O‘ZBEK–INGLIZ PARALLEL KORPUSIDA TARJIMA TRANSFORMATSIYALARI VA MOSLASHTIRISH XATOLARINI FARQLASH

M.A.Amirkulov[1]

1. 2-kurs tayanch doktorant, Toshkent davlat o‘zbek tili va adabiyoti universiteti

Mas'ul muallif: Amirkulov Ma’rufjon Alikulovich, 2-kurs tayanch doktorant, Toshkent davlat o‘zbek tili va adabiyoti universiteti.

e-mail: amirkulov.edu01@gmail.com

Received: 26 Oktabr 2026
Accepted: 31 Oktabr 2026
Published: 05 Noyabr 2026

Copyright: © 2026 by the author(s). Licensee Buxoro xalqaro universiteti ilmiy axborotnomasi.

Iqtibos uchun: Amirkulov, M. A. O‘ZBEK–INGLIZ PARALLEL KORPUSIDA TARJIMA TRANSFORMATSIYALARI VA MOSLASHTIRISH XATOLARINI FARQLASH / M. A. Amirkulov // Buxoro xalqaro universiteti ilmiy axborotnomasi. — 2026. — T. 11, № 1. — B. 65–70.

Annotatsiya. Dolzarblik. Parallel korpusda bir-biriga yaqin ko‘ringan gaplar har doim ham mazmunan muvofiq bo‘lmaydi; ayni paytda grammatik tuzilishi farqli segmentlar bir xil axborotni yetkazishi mumkin. **Maqsad.** O‘zbek–ingliz gap juftliklarida tarjima transformatsiyasi, mazmuniy tafovut va noto‘g‘ri moslashtirishni farqlashning tekshiriladigan mezonlarini belgilash. **Metodlar.** UzEnParC loyihasiga oid ilmiy ishlar va xalqaro adabiyotlar qiyoslandi; O‘zbekiston Respublikasi Prezidentining rasmiy veb-saytidagi o‘zbekcha hamda inglizcha xabardan tanlangan ikki tabiiy gap juftligi predikat, ishtirokchilar, miqdor, zamon va harakat holati bo‘yicha tahlil qilindi. Qo‘shimcha ravishda muallif tuzgan nazorat misollaridan foydalanildi. **Natijalar.** Shakliy o‘zgarishni xato deb baholamasdan, moslashtirish holati, mazmuniy muvofiqlik va transformatsiya turini alohida qayd etuvchi annotatsiya tartibi ishlab chiqildi. Birinchi tabiiy juftlikda miqdor va o‘lchov birligi saqlangani, ikkinchisida esa tegishli gaplar bog‘lanishi saqlangan holda harakatning amalga oshish holati farqlanishi aniqlandi. **Xulosa.** Taklif etilgan mezonlar muharrir va annotator qarorlarini izchil asoslash uchun xizmat qiladi; ularning barqarorligi katta tanlanmada mustaqil sinovdan o‘tkazilishi lozim. Nazorat misollarini haqiqiy korpus dalillaridan ajratish hamda har bir tahririy qaror uchun qayta tekshiriladigan izoh yozish zarurati alohida asoslanadi.

Kalit so'zlar: parallel korpus, tarjima transformatsiyasi, gaplarni moslashtirish, mazmuniy muvofiqlik, predikat, modallik, annotatsiya, korpus sifati, kontekst.

РАЗГРАНИЧЕНИЕ ПЕРЕВОДЧЕСКИХ ТРАНСФОРМАЦИЙ И ОШИБОК ВЫРАВНИВАНИЯ В УЗБЕКСКО-АНГЛИЙСКОМ ПАРАЛЛЕЛЬНОМ КОРПУСЕ

Аннотация. Актуальность. В параллельном корпусе внешне сходные предложения не всегда передают одинаковое содержание, тогда как грамматически различающиеся сегменты могут выражать эквивалентную информацию. **Цель.** Разработать проверяемые критерии разграничения переводческих трансформаций, смысловых расхождений и ошибок выравнивания узбекско-английских предложений. **Методы.** Проведён сравнительный анализ публикаций проекта UzEnParC и международной научной литературы. Две аутентичные пары предложений из узбекской и английской версий сообщения официального сайта Президента Республики Узбекистан исследованы по предикату, участникам ситуации, количественным выражениям, времени и характеру протекания действия. Дополнительно использованы специально составленные контрольные примеры. **Результаты.** Предложена процедура аннотирования, при которой корректность выравнивания, смысловое соответствие и тип трансформации фиксируются независимо. В первой аутентичной паре сохранены количественное значение и единица измерения; во второй соответствующие сегменты связаны правильно, но различается представление о завершённости действия. Обсуждение. Такое разграничение позволяет не смешивать формальные перестройки перевода с изменением фактической информации и ошибочным выбором сегмента. **Выводы.** Процедура может способствовать обоснованности экспертной разметки, однако её воспроизводимость и согласованность оценок требуют последующей проверки на более широком материале.

Ключевые слова: параллельный корпус, переводческая трансформация, выравнивание предложений, смысловое соответствие, предикат, модальность, аннотирование, качество корпуса, контекст.

DISTINGUISHING TRANSLATION TRANSFORMATIONS FROM ALIGNMENT ERRORS IN UZBEK–ENGLISH PARALLEL CORPORA

Abstract. Background. In a parallel corpus, superficially similar sentences do not necessarily convey the same information, whereas structurally different segments may remain semantically equivalent. **Objective.** This study develops verifiable criteria for distinguishing translation transformations, semantic discrepancies, and sentence-alignment errors in Uzbek–English data. **Methods.** Studies associated with the UzEnParC project and international methodological literature were compared. Two authentic

sentence pairs taken from the Uzbek and English versions of an announcement on the official website of the President of Uzbekistan were examined with regard to predicates, participant roles, quantities, tense, and event status. Constructed contrastive examples served solely to illustrate individual criteria. Findings. A three-part annotation procedure is proposed to record alignment correctness, semantic correspondence, and transformation type independently. In the first authentic pair, the numerical value and unit of measurement are preserved despite structural differences. In the second, the corresponding sentences are properly linked, but the description of whether the action has been completed differs. **Conclusions.** These criteria support transparent and reviewable annotation decisions without treating all linguistic variation as noise. Their reliability and inter-annotator consistency remain to be evaluated on a larger and more varied sample.

Keywords: parallel corpus, translation transformation, sentence alignment, semantic correspondence, predicate, modality, annotation, corpus quality, context.

Kirish. Parallel korpusdagi gap juftligi ikki tildagi matnni qiyoslash uchun tayanch birlikdir. Uning to‘g‘ri tuzilishi lug‘aviy muqobillarni aniqlash, grammatik hodisalarni o‘rganish va tarjima materialini tanlashga bevosita ta’sir qiladi. Biroq matnlarni yonma-yon joylashtirishning o‘zi mazmuniy muvofiqlikni ta’minlamaydi. Bir voqeaga bag‘ishlangan ikki gap boshqa-boshqa tafsilotlarni ifodalashi mumkin. Aksincha, shakli ancha farq qiladigan gaplarda ayni axborot saqlanishi ham kuzatiladi.

O‘zbek va ingliz tillarining grammatik qurilishi bu masalani yanada murakkablashtiradi. O‘zbek tilida qo‘shimcha bilan beriladigan egalik, kelishik yoki inkor ma’nosi ingliz tilida alohida so‘zlar yordamida ifodalanishi mumkin. Eganing tushirilishi, kesimning joylashuvi, aniqlovchilarning tuzilishi va gaplarni birlashtirish imkoniyatlari ham bir xil emas. Shu sababli so‘zlar soni, tartibi va gap uzunligi yordamchi belgilar bo‘lib xizmat qiladi; yakuniy qaror esa ifodalangan mazmun va uning matndagi vazifasiga tayanishi lozim.

O‘zbek–ingliz parallel korpusini yaratish bosqichlari avvalgi tadqiqotda yig‘ishdan annotatsiyalashgacha bo‘lgan izchil jarayon sifatida yoritilgan [1]. Ekstralingvistik talablar haqidagi ishda kontekst, janr va tarjima kelib chiqishi korpus sifatining zarur shartlari sifatida ko‘rsatilgan [2]. Mazkur maqolada ana shu yondashuvlar asosida torroq masala qo‘yiladi: tarjimadagi tabiiy qayta ifodalashni noto‘g‘ri bog‘langan segmentdan qanday farqlash mumkin? Maqsad ushbu farqni izohlash va takror qo‘llash mumkin bo‘lgan annotatsiya qarorlariga aylantirishdan iborat.

UzEnParC 1.0 tavsifida asl va me’yorlashtirilgan o‘zbekcha matnni alohida saqlash, segmentni hujjatga bog‘lash hamda tekshiruv holatini qayd etish tamoyillari belgilangan [3]. Bunday tuzilma bahsli juftlikning kelib chiqishini tekshirishga imkon beradi. Gibrid tozalash haqidagi tadqiqotda esa semantik yaqinlik, moslashtirish belgilari va o‘zbek tiliga xos morfologik axborot birgalikda qo‘llanadi. Unda qisman parallel juftliklar saralashning murakkab toifasi sifatida ajratilgan [4]. Biroq qismanlikning har bir ko‘rinishi bir xil tahririy qarorni talab qilmaydi.

Baker va Jones tarjimada muqobillikni so‘z, birikma, grammatika, matniy bog‘lanish va pragmatika doirasida ko‘rib chiqadi [5]. Ushbu yondashuvdan korpus tahlili uchun kelib chiqadigan xulosa shuki, alohida so‘zning tashqi muqobili yo‘qligi butun xabarning yo‘qolganini bildirmaydi. Mazmun boshqa grammatik vositada yoki qo‘shni gapda ifodalangan bo‘lishi mumkin. Shu bilan birga, erkinroq ifoda mavjudligi har qanday qo‘shish yoki tushirishni oqlamaydi. Uning asoslanganligi aynan tahlil qilinayotgan matn orqali ko‘rsatilishi kerak.

Avtomatik moslashtirishda Gale va Church gap uzunligiga asoslangan statistik usulni taklif etgan [6]. Vecalign esa ko‘p tilli vektor ifodalaridan foydalanadi [7]. So‘z darajasidagi bog‘lanishlarni aniqlashda Dou va Neubig kontekstual vektorlarni parallel matnlarda moslashtirib o‘qitish usulini ishlab chiqqan [8]. Mazkur usullar nomzod bog‘lanishni topish uchun muhim. Lekin algoritm qaytargan bog‘lanishning lingvistik talqini alohida vazifa bo‘lib qoladi: yaqinlik bali qaysi ma’no saqlanganini yoki o‘zgarganini o‘z-o‘zicha tushuntirmaydi.

OpusFilter turli sifat belgilarini bir jarayonda qo‘llash imkonini beradi [9]. Tarjima sifatini tahliliy baholashga mo‘ljallangan MQM yondashuvi esa nuqsonlarni turlarga ajratishga asoslanadi [10]. Freitag va hammualliflar ekspert bahosi hamda hujjat kontekstining ahamiyatini ko‘rsatgan [11]. Bu ishlardan foydalanishda ikki vazifani aralashtirmaslik zarur: tarjima bahosi maqsad matnning asliyatni qay darajada ifodalaganini tekshiradi; moslashtirish bahosi esa tegishli parchalar bir-biriga to‘g‘ri bog‘langanini aniqlaydi. Ular o‘zaro aloqador bo‘lsa-da, bir-birining o‘rnini bosa olmaydi.

Metodlar.

Tadqiqot nazariy-metodologik xususiyatga ega. Materialni UzEnParC loyihasiga oid to‘rtta maqola [1–4], xalqaro uslubiy adabiyotlar hamda O‘zbekiston Respublikasi Prezidentining rasmiy saytida 2025-yil 30-dekabrda e’lon qilingan xabarning o‘zbekcha va inglizcha shakllari tashkil etadi [14]. Rasmiy xabardan miqdor ifodasi va harakatning amalga oshish holatini qiyoslashga imkon beradigan ikki juftlik maqsadli tanlandi. Bu kichik material butun korpusdagi hodisalar chastotasini hisoblash uchun qo‘llanmaydi.

Qiyosiy tahlil, mazmuniy tarkibni ajratish va minimal qarama-qarshi misollar usulidan foydalanildi. Rasmiy matndan olingan juftliklar manbasi bilan beriladi. Qolgan qisqa misollar mezonning ishlashini tushuntirish uchun tuzilgan nazorat misollaridir; ular korpusdan olingan dalil hisoblanmaydi. Bunday ajratish tabiiy matnda kuzatilgan holat bilan nazariy imkoniyatni aralashtirmaslikka xizmat qiladi. Taklif etilayotgan tartibning aniqligi yoki annotatorlar kelishuvi bo‘yicha yangi statistik natija ilgari surilmaydi.

Har bir juftlikda avvalo umumiy voqea va predikat, keyin ishtirokchilar, grammatik ma’nolar va tafsilotlar tekshiriladi. Tafovut aniqlansa, qo‘shni gaplarga murojaat qilinadi. Shundan keyingina uning sababi segment chegarasi, tarjimaviy qayta ifodalash yoki mazmuniy nomuvofiqlik bilan bog‘lanadi. Ushbu ketma-ketlikda matnning asl shakli tahrir qilinmaydi. Tahlilchining xulosasi alohida annotatsiyada saqlanadi, chunki korpus nashr qilingan matnni tekshirish va

qayta talqin qilish imkonini ham saqlashi kerak.

Tahlilning birligi sifatida alohida so‘z emas, nashrda mustaqil chegaraga ega bo‘lgan gap va uning mazmuniy tarkibi olindi. Har bir nomzod juftlik uchun dastlab hujjat manbasi va nashr sanasi tekshirildi, undan keyin gapdagi asosiy predikat, uning ishtirokchilari hamda miqdor, inkor, modallik va zamonga doir belgilar ajratildi. So‘ng mazmun qo‘shni segmentga ko‘chgan-ko‘chmagani kuzatildi. Bu tartibning afzalligi shundaki, bir xil so‘zlar mavjudligi dastlabki belgi bo‘lib qoladi, ammo yakuniy qaror uchun yetarli dalilga aylantirilmaydi. Natijalarni sharhlashda «manbada kuzatildi», «nazorat misolida ko‘rsatildi» va «keyingi tekshiruv uchun taklif etildi» degan uch xil dalil maqomi qat’iy ajratildi.

Tadqiqotda asosiy ko‘rsatkich sifatida moslashtirilgan gaplar sonidan kelib chiqadigan statistik natija hisoblanmadi. Ikki tabiiy juftlik maqsadli tanlangani sababli ularning ulushi, xatolar chastotasi yoki taklif etilgan tartibning aniqlik foizi haqida umumlashtiruvchi xulosa chiqarish mumkin emas. Buning o‘rniga, takror tekshiriladigan qaror yozuvi modeli ishlab chiqildi: unda manba havolasi, segment chegarasi, kuzatilgan lingvistik belgi, muqobil izohlar hamda yakuniy annotatsiya alohida beriladi. Mazkur model keyinchalik bir nechta mutaxassis ishtirokida tekshirish uchun metodik asos vazifasini bajarishi mumkin.

Natijalar.

Moslashtirishning to‘g‘riligi deganda, manba segmentga uning tarjima matnidagi tegishli o‘rni bog‘langanligi tushuniladi. Mazmuniy muvofiqlik esa shu o‘rinda axborotning qay darajada saqlanganini bildiradi. Masalan, tarjimon asliyatdagi inkorni tushirib qoldirgan bo‘lsa, tegishli gaplar baribir to‘g‘ri bog‘langan bo‘lishi mumkin. Bunda nuqson tarjima mazmuniga tegishlidir. Gapni boshqa segmentga bog‘lash orqali bunday tafovutni bartaraf etib bo‘lmaydi. Noto‘g‘ri moslashtirishda esa kerakli mazmun tarjimaning boshqa o‘rnida uchraydi yoki tanlangan gap umuman tegishli bo‘lmaydi.

Shu asosda bitta umumiy «sifat» yorlig‘i o‘rniga mustaqil belgilarni qayd etish maqsadga muvofiq. Moslashtirish holati to‘g‘ri, chegarasi qayta ko‘riladigan, noto‘g‘ri yoki noaniq bo‘lishi mumkin. Mazmuniy muvofiqlik yetarli, qisman, zid yoki baholash uchun kontekst yetishmaydigan holat sifatida belgilanadi. Transformatsiya turi esa so‘z tartibining o‘zgarishi, grammatik almashtirish, gapni bo‘lish, birlashtirish yoki izohlash singari alohida tavsiflanadi. Transformatsiya borligi mazmuniy nuqson borligini anglatmaydi.

Tarjima transformatsiyalarini farqlash mezonlari

Predikativ mazmun va ishtirokchilar. Dastlab har ikki segmentda nima sodir bo‘lgani va voqeada kimlar qatnashgani aniqlanadi. Nazorat misolida «Tadqiqotchi muharrirni tanqid qildi» gapining «The editor criticised the researcher» shaklida berilishi deyarli ayni lug‘at tarkibini saqlaydi, biroq harakat bajaruvchisi bilan obyekti almashadi. Demak, umumiy so‘zlar ko‘pligi yetarli dalil emas. «The editor was criticised by the researcher» variantida esa nisbat va so‘z tartibi o‘zgarsa-da, ishtirokchilarning mazmuniy roli saqlanadi.

Inkor va modallik. «U hujjatni imzolashi mumkin» hamda «He must sign the document» nazorat juftligida harakat va ishtirokchi mos, lekin imkoniyat majburiyatga aylangan. Xuddi shuningdek, inkorning yo‘qolishi yoki uning boshqa bo‘lakka tegishli bo‘lib qolishi xabar mazmunini o‘zgartiradi. Shu sababli «emas», «yo‘q», inkor qo‘shimchalari, ehtimollik va zaruriyat ifodalari alohida tekshiriladi. Inglizchadagi yordamchi fe’lning o‘zbekchada mustaqil so‘z bilan berilmagani esa uning ma’nosi yo‘qolganini anglatmaydi; bu ma’no fe’l shaklida ifodalanishi mumkin.

Zamon va harakat holati. Rejalashtirilgan, davom etayotgan va tugallangan ish bir voqeaning turli holatlaridir. Ular orasidagi farq axborot matnida muhim bo‘lishi mumkin. Biroq o‘zbekcha va inglizcha zamon shakllarini faqat grammatik nomi bo‘yicha tenglashtirish ham to‘g‘ri emas. Qiyoslashda vaqt ravishlari, matnning hikoya nuqtasi va harakat natijasining mavjudligi hisobga olinadi. Tafovut aniqlansa, avvalo uning matn kontekstida izohi bor-yo‘qligi tekshiriladi; izoh topilmasa, holat mazmuniy tekshiruv uchun belgilanadi.

Miqdor va chegaralovchi birliklar. Sonlarning yozilishi turlicha bo‘lsa ham, qiymati va o‘lchov birligi bir xil qolishi mumkin. Aksincha, ayni raqam yonidagi «kamida», «qariyb», «ortiq» kabi birliklarning tushirilishi miqdor haqidagi hukmni o‘zgartiradi. «Yigirmadan ortiq» bilan «twenty» o‘rtasidagi farq faqat so‘zlar soniga tegishli emas. Foiz va foiz punkti, ming va million, muddatning boshlanishi va tugashi ham farqlanadi. Avtomatik tekshiruv sonni topish bilan cheklanmay, uni aniqlab kelgan unsurga ham murojaat qilishi lozim.

Kontekst va referensiya. Olmoshning muqobili uning qaysi shaxs yoki predmetga ishora qilishiga bog‘liq. «U keldi» kabi gap alohida olinganda shaxsning jinsini ko‘rsatmaydi; inglizcha variantni baholash uchun oldingi matn kerak bo‘lishi mumkin. Kontekst yetishmayotgan vaziyatda xato to‘g‘risida qat’iy qaror chiqarish asossizdir. Yana bir xavf shundaki, har ikki tilda bir xil olmosh ishlatilsa ham, ular turli shaxslarga ishora qilishi mumkin. Tekshiriladigan narsa olmoshning tashqi shakli emas, uning matndagi referentidir.

Mazmunning segmentlar orasida taqsimlanishi. Tarjimada ma’lumot qo‘shni gapga ko‘chgan bo‘lishi mumkin. Nazorat misolida «Yig‘ilish tugagach, ishtirokchilar tarqaldi» gapi «The meeting ended. The participants then left» tarzida berilsa, mazmun ikki gapga taqsimlanadi. O‘zbekcha gapni faqat inglizcha birinchi gap bilan bog‘lash axborotni qisman qamrab oladi. Shu o‘rinda 1x2 moslashtirish asosli. Inglizcha ikki gapni sun’iy ravishda bir gapga aylantirish shart emas: korpus asl chegaralarni saqlab, bog‘lanishni kengaytirishi mumkin.

Miqdor ifodasiga oid birinchi juftlik rasmiy xabarda quyidagicha berilgan: «Birinchi bosqichda 6 ming gektar hududda qurilish ishlari bajariladi.» va «In the first stage, construction work will cover 6,000 hectares of land.» [14]. O‘zbekcha son aralash yozuvda, inglizchasi raqam bilan ifodalangan. Shunga qaramay, qiymat hamda o‘lchov birligi saqlangan. Kesimning shakli ham aynan takrorlanmaydi: o‘zbekcha majhul qurilma inglizchada boshqa sintaktik yechim bilan berilgan. Mazkur farqlar juftlikni rad etishga asos bo‘la olmaydi.

Ikkinchi juftlikda o‘zbekcha «Hududda o‘n kilometrlik sohil bog‘i tashkil etildi.» gapiga inglizcha «A ten-kilometer

coastal park is being created on the territory.» gapi muvofiq keladi [14]. Har ikki gapda predmet, uning uzunligi va joyga ishora saqlangan. Biroq o‘zbekcha kesim ishning amalga oshganini, inglizcha shakl esa davom etayotgan jarayonni bildiradi. Bu tafovutni oddiy so‘z tartibi almashishi yoki majhul nisbat bilan izohlab bo‘lmaydi. Tekshirishni talab qiladigan nuqta harakatning amalga oshish holatidir.

Shu bilan birga, mazkur tafovut gaplar noto‘g‘ri bog‘langanini isbotlamaydi. Ikkala segment bir predmet haqidagi tegishli axborotni beradi. Farqning tarjima jarayoni, matn tahriri yoki turli yangilanishlarga borib taqalishini faqat ochiq sahifa matnidan aniqlab bo‘lmaydi. Shuning uchun moslashtirish «to‘g‘ri», mazmuniy muvofiqlik esa «harakat holati bo‘yicha tafovut mavjud» deb qayd etiladi. Sabab haqidagi hukm ochiq qoldiriladi. Bu usul aniqlangan farqni yashirmaydi va tekshirilmagan sababni ham muallifga nisbat bermaydi.

Rasmiy manba misollarini uch mustaqil mezon bo‘yicha qayd etish natijasida bir muhim tafovut aniq ko‘rinadi: «moslashtirish to‘g‘ri» degan hukm «mazmun to‘liq aynan» degan hukmga teng emas. Birinchi juftlikda 6 ming gektar va 6,000 hectares bir xil miqdorni bildiradi; qurilish ishining grammatik ifodalanishi turlicha bo‘lsa-da, asosiy xabar mos keladi. Ikkinchi juftlikda esa o‘n kilometrlik sohil bog‘i haqidagi ishora bir xil, biroq «tashkil etildi» va «is being created» kesimlari bajarilgan ish bilan davom etayotgan jarayonni farqlaydi. Shuning uchun ikkinchi holatda qayta moslashtirish emas, mavjud bog‘lanish uchun mazmuniy ogohlantirish zarur.

Tahlilning amaliy ko‘rinishi 2-jadvalda jamlandi. Jadvaldagi baholar faqat ikki ko‘rsatilgan gap juftligiga tegishli; ular UzEnParC korpusining barcha yozuvlariga nisbatan qo‘llangan avtomatik tasnif natijasi emas. Nazorat misollarida esa subyekt va obyekt o‘rni almashishi, modal ma’noning kuchayishi va mazmunning ikkita gapga taqsimlanishi alohida xavf sifatida namoyon bo‘ladi. Ushbu farqlar annotatorning birgina «to‘g‘ri/noto‘g‘ri» belgisiga tayanishini cheklaydi va tekshirish sababini qayd etishni talab qiladi.

1-jadval.

Tafovutni talqin qilish va annotatsiya qarori

Kuzatilgan holat	Talqin	Annotatsiya qarori
Shakl o‘zgargan, mazmun saqlangan	Asoslangan transformatsiya	Muvofiq juftlik sifatida saqlash
Mazmun qo‘shni segmentda berilgan	Bog‘lanish chegarasi muammosi	1x2 yoki 2x1 imkoniyatini tekshirish
Tegishli gapda muhim ma’no farqlangan	Mazmuniy nomuvofiqlik	Bog‘lanishni saqlab, farqni belgilash
Boshqa voqea yoki boshqa parcha bog‘langan	Moslashtirish xatosi	Bog‘lanishni tuzatish yoki rad etish
Kontekst yoki nashr ma’lumoti yetishmaydi	Qaror uchun dalil yetarli emas	Noaniq holat sifatida tekshiruvga qoldirish

Manba: muallifning tarjima moslashtirish mezonlari asosidagi ishlanmasi.

2-jadval.

Rasmiy matndan olingan gap juftliklarining qiyosiy bahosi

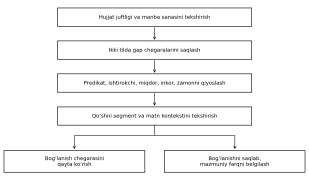
Tekshirilgan belgi	Birinchi juftlik (6 ming gektar)	Ikkinchi juftlik (sohil bog‘i)
Predmet / asosiy voqea	Qurilish ishlari — mos	Sohil bog‘i — mos
Miqdor	6 ming = 6,000 gektar; mos	O‘n kilometr = ten-kilometer; mos
Harakat holati	Rejalashtirilgan ish; mazmun mos	Tugallangan / davom etayotgan; tafovut
Moslashtirish qarori	Tegishli gaplar bog‘langan	Tegishli gaplar bog‘langan
Annotatsiya izohi	Grammatik qayta ifodalash	Harakat holati bo‘yicha mazmuniy farq

Manba: O‘zbekiston Respublikasi Prezidentining 30.12.2025-yildagi o‘zbekcha va inglizcha rasmiy xabarlari [14] asosida muallif tahlili.

Tekshiruv hujjat darajasidan boshlanadi. Sarlavha, nashr sanasi, manba manzili va matnning tuzilishi qiyoslanadi. Bitta mavzudagi ikki xabar tarjima juftligi deb avtomatik qabul qilinmaydi. Asliyat tili yoki tarjima yo‘nalishi noma’lum bo‘lsa, bu ma’lumot taxmin bilan to‘ldirilmaydi. Bender va Friedman tavsiya etgan ma’lumotni hujjatlashtirish yondashuvi korpus haqidagi xulosalarning qo‘llanish doirasini aniq ko‘rsatishga xizmat qiladi [13]. Shu jihatdan kelib chiqish ma’lumoti annotatsiya qarorining tekshiriladigan asosidir.

Keyingi bosqichda har ikki tilda gap chegaralari mustaqil saqlanadi. Annotator nomzod juftlikni o‘qib, yuqoridagi mezonlar bo‘yicha farqlarni aniqlaydi. Mazmunning bir qismi yetishmasa, avvalo oldingi va keyingi segmentlar ko‘riladi. Kerakli axborot shu yerda topilsa, bog‘lanish chegarasi qayta belgilanadi. Topilmasa, tarjimadagi tushirish, matn versiyalaridagi farq yoki noto‘g‘ri juftlash ehtimoli ajratib tekshiriladi. Birinchi taxminni tasdiqlovchi dalil bilan cheklanish o‘rniga, muqobil izohlar ham qiyoslanadi.

1-rasm. Moslashtirish va mazmuniy tafovutni ajratish bosqichlari



Manba: muallif tomonidan ishlab chiqilgan tekshiruv sxemasi.

Qaror yozuvida hujjat va segment identifikatori, moslashtirish turi, mazmuniy holat, aniqlangan unsur hamda qisqa asos bo‘lishi yetarli. Masalan, «son mos emas» o‘rniga «manbada quyi chegara ifodalangan, tarjimada aniq miqdor berilgan» kabi tekshiriladigan izoh yoziladi. Xuddi shuningdek, «tarjima erkin» degan umumiy baho o‘rniga qaysi mazmun boshqa shaklga ko‘chirilgani ko‘rsatiladi. Izohni keyingi tekshiruvchi ayni matnga murojaat qilib tasdiqlay olishi kerak.

Korpusning asosiy qidiruv qatlamiga faqat tasdiqlangan bog‘lanishlar chiqarilishi mumkin, biroq rad etilgan nomzodlar va tekshiruv sabablari alohida audit yozuvida saqlanishi foydali. Bu tanlov xatolarini keyinchalik o‘rganishga imkon beradi. Asl matn tahrir qilinmaydi; texnik me’yorlashtirish alohida qatlamda bajariladi [3]. Annotator taklif etgan yaxshi tarjimani manbadagi nashr qilingan tarjima o‘rniga yozish esa tadqiqot materialini o‘zgartirib yuboradi. Bunday tuzatish kerak bo‘lsa, u mustaqil tahririy taklif sifatida saqlanishi lozim.

Bir necha annotator ishlaganida dastlab kichik kalibrlash tanlanmasi yordamida qoida talqini muvofiqlashtiriladi. Shundan so‘ng mustaqil belgilash o‘tkazilib, kelishmovchiliklar sabablar bo‘yicha tahlil qilinadi. Moslashtirish holati va mazmuniy muvofiqlik bo‘yicha kelishuv alohida hisoblanishi maqsadga muvofiq: ikki bahoni birlashtirish aynan qaysi qaror qiyin bo‘lganini yashirishi mumkin. Yakuniy etalon yorliq muhokama orqali belgilanadi. Ushbu maqolada bu tartib keyingi empirik tekshiruv uchun taklif etilmoqda.

Muhokama.

Taklif etilgan yondashuvning asosiy natijasi turli sababli tafovutlarni bitta «shovqin» toifasiga birlashtirmaslikdir. Keng ko‘lamli veb-korpuslar auditi ma’lumotlarning tili va foydalanishga yaroqliligini alohida tekshirish zarurligini ko‘rsatadi [12]. O‘zbek–ingliz juftliklarida bunga grammatik ma’no va segment chegarasi tahlili qo‘shiladi. Avtomatik saralash tekshiriladigan nomzodlarni kamaytirishi mumkin; lingvistik annotatsiya esa saralash qarorining sababini ochib beradi. Mazkur maqolada shu sababni yozib borishning amaliy shakli taklif qilindi.

Korpusdan foydalanish maqsadi ham hisobga olinadi. Tarjima xatolarini o‘rganish uchun mazmuniy tafovuti bor, lekin to‘g‘ri moslashtirilgan juftlik qimmatli materialdir. Mashina tarjimasiga o‘quv ma’lumoti tanlashda esa bunday juftlik alohida cheklovga tushishi mumkin. Shu sababli «korpusda saqlash» va «muayyan tajribaga kiritish» qarorlari ajratiladi. Foydalanish sharti aniq qayd etilsa, turli tadqiqotlar bir manbadan o‘z vazifasiga mos tanlanma tuza oladi.

Janrga bog‘liq tafovutlar ham muhim. Badiiy tarjimada obraz, ohang va murojaat shaklining vazifasi kengroq kontekstda ochiladi; rasmiy xabarda esa miqdor, muddat va harakatning amalga oshish holati ko‘pincha bevosita tekshiriladi. Biroq janr o‘zgarishi dalil talabini bekor qilmaydi. Har qanday qayta ifodalashning qaysi mazmunni saqlagani ko‘rsatilishi zarur. Shuning uchun tavsiya etilgan mezonlar janrga qarab to‘ldirilishi mumkin, lekin tekshirilmagan tafovutni «uslubiy erkinlik» bilan oqlash uchun ishlatilmasligi kerak.

Ishning chegarasi maqsadli tanlangan ikki tabiiy juftlik va nazorat misollari bilan bog‘liq. Taklif etilgan toifalar katta tanlanmada sinovdan o‘tkazilgandagina ularning barqarorligi, o‘zaro farqlanishi va annotatorlarga tushunarliligi baholanadi. Keyingi tekshiruv turli janrlarni, ko‘p gapli bog‘lanishlarni va avtomatik tizim asossiz rad etishi mumkin bo‘lgan muvofiq juftliklarni qamrab olishi zarur.

Taklif qilingan belgilar tizimining amaliy qiymati uning tahririy qarorni izohlashga xizmat qilishida ko‘rinadi. Masalan, shakliy jihatdan nomuvofiq bo‘lgan, ammo axboroti saqlangan juftlik tilshunoslik tahlili uchun ahamiyatli bo‘lishi mumkin; mazmuni o‘zgargan, lekin manbaga tegishli gap esa tarjima sifatini o‘rganishga yordam beradi. Noto‘g‘ri segmentga bog‘langan gap bunday vazifalarning har ikkisida yanglish xulosa tug‘diradi. Shu sababli korpus yozuvining o‘zi, uning lingvistik bahosi va tadqiqotga kiritish mezoni bir xil ma’lumot sifatida saqlanmasligi lozim.

Xulosa va tavsiyalar.

O‘zbek va ingliz tillari parallel korpusida tarjima transformatsiyasini moslashtirish xatosidan farqlash uchun mazmunni segment chegarasi bilan birga tekshirish lozim. Gaplarning turlicha qurilishi yoki so‘zlar sonining farqlanishi xatoni tasdiqlamaydi. Predikat, ishtirokchilar, inkor, modallik, miqdor va kontekstga oid dalillar qarorning lingvistik asosini tashkil etadi. Mazmun qo‘shni gapga ko‘chgan vaziyatda esa asl gap chegaralarini buzmasdan moslashtirish doirasini kengaytirish mumkin.

Maqolada moslashtirish holati, mazmuniy muvofiqlik va transformatsiya turini mustaqil qayd etish taklif qilindi. Rasmiy matn misollari to‘g‘ri bog‘langan juftlik ichida ham harakat holatiga oid tafovut bo‘lishi mumkinligini ko‘rsatdi. Bunday tafovutning mavjudligi bilan uning sababi haqidagi xulosa ajratilishi kerak. Taklif etilgan annotatsiya tartibi korpus yozuvini asosli tanlash, bahsli qarorni qayta ko‘rish va asl matnni saqlashga xizmat qiladi. Uning amaliy samaradorligi keyingi bosqichda mustaqil annotatsiya va kengroq material asosida tekshirilishi lozim.

Foydalanilgan adabiyotlar ro'yxati

[1] Elov B. B., Kurbanova M. M., Yusupova T. A., Amirkulov M. A., Barnoyeva Z. S., Shodmonova M. B. Developing and Processing the Uzbek-English Parallel Corpus: Approaches and Computational Tools // 2026 5th International Informatics and Software Engineering Conference (IISEC). Ankara: IEEE, 2026. P. 204–208. DOI: 10.1109/IISEC69317.2026.11418506.

[2] Elov B. B., Amirkulov M. A., Tulanova N. M. Extralinguistic Requirements for the Uzbek-English Parallel Corpus // 2026 5th International Informatics and Software Engineering Conference (IISEC). Ankara: IEEE, 2026. P. 250–255. DOI: 10.1109/IISEC69317.2026.11418459.

[3] Elov B., Amirkulov M. UzEnParC 1.0: A Cross-Script, Metadata-Rich Uzbek-English Parallel Corpus for Low-Resource NLP // Speech and Computer. SPECOM 2026. – Cham : Springer, 2027. – Vol. 16935. – P. 225–238. – DOI: 10.1007/978-3-032-37870-5_16.

[4] Elov B. B., Toirova G. I., Fayziyev Z. X., Amirkulov M. A., Sultonova I. Sh. Data-Centric Bitext Cleaning for Uzbek-English: Hybrid Quality Estimation and Noise Filtering for Low-Resource Parallel Corpora // 2026 International Conference on Computer Science and Engineering (UBMK). IEEE, 2026. P. 1507–1512.

[5] Baker M., Jones H. In Other Words: A Coursebook on Translation. 4th ed. London: Routledge, 2026. 410 p. DOI: 10.4324/9781003440857.

[6] Gale W. A., Church K. W. A Program for Aligning Sentences in Bilingual Corpora // Computational Linguistics. 1993. Vol. 19, no. 1. P. 75–102. URL: https://aclanthology.org/J93-1004/.

[7] Thompson B., Koehn P. Vecalign: Improved Sentence Alignment in Linear Time and Space // Proceedings of EMNLP-IJCNLP 2019. Hong Kong: Association for Computational Linguistics, 2019. P. 1342–1348. DOI: 10.18653/v1/D19-1136.

[8] Dou Z.-Y., Neubig G. Word Alignment by Fine-tuning Embeddings on Parallel Corpora // Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics. Main Volume. 2021. P. 2112–2128. DOI: 10.18653/v1/2021.eacl-main.181.

[9] Aulamo M., Virpioja S., Tiedemann J. OpusFilter: A Configurable Parallel Corpus Filtering Toolbox // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. 2020. P. 150–156. DOI: 10.18653/v1/2020.acl-demos.20.

[10] Mariana V., Cox T., Melby A. The Multidimensional Quality Metrics (MQM) Framework: A New Framework for Translation Quality Assessment // The Journal of Specialised Translation. 2015. No. 23. P. 137–161. DOI: 10.26034/cm.jostrans.2015.343.

[11] Freitag M., Foster G., Grangier D., Ratnakar V., Tan Q., Macherey W. Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation // Transactions of the Association for Computational Linguistics. 2021. Vol. 9. P. 1460–1474. DOI: 10.1162/tacl_a_00437.

[12] Kreutzer J., Caswell I., Wang L. va boshqalar. Quality at a Glance: An Audit of Web-Crawled Multilingual Datasets // Transactions of the Association for Computational Linguistics. 2022. Vol. 10. P. 50–72. DOI: 10.1162/tacl_a_00447.

[13] Bender E. M., Friedman B. Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science // Transactions of the Association for Computational Linguistics. 2018. Vol. 6. P. 587–604. DOI: 10.1162/tacl_a_00041.

[14] O‘zbekiston Respublikasi Prezidenti. Prezidentimiz Yangi Toshkentda amalga oshirilayotgan loyihalar taqdimoti bilan tanishdi // Prezidentning rasmiy veb-sayti. – 30.12.2025. – URL: https://www.president.uz/oz/lists/view/8845; The President reviews the presentation of ongoing projects in New Tashkent. – URL: https://www.president.uz/en/lists/view/8845 (murojaat qilingan sana: 31.10.2026).