



II SHO‘BA. KORPUS LINGVISTIKASI

KO‘P TILLI KORPUSLAR ASOSIDA MASHINA TARJIMASI TIZIMLARINI TAKOMILLASHTIRISH BOSQICHLARI

Abduraxmonova Nilufar Zaynobiddin qizi,
Filologiya fanlari doktori, professor
n.abduraxmonova@nuu.uz
O‘zMU

Shamsiyeva Gulshoda Asliddin qizi,
Gumanitar fanlar kafedrası o‘qituvchisi
shamsiyeva@perfectuniversity.uz
Perfect-university

Annotatsiya. Sun‘iy intellekt va tabiiy tilni qayta ishlash texnologiyalarining jadal rivojlanishi mashina tarjimasini sohasini sezilarli darajada o‘zgartirdi. So‘nggi yillarda ko‘p tilli korpuslar neyron mashina tarjima tizimlarining sifatini, kontekstual aniqligini va semantik izchilligini oshirish uchun eng muhim resurslardan biriga aylandi. Ushbu maqolada ko‘p tilli korpuslarga asoslangan mashina tarjima tizimlarini takomillashtirishdagi asosiy bosqichlar va texnologik yondashuvlar tahlil qilinadi. Mazkur tadqiqot, xususan, transformatorga asoslangan neyron arxitekturalari, parallel korpuslarda alignment dasturlari, kontekstga asoslangan tarjima mexanizmlari va ko‘p tilli model integratsiyasiga qaratilgan.

Maqolada ko‘p tilli neyron mashina tarjimasida qo‘llaniladigan zamonaviy yondashuvlar, jumladan, transfer o‘rganish (transfer learning), teskari tarjima (back-translation), ko‘p tilli dastlabki o‘qitish (multilingual pretraining) va ma‘lumotlarni qayta tahlil qilish asosida boyitilgan tarjima (retrieval-augmented translation) texnologiyalari tahlil qilinadi. Asosiy e‘tibor kam resursli tillarga qaratilgan bo‘lib, bunda ko‘p tilli korpuslar lekisik noaniqlikni kamaytirish hamda tarjimaning semantik va kontekstual sifatini oshirishda muhim ahamiyat kasb etadi. Bundan tashqari, maqolada zamonaviy tarjima tizimlariga OPUS, CCMatrix, mC4 va No Language Left Behind (NLLB) korpuslari kabi yirik ko‘p tilli ma‘lumotlar to‘plamlarining integratsiyasi muhokama qilinadi.



Kalit soʻzlar: *koʻp tilli korpus, neyron mashina tarjimasi, transformator arxitekturasi, NLP, kam resursli tillar, til modellari, kontekstga asoslangan tarjima.*

Annotation. The rapid development of artificial intelligence and natural language processing technologies has significantly changed the field of machine translation. In recent years, multilingual corpora have become one of the most important resources for improving the quality, contextual accuracy, and semantic consistency of neural machine translation systems. This article analyzes the main stages and technological approaches in improving machine translation systems based on multilingual corpora. In particular, this research focuses on transformer-based neural architectures, alignment programs in parallel corpora, context-based translation mechanisms, and multilingual model integration.

The article analyzes modern approaches used in multilingual neural machine translation, including transfer learning, back-translation, multilingual pretraining, and retrieval-augmented translation technologies. The focus is on low-resource languages, where multilingual corpora play a key role in reducing lexical ambiguity and improving the semantic and contextual quality of translation. In addition, the paper discusses the integration of large multilingual datasets such as OPUS, CCMatrix, mC4, and No Language Left Behind (NLLB) corpora into modern translation systems.

Keywords: *multilingual corpus, neural machine translation, transformer architecture, NLP, low-resource languages, language models, context-based translation.*

Tabiiy tillarni qayta ishlash va kompyuter lingvistikasidagi yutuqlar soʻnggi oʻn yillikda mashina tarjimasining metodologik asoslarini tubdan oʻzgartirdi. Anʼanaviy qoidaga asoslangan va statistik yondashuvlar asta-sekin semantik va kontekstual aniqlikning yuqori darajalariga erishish uchun chuqur oʻrganish (deep



learning) arxitekturalari va ko'p tilli korpuslardan foydalanadigan neyron mashina tarjimasi (NMT) tizimlari bilan almashtirildi.

Zamonaviy mashina tarjimasi tizimlari endi to'g'ridan-to'g'ri tarjima yoki frazaga asoslangan statistik modellashtirish bilan cheklanib qolmayapti. So'nggi o'n yilliklarda mashina tarjimasi texnologiyalarining rivojlanishi bir qancha muhim metodologik o'zgarishlarni boshdan kechirdi. Dastlabki mashina tarjimasi tizimlari asosan qoidaga asoslangan yondashuvlarga tayangan bo'lib, bu yerda lingvistik qoidalar, grammatik tuzilmalar va qo'lda yaratilgan lug'atlar tarjima yaratishning asosini tashkil etgan bo'lib, bunday mashina tarjimasi tizimlari boshqariladigan lingvistik muhitda qisman muvaffaqiyatga erishgan bo'lsa-da, ular kengaytirish imkoniyati, kontekstual tushunish va semantik moslashuvchanlik jihatidan cheklangan edi. Martin Kayning ta'kidlashicha, dastlabki kompyuterlashtirilgan tarjima tizimlari qo'lda tuzilgan lingvistik qoidalarning qat'iy tuzilishi sababli noaniqliklarni, idiomatik ifodalarni hamda sintaktik variativlikni qayta ishlashda jiddiy qiyinchiliklarga duch kelgan [7:3-23].

Qoidaga asoslangan tizimlarning cheklovlari asta-sekin tarjima jarayonlariga ehtimollik modellashtirishni kiritgan statistik mashina tarjimasi (SMT) ning paydo bo'lishiga olib keldi. Statistik yondashuvlar tarjima ehtimolliklarini va manba va maqsadli tillar o'rtasidagi leksik moslashuvlarni baholash uchun ikki tilli parallel korpuslardan foydalangan. SMT bo'yicha asosiy tadqiqotlardan biri Braun va boshqalar tomonidan olib borilgan bo'lib, ular tarjima mosliklarini to'g'ridan-to'g'ri ikki tilli korpuslardan o'rganish imkoniyatiga ega bo'lgan moslashtirishga asoslangan ehtimollik tarjima modellarini taklif qilishadi [2:263-311]. Keyinchalik iboraga asoslangan statistik mashina tarjimasi (Phrase-based SMT) lokal kontekstual bog'liqliklarni va ibora darajasidagi semantikasi saqlab qolish imkoniyatining yaxshilangani sababli ustun yondashuvga aylanadi. Biroq, statistik tizimlar hanuzgacha fragmentatsiyalashgan kontekstual tasvirlash va uzoq masofali



bog'liqliklarni yetarli darajada modellashtira olmaslik muammolariga duch kelmoqda.

Statistik paradigmalardan neyron mashina tarjimasiga (NMT) o'tish mashina tarjimasini tadqiqotlari tarixidagi eng inqilobiy bosqichlardan birini belgilab berdi. Neyron modellari ko'p tilli ma'lumotlardan uzluksiz semantik tasvirlarni (semantic representations) o'rganishga qodir chuqur o'rganish (deep learning) arxitekturalarini joriy etdi. Takrorlanuvchi neyron tarmoqlariga (Recurrent Neural Networks, RNN) asoslangan enkoder-dekoder arxitekturalari iboraga asoslangan SMT tizimlariga nisbatan tarjimaning ravonligi hamda semantik izchilligini sezilarli darajada yaxshiladi. Bahdanau va boshqalar tomonidan taklif etilgan diqqat mexanizmi (attention mechanism) esa neyron modellarga tarjima generatsiyasi jarayonida e'tiborni tegishli manba tilining muhim segmentlariga qaratish imkonini yaratdi, bu esa kontekstual bog'liqliklarni modellashtirish samaradorligini oshirdi [1:1-15].

Ashish Vaswani va boshqalar tomonidan transformer arxitekturasining taklif etilishi neyron tarjima tizimlari rivojida muhim burilish nuqtasi bo'ldi. Transformer modellari takroriy ketma-ket ishlov berish (recurrent sequential processing) mexanizmlarini o'z-o'ziga e'tibor (self-attention) mexanizmlari bilan almashtirdi, bu parallel hisoblash va uzoq masofali lingvistik bog'liqliklarni yanada samarali modellashtirish imkonini yaratdi. Ushbu innovatsiya tarjima sifatini, o'qitish samaradorligi (training efficiency) ni va ko'p tilli muhitlarda tizimning kengayuvchanligi imkoniyatlarini sezilarli darajada yaxshiladi [17:5998-6008]. Keyinchalik transformer arxitekturasi BERT, mBART, T5, GPT va NLLB asosidagi modellarni o'z ichiga olgan aksariyat zamonaviy ko'p tilli tarjima tizimlari uchun asos bo'ldi.

Ko'p tilli korpuslarning tobora keng miqyosda shakllanishi va foydalanishga joriy etilishi zamonaviy tarjima tizimlari rivojida muhim ahamiyat kasb etadi. Ko'p tilli korpuslar bir nechta tillarda semantik jihatdan ekvivalent mazmunni o'z ichiga



olgan moslashtirilgan matnli resurslardan tashkil topadi. Bunday ma'lumotlar to'plami neyron modellarga tillararo semantik tasvirlar (semantic representations) va umumiy lingvistik strukturalarni o'rganish imkonini beradi. Philipp Koehn ko'p tilli korpuslar nafaqat o'qitish ma'lumotlari (training data) sifatida, balki neyron tizimlariga turli tillar orasidagi sintaktik va kontekstual bog'liqliklarni aniqlash imkonini beruvchi semantik infratuzilma sifatida ham xizmat qilishini ta'kidlaydi [8:71-95].

So'nggi tadqiqotlar tobora ko'proq ko'p tilli neyron mashina tarjimasiga (MNMT) qaratilgan bo'lib, bu yerda bitta model bir vaqtning o'zida bir nechta til juftliklarida o'qitiladi. Jonson va boshqalar ko'p tilli neyron tizimlari zero-shot (nol zarbali) tarjimani amalga oshirishi mumkinligini ko'rsatdilar, bu esa modellarga o'qitish paytida hech qachon aniq uchramagan til juftliklari o'rtasida tarjima qilish imkonini beradi [6:339-351]. Mazkur yondashuv ko'p tilli tarjima tizimlarining amaliy qo'llanilish imkoniyatlarini, ayniqsa, kam resursli tillar bilan bog'liq holatlarda sezilarli darajada kengaytirdi. Natijada kam resursli tillar muammosi ko'p tilli mashina tarjimasida sohasidagi eng muhim tadqiqot yo'nalishlaridan biriga aylandi. Ko'plab tillarda raqamli matn resurslari cheklangan, yetarli emas. Buning oqibatida kam resursli tillar uchun mo'ljallangan tarjima tizimlarida semantik aniqlik va kontekstual izchillik darajasi pasayishi kuzatiladi. So'nggi ko'p tilli o'qitish yondashuvlari (transfer learning) umumiy ko'p tilli embeddinglar hamda parametrlarni ulashish mexanizmlari orqali yuqori resurli tillarda shakllangan lingvistik bilim bazasini kam resursli tillarga ko'chirish yordamida ushbu cheklovni bartaraf etishga urinmoqda. Angela Fan va boshqalarning fikriga ko'ra, ko'p tilli masshtablash strategiyalari sun'iy intellekt tizimlarida lingvistik qamrovni kengaytirish bilan birga, yetarli darajada reprezentatsiya qilinmagan tillar uchun tarjima sifatini sezilarli darajada yaxshilaydi [5:1-48].



Sohadagi yana bir muhim rivojlanish bu oldindan o'qitilgan ko'p tilli til modellarini mashina tarjima tizimlariga integratsiyalashdir. mBART, ko'p tilli T5, XLM-R va NLLB kabi modellar katta hajmdagi ko'p tilli korpuslarda o'qitilgan keng ko'lamli o'z-o'zini nazorat qiladigan (self-supervised) o'rganish texnikasidan foydalanadi. Ushbu arxitekturalar kontekstual semantikani avvalgi kodlovchi-dekoder tizimlariga qaraganda samaraliroq qamrab oladi. Liu va boshqalarning ta'kidlashicha, ko'p tilli oldindan o'rgatilgan modellar tillar bo'ylab umumiy bo'lgan universal semantik tasvirlarni o'rganish orqali tillararo umumlashtirishni yaxshilaydi [10:726-742]. Shu bilan birga, katta til modellarining (LLM) tez rivojlanishi mashina tarjimasida tadqiqotlari uchun yangi imkoniyatlarni yaratdi. LLM asosidagi tarjima tizimlari kontekstual fikrlash, semantik moslashuv va diskurs darajasida tushunishda ajoyib imkoniyatlarni namoyish etadi. An'anaviy NMT tizimlaridan farqli o'laroq, katta til modellari tarjima yaratish jarayonida kengroq kontekstual ma'lumot, pragmatik talqin va sohaga xos moslashuvni birlashtirishi mumkin. OpenAI va boshqa tadqiqot tashkilotlari tomonidan olib borilgan so'nggi tadqiqotlar shuni ko'rsatadiki, LLM asosidagi ko'p tilli tizimlar murakkab kontekstual tarjima vazifalari va past chastotali lingvistik hodisalarni hal qilishda avvalgi neyron arxitekturalaridan sezilarli darajada ustun turadi. Arxitekturaviy yaxshilanishlardan tashqari, so'nggi adabiyotlarda ko'p tilli mashina tarjimasida ma'lumotlar sifati va korpusni oldindan qayta ishlashning muhimligi ta'kidlangan. Korpusni filtrlash, jumalarni tekislash, tokenizatsiya, terminologiyani normallashtirish va shovqinni kamaytirish (noise reduction) tarjima samaradorligiga bevosita ta'sir qiladi. Tiedemann va Thottingalning ta'kidlashicha, OPUS kabi keng ko'lamli ochiq ko'p tilli ma'lumotlar to'plamlari zamonaviy mashina tarjimasida tadqiqotlari uchun muhim resurslarga aylangan, chunki ular yuzlab tillar va sohalarni qamrab oluvchi moslashtirilgan korpuslarni taqdim etadi [16:479-480].



Bundan tashqari, so‘nggi yillarda qidiruv-kengaytirilgan generatsiya (retrieval-augmented generation, RAG) va kontekstga asoslangan tarjima tizimlari istiqbolli tadqiqot yo‘nalishlari sifatida rivojlanmoqda. Qayta tiklash-kengaytirilgan tarjima tashqi bilimlarni qidirish mexanizmlarini faktik izchillik va terminologiyani saqlashni yaxshilash uchun neyron generatsiya modellari bilan birlashtiradi. Bunday yondashuvlar, ayniqsa, tibbiyot, huquq, ilmiy adabiyotlar va texnik hujjatlar kabi ixtisoslashgan sohalarda foydalidir. Lewis va boshqalarning fikriga ko‘ra, qayta tiklash bilan takomillashtirilgan neyron tizimlari dinamik ravishda olingan kontekstual ma’lumotlarni o‘z ichiga olish qobiliyati tufayli bilimga asoslangan tabiiy til yaratish vazifalarida yuqori natijalarga erishadi [9:9459-9474].

Ko‘p tilli mashina tarjimasi tadqiqotlarida sezilarli yutuqlarga qaramay, bir qator hal qilinmagan muammolar mavjud. Masalan, mavjud tizimlarda madaniy moslashuv, idiomatik ifodalar, nutqning uyg‘unligi, semantik gallyutsinatsiya va sohaga xos noaniqliklar uchraydi. Bundan tashqari, ko‘p tilli modellar ko‘pincha yuqori resursli va kam resursli tillar o‘rtasida ishlash nomutanosibligini namoyish etadi. Texnologik taraqqiyotlarning sezilarli darajada oshishiga qaramay, natijalar ko‘p tilli mashina tarjimasi tadqiqotlarida bir nechta muhim muammolar haligacha hal qilinmaganligini ko‘rsatadi.

Birinchidan, ko‘p tilli modellar yuqori resursli va kam resursli tillar samaradorligi o‘rtasidagi nomutanosiblikni ko‘rsatadi. Keng hajmdagi raqamli korpuslarga ega tillar ko‘p tilli reprezentatsiyalarni o‘rganishda ustunlik qilishda davom etmoqda, yetarli darajada reprezentatsiya qilinmagan tillar esa nisbatan zaifroq semantik modellashtirishga bilan cheklanib qolmoqda..

Ikkinchidan, semantik xatoliklar katta til modellari yordamida ishlovchi tarjima tizimlarida jiddiy muammolardan biri bo‘lib qolmoqda. Neyron modellar, ayniqsa ixtisoslashgan ilmiy va texnik sohalarda ba’zan ravon, ammo faktik jihatdan noto‘g‘ri tarjimalarni generatsiya qilishi mumkin.



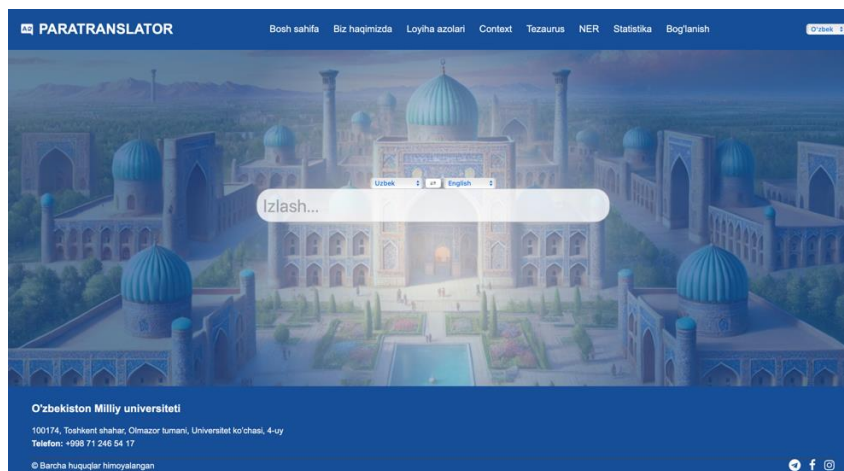
Uchinchidan, ko'p tilli transformator tizimlari lingvistik jihatdan uzoq til juftliklarida madaniy nuanslarni, idiomatik ifodalarni va pragmatik talqinni saqlashda qiyinchiliklarga duch kelmoqda.

Shunga qaramay, ushbu tadqiqot natijalari ko'p tilli korpuslar, transformatorga asoslangan neyron arxitekturalari va kontekstga asoslangan semantik modellashtirish mashina tarjimasini takomillashtirishning eng samarali zamonaviy yondashuvini ifodalashini aniq ko'rsatib turibdi. Ko'p tilli vakillikni o'rganish, qayta tiklash orqali kengaytirilgan tarjima va katta til modeli yordamida kontekstual takomillashtirishning integratsiyasi zamonaviy ko'p tilli tarjima texnologiyalarining semantik intellektini va amaliy qo'llanilishini sezilarli darajada oshiradi.

Ko'p tilli korpusga asoslangan tarjimaning amaliy qo'llanilishi: Paratranslator platformasi. Ko'p tilli korpusga asoslangan tarjima texnologiyalarining samaradorligini namoyish etuvchi amaliy misollaridan biri Paratranslator platformasidir. Platforma parallel korpuslar, semantik o'xshashlik tahlili va zamonaviy tabiiy tilni qayta ishlash texnologiyalari asosida ishlab chiqilgan kontekstga yo'naltirilgan ko'p tilli tarjima muhitini ifodalaydi. An'anaviy lug'atga asoslangan tarjima tizimlaridan farqli o'laroq, platforma ko'p tilli kontekstual qidiruv mexanizmlaridan foydalanadi, bu esa moslashtirilgan ko'p tilli korpuslardan olingan semantik jihatdan tegishli tarjima variantlarini yaratishga qodir. Platforma kam resursli va morfologik jihatdan boy tillarni qo'llab-quvvatlaydigan ko'p tilli lingvistik infratuzilma sifatida ishlab chiqilgan bo'lib, kontekstual semantik qidiruv, neyron tilni qayta ishlash, nomlangan obyektning aniqlash (Named Entity Recognition, NER) va leksik-semantik tarmoq tahlili texnologiyalarini integratsiya qiladi [11:1800-1804].

Platformaning texnik arxitekturasi zamonaviy veb-yo'naltirilgan NLP infratuzilmasiga, jumladan, Python Flask backend texnologiyalariga, ko'p tilli

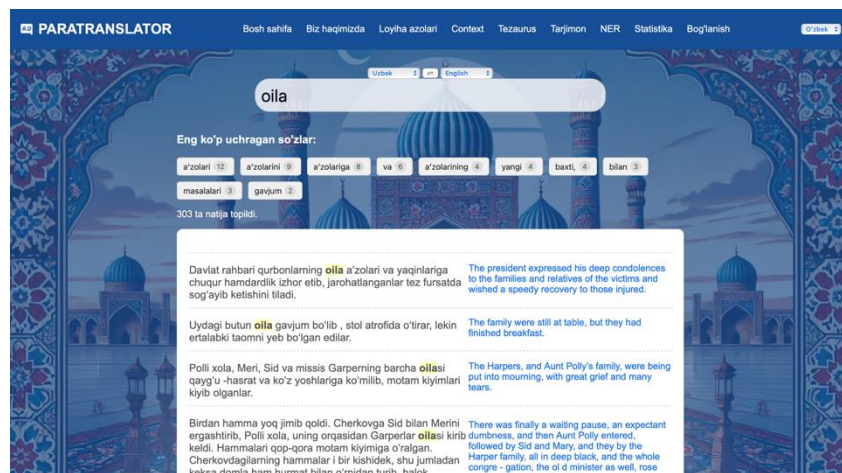
korpus ma'lumotlar bazalariga, HuggingFace transformer modellariga, word2vec semantik ifodalash algoritmlariga va kontekstual semantik moslashtirish uchun kosinus o'xshashlik mexanizmlariga asoslangan. Platforma hozirda o'zbek, ingliz, rus, turk, yapon, koreys, fransuz, ispan va qoraqalpoq tillarini o'z ichiga olgan bir nechta tillarni qo'llab-quvvatlaydi. Platformaga integratsiyalashgan ko'p tilli korpus turli xil lingvistik manbalardan to'plangan taxminan 960 000 ta moslashtirilgan matn segmentlarini o'z ichiga oladi. 1-rasmda Paratranslator platformasining asosiy interfeysi ko'rsatilgan. Tizim foydalanuvchilarga turli til juftliklari bo'yicha kontekstga asoslangan tarjima misollarini olish imkonini beruvchi ko'p tilli qidiruv funksiyasini taqdim etadi. Interfeys semantik tarjima tahlili va kontekstual korpusni o'rganish uchun mo'ljallangan foydalanuvchiga yo'naltirilgan ko'p tilli qidiruv muhitini namoyish etadi.



1-rasm. Paratranslator ko'p tilli tarjima platformasining asosiy interfeysi

Platformaning eng muhim texnologik xususiyatlaridan biri parallel korpuslarga asoslangan kontekstual tarjima qidiruvi mexanizmi bilan bog'liqdir. Tizim alohida leksik ekvivalentlarni generatsiya qilish o'rniga ko'p tilli korpuslardan semantik jihatdan moslangan jumla juftliklarini oladi. Ushbu yondashuv tarjima talqini paytida kontekstual yetarlilikni sezilarli darajada yaxshilaydi va leksik noaniqlikni kamaytiradi.

2-rasmda platformada “oila” leksik birligidan foydalangan holda amalga oshirilgan kontekstual ko‘p tilli qidiruv mexanizmi ko‘rsatilgan. Tizim moslashtirilgan ko‘p tilli jumla juftliklarini oladi va bir vaqtning o‘zida korpus muhitida eng ko‘p bog‘liq bo‘lgan leksik birliklarni aniqlaydi. Bunday funktsionallik foydalanuvchilarga semantik kollokatsiyalarni, kontekstual foydalanish va diskurs darajasidagi tarjima mosliklarini tahlil qilish imkonini beradi.



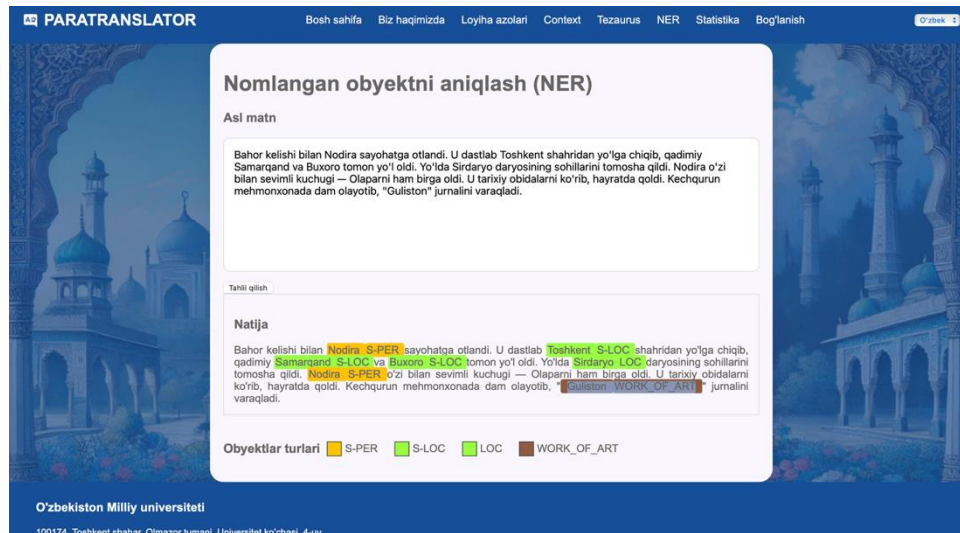
2-rasm. Kontekstga asoslangan ko‘p tilli korpus qidiruvi va moslashtirilgan jumla tarjimasi

Natijalar shuni ko‘rsatadiki, korpusga asoslangan kontekstual qidiruv ko‘p tilli tarjima vazifalarida semantik talqinni sezilarli darajada yaxshilaydi. Ushbu kuzatuv ko‘p tilli NLP muhitida kontekstga asoslangan neyron tarjima tizimlari va diskurs darajasidagi semantik modellashtirishning ahamiyatini ta’kidlaydigan zamonaviy tadqiqotlarga mos keladi [10:726-742].

Platformaning yana bir muhim komponenti nomlangan obyektни aniqlash (NER) texnologiyalarini integratsiyalashidir. Nomlangan obyektни aniqlash tabiiy tilni qayta ishlashning asosiy vazifalaridan biridir, chunki u matn muhitida atoqli otlar: shaxs ismlari, geografik joylashuvlar, tashkilotlar nomlari kabi semantik obyektlarni avtomatik ravishda aniqlash imkonini beradi.

3-rasmda Paratranslator tizimiga integratsiyalashgan NER moduli ko‘rsatilgan. Platforma o‘zbek tilidagi matn kiritishda shaxslar, joylashuvlar va tashkilot nomlarini o‘z ichiga olgan semantik obyektlarni avtomatik ravishda

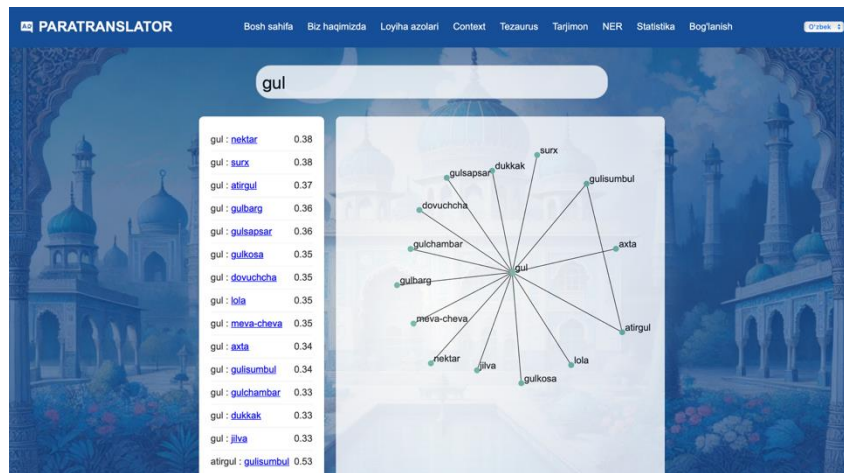
aniqlaydi va tasniflaydi. Bunday funktsionallik kontekstual tarjima aniqligiga sezilarli hissa qo'shadi, chunki nomlangan obyektlarni to'g'ri semantik aniqlash ko'p tilli tarjima yaratish jarayonida noaniqlikni kamaytiradi.



3-rasm. Paratranslator platformasida amalga oshirilgan Nomlangan obyektning aniqlash (NER) moduli

NER texnologiyalarini ko'p tilli tarjima tizimlariga integratsiyalashuvi aqlli mashina tarjimasi va semantik ma'lumotlarni olish tadqiqotlaridagi so'nggi yutuqlarga mos keladi. Devlin va boshqalarning fikriga ko'ra, kontekstual neyron tasvirlari ko'p tilli muhitda mavjudlikni aniqlash va semantik tasniflash vazifalarini sezilarli darajada yaxshilaydi [4:4141-4186].

Bundan tashqari, platforma semantik o'xshashlik tahlili va grafik tasvirlash texnologiyalariga asoslangan leksik-semantik vizualizatsiya mexanizmlarini o'z ichiga oladi. 4-rasmda “gul” leksik birligi uchun yaratilgan semantik tarmoq vizualizatsiya moduli ko'rsatilgan. Grafik asosida tasvirlash vektor o'xshashlik algoritmlaridan foydalangan holda ko'p tilli korpus ma'lumotlaridan olingan leksik birliklar o'rtasidagi semantik yaqinlik munosabatlarini ko'rsatadi.



4-rasm. Ko‘p tilli korpus tahlili orqali yaratilgan leksik-semantik tarmoq vizualizatsiyasi

Semantik tarmoq moduli korpus lingvistikasi tadqiqotlari, leksik semantika tahlili va ko‘p tilli kontseptual modellashtirish uchun muhim imkoniyatlar yaratadi. Semantik munosabatlarni vizualizatsiya qilish ko‘p tilli matn muhitida kontekstual leksik assotsiatsiyalar va semantik klasterlash chizmalarini yaxshiroq tushunishga yordam beradi. Bundan tashqari, so‘zlarni joylashtirish texnologiyalari orqali semantik o‘xshashlik tahlilini amalga oshirish ko‘p tilli korpus lingvistikasi va sun‘iy intellektga asoslangan til modellashtirish o‘rtasidagi o‘sib borayotgan yaqinlikni ko‘rsatadi. Bunday yondashuvlar tobora ko‘proq an’anaviy lug‘atga asoslangan leksik yozishmalardan tashqari kontekstual ma’noni aks ettirish imkoniyatiga ega bo‘lgan taqsimlangan semantik tasvirlash modellariga tayanadi.

Paratranslator platformasi asosida olingan amaliy natijalar ko‘p tilli mashina tarjimasini tizimlarini rivojlantirishga oid bir qator muhim xulosalarni tasdiqlaydi. Birinchidan, ko‘p tilli korpuslar kontekstual-semantik talqin va tarjima adekvatligini sezilarli darajada yaxshilaydi. Ikkinchidan, kontekstual qidiruv mexanizmlarining integratsiyasi diskurs darajasidagi semantik muvofiqlikni oshirish va leksik noaniqlikni kamaytirishga xizmat qiladi. Uchinchidan, NER texnologiyalari va semantik vizualizatsiya modullarini birlashtirish ko‘p tilli tarjima tizimlarining tahliliy imkoniyatlarini an’anaviy leksik tarjima funksiyalaridan ancha kengaytiradi.



Umuman olganda, Paratranslator platformasi ko‘p tilli korpuslar, kontekstual semantik qidiruv, nomlangan obyektни aniqlash va semantik tarmoq tahlilini yagona intellektual tarjima ekotizimiga integratsiya qilish orqali zamonaviy ko‘p tilli tarjima tadqiqotlariga muhim amaliy hissa qo‘shadi. Ushbu platforma ko‘p tilli korpusga asoslangan yondashuvlar kelajakda kontekstga asoslangan, semantik jihatdan moslashuvchan va lingvistik qamrovi keng mashina tarjiması texnologiyalarini rivojlantirishga qanday hissa qo‘shishi mumkinligini ko‘rsatadi.

Xulosa

Sun‘iy intellekt, tabiiy tillarni qayta ishlash va kompyuter lingvistikasining jadal rivojlanishi zamonaviy mashina tarjima tizimlarining metodologik va texnologik asoslarini tubdan o‘zgartirdi. Ushbu tadqiqotda ko‘p tilli korpuslar va zamonaviy neyron arxitekturalariga asoslangan mashina tarjima tizimlarini takomillashtirishning asosiy bosqichlarini o‘rganib chiqildi. Tadqiqot natijalari shuni ko‘rsatadiki, ko‘p tilli korpuslar, transformatorga asoslangan neyron modellari va kontekstga asoslangan semantik o‘rganish zamonaviy ko‘p tilli mashina tarjimasining asosiy texnologik asosini tashkil qiladi.

Yana bir muhim xulosa shuki, kontekstga asoslangan neyron tizimlari terminologiya izchilligini, nutqning uzluksizligini, semantik farqlashni va pragmatik talqinni saqlashda yaxshilangan qobiliyatni namoyish etadi. Qayta tiklash bilan kengaytirilgan tarjima strategiyalari va LLM yordamidagi tahrirdan keyingi yondashuvlar, ayniqsa ilmiy va texnik tarjima vazifalarida, faktik izchillik va semantik ravonlikni yanada oshiradi.

Shuningdek, kelgusidagi tadqiqotlar yetarli darajada reprezentatsiya qilinmagan tillar uchun ko‘p tilli korpuslarni kengaytirish, semantik xatoliklarni nazorat qilish mexanizmlarini takomillashtirish, retrieval-augmented reasoning frameworklarini integratsiya qilish hamda hisoblash samaradorligi yuqori bo‘lgan ko‘p tilli transformer arxitekturalarini ishlab chiqishga qaratilishi mumkin.

Foydalanilgan adabiyotlar ro‘xati

1. Bahdanau D., Cho K., Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate // Proceedings of the International Conference on Learning Representations (ICLR). – San Diego, 2015. – P. 1–15.
2. Brown P.F., Della Pietra S.A., Della Pietra V.J., Mercer R.L. The Mathematics of Statistical Machine Translation: Parameter Estimation // Computational Linguistics. – 1993. – Vol. 19, No. 2. – P. 263–311.
3. Costa-jussà M.R., Cross J., Çelebi O., Heafield K. et al. No Language Left Behind: Scaling Human-Centered Machine Translation. – Meta AI Research, 2022. – 40 p.
4. Devlin J., Chang M.W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of NAACL-HLT. – Minneapolis, 2019. – P. 4171–4186.
5. 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.
6. Johnson M., Schuster M., Le Q.V., Krikun M. et al. Google’s Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation // Transactions of the Association for Computational Linguistics. – 2017. – Vol. 5. – P. 339–351.
7. Kay M. The Proper Place of Men and Machines in Language Translation. – Stanford: CSLI Publications, 1997. – 23 p.
8. Koehn P. Neural Machine Translation. – Cambridge: Cambridge University Press, 2020. – 393 p.
9. Lewis P., Perez E., Piktus A., Petroni F. et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – P. 9459–9474.



10. Liu Y., Gu J., Goyal N., Li X. et al. Multilingual Denoising Pre-training for Neural Machine Translation // Transactions of the Association for Computational Linguistics. – 2020. – Vol. 8. – P. 726–742.
11. 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
12. 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.
13. N.Abdurakhmonova, I.Alisher and G.Toirova, “Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing,” 2022 7th International Conference on Computer Science and Engineering (UBMK), Diyarbakir, Turkey, 2022, pp. 73-75, doi: 10.1109/UBMK55850.2022.9919521.
14. N.Abdurakhmonova, A.Z.Mohirdev, M.Salokhiddinov, A.Narzullayev and A.Gatiatullin, “NLLB-Based Uzbek NMT: Leveraging Multisource Data,” 2024 9th International Conference on Computer Science and Engineering (UBMK), Antalya, Turkiye, 2024, pp. 1-5, doi: 10.1109/UBMK63289.2024.10773423.
15. N.Abdurakhmonova and N.Shamieva, “Creating an English-Uzbek Bilingual Thesaurus of Frequently Used Adjectives in Uzbek Corpus,” 2024 IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE), Novosibirsk, Russian Federation, 2024, pp. 1640-1644, doi: 10.1109/PIERE62470.2024.10804960.



16. 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). – Lisboa, 2020. – P. 479–480.

17. 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.