On the Impact of Machine Translation on the Quality of the final Review outputs

Brita Dorer, GESIS
Diana Zavala-Rojas, UPF
Veronika Keck & Dorothée Behr, GESIS
Danielly Sorato, UPF

ESRA conference 2021
Session: Investigating effects of Machine Translation and Post-editing on TRAPD
July 9th, 2021, virtual conference

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

Overview

- Research questions
- Coding process
- Error scheme
- Data preparation
- Analytical approach
- Preliminary results
- Conclusions / Limitations

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Research questions

- Is the translation quality (at textual level) of the final review output affected by introducing Machine Translation (MT) & Post-Editing (PE)?
  If so, is it to the better or to the worse?
- Are the differences conditional to group effects, a group defined as the combination of language and type of Post-Editing?

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Coding process

- For both languages GER + RUS identical
- Each 2 experienced and qualified linguists / questionnaire translators coded separately
- Then the 2 persons agreed on coding in “coding harmonization meeting”
- 3rd person in the coding harmonization meeting = adjudicator, involvement in case of uncertainties
- Error subcategory + Severity level
- Additional code: “Fixed source”

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

Error scheme

- Developed based on Multidimensional Quality Metrics (DQF-MQM), developed to assess translation quality
- Adapted to questionnaire translation by the project team
- First level Error Categories:
  Accuracy – Fluency - Survey-specific terminology/phrases and features – Style – Locale convention – Verity – Other
- Second level Error Subcategories, examples: *Omission – Register – Mistranslation of survey-specific terminology/phrases*

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Coding process: Error scheme + severity levels

Severity levels:

- **Major:** Major level of severity means that the translation completely changes the meaning, likely misleads the respondent, or provides incorrect, missing and/or contradictory information.
- **Minor:** Minor errors may affect the respondent's comprehension of translated text and increase the time required to read and to understand the translation.
- **Neutral:** Neutral errors include those that might make the translation a bit harder to understand, but ultimately do not stop the respondent from overall understanding and using the translation in terms of the measurement goal.

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Data preparation

- Translations resulting from the Review meetings
- Baseline (2 human translators)
  Experimental condition 1 (human translator + Full Post-editing)
  Experimental condition 2 (human translator + Light Post-editing)
- Both languages separately (GER + RUS)
- Spelling and Punctuation errors taken out:
  - for Russian, quite a lot of these rather minor errors;
  - they would possibly have biased the results;
  - therefor decision to remove these;
  - this is also in line with reality, as in real-life, these would be corrected after the Review session

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Analytical approach

- Null hypothesis: H0: p1 - p2 = 0
  (There is no difference in the number of errors in the final Review output between the baseline and the experimental conditions.)
- Equation Z-test:

$$\frac{(\bar{p}_1 - \bar{p}_2) - 0}{\sqrt{\bar{p}(1-\bar{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC social sciences & humanities open cloud

Preliminary results

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Overall number of errors

- Number of errors in Baseline review GER: 30
- Number of errors in Baseline review RUS: 45

- Number of errors in Full PE review GER: 33
- Number of errors in Full PE review RUS: 30

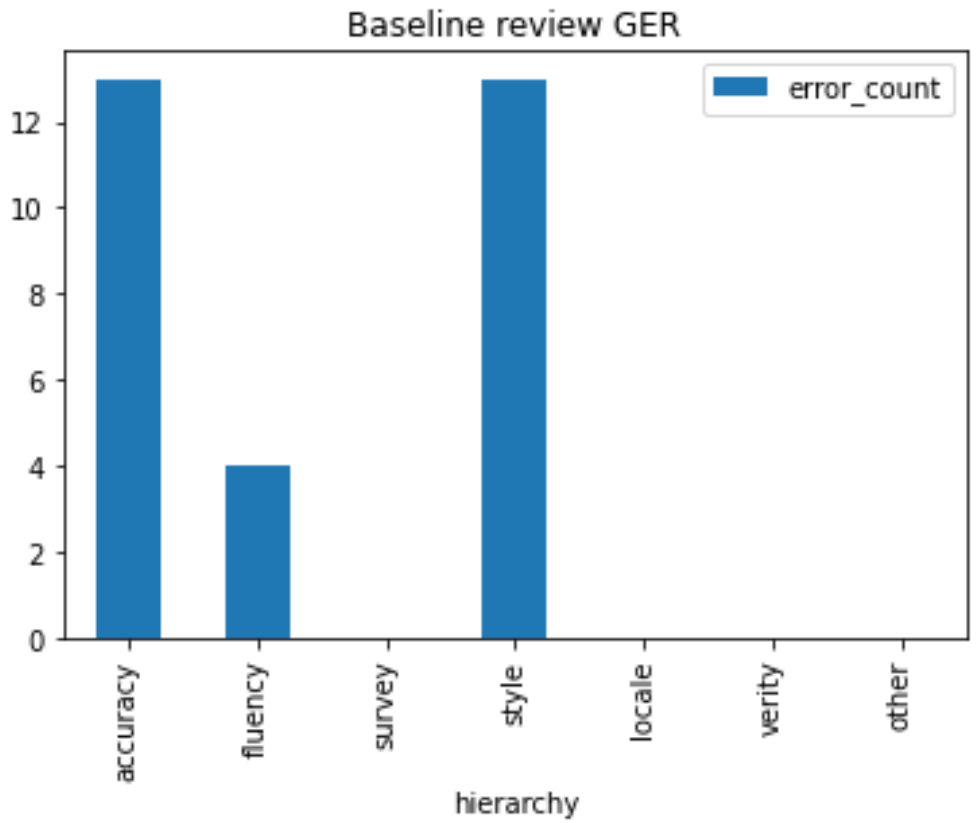
- Number of errors in Light PE review GER: 68
- Number of errors in Light PE review RUS: 39

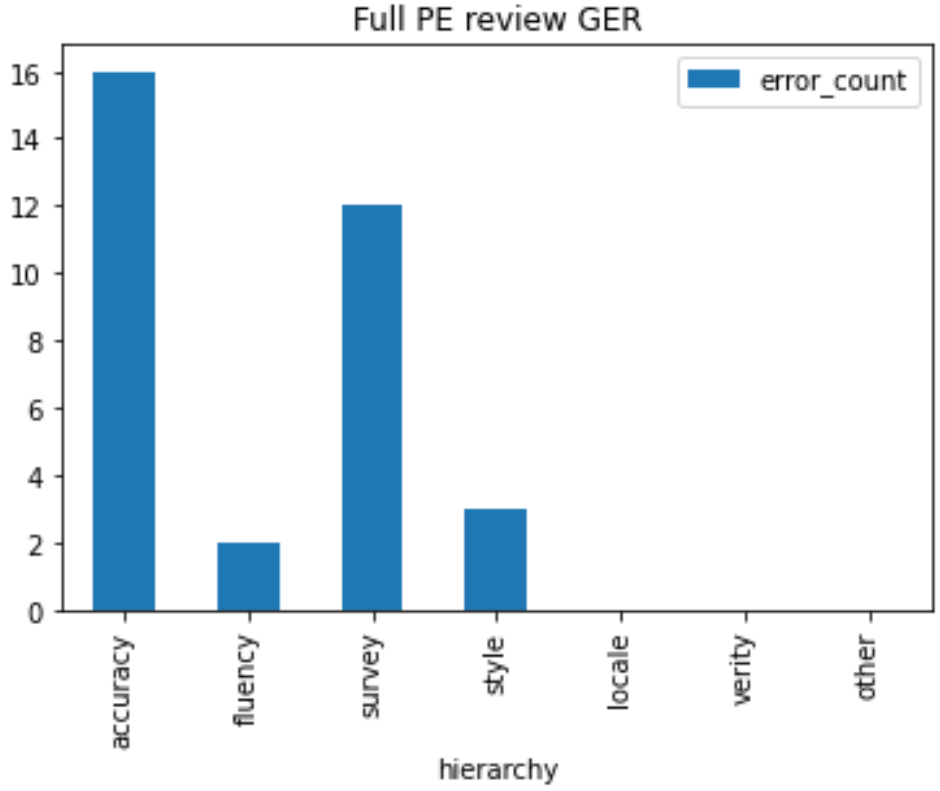
This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Example BSLN GER vs FULL PE GER

- The value of p is .68916.
- The result is not significant at p < .05.
- **NO difference**

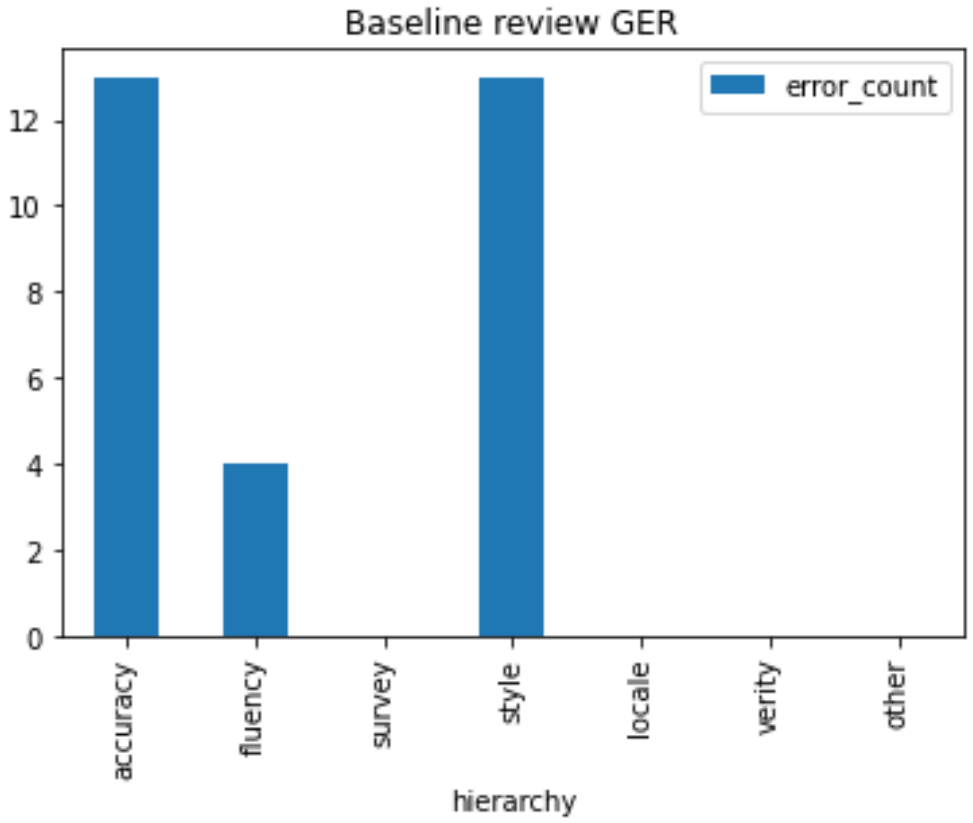


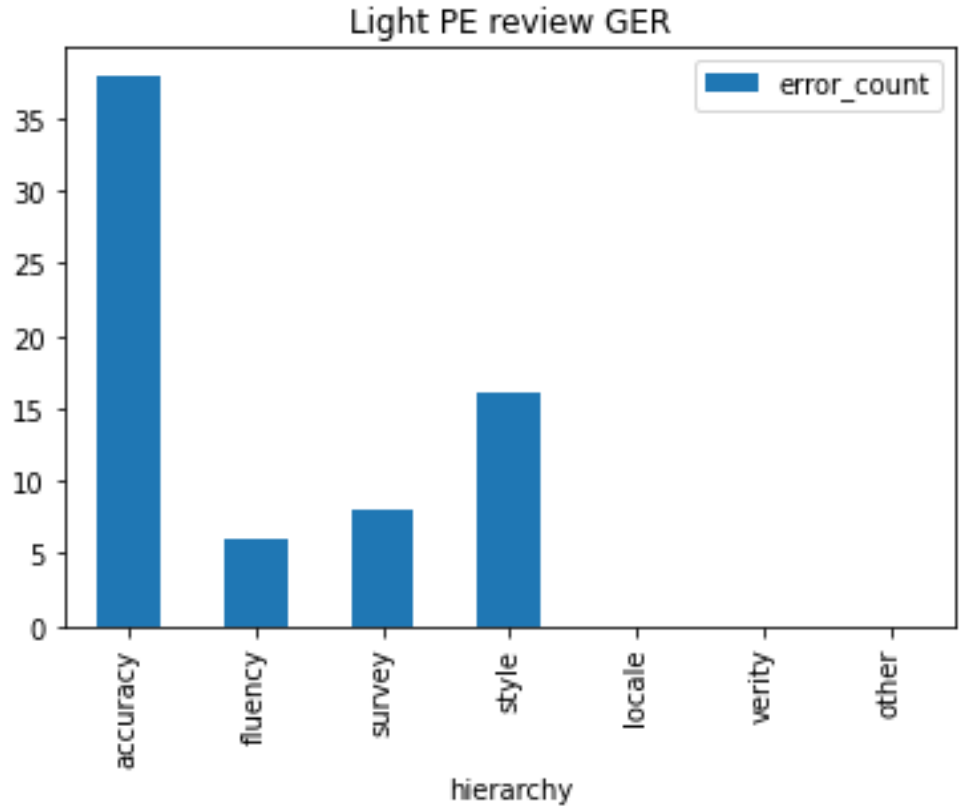


This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

Example BSLN GER vs LIGHT PE GER

- The value of p is < .00001.
- The result is significant at p < .05.
- **Difference = LIGHT has more errors**

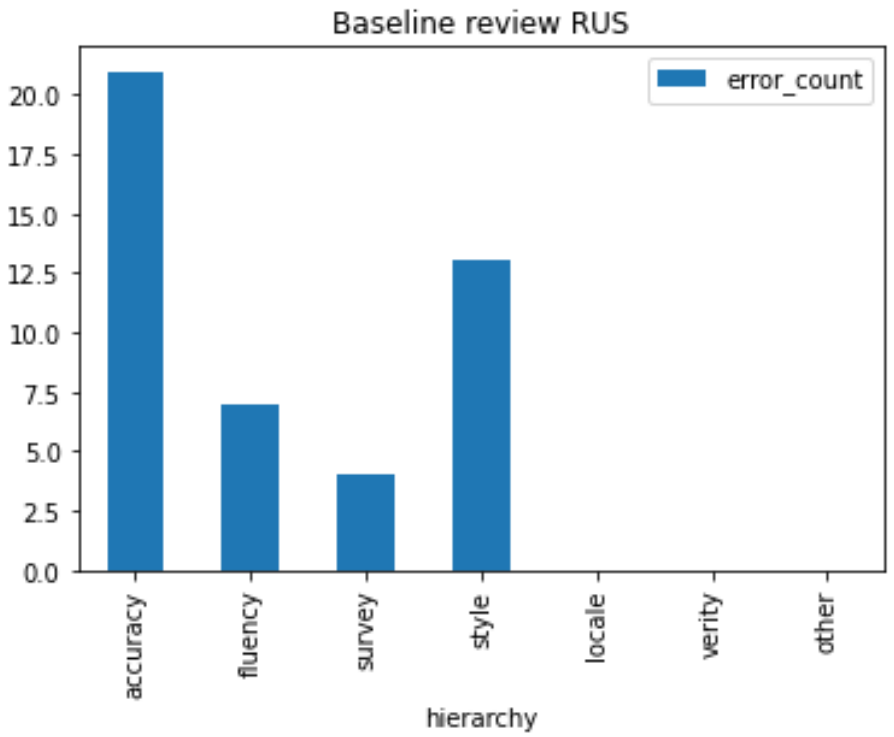


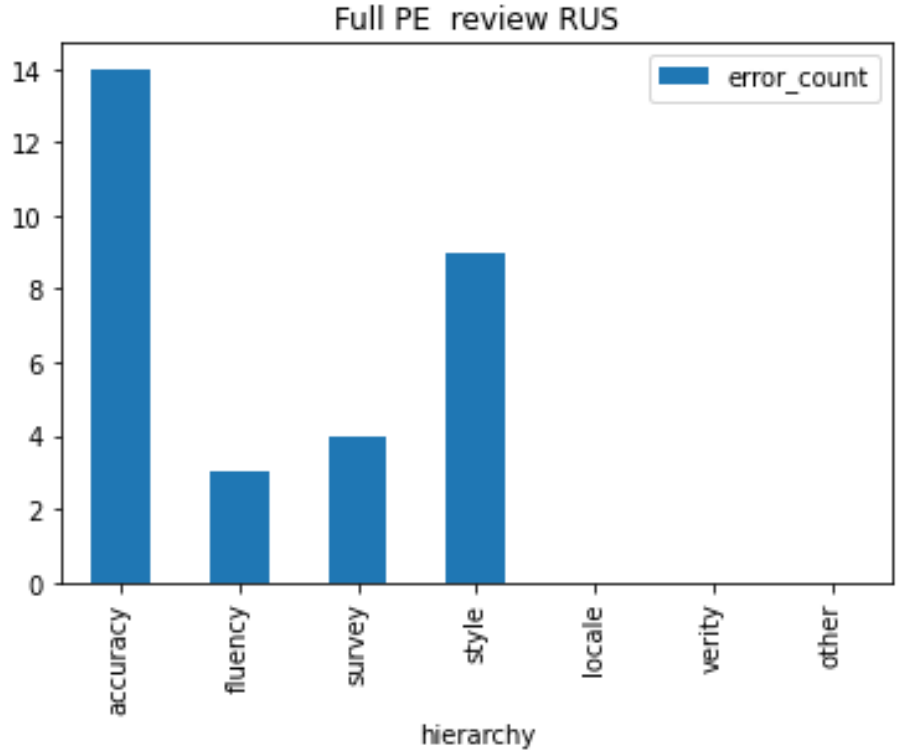


This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

Example BSLN RUS vs FULL PE RUS

- The value of p is < .00001.
- The result is significant at p < .05.
- **Difference = BASELINE has more errors**



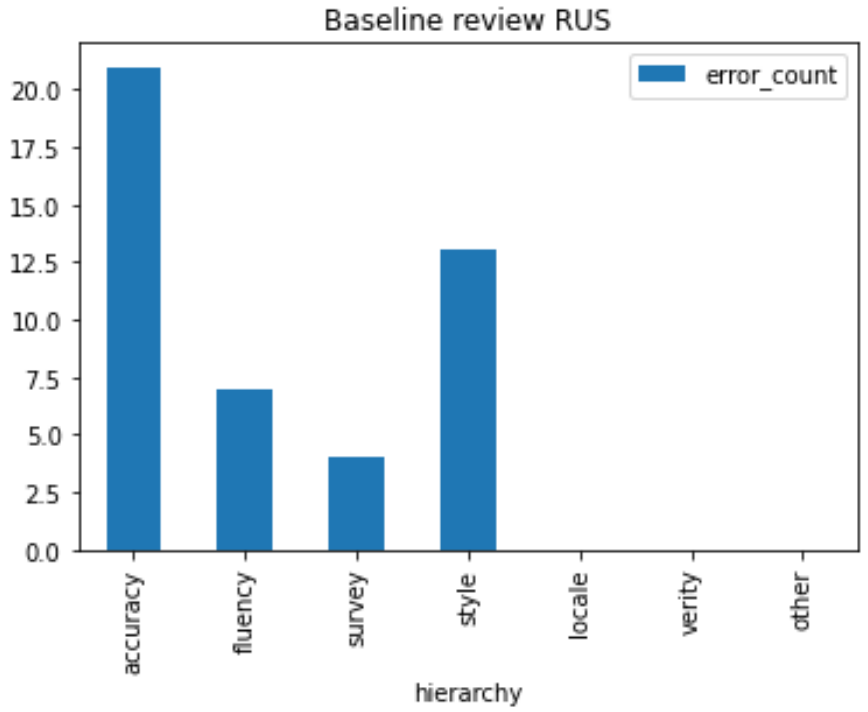


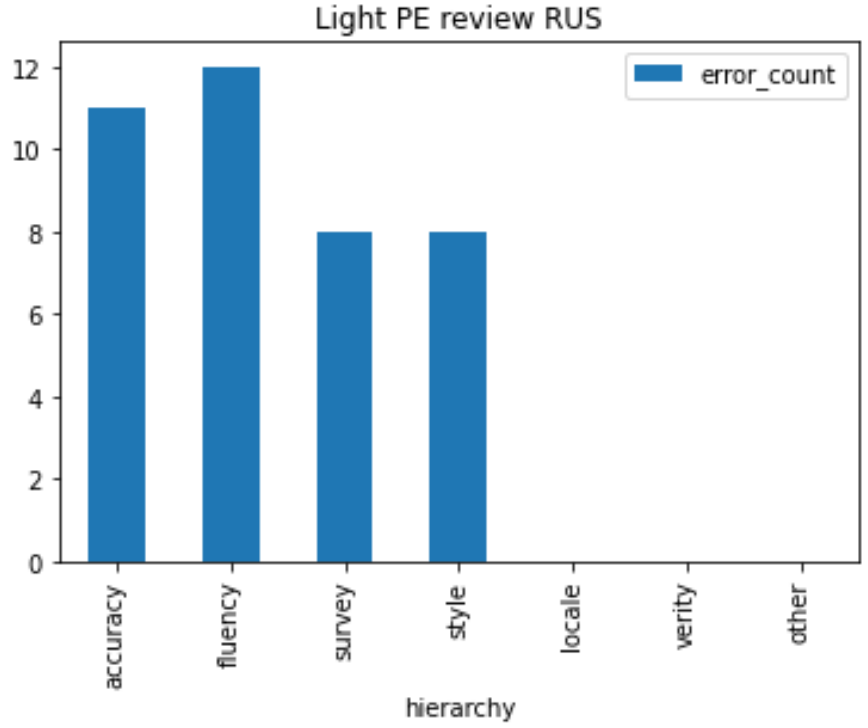
This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Example BSLN RUS vs LIGHT PE RUS

- The value of p is < .00001.
- The result is significant at p < .05.
- **Difference = BASELINE has more errors**





This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Conclusions / Limitations

- In Russian, the involvement of MT had overall a positive effect on the final quality of the Review version.
- In German, the effect of adding MT is not clear because in both Review results we had different effects.
- Details will be presented in later presentations in this ESRA session.

This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud

Thank you for your attention!

contact: brita.dorer@gesis.org

Join our community

https://www.sshopencloud.eu

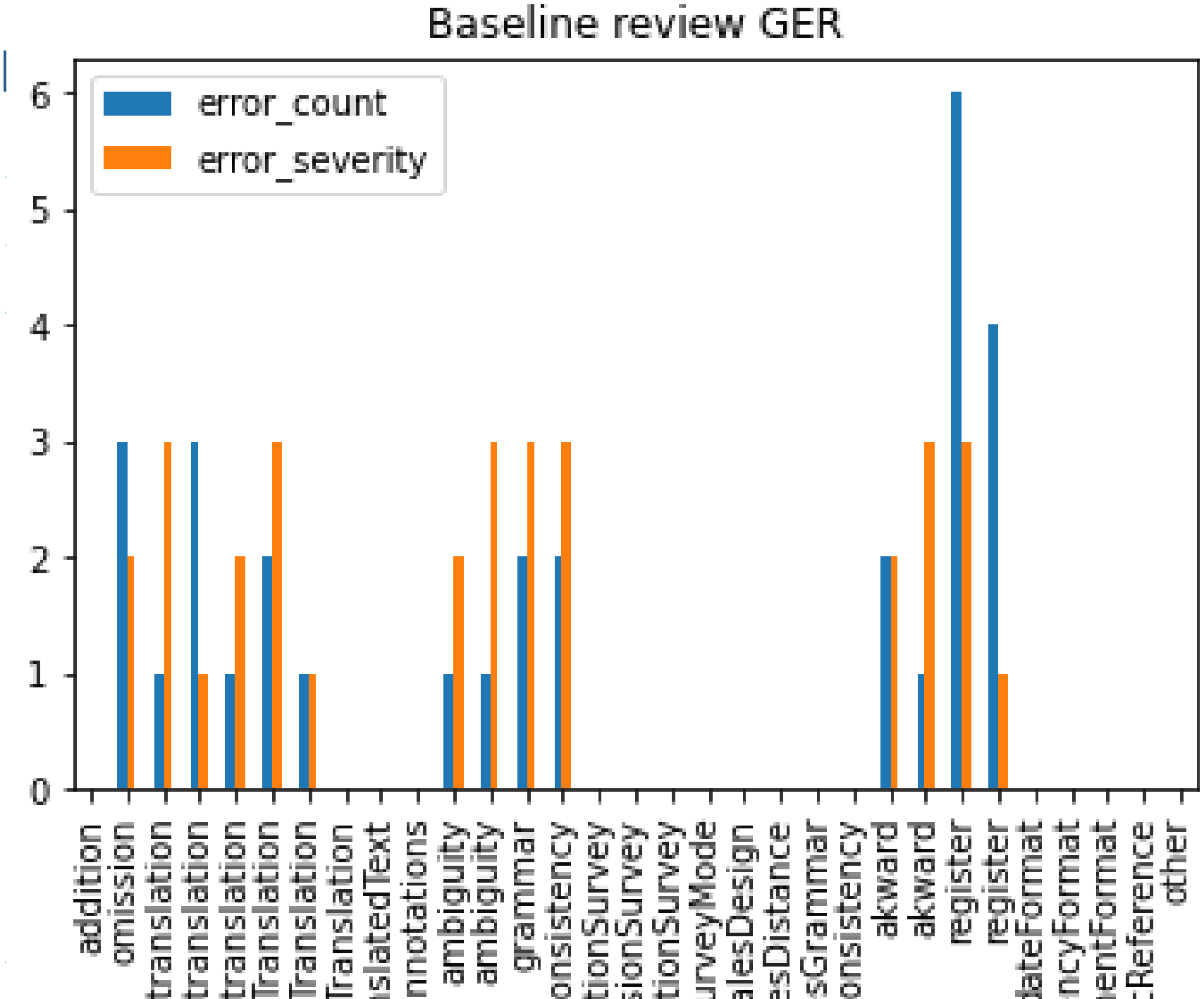
@SSHOpenCloud

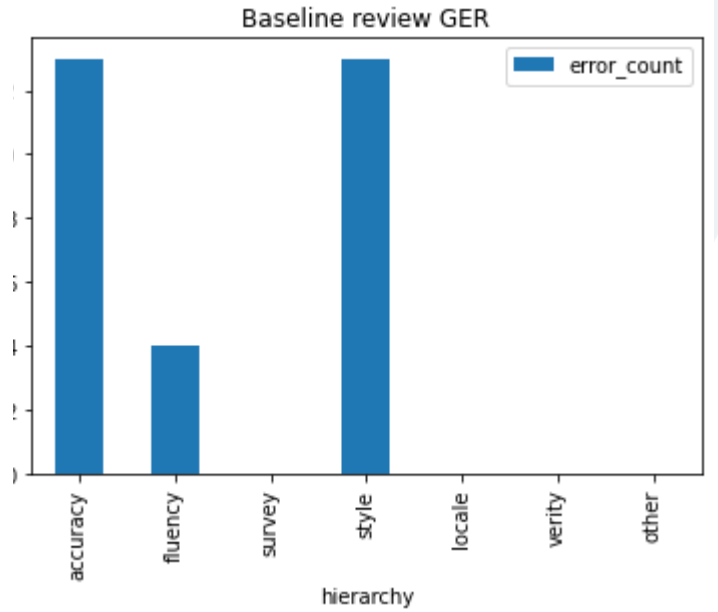
info@sshopencloud.eu

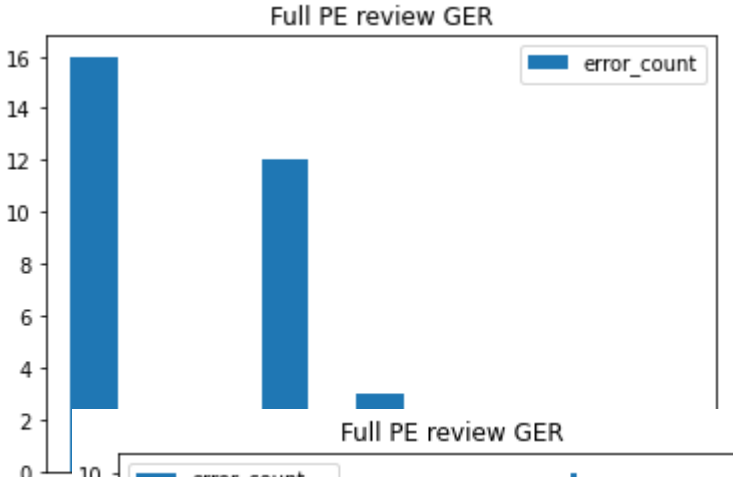
/in/sshopencloud

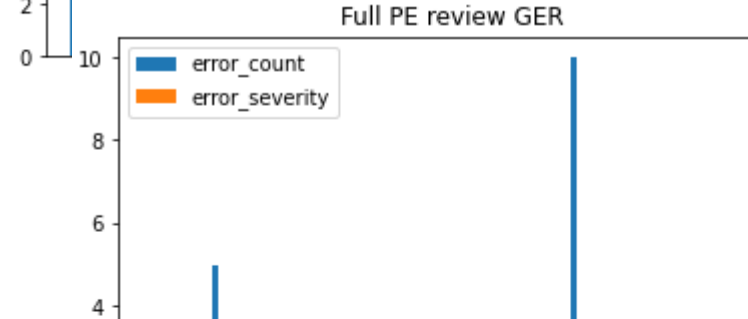
This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

SSHOC
social sciences & humanities open cloud


Baseline review GER
error_count
error_severity
0
1
2
3
4
5
6
addition
omission
translation
translation
translation
Translation
Translation
Translation
ambiguity
ambiguity
grammar
akward
akward
register
register
other

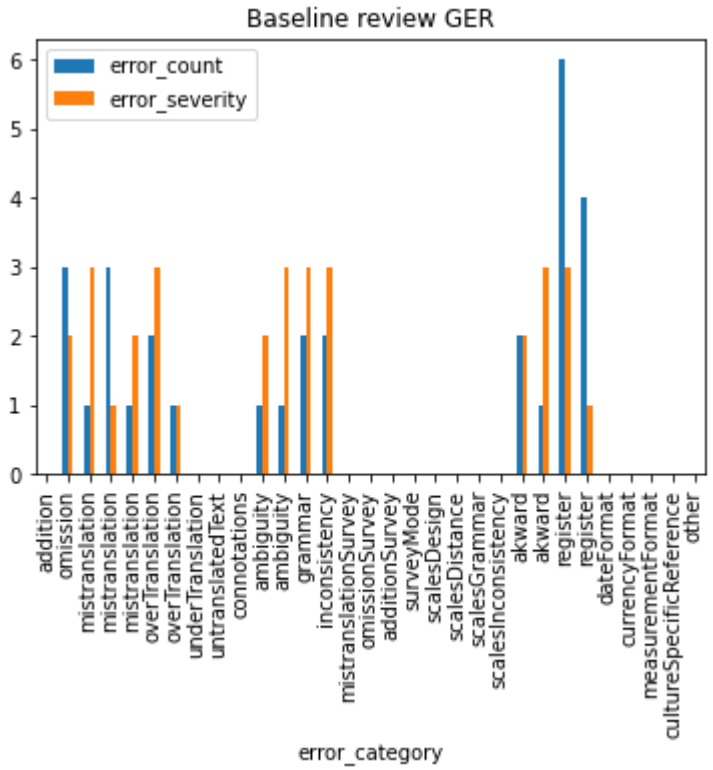

Baseline review GER
error_count
accuracy
fluency
survey
style
locale
verity
other
hierarchy

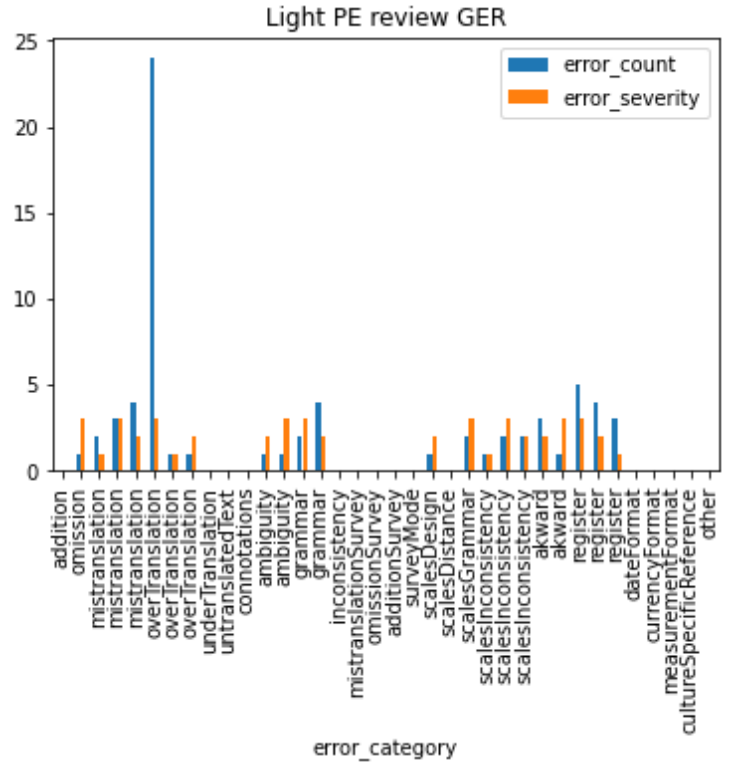

Full PE review GER
error_count
0
2
4
6
8
10
12
14
16

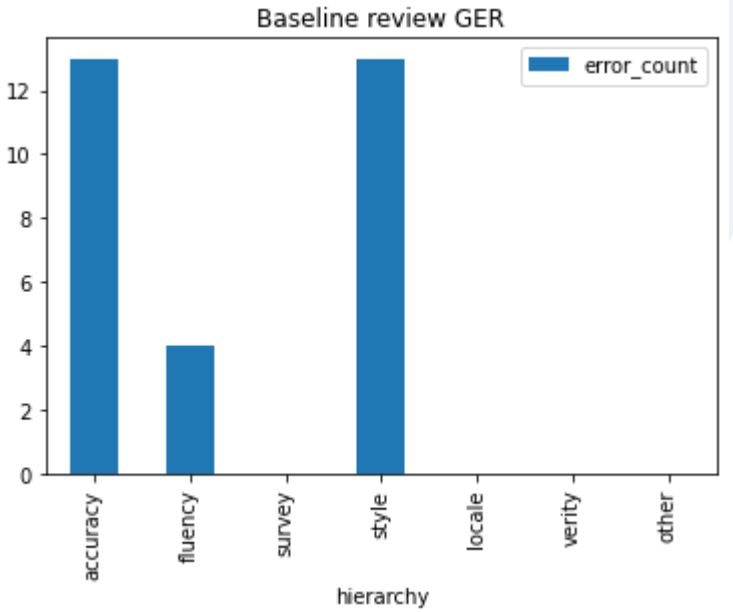

Full PE review GER
error_count
error_severity
0
4
6
8
10

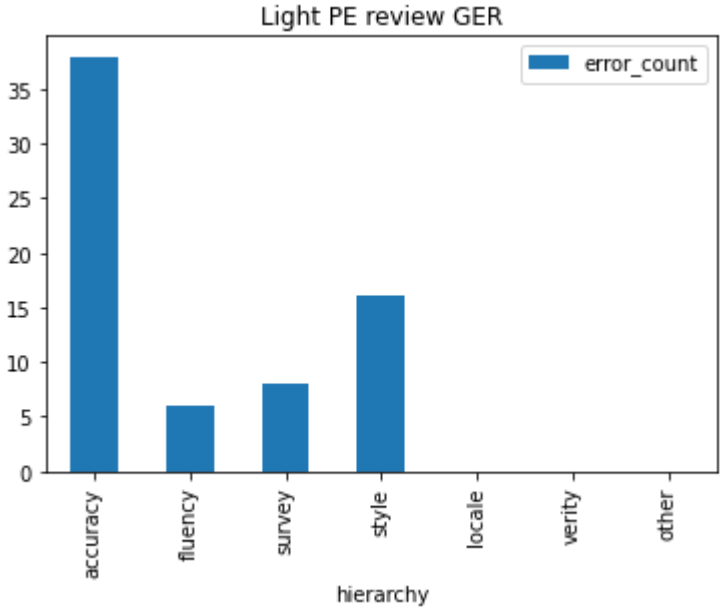
Example BSLN GER vs LIGHT PE GER

- The value of p is < .00001.
- The result is significant at p < .05.
- **Difference = LIGHT has more errors**







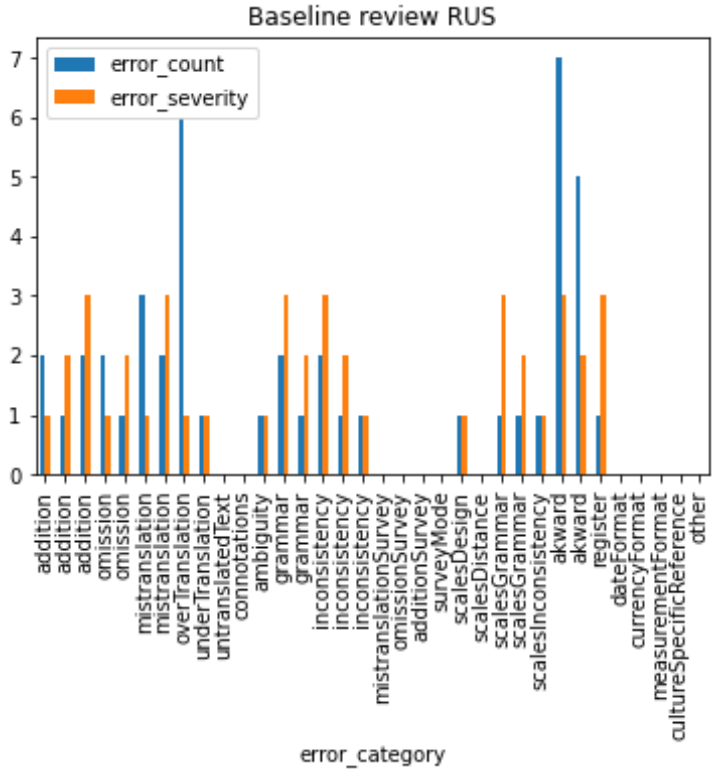


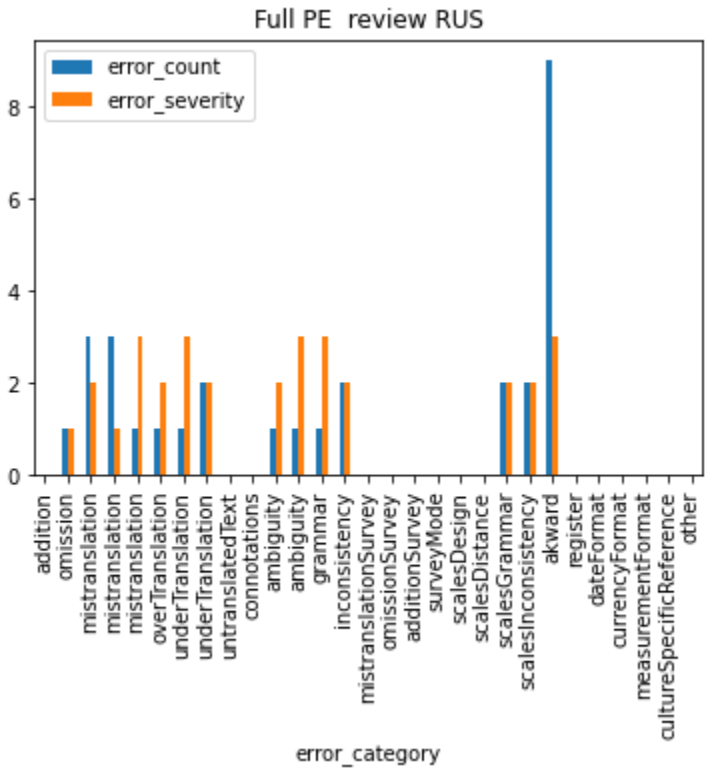
This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

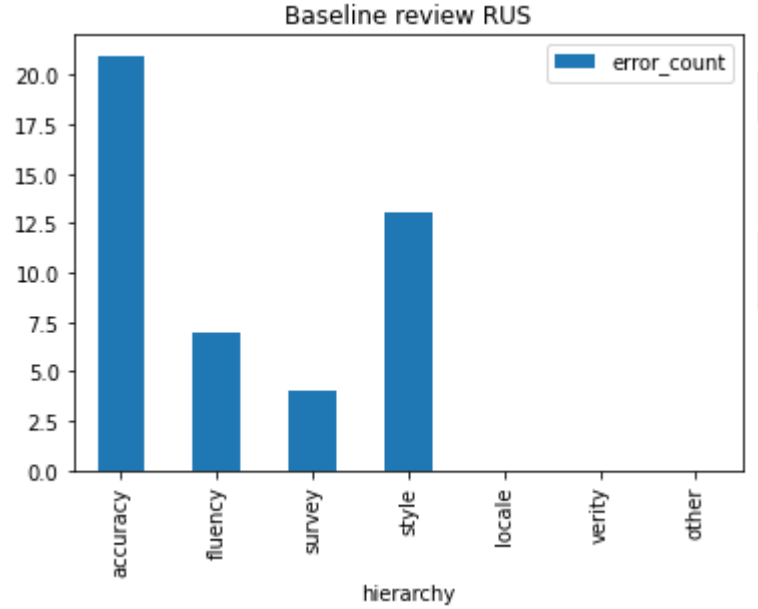
SSHOC
social sciences & humanities open cloud

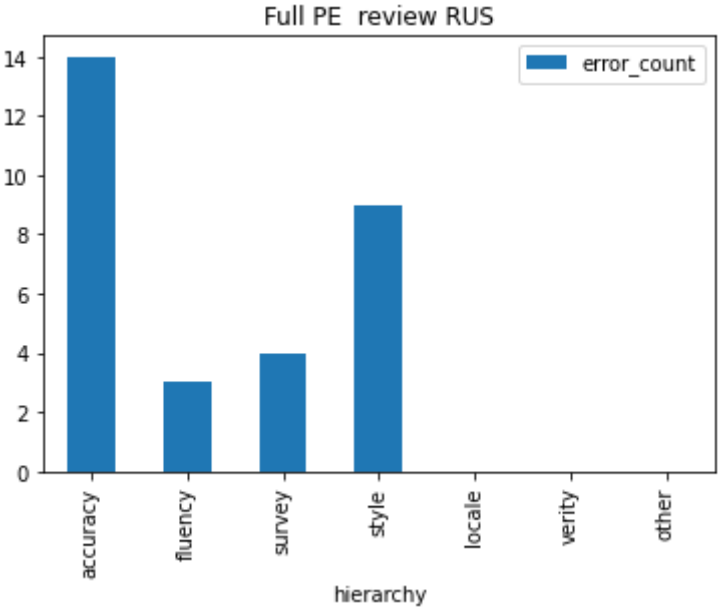
Example BSLN RUS vs FULL PE RUS

- The value of p is < .00001.
- The result is significant at p < .05.
- **Difference = BASELINE has more errors**





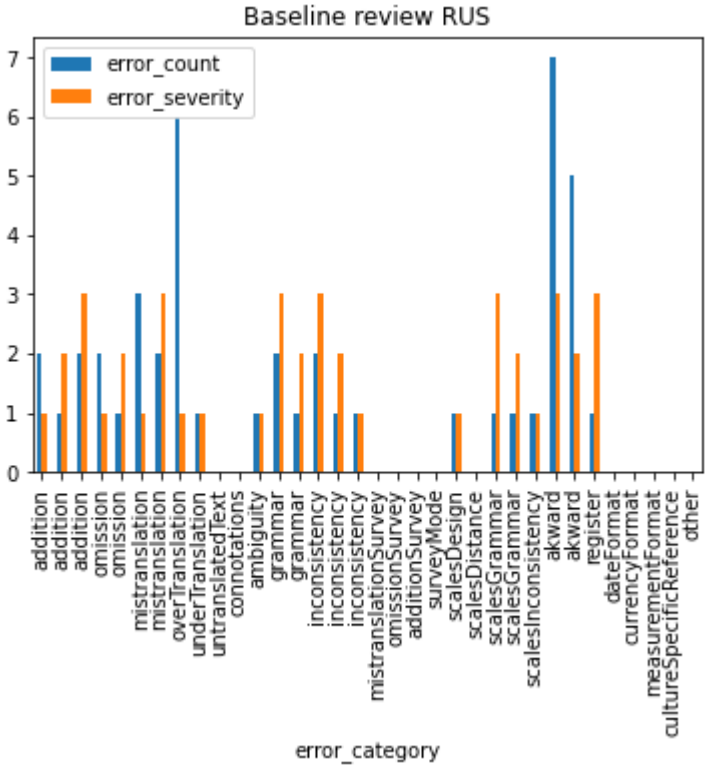


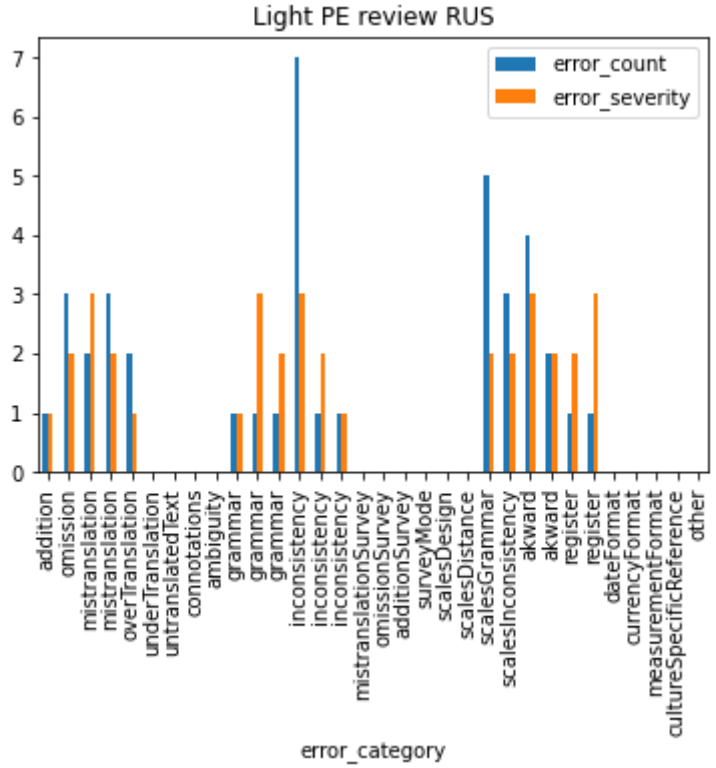


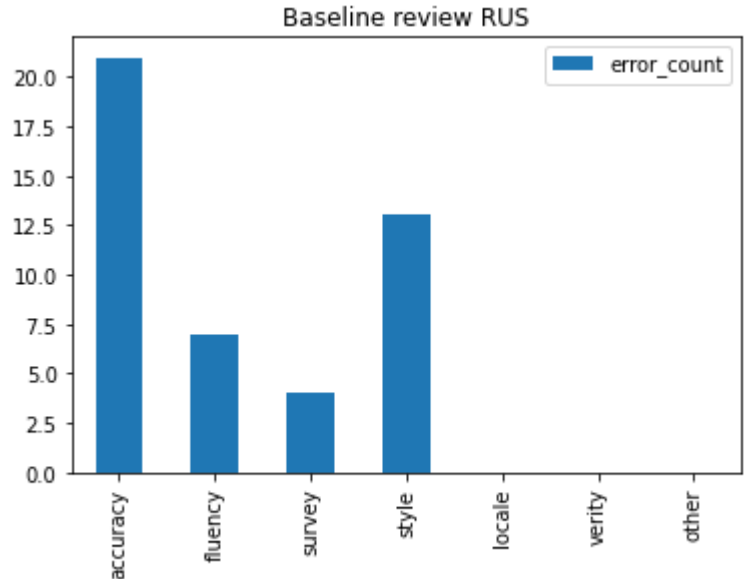
This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782

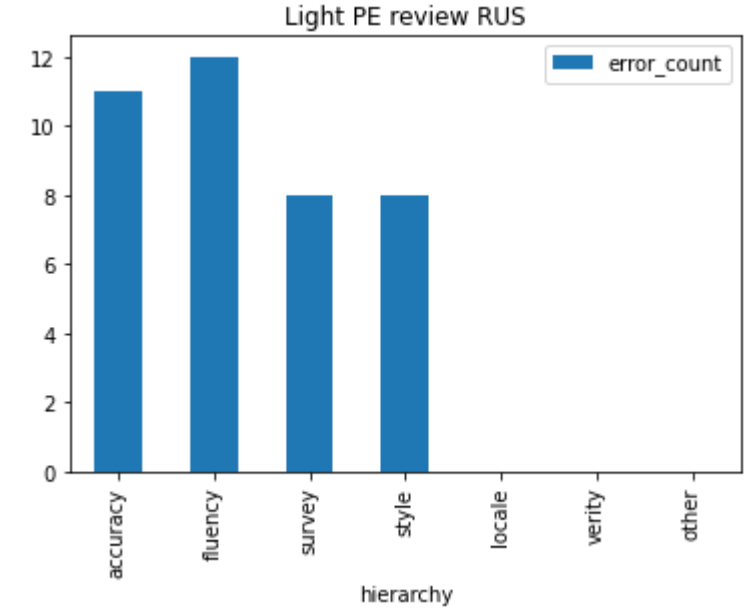
Example BSLN RUS vs LIGHT PE RUS

- The value of p is < .00001.
- The result is significant at p < .05.
- **Difference = BASELINE has more errors**









This project is funded from the EU Horizon 2020 Research and Innovation Programme (2014-2020) under Grant Agreement No. 823782