



The influence of a machine translation system on sentiment levels

Jaroslav Reichel and Ľubomír Benko
Constantine the Philosopher University in Nitra



Aim of the research

- Research problem:
 - The aim of the research was to verify whether the sentiment of the text remains at the same level in machine translation in the two most widely used translation systems.
 - The validation was part of a larger research aimed at identifying the level of sentiment in machine translation from different translators.



Experimental setup

- Research question
 - Is there a significant difference between the translation systems Google Translate and DeepL in the accuracy of identifying sentiment scores compared to human texts?
- H0
 - There is no statistically significant difference between the correlation of sentiment level in human text and in Google Translate machine translation compared to the correlation of sentiment level in human text and in DeepL machine translation.



Data preparation

1. Source corpus preparation
 1. Alignment of Slovak and English subtitles into a coherent parallel corpus.
 2. Removal of erroneous, inconsistent, repetitive, or unnecessary records.
 3. Segmentation - Merging sentences that have been split to multiple subtitles back into a single segment.
2. Generating a machine translation for each of the subtitles using Google Translate and DeepL machine translation systems.
3. Identification of keywords and their sentiment using IBM Watson NLU service.
4. Transforming the sentiment of the keywords into a coherent dataset of sentiment scores of each segment for the three sets:
 1. Human text (EN),
 2. Machine translation from Google Translate (GT),
 3. Machine translation from DeepL (DL).



Dataset

- The dataset contained 11601 subtitles from 10 movies (SK, EN),
- After cleaning 8551 segments (pairs),
- Assigning machine translations,
 - Google Translate,
 - DeepL.



Machine translation from GT and DL

id	Text_sk	Text_en	Text_gt	Text_dl
0	Blake.	Blake.	Blake .	Blake.
5	Zabalili nám jedlo?	Did they feed us?	Did they pack our food?	Did they pack us food?
6	Nie. Len poštu.	No. Just mail.	Not. Just mail.	No. Just mail.
7	Je čas na čaj!	Time for some tea! Tea's up!	It's tea time!	It's tea time!
8	Myrtle bude mať šteniatka.	Myrtle's having puppies.	Myrtle will have puppies.	Myrtle will have puppies.
10	Som hrozne hladný. Ty nie?	Oh, I'm bloody starving. Aren't you?	I'm terribly hungry. You do not?	I'm terribly hungry. Aren't you?



Sentiment analysis

- IBM Natural Language Understanding for sentiment analysis,
- Sentiment assigned to each segment (if IBM NLU has identified the keywords)
 - Text_en – 8419 segments,
 - Text_gt – 6886 segments,
 - Text_dl – 6923 segments,
- Combined - 4076 segments.



Sentiment analysis output (text_en)

id	keyword	text_en	sentiment_score
5	food	Did they feed us?	0
6	mail	No. Just mail.	0
7	tea time	Time for some tea! Tea's up!	0.842084
8	Myrtle	Myrtle's having puppies.	0.849348
8	puppies	Myrtle's having puppies.	0.849348
12	priesthood	It was the only reason I decided against the priesthood.	-0.839655



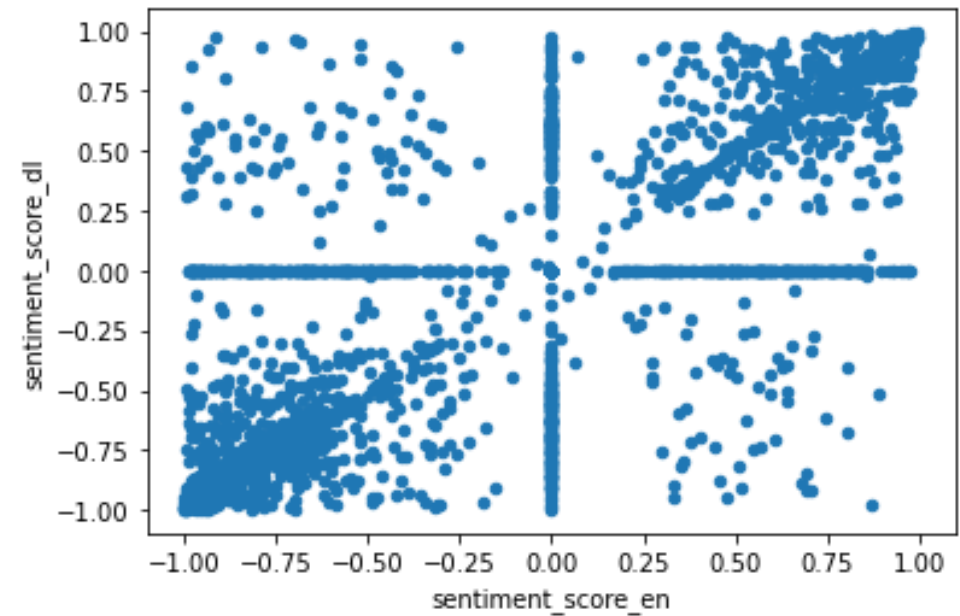
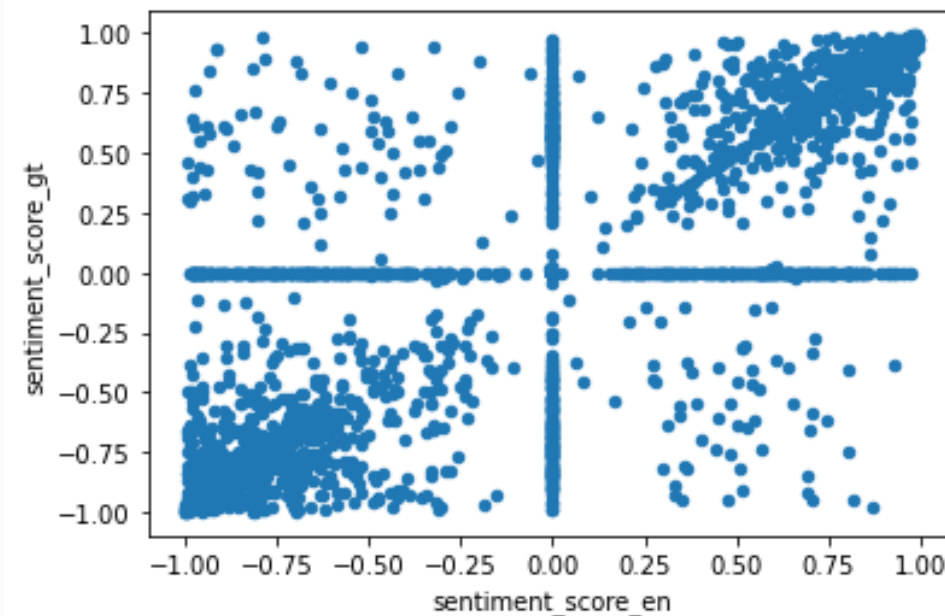
Data analysis

1. Verification of the level of correlation of the identified sentiment of the machine translations (GT, DL) with the reference sentiment from the human text (EN).
2. Comparison of results from Google Translate and DeepL.



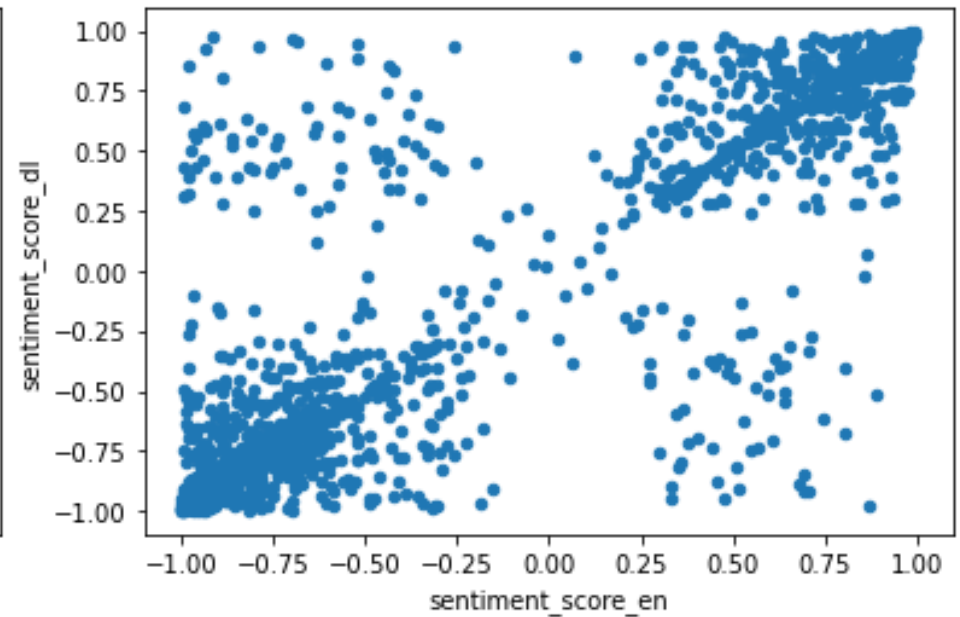
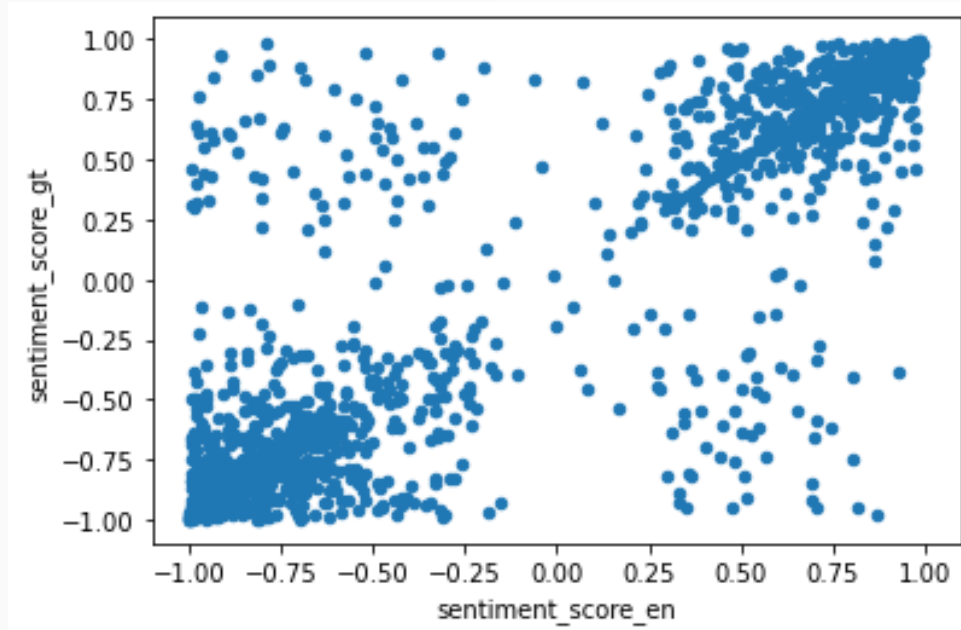
Results

- Results of correlation analysis:
 - EN/GT: $r(3768) = 0.73$, $p < 0.01$,
 - EN/DL: $r(3768) = 0.74$, $p < 0.01$.



Results

- Results of correlation analysis:
 - EN/GT: $r(1497) = 0.86$, $p < 0.01$,
 - EN/DL: $r(1539) = 0.86$, $p < 0.01$.



Conclusion

- H0 is not reject
 - There is no significant difference in the level of sentiment preservation between Google Translate and DeepL.
- Therefore, when analyzing sentiment preservation, it is possible to consider one of the tools in machine translation and generalize the results to machine translation systems.
 - between English and Slovak.
- Further research
 - analyze the characteristics of specific cases (segments) where opposite sentiments were identified in individual translations
 - e.g., GT positive and DL negative.





Thank you for your attention!

