

Session: Investigating the effects of machine translation and post-editing in the TRAPD: an experimental approach

Session organisers:

Diana Zavala-Rojas (UPF-RECSM, ESS ERIC)

Dorothee Behr (GESIS)

European Survey Research Association Conference 2021

July 9, 2021 13:15 - 14:45

Session sponsor:



Project:



Horizon 2020
European Union Funding
for Research & Innovation

Type of action & funding:
Research and Innovation action
(INFRAEOSC-04-2018)

Partners: 45
(20 beneficiaries + 25 LTPs)

SSH ESFRI Landmarks and Projects
& international SSH data infrastructures

Project budget:
€ 14,455,594.08

Duration: 40 months
(January 2019 – 30 April 2022)

Project website:
www.SSHopencloud.eu



Objectives:

- creating the social sciences and humanities (**SSH**) part of European Open Science Cloud (**EOSC**)
- maximising **re-use** through **Open Science** and **FAIR** principles (standards, common catalogue, access control, semantic techniques, training)
- interconnecting existing and new infrastructures (clustered cloud infrastructure)
- establishing appropriate **governance model** for SSH-EOSC

Welcome to this session

- Social Sciences & Humanities Open Cloud (SSHOC) and Machine Translation research
 1. Investigating the effects of Machine Translation and Post-editing in the TRAPD: experimental design and methodological considerations (Zavala-Rojas, D. et al).
 2. On the Impact of Machine Translation on the Quality of the final Review outputs (Dorer, B. et al).
 3. Survey translation according to the team approach: On the impact of post-edited translations on final review output (Behr, D. et al.)
 4. Assessment of machine translations of survey questions and response scales (Sorato, D. et al.).
 5. Usability of neural machine translation application for translation of measurement instruments (Keck, V. et al.).

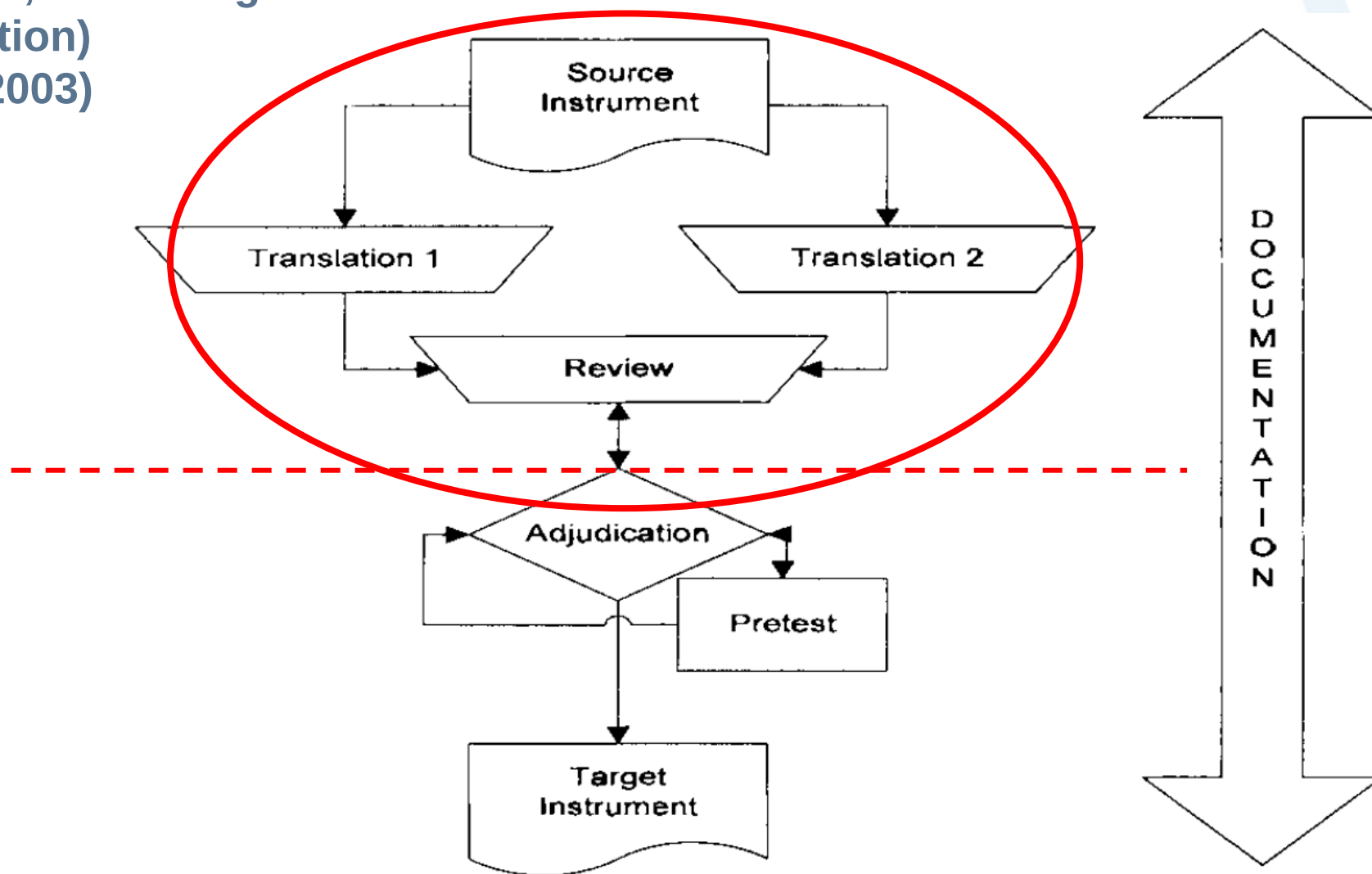


Investigating the effects of Machine Translation and Post-editing in the TRAPD: experimental design and methodological considerations



**TRAPD (Translation, Review,
Adjudication, Pretesting and
Documentation)
(Harkness 2003)**

TRAPD Model



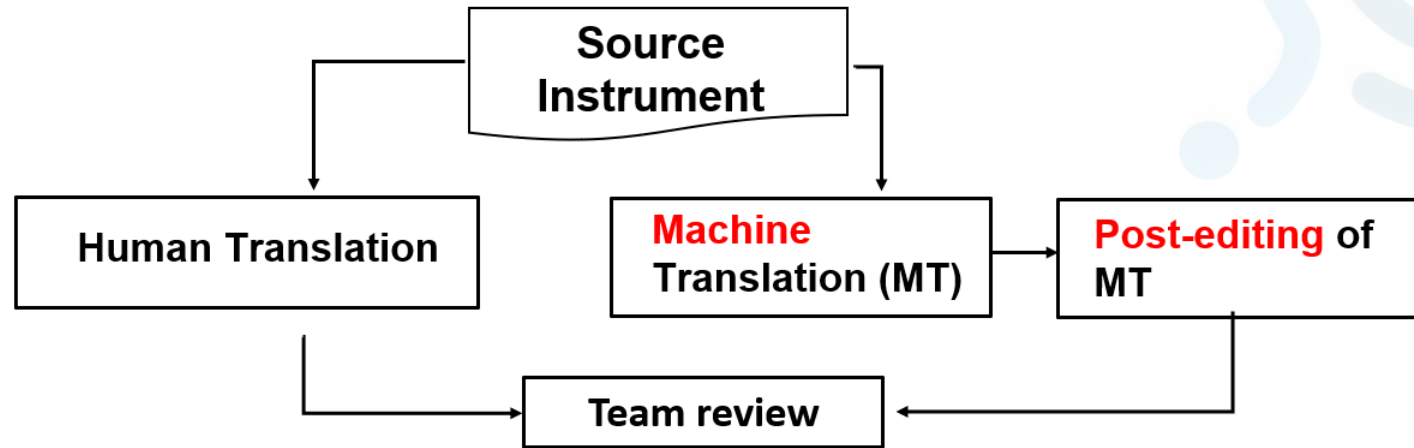
TRAPD Model and Machine translation

- 🌀 TRAPD is focused on the translation process, not on translation tools (Harkness 2003).
- 🌀 Human work intense, increasingly adopted by large survey projects.
- 🌀 Quality is considered when developing translation in a committee approach and in the Pretest.
- 🌀 In practice increasingly additional quality assessment methods of translations are used as a complement to the procedure.
- 🌀 Machine translation evolved and improved in quality since 2014, when the paradigm changed from statistical, rule-based or hybrid machine translation models to (artificial) neural networks modelling.

Experiments making use of machine translation and post-editing in TRA(PD)

- 🌀 Pilot and pioneer research project to test for potential effects of machine translation and post-editing when implementing the TRAPD approach to survey translation.
- 🌀 Testing was conducted under a highly controlled experimental setting. Difficult with humans!
- 🌀 Two experimental interventions compared to a baseline/control group:
 - 🌀 Control group with only human translation
 - 🌀 Machine translation and full post editing
 - 🌀 Machine translation and light post editing

Adapted TRA(PD) Model & Experimental Goal

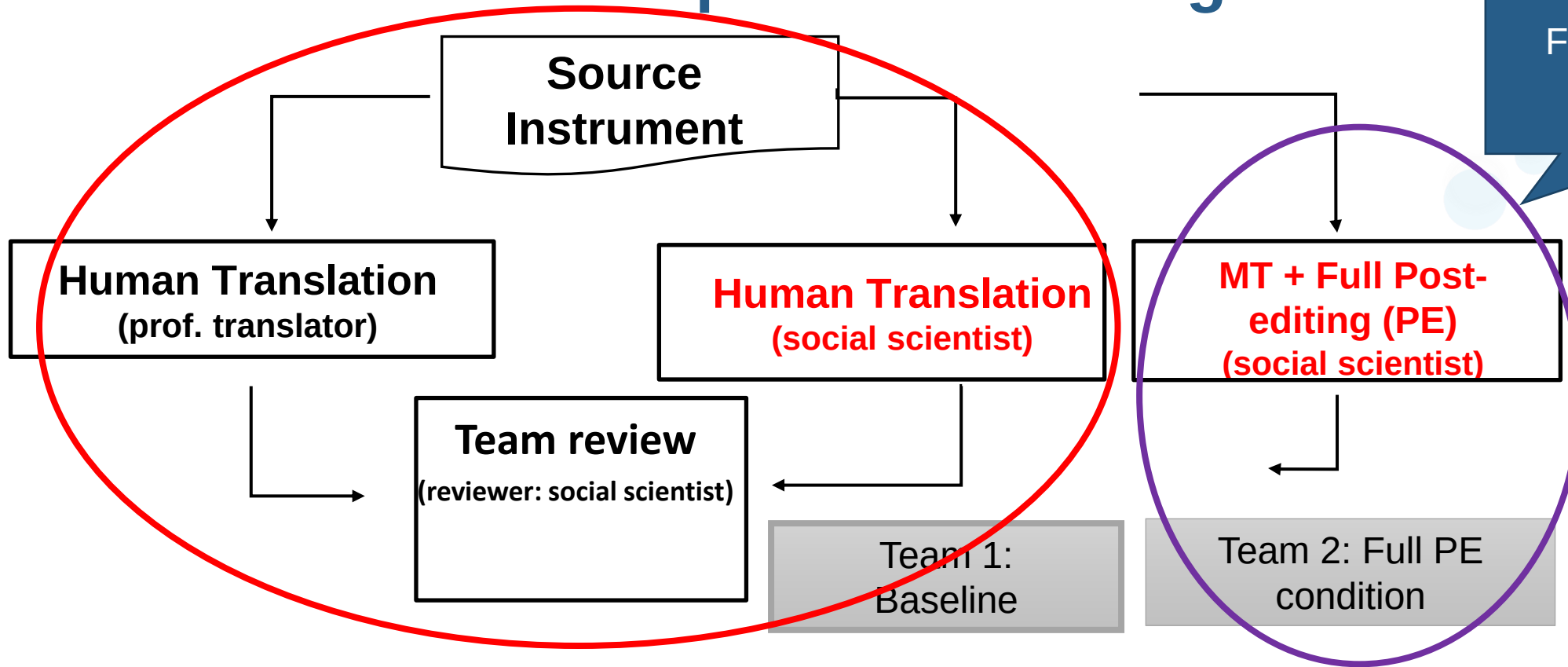


- Making use of machine translation and post editing at the translation stage
- Obtaining evidence on the potential impact of employing machine translation and different types of post editing in the translation of survey questionnaires

Post-editing

- A process in which human translators correct and improve a machine translated text.
- **Light post-editing** covers the revision of the raw machine translation with only a few modifications required to make it understandable for the end user.
- No attempt is made to produce a result comparable to a text written by a human translator.
- **Full post-editing** implies a detailed in-depth revision of the machine translation with the purpose of producing an “end result which is of a comparable quality to a human translation” (ISO 18587: 2017)

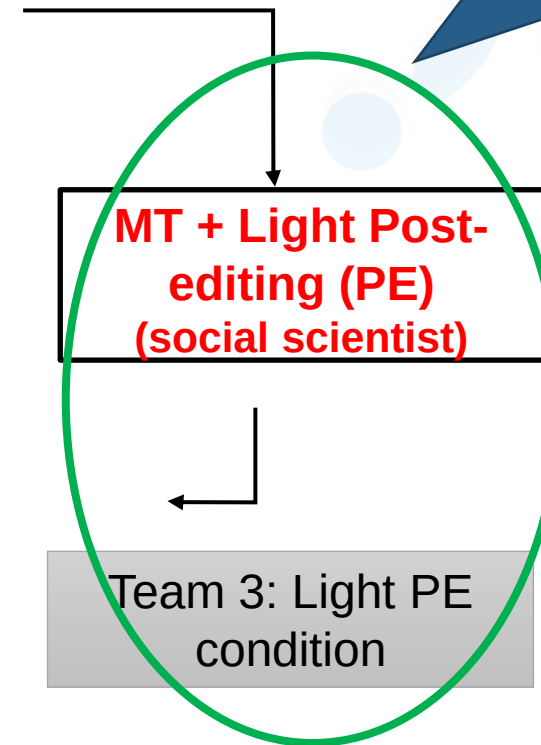
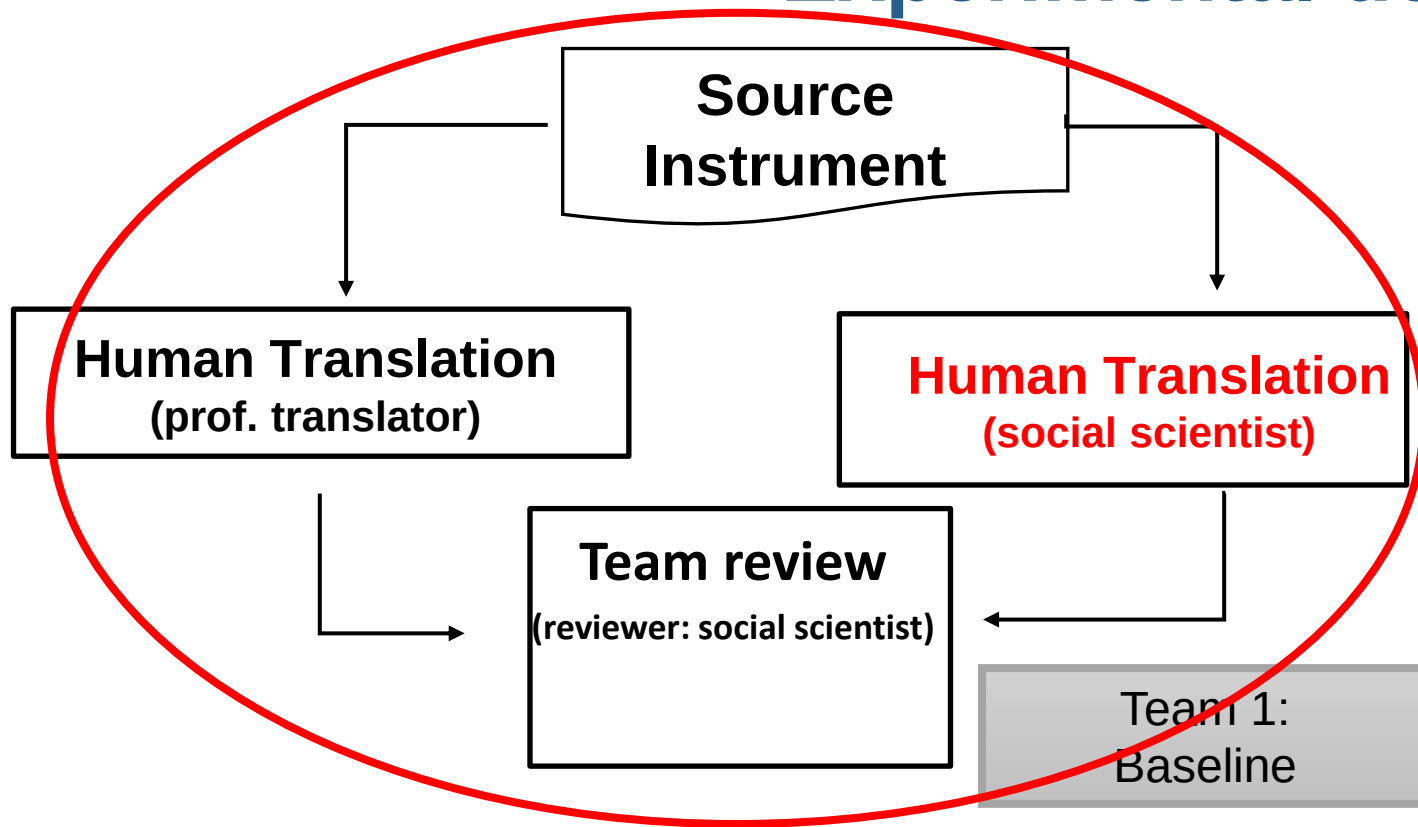
Experimental design



Full PE: human-like

Experimental design

Light PE: accurate but not necessarily correct language-wise



Questionnaire for the experiment

- 40 questions sampled from the European Social Survey and European Values Study questionnaires
- Two-fold selection process:
 - Random sampling of about 300 questions from EVS Waves 1-5 and ESS Rounds 1-9. In the case of grids, the parent/first question was attached to the one selected in the random draw.
 - Team selected final questions based on problematic terms and other translation related criteria.
 - Some questions were slightly amended to make the “lab” questionnaire fluent

Languages in the experiments

🌀 English is the source language and German and Russian are the target languages.

🌀 Because:

- 🌀 Russian is the language translated in more countries in the survey projects. Only in ESS: Estonia, Israel, Latvia, Lithuania, Russia, Georgia and Ukraine. And can provide information for Slavic languages.
- 🌀 German is used by a high number of countries, in ESS: Austria, Germany, and Switzerland, and can provide insight for languages from the Proto-Germanic family which includes Dutch, Norwegian and English.
- 🌀 Both have several degrees of complexity for MT engines (gender, inflections, alphabets among Latin and Cyrillic).
- 🌀 The research team speaks these languages.

Participants in the experiment

- Participants had unique roles in the experiments among:
 - Translator using human translation,
 - Translator using MT and post editing
 - Reviewer in the TRAPD.
- Recruitment was done in July and August 2020.
- Backgrounds of participants were selected according to the TRAPD literature: social scientists and professional translators
- Participants received a fee for collaborating in the experiments
- Participants answered 2 or 3 questionnaires depending on the role in the experiments, before and after the experiments.
- Participants had different degrees of knowledge about survey translation.

Profiles of participants

Professional translators:

-  Degree in translation/linguistics (other degree possible in combination with translation experience), at least 5 years translation experience, experience in survey translation, native speaker of German/Russian, proficiency in English.

Social scientists as translator/post-editor:

-  Degree in social sciences (other degree possible in combination with work experience in social sciences), at least 2 years of work experience in the social sciences, native speaker of German/Russian, proficiency in English, some form of questionnaire translation experience.

Social scientists as reviewers:

-  As above, but with at least 5 years of work experience in the social sciences and experience in review sessions ideally as the lead.

Fieldwork

- 🌀 The experiments took place in September and October 2020.
- 🌀 The source questionnaire was imported into MateCat, an open-source translation tool.
- 🌀 Participants produced their translations in this tool.
- 🌀 Participants received training of approximately 2.5 hours.
- 🌀 Due to the pandemic the committee meetings were virtual instead of in person.

Limitations

- Pilot project: only possible to include 40 items and 2 language pairs (English-German, English-Russian).
- Design does not allow for generalization of conclusions.
- Experiments with human subjects are difficult by design, as it is difficult to control and monitor behavioural aspects (eg. did they follow instructions 100%?), personality tendencies (eg. do some participants participate more actively in the committee meetings?), learning curve.

Configuration of the groups in Experiment 1

- Two language pairs: English-German, English-Russian.
- Baseline group:
 - T1: Human Translator (professional translator).
 - T2: Human Translator (social scientist).
 - R: Reviewer (social scientist).
- Treatment 1:
 - T1: Human Translator (professional translator).
 - T2: **Machine translation + full post-editing** (social scientist).
 - R: Reviewer (social scientist).

Configuration of the groups in Experiment 2

- Two language pairs: English-German, English-Russian.
- Baseline group:
 - T1: Human Translator (professional translator).
 - T2: Human Translator (social scientist).
 - R: Reviewer (social scientist).
- Treatment 1:
 - T1: Human Translator (professional translator).
 - T2: **Machine translation + light post-editing** (social scientist).
 - R: Reviewer (social scientist).

Analysis strategy: interdisciplinary array of methods and approaches

- 🌀 **Error scheme** defining categories of potential errors (Pres by Dorer, et al).
 - 🌀 The **translations** were **evaluated** by **independent coders** (not the same participants in the experiments).
 - 🌀 **Blind coding process**: Coders did not know which texts use machine translation in the process and which not.
- 🌀 Analysis of **anchor terms** to assess the impact of post-edition at the review step (Behr, et al.).
- 🌀 Analysis of the **raw machine translation** outputs with **automated metrics**, such as similarity measures, BLEU, METEOR, etc.. (Sorato et al.).
- 🌀 Analysis of the **questionnaires and the post-editing output** using the framework of **usability of machine translation** (Keck et al).
- 🌀 Future work: **Decision making** in the committee meetings, analysis of **transcripts** (Dorer, et al.).



Any question about the design of the experiments?

Let's look at the results...



Thank you for your attention!

contact: diana.zavala@upf.edu

Join our community



<https://www.sshopencloud.eu>



@SSHOpenCloud



info@sshopencloud.eu



/in/sshopencloud