

UDK: 811.512.133`42

MASHINA TARJIMASINING SIFATINI OSHIRISHDA PARALLEL KORPUSNING AHAMIYATI

Hamroyeva Shahlo Mirdjonovna,
Filologiya fanlari doktori, professor v.b.
shaxlo.xamrayeva@navoiy-uni.uz
ToshDO`TAU

Xolmonova Iqbola Alisher qizi,
tayanch doktoranti
iqbolabintualisher@gmail.com
ToshDO`TAU

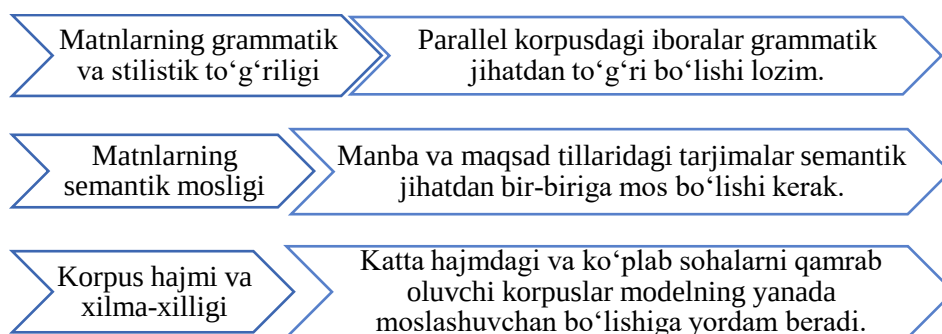
Annotatsiya: Ushbu maqolada mashina tarjimasi sifatiga ta'sir etuvchi asosiy omillar, parallel korpusning tarjima jarayonidagi roli va uning optimallashtirish usullari tahlil qilinadi. Maqolada parallel korpusni optimallashtirishning turli usullari, jumladan, ma'lumotlarni tozalash, balanslash va ko'paytirish jarayonlari tahlil qilinadi. Korpusning sifatini yaxshilash orqali tarjima tizimlarining samaradorligini oshirish mumkin. Shuningdek, tilning grammatik va struktural farqlarini hisobga olish, ma'lumotlarni optimallashtirish va ko'paytirishning ahamiyati ko'rsatilgan.

Kalit so'zlar: *Mashina tarjimasi, parallel korpus, optimallashtirish, ma'lumotlarni tozalash, ma'lumotlarni ko'paytirish, tarjima tizimi, til farqlari, korpus sifati, korpusni balanslash*

Zamonaviy mashina tarjimasi texnologiyalari til to'sig'ini bartaraf etish va turli tillar o'rtasida samarali kommunikatsiyani ta'minlashga xizmat qilmoqda. Mashina tarjimasining sifati ko'plab omillarga bog'liq bo'lib, ularning eng asosiylaridan biri o'quv ma'lumotlari sifatidir. Xususan, parallel korpuslar neyron mashina tarjimasi modellarining samaradorligini oshirishda muhim o'rin tutadi. Parallel korpusning sifati, hajmi va xilma-xilligi tarjima natijalarining aniq va tabiiy chiqishini ta'minlaydi. Sifatli parallel korpusning yetishmovchiligi esa mashina tarjimasining aniqligi va grammatik to'g'riligiga salbiy ta'sir ko'rsatishi mumkin. Shuning uchun parallel korpusni to'plash, tozalash va optimallashtirish jarayonlari mashina tarjimasi sifatini oshirishning muhim bosqichlaridan biri hisoblanadi.

Mashina tarjimasi sun'iy intellekt va tabiiy tilni qayta ishlash sohasidagi muhim yo'nalishlardan biri bo'lib, turli tillar o'rtasida avtomatik tarjima qilish imkonini beradi. Hozirgi kunda neyron mashina tarjimasi (NMT) modellarining rivojlanishi tufayli tarjima sifatini sezilarli darajada oshirishga erishilmoqda. Biroq, ushbu tizimlarning samaradorligi turli omillarga bog'liq bo'lib, ularning eng

muhimlaridan biri o'quv ma'lumotlari sifatidir. Parallel korpus – neyron mashina tarjimasida uchun asosiy ma'lumot manbai bo'lib, tarjimaning aniqligi, tabiiyligi va semantik mosligini belgilaydi. Mashina tarjimasida tizimlari turli bosqichlardan o'tib ishlab chiqiladi va natijalarning sifati ko'plab omillarga bog'liq. Masalan, parallel korpus sifati va hajmi. Parallel korpus mashina tarjimasida modelini o'rgatish uchun asosiy vosita hisoblanadi. Masalan, o'zbek-turk neyron mashina tarjimasida uchun tuzilayotgan parallel korpus yetarlicha hajmga ega bo'lishi va har ikkala til tizimiga mos kelishi lozim. Bunday korpusning sifati quyidagi jihatlarga bog'liq bo'ladi:

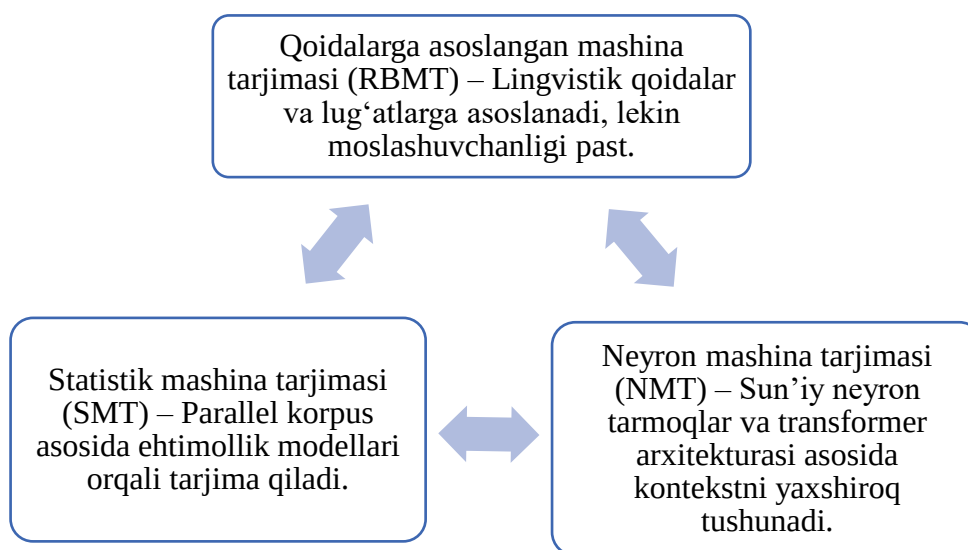


1-chizma. Sifatli korpusning asosiy jihatlari

Mashina tarjimasida tizimlarining samaradorligini oshirish uchun til juftligi o'rtasidagi grammatik tafovutlarni to'g'ri aniqlash va ularni hisobga olish juda muhimdir. Har bir til o'zining morfologik, sintaktik va semantik xususiyatlariga ega bo'lib, bu xususiyatlar ikki til o'rtasidagi tarjimani murakkablashtirishi yoki osonlashtirishi mumkin. Masalan, o'zbek va turk tillari o'rtasidagi so'z tartibidagi grammatik tafovutlar. O'zbek va turk tillari SOV (Subject-Object-Verb) tuzilmasiga asoslangan bo'lsa-da, so'z tartibida ba'zi farqlar mavjud. Misol uchun, o'zbek tilida ba'zi hollarda so'z tartibi sintaktik funktsiyalarni aniq ifodalash uchun o'zgartirilishi mumkin, ammo turk tilida bu nisbatan qat'iyroq. Turk tilida ham so'z tartibi asosan SOV, lekin ba'zi holatlarda SVO (Subject-Verb-Object) ham ishlatiladi, ayniqsa, urg'u qayerga qo'yilishi kerakligi bilan bog'liq bo'lgan kontekstlarda. Mashina tarjimasida tizimida bu farqlarni hisobga olish kerak, chunki so'zlar o'zaro joylashuvini noto'g'ri tarjima qilish tarjima sifatining pasayishiga olib kelishi mumkin. Bundan tashqari sinonimlar va idiomatik iboralarni ishlatishdagi tafovutlar ham mavjud. O'zbek va turk tillarida ba'zi sinonimlar va idiomatik ifodalar bir-biriga o'xshash bo'lishi mumkin, ammo ularning ishlatilishi va konteksti farq qiladi. Masalan: o'zbek tilida “Yurtimizning kelajagi” (The future of our country) turk tilida “Ülkemizin geleceği” (The future of our country) shaklida bo'ladi. Ammo, ba'zi hollarda bu so'zlarning kontekstual ishlatilishi bir-biridan farq qilishi mumkin, masalan, o'zbek tilidagi “yurakdan” (from the heart) turk tilida “kalpten” sifatida tarjima qilinadi, lekin ba'zi hollarda “gönülden” ham ishlatiladi. Mashina tarjimasida tizimining sifatini oshirish uchun, sinonimlar va idiomatik iboralarni to'g'ri

tushunish va tarjima qilish zarur. Bu soʻzlarning kontekstual maʼnosi va ular ishlatilgan muhitni hisobga olish muhim. Oʻzbek va turk tillari oʻrtasidagi grammatik tafovutlar mashina tarjimasi tizimining samaradorligini taʼminlashda katta ahamiyatga ega. Soʻz tartibi, morfologiya, feʼl tizimi va idiomatik iboralar kabi grammatik xususiyatlarni tushunish va toʻgʻri aks ettirish mashina tarjimasi tizimlarining sifatini oshiradi. Bunday tafovutlarni hisobga olgan holda, parallel korpuslar tuzish va mashina tarjimasi modellarini oʻrgatishda bu farqlarni inobatga olish zarur. Shuningdek, oʻzbek-turk til juftligi uchun oʻziga xos grammatik strategiyalar ishlab chiqilishi kerak, bu esa tarjima sifatining yaxshilanishiga olib keladi.

Mashina tarjimasi tizimlarining sifatini oshirishda ikki asosiy omil muhim rol oʻynaydi: parallel korpus sifati va model arxitekturasini. Ularning oʻzaro taʼsiri va toʻgʻri ishlashiga bogʻliq holda, tarjima tizimlarining aniqligi va samaradorligi sezilarli darajada yaxshilanadi. Shuni hisobga olgan holda mashina tarjimasi texnologiyalarining rivojlanishi natijasida turli yondashuvlar ishlab chiqilgan. Bular:



2-chizma. Mashina tarjimasi yondashuvlari

Hozirgi kunda NMT modellari, ayniqsa transformer arxitekturasini, eng samarali deb hisoblanadi[1]. Mashina tarjimasi tizimlarining samaradorligini oshirishda parallel korpuslardan toʻgʻri foydalanish juda muhim. Korpusdan maksimal samaradorlik olish uchun maʼlumotlarni oldindan qayta ishlash va optimallashtirish zarur. Ushbu jarayonlar mashina oʻrganish modellari uchun sifatli va optimal trening maʼlumotlarini taʼminlashga yordam beradi.

Parallel korpusdan samarali foydalanish uchun ma'lumotlarni oldindan qayta ishlash va optimallashtirish zarur. Oldindan qayta ishlashda matnni tokenizatsiya qilish, stop-so'zlarni olib tashlash, tozalash va lematizatsiya kabi jarayonlar amalga oshiriladi. Bu jarayonlar matnni modellar uchun to'g'ri formatda taqdim etishga yordam beradi. Optimallashtirish bosqichida, korpusning balanslanishi, ma'lumotlarni ko'paytirish va modelning giperparametrlari ustida ishlash orqali modelning samaradorligini oshirish mumkin.

Mashina tarjima tizimlarining samaradorligini aniqlash va ularning sifatini baholash sohada muhim vazifa hisoblanadi. Tarjima sifatini baholashda parallel korpuslar katta ahamiyatga ega, chunki ular turli tillar o'rtasidagi to'g'ri moslikni aniqlash va tahlil qilish imkonini beradi. Parallel korpusdan foydalanish, mashina tarjima tizimlarining samaradorligini o'lchashda aniq va obyektiv mezonlarni taqdim etadi. Tarjima sifatini baholashda bir qator avtomatik baholash mezonlari mavjud bo'lib, ulardan eng keng tarqalganlari BLEU, TER, METEOR, NIST va inson baholashidir. Har bir mezon tarjimaning turli jihatlarini – sintaktik moslik, semantik aniqlik, grammatik tuzilish va o'zaro aloqadorlikni – baholaydi. Parallel korpus orqali bu mezonlar yordamida mashina tarjima tizimlarining samaradorligini doimiy ravishda baholash va yaxshilash imkoniyati mavjud. Bu jarayon, o'z navbatida, til juftliklari orasidagi tafovutlarni hisobga olishga ham yordam beradi, bu esa tarjimaning aniqligi va sifatini oshiradi.

Parallel korpuslar mashina tarjima tizimlarining asosiy tayanchidir. Bu korpuslar bir nechta tillarda bir xil ma'lumotlarni o'z ichiga oladi va mashina tarjima modellari uchun o'rganish materialini taqdim etadi. Parallel korpusning hajmi va sifati mashina tarjima tizimlarining umumiy samaradorligi va sifatiga katta ta'sir ko'rsatadi. Parallel korpus hajmi mashina tarjima tizimining samaradorligini belgilovchi asosiy omillardan biridir. Korpus hajmi katta bo'lsa, model ko'proq ma'lumotni o'zlashtirib, o'z tarjima qobiliyatini yaxshilaydi. Katta hajmdagi parallel korpus mashina tarjima modelining generalizatsiya qobiliyatini oshiradi, chunki model turli kontekstlarda so'zlar va iboralarni qanday ishlatishni o'rganadi. Bir qator tadqiqotlar shuni ko'rsatadiki, parallel korpus hajmi ortgan sari, mashina tarjima tizimining umumiy sifatida yaxshilanish kuzatiladi. Masalan, Bengio va boshqalarning tadqiqotlarida katta hajmdagi parallel korpuslar yordamida o'rgangan neyron tarmoq modellari kichik hajmdagi korpuslarga nisbatan ancha yuqori sifatga ega ekanligi aniqlangan[2]. Korpus hajmi ortgan sari modelning tarjima xatolari kamayadi, ayniqsa, grammatik va semantik jihatlardan. Biroq, hajmning ortishi bilan modellarning o'rganish va hisoblash resurslari talablarining oshishi mumkin. Bu holatni optimallashtirish uchun ma'lumotlarni tozalash, filtrlash va boshqa texnikalardan foydalanish muhimdir.

Parallel korpusning sifati uning mazmuni va tuzilishi bilan bevosita bog'liq. Sifatli parallel korpuslar o'z ichiga aniq, to'liq va grammatik jihatdan to'g'ri

tarjimalar kiradi. Agar korpusda ko'p miqdordagi noto'g'ri yoki noaniq tarjimalar bo'lsa, mashina tarjimasi tizimi o'rganishda xato natijalar beradi. Shuning uchun korpusning sifati tarjima sifatining yuqori bo'lishi uchun hal qiluvchi omil hisoblanadi. Quyida korpusning sifatini baholashda asosiy faktorlarni keltirib o'tamiz:

Grammatik to'g'rilik:	Semantik aniqlik:	Kontekstual moslik:
Korpusdagi tarjimalarning grammatik jihatdan aniq va to'g'ri bo'lishi talab etiladi. Grammatik xatolar, so'zning noto'g'ri ishlatilishi yoki sintaktik xatolar mashina tarjimasi tizimining sifatiga salbiy ta'sir ko'rsatadi.	Tarjimada semantik ma'no xatolari bo'lmasligi kerak. Agar parallel korpusda semantik jihatdan noaniq yoki chalkash tarjimalar mavjud bo'lsa, mashina tarjimasi tizimi bu xatolarni o'rganib, noaniq natijalar beradi.	Tarjimalar kontekstga mos bo'lishi kerak. Yaxshi sifatli parallel korpus kontekstual farqlarni to'g'ri aks ettiradi, bu esa tarjima tizimining aniq va mantiqiy natijalar berishiga yordam beradi.

3-chizma. Korpusning sifatini baholashdagi asosiy faktorlar

Tadqiqotlar shuni ko'rsatadiki, parallel korpusda sifatli tarjimalar mavjud bo'lsa, mashina tarjimasi tizimining sifat ko'rsatkichlari sezilarli darajada yaxshilanadi. Koehn o'z tadqiqotida parallel korpusning sifati yuqori bo'lgan hollarda tarjima tizimlarining aniq va mantiqiy natijalar berishini ta'kidlagan[3]. Shuningdek, Liu va boshqalarning tadqiqotlari parallel korpusning sifati va mashina tarjimasi tizimi o'rtasidagi to'g'ri moslikni ko'rsatadi[4]. Bundan tashqari, parallel korpusning hajmi va sifati o'rtasidagi muvozanat ham muhimdir. Katta hajmdagi korpuslarning sifati yomon bo'lsa, mashina tarjimasi tizimi yomon natijalar beradi, chunki model xatoliklar bilan to'ldirilgan ma'lumotlardan o'rganadi. Aksincha, kichik hajmdagi yuqori sifatli korpuslar ham samarali bo'lishi mumkin, lekin bu holatda modelning umumiy o'rganish imkoniyatlari cheklangan bo'ladi. Yuqori sifatli va katta hajmdagi parallel korpuslar mashina tarjimasi tizimlarining samaradorligini maksimal darajada oshiradi. Sennrich va boshqalar o'z tadqiqotida katta hajmdagi yuqori sifatli parallel korpuslar bilan mashina tarjimasi tizimining aniq va samarali ishlashini ko'rsatgan[5].

Parallel korpusning hajmi va sifati mashina tarjimasi tizimlarining samaradorligi va sifatiga bevosita ta'sir ko'rsatadi. Korpus hajmi ortgan sari modelning o'rganish imkoniyatlari kengayadi, lekin uning sifati yaxshilanishi kerak. Katta hajmdagi, lekin sifatsiz korpuslardan foydalanish natijada noaniq va xato tarjimalarga olib kelishi mumkin. Shuning uchun, mashina tarjimasi tizimlarining

muvaffaqiyatli ishlashi uchun sifatli va katta hajmdagi parallel korpuslarni yaratish zarur.

Parallel korpusda noto'g'ri, chala, shovqinli (noisy) yoki takroriy (redundant) ma'lumotlar bo'lishi mumkin. Bu ma'lumotlar mashina tarjimasi tizimining sifatini pasaytiradi. Korpusni optimallashtirishning birinchi qadamlaridan biri – ma'lumotlarni tozalash va filtrlashdir. Ma'lumotlarni tozalash jarayonida, noto'g'ri tarjimalar, grammatik xatolar yoki noaniqliklarni bartaraf etish kerak. Ma'lumotlarni tozalash va filtrlashning bir necha asosiy usullari mavjud. Masalan, stop-so'zlarni olib tashlash. Parallel korpusda har xil tillar o'rtasidagi umumiy so'zlar va iboralarning tarjimasi ko'p bo'lishi mumkin. Bunday so'zlar ko'pincha tarjima sifatiga salbiy ta'sir ko'rsatadi. Stop-so'zlarni olib tashlash orqali modelni faqat muhim va kontekstual so'zlarga yo'naltirish mumkin. Bundan tashqari, grammatik tozalash ya'ni korpusdagi grammatik jihatdan xato yoki noto'g'ri tarjimalarni aniqlash va ularni olib tashlash mashina tarjimasi tizimining sifatini yaxshilaydi. Noto'g'ri tarjimalar modelga noto'g'ri ma'lumotlar kiritilishiga olib kelishi mumkin, bu esa xatoliklarni oshiradi. Shu bilan birga, parallel korpusda takroriy tarjimalar yoki o'xshash ma'lumotlar bo'lishi mumkin. Bu takroriy ma'lumotlar mashina tarjimasi tizimining o'rganish jarayonini zaiflashtiradi. Shuning uchun redundant ma'lumotlarni olib tashlash korpusni optimallashtirishga yordam beradi[6].

Parallel korpusda tillar o'rtasidagi so'zlarning chastotalari va iboralarning turlari o'rtasidagi farqlar ko'p uchraydi. Ba'zi tillarda ba'zi so'zlar yoki iboralar ko'proq ishlatiladi, boshqalari esa kamroq ishlatiladi. Korpusni balanslash usuli, tillar o'rtasidagi nisbatlarni muvozanatlash, tarjima tizimining umumiy ishlashini yaxshilaydi. Korpusni balanslash quyidagi usullarni o'z ichiga oladi:

Sof ma'lumotlar olish:

Parallel korpusda ba'zi tillar juda yuqori chastotali so'zlarga ega bo'lishi mumkin. Korpusdagi bu yuqori chastotali so'zlarni olib tashlash, ular bilan bo'lgan yuqori chastotali takrorlanishlarni kamaytirishga yordam beradi.

So'zlarning hamkorligini oshirish:

Korpusda biron-bir so'z boshqa tilga tarjima qilinganda kamdan-kam uchrasa, u holda bunday ma'lumotlarni kiritishni ko'paytirish mumkin. Korpusda so'zlar va iboralar o'rtasidagi hamkorlikni oshirish mashina tarjimasi tizimining xatoliklarini kamaytiradi.

O'xshash ma'lumotlarni qo'shish:

Ba'zi tillarda so'zlarning chastotalari juda past bo'lishi mumkin. Bunday hollarda, o'xshash ma'lumotlardan foydalanib, korpusni kengaytirish mumkin.

4-chizma. Korpusni balanslash usullari

Korpusni balanslash orqali mashina tarjimasi tizimi tilning barcha sohalarini o'rganib, ko'proq kontekstual ma'lumotlarni o'zlashtiradi[7].

Parallel korpusni optimallashtirishning yana bir muhim usuli – bu ma'lumotlarni ko'paytirishdir. Korpus hajmi kichik bo'lganda, mashina tarjimasi tizimi o'rganishda yetarli ma'lumotga ega bo'lmashligi mumkin. Ma'lumotlarni ko'paytirish yordamida bu muammoni hal qilish imkoniyati mavjud. Quyida ma'lumotlarni ko'paytirishning ba'zi usullarini ko'rib chiqamiz:

Parafrazlash (Paraphrasing):	Sintetik ma'lumotlar yaratish:	Back-translation:
Parafrazlash usulida, tarjima matnlarini qayta yozish orqali yangi tarjimalar hosil qilinadi. Bu usul yordamida yangi tarjimalar yaratiladi, bu esa modelga ko'proq o'rganish materialini taqdim etadi.	Sintetik tarjimalar yaratish orqali parallel korpusni sun'iy ravishda kengaytirish mumkin. Sintetik ma'lumotlarni yaratish mashina tarjimasi tizimining o'rganish imkoniyatlarini kengaytiradi.	Back-translation usulida birinchi tildan ikkinchi tilga tarjima qilingan matn qaytadan birinchi tilga tarjima qilinadi. Bu usul yordamida yangi tarjimalar hosil qilinadi va ularni modelga qo'shish orqali o'rganishni yanada mustahkamlash mumkin

5-chizma. Korpus ma'lumotlarini ko'paytirishning usullari

Bu usullar yordamida, parallel korpusning hajmi oshiriladi va mashina tarjimasini tizimining umumiy samaradorligi yaxshilanadi[5].

Parallel korpusda turli xil matnlar bo'lishi mumkin, masalan, ilmiy, texnik, adabiy yoki so'zlashuv uslubiga oid matnlar. Korpusni optimallashtirishda matnlarni kategoriyalash va tasniflash ham juda muhim rol o'ynaydi. Bu usul, ma'lumotlarni til turiga ko'ra ajratib, har bir turdagi matn uchun maxsus modellar yaratishga yordam beradi. Korpusni kategoriyalashning asosiy afzalliklariga matn turlari bo'yicha optimallashtirish kiradi. Ya'ni har bir matn turi uchun alohida model yaratish, mashina tarjimasini tizimining o'rganish jarayonini yanada optimallashtiradi. Masalan, ilmiy matnlar uchun maxsus model, adabiy matnlar uchun esa alohida model yaratilishi mumkin. Bundan tashqari, tizimni tezlashtirish mumkin. Bunda kategoriyalash yordamida, mashina tarjimasini tizimining o'rganish jarayoni tezlashadi, chunki model faqat o'sha turdagi matnlarni o'rganadi va noaniqliklar kamayadi [1]. Parallel korpusni optimallashtirishda o'quv algoritmlarini takomillashtirish ham muhimdir. Mashina tarjimasini tizimlarining samaradorligini oshirish uchun o'rganish algoritmlarini optimallashtirish zarur. Yangi arxitekturalar, masalan, neyral tarmoq va attention mechanism yordamida, model ko'proq kontekstual ma'lumotlarni o'zlashtirishi mumkin. Bu esa mashina tarjimasini tizimining to'g'ri ishlashiga yordam beradi[1]. Parallel korpusni optimallashtirish, mashina tarjimasini tizimlarining samaradorligini oshirish uchun juda muhim jarayon hisoblanadi. Korpusni tozalash, filtrlash, balanslash, ko'paytirish va kategoriyalash kabi usullar yordamida modelning sifatini yaxshilash mumkin. Shuningdek, o'quv algoritmlarini optimallashtirish va yangi texnologiyalarni qo'llash orqali mashina tarjimasini tizimlari yanada samarali ishlashi mumkin. Bu optimallashtirish usullari parallel korpusdan maksimal foydalanish imkoniyatini yaratadi va mashina tarjimasini tizimining sifatini sezilarli darajada oshiradi.

Mashina tarjimasini tizimlarining samaradorligini oshirishda parallel korpuslarning ahamiyati juda katta. Parallel korpus, ikki yoki undan ortiq tillarda tarjima qilingan matnlar to'plamidir, va bu korpuslar mashina tarjimasini tizimlarini o'rgatishda asosiy ma'lumot manbai hisoblanadi. Korpusning hajmi, sifati va tarkibi tarjima tizimlarining natijalariga bevosita ta'sir qiladi.

Parallel korpusni optimallashtirish orqali mashina tarjimasini tizimining sifatini sezilarli darajada oshirish mumkin. Korpusni tozalash, filtrlash va balanslash kabi usullar yordamida sifatni yaxshilashga erishish mumkin. Korpusdagi redundant yoki noaniq ma'lumotlarni olib tashlash, grammatik xatolarni tuzatish va o'xshash ma'lumotlarni kiritish orqali modelga aniq va ishonchli ma'lumotlar taqdim etiladi. Bundan tashqari, ma'lumotlarni ko'paytirish va back-translation kabi metodlar yordamida korpusning hajmi kengaytirilishi mumkin, bu esa o'rganish jarayonini kuchaytiradi. Korpusni kategoriyalash va turli matn turlariga mos ravishda modellarni yaratish, mashina tarjimasini tizimining muvaffaqiyatini oshirishda muhim

o'rin tutadi. Shuningdek, o'quv algoritmlarini optimallashtirish va neyron tarmoqlarni qo'llash orqali tarjima tizimlarining samaradorligi oshirilishi mumkin. Bundan tashqari, parallel korpusning lingvistik xususiyatlarini, grammatika, so'z tartibi va til strukturasidagi farqlarni hisobga olish, modelni yanada aniqroq va moslashtirilgan tarjimalar yaratishga yo'naltiradi. Korpusning xilma-xilligi va sifatli tarkibi mashina tarjimasi tizimining tillararo o'rganishni yaxshilaydi. Shu bilan birga, parallel korpusdan foydalanish nafaqat mashina tarjimasi tizimlarini rivojlantirishga, balki boshqa tilni o'rganish va lingvistik tadqiqotlar uchun ham katta imkoniyatlar yaratadi. Korpusning sifatini oshirish va optimallashtirish, kelajakda mashina tarjimasi tizimlarining yanada mukammal va ishonchli bo'lishiga xizmat qiladi.

Foydalanilgan adabiyotlar:

1. Vasvani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L., Polosukhin I. Attention Is All You Need. Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017–P. 5998–6008
2. Bengio Y., Ducharme R., Vincent P., Jauvin C. A Neural Probabilistic Language Model. Journal of Machine Learning Research, 3. 2003–P. 1137–1155
3. Koehn P. Europarl: A Parallel Corpus for Statistical Machine Translation. Proceedings of the 10th Machine Translation Summit. 2005–P. 79–86.
4. Liu Y. Automatic Evaluation of Machine Translation Quality: A Survey. 2016
5. Sennrich R., Haddow B., Birch A. Neural Machine Translation of Rare Words with Subword Units. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2016–P. 1715–1725.
6. Koehn P. Statistical Machine Translation. Cambridge University Press. 2016
- Tiedemann J. Parallel Data, Tools and Interfaces in OPUS. Proceedings of the 8th International Conference on Language Resources and Evaluation. 2012–P. 2214–2218