

MACHINE TRANSLATION OF LITERARY TEXTS

A comparative study on the translation of two different literary genres by three neural machine translation systems

*Ana Isabel Cespedosa Vázquez
Ruslan Mitkov*

01. Rationale

02. Aim

03. Methodology



04. Results and discussion

05. Limitations

06. Conclusion

01

Rationale

01 Rationale

PRESENT

Recent advances in Artificial Intelligence and MT

Negative perception when applied to literary texts



RESEARCH

Neural MT systems implementation

Lack of studies on literary MT

02

Aim

02 Aims



Compare the performance of 3 recent NMT systems on literary texts



Identify whether MT performance is influenced by genre and time period

03

Methodology

03 Methodology

a. Dataset: texts and translations

Pride and prejudice (Jane Austen)

1984 (George Orwell)

I Felt a Funeral, in my Brain (Emily Dickinson)

Siren Song (Margaret Atwood)

b. Selection of Neural Machine Translation systems:

→ DeepL

→ Systran

→ Yandex

03 Methodology

c. BLEU metric implementation:

- | | |
|--|--|
| <ul style="list-style-type: none">❑ Installing Python❑ Downloading necessary complements:<ul style="list-style-type: none">→ NLTK library→ Corpus function | <ul style="list-style-type: none">❑ Creating the code to be run in Python❑ Defining the required commands |
|--|--|

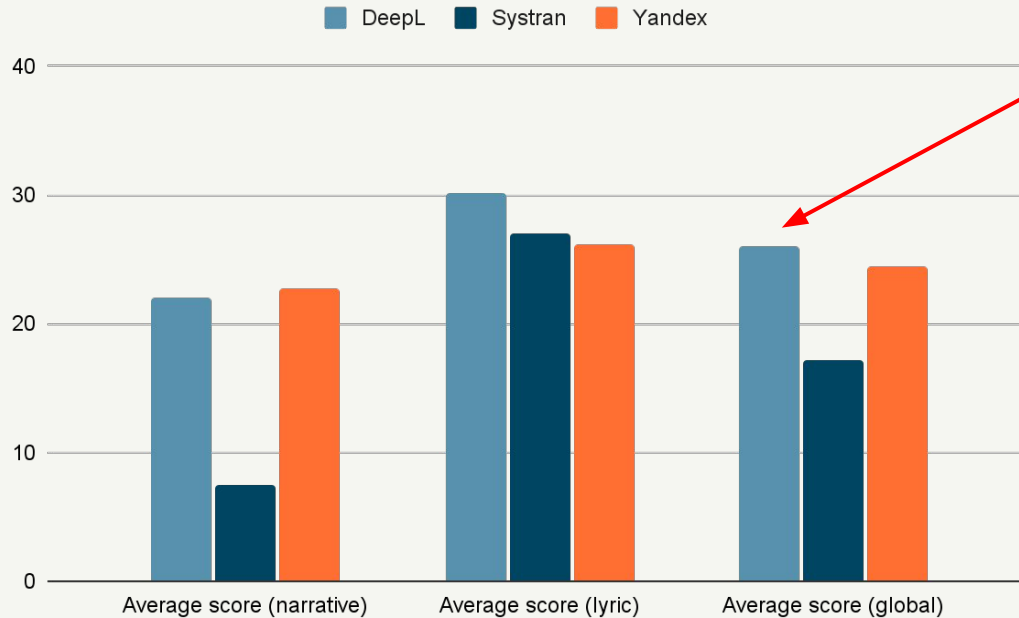
```
import nltk, io
from nltk.translate.bleu_score import corpus_bleu, sentence_bleu, SmoothingFunction
from nltk.tokenize import sent_tokenize, word_tokenize
```

Image: Excerpt from the code

04

Results and
discussion

BLEU results



DeepL had the best performance

The MT systems had a better performance for lyric rather than narrative genre

Average score for each genre and global score for each MT system.

BLEU results

Work	1984	Pride and prejudice	I Felt a Funeral, in my Brain	Siren Song
Average score <i>(for 3 MT systems)</i>	22.08	19.13	17.16	38.37

Average score according to the time of the work for each MT system.

BLEU results

	DeepL	Systran	Yandex
<i>Average score for classical literature</i>	15.07	14.55	15.25
<i>Average score for contemporary literature</i>	37.05	19.91	33.72

Average score according to the time of the work for each MT system.

05

Limitations

Limitations of the study

DATA SIZE

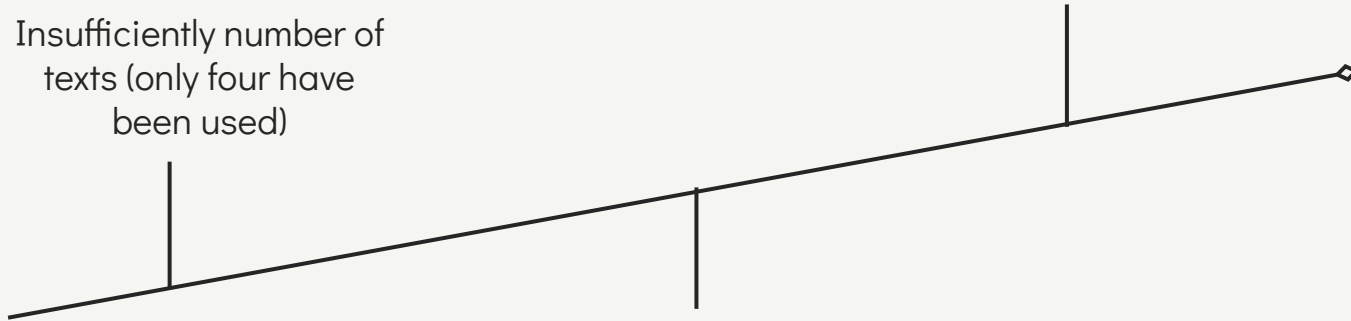
Insufficiently number of texts (only four have been used)

BLEU LIMITATIONS

The system compares between MT and HT rather than analysing context or meaning

CORPUS REPRESENTATIVENESS

A greater range of linguistic features should be taken into consideration



06

Conclusionu

Conclusion

According to BLEU:
The MT system DeepL had the
best performance

Genre and time are related
(and should be taken into
consideration) when
translating a text

Further research on MT
applied to literary texts
should be carried out

“

It is not the nature of translation technology itself that should be criticized but the discourse surrounding it and the lack of inclusion of all interested parties in the innovation process.

Paola Ruffo

Thank you very much for
your time and attention!

Ana Isabel Cespedosa Vázquez
anabelcespedosa@gmail.com

Ruslan Mitkov
r.mitkov@lancaster.ac.uk