

A CAUSATIVE-SPECIFIC ERROR TAXONOMY FOR UZBEK-ENGLISH MACHINE TRANSLATION

Bahriddinova Raximaxon Bahriddin qizi¹Husanov Ilyos Jamoliddin o'g'li²¹Department of Teaching English to Speakers of Other Languages (TESOL), Webster University in Tashkent, Tashkent, Uzbekistan.*E-mail:* rbaxriddinova@webster.edu²Department of Business Management, Turan International University, Namangan, Uzbekistan.*E-mail:* husanov_ilyos@turan-edu.uz<https://doi.org/10.5281/zenodo.23139698>

Abstract. A translation score alone does not tell us whether causative meaning has survived. This article presents a procedure for evaluating Uzbek causative verbs translated into English. It combines an operational definition and inclusion rules with a targeted test set, two layers of error codes, and blind assessment against a short statement of the required meaning. Each target receives one causative code (C0–C7); general errors elsewhere in the sentence are recorded separately (G1–G7). Coding 112 outputs from four systems exposed several weaknesses in the procedure. Sixteen decisions (14%) were ties, all in sentences with uncertain class labels. Ten of the 27 sentences fitted none of the predefined classes. A reference set drafted after the outputs had been collected also closely resembled one system's translations. These findings guide the proposed revisions.

Keywords: causative verbs; Uzbek; machine translation evaluation; error taxonomy; annotation; MQM.

Annotatsiya. Umumiy sifat ko'rsatkichlari grammatik ma'no tarjimada saqlanganini ko'rsatmaydi. Maqolada o'zbek tilidagi kauzativ fe'llarning ingliz tiliga mashina tarjimasini baholash metodikasi taqdim etiladi: kiritish qoidalariga ega operatsion ta'rif, kichik maqsadli test to'plami, ikki qatlamli kodlash tizimi (har bir fe'lga bitta kauzativ kod C0–C7 va umumiy kodlar G1–G7) hamda andoza tarjima o'rniga talab qilinadigan ma'no bayoniga tayangan, tizim nomi yashirilgan baholash. To'rt tizimning 112 ta tarjimasiga qo'llanganda tasnif kauzativ xatolarni boshqa xatolardan ajratdi. Zaif jihatlari ham aniqlandi: 16 ta qarorda (14%) ikki kod teng asosli bo'lib, barchasi sinf belgisi noaniq gaplarda uchragan; 27 ta gapdan o'ntasi oldindan belgilangan sinflarga mos kelmagan; tarjimalardan keyin tayyorlangan andoza tarjima bitta tizim natijasini takrorlagan. Tuzatishlar taklif etiladi.

Kalit so'zlar: kauzativ fe'llar; o'zbek tili; mashina tarjimasini baholash; xatolar tasnifi; annotatsiya; MQM.

Аннотация. Общие показатели качества не показывают, сохранилось ли грамматическое значение при переводе. В статье предлагается методика оценки машинного перевода узбекских каузативных глаголов на английский язык: операциональное определение с правилами включения, небольшой целевой тестовый набор, двухуровневая схема кодирования (один каузативный код C0–C7 для каждого целевого глагола и общие коды G1–G7) и слепая оценка по описанию требуемого значения вместо эталонного перевода. При применении к 112 переводам четырёх систем схема отделила каузативные ошибки от прочих. Выявлены и слабые места: 16 решений (14%) оказались спорными, все – в предложениях с неуверенной классовой меткой; десять из 27 предложений не вписались в заданные классы; эталон, составленный после получения переводов, повторял одну из систем. Предлагаются изменения.

Ключевые слова: каузативные глаголы; узбекский язык; оценка машинного перевода; классификация ошибок; аннотирование; MQM.

1. Introduction

A single translation score can conceal quite different errors. Automatic metrics such as BLEU, TER and COMET compare outputs with reference translations; recent Uzbek work combines these scores with MQM error annotation [10]. Error typologies provide more detail [7, 2], but a broad category such as “mistranslation” can include both reversed participant roles and a causative verb that is too weak. Those errors change the reader’s understanding in different ways. Uzbek causatives make the distinction especially useful: Uzbek adds a suffix, as in *yoz-dir-* ‘have (someone) write’, while English uses a lexical verb or make, have, let or get [1, 4]. Uzbek MT studies have so far assessed whole sentences through general categories. Sultonova [10] identifies suffix dropping as a major source of error but does not examine individual verb categories. Here, I describe a procedure for causatives and examine what its pilot application revealed.

2. Methods

2.1. Definition and test set

A correct translation must preserve causation, the roles of causer and causee, and the type of causation. Contrastive studies distinguish these meanings in both languages [5, 9]. An appropriate English lexical causative, such as teach or kill, meets these requirements. The core set includes verbs with a transparent base and a productive suffix: *-t*, *-tir*, *-dir* or *-giz/-qiz/-g'iz*. These suffixes account for 189 of the 211 suffix-derived causatives counted by Karimjonova [5, p. 19]. Lexicalised pairs, including *o'ldirmoq* ‘kill’, stacked forms and noun-derived forms are included with separate tags. Analytic causatives such as *ijozat bermoq* ‘give permission’ serve as controls; idioms and verbs without a modern base are excluded. Each sentence receives a semantic class—inanimate causee, coercion, permission, delegation, informing or unclear—and a confidence level. The pilot comprises 27 sentences with 28 target verbs from modern prose [6, 8, 3], plus one control sentence. I rebuilt an initial 30-sentence set because it included archaic and decree-style passive forms in which the causative had become a fixed formula rather than an active relation.

2.2. Coding

The coding unit is one target verb in one translation. The first layer assigns exactly one causative code, checked in the order shown in Table 1. The second records any general errors elsewhere in the sentence: G1 mistranslation, G2 omission, G3 addition, G4 untranslated word, G5 grammar, G6 names and realia, and G7 pronoun gender. Gender is logged but does not count as an error. Google’s “lay her naked and bathe her” for *yalang'och yotqizib... savalasayam* shows why the layers are separate. Lay preserves the causative and receives C0; bathe mistranslates ‘beat’ and receives G1. When two causative codes fit equally well, the earlier code is assigned and the tie is recorded, allowing both strict and lenient counts. Each error is rated minor or major. During coding, shuffled letters conceal the system names.

Table 1. Causative codes in testing order and their use in the pilot (n = 112 outputs)

Code	Label	Uses
C7	Untranslated, omitted or garbled verb	6
C6	Wrong verb (unrelated meaning)	3
C2	Roles of causer and causee confused	6
C1	Causation lost	4
C3	Wrong type of causation (e.g. <i>make</i> for delegation)	10
C4	Recast as a passive or other structure	3
C5	Double causative flattened	0
C0	Correct (lexical causatives included)	80

Each target is assessed against a short statement of the required meaning, for example ‘cause the two farm chairmen to hand over two sheep’, rather than a reference translation. The pilot exposed a reason for this choice. After the outputs had been collected, an AI assistant drafted a reference set, which was then edited. Its mean character similarity, measured with Python difflib, was 0.90 to one language model’s outputs, compared with 0.63–0.71 for the other systems. Two reference sentences were identical to that model’s translations. Using this set as the benchmark would have favoured the system it most closely resembled.

3. Results

Sixteen coding decisions (14%) were ties. Seven fell between C3 and C0, three between C1 and C0 when causation became accompaniment (“came with his gang”), and three between C4 and C0 for a natural passive rendering of o’g’irlatgan. The remaining three involved other boundaries. All sixteen came from sentences whose class labels had medium or low confidence; none occurred in the ten high-confidence main sentences. The suffix -dir rarely specifies whether someone was forced, allowed or asked to act, so uncertainty in the class label carries over into C3. I therefore propose assigning C3 only when the class label has high confidence.

The semantic classes also left too much unclassified. Ten of the 27 sentences were labelled “unclear”. They involved caretaking, such as serving a drink, dressing, nursing or grazing; direct handling, such as laying down or killing; accompaniment; and abstract causers. Karimjonova’s distinction between physical influence and speech-act types such as command, request and permission offers a better fit [5, pp. 10, 23]. I propose adding a class for direct causation of an animate causee and a flag for non-volitional uses, including o’g’irlatgan ‘had (it) stolen’.

Two parts of the scheme contributed little in this pilot. C5 was never assigned: the only sentence with a stacked causative had no intermediary. Severity ratings also followed the codes without exception. Every C1, C2 and C7 error was major, while every C3 and C4 error was minor. These features could be removed or tested again on a larger set.

The two coding layers served a clearer purpose. Systems supplied pronoun gender in 55 of the 112 outputs. Treating each choice as an error would have flagged almost half the data for information that the Uzbek source does not provide.

In a blind re-coding of 24 randomly selected outputs one week later, three codes changed. All changes were between C3 and C0, the same boundary that produced the largest group of ties.

The next step is to involve a second annotator and measure agreement with a chance-corrected coefficient such as Cohen's kappa.

4. Conclusion

The procedure separates errors in causative meaning from errors elsewhere in a translation. Its pilot application also showed what needs to change before it is used more widely: uncertain class labels make C3 difficult to assign, the semantic classes miss several common uses, and some codes and severity ratings add little on this set. References drafted after the outputs have been seen introduce a further risk of bias. Revising these points would prepare the scheme for larger Uzbek datasets. Extending it to other Turkic languages, or to English–Uzbek translation, would require further testing and, for the reverse direction, new codes.

References

1. Alimova, M. Kh., & Jilina, O. Yu. (2019). Expression of causative meaning in the structures of the Uzbek and English languages. *Central Asian Journal of Education*, 4(1). <https://uzjournals.edu.uz/cjedu/vol4/iss1/1/>
2. Azimova, S. O. (2026). Post-editing (MTPE) problems and their linguistic solutions in rendering Uzbek national realia into English through artificial intelligence and neural machine translation. *BG Pulse Journal*, 2(8), 702–712.
3. Ibrohimov, A. (2013). *U – siz emas...* [Story]. Ziyouz. <https://ziyouz.uz/uzbek-nasri/abduqahhor-ibrohimov/abdu-a-or-ibro-imov-u-siz-emas-ik/>
4. Karimjonova, S. R. (2023). The meaning of causation in linguocognitive aspect. *International Journal of Information Sciences and Techniques*, 13(1), 1–5. <https://doi.org/10.5121/ijist.2023.13101>
5. Karimjonova, Sh. R. (2024). *O'zbek va ingliz tillarida kauzativ fe'llar tadqiqi* [Abstract of PhD dissertation]. Fergana State University.
6. Obidjon, A. (2019). *Mo'tivoymisan, Mittivoymisan?* [Novella]. Ziyouz. <https://ziyouz.uz/uzbek-nasri/anvar-obidjon/anvar-obidjon-mottivoymisan-mittivoymisan-qissa/>
7. Olimova, S. M. (2026). Google Translate-ning eng ko'p yo'l qo'yadigan lingvistik xatolari [The most frequent linguistic errors of Google Translate]. *International Conference "Pedagogical Reforms and Their Solutions"*, 15(2), 166–168.
8. Rahmon, G'. (2021). *Uch oyoq bo'ri izidan* [Story]. Ziyouz. <https://ziyouz.uz/uzbek-nasri/gozi-rahmon/gozi-rahmon-uch-oyoq-bori-izidan-hikoya/>
9. Rasulov, N. A. (2026). Structural and semantic patterns of causative verbs in fiction: Cross-linguistic evidence from English and Uzbek. *Arabian Journal of Science*, 3(6), 46–48.
10. Sultonova, S. B. (2026). The impact of artificial intelligence on the quality of English–Uzbek translation. *International Journal of Technology and Academic Research*, 6(7), 326–330.