

SUN'IY INTELEKT ASOSIDAGI TARJIMA TIZIMLARIDA SIFAT VA SEMANTIK ANIQLIK MUAMMOLARI VA YECHIMLAR

Madina Shixnazarovna ATANAZAROVA

Ma'mun universiteti

magistranti

madinaatanazarova@gmail.com

Annotatsiya: Ushbu maqola sun'iy intellekt asosidagi tarjima tizimlarida sifat va semantik aniqlik muammolarini tahlil qiladi, baholash mezonlari hamda yechimlarni taqqoslaydi. Bizning yondashuvimiz ma'lumotlar sifati, kontekstli modellashtirish va inson-aralashuvini birlashtirib, terminologik izchillik, iboraviy birliklar va hujjat darajasidagi uyg'unlikni kafolatlashga qaratiladi.

Kalit so'zlar: sun'iy intellekt, avtomatik tarjima, semantik aniqlik, baholash mezonlari, terminologiya, hujjat darajasi.

Sun'iy intellekt asosidagi tarjima tizimlari (AT) matnlarni tez va arzon tarjima qilish imkonini kengaytirdi, biroq "sifat" va "semantik aniqlik" masalalari hanuz dolzarb. Bunday sifat masalalarini faqat sirtqi ravonlik bilan o'lchash yetarli emas: tarjimada mazmuniy ekvivalentlik, diskursiv bog'liqlik va vazifa-yo'naltirilgan moslik (masalan, texnik hujjatdagi terminlar aniqligi) birgalikda baholanishi lozim. Bizning fikrimizcha, AT tizimlarini tilshunoslikdagi sifat paradigmalari bilan bog'lash, xususan, vazifaga yo'naltirilgan ekspertiza va ko'p o'lchovli tahlilni qo'llash, amaliy samaradorlikni oshirishning asosiy sharti hisoblanadi [1].

Semantik aniqlikning ildizi leksik-semantik tizimda yotadi: ko'pma'nolilik, sinonimiya, kollokatsion cheklovlar va semantik ro'llar AT uchun muammoli zonalar yaratadi. Masalan, bir so'zning variantlari nutq vaziyatiga ko'ra turlicha talqin qilinadi, iboraviy birikmalar esa kompozitsion bo'lmagan ma'noga ega. Bizning yondashuvimizda disambiguatsiya faqat statistik ehtimollarga emas, balki tizimli leksikografiya va integral ma'no tavsiflaridan foydalanishga tayanishi kerak, bunday manzara semantik toifalash va izchil sinonimik qatorlar yordamida tarjima qarorini barqarorlashtiradi [2].

AT evolyutsiyasi statistika asosidagi yondashuvlardan chuqur neyron modellariga o'tish orqali sezilarli sifat sakrashiga erishdi. Neyronli, xususan, transformerlarga asoslangan arxitekturalar uzoq masofali bog'lanishlarni modellashtirishi va moslashuvchan kontekstni ushlashi tufayli ravonlikni oshirdi. Biroq, bizning fikrimizcha, ravonlikning ko'tarilishi ba'zan mazmuniy siljish va noto'g'ri dalillar keltirish (hallucinatsiya) xavfi bilan birga keladi, shu bois domenga moslashtirish, terminologik cheklovlar va hujjat darajasida kontekstni kengaytirish zarur [3, 4].

Baholash mezonlari masalasida avtomatik ko'rsatkichlar, xususan, n-gramm o'xshashligiga tayanuvchi BLEU keng qo'llanadi, ammo u semantik ekvivalentlikni to'liq tutib bera olmaydi: sinonimik variantlar, iboraviy moslik va diskursiv bog'liqlik past darajada aks etadi. Bizning yondashuvimiz avtomatik ko'rsatkichlarni vazifa-yo'naltirilgan insoniy ekspertiza bilan birlashtirishni talab etadi: adekvatlik, terminologik to'g'rilik,



stilistik moslik va hujjat ichidagi referensial izchillik mezonlarini tizimli ravishda aniqlash lozim [5, 1].

Til juftliklarining o'ziga xosligi ham muhim. O'zbek tilida agglutinatív morfologiya, so'z tartibining nisbatan erkinligi va boy frazeologiya AT uchun sinovdir. Iboraviy birliklar va maqollarni so'zma-so'z tarjima qilish semantik xatolarga olib keladi, demak, frazeologik lug'atlar va iboralar bankini modelga singdirish zarur. Bizning yondashuvimiz iboralarning ma'no yadrosini, qo'llanish doirasini va funksional muqobillarini e'tiborga olgan holda, terminologiya bilan bir qatorda frazeologik resurslardan majburiy foydalanishni ko'zda tutadi [6, 2].

Yechimlar majmui bir necha qatlamdan iborat. Birinchidan, ma'lumotlar sifati: domenli, toza va tekshirilgan parallel korpuslar, shuningdek, sintetik parallel ma'lumotlarni ehtiyotkor balansda qo'shish modelning adekvatlikni ushlab qobiliyatini oshiradi. Ikkinchidan, hujjat darajasida tarjima: kontekst oynasini kengaytirish, ko'p jumla bog'liqlikni modellovchi mexanizmlar va referensial izchillikni kuzatish vositalari zarur. Uchinchidan, dekodlash cheklavlari: terminologiya glosariylari, nomlar va o'lchov birliklari uchun cheklangan lug'at, son-miqdoriy birliklar va formulalarni saqlovchi qoidalar muhim. To'rtinchidan, sifatsiz chiqishni erta aniqlovchi avtomatik sifat-baholash (QE) modullari va kalit segmentlar bo'yicha ikki bosqichli verifikatsiya "katta xatolar"ni kamaytiradi [3, 4].

Bizning yondashuvimizda tilshunoslikka asoslangan resurslar va neyron modellarning kuchli tomonlari integratsiyalanadi: (a) termin va iboralar uchun majburiy moslik qoidalari, (b) ixtisoslashgan leksik-semantik toifalar bilan sens disambiguatsiya, (c) domen moslashuvi uchun nozik sozlash va ko'p bosqichli o'qitish, (d) hujjat darajasida koherensiya chek-listi. Shu bilan birga, baholashda avtomatik ko'rsatkichlar faqat filtrlovchi rol o'ynaydi, yakuniy qaror ekspert tahriri, vazifaviy moslik va foydalanuvchi talablariga asoslanishi lozim [1, 5].

O'zbek tili uchun alohida amaliy chora-tadbirlar taklif etamiz. Birinchidan, so'z yasovchi affikslarning ko'pligi tufayli subtokenga ajratish strategiyasi semantik yaxlitlikni buzmasligi kerak, shu sababli morfologik xabardor segmentatsiya qulay. Ikkinchidan, iboraviy birliklar ro'yxatini domen kesimida boyitish, tarjimada ularni birlamchi sinf sifatida tutish zarur. Uchinchidan, nomlar, toponimlar va realiyalar uchun transkripsiya/transliteratsiya siyosati qat'iy boshqarilishi lozim. Bizning fikrimizcha, shu yo'l bilan ravonlik ortidan quvib, mazmuniy aniqlikdan voz kechmaslik mumkin.

Operatsion jarayonda "inson-aralashuvi"ni (human-in-the-loop) markazga qo'yish muhim: AT dastlabki loyihani yaratadi, so'ng redaktor terminlar va muhim segmentlarni tekshiradi, yakunda esa hujjat darajasida izchillik, uslub va murojaat shakllari bir xilligi ko'rib chiqiladi. Bu bosqichlarda xatolarni kategoriyalash va qayta o'qitish uchun relabelling usullari tizimning keyingi versiyalarini yaxshilaydi [1].

Xulosa sifatida aytish mumkinki, sifat va semantik aniqlik masalasi yagona texnik yechim bilan hal bo'lmaydi: ma'lumotlar muhandisligi, kontekstual modellashtirish, tilshunoslik resurslari va ekspert tahririning uyg'unligi talab qilinadi. Bizning yondashuvimiz ATni tezlik va arzonlik bilan cheklamasdan, ishonchlilik, javobgarlik va domen darajasida foydalilik mezonlariga moslashtirishni maqsad qiladi, shu tariqa 10.00.02 ixtisosligining nazariy va amaliy talablari bilan uyg'un natija olinadi.



FOYDALANILGAN ADABIYOTLAR RO'YXATI

1. House J. Translation Quality Assessment: Past and Present. London, New York: Routledge, 2015. 194 p.
2. Апресян Ю. Д. Избранные труды. Т. II. Интегральное описание языка и системная лексикография. Москва: Языки русской культуры, 1995. 782 с.
3. Koehn P. Statistical Machine Translation. Cambridge: Cambridge University Press, 2010. 446 p.
4. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention Is All You Need // Advances in Neural Information Processing Systems 30. Red.: I. Guyon va boshq. Red.: Curran Associates, Inc., 2017. P. 5998–6008.
5. Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: a Method for Automatic Evaluation of Machine Translation // Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia: Association for Computational Linguistics, 2002. P. 311–318.
6. Рахматуллаев Ш. Ўзбек тилининг изоҳли фразеологик луғати. Тошкент: Ўзбекистон миллий энциклопедияси, 2003. 592 б.