

PARALLEL KORPUS VA KORPUS LINGVISTIKASI

Qo'chqarova Shakhzoda Jalolkhon qizi

O'zbekiston davlat jahon tillari universiteti Tarjimonlik fakulteti 1-bosqich talabasi

Imliy rahbar: Ahmedova Anorkhan Nasibaliyevna

<https://doi.org/10.5281/zenodo.10597822>

Annotatsiya: Ushbu maqolada korpus, korpus lingvistikasi, uning tuzilishi, korpus turlari, token, lemma va stem haqida so'z boradi. Bugungi kunda korpusning nazariy va amaliy ahamiyati o'zbek tilshunosligida tilning imkoniyatlarini kengroq o'rganishda, tilshunoslikdagi muammoli tomonlarini aniqlashda, elektron lug'atlar yaratishda, zamonaviy texnologiyalardan foydalanish imkoniyatini oshirishda ahamiyati katta hisoblanadi.

Kalit so'zlar: Korpus, korpus lingvistikasi, parallel korpus, tarjima korpus, taqqoslanadigan korpus

PARALLEL CORPORA AND CORPUS LINGUISTICS

Abstract: This article discusses corpus, corpus linguistics, its structure, corpus types, token, lemma and stem. Today, the theoretical and practical importance of the corpus in Uzbek linguistics is considered to be of great importance in the wider study of the possibilities of the language, in identifying its problematic aspects in linguistics, in creating electronic dictionaries, and in increasing the possibility of using modern technologies.

Keywords: Corpus, corpus linguistics, parallel corpus, translation corpus, comparative corpus

ПАРАЛЛЕЛЬНАЯ КОРПУСНАЯ И КОРПУСНАЯ ЛИНГВИСТИКА

Аннотация: В данной статье рассматривается корпус, корпусная лингвистика, его структура, типы корпуса, токен, лемма и основа. Сегодня теоретическое и практическое значение корпуса в узбекском языкознании считается имеющим большое значение для более широкого изучения возможностей языка, выявления его проблемных сторон в языкознании, создания электронных словарей, увеличения возможностей с использованием современных технологий.

Ключевые слова: Корпус, корпусная лингвистика, параллельный корпус, корпус переводов, сравнительный корпус.

KIRISH

Dunyodagi jadal globallashuv tufayli tarjimaga va uning qulay hamda sifatli va xohlagan joydan turib foydalana oladigan shakliga bo'lgan talabimiz yanada ortmoqda. Yuqoridagi talablarni sifatli ravishda qondira oladigan va foydalanuvchiga eng sifatli tarjimalarni yetkazib beradigan parallel korpuslar ustidagi dastlabki ishlar 1980-yildan boshlab jadal rivojlana boshladi. Bugungi kunga kelib esa uning salmog'i tobora kuchaymoqda. Parallel korpus ya'ni Parallel Corpora- Parallel tarjima matnlarining elektron analogi. Ko'plab "original matn va ularning bir/bir necha tarjimasi" bloklaridan iborat. Korpusdagi elektron matnlar orimatnning o'zi yoki uning bir qismi bo'lishi mumkin. Parallel korpus- original matn va uning tarjimasi yig'ilgan korpus hisoblanadi. Parallel korpuslar deganda bir tildagi manbaning ikki yoki undan ortiq tillarda aks etishi tushuniladi hamda u manba tili va uning muvofiq tarjimasini jamlagan tarjima tili tizim hisoblanadi. Uning kuchli qidiruv, taqqoslash, xotira va boshqa jihatlari, shuningdek korpusning ishonchliligi, uning leksik ma'lumotlar bilan boyitilganligi, ma'lumotlarning aniqligi an'anviy tarjimadagi kamchiliklarni bartaraf etuvchi asosiy omil bo'lib xizmat qiladi. Parallel korpuslar istalgan bir tilning asli va uning tarjimasini o'zida jamlagan elektron tizim hisoblanadi.

ASOSIY QISM

Til korpuslari tarixiga nazar soladigan bo'lsak, XX asrning so'nggi yillarida korpus tilshunosligi alohida qism sifatida ajralib chiqdi va 1950-1960-yillarda dastlabki muhim tadqiqotlar amalga oshirildi. Korpus tilshunosligi- hisoblash tilshunosligining bir qismi bo'lib, u kompyuter texnologiyalaridan foydalangan holda lingvistik korpuslarni qurish va ishlatishning umumiy tamoyillaridan ishlab chiqadi. Matnlarning lingvistik yoki til korpusi deganda, muayyan til muammolarini hal qilish uchun mo'ljallangan, mashinada o'qiladigan, birlashtirilgan, tuzilgan, belgilangan, fiziologik jihatdan mukammal lingvistik ma'lumotlar to'plami tushuniladi. Matn korpuslari parallelligiga ko'ra bir tilli, ikki tilli va ko'p tilli turlarga bo'linadi. Parallel korpuslar yoki bitekstlar manba tilidagi ma'lumotlarning tarjima tiliga o'tkazilgan ko'rinishi bo'lib, bunda matnlar, paragraflar va gaplar tarkibi so'zlarga ajratilgan holda ikki tildagi bir-biriga mos so'zlarning o'zaro bo'g'lanishidir.

Taqqoslanadigan korpus- matnlar janri va tarkibiga ko'ra bir xil, ammo ular bir-birlarining tarjimalari emas. Parallel matnlardan foydalanish uchun matn segmentlarini identifikatsiyalash va matn turlari va janrlarini o'zaro moslashtirish, muvofiqlashtirish eng birinchi shartlardan hisoblanadi. Ikki til o'rtasida tarjima qilish ishlarini amalga oshirish uchun mashina tarjimasini algoritmlari ko'pincha birinchi til korpusining elementli tarjimasini bo'lgan birinchi til korpusi va ikkinchi til korpusidan tashkil topgan parallel qismlar yordamida amalga oshiriladi. **Tokenlash-** tabiiy tildagi nutq birliklari alohida tarzda grammatik ma'nolarga ajratib beriladi, so'z shakllari aniqlanadi. **Lemmalash-** so'z shakllarining dastlabki qismi, ya'ni lug'atdagi ko'rinishi aniqlanadi. **Stemming-** so'zning o'zak qismi aniqlanadi.

XULOSA

Bugungi kunda korpus vaqtini va mehnatni tejaydigan, istalgan joyda foydalanish mumkin bo'lgan ish quroliga aylanib ulgurdi desak xato bo'lmaydi. Parallel korpuslar nafaqat tillarni o'qitishda, balki mashina tarjimasini uchun ham xizmat qiladi. O'zbek va ingliz tili korpuslari haqida gap ketar ekan, ularning turli xil til oilalariga mansubligi, o'zaro tarjima manbalarining tanqisligi esa parallel korpus yaratish jarayonini yana bir marta qiyinlashtiradi. Shunga qaramay, mavjud manbalar asosida o'zbek-ingliz tillari korpusini yaratish yuqori natija sari ilk qadam bo'lib xizmat qilmoqda. Shundan kelib chiqamiz-ki, yaqin orada o'zbek-ingliz korpusini yaratish ham o'z yechimini topadi.

Foydalanilgan adabiyotlar:

1. Umerova.M.-V.: Parallel corpora in translation studies. Sciences of Europe.2018,29-39. <https://doi.org/10.2991/iceme-17.2017.71>
2. Ro'ziyev H, Xodjayeva N. (2021). Matnlarning parallel korpuslarini tuzish muammolari. ACADEMIC RESEARCH IN EDUCATIONAL SCIENCES
3. Abdurakhmonova N. Tuliyeu U. Morphological analysis by finite state transducer for Uzbek-English machine translation/ Foreign philosophy: Language, Literature, Education
4. Alberto Simoes, Jose Joao Almeida.2007. Parallel Corpora based Translation Resources Extraction. Procesamiento del Lenguaje Natural https://www.academia.edu/3219071/Parallel_Corpora_based_Translation_Resources_Extraction
5. Nazira G'anijon qizi Sobirova, Madina Akramajanova, Shohjahon G'anijon o'g'li Sobirov. (2022). Parallel korpus xususiyati, tuzilishi, interfeysi. ACADEMIC RESEARCH IN EDUCATIONAL SCIENCES