NEYRON MASHINA TARJIMI: TARIX, TAMOYILLAR VA BADIY MATN MASALASI

Ismatova Yulduz Nu'mon qizi
PhD, Samarqand davlat chet tillar instituti katta o'qituvchisi
https://orcid.org/0000-0001-8698-6250

Annotatsiya. Ushbu maqolada mashina tarjimasi tizimlarining tarixiy evolyutsiyasi — qoidaga asoslangan (rule-based), statistik (statistical) va neyron (neural) bosqichlar — tahlil qilinadi. Neyron mashina tarjimasi arxitekturasining asosiy tamoyillari (koder-dekoder modeli, transformer arxitekturasi, e'tibor mexanizmi) tarjimonlik nuqtai nazaridan soddalashtirilgan tarzda yoritiladi. Shuningdek, neyron tizimlarning kuchli va zaif tomonlari, xususan, badiiy matn tarjimadagi cheklovlari ingliz-o'zbek til juftligi misolida ko'rib chiqiladi. Tadqiqot shuni ko'rsatadiki, neyron tizimlar normativ va takroriy matnlarda yuqori natija bersa-da, badiiy matnning estetik va pragmatik qatlamlarini hali ham inson tarjimon arbitri zimtasida qoldiradi — bu posttahrirlash zaruratining obyektiv asosidir.

Kalit so'zlar: neyron mashina tarjimasi, transformer, sun'iy intellekt, badiiy matn, mashina tarjima tarix, ingliz-o'zbek til juftligi.

Аннотация. В статье анализируется историческая эволюция систем машинного перевода — правовые, статистические и нейронные этапы; рассматриваются архитектурные принципы нейронного машинного перевода и ограничения художественного текста на материале англо-узбекской языковой пары.

Ключевые слова: нейронный машинный перевод, трансформер, искусственный интеллект, художественный текст.

Abstract. The article examines the historical evolution of machine translation systems — rule-based, statistical and neural stages — and analyses the architectural principles of neural machine translation as well as the constraints of literary text, exemplified by the English–Uzbek language pair.

Keywords: neural machine translation, transformer, artificial intelligence, literary text.

Mashina tarjimasi (machine translation — MT) insoniyatning eng qadimiy sun'iy intellekt orzularidan biridir. 1950-yillarda chorak asrdan kam vaqt ichida mashina tarjimasi «ikki yilda yechiladi» degan optimistik bashoratlar qilingan edi; aslida esa bu masala o'n yilliklar davomida xalqaro tadqiqot dasturlarining markazida qoldi va har bir yangi bosqich avvalgisining yecholmagan muammolarini yangicha ko'rib chiqishga majbur qildi. Bugun neyron mashina tarjimasi (NMT) tizimlari kundalik hayotning ajralmas qismiga aylangan: Google Translate kuniga yuzlab milliard so'zni tarjima qiladi. Biroq tizimlar qanchalik takomillashgan bo'lishidan qat'i nazar, badiiy matn ularning eng zaif nuqtasidir. Ushbu maqolaning maqsadi — mashina tarjimasi tarixining uch asosiy paradigmasini yoritish, neyron tizimlarning ishlash tamoyillarini tarjimashunoslik nuqtai nazaridan tushuntirish hamda nega aynan badiiy matn masalasi inson arbitrijini saqlab qolishini asoslashdan iborat.

Mashina tarjimasi tarixi shartli ravishda uch katta paradigmaga bo'linadi. Birinchi paradigm — qoidaga asoslangan mashina tarjimasi (rule-based MT, RBMT). 1950–1980-yillarda hukm surgan ushbu yondashuvda tilshunoslar ikki til o'rtasidagi grammatik qoidalarni algoritmlarga aylantirib, lug'at va morfologik analiz modullarini qo'lda tuzishdi. 1954-yilda Georgetown–IBM tajribasi (49 ruscha jumlaning inglizchaga tarjimasi) ommaviy e'tiroz uyg'otgan bo'lsa-da, natijalar ko'rgazmali darajada tanlab tayyorlangan edi. RBMTning asosiy kamchiligi — qoidalarni yaratish va ularni

Mualliflik huquqi egasi tomonidan ushbu material Creative Commons Attribution 4.0 (CC BY 4.0) litsenziyasi asosida tarqatiladi.

yangilab turishning o'ta mehnat talab qilishi, til realiyasidagi istisnolarni qamrab ololmasligi edi. Shu bilan birga, mana shu davrda mashina tarjimasi nazariyasining poydevori qurildi: dastlabki matematik modellar, interlingva g'oyasi, transfer yondashuvi shu davrga oid (Po'latov, 2011).

Ikkinchi paradigm — statistik mashina tarjimasi (statistical MT, SMT). 1990-yillarda, ayniqsa 2006-yilda Google Translate statistik yondashuvga o'tgach, MT amaliyotida tub burilish yasaldi. SMT qoidalarga tayanmaydi; u instead katta hajmdagi parallel korpuslar asosida, hisoblash statistikasi yordamida, bir tildagi so'zlar ketma-ketligining ikkinchi tildagi eng ehtimoliy mos keluvchi ketma-ketligini topadi. Model «tarjima o'zini o'zi o'rgatadi»: odam tarjima qilgan millionlab jumla juftliklaridan ehtimolliklar statistikasi xulosalanadi. SMTning yutug'i — qo'lda qoida yozish zaruriyatining yo'qolishi; kamchiligi — lokal (so'z va ibora darajasidagi) ehtimollarga tayanishi natijasida uzoq masofadagi bog'liqliklarni ushlay olmasligi, jumla bo'ylab «mukammal» so'z tanlovlari umumiy ma'noni buzishi mumkinligi. Yirik SMT tizimlarida necha minglab iboralar ustuni (phrase table) va tillararo tilshunoslik bo'yicha murakkab algoritmlar qo'llanardi (Abduraxmonova, 2020).

Uchinchi paradigm — neyron mashina tarjimasi (neural MT, NMT). 2014-yildan boshlab rekurrent neyron tarmoqlar asosidagi koder–dekoder modellari, 2017-yildan esa transformer arxitekturasi MT sifatida keskin sakrashga sabab bo'ldi. NMT end-to-end o'qitiladi: parallel korpus beriladi va tarmoq o'zi tarjima qonuniyatlarini vektor ko'rinishida «o'zlashtiradi». Bu bosqichda mashina tarjimasi birinchi marta odam tarjimasiga bevosita yaqinlashdi — grammatik jihatdan to'g'ri, tabiiy o'qiladigan jumlalar paydo bo'ldi. Moorkens va hamkorlari (2018) tarjimonlarning badiiy posttahrirlashga oid tasavvurlarini o'rganishda aynan shu davr SMT va NMT tizimlarini qiyoslagan va NMT matnining kamroq og'ir tahrir talab qilinishini, lekin badiiy sifat bo'yicha ikkalasining ham yetarli emasligini ko'rsatdi. Toral, Wieling va Way (2018) butun bir roman ustida o'tkazgan tajribada NMT posttahrirlash mehnatining SMT dan pastligini miqdoriy tasdiqladi.

Neyron mashina tarjimasi chuqur o'rganish (deep learning)ning amaliy muvaffaqiyatlaridan biridir. Uning arxitekturasi tushunish uchun ikki asosiy tamoyilni bilish kifoya: koder–dekoder tuzilma va e'tibor (attention) mexanizmi.

Koder–dekoder modelida manba jumla avval koder nomli neyron tarmoq qismida ishlab chiqiladi: har bir so'z vektor (sonlar qatori) ko'rinishida ifodalanadi va jumla umumiy ma'no vektoriga siqiladi — bu jarayon xotira holati deb tushunilishi mumkin. So'ng dekoder ushbu holatdan foydalanib, birma-bir maqsad tilidagi so'zlarni ehtimoliy tarzda generatsiya qiladi. Har bir yangi so'z tarkibida oldingi so'zlar va umumiy kontekst hisobga olinadi. Bu model tarjimani «tiklash» emas, «generatsiya» jarayoni sifatida ko'radi — tizim qoida bilan emas, millionlab namunadagi statistik namoyandagi bilan ishlaydi (Qodirova, 2024).

Transformer arxitekturasi (2017-yildan) e'tibor mexanizmini olib kirish bilan NMT sifatini yana bir pog'ona oshirdi. E'tibor mexanizmi modelga tarjima qilishda manba jumlaniing qaysi qismlariga «qarab» ishlashini boshqarish imkonini beradi: masalan, ingliz tilidagi zamir so'z (pronoun) tarjimasi o'zbekchada otga bog'liq bo'lganda, model zamirni qaysi otga qaytarishini kontekst orqali aniqlaydi. Transformer parallel hisoblashga yaroqli bo'lgani uchun u korpuslarni ancha tez o'rgatadi va uzoq masofadagi bog'liqliklarni SMTdan ancha yaxshi ushlaydi.

Tarjimon uchun muhim xulosa: NMT modeli lingvistik qoidalar to'plami emas, balki statistik namoyandadir. U «grammatikni bilmaydi» — u qanday tarjima qilish mumkinligini namunalardan umumlashtiradi. Shu sababli: a) u o'quv korpusida tez uchraydigan konstruktsiyalarda juda ishonchli; b) noyob, badiiy, kontekstga chuqur bog'liq birliklarda xavfli; v) u xatolarni «bilmasdan» qiladi — tizim o'ziga ishonch hosil qiladi va noto'g'ri variantni ham ravon tarzda taqdim etadi. Aynan shu soxta

Mualliflik huquqi egasi tomonidan ushbu material Creative Commons Attribution 4.0 (CC BY 4.0) litsenziyasi asosida tarqatiladi.

ishonchlilik (fluent but wrong) badiiy matn posttahrirlashidagi eng xavfli omildir, chunki tahrirgacha o'qiyotgan inson e'tibori ravon matnda pasayadi (Koponen, 2016).

Badiiy matn o'z-o'ziga xos bir necha xususiyatga ega bo'lib, ularning har biri NMT uchun alohida qiyinchilik tug'diradi. Birinchidan, badiiy matn «ko'p qatlamli» matndir: unda syujet qatlami (voqea), psixologik qatlam (qahramon kechinmalari), til qatlami (muallif uslubi) bir vaqtning o'zida ishlaydi. Mashina modeli asosan syujet-ma'lumot qatlamiga tayanadi; psixologik nuans va uslubiy ohang model uchun ikkinchi darajali signal bo'lib qoladi.

Ikkinchidan, badiiy matn subtekstga boy: qahramonning so'zlari aytilgani bilan emas, aytilmagan ma'noda ham ishlaydi. Iroinity, kinoya, hakoya (understatement) kabi vositalar kontekstning keng ma'lumotlariga tayanadi; NMT esa jumlani segment-segment ko'rib tarjima qiladi va subtekstni ko'pincha yo'qotadi. Yuqorida ko'rilgan Moorkens va hamkorlari (2018) tadqiqoti shuni ko'rsatadiki, tarjimonlar badiiy matn posttahrirlashida eng ko'p e'tiroz bildirgan joy aynan ushbu subtekstning yo'qolishidir.

Uchinchidan, badiiy matn ovozli san'at asariga yaqin: ritm, ohang, takror, aliteratsiya. Tarjima bunday vositalarni maqsad tilida qayta yaratishi kerak — bu NMTdan talab qilinadigan generativ vazifaning eng murakkab turi. Model biror naqshni «eslaydi» (agar o'quv korpusida shunga o'xshash namuna uchrasa), lekin muallifning individual ovozini tiklay olmaydi. To'rtinchidan, badiiy matnda frazeologizm, realizm, madaniy-markerlar zich joylashadi — ularning har biri lokal statistikaga tushadi va model noyob variantni «topa olmaydi». Yusufjonova (2023) ko'rsatganidek, frazeologizmlarning milliy xususiyati badiiy uslubning ajralmas qismi bo'lib, bevosita ko'chirish komizm yoki notushunlik tug'diradi.

Ingliz-o'zbek til juftligi NMT uchun uch qatlamli qiyinchilik tug'diradi. Birinchidan — kam resursilik. NMT sifati parallel korpus hajmiga to'g'ridan-to'g'ri bog'liq: ingliz-fransuz juftligi uchun milliardlab so'z parallel matn mavjud bo'lsa, ingliz-o'zbek uchun bunday resurs ancha cheklangan. Olimjonov va Abidova (2023) ochiq parallel korpus bazalarini yaratish zaruratini aynan shu bo'shliqdan kelib chiqqan holda qo'yadi. Kam resurslilik modelning noyob leksik va grammatik hodisalardagi xatoligini keskin oshiradi (Yendirboyeva, 2025).

Ikkinchidan — tipologik masofa. Ingliz tili analitik, o'zbek tili agglutinativ: o'zbekchada bitta fe'l shakli (men bordim, siz bordingiz, u boribdi, u borar ekan) inglizchada bir nechta alohida yordamchi vositalar yoki kontekst talqiniga teng keladi. Model bu transformatsiyalarni statistik o'rganadi, ammo korpusda noyob bo'lgan shakllarni noto'g'ri «sodralashtiradi». Uchinchidan — o'zbek tili xalqaro tizimlarda ko'pincha umumiy turkiy model sifatida qayta ishlanadi, bu esa o'zbekchaga xos leksikaro (synonym) boylikni va sinonimik tanlov imkoniyatlarini kamaytiradi. Sotvoldiyeva (2024) sinonimlarni tarjima qilish turlarini tahlil qilib, aynan mazkur masalada inson tarjimonning mas'uliyat maydoni mashina tarjimasi qamrab ola olmaydiganini ko'rsatadi.

Amaliy kuzatish shuni ko'rsatadiki, Google Translate va DeepL tizimlari inglizcha badiiy matnni o'zbekchaga tarjima qilganda quyidagi qonuniyatlar kuzatiladi: dialog qahramonlarining ohangdorligi bir xillashtiriladi; frazeologik birliklar so'zma-so'z tarjima qilinadi (to let the cat out of the bag — mushukni xaltadan chiqarish); o'zbek morfologiyasining boy shakllari soddalashtiriladi (uch soat bo'di kutyapti o'rniga uch soatdan beri kutmoqda kabi shakl qo'llaniladi); realizmlar transkripsiya qilinadi yoki noto'g'ri moslashtiriladi. Isomiddinova (2024) umumlashtirganidek, mashina tarjimasi tezlik va hajmda ustundir, inson tarjimasi esa sifat va madaniy adekvatlikda — ikkalasi posttahrirlash orqali birlashtirilgandagina optimal natija hosil qiladi.

Mualliflik huquqi egasi tomonidan ushbu material Creative Commons Attribution 4.0 (CC BY 4.0) litsenziyasi asosida tarqatiladi.

Koder–dekoder mexanizmini tarjima jarayoniga o'xshatib tasavvur qilish mumkin: koder manba matnni «tushunib», uni zichlashgan xotira holatiga aylantiradi; dekoder esa shu holatdan maqsad tilida matnni qayta quradi. Amalda bu jarayon rekurrent neyron tarmoqlar (LSTM, GRU) yoki transformer blokalari orqali amalga oshiriladi. Rekurrent modellarda ma'lumot ketma-ketlik bo'ylab qayta ishlanadi — har bir so'z o'zidan oldingi so'zlarning xotirasiga tayanadi; bu texnik jihatdan samaradorlikni pasaytiradi, chunki uzoq jumlalarda dastlabki ma'lumotlar «unutiladi». Transformer arxitekturasi bu muammoni e'tibor mexanizmi orqali yechadi: model jumladagi istalgan so'z bilan istalgan boshqa so'z o'rtasidagi bog'liqlikni bevosita hisoblaydi, shu bois uzoq masofadagi zamirsoz bog'lanishlar, ko'chirma boshlanishlari va murakkab modifikator tuzilmalari ancha ishonchli tarjima qilinadi (Qodirova, 2024).

Tarjimon uchun arxitekturaning amaliy ahamiyati shundaki: neyron model «qaysi so'z qaysi so'zga mos keladi» degan savolni emas, «ushbu kontekstda keyingi so'z ehtimoli eng yuqori bo'lgan variant qaysi» degan savolni yechadi. Shuning uchun model ba'zan ma'nodan uzoq, lekin statistik jihatdan «tabiiy» variantni tanlaydi. Masalan, ingliz tilidagi common knowledge iborasi o'zbek badiiy matnda kundalik fikr tarzida qayta yaratilishi kerak bo'lgan joyda, model o'quv korpusidagi eng tez uchraydigan — ko'pincha texnik uslubdagi — variantni tanlashi mumkin. Bu hodisa NMT xatolarining tizimli xususiyatidir va uni faqat qoida bilan emas, kontekstni qayta o'qib chiqish orqali tuzatish mumkin (Ustinova, 2025).

Mashina tarjimasi tarixining uch paradigmasi — qoidaviy, statistik va neyron — har biri avvalgisining cheklovlarini yengishga intilgan, biroq hech biri badiiy matn muammasini to'liq yechmagan. Neyron tizimlar tabiiy, ravon matn generatsiya qila oladi, lekin ularning xatolari endi ko'rinmas qolgani uchun yanada xavfli: «oq oqadigan» noto'g'ri tarjima tahrirgacha o'qishni va e'tiborni susaytiradi. Badiiy matnning ko'p qatlamli subyekti, subtekstga boyligi, ovoziyohususi va madaniy realizmlari NMTning sof statistik tabiatiga zid keladi; ingliz-o'zbek juftligida esa kam resurslik va tipologik masofa masalani keskinlashtiradi. Demak, zamonaviy badiiy tarjima amaliyotida neyron tizim «tarjima qiluvchi» emas, «matn gipotezasi taklif qiluvchi vosita» sifatida qabul qilinishi va posttahrirlash jarayoni tizimli metodik asosda tashkil etilishi lozim. Kelgusi tadqiqotlar o'zbek tiliga moslashtirilgan neyron modellar, badiiy matn uchun predtahrirlash strategiyalari va posttahrirlash mehnat sarfining miqdoriy mezonlariga bag'ishlanishi mumkin.

ADABIYOTLAR

1. Abduraxmonova N. M. Mashina tarjimasing lingvistik ta'minoti. Toshkent: Lambert Academic Publishing, 2020.
2. Hamroyeva Sh., Matyakubova N. Mashina tarjimasida matnni moslashtirish usullari // Uzbekistan: Til va madaniyat. 2023. № 2. B. 6–19.
3. Isomiddinova L. Q. Mashina tarjimasi va inson tarjimasi: farqlar, afzalliklar va cheklovlar // Modern Educational System and Innovative Teaching Solutions: xalqaro ilmiy konferensiya materiallari. European Science International Conference, 2024.
4. Koponen M. Is machine translation post-editing worth the effort? A survey of research into post-editing and effort // The Journal of Specialised Translation. 2016. № 25. P. 131–148.
5. Moorkens J., Toral A., Castilho S., Way A. Translators' perceptions of literary post-editing using statistical and neural machine translation // Translation Spaces. 2018. Vol. 7, № 2. P. 240–262.
6. Olimjonov A. F., Abidova Sh. B. Mashina tarjimasi uchun ochiq korpusli parallel korpus bazalari yaratish // Educational Research in Universal Sciences. 2023. № 2 (2). B. 44–50.
7. Po'latov A. Kompyuter lingvistikasi. Toshkent: O'zbekiston Milliy universiteti nashriyoti, 2011.

Mualliflik huquqi egasi tomonidan ushbu material Creative Commons Attribution 4.0 (CC BY 4.0) litsenziyasi asosida tarqatiladi.

8. Qodirova M. Tarjima dasturlarining ishlash prinsipi // Journal of Effective Learning and Sustainable Innovation. Toshkent, 2024.
9. Sotvoldiyeva N. Sinonimlarni ifodalashda tarjima qilish turlari // O'zMU xabarlari. 2024. № 1.9.1. B. 344–346.
10. Toral A., Wieling M., Way A. Post-editing effort of a novel with statistical and neural machine translation // Frontiers in Digital Humanities. 2018. Vol. 5. Art. 9.
11. Yendirboyeva Z. Z. Mashina tarjimasi: til va texnologiya kesishmasidagi dolzarb masalalar // American Journal of Social Science. 2025. Vol. 3, № 9.
12. Yusufjonova Sh. M. Tarjimada frazeologik iboralarning milliy xususiyati va ularning badiiy asarlardagi tasviriy ifodasi // Current Approaches and New Research in Modern Sciences. 2023. № 2 (2). B. 112–116.

Mualliflik huquqi egasi tomonidan ushbu material Creative Commons Attribution 4.0 (CC BY 4.0) litsenziyasi asosida tarqatiladi.