

# **ERROR ANALYSIS FOR MACHINE- TRANSLATED TEXTS FROM ENGLISH INTO KAZAKH**

---

**Dinara Akmurzina**

**Master's Student**

**European Master in Technology for Translation  
and Interpreting**

**(New Bulgarian University, Ghent University)**

# European Master in Technology for Translation and Interpreting

- Year 1 University: New Bulgarian University (NBU), 2021-2022 academic year
- Year 2 University: Ghent University (UGent), 2022-2023 academic year
- Master thesis topic: Error Analysis for Machine Translation into Kazakh as a Low-resource Language - ongoing
- Supervised by: Prof. A.S. Doğruöz (UGent), Prof. M. Stambolieva (NBU)



# Plan

1. Research question and hypothesis
2. Background knowledge on error analysis
3. Methodology and examples of error analysis
4. Conclusion

# Research question and hypothesis

*What are the MT errors for translations from English into Kazakh?*

The majority of errors in MT output could be of a linguistic nature, due to the linguistic differences between English and Kazakh, as well as the lack of language resources for Kazakh.

# Differences between Kazakh and English

Kazakh	English
Agglutinative	Analytical language
SOV order of words	SVO order of words
No articles	Definite and indefinite articles
Morphologically rich	Less complex morphology
Genderless language	Natural gender language

A *translation error* is defined as some non-equivalence between the source text (ST) and the target text (TT) or the TT's inadequacy.

(Hansen, 2010)

*A translation error* is defined as some non-equivalence between the source text (ST) and the target text (TT) or the TT's inadequacy.

(Hansen, 2010)

*A translation error* is a 'semantic component not shared by source text and target text'.

(Koponen, 2010)

A *translation error* is defined as some non-equivalence between the source text (ST) and the target text (TT) or the TT's inadequacy.

(Hansen, 2010)

A *translation error* is a 'semantic component not shared by source text and target text'.

(Koponen, 2010)

A *translation error* is some non-equivalence between the ST and TT and/or linguistically bad-formed TT that causes difficulties of understanding.



# A type of error identified: A missing word

Example (1) is taken from the subtitles for the American film *Home Alone*.

English Subtitle: *No, we put the dog in the kennel.*

## Expert Translation:

Kazakh:	<b>Joq,</b>	<b>biz</b>	<b>it-ti</b>	<b>köşettik-ke</b>	<b>ornalastyr-diq</b>
Gloss:	No,	we	dog-ACC	kennel-DAT	put-1PL

## Open Subtitles Translation:

Kazakh:	<b>Janwar-lar</b>	<b>pana-sı-na</b>	<b>qaldır-ayıq</b>
Gloss:	Animal-PL	kennel-3SG.POSS-DAT	leave-PRES.PTCP

## Types of errors identified:

### A morphological error, a foreign word

Example (1) is taken from the subtitles for the American film *Home Alone*.

English Subtitle: *No, we put the dog in the kennel.*

#### Expert Translation:

Kazakh: Joq, biz it-ti köşettik-ke ornalastyr-diq

Gloss: No, we dog-ACC kennel-DAT put-1PL

#### Google Translate Output:

Kazakh: Joq, biz it-ter-di pïtomnik-ke ornalastır-dıq

Gloss: No, we dog-PL-ACC kennel-DAT put-1PL

## Types of errors identified:

A missing word, an extra word, a lexical error

Example (2) is taken from the subtitles for the American film *Home Alone*.

**English Subtitle:** *The **dope** was **whining** about a suitcase*

### Expert Translation:

Kazakh: <b>Aqımaq</b>	<b>şabadan-ı</b>	<b>jöninde</b>	<b>qıñsıla-dı.</b>
Gloss: The dope	suitcase-3SG.POSS	about	whine-PST-3SG

### Open Subtitles Translation:

Kazakh: <b>Endi</b>	<b>söz</b>	<b>şabadan</b>	<b>jaylı</b>	<b>bol-atın.</b>
Gloss: Now	word	suitcase	about	be-PRES.PTCP

# Types of errors identified:

## A lexical error, foreign words

Example (2) is taken from the subtitles for the American film *Home Alone*.

English Subtitle: *The **dope** was **whining** about a **suitcase***

### Expert Translation:

Kazakh: <b>Aqımaq</b>	<b>şabadan-ı</b>	<b>jöninde</b>	<b>qıñsıla-dı.</b>
Gloss: The dope	suitcase-3SG.POSS	about	whine-PST-3SG

### Google Translate Output:

Kazakh: <b>Bul</b>	<b>prıdwrok</b>	<b>skwlil</b>	<b>arqasında</b>	<b>çemodan.</b>
Gloss: This	dope	whine	thanks to	suitcase

# Conclusion

1. Google Translate sometimes produced incorrect output when translating subtitles from English into Kazakh;
2. The most frequent errors were of a lexical and morphological nature;
3. An additional error type was the use of foreign words.

# Conclusion

1. The subtitles from the Open Subtitles database also had some errors.
2. The analysis identified some accuracy errors in them:
  1. adding a word that was not in the source segment;
  2. omitting a word that was included in the source segment, but not translated.

# References

- Dotton, Z. and Wagner, J.D. (2018) A Grammar of Kazakh. PhD thesis. Duke University Slavic and East European Language Resource Centre. Available at: <https://slaviccenters.duke.edu/sites/slaviccenters.duke.edu/files/file-attachments/kazakh-grammar.pdf>.
- Hansen, G. (2010) Translation ‘errors’, in Y. Gambier and L. V. Doorslaer (Eds.) Handbook of translation studies. (“Translation "Errors" — CBS Research Portal”) Amsterdam: John Benjamins, pp. 385–388.
- James, C. (2013) Errors in language learning and use: Exploring error analysis, New York: Routledge.
- Koponen, M. (2010) ‘Assessing Machine Translation Quality with Error Analysis’, MikaEL: Electronic proceedings of the KäTu symposium on translation and interpreting studies, 4. p.14 Suomen kääntäjien ja tulkkien liitto. Available at: [https://sktl-fi.directo.fi/@Bin/40701/Koponen\\_MikaEL2010.pdf](https://sktl-fi.directo.fi/@Bin/40701/Koponen_MikaEL2010.pdf). (Accessed: 17 September 2022).
- Sundetova, A., Forcada, M. and Tyers, F. (2015) ‘A free/open-source machine translation system from English to Kazakh’, Proceedings of the International Conference “Turkic Languages Processing” (TurkLang 2015), Kazan, Tatarstan, Russia, 17-19 September. pp. 78.
- **Web resources**
- The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses. Available at: <https://www.eva.mpg.de/lingua/pdf/Glossing-Rules.pdf> (Accessed: 16 October 2022)