Exploring different speech recognizers for post-editing translation outputs: A pilot study in an international organisation

Jeevanthi Liyanapathirana, Faculty of Translation and Interpreting (FTI), University of Geneva

Prof. Pierrette Bouillon, Faculty of Translation and Interpreting (FTI), University of Geneva

24/11/2022
Background of our research study

- Improvements in Machine Translation (MT) quality, increasing demand for translations, post-editing
  - Larger volumes of translations while saving time and costs
  - Speech technology integration is likely to contribute to further innovation

How would a combination of Speech Recognition + MT help?
Question: What are the similarities and differences between written post-editing and spoken post-editing?

Q1) Does written post-editing of machine translation demand more temporal and technical effort (Krings, 2001) than spoken post-editing?

Q2) Is spoken post-editing more satisfying than written post-editing for the professional?

Q3) Are speech techniques better when used for re-speaking the translation or for post-editing the machine translation suggestions?
Previous Work

a) Surveying the potential of using speech technologies for post-editing purposes in the context of international organizations: What do professional translators think? (1,2,3)

b) Using speech technology in the translation process workflow in international organizations: A quantitative and qualitative study (1,2)

1) Jeevanthi Liyanapathirana
Fac. de traduction et d’interprétation
University of Geneva, 1211 Geneva, Switzerland
juliyanapathirana@gmail.com

2) Pierrette Bouillon
Fac. de traduction et d’interprétation
University of Geneva, 1211 Geneva, Switzerland
pierrette.bouillon@unige.ch

Bartolomé Mesa-Lao
Universitat Oberta de Catalunya
Av. Tibidabo, 39-43
08035 Barcelona, Spain
bmesa@uoc.edu
Previous Work

c) Integrating post-editing with Dragon speech recognizer: a use case at international organizations (1,2)

d) Integrating Speech in Post-Editing (PE) - Comparison of two PE Interfaces (1,2)

1) Jeevanthi Liyanapathirana
Fac. de traduction et d’interprétation
University of Geneva, 1211 Geneva, Switzerland
juliyanapathirana@gmail.com

2) Pierrette Bouillon
Fac. de traduction et d’interprétation
University of Geneva, 1211 Geneva, Switzerland
pierrette.bouillon@unige.ch
Today’s pilot study / starting point

Q4) How would different parameters affect the choice of using spoken based post-editing or written post-editing (based on translator satisfaction and effort)?

- The type of speech recognition tool
- Translator profile

We conducted a pilot study to start investigating on using different types of speech recognizers for post-editing (in an international organisation)

Aim was to do a preliminary background check to identify potential tools/methods to perform an actual study
- With some preliminary results!
Resources

- MS Word, Google Docs

- Dragon Professional, Google Speech and Microsoft Dictate

- Neural machine translation engines trained specifically using trade domain English and French parallel data were used as MT suggestions
1) Setup using Speech Recognition: MS WORD + MS Dictate

- The user can train his or her voice on Windows OS and start using e.g. Word document to transcribe their speech.

- Less correction commands based on the language

- E.g. Way more commands for English, than for French
2) Setup using Speech Recognition: Google Docs + Google Voice

- The user does not have to train the voice

- Less correction commands based on the language

- E.g. Way more commands for English, than for French
3) Setup using Speech Recognition: Dragon + MS Word

- **Training translator profiles, adding domain specific vocabulary, using built-in commands** as well as **training new commands** to navigate through Trados using Dragon speech

- Dragon can easily be used with MS Word
Experiment

- Four professional translators (EN-FR) were asked to translate four comparable texts (average length of 220 words) of the trade domain using:
  1. Post-editing via typing (M1)
  2. Spoken Translation (MSD) + Revision by typing (M2)
  3. Spoken Translation (GS) + Revision by typing (M3)
  4. Spoken Translation (DP) + Revision by speech/typing (M4)

- Participants were given a short training and manuals on voice commands

- Pre and post-experiment questionnaires on user satisfaction (qualitative analysis), Screen recording software (quantitative analysis)
Results (BLEU and Term Error Rate)

Translation performances of each of the four methods were then compared against using BLEU and Translation Error Rate (TER: between post-edited translation and reference translation) scores.

<table>
<thead>
<tr>
<th>Method</th>
<th>BLEU (PE-Ref)</th>
<th>TER (PE-Ref)</th>
<th>Average Minutes per segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>51.55</td>
<td>0.34</td>
<td>2.44</td>
</tr>
<tr>
<td>M2</td>
<td>57.67</td>
<td>0.304</td>
<td>1.87</td>
</tr>
<tr>
<td>M3</td>
<td>47.59</td>
<td>0.401</td>
<td>2.06</td>
</tr>
<tr>
<td>M4</td>
<td>54.65</td>
<td>0.37</td>
<td>2.33</td>
</tr>
</tbody>
</table>
Quantitative Observations

- Translating using MSD followed by revising via typing (M2) provides the best BLEU score with lesser edits while taking the least time compared to other methods.

- Post-editing via typing only (M1) took the longest time and translating using GS followed by post-editing via typing (M3) did not perform well in spite of a higher edit rate.

- Dragon Professional method (M4) performed second best in BLEU score
Screen Recording, Keystroke analysis

- Screen recording showed participants using both voice commands and typing to revise the spoken translation in M4 (Dragon Professional).

- M3 (Google Voice + MS word) requiring extensive post-editing using typing - demonstrating inadequate speech recognition support in GS.
Qualitative Observations

- All four participants were pro towards using speech recognition for drafting the translation and then to use the keyboard and mouse for post-editing.

- Participants preferred to use Dragon Professional for speech recognition, stating “resourceful”, “grammatically better French phrases”, “better alternative suggestions” and “more voice commands for French”.

- “Review” feature in MSWord has aided them in their task, Terminology supports in their tasks.
Conclusion and Future work

- A starting point on how different speech recognizers would work when used for translation/post-editing purposes
  - Provides insight for a larger experiment.

- Commercial tools still are on the front (with additional support, resource integration)
  - Speech is deemed attractive for users when given with a mix-and-match choice with typing, performs well in accuracy and takes less time.

Free tools, while attractive, can appear with unexpected hiccups

- Future work
  - Configure the resource setup based on ideal parameters (language pair, translator profile) to investigate speech based translation/post editing.
thank you