiga binoan faqat grafik omonimlar uchun ko'rishingiz mumkin. Olingan ro'yxat uchun saralash variantini tanlash ham mumkin. Bundan tashqari, foydalanuvchi tomonidan tanlangan fayllar uchun ELS dasturi ba'zi qo'shimcha statistik ma'lumotlarni hisoblab chiqadi: jumlar soni, jumladagi so'zlar soni va boshqalar. Dastur xususiyatlari so'zlarni bo'g'inlarga avtomatik ravishda ajratishni o'z ichiga oladi.

Korpus kompilyatorlari o'z ishlarida bir qator muammolarga duch kelishadi. Avvalo, korpusning vakillik muammosini ta'kidlash kerak, ya'ni. muammoli hududning barcha xususiyatlarini aks ettirish qobiliyati. Reprezentativlik fonetik, morfologik, sintaktik, uslub parametrlari bilan belgilanadi. Korpusni yaratuvchisi birinchi navbatda qanday korpusni va kim uchun yaratadi, degan savolni ko'taradi. Ma'lum bir tilning barcha matnlarini yoki barcha suhbatlarini kompyuterga taqdim etishning iloji yo'q, shuning uchun korpusni yaratuvchilar ushbu korpus mo'ljallangan tadqiqotchilar tomonidan boshqariladi. Leksikograflarga kam qo'llaniladigan so'zlar va/yoki ularning shakllari misollari bo'lgan katta korpus kerak, lekin grammatikachilar, stilistlar va boshqalarning maxsus tadqiqotlari uchun milliy tilning turli uslublari va janrlarini qamrab olish zarur. Va bu erda kompilyatorlar korpus uchun matnlarni tanlash muammosiga duch kelishadi.

Masalan, V. Frensis va G. Kuchera aniq va aniq tanlov mezonlariga javob beradigan matnlar korpusini taqdim etishni maqsad qilgan:

1. Matnning kelib chiqishi va tarkibi (muallif amerikacha ingliz tilida so'zlashuvchi bo'lishi, dialog matnning yarmidan kam qismini egallashi kerak edi).
2. Sinxronizatsiya (birinchi marta 1961 yilda nashr etilgan matnlar kiritilgan).

3. Turli janrlarni sonli tasvirlashning puxta o'ylangan nisbati va alohida ehtimollik protsedurasi yordamida alohida matnlarni tanlash.

4. Kompyuterda ishlov berishning mavjudligi (matnning grafik xususiyatlarini etkazish uchun maxsus teglar va boshqalar).

Alohida matnning hajmi statistik jihatdan uning stilistik xususiyatlarini aks ettirishi kerak va janrlarning son tarkibi va nisbati janrlarning stilistik xususiyatlarini va ularning nisbiy vaznini etarli darajada ifodalashi kerak. Shu bilan birga, korpus bo'lgan ob'ektiv dasturiy vosita sub'ektiv tanlashga asoslanganligi va inson omili bilan belgilanadiganligi qarama-qarshi ko'rinadi. Ushbu sharhni yakunlar ekanmiz, korpus lingvistikasining istiqbollari va ulardan kelib chiqadigan muammolarni ajratib ko'rsatishimiz mumkin.

Birinchidan, ma'lumotlarning katta miqdori uni qidirishni qiyinlashtiradi, bu esa ushbu korpus maqsadlari uchun maxsus ishlab chiqilgan maxsus korpus va qo'shimcha qidiruv vositalarini yaratishni talab qiladi.

Ikkinchidan, foydalanuvchi so'rovini yanada samaraliroq qilish uchun mo'ljallangan kontseptual qidiruv, aslida sinonimlarni izlash va ularni semantik uyalariga tartibga solishga to'g'ri keladi, ya'ni. mantiqdan ko'ra ko'proq semantika sohasiga tegishli bo'lib, uni rasmiylashtirish qiyin.

Uchinchidan, korpuslar ob'ektiv ma'lumot berish uchun qurilgan, shuning uchun korpus yaratuvchilari korpus uchun matnlarni tanlashda sub'ektivlik omilini kamaytirish va qat'iy va aniq tanlash mezonlarini ishlab chiqish vazifasini bajaradilar. Bularning barchasi korpus lingvistikasi sohasidagi tadqiqotlarning ahamiyati va istiqbollarini kamaytirmaydi va amaliy tadqiqotlarning yangi yo'nalishlariga ishora qiladi.

Adabiyotlar:

1. Bakirova X.B. Nolingvistik universitetlarda xorijiy tilda terminologiyani o'qitish. Innovatsion nutq xalqaro jurnali. Integratsiya va ta'lim. Jildi: 01-son: 01. 2020. <http://summusjournals.uz/index.php/ijdiie>.
2. Belyaeva L.N. Zamonaviy tarmoq makonidagi lingvistik texnologiyalar: mahalliyashtirish sohasidagi tilshunos: monografiya. Kitob uyi, 2016. - S. 134.
3. Delpech E., Dyle B. Taqqoslanadigan korpuslardan olingan lug'at bilan ishlash: tekshirish va almashish // Terminologiya va bilim muhandisligi bo'yicha 9-konferentsiya materiallari. Fiontar: Dublin shahar universiteti, 2010, 229-223-betlar.
4. Finegan E. TILI: tuzilishi va ishlatilishi. - Nyu-York: Harcourt Brace College Publishers, 2004. Fung P. Parallel bo'lmagan korpuslardan terminologiyani tarjimalarini topish. – URL: http://www.cs.columbia.edu/nlp/papers/1997/fung_mckeown_97.pdf.
5. Gerd A.S. Maxsus maqsadlar uchun til o'rganishga kirish: o'quv qo'llanma. Filologiya fakulteti, 2011. - B. 60.
6. Xurshid A., Rojers M. Terminologiyani boshqarish: korpus yondashuvi. URL: <http://www.mt-archive.info/90/Aslib-1992-Ahmad.pdf>.