

This translation is not too bad:  
An analysis of post-editor choices in a machine  
translation post-editing task

Maarit Koponen

University of Helsinki

*maarit.koponen@helsinki.fi*

2nd Workshop on Post-Editing Technology and Practice  
MT Summit, Nice, September 2, 2013



## Background and the purpose of this study

Post-editing is an inherently subjective task – meaning can be expressed in more than one way.

Questions explored in this study:

## Background and the purpose of this study

Post-editing is an inherently subjective task – meaning can be expressed in more than one way.

Questions explored in this study:

- 1 How often do post-editors agree on the best MT suggestion?
- 2 How often do they produce the same PE version?
- 3 How do individual editors differ from each other?
- 4 What kinds of differences are observed in the PE versions?

# The post-editing experiment

## Material

Pilot evaluation material from EU MOLTO evaluation of a rule-based, multi-lingual controlled language generation and MT tool.

- 139 English source segments (2–15 words per segment, total 827)
- 3 Finnish MT versions (MOLTO GF, Google, Bing)
- Post-editing data from 11 post-editors

Post-editing was carried out by 11 native Finnish speakers using the Appraise tool (Federmann 2012).

# The post-editing experiment

## Material

Pilot evaluation material from EU MOLTO evaluation of a rule-based, multi-lingual controlled language generation and MT tool.

- 139 English source segments (2–15 words per segment, total 827)
- 3 Finnish MT versions (MOLTO GF, Google, Bing)
- Post-editing data from 11 post-editors

Post-editing was carried out by 11 native Finnish speakers using the Appraise tool (Federmann 2012).

## Instructions for the post-editing task:

- 1 Select the MT suggestion you consider the best.
- 2 Edit the suggestion as necessary, or accept as-is.
- 3 If none of the suggestions are acceptable, translate from scratch.

# Post-editor agreement of best MT and acceptability

- The same MT suggestion is selected by all.  
All accept without editing.
- The same MT suggestion is selected by all.  
None accept without editing.
- The same MT suggestion is selected by all.  
Some accept without editing.

# Post-editor agreement of best MT and acceptability

- The same MT suggestion is selected by all.  
All accept without editing.
- The same MT suggestion is selected by all.  
None accept without editing.
- The same MT suggestion is selected by all.  
Some accept without editing.
- Different MT suggestions are selected.  
All accept without editing.
- Different MT suggestions are selected.  
None accept without editing.
- Different MT suggestions are selected.  
Some accept without editing.

# Results: Selection and acceptance

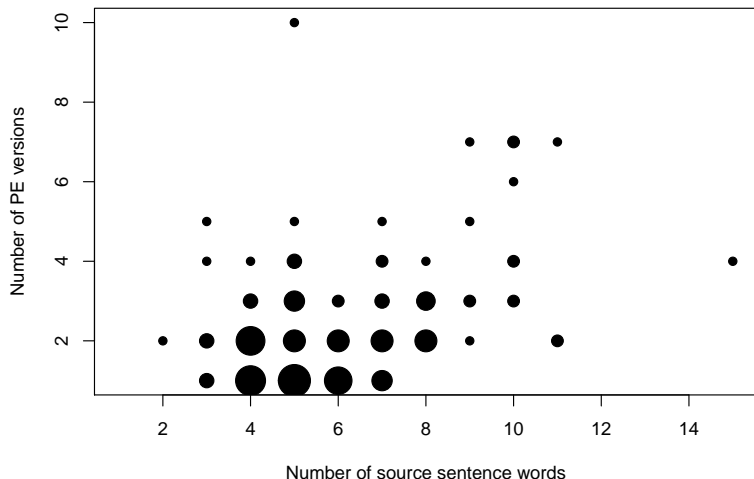
	MT suggestions selected		
	Same	Different	Total
All accept	44	15	59
None accept	1	5	6
Some accept	33	41	74
Total	78	61	139



# Results: Different PE versions per segment

		Number of PE versions				
		1	2	3	$\geq 4$	Total
Same MT	All accept	44	0	0	0	44
	None accept	0	0	0	1	1
	Some accept	0	21	7	5	33
Different MT	All accept	0	15	0	0	15
	None accept	1	0	3	1	5
	Some accept	3	11	13	14	41
Total		48	47	23	21	139

# Versions vs segment length

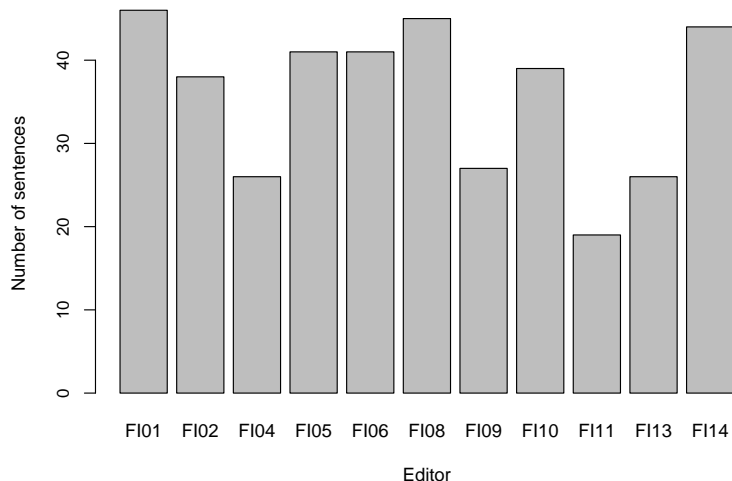


# Comparison of individual editors

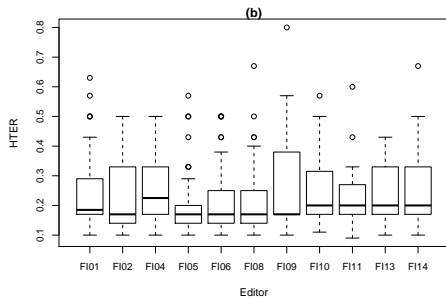
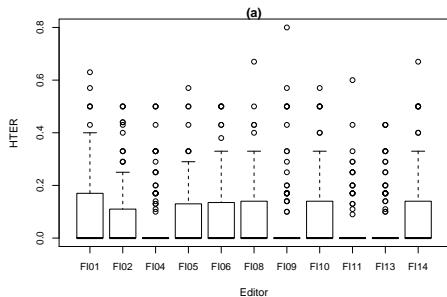
Individual editors were compared in terms of

- Number of segments edited
- HTER : number of edited tokens divided by number of PE tokens – calculated using TERplus (Snover et al. 2009)
- Post-editing time recorded by Appraise (Federmann 2012)
- Number of MT selections differing from the most common choice
- Number of PE versions differing from the most common choice

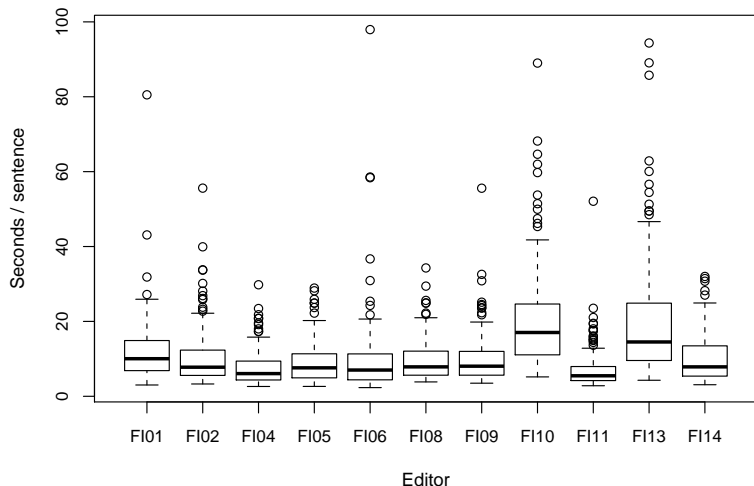
# Results: Number of sentences edited by each editor



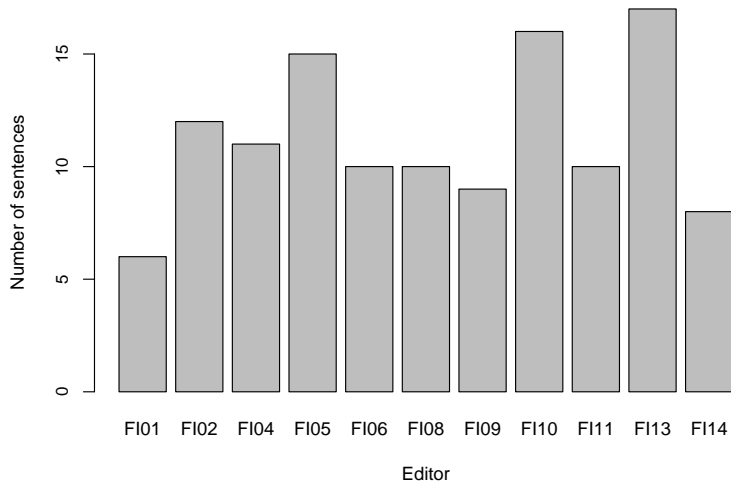
# Results: HTER scores for each editor



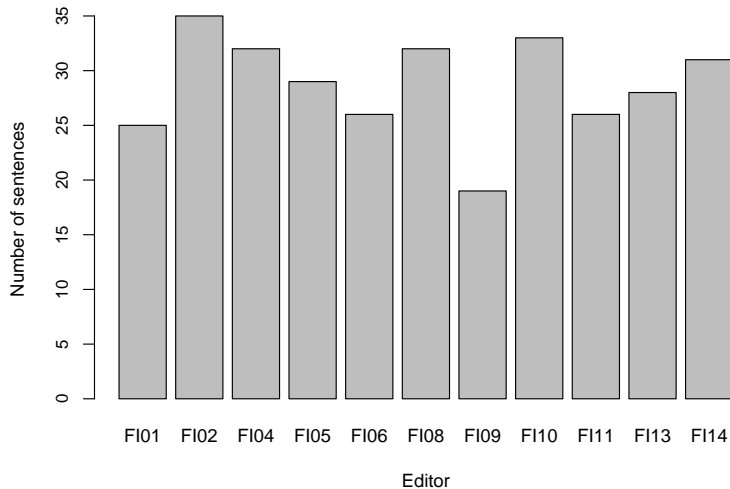
# Results: Editing time for each editor



# Results: Number of MT selections differing from the most common



# Results: Number of PE versions differing from the most common





# Observations about accepted/rejected suggestions and PE differences

- Most rejected suggestions contain multiple errors affecting both meaning and language.
- Differences in selections and PE versions generally involved:
  - ▶ Synonymous words and expressions
  - ▶ Word order
  - ▶ Punctuation
  - ▶ Alternate spellings, alternate suffix forms
  - ▶ Editor errors
- Some particularly strong preferences were observed:
  - ▶ Correct MT edited to include a preferred expression
  - ▶ Incorrect MT with a preferred expression selected over correct MT
- Some sentences attracted a particularly large number of PE versions
  - ▶ Often idiomatic expressions



## Example: Editor preferences

ST Does his daughter want to go to the most popular museum?

MT1 Tahtooko hänen tyttärensä mennä suosituimpaan museoon?

'Does his daughter want to go to the most popular museum?'

MT2 Haluaako tyttärensä Siirry suosituimmista museo?

'Does daughter (of?) want Move (sg2 imperative) from most popular (plural) museum?'

MT3 Ei hänen tyttärensä haluaa lähteä suosituin museo?

'No his daughter want to leave most popular museum?'



## Example: Editor preferences

**ST** Does his daughter want to go to the most popular museum?

**MT1** Tahtooko hänen tyttärensä mennä suosituimpaan museoon?  
'Does his daughter want to go to the most popular museum?'

**MT2** Haluaako tyttärensä Siirry suosituimmista museo?  
'Does daughter (of?) want Move (sg2 imperative) from most popular (plural) museum?'

**MT3** Ei hänen tyttärensä haluaa lähteä suosituin museo?  
'No his daughter want to leave most popular museum?'

8 out of 11 select MT1 as the best suggestion – 6 accept as-is.

3 out of 11 prefer to edit MT2, although it contains multiple errors.



# Example: Differing PE versions

- ST This apple is not too bad.
- FI01 Tämä omena ei ole hullumpi.
- FI02 Tämä omena ei ole hassumpi.
- FI04 Tämä omena ei ole hullumpi.
- FI05 Tämä omena ei ole niin pahaa.
- FI06 Tämä omena ei ole niin huono.
- FI08 Tämä omena ei ole hullumpaa.
- FI09 Tämä omena ei ole kummempaa.
- FI10 Tämä omena ei ole ihan huono.
- FI11 Tämä omena ei ole liian pahaa.
- FI13 Tämä omena ei ole kovin huono.
- FI14 Tämä omena on ihan hyvä.

# Conclusions

- For most source sentences, all editors or all but one select the same MT suggestion.
- For most source sentences, there is one or two PE versions.
  - ▶ This result is likely affected by the nature of the controlled language sentences and good MT quality (large number of segments accepted without editing).



# Conclusions

- For most source sentences, all editors or all but one select the same MT suggestion.
- For most source sentences, there is one or two PE versions.
  - ▶ This result is likely affected by the nature of the controlled language sentences and good MT quality (large number of segments accepted without editing).
- Different editors show differing editing patterns and preferences.
- Preferences for some words or expressions appeared particularly strong.

Thank you! Kiitos!

*maarit.koponen@helsinki.fi*

