



Anorxon AXMEDOVA,
O'zbekiston davlat jahon tillari universiteti o'qituvchisi, PhD
E-mail: anorxonimomaliyeva@mail.ru

O'zMU professori H.Xudaykulov taqrizi asosida

YOZMA TARJIMA KO'NIKMASINI TURLI TIL KORPUSLARI ASOSIDA RIVOJLANTIRISHNING NAZARIY ASOSLARI

Аннотация

So'nggi yillarda til korpuslaridan til instrumenti sifatida foydalanishga bo'lgan talab yildan-yilga o'sib bormoqda. Til korpuslaridan foydalanish uchun birinchi navbatda o'qituvchi kompyuter texnologiyalaridan foydalanishni puxta egallagan bo'lishi lozim. Korpus lingvistikasi kontekstida "tabiiy ravishda yuzaga kelgan matnlar" iborasi ko'p ishlatiladi. Tabiiy ravishda yuzaga kelgan matnlar o'zi nima degan savol tug'iladi. Tabiiy ravishda yuzaga kelgan matnlar bu – o'rganish maqsadida sun'iy yaratilmagan, tabiiy kommunikatsion sharoitlarda hosil bo'lgan matnlar tushuniladi. Bu matnlar kundalik muloqotlar, gazeta maqolalari, kitoblar, elektron pochta xabarlari, ijtimoiy tarmoqlar yozishmalari va boshqa ko'plab narsalarni o'z ichiga oladi. Ular turli sharoitlarda tilning haqiqiy qo'llanilishini aks ettiradi.

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ РАЗВИТИЯ НАВЫКОВ ПИСЬМЕННОГО ПЕРЕВОДА НА ОСНОВЕ РАЗЛИЧНЫХ ЯЗЫКОВЫХ КОРПУСОВ

Аннотация

В последние годы наблюдается постоянно возрастающий интерес к использованию языковых корпусов как инструмента языковедения. Первоначально, для использования языковых корпусов, необходимо, чтобы преподаватель в совершенстве овладел компьютерными технологиями. В контексте корпусной лингвистики часто используется фраза "естественно возникшие тексты". Возникает вопрос: что такое естественно возникшие тексты? Под естественно возникшими текстами понимаются тексты, которые не созданы искусственно в учебных целях, а сформированы в естественных коммуникативных условиях. Эти тексты включают в себя повседневные разговоры, статьи в газетах, книги, сообщения в электронной почте, записи в социальных сетях и многие другие виды материалов. Они отражают реальное использование языка в различных условиях.

THEORETICAL FOUNDATIONS FOR DEVELOPING WRITTEN TRANSLATION SKILLS BASED ON VARIOUS LANGUAGE CORPORA

Annotation

In recent years, the demand for using language corpora as a tool for linguistic analysis has been steadily increasing. Initially, to utilize language corpora effectively, educators must master computer technologies. Within the context of corpus linguistics, the phrase "naturally occurring texts" is frequently employed. This raises the question: what are naturally occurring texts? Naturally occurring texts are understood to be those that are not artificially created for study purposes but are formed in natural communicative settings. These texts include everyday conversations, newspaper articles, books, email messages, social network posts, and many other items. They reflect the actual use of language in various conditions.

Mavzuga oid adabiyotlar tahlili. O'zbek tilining ta'limiy korpusida til korpuslarini ta'limga integratsiya qilish bo'yicha quyidagi fikrlar berilgan: "Ta'limiy korpus ona tili hamda xorijiy tilni o'rganishda birdek ahamiyatli. Tilni o'rgatishda lug'at boyligining kattaligini ko'rsata olish, so'zning qo'llanish imkoniyatini u yoki bu grammatik konstruktsiya orqali tushuntirish uchun misollar massivini ko'rsatishda korpus juda qo'l keladi. Til ta'limi uchun muhim bo'lgan misolning doimiy yangilanib borishi, buni aks ettirib turish xususiyati hamda imkoniyati faqat korpusda mavjud. O'qituvchi yangi, ishonarli, cheksiz hamda xilma-xil misollarni shu yerdan topa oladi, topshiriq, mashqlarni belgilashda qiynalmaydi, bir necha daqiqada mavzu bo'yicha yangi-yangi misollardan iborat topshiriqlarni tayyorlay oladi.

Ta'limda korpusdan foydalanish quyidagi jarayonlar uchun zamonaviy instrumentariy vazifasini bajaradi;

Tilni o'rganuvchi, uning tashuvchisi bo'lgan kishi nutqini qiyoslash;

Til uslublarini kitob, darslik, nashr materiali bo'yicha qiyoslash, farqini aniqlash;

Gaplarda so'z tartibini aniqlash;

Adabiy og'zaki/yo'zma tilni chog'ishtirish"[1].

1-jadval. Ixtisoslashgan va sohaviy korpuslarning me'zonlaridagi farqlari

Mezonlar	Ixtisoslashgan korpus	Sohaviy korpus
Maqsad	Tilning maxsus xususiyatlarini o'rganish	Muayyan sohadagi tilni o'rganish
Tarkib	Tilga oid har xil mavzulardagi matnlar	Ma'lum bir sohadagi matnlar
Qamrov	Tilning tor bir aspektiga qaratilgan	Sohaga oid matnlarning keng doirasi
Misol	Frazeologizmlar korpusi	Iqtisodiy matnlar korpusi

Ixtisoslashgan korpuslar biror tilning muayyan jihatini chuqur tahlil qilish uchun yaratiladi. Tilshunoslik, grammatika, uslub yoki boshqa bir til xususiyatini o'rganishga mo'ljallangan korpuslar. Masalan, frazeologizmlar, murakkab sintaksis yoki dialektlarni o'rganish uchun maxsus yig'ilgan matnlarni o'z ichiga oladi. Har xil janrlar va mavzulardan iborat bo'lishi mumkin, lekin bitta xususiyatga moslashtirilgan bo'ladi.

Bu turdagi korpuslar akademik tadqiqotlar, til xususiyatlarini aniqlash yoki qiyosiy tahlil uchun ishlatiladi. Masalan, o'zbek tilidagi shevalarni o'rganish uchun maxsus shevalar korpusi.

Child Language Data Exchange System [5] (CHILDES) bolalar tilini o'rganish uchun maxsus yaratilgan korpus bo'lib, bolalar tilini o'rganish bo'yicha eng yirik ochiq ma'lumotlar bazalaridan biri bo'lib, u bolalar tilini rivojlantirishni ilmiy tahlil qilish uchun mo'ljallangan. Mazkur korpusning tarkibida 26 tildagi 230 dan ortiq korpuslar, jumladan, 1960-yildan beri yig'ilgan transkriptlar, audio va video materiallar mavjud. CHILDES tizimida transkripsiyalarni standartlashtirish uchun CHAT (Codes for the Human Analysis of Transcripts) formati qo'llaniladi. Bu format fonologik va morfologik tahlil uchun maxsus ishlab chiqilgan.

Bolalar va kattalar nutqini transkripsiya qilish, shuningdek, ikkinchi til o'rganish jarayonini tahlil qilish imkoniyatini beradi. CLAN (Computerized Language Analysis) – CHILDES uchun maxsus ishlab chiqilgan dastur bo'lib, u transkripsiyalarni tahlil qilish, so'z chastotasi va grammatik tuzilmani o'rganish uchun ishlatiladi. CLAN nafaqat transkripsiyalarni ko'rish, balki tahlil va statistik ma'lumotlar tuzishga ham imkon beradi. O'z tarkibiga quyidagi ma'lumot turlarini qamrab oladi:

Bolalar nutqi (yakka tartibda yoki guruhlarda).

Kattalar tomonidan bolalarga yo'naltirilgan nutq.

Ikkinchi tilni o'zlashtirishga oid ma'lumotlar.

Maxsus ehtiyojli bolalar (afaziya (nutqdagi kamchiliklar), nutq buzilishlari va boshqalar) bilan bog'liq materiallar.

CHILDES korpusi quyidagi tadqiqotlar uchun moslashgan:

Lingvistik tadqiqotlar.

Pedagogik tadqiqotlar.

Psixologik va kognitiv rivojlanish tahlillari.

CHILDES quyidagi amaliy tadqiqotlarda qo'llaniladi:

1. Tilni rivojlantirish tadqiqotlarida. Bolalar qanday qilib o'z ona tillarini o'zlashtirishini kuzatish va tahlil qilish.

2. Ikkinchi tilni o'rganish uchun. Ikkinchi tilni o'rganuvchilar uchun qiyinchiliklarni aniqlash va ularni yengib o'tish bo'yicha strategiyalarni ishlab chiqish.

3. Nutq buzilishlarini o'rganishda. Nutq va til bilan bog'liq muammolarni tahlil qilish, jumladan, maxsus ehtiyojli bolalar uchun strategiyalar yaratish maqsadida.

4. Lingvistik qiyosiy tahlil. Turli tillardagi grammatik va semantik xususiyatlarni solishtirishda ishlatiladi.

Shakespeare Corpus [6] – bu Uilyam Shekspirning asarlarini tahlil qilish va o'rganish uchun yaratilgan lingvistik korpus bo'lib, ushbu korpus Shekspir tilining boyligini, uslubini va grammatik xususiyatlarini chuqur o'rganishga imkon beradi. U adabiyotshunoslik, lingvistika va tarjimonlik sohalarida keng qo'llaniladi.

Tadqiqot metodologiyasi. Sohaviy korpuslar ma'lum bir soha yoki kasbga oid matnlarni o'rganishga yo'naltirilgan bo'lib, asosan, terminologiya, uslub va tilni kasbiy kontekstda o'rganish uchun ishlatiladi. Masalan, tibbiyot, huquq, muhandislik yoki iqtisodiyot sohalaridagi matnlar korpusi. Bu kabi korpuslar amaliy maqsadlarda ishlatiladi. Masalan, kasbiy tilni o'rgatish (huquqshunoslik, diplomatik so'zlashuvlari). Sohaviy korpuslar maxsus soha uchun lug'atlar tuzish, soha tilidagi tarjimonlarni tayyorlash maqsadida yaratiladi.

English Medical Corpus [7] tibbiyotdagi matnlarni qamrab oluvchi korpus.

English Medical Corpus (EMC) – ingliz tibbiyot sohasidagi matnlarni o'z ichiga oluvchi maxsus til korpusi bo'lib, u tibbiyot sohasidagi tilni o'rganish, tarjima qilish va tadqiqotlar olib borishda qo'llaniladi. Ushbu korpus tibbiy terminologiyani,

nutq uslublarini va ilmiy yozish odatlarining tahlilini amalga oshirish uchun mo'ljallangan.

EMCning tarkibi quyidagilardan iborat; tibbiy ilmiy maqolalar, tadqiqot maqolalari, jurnal nashrlari, klinik hujjatlar, bemor kartalari, tibbiy hisobotlar, retseptlarni o'quv materiallari, tibbiy darsliklar va qo'llanmalar, ijtimoiy axborot matnlari, sog'liqni saqlash bo'yicha xalqaro tashkilotlarning rasmiy e'lonlari va ma'lumotlari, bemorlarning fikr-mulohazalari, klinik tadqiqotlarga asoslangan yozma sharhlarni o'z ichiga oladi.

Mazkur korpus orqali o'quvchilar terminlarni qidirish, chastotasini o'rganish va matnlarda ularning qo'llanilishini tahlil qilish imkoniyatiga ega. EMC tibbiy terminlarning turli kontekstlarda qanday ishlatilishini aniqlashga yordam beradi.

Masalan: "diagnosis" yoki "prognosis" kabi atamalarining turli ilmiy va amaliy qo'llanilishini tahlil qilish, tibbiy yozishmalardagi qisqartmalarni va terminologiyani tushunish ko'nikmasini rivojlantiradilar. EMC tarjimonlar uchun muhim vositadir. U tibbiyot matnlarini bir tilga tarjima qilishda kontekstual ma'lumot beradi.

EMC so'z chastotasi va kollokatsiyalarni o'rganishda foydalidir.

English Medical Corpus tibbiyot sohasidagi tilni o'rganish, tahlil qilish va targ'ib qilish uchun noyob resurs hisoblanadi. U tilshunoslik, tarjimonlik va tibbiy ta'limda samarali foydalaniladi. EMC orqali tibbiyot sohasida aniq va mazmunli kommunikatsiya ko'nikmalarini shakllantirish mumkin.

Tahlil va natijalar. 3) Parallel yoki tarjima korpuslari: Bu turdagi korpuslar ikki yoki undan ortiq tillardagi matnlarni o'z ichiga oladi. Bu esa bir til matnining boshqa tilga qanday tarjima qilinganini tahlil qilish imkonini beradi. Parallel korpuslar tarjimonlar va lingvistlar uchun til tuzilmalarini solishtirish va tarjima strategiyalarini o'rganish imkonini beradi.

Parallel korpusga misol tariqasida biz ish olib borayotgan o'zbek tili korpusini aytishimiz mumkin. Mazkur korpusga <https://uzbekcorpus.uz/> linki orqali kiriladi. Bizning korpusimizda o'zbekcha-inglizcha, inglizcha-o'zbekcha, o'zbekcha-ruscha, ruscha-o'zbekcha; o'zbekcha-turkcha, turkcha-o'zbekcha, o'zbekcha-loreyscha, koreyscha-o'zbekcha parallel matnlardan tashkil topgan.

UM-Corpus (Michigan Universiteti Korpusi) – bu ingliz-xitoy tilidagi statistik mashina tarjimasini tadqiqotlari uchun mo'ljallangan katta hajmdagi parallel korpusdir. Ushbu korpus kompyuter lingvistikasi va mashina tarjimasini sohasida ishlayotganlar uchun juda qimmatli, chunki u ingliz va xitoy tillari o'rtasidagi tarjima algoritmlarini ishlab chiqish va sinovdan o'tkazish uchun katta ma'lumotlar to'plamini taqdim etadi.

English-Spanish EuroParl Corpus – bu korpus Yevropa Parlamenti sessiyalarining ingliz va ispan tillaridagi yozuvlaridan iborat. Bu korpus siyosiy matnlar tarjimasini tahlil qilish va ikki til o'rtasidagi uslubiy va terminologik mosliklarni o'rganish uchun juda qimmatli [8].

JRC-Acquis Multilingual Parallel Corpus [9] – mazkur korpus Yevropa Ittifoqi rasmiy hujjatlarining 22 tilda tarjimasini o'z ichiga oladi. Ushbu korpus, yuridik matnlar tarjimasida qo'llaniladigan uslublar va terminologiyalarni o'rganishda foydali hisoblanadi.

Parallel korpuslar turli tillar o'rtasidagi lingvistik va tarjima tadqiqotlarida muhim rol o'ynaydi, chunki ular turli tildagi matnlarning qiyosiy tahlilini ta'minlaydi va tilshunoslar hamda tarjimonlar uchun zarur ma'lumotlar manbai hisoblanadi.

4) Dinamik yoki yangilanib turadigan korpuslar vaqt o'tishi bilan yangilanib turadi va zamonaviy til foydalanish tendentsiyalarini kuzatish imkonini beradi.

Dinamik yoki yangilanib turadigan korpuslar, vaqt o'tishi bilan muntazam ravishda yangilanadigan va kengaytiriladigan til materiallari to'plamidir. Bu turdagi korpuslar tilning o'zgaruvchanligi va rivojlanishini kuzatish maqsadida foydalaniladi, ayniqsa zamonaviy so'zlarning paydo bo'lishi, tilning ijtimoiy va madaniy o'zgarishlariga moslashuvi kabi jarayonlarni tahlil qilishda foydali manba sifatida xizmat qiladi.

Corpus of Contemporary American English (COCA) – bu Amerika ingliz tilida 1990-yillardan boshlab ishlatilayotgan so'zlar va iboralarni o'z ichiga olgan korpusdir. COCA har yili

yangilanib turiladi va 560 milliondan ortiq so'zni o'z ichiga oladi, bu esa tilshunoslar va tadqiqotchilarga Amerika ingliz tilidagi so'nggi o'zgarishlarni kuzatish imkonini beradi.

British National Corpus (BNC) – aslida BNC yaratilish jarayonida dinamik korpus sifatida rejalashtirilmagan bo'lsa-da, u vaqt o'tishi bilan yangilanib turilgan korpuslarni yaratish g'oyasiga zamin yaratgan. BNCning yangi formati – BNC2014, Britaniya ingliz tilidagi o'zgarishlarni kuzatish uchun ishlatiladi.

The Hansard Corpus – bu Buyuk Britaniya Parlamentining yozuvlari korpusi bo'lib, 1800-yillardan to hozirgacha bo'lgan davrni qamrab oladi. Hansard muntazam ravishda yangilanib turadi va siyosiy tilning o'zgarishlarini kuzatishda qimmatli manba hisoblanadi.

5) Multimodal korpuslar: Ushbu korpuslar faqat yozma matnlar emas, balki audio, video va boshqa multimodal materiallarni ham o'z ichiga oladi. Ular tilni kengroq kontekstda, masalan, nutqiy kommunikatsiya yoki noverbal elementlar bilan birgalikda tahlil qilish imkonini beradi.

TalkBank – bu multimodal korpus nutqiy kommunikatsiyani tadqiq qilish uchun yaratilgan bo'lib, turli yoshdagi va turli tillarni so'zlovchi shaxslarning audio va video yozuvlarini o'z ichiga oladi. Bu korpus, o'z ichiga olgan ma'lumotlar orqali nutqiy rivojlanish, til o'rganish jarayonlari va tilning ijtimoiy foydalanishini tahlil qilish imkonini beradi.

TalkBank – bu o'z ichiga turli tillardagi til ma'lumotlarini qamrab olgan, dunyoning eng yirik ochiq kirishli integratsiya qilingan til ma'lumotlar bazasidir. TalkBank, tilshunoslik, psixologiya, ta'lim, kompyuter fanlari va nutq patologiyasi sohalaridagi tadqiqotchilarni qo'llab-quvvatlash uchun til korpuslari va boshqa audio resurslarni taqdim etadi.

TalkBankning yana bir muhim jihati, CHAT transkripsiya formatidan foydalanishi va CLAN dasturini til tahlili uchun ishlatilishi hisoblanadi. Bu dastur tilni tahlil qilishda turli xil foydalanish xususiyatiga ko'ra avtomatik hisoblab chiqarish imkonini beradi.

AMI Meeting Corpus [10] – bu korpus ishbilarmonlik uchrashuvlarining audio va video yozuvlaridan iborat bo'lib,

ishtirokchilarning muloqotlarini, imo-ishoralarini va mimikalarini qamrab oladi. Tadqiqotchilar bu ma'lumotlardan jamoaviy ish jarayonlarini va kommunikativ strategiyalarni o'rganish uchun foydalanadilar.

HCRC Map Task Corpus: Bu korpus ikki shaxs o'rtasidagi muloqotni audio va video formatlarda yozib oladi, ular bir-biriga xarita orqali yo'nalish ko'rsatish vazifasini bajarishadi. Ushbu korpus, tilning interaktiv ishlatilishini va kommunikativ vazifalarni bajarishdagi rolni tahlil qilish imkonini beradi.

Xulosa va takliflar. Xulosa qilib aytganda, multimodal korpuslar, nutqiy va yozma tilni tahlil qilish bilan bir qatorda, non-verbal kommunikatsiya elementlarini (masalan, imo-ishoralar, yuz ifodalari va ko'z harakatlari) ham qamrab oladi. Bu korpuslar, tilshunoslik, psixologiya, antropologiya va boshqa sohalarida qo'llaniladi va tilni kengroq madaniy va ijtimoiy kontekstlarda o'rganish imkonini beradi.

Har bir korpus turining o'ziga xos xususiyatlari va tadqiqot maqsadlari uchun alohida ahamiyati bor. Tilshunoslikda korpuslardan foydalanish, tilni chuqurroq tushunish va turli lingvistik fenomenlarni aniqroq tahlil qilish imkonini beradi.

Korpus lingvistikasida og'zaki nutqdan tashkil topgan korpusga tilning yozma asarlari to'planib, haqiqiy hayotiy suhbatlar, nutqlar va boshqa og'zaki faoliyatlarning audio yozuvlari tahlil qilinadi. Bu yozuvlar transkripsiya qilinadi va ko'pincha og'zaki tilning qo'llanilishini tushunishga yordam beradigan belgilarni o'z ichiga oladi[11].

Korpus lingvistikasida "a principled collection" atamasi ishlatilib mazkur ibora "Prinsipga asoslangan to'plam" degan ma'noni ifodalaydi[12]. Ya'ni aniq bir til shaklini o'rganish maqsadida qurilgan korpusni ma'lum mezonlar va maqsadlar asosida tuzilganini anglatadi. Bu, korpus hajmi, turli matn turlari (masalan, rasmiy va norasmiy) kiritilishi, so'zlashuvchilar yoki yozuvchilarning metama'lumoti va boshqa omillar kabi, tilni muvozanatli va maqsadli tarzda ifodalashga yordam beradigan asoslar tushuniladi.

ADABIYOTLAR

1. <https://uzschoolcorpara.uz/>
2. <https://www.english-corpora.org/bnc/>
3. <https://uzschoolcorpara.uz/>
4. <https://www.english-corpora.org/coca/>
5. <https://childes.talkbank.org/>
6. <https://www.opensourceshakespeare.org/>
7. <https://www.sketchengine.eu/english-medical-corpus/>
8. <https://www.statmt.org/europarl/>
9. https://joint-research-centre.ec.europa.eu/language-technology-resources/jrc-acquis_en
10. <https://groups.inf.ed.ac.uk/ami/corpus/>
11. Abduraxmonova N.Z. Inglizcha matnlarni o'zbek tiliga tarjima qilish dasturining lingvistik ta'minoti (sodda gaplar misolida): Filol. fan. bo'yicha falsafa dokt. (PhD)...disc. –Toshkent, 2018. – 165 b.
12. Abdurakhmonova N. Linguistic Issues of Creating Parallel Corpora for Uzbek Multilingual Machine Translation System // BuxDU ilmiy axborotnomasi №6, 2020. – P. 60-68