

Oriental Journal of Philology**ORIENTAL JOURNAL OF PHILOLOGY**

journal homepage:

<http://www.supportscience.uz/index.php/ojp/about>**REQUIREMENTS FOR PARALLEL CORPORA USED IN TRAINING MACHINE TRANSLATION MODELS*****Gulshoda Shamsiyeva****Lecturer, Department of Humanities, Perfect University**E-mail: shamsiyeva@perfectuniversity.uz**Tashkent, Uzbekistan***ABOUT ARTICLE**

Key words: parallel corpus, modeling, computational linguistics, machine translation, multilingual corpus, semantic matching, intelligent translation systems, NLP.

Received: 11.06.26**Accepted:** 12.06.26**Published:** 13.06.26

Abstract: The rapid development of artificial intelligence, natural language processing, and multilingual machine translation technologies has significantly increased the importance of parallel corpora in translation modeling and translation education. Modern intelligent translation systems increasingly rely on semantically adapted multilingual corpora for contextual interpretation, semantic image learning, and adaptive translation generation. This study examines the linguistic, semantic, technological, and pedagogical requirements for parallel corpora in teaching translation modeling in a multilingual computing environment.

TARJIMA MODELINI O'QITISHDA PARALLEL KORPUSLARGA QO'YILADIGAN TALABLAR***Gulshoda Shamsiyeva****Perfect-university Gumanitar fanlar kafedrasi o'qituvchisi**e-mail: shamsiyeva@perfectuniversity.uz**Toshkent, O'zbekiston***MAQOLA HAQIDA**

Kalit so'zlar: parallel korpus, modellashtirish, kompyuter lingvistikasi, mashina tarjimasini, ko'p tilli korpus, semantik moslashtirish, aqlli tarjima tizimlari, NLP.

Annotatsiya: Sun'iy intellekt, tabiiy tilni qayta ishlash va ko'p tilli mashina tarjimasini texnologiyalarining jadal rivojlanishi tarjima modellashtirish va tarjima ta'limida parallel korpuslarning ahamiyatini sezilarli darajada oshirdi. Zamonaviy intellektual tarjima tizimlari kontekstual talqin, semantik tasvirni o'rganish va adaptiv tarjima yaratish uchun

semantik jihatdan moslashtirilgan ko'p tilli korpuslarga tobora ko'proq tayanmoqda. Ushbu tadqiqot ko'p tilli hisoblash muhitida tarjima modellashtirishni o'qitishda parallel korpuslarga qo'yiladigan lingvistik, semantik, texnologik va pedagogik talablarni o'rganadi.

ТРЕБОВАНИЯ К ПАРАЛЛЕЛЬНЫМ КОРПУСАМ ПРИ ОБУЧЕНИИ МОДЕЛЕЙ МАШИННОГО ПЕРЕВОДА

Гульшода Шамсиева

Преподаватель кафедры гуманитарных наук Университета Perfect

E-mail: shamsiyeva@perfectuniversity.uz

Ташкент, Узбекистан

О СТАТЬЕ

Ключевые слова: параллельный корпус, моделирование, компьютерная лингвистика, машинный перевод, многоязычный корпус, семантическое согласование, интеллектуальные системы перевода, обработка естественного языка (NLP).

Аннотация: Стремительное развитие переводческих технологий значительно повысило важность параллельных корпусов в моделировании перевода и обучении переводу. Современные интеллектуальные системы машинного перевода всё больше опираются на семантически согласованные многоязычные корпуса для контекстной интерпретации, изучения семантических представлений и создания адаптивного перевода. В данном исследовании рассматриваются лингвистические, семантические, технологические и педагогические требования, предъявляемые к параллельным корпусам, используемым для обучения моделей перевода в многоязычной вычислительной среде.

Kirish. Ko'p tilli aloqaning tobora ortib borayotgan raqamlashtirilishi tarjima tadqiqotlari va kompyuter lingvistikasiga zamonaviy yondashuvlarni tubdan o'zgartirdi. Sun'iy intellekt va tabiiy tillarni qayta ishlash texnologiyalaridagi so'nggi yutuqlar ko'p tilli korpuslarning aqli tarjima tizimlarida, tarjima ta'limida va semantik til modellashtirishda rolini sezilarli darajada kengaytirdi. Zamonaviy ko'p tilli muhitlarda parallel korpuslar endi faqat tarjima qilingan matnlar to'plami sifatida qaralmaydi, buning o'rniga ular tobora ko'proq kontekstual talqin, ko'p tilli vakillikni o'rganish va moslashuvchan tarjima yaratishni qo'llab-quvvatlovchi semantik infratuzilma sifatida ishlaydi.

Zamonaviy mashina tarjima tizimlari kontekstual semantik ishlov berishga qodir transformatorga asoslangan neyron arxitekturalarini o'qitish uchun keng hajmdagi ko'p tilli

parallel korpuslarga tayanadi. Philipp Kohenning soʻzlariga koʻra, koʻp tilli korpuslar neyron tarjima tizimlarining markaziy komponentlaridan birini tashkil qiladi, chunki moslashtirilgan matn maʼlumotlari kompyuter modellariga tillar orasidagi semantik, sintaktik va kontekstual mosliklarni aniqlash imkonini beradi [6, 45–67].

Katta til modellari va koʻp tilli transformator arxitekturalarining tarjima texnologiyalariga tobora integratsiyalashuvi semantik jihatdan tuzilgan korpuslarning ahamiyatini yanada oshirdi. mBART, koʻp tilli T5, GPT asosidagi arxitekturalar va No Language Left Behind (NLLB) modellari kabi zamonaviy tizimlar turli til muhitlarida kontekstual lingvistik munosabatlarni ifodalashga qodir keng koʻlamli koʻp tilli maʼlumotlar toʻplamlarini talab qiladi. Soʻnggi koʻp tilli oʻqitish yondashuvlari shuni koʻrsatadiki, semantik jihatdan moslashtirilgan korpuslar, ayniqsa, kam resursli tillar va morfologik jihatdan murakkab lingvistik tizimlar uchun tarjima sifatini sezilarli darajada yaxshilaydi [1].

Parallel korpuslar qoʻshimcha ravishda tarjima taʼlimi va tarjima modellashtirishda muhim rol oʻynaydi. Korpusga asoslangan tarjima oʻqitish talabalar va tadqiqotchilarga haqiqiy koʻp tilli muhitda kontekstual semantik mosliklarni, leksik oʻzgaruvchanlikni, diskurs darajasidagi tarjima strategiyalarini va pragmatik moslashuv mexanizmlarini tahlil qilish imkonini beradi. Bunday yondashuvlar lingvistik tahlilni hisoblash semantik ishlov berish bilan integratsiya qilish orqali tarjima modelini oʻqitishni yanada samaraliroq qilishga hissa qoʻshadi.

Shu bilan birga, koʻp tilli tarjima tizimlarining rivojlanishi korpus sifati, semantik moslashuv, kontekstual izchillik, annotatsiya aniqligi va koʻp tilli maʼlumotlardagi nomutanosibliklar bilan bogʻliq yangi lingvistik va texnologik muammolarni keltirib chiqardi. Bu muammolar, ayniqsa, boy morfologiya, moslashuvchan sintaktik tuzilmalar va raqamli lingvistik resurslarning yetarli emasligi bilan ajralib turadigan kam resursli tillari uchun muhimdir. Oʻzbek, qozoq, qoraqalpoq, turk va boshqa agglyutinativ tillar kabi tillar koʻpincha sezilarli leksik noaniqlik va kontekstual oʻzgaruvchanlikni namoyon etadi, bu esa koʻp tilli semantik modellashtirishni murakkablashtiradi. Natijada, tarjima modellashtirishning samaradorligi koʻp jihatdan oʻqitish va semantik tahlil qilish uchun ishlatiladigan parallel korpuslarning lingvistik va texnologik sifatiga bogʻliq.

Shu nuqtai nazardan, parallel korpuslar uchun asosiy talablarni aniqlash tarjimini modellashtirish va koʻp tilli tarjima taʼlimini takomillashtirish uchun juda muhim ahamiyat kasb etadi. Yuqori sifatli parallel korpuslar semantik ekvivalentlik, kontekstual adekvatlik, nutq izchilligi, sintaktik moslashuv, terminologik barqarorlik va koʻp tillilikni taʼminlashi kerak.

Ushbu tadqiqot tarjima modelini oʻqitishda parallel korpuslar uchun lingvistik, semantik va texnologik talablarni oʻrganishga qaratilgan. Koʻp tilli semantik moslashuv, kontekstual vakillikni oʻrganish, korpus annotatsiyasi sifati va transformatorga asoslangan tarjima texnologiyalariga

alohida e'tibor qaratilgan. Tadqiqot qo'shimcha ravishda Paratranslator platformasidan ko'p tili tarjima infratuzilmasi modeli sifatida foydalangan holda korpusga asoslangan aqlli tarjima tizimlarining amaliy qo'llanilishini o'rganadi.

Tadqiqotning nazariy ahamiyati tarjimashunoslik, korpus lingvistikasi, kompyuter lingvistikasi va sun'iy intellekt metodologiyalarining fanlararo integratsiyasida yotadi. Tadqiqotning amaliy ahamiyati ko'p tili tarjima ta'limini takomillashtirish va yuqori resursli va kam resursli til muhitlari uchun semantik jihatdan moslashuvchan aqlli tarjima tizimlarini ishlab chiqish bilan bog'liq.

Metodologiya

Tadqiqotning metodologik tuzilishi quyidagi bosqichlardan iborat:

1. Ko'p tili parallel korpuslarni to'plash va tahlil qilish
2. Kontekstual semantik moslashtirish tahlili
3. Lingvistik va semantik korpus talablarini o'rganish
4. Ko'p tili tarjima modelini o'qitish yondashuvlarini tahlil qilish
5. Transformatorga asoslangan ko'p tili arxitekturalarni o'rganish
6. Paratranslator platformasiga asoslangan amaliy tahlil.

Tadqiqotda neyron mashina tarjimasi va ko'p tili semantik modellashtirish tadqiqotlarida keng qo'llaniladigan xalqaro miqyosda tan olingan lingvistik resurslardan olingan ko'p tili parallel korpuslardan foydalaniladi. Tanlangan ko'p tili ma'lumotlar to'plamlariga OPUS, CCMatrix, WikiMatrix va No Language Left Behind (NLLB) korpuslari kiradi. Ushbu korpuslar ilmiy munozaralar, yangiliklar maqolalari, suhbat matnlari, texnik hujjatlar va o'quv materiallari kabi bir nechta matn sohalari ifodalovchi semantik jihatdan moslashtirilgan ko'p tili jumla juftliklarini o'z ichiga oladi. Ko'p tili ma'lumotlar to'plamlari quyidagi mezonlarga muvofiq tanlandi:

- semantik ekvivalentlik;
- kontekstual izchillik;
- ko'p tili xilma-xillik;
- munozara darajasidagi izchillik;
- izoh sifati;
- moslashish aniqligi.

J.Tiedemannning fikriga ko'ra, yuqori semantik moslanish sifatiga ega ko'p tili korpuslar kontekstual tarjima modellashtirish va ko'p tili semantik ifodalashni o'rganishni sezilarli darajada yaxshilaydi [7].

Semantik moslashtirish tadqiqotning markaziy metodologik tarkibiy qismlaridan birini tashkil qiladi. Tadqiqotda semantik jihatdan ekvivalent matn segmentlari ko'p tili korpuslarda

qanday ifodalanganligi va kontekstual moslashuvlar tarjima modellashtirish sifatiga qanday ta'sir qilishi tahlil qilinadi.

An'anaviy leksik moslashtirish yondashuvlaridan farqli o'laroq, ushbu tadqiqot kontekstga asoslangan semantik moslashtirish metodologiyasini qo'llaydi. Semantik moslashtirish quyidagilarga ko'ra baholanadi:

- leksik moslashuv;
- kontekstual ekvivalentlik;
- nutq izchilligi;
- pragmatik adekvatlik;
- sintaktik moslik;
- semantik izchillik.

Yaqinda o'tkazilgan ko'p tilli tarjima tadqiqotlari shuni ko'rsatadiki, kontekstual semantik moslashuv ko'p tilli neyron tarjima samaradorligini va semantik ifodani o'rganishni sezilarli darajada yaxshilaydi [4].

Tadqiqotda transformatorga asoslangan ko'p tilli neyron arxitekturalariga asoslangan zamonaviy tarjima modellashtirish yondashuvlari qo'llaniladi. Ushbu tadqiqotda tarjima modellashtirish ko'p tilli matn tuzilmalari o'rtasidagi semantik, kontekstual va lingvistik munosabatlarni hisoblash orqali ifodalash jarayoni sifatida tushuniladi.

Tadqiqotda ko'p tilli tarjima modellashtirish quyidagilar orqali tahlil qilinadi:

- ko'p tilli qo'shimchalar;
- kontekstual semantik ifodalar;
- transformatorga asoslangan diqqat mexanizmlari;
- nutqqa sezgir tarjima modellashtirish;
- ko'p tilli semantik o'rganish.

mBART, ko'p tilli T5, GPT asosidagi tizimlar va NLLB modellarini o'z ichiga olgan zamonaviy transformator arxitekturalari kontekstga asoslangan ko'p tilli tarjima tizimlarining namunalari sifatida nazariy jihatdan o'rganildi. Ashish Vasvani va hamkasblarining fikriga ko'ra, diqqat mexanizmlari neyron tarjima tizimlarida uzoq masofali bog'liqlik modellashtirish va ko'p tilli kontekstual ishlov berishni sezilarli darajada yaxshilaydi [9].

Tadqiqotning asosiy maqsadlaridan biri tarjima ta'limi va tarjima modellashtirishda samarali parallel korpuslar uchun zarur bo'lgan lingvistik va strukturaviy talablarni aniqlashdir.

Tadqiqot quyidagi korpus talablarini baholaydi:

- semantik adekvatlik;
- kontekstual moslik;
- diskurs darajasidagi moslashtirish;

- terminologik izchillik;
- ko‘p tilli vakillik;
- izoh sifati;
- metama'lumotlarning to'liqligi;
- sintaktik moslashtirish;
- morfologik izoh;
- semantik teglash.

Korpus izohlash jarayoniga alohida e'tibor qaratildi, chunki izohlangan korpuslar semantik talqin va kontekstual tarjima sifatini sezilarli darajada yaxshilaydi. Tadqiqotda quyidagilarning roli, xususan, tahlil qilinadi:

- nomlangan shaxsni aniqlash (NER);
- semantik teglash;
- morfologik tahlil;
- leksik-semantik tasniflash.

O'tkazilgan so'nggi tadqiqotlarga ko'ra ko'p tilli NLP tadqiqotlariga ko'ra, izohli ko'p tilli korpuslar kontekstual semantik modellashtirish va intellektual tarjima samaradorligini sezilarli darajada yaxshilaydi [5].

Kontekstual semantik tahlil tadqiqotning yana bir muhim metodologik tarkibiy qismini tashkil etadi. Tadqiqot ko'p tilli korpuslar turli til muhitlarida kontekstual ma'no va semantik o'zgaruvchanlikni qanday saqlab qolishini o'rganadi.

Tadqiqotning amaliy komponenti ko'p tilli korpusga asoslangan aqlli tarjima tizimining namunasi sifatida Paratranslator platformasiga asoslangan. Platforma kontekstual semantik qidiruv, ko'p tilli korpus moslashuvi, nomlangan obyektни aniqlash va yagona tarjima ekotizimi ichida semantik tarmoq vizualizatsiyasini birlashtiradi.

Amaliy tahlil quyidagilarni o'z ichiga oladi:

- kontekstual tarjima qidiruvi;
- ko'p tilli jummalarni moslashtirish;
- semantik o'xshashlik tahlili;
- leksik-semantik tarmoq vizualizatsiyasi;
- ko'p tilli korpus qidiruvi;
- nomlangan obyektни aniqlash tahlili.

Platforma qo'shimcha ravishda ko'p tilli matn muhitida kontekstual semantik mosliklarni tahlil qilish imkoniyatlarini taqdim etadi. Platformadan olingan amaliy misollar kontekstual adekvatlik, semantik izchillik va diskurs darajasidagi tarjima modellashtirishni o'rganish uchun ishlatilgan.

Platforma arxitekturasiga oid soʻnggi tadqiqotlar shuni koʻrsatadiki, koʻp tilli kontekstual qidiruv tizimlari kam resursli til muhitida semantik talqin va koʻp tilli tarjima yetarliligini sezilarli darajada yaxshilaydi [6].

Koʻp tilli korpuslar va zamonaviy neyron tarjima texnologiyalarining integratsiyasiga qaramay, bir qator metodologik cheklovlar saqlanib qolmoqda. Kam resursli tillarda koʻp tilli korpuslarning yetarli emasligi, cheklangan annotatsiya sifati va koʻp tilli maʼlumotlar toʻplamlari ichida semantik nomutanosiblik kabi muammolar yuzaga kelmoqda.

Bundan tashqari, leksik noaniqlik, nutq oʻzgaruvchanligi va madaniy jihatdan shartli semantik tuzilmalar tufayli kontekstual semantik talqin qilish qiyinligicha qolmoqda. Transformatorga asoslangan koʻp tilli tizimlar qoʻshimcha ravishda yuqori darajada ixtisoslashgan sohalarida semantik jihatdan notoʻgʻri yoki kontekstual jihatdan nomuvofiq natijalarni yaratishi mumkin.

Shunga qaramay, taklif qilingan metodologik asos koʻp tilli intellektual tarjima muhitida tarjima modellashtirishni oʻqitishda parallel korpuslar talablarini oʻrganish uchun keng qamrovli fanlararo yondashuvni taqdim etadi.

Muhokama va natijalar. Koʻp tilli parallel korpuslarni tahlil qilish shuni koʻrsatadiki, kontekstual semantik moslashuv tarjima modellashtirish sifatini va koʻp tilli semantik tasvirni oʻrganishni yaxshilashda hal qiluvchi rol oʻynaydi. Anʼanaviy leksik moslashuv yondashuvlaridan farqli oʻlaroq, kontekstga asoslangan moslashuv mexanizmlari aqlli tarjima tizimlariga koʻp tilli matn muhitida semantik izchillikni, nutqning izchillikligini va pragmatik ekvivalentlikni saqlash imkonini beradi.

Natijalar shuni koʻrsatadiki, semantik jihatdan moslashtirilgan koʻp tilli korpuslar kontekstual tarjimaning yetarliligini sezilarli darajada yaxshilaydi, ayniqsa morfologik jihatdan boy va kam resursli tillarda. Oʻzbek, turk va qozoq tillari kabi tillar koʻpincha kontekstga bogʻliq leksik oʻzgaruvchanlikni namoyish etadi, bu yerda semantik talqin nafaqat alohida leksik birliklarga, balki sintaktik tuzilmalar, nutq konteksti va pragmatik foydalanish maʼno qirralariga ham bogʻliq.

Tadqiqot shuni koʻrsatadiki, samarali tarjima modellashtirish uchun jumla darajasidagi moslashuvning oʻzi yetarli emas, chunki semantik moslashuv koʻpincha alohida leksik birliklardan tashqariga chiqadi. Diskursga sezgir semantik moslashuv kontekstual talqinni sezilarli darajada yaxshilaydi va koʻp tilli neyron tizimlarida tarjimaning noaniqligini kamaytiradi.

Oʻtkazilgan soʻnggi koʻp tilli NLP tadqiqotlari shunga oʻxshash tarzda kontekstual semantik moslashuv transformator arxitekturalariga tillararo semantik bogʻliqliklarni yanada samarali modellashtirish imkonini berish orqali koʻp tilli tasvirni oʻrganish va tarjima sifatini yaxshilaydi, deb taʼkidlaydi [4].

Shuningdek, yuqori sifatli parallel korpuslar intellektual tarjima modellashtirish va tarjima ta'limini samarali qo'llab-quvvatlash uchun bir nechta muhim lingvistik va semantik talablarga javob berishi kerak.

Birinchiidan, semantik ekvivalentlik ko'p tilli korpuslar uchun eng muhim talabni tashkil qiladi. Manba va maqsadli til segmentlari faqat to'g'ridan-to'g'ri leksik mosliklarga tayanish o'rniga kontekstual ma'no, nutq niyatini va pragmatik tuzilmani saqlab qolishi kerak.

Ikkinchiidan, kontekstual yetarlilik ko'p tilli tarjima samaradorligiga sezilarli darajada ta'sir qiladi. Kontekstual jihatdan izolyatsiya qilingan jumla juftliklarini o'z ichiga olgan korpuslar ko'pincha parchalangan semantik tasvirlarni va neyron tarjima tizimlarida yetarli bo'lmagan nutq izchilligini keltirib chiqaradi.

Uchinchiidan, terminologik izchillik ilmiy, texnik va ta'lim tarjima muhitlarida juda muhim bo'lib qolmoqda. Noto'g'ri ko'p tilli terminologiya semantik talqin va tarjima ishonchligiga salbiy ta'sir qiladi.

To'rtinchiidan, ko'p tilli vakillik ko'p tilli AI tizimlarida til nomutanosibligini kamaytirish uchun juda muhimdir. Yuqori resursli tillar ko'p tilli ma'lumotlar to'plamlarida ustunlik qilishda davom etmoqda, past resursli tillar esa ko'pincha yetarlicha vakillik qilinmayapti. Bu nomutanosiblik ko'p tilli vakillikni o'rganish va semantik uzatish mexanizmlariga salbiy ta'sir qiladi.

An'anaviy tarjima tizimlari ko'pincha leksik birliklarni nutq kontekstidan mustaqil ravishda talqin qiladi va shu bilan semantik jihatdan noaniq yoki pragmatik jihatdan noto'g'ri tarjimalarni yaratadi. Aksincha, kontekstga e'tiborli ko'p tilli tizimlar lingvistik birliklarni kengroq semantik muhitda qayta ishlaydi, bu esa aniqroq kontekstual talqin qilishga imkon beradi.

Kontekstual tarjima modellashtirish quyidagilar uchun ayniqsa muhimdir:

- polisemantik leksik birliklar;
- idiomatik iboralar;
- madaniy jihatdan shartli iboralar;
- nutqqa bog'liq semantik tuzilmalar;
- pragmatik jihatdan moslashuvchan tarjima.

Masalan, bir nechta semantik ma'nolarga ega leksik birliklar nutqqa sezgir semantik mosliklarni aniqlashga qodir kontekstual talqin mexanizmlarini talab qiladi. Kontekstga e'tiborli transformator arxitekturalari ko'p tilli matn tuzilmalarida kontekstual ma'no va nutqning uzluksizligini saqlashda sezilarli darajada yaxshilangan samaradorlikni namoyish etdi.

Angela Fan va hamkasblarining fikriga ko'ra, ko'p tilli kontekstual o'rganish ko'p tilli tarjima tizimlarida semantik izchillik va lingvistik qamrovni sezilarli darajada yaxshilaydi [2].

Diqqat mexanizmlariga asoslangan transformator arxitekturalari uzoq masofali semantik bog‘liqliklar va ko‘p tilli kontekstual munosabatlarni samarali ravishda modellashtiradi. Bunday tizimlar ko‘p tilli matnli ma’lumotlarni parallel ravishda qayta ishlaydi va bir nechta tillarda umumiy semantik fazolarni yaratadi.

Natijalar shuni ko‘rsatadiki, transformatorga asoslangan tizimlar quyidagilarni sezilarli darajada yaxshilaydi:

- semantik izchillik;
- kontekstual adekvatlik;
- ko‘p tilli semantik uzatish;
- nutq izchillik;
- tarjima ravonligi.

Paratranslator platformasiga asoslangan amaliy tahlil natijalari

Paratranslator platformasi orqali o‘tkazilgan amaliy tahlil ko‘p tilli korpusga asoslangan kontekstual tarjima texnologiyalarining tarjima modellashtirish va semantik talqinda samaradorligini ko‘rsatdi.

Platforma quyidagilarni birlashtiradi:

- ko‘p tilli korpusni qidirish;
- kontekstual semantik tahlil;
- nomlangan obyektни aniqlash;
- semantik o‘xshashlik tahlili;
- leksik-semantik tarmoq vizualizatsiyasi.

Platformada joriy etilgan kontekstual qidiruv mexanizmi ajratilgan leksik tarjimalar o‘rniga semantik jihatdan moslashtirilgan ko‘p tilli jumla juftliklarini olish orqali kontekstual yetarlilikning yaxshilanganligini ko‘rsatadi. Ushbu yondashuv leksik noaniqlikni sezilarli darajada kamaytirdi va nutq darajasidagi semantik talqinni yaxshiladi.

Tahlil shuningdek, ko‘p tilli korpus qidiruvi foydalanuvchilarga turli lingvistik muhitlarda kontekstual semantik o‘zgaruvchanlikni aniqlash imkonini berishini ko‘rsatadi. Bunday funkcionallik, ayniqsa, tarjima ta’limida muhimdir, chunki u talabalarga haqiqiy ko‘p tilli kontekstual foydalanish naqshlarini kuzatish imkonini beradi.

Yana bir muhim natija nomlangan obyektни aniqlash texnologiyalarining integratsiyasi bilan bog‘liqligidadir. Platforma quyidagi semantik obyektlarni muvaffaqiyatli aniqlaydi:

- shaxslar nomlari;
- tashkilotlar nomlari;
- geografik joylashuv nomlari.

NER integratsiyasi ko'p tilli muhitda semantik talqin va kontekstual tarjima aniqligini sezilarli darajada yaxshilaydi.

Bundan tashqari, platformada joriy etilgan semantik tarmoq vizualizatsiya mexanizmlari ko'p tilli korpus ma'lumotlariga asoslangan leksik-semantik munosabatlarni tahlil qilish imkonini beradi. Bunday grafik asosidagi semantik tasvirlar kontekstual leksik assotsiatsiyalar va ko'p tilli semantik tuzilmalarni yaxshiroq tushunishga yordam beradi.

Korpusga asoslangan tarjima o'rganish talabalarga quyidagilarga imkon beradi:

- haqiqiy ko'p tilli kontekstlarni tahlil qilish;
- semantik o'zgaruvchanlikni kuzatish;
- diskursga sezgir tarjima strategiyalarini aniqlash;
- kontekstual ekvivalentlikni o'rganish;
- pragmatik moslashuv mexanizmlarini o'rganish.

An'anaviy lug'atga asoslangan tarjima o'qitishdan farqli o'laroq, korpusga asoslangan tarjima modellashtirish kontekstual semantik mulohaza yuritish va ko'p tilli diskurs tahlilini qo'llab-quvvatlaydi.

Natijalar shuni ko'rsatadiki, ko'p tilli korpus texnologiyalari nafaqat mashina tarjimasini rivojlantirishga, balki tarjima pedagogikasi va ko'p tilli lingvistik ta'limni modernizatsiya qilishga ham hissa qo'shadi. Kontekstga asoslangan ko'p tilli korpuslar semantik talqin, kontekstual o'rganish va moslashuvchan ko'p tilli muloqot ko'nikmalarini qo'llab-quvvatlovchi intellektual ta'lim infratuzilmasi sifatida tobora ko'proq faoliyat yuritmoqda.

Umuman olganda, natijalar yuqori sifatli ko'p tilli parallel korpuslar intellektual tarjima modellashtirish, ko'p tilli semantik o'rganish va zamonaviy tarjima ta'limi tizimlarining eng muhim asoslaridan birini tashkil etishini tasdiqlaydi.

Ushbu tadqiqot natijalari shuni ko'rsatadiki, ko'p tilli parallel korpuslar zamonaviy intellektual tarjima tizimlarining eng muhim lingvistik va texnologik asoslaridan birini tashkil qiladi. Tadqiqot shuni tasdiqlaydiki, semantik jihatdan tuzilgan ko'p tilli korpuslar transformatorga asoslangan neyron muhitida kontekstual tarjima modellashtirish, semantik talqin va ko'p tilli tasviriy o'rganishni sezilarli darajada yaxshilaydi.

Tahlildan kelib chiqadigan asosiy xulosalardan biri parallel korpuslarning oddiy ikki tilli matn to'plamlaridan ko'p tilli sun'iy intellekt tizimlarini qo'llab-quvvatlovchi murakkab semantik infratuzilmalarga o'sib borayotgan transformatsiyasi bilan bog'liq. Zamonaviy tarjima texnologiyalari endi faqat alohida leksik moslik mexanizmlariga tayanmaydi, balki, buning o'rniga ular tobora ko'proq kontekstual semantik ishlov berish, nutq darajasidagi talqin va ko'p tilli semantik tasviriy o'rganishga bog'liqdir.

Natijalar qo‘shimcha ravishda semantik moslashtirish tarjimaning adekvatligi va kontekstual izchillikka ta‘sir qiluvchi hal qiluvchi omil ekanligini ko‘rsatadi. An‘naviy jumla darajasidagi moslashtirish usullari ko‘pincha nutqning uzluksizligi va pragmatik ma‘nosini saqlab qola olmaydi, ayniqsa morfologik jihatdan boy va kontekstga sezgir tillarda. Aksincha, kontekstga asoslangan semantik moslashtirish ko‘p tilli kontekstual o‘rganish va semantik izchillikni sezilarli darajada yaxshilaydi.

Bu masala o‘zbek, turk, qozoq va qoraqalpoq kabi agglyutinativ tillarda ayniqsa muhim ahamiyat kasb etadi. Bu tillar boy morfologik tuzilmalarga, moslashuvchan sintaktik tuzilishga va kontekstga bog‘liq semantik o‘zgarishlarga ega bo‘lib, shu bilan ko‘p tilli tarjima tizimlari uchun qo‘shimcha qiyinchiliklar tug‘diradi. Tadqiqot shuni ko‘rsatdiki, bunday tillarda kontekstual talqin ko‘pincha ajratilgan leksik ekvivalentlikka emas, balki diskurs muhitiga, morfologik belgilarga va pragmatik tuzilmalarga bog‘liq hisoblanadi..

Yana bir muhim masala zamonaviy tarjima texnologiyalaridagi ko‘p tilli nomutanosiblik bilan bog‘liq. Yuqori resursli tillar ko‘p tilli korpus va transformator o‘qitish muhitlarida ustunlik qilishda davom etmoqda, ko‘plab kam resursli tillar esa yetarlicha foydalanilmayapti. Bu nomutanosiblik ko‘p tilli semantik uzatish va kontekstual vakillikni o‘rganishga salbiy ta‘sir ko‘rsatadi.

Paratranslator platformasi orqali o‘tkazilgan amaliy tahlil aqlli tarjima tizimlarida kontekstual semantik qidiruvning muhimligini yana bir bor tasdiqlaydi. Platforma ko‘p tilli korpusga asoslangan kontekstual qidiruv leksik tarjima mexanizmlariga nisbatan semantik talqinni sezilarli darajada yaxshilaganini ko‘rsatadi. Statik lug‘at ekvivalentlarini ishlab chiqarish o‘rniga, tizim ko‘p tilli korpuslardan semantik jihatdan moslashtirilgan kontekstual misollarni oladi va shu bilan diskursga sezgir tarjima modellashtirishni qo‘llab-quvvatlaydi.

Platformaning yana bir muhim hissasi semantik tarmoq vizualizatsiyasi va kontekstual semantik tahlil bilan bog‘liq. Tadqiqot natijalari shuni ko‘rsatadiki, grafikaga asoslangan leksik-semantik tuzilmalar ko‘p tilli semantik munosabatlar va kontekstual leksik assotsiatsiyalarni samaraliroq tahlil qilish imkonini beradi. Bunday texnologiyalar korpusga asoslangan tarjima ta‘limi va ko‘p tilli semantik tadqiqotlarga sezilarli hissa qo‘shishi mumkin.

Shuningdek, tarjima modellashtirishda transformatorga asoslangan ko‘p tilli arxitekturalarning ahamiyati tobora ortib bormoqda. mBART, ko‘p tilli T5, GPTga asoslangan modellar va NLLB arxitekturalari kabi tizimlar diqqat mexanizmlari va ko‘p tilli qo‘shimchalar orqali ko‘p tilli kontekstual ishlov berish va semantik tasvirni o‘rganishni sezilarli darajada yaxshilaydi.

Xulosa. Tadqiqot shuni tasdiqlaydiki, parallel korpuslar shunchaki ikki tilli matn omborlari sifatida emas, balki kontekstga sezgir tarjima modellashtirish va moslashuvchan ko‘p tilli

o'rganishni qo'llab-quvvatlashga qodir semantik jihatdan tuzilgan ko'p tilli ekotizimlar sifatida qaralishi kerak. Samarali ko'p tilli korpuslar ko'p tilli matn muhitlarida semantik ekvivalentlik, kontekstual izchillik, nutq uzluksizligi va pragmatik adekvatlikni saqlaydi.

Tadqiqot shuningdek, annotatsiyalangan ko'p tilli korpuslar aqlli tarjima modellashtirish va tarjima ta'limini sezilarli darajada yaxshilaydi, deb ko'rsatadi. Semantik teglash, morfologik tahlil, sintaktik tahlil va nomlangan obyektни aniqlash kabi annotatsiya mexanizmlari ko'p tilli muhitda kuchli semantik tasvirlar va aniqroq kontekstual talqinga hissa qo'shadi.

Paratranslator platformasi orqali o'tkazilgan amaliy tahlil ko'p tilli tarjima modellashtirishda kontekstual semantik qidiruv texnologiyalarining samaradorligini tasdiqlaydi. Tadqiqotning pedagogik natijalari shuni ko'rsatadiki, korpusga asoslangan tarjima o'qitish ko'p tilli semantik tushunishni, kontekstual fikrlashni va diskursga sezgir tarjima kompetensiyasini sezilarli darajada yaxshilaydi. Tarjima ta'limi tobora ko'p tilli korpuslar, sun'iy intellekt texnologiyalari, semantik tahlil tizimlari va kontekstga sezgir tarjima muhitlarini integratsiyalashni talab qilmoqda.

Tadqiqotning nazariy ahamiyati filologiya, korpus lingvistikasi, tarjimashunoslik, hisoblash lingvistikasi va sun'iy intellekt metodologiyalarining fanlararo integratsiyasida yotadi. Tadqiqotning amaliy ahamiyati ko'p tilli tarjima ta'limini takomillashtirish va kam resursli va ko'p tilli lingvistik muhitlar uchun semantik jihatdan moslashuvchan aqlli tarjima tizimlarini ishlab chiqish bilan bog'liq.

Foydalanilgan adabiyotlar ro'yxati:

1. Costa-Jussa M. R., Cross J., Çelebi O., Elbayad M., Heafield K. et al. No Language Left Behind: Scaling Human-Centered Machine Translation. – Meta AI Research, 2022. – 40 p.
2. Fan A., Bhosale S., Schwenk H., Ma Z. et al. Beyond English-Centric Multilingual Machine Translation // Journal of Machine Learning Research. – 2021. – Vol. 22. – P. 1–48.
3. Koehn, P. Neural Machine Translation. – Cambridge: Cambridge University Press, 2020. – 393 p.
4. Liu Y., Gu J., Goyal N., Li X., Edunov S. et al. Multilingual Denoising Pre-training for Neural Machine Translation // Transactions of the Association for Computational Linguistics. – 2020. – Vol. 8. – P. 726–742.
5. Rei R., Stewart C., Farinha A. C., Lavie A. COMET: A Neural Framework for MT Evaluation // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). – 2020. – P. 2685–2702.
6. N.Abdurakhmonova, Shamsiyeva G. "Context-based multilingual translation technology: on the example of the paratranslator platform". IEEE_UBMK-2025 International Conference on Computer Science and Engineering, 1800-1804 pp, 2025

7. Tiedemann J. Parallel Data, Tools and Interfaces in OPUS // Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC). – 2012. – P. 2214–2218.

8. Tiedemann J., Thottingal S. OPUS-MT – Building Open Translation Services for the World // Proceedings of the 22nd Annual Conference of the European Association for Machine Translation (EAMT). – 2020. – P. 479–480.

9. Vaswani A., Shazeer N., Parmar N., Uszkoreit J. et al. Attention Is All You Need // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 5998–6008.

10. N. A. Zaynobiddin qizi and S. G. Asliddin qizi, “Theoretical Foundations of Corpus-based Uzbek-English Machine Translation,” 2024 IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE), Novosibirsk, Russian Federation, 2024, pp. 1650-1653, doi: 10.1109/PIERE62470.2024.10805010.