

BADIIY ASARLAR VA ULARNING ELEKTRON PARALLEL KORPUSIDAGI USLUBLAR

Shahzoda Dilmurodova Istam qizi

Buxoro davlat universiteti Filologiya fakulteti 2- kurs talabasi

Annotatsiya: Maqolada badiiy asarlardan tarkib topgan elektron parallel korpuslarning xususiyatlari va ulardagi uslubiy yo'nalishlar tahlil qilinadi. Badiiy matnlarda o'ziga xos badiiy uslublarning namoyon bo'lishi, ularni korpus texnologiyalari orqali tadqiq qilish va parallel korpuslarda turli tillar o'rtasidagi o'xshashlik va farqlarni aniqlash masalalari o'rganiladi. Maqolada, shuningdek, elektron korpuslardagi badiiy matnlarning leksik, grammatik va stilistik jihatlarini ham tahlil qilinadi. Natijada, badiiy asarlardagi uslublar va ularning turli tillardagi tarjimalari orqali uslubiy elementlarni avtomatik aniqlash imkoniyatlari muhokama qilinadi. Ushbu tadqiqot badiiy matnlarning korpus lingvistikasi orqali tahlilining nazariy va amaliy ahamiyatini ochib beradi.

Kalit so'zlar: Korpus, elektron korpus, afzalliklari, tadqiqotlar, lingvistika, mashina tarjimasi, parallel korpus.

Parallel korpusning qimmatini uning hajmi, tillarning miqdori bilan belgilanadi. Korpusdagi elektron matnlar original matnning o'zi yoki uning bir qismi bo'lishi mumkin. Parallel korpus – original matn va uning tarjimasi yig'ilgan korpus. Parallel korpuslardan kutilayotgan natija - matnni to'liq avtomatizatsiya qilish, bu orqali nafaqat tarjimonlikda, balki mashina tarjimasi uchun ham qo'llashdir.

Elektron korpusning afzalliklaridan biri - tilda so'zning o'zgarishi, istorizm, neologizm, ma'no kengayishi va torayishi, yangi frazeologizmlarning paydo bo'lishini kuzatish mumkin. Davrlararo tilning boyib borish jarayonini ko'rsatib turuvchi elektron manba bo'lgan til korpuslariga doimiy ravishda yangiliklarning kiritilishi, joylanib turishi inson mehnatini yengillatadi. Kitob shaklidagi lug'atlarga yil o'tgani sari tilda vujudga keladigan o'zgarishlarni kiritish ancha kuch va mablag' sarf etishni talab qiladi. Korpus esa bu muammoni qisqa muddatda, hech qanday xarajatsiz bajarilish imkoniyatini beradi. Elektron korpusning parallel korpus avtonom qismi juda ko'plab kerakli axborotlarni o'zida jamlay olish xususiyati bilan ahamiyatga molikdir. Mashina tarjimasi yo'nalishida yonma - yon taqqoslash uchun maxsus formatlangan ko'p tilli korpuslar mavjud bo'lib, ular tizimlangan parallel korpus deb ataladi. Parallel matnlar korpusining dastlabki namunasi sifatida 1799-yilda Rozetta shahri yaqinidagi Nil deltasida topilgan mil.av 196-yilga borib taqaladigan, Misr ibodatxonalari tomonidan ikki tilda (yunon va misr) va uchta yozuv tizimida (Misr iyerogliflari va demotik tilda yozilgan) shoh Ptolomey Vga berilgan sharaflar haqida so'z yuritilgan tosh hisoblanadi. Parallel korpusning tuzilishi, tarkibi va imkoniyati

“IZLANISH SAMARALARI”

№9 mavzusidagi respublika an'anaviy ilmiy-nazariy anjumani

haqidagi ma'lumotlarni D.O.Dobrovolskiy, Yu.Tao, V.Zaxarov, A.A.Kokoreva, E.P.Sosninalarning ishlarida kuzatamiz . [3.1]

Parallel korpus, ya'ni asl nusxalar to'plami va ularning tarjimalari tarjimashunoslik, mashina tarjimasi, lingvistika, hisoblash tilshunosligi yoki oddiygina inson tarjimoni manfaati uchun turli yo'llar bilan ishlatilishi mumkin. Hisoblash tilshunosligida tarjima korpuslari mashina tarjimasi uchun, shuningdek, atamalarni ajratib olish, so'z ma'nosini aniqlash va hokazolar uchun 1980-yillarning boshida qo'llanilgan. Birinchi parallel matnlar sifatida Shveysariyada nemis, fransuz, italyan tillarida to'plangan qor ko'chkisi haqidagi hisobotlar, Kanada ommaviy axborot vositalari ingliz, fransuz tillarida berilgan ob-havo ma'lumotlari 1980-yillarning oxiri 1990-yillarning boshlarida paydo bo'lgan. Dastlabki elektron resurslardan biri Kanada Hansard bo'lib, u dastlab jummalarni moslashtirishni amalga oshirish uchun ishlatilgan. [14.4]

Bu vazifa hozirda tarjima xotiralari kabi ilovalarning standart xususiyati hisoblanadi. Bundan tashqari, parallel korpuslar ko'p tilli grammatik induksion, avtomatik leksikografiya va turli tillarda ma'lumot olish va tilni qayta ishlashda boshqa ko'plab vazifalar uchun ma'lumotlar bazasi sifatida ishlatiladi . Ushbu yillarda yaratilgan loyihalarning barchasidan oxirgi maqsad mashina tarjima tizimini yaratish bo'lgan. [3.4]

Dastlabki lingvistik korpus o'tgan asrning 60-yillarda paydo bo'lgan. 1963-yilda Braun universitetida (AQSh) birinchi marta mashina muhitida matnlarning katta korpusi yaratilgan (Brown corpus). Korpus mualliflari V. Frensis va X. Kucher hisoblanadi. [15.28]

Ushbu korpusga turdagi janrlardan tashkil topgan nasriy matnlardan iborat bo'lgan. Barun korpusining ommallashuvi shu sohaga doir olimlar orasida turli muhokamalarni yuzaga keltirdi. 90-yillarning birinchi yarmida korpus lingvistikasi til haqidagi alohida soha sifatida shakllanib ulgurdi. Shu bilan birga, u kompyuter lingvistikasi bilan birga bilan chambarchas aloqada bo'lib, uning yutuqlaridan foydalanadi va o'z navbatida kompyuter lingvistikasi uchun lingvistik resurs sifatida turli dasturiy ta'minotlar uchun tadqiqot obyekti bo'lib hisoblanadi. Korpus lingvistikasi matnlarni yi g'ish, tasniflash, annotatsiyalash kabi vazifalarni bajaradi. Korpus lingvistikasini kompyuter lingvistikasining asosiy yo'nalishi sifatida e'tirof etish mumkin, chunki matnni qayta ishlovchi dasturlar korpusga asoslanadi. NLP tizimi uchun lingvistik bilimlar juda muhim. Ular qabul qilingan muayyan lisoniy modellar, qoidalar, lug'atlar shaklida ifoda etilsada, biroq tilning nutqiy omili bilan bog'liq bo'lgan ekstralingvistik ma'lumotlarga ham ehtiyoji katta. Shu ma'noda

“IZLANISH SAMARALARI”

№9 mavzusidagi respublika an’anaviy ilmiy-nazariy anjumani

tilning konseptual tasviri aks etgan semantik tarmoqlar, ontologiya ko‘rinishidagi bilimlar ta’minoti zarur. Demak, tilning murakkab funksional imkoniyatlarini pragmatik, neyrolingvistik, psixolingvistik yoki diskursiv parametrlarini baholashda korpus texnologiyasi oldida muhim vazifalar mavjud. Kaliforniya universiteti mutaxassisi Stefan Gries o‘zining ilmiy qarashlarida korpus uchun berilgan ta’riflar “u metod, nazariya yoki modelmi”, degan savolga nisbatan uni metod(ologiya) deb baholaydi. Olim fikrini asoslab, til nazariyasining generativ nuqtayi nazaridan tilga aloqasi yo‘q, degan to‘xtanga keladi. U deskriptiv va amaliy jihatdan ayrim metodlardan foydalanib, misollar yordamida o‘z fikrini isbotlashga harakat qiladi. Unga ko‘ra, agar tilshunos biror leksemani korpus ichida grammatik birlik sifatida o‘rganmoqchi bo‘lsa, demak u grammatik nazariyaga, agar u yoki bu millatning ikkinchi til sifatida murakkab konstruksiyalardan qay darajada qo‘llash imkoniyatini baholamoqchi bo‘lsa, ikkinchi tilni o‘rganish nazariyasiga asoslanadi, ya’ni har bir izlanuvchi o‘ziga xos metodni tanlaydi, degan munosabatni bildiradi. Dastlabki ta’limiy korpuslarni yaratilishi Sylviane Granger va uning ilmiy jamoasi tomonidan “Ingliz tili o‘rganuvchilari uchun xalqaro korpus” yaratilgan. Ushbu korpusning maqsadi til o‘rganuvchilarining umumiy yuzaga keluvchi xatolarini tasniflash orqali ularni tuzatish metodlarini yaratishga qaratilgan. Korpus lingvistikasining asosiy maqsadi til tizimining lingvistiktavsifi sanaladi, bunda muloqotni o‘rganish jarayonida olingan nutqiy materiallar matnlarikorpusida aks etishie ‘tiborga olinadi.

[1.29]

Korpus lingvistikasining vazifalari quyidagilar hisoblanadi: 1)ushbu sohaning nazariy asoslarini ishlab chiqish; 2) har xil turdagikorpuslarni yaratish va ulardan foydalanish tajribasini tahlil qilish; 3)korpusga qo ‘yiladigan umumiy talablarni shakllantiri hash; 4) turli xiltadqiqot va o ‘quv vazifalari uchun korpuslarni yaratish; 5)tilshunoslikning turli sohalarida matn korpuslarini qo ‘llashning samarali usullarini shakllantirish. [2.8]

Badiiy matnlarda har xil uslublar (emosional, obrazli, subyektiv va shu kabilar) bo‘lib, ular asarning umumiy mazmunini to‘ldirib boradi. Ular turli tillarda turlicha ifodalanishi mumkin va shu bois, parallel korpusda uslubiy yondoshuvni tahlil qilish muhimdir. Korpus lingvistikasi turli tillardagi matnlarning tuzilishini va ularning o‘zaro aloqasini tahlil qilishga yo‘naltirilgan. Badiiy matnlarni tadqiq qilishda elektron korpuslardan foydalanish turli leksik va uslubiy elementlarni o‘rganishda muhim. Parallel korpuslar turli tillardagi badiiy matnlarning o‘xshash qismlarini topish va tahlil qilishda yordam beradi. Bu asarlar orasidagi uslubiy va mazmunli o‘xshashliklarni va farqlarni aniqlash imkonini beradi. Badiiy matnlarning korpusdagi

“IZLANISH SAMARALARI”

№9 mavzusidagi respublika an’anaviy ilmiy-nazariy anjumani

tahlili quyidagicha amalga oshiriladi: 1. Leksik tahlil: Korpusda badiiy asarlarda qo'llaniladigan so'zlar va lug'at birliklarining ko'rsatkichlari, ularning to'qilishi va aniq tarzda qaysi uslubiy maqsadga xizmat qilishi. 2. Grammatik tahlil: Turli tillardagi matnlarning grammatik tuzilishi va ularning o'ziga xosliklarini aniqlash. 3. Stil va uslubiy elementlar: Badiiy matnlarning korpusdagi uslubiy elementlari, masalan, metaforalar, tashbehtar, sintaksik vositalar va boshqa uslubiy elementlarni korpusda tahlil qilish.

Korpusda badiiy asarlarda qo'llaniladigan so'zlar va lug'at birliklari o'zining semantik, uslubiy va emosiyaga boy jihatlari bilan matnning umumiy mazmuniga ta'sir ko'rsatadi. Ular asarning badiiy qiymatini oshirish, o'quvchiga kuchli ta'sir o'tkazish va asarning asl maqsadini ifodalashga yordam beradi. Masalan, badiiy asarda ishlatilgan lug'at birliklari va so'zlar quyidagi tarzda turli uslubiy maqsadlarga xizmat qilishi mumkin:

Ma'naviy va mazmunli boylik yaratish: Badiiy asarlardagi so'zlar o'zida chuqur ma'nolarni mujassam etib, lug'at boyligini ko'rsatadi. Bu birliklar o'quvchida asarning mavzusi va g'oyasiga oid tushunchalarni shakllantirishda xizmat qiladi. Masalan, qahramonning hissiyotlari va ichki kechinmalari ifodalanganda, to'g'ri tanlangan so'zlar va tuzilmalar matnni hissiy jihatdan boyitadi.

O'ziga xos uslubiy yondoshuvni yaratish: Badiiy asarning o'ziga xos uslubini yaratish uchun muayyan so'zlar va lug'at birliklari, ayniqsa, biror muayyan davr, madaniyat yoki guruhga xos til vositalari qo'llaniladi. Masalan, tarixiy asarlarda o'sha davrga xos til, obrazlar va iboralar ishlatilishi asarning o'ziga xosligini va stilistik to'liqligini ta'minlaydi.

Qahramonlarning xarakterini yaratish: Qahramonning xususiyatlarini yaratishda maxsus so'zlar va lug'at birliklari qo'llaniladi. Ular orqali yozuvchi qahramonning ichki olamini, uning qarashlari va hissiyotlarini namoyon etadi. Misol uchun, shijoatli yoki yumshoq qahramonlarning nutqida qo'llaniladigan so'zlar, ularning ruhiyatini va yozuvchining ulug'lab yoki tanqidiy yondashuvini ko'rsatadi.

So'z va iboralarning semantik tarmoqlari: Badiiy asarlarda ko'pincha bir so'z turli ma'nolar bilan kelib, muloqot va matn ohangini boyitadi. Bu lug'at birliklarining turli ma'noda ishlatilishi asarning lirik yoki satirik yondoshuvini kuchaytirishi mumkin. Korpusda ushbu turdagi so'zlarning semantik tarmoqlari orqali asarning semantik strukturasi tahlil qilinishi mumkin.

Badiiy simbolika va obrazli yondoshuv: Ko'p hollarda asarlarda muayyan so'zlar va lug'at birliklari orqali obrazli yondoshuv yaratiladi. Ushbu obrazlar, metaforalar yoki simvollar o'quvchiga to'g'ridan-to'g'ri aytilmagan ma'nolarni ifodalaydi. Mazkur

“IZLANISH SAMARALARI”

№9 mavzusidagi respublika an'anaviy ilmiy-nazariy anjumani

birliklar yordamida asardagi obrazlar o'rtasidagi bog'liqlikni topish va ularning ma'nolarini avtomatik aniqlash imkoniyati yuzaga keladi.

Emosional ta'sirni kuchaytirish: Badiiy matnda qo'llanilgan so'zlar va lug'at birliklari o'quvchida muayyan hissiyotlarni uyg'otishga mo'ljallangan bo'ladi. Ular orqali asardagi voqealar va qahramonlar his-tuyg'ularini o'quvchiga ta'sirli tarzda yetkazish mumkin. Korpusda emosiyaga boy so'zlarni aniqlash, ularning qaysi asarlarda ko'proq uchrashi va ularning umumiy uslubiy ta'sirini o'rganish mumkin.6. Uslubiy o'zgarishlarni aniqlash uchun statistik modellar.Latent Dirichlet Allocation (LDA): LDA modeli matndagi turli mavzular va ularning qaysi uslubiy maqsadga xizmat qilishini aniqlashga qaratilgan. Bu modelning yordamida matndagi mavzular to'plamini, ularning semantik rivojlanishini va uslubiy yondoshuvni kuzatish mumkin. Hidden Markov Model (HMM): HMM statistik modeli orqali matndagi uslubiy elementlarning qaysi joylarda ko'proq yoki kamroq qo'llanganini aniqlash mumkin. Bu uslubiy elementlar ko'proq takrorlanadigan joylarni tahlil qilish imkoniyatini yaratadi.

7. Mashina o'rganishi va transformer arxitekturalari. Transformer modeli: Transformer asosidagi arxitekturalar, ayniqsa, uslubiy elementlarning turli tillardagi transformasiyasini aniqlash uchun samarali. Masalan, GPT va BERT kabi modellar matnni kontekstda tushunish va uslubiy elementlarni masofaviy aniqlashda foydalaniladi. Clustering Techniques (klasterlash usullari): Matndagi o'xshash uslubiy elementlarni topish va ularni bir klasterga to'plashda klasterlash usullari, jumladan K-means algoritmi, keng qo'llaniladi. Bu uslubiy aloqador elementlarning joylashuvini tahlil qilish uchun qulay.

8. Badiiy matnlarni korpusda uslubiy tahlil qilish. Corpus Analysis Tools: Badiiy matnlarni korpus asosida tahlil qilish uchun Sketch Engine, AntConc va NLPwin kabi maxsus vositalardan foydalaniladi. Bu vositalar leksik yondoshuvlar, frazalar, simbollar va qaysidir to'plamdagi uslubiy elementlarni birdaniga ko'rish imkonini beradi.

Ushbu usullardan foydalanish orqali badiiy matnlardagi uslubiy elementlarni avtomatik aniqlash jarayonidagi tadqiqotlarning samaradorligi ortishi mumkin.[**4,5,6,7,8,9,10,11,12,13**]

Xulosa qilib aytganda, uslubiy elementlarni avtomatik aniqlashda zamonaviy NLP metodlari va algoritmlaridan foydalanish, badiiy matnlarning to'liq va aniq tahlilini amalga oshirishga yordam beradi. Badiiy asarlardagi uslubiy xususiyatlarni aniqlash uchun qo'llaniladigan asosiy usullar va texnologiyalar, jumladan neyron tarmoqlar, sintaktik va semantik tahlil, klasterlash usullari, semantik embeddinglar va

“IZLANISH SAMARALARI”

№9 mavzusidagi respublika an'anaviy ilmiy-nazariy anjumani

stil transferlari, tadqiqotning samaradorligini oshirish uchun muhim ahamiyatga ega. Uslubiy elementlarni avtomatik aniqlash jarayonida metodlar va algoritmlarni avtomatlashtirish, tadqiqotning samaradorligini va aniqligini oshirishga yordam beradi. Badiiy matnlar orasidagi uslubiy aloqalarni aniqlash, milliy va xalqaro adabiyotdagi o'zgarishlarni o'rganishga imkon beradi.

Foydalanilgan adabiyotlar ro'yxati:

I. Ilmiy - nazariy adabiyotlar:

1. "Korpus lingvistikasi" Abdurahmonova Nilufar. Toshkent- 2023
2. "Korpus lingvistikasi" Manzura Abjalova Toshkent -2022
3. "Parallel korpus xususiyati " Nazira Sobirova, Madina Akramajanova, Shohjahon Sobirov.

II. Internet saytlari:

1. Jurafsky, D., & Martin, J. H. (2020). Speech and Language Processing
2. Manning, C. D., & Schütze, H. (1999). Foundations of Statistical Natural Language Processing
3. Biber, D., Conrad, S., & Reppen, R. (1998). Corpus Linguistics: Investigating Language Structure and Use
4. Cambria, E., & White, B. (2014). Jumping NLP Curves: A Review of Natural Language Processing Research
5. Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing
6. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information Koehn, P. (2020). Neural Machine Translation AntConc (Software)
7. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space
8. Natural Language Toolkit (NLTK) va spaCy
9. Chinese-Uyghur sentence alignment January 2009 DOI:10.3115/1690339.1690350 Conference: the 2nd Workshop Project director: Prof. Sylviane Granger Contact: sylviane.granger@uclouvain.be
10. How to cite : Granger, S., Dupont, M., Meunier, F., Naets, H. & Paquot, M. (2020) The International Corpus of Learner English. Version 3. Louvain-la-Neuve: Presses universitaires de Louvain. <https://dial.uclouvain.be/pr/boreal/object/boreal:229877>
11. Trial interface : <https://corpora.uclouvain.be/cecl/icle/trial/>