

PARALLEL MATNLAR KORPUSINI YARATISHNING AMALIY VA NAZARIY AHAMIYATI

Akramdjanova Madina,

Toshkent davlat O'zbek tili va adabiyoti universiteti tayanch doktoranti

Annotatsiya. Korpuslar, birinchi navbatda, lingvistik tadqiqotlar uchun ma'lumotlar bazasidir. Tilning mavqeini oshirish, ijtimoiy funksiyalarini kengaytirish, kommunikatsiyadagi o'rnini belgilash, axborot asri talablariga muvofiq ravishda takomillashtirishga doir tadqiqotlar olib borilmoqda. Hozirgi texnik imkoniyatlardan foydalangan holda milliy tillarning umumiy ma'lumotlar bazasini yaratish, shu asosda tilning semantik imkoniyatlarini, mazmun ifodalash ko'lamini belgilash global taraqqiyotning dolzarb masalalaridan biridir. Har bir tilda yaratilgan matnlar korpusi, birinchi navbatda, shu tilning lug'at boyligini aks ettirishga xizmat qiladi. Grammatik va uslubiy xususiyatlarini aks ettiradi. Jumladan, parallel matnlar korpusi tarjima qilinayotgan tillar o'rtasidagi lakunar hamda realiya birliklarni to'liq ifodalash imkonini beradi. Ushbu maqolada parallel matnlar korpusining ahamiyati haqida fikr yuritilgan.

Kalit so'zlar: Korpus, korpus lingvistikasi, kompyuter lingvistikasi, parallel matnlar korpusi, chog'ishtirma aspekt, badiiy asar bo'yicha parallel matnlar korpusi.

Annotation. Corpora, first and foremost, serve as databases for linguistic research. Extensive studies are being conducted to enhance the status of language, expand its social functions, determine its role in communication, and improve it in accordance with the demands of the information age. Utilizing modern technological capabilities to create comprehensive databases of national languages, and on this basis to identify the semantic potential of a language and the scope of meaning expression, is one of the pressing issues of global development. A corpus of texts created in any language primarily reflects the lexical richness of that language, as well as its grammatical and stylistic features. In particular, a parallel text corpus provides the opportunity to adequately represent lacunae and realia between the languages involved in translation. This article discusses the significance of parallel text corpora.

Keywords: Corpus, corpus linguistics, computer linguistics, parallel texts corpus, corpus aspect, parallel texts corpus on the work of art.

Аннотация. Корпусы – это, прежде всего, база данных для лингвистических исследований. Исследования проводятся с целью повышения статуса языка, расширения его социальных функций, определения его места в коммуникации и его совершенствования в соответствии с требованиями информационной эпохи. Создание общей базы данных национальных языков с использованием современных технических возможностей и на этой основе определение семантических возможностей языка и объема содержательного выражения является одной из актуальных задач глобального развития. Корпус текстов, созданный на каждом языке, служит, прежде всего, для отражения лексики этого языка. Он отражает его грамматические и стилистические особенности. В частности, корпус параллельных текстов позволяет в полной мере выразить лакунарные и реальные единицы между переводимыми языками. В данной статье рассматривается значение корпуса параллельных текстов.

Ключевые слова: Корпус, корпусная лингвистика, компьютерная лингвистика, корпус параллельных текстов, сравнительный аспект, корпус параллельных текстов по произведению искусства.

Kirish. Korpus lingvistikasi kompyuter lingvistikasi yo'nalishi sifatida shakllandi. Korpus lingvistikasi hal qilinishi lozim bo'lgan muammolar hajmining kengligi, vazifalarning keng ko'lamdaligi bois alohida soha sifatida rivojlanmoqda. Kompyuter lingvistikasi, korpus lingvistikasi amaliy tilshunoslik yo'nalishlari hisoblanadi.

Har bir tilning mavqeyi axborot-kommunikatsiyadagi o'ri bilan belgilanadi. Tilning axborot almashuv funksiyasini ta'minlashda kompyuter lingvistikasining, korpus lingvistikasining ahamiyati katta. Korpuslar muayyan tilning xususiyatlarini yoritishda, imkoniyatlarini aks ettirishda, tilshunoslik sohalarini, xususan, kompyuter leksikografiyasini takomillashtirishda, ijtimoiy soha tushunchalarini ommalashtirishda amaliy ahamiyat kasb etadi.

Tadqiqot metodologiyasi. Kompyuter lingvistikasi taraqqiyotini ta'minlovchi asosiy lingvistik omil elektron matnlar korpusi yoki parallel matnlar korpusidir. *Korpus* (korpus) lotincha so'z bo'lib, "tana" degan ma'noni bildiradi. "Korpus so'z, so'z birikmasi, grammatik shakllarni, so'z ma'nosini muayyan qidiruv tizimi orqali topishni anglatuvchi elektron ko'rinishdagi matnlar jamlanmasidir" [1]. Korpuslar barcha yo'nalishdagi tadqiqotlar uchun muhim manba sifatida xizmat qiladi. Korpus lingvistikasining kommunikatsiya, axborot almashuv, ilmiy tadqiqotlardagi ahamiyatli o'ri tadqiqotchidan tavsiya qilingan adabiyotlardagi ma'lumotlarni tahlil qilish, qiyoslash va baho berish ko'nikmalariga, korpus lingvistikasining dasturiy vositalari va axborot resurslari bilan ishlash malakasiga ega bo'lishni talab etadi.

Adabiyotlar sharhi. Korpus lingvistikasining shakllanishi, taraqqiyoti va nazariy asoslari tadqiqotlarda yoritilgan. Korpus lingvistikasi, uning predmeti izohlangan, ilmiy adabiyotlardagi ta'riflari til korpusi ma'lum tilning belgilangan davrdagi, xilma-xil janr, rang-barang uslub, hududiy hamda ijtimoiy variantdagi matnlarning elektron shakldagi maxsus dasturiy ta'minot asosidagi yig'indisi sifatida umumlashtirilgan [2].

Natijalar. Korpus lingvistikasida parallel matnlar korpusi muhim ahamiyat kasb etadi. Parallel matnlar korpusi badiiy asar, qo'llanma, ommaviy axborot vositalari, turli xil hujjatlarning ikki yoki undan ko'p tillardagi elektron ko'rinishlaridir. Parallel matnlar korpusi chog'ishtirma aspektdagi tadqiqotlarda o'ziga xos o'rin tutadi. Parallel korpuslar yordamida bir so'z, jumla, abzats, supersintaktik butunliklarning turli tillardagi variantlarini bilish mumkin. Parallel korpuslar – madaniyatlararo muloqot keng yoyilgan hozirgi davr uchun muhim bo'lgan ma'lumotlar bazasi. Parallel korpuslar orqali turli til muhitidagi, madaniyatlaridagi universalialar hamda tillarning o'ziga xos mental xususiyatlari, realiya va lakunar birliklarni aniqlash mumkin bo'ladi. Parallel matnlar korpusi avtomatik tarjima rivoji uchun ham xizmat qiladi, kompyuter leksikografiyasining taraqqiyotini ta'minlaydi. Parallel matnlar korpusi yordamida maxsus konkordanser dasturlar ishlab chiqiladi va turli xil ixtisoslik lug'atlari tuzish imkoniyati yuzaga keladi.

"Parallel matnlar korpusi ilmiy va amaliy maqsadlarda (shu jumladan, xorijiy tillarni o'qitish maqsadida) amalga oshirilmoqda. Parallel matnlar strukturasi ko'ra manba tildan maqsad tilga tarjima matnlardan iborat. Masalan, "Alie in Wonderland" ingliz matni va uning nemis, fransuz, rus tillaridagi tarjimasini parallel matnlarni yaratishga asos bo'ladi" [3].

Parallel matnlar korpusining shakllanishi tilning nufuzini oshirish, mavqeini mustahkamlashga xizmat qiladi. Avvalo, parallel matnlar korpusi ma'lumotlar bazasi sifatida kommunikativ funksiya bajaradi. Bir tildagi matnning boshqa tildagi tarjimasida madaniyatlararo munosabatlarni aniqlashga, farqlarni ko'rsatishga yo'naltiriladi. Leksik birliklarni boshqa tillardagi tarjima variantlari orqali semantik tahlil qilishga imkoniyat yaratiladi. Axborot uzatishning turli tillardagi umumiy va spetsifik xususiyatlarini aniqlash imkoni yuzaga keladi.

Badiiy asar bo'yicha parallel matnlar korpusi, avvalo, ikki yoki bir necha tillar grammatikasidagi o'xshash va farqli jihatlarni tahlil qilish imkonini beradi. Parallel matnlar korpusi har ikki tildagi badiiy uslub xususiyatlarini qiyoslashda ahamiyatlidir. Badiiy uslubda tasvir vositalari, ko'chimlar, obrazli ifodalar har bir til an'anasiga mos ravishda aks etadi [4]. Muallif nutqida tillarning grammatik xususiyatlari o'z ifodasini topsa, qahramonlar nutqida adabiy til va sheva so'zlarining uyg'un ravishda qo'llanishi obrazli tasvir mahoratini yoritishga xizmat qiladi.

Badiiy matnlar asosidagi korpus muayyan tillarga xos birliklar – realiya birliklarni ham o'z ichiga oladi. Parallel matnlar korpusi realiya birliklar tarjimasidagi tamoyillarni aniqlashga imkon beradi. Realiza birliklarning kalkalash, semantik kalkalash, ekvivalent so'zni tanlash kabi usullardagi tarjimalari borasida ma'lumotlar olish mumkin bo'ladi.

Badiiy asarlar asosidagi parallel matnlar korpusida frazemalar bilan bog'liq muammolar bo'lishi aniq. O'zbek tilidagi frazemalar ikki yoki undan ortiq so'zdan tuzilib, shu so'zlar semantikasi asosida yangi ko'chma lug'aviy ma'noni shakllantirishga xizmat qiladi. Shu bois parallel matnlar korpusini shakllantirishda to'g'ridan to'g'ri mashina tarjimasidan foydalanilsa, semantikani to'g'ri yoritishda muammolar kelib chiqadi. Muammolarni bartaraf etish uchun frazemalar sodda va qo'shma leksik birliklar, so'z birikmalaridan farqlashi, maxsus teglar, razmetkalar bilan belgilanishi lozim. Frazemalar tarjimasida, albatta, mutaxassis nazorati kerak bo'ladi.

Parallel matnlar korpusi turli tillar asosida xilma-xil madaniyatlarni solishtirish, madaniy munosabatlarni ifodalovchi leksik birliklar mazmun-mohiyatini o'zlashtirish imkoniyatini beradi. Parallel matnlar korpusi orqali tillarning fonetik, leksik, morfologik, sintaktik xususiyatlarini qiyoslash, chog'ishtirish mumkin. Bunday korpuslar so'z turkumlariga xos kategoriyalar, grammatik ma'noning ifodalanishi, so'z yasaliş tizimiga oid ma'lumotlar bera olishi bilan ham e'tiborga molikdir.

Qo'llanma, o'quv adabiyotlari matni bo'yicha tuzilgan korpuslar ilmiy uslub xususiyatlarini yoritishda, muayyan soha bo'yicha nazariy ma'lumotlar berishga mo'ljallangan. Ommaviy axborot vositalari doirasida tuzilgan parallel matnlar korpusi ommaviy nashrlarning turi, mazmuni haqida to'liq ma'lumot beradi [5]. Matbuot nashrlarida yoritilayotgan masalalar mazmunini o'zlashtirishga yordam beradi. Parallel matnlar korpusiga televidenie va radiomatnlarning og'zaki varianti (audiomatnlarni) joylashtirish ma'lumotlar bazasining illyustrativligini oshiradi.

Rasmiy, ish yuritish hujjatlariga oid parallel matnlar turli tillardagi rasmiy ish yuritish uslubi, me'yoriy-huquqiy hujjatlar qolipini aniqlashda ahamiyatlidir. Parallel matnlarning bu ko'rinishi turli tillardagi rasmiy uslub xususiyatlarini qiyoslash, universal hamda farqli jihatlarini yoritish maqsadi qo'yilgan tadqiqotlar uchun manba vazifasini o'taydi.

Xulosa. Xulosa qilib aytganda, Parallel matnlar korpusidan til ta'limi, til o'qitish tizimida lingvistik ma'lumotlar bazasi sifatida foydalanish mumkin. Badiiy asarlar mazmunini o'zlashtirish, konseptual tahlil qilish, matn lingvistikasi asoslarini, turli tillarga xos matn xususiyatlarini o'rganishda parallel matnlar korpusining ahamiyati katta. Shuningdek, parallel matnlar korpusi lingvistik ma'lumotlar bazasi sifatida ham katta ilmiy qiymatga ega. U orqali morfologik, sintaktik va semantik darajadagi o'zgarishlar, ekvivalentlik muammolari, interferensiya holatlari va til birliklarining funksional yuklamasi tizimli ravishda tadqiq etilishi mumkin. Bu esa nazariy tilshunoslik bilan bir qatorda amaliy yo'nalishlar — tarjimashunoslik, korpus lingvistikasi va kompyuter lingvistikasi rivojiga ham xizmat qiladi.

Foydalanilgan adabiyotlar ro'yxati:

1. <http://rusorpora.ru>
2. Hamroeva Sh. M. O'zbek tili mualliflik korpusini tuzishning lingvistik asoslari: Filol. fan. f. d. (PhD) dis. avtoref. – Qarshi, 2018. – B.9
3. Баранов А.Н. Введение в прикладную лингвистику. - М.: Эдиториал УРСС, 2001.-S.61.
4. Po'latov A. Kompyuter lingvistikasi. –T., 2011;
5. Rahimov A. Kompyuter lingvistikasi asoslari. –Toshkent: Akademnashr, 2011. –B.117