

# EVALUATING MACHINE TRANSLATION BY POST-EDITING

DR. MIHAELA VELA

SoSe 2018

# Overview

2

- ❑ Defining post-editing
- ❑ Factors influencing the use of post-editing
- ❑ Post-editing and types of translation
- ❑ Quality expectation
- ❑ Quality measurement
- ❑ Training post-editors
- ❑ Feedback from post-editors
- ❑ Literature

# Post-Editing - definition

3

- The „term used for the correction of machine translation output by human linguists/editors (Veale & Way, 1985)
- „... the process of improving a machine-generated translation with a minimum of manual labor“ (TAUS report, 2010)
- A process of modification rather than revision (Loffler-Laurian, 1985)
- Repairing Texts (Krings, 2001)
- Correction of already translated text rather than translation (Wagner, 1985)

# Post-Editing - definition

4

- *Edit, modify and correct* pre-translated text that has been processed by a machine translation system from source language into target language

# Post-Editing - Usage

5

- Used in big translation companies
  - ▣ Same re-occurring domain
  - ▣ Big amount of data

# Post-Editing vs. Pre-Editing

6

- Pre-editing definition
  - ▣ modifying the input text before automatic translation to facilitate processing
  
- Pre-editing techniques
  - ▣ Use of style guides
  - ▣ Use of controlled terminology
  - ▣ Use of controlled language

# Post-Editing – Controlled Languages

## (Uwe Muegge)

7

- A controlled language has stricter rules than the general language
- Controlled languages enable authors to write texts that are easily comprehensible
- Grammar rules are different for each language
  - ▣ the rules for controlled languages differ from language to language
- The following text types are particularly suitable for the use of a controlled language:
  - ▣ Help systems
  - ▣ Technical specifications
  - ▣ Technical reports
  - ▣ Technical documentation

# Post-Editing – Controlled Languages (Uwe Muegge)

8

## □ Rules:

- Write sentences that are shorter than 25 words
- Write sentences that express only one idea
- Write the same sentence if you want to express the same content
- Write sentences that are grammatically complete
- Write sentences that have a simple grammatical structure
- Write sentences in the active form
- Write sentences that repeat the noun instead of using a pronoun
- Write sentences that use articles to identify nouns
- Write sentences that use words from a general dictionary
- Write sentences that use only words with correct spelling



# Post-Editing vs. Revision

9

- Revision definition
  - ▣ the act of rewriting something
  - ▣ editing that involves writing something again
  
- Differences
  - ▣ Types of errors
  - ▣ Time available
  - ▣ Level of final quality

# Post-Editing vs. Revision

10

## □ Similarities

- Accuracy
- Completeness
- Logic
- Facts
- Smoothness (cohesion)
- Tailoring (target audience)
- Style
- Mechanism (grammar, etc.)
- Layout
- Typography
- Organisation

# Post-Editing – Translation Approach Types (Allen, 2003)

11

- Inbound translation approach (for internal use)
  - ▣ Apply to raw translation appropriate corrections for not public & perishable documents
    - MT with no post-editing (gisting/rough translation)
      - Rough comprehensible translation
      - Not for public use & perishable
    - Rapid post-editing
      - Minimal amount of corrections in order to be understandable
      - Remove significant errors
      - Stylistics is not taken into consideration
      - Not for public use & perishable

# Post-Editing – translation approach (Allen, 2003)

12

- Outbound translation approach (for external use)
  - ▣ Apply to raw translation appropriate corrections for published documents
    - MT with no post-editing
      - Not available, yet
    - MT with minimal post-editing
      - Depends on the definition of „minimum“
      - Is a learning process
    - MT with full post-editing
      - Higher quality of the resulting texts
      - Translating from scratch vs. post-editing

# Degrees of Post-Editing (O'Brien 2009)

13

- Fast post-editing
  - ▣ Quick turn around
  - ▣ Essential corrections only
  - ▣ Also called
    - Gist post-editing
    - Rapid post-editing
    - Light post-editing
  
- Conventional Post-Editing
  - ▣ Slower turn around
  - ▣ More corrections leading to higher quality
  - ▣ Also called
    - Full post-editing

# Post-editing – Influencing factors (Allen 2003)

14

- User/client requirements
- Volume of data to be processed
- Level of quality
- Time
- Use of document in time/perishability
- Text function/use of the document for gisting to publishable

# Post-Editing: Quality Expectations

15

- MT + PE will generally not produce the same high level quality as HT + revision
  
- Raw MT quality & PE effort will vary depending on
  - ▣ System
  - ▣ Language pair
  - ▣ Domain
  - ▣ Text type
  - ▣ Degree of control of input text

# Post-Editing: Quality Expectations

## System-Type Dependencies

16

### RBMT Systems

- ❑ Level of dictionary coding
- ❑ Level of linguistic coding via rules
- ❑ Customisability
- ❑ Quality of source input

### SMT Systems

- ❑ Quality of training data
- ❑ Domain of training data
- ❑ Volume of training data
- ❑ Linguistic rules

## Terminology



# Post Editing: Quality Expectations

## System-Type Errors

17

### RBMT Systems

- ❑ Incorrect word/term selected
- ❑ Incorrect attachment (e.g. of PPs)
- ❑ Meaning is not disambiguated

### SMT Systems

- ❑ Words added
- ❑ Words omitted
- ❑ Loss of capitalization
- ❑ Loss/incorrect punctuation
- ❑ Some phrases very fluent, other not at all

# Post-Editing: MT Quality

## User Perspectives (TAUS Report 2010)

18

Role	Method	Tools
Developer	Automatic metrics	BLEU, NIST, TER, GTM, etc.
User	Utility, acceptability	User surveys, crowd consensus
Buyer	Financial, practical	ROI, throughput, standard quality measurements
Linguist/Language Service Provider (LSP)	Financial, human evaluation	Word rate, productivity

# Post-Editing:

## Measuring MT Quality for PE

19

- Types of errors
  - ▣ Compares source text with raw MT output
  
- Changes made
  - ▣ Compares post-edited text with raw MT output
  
- Estimated effort
  - ▣ Compares source text with raw MT output and qualitatively estimates PE effort

# Post-Editing: MT Quality Measurement

20

- Which method ist best
  - ▣ Types of errors
    - Good for system development
  - ▣ Changes made
    - Good for system developments
    - Good for post-task assessment of effort
  - ▣ Estimated effort
    - Good for estimating PE productivity prior to task commencement

# Post-editing: MT Quality Measures

21

- Human evaluation metrics
- Automated evaluation metrics
  - ▣ BLEU, METEOR
  - ▣ Precision, recall and F-measure
  - ▣ Unigram-based F-measure → good correlation with human judgements

# Post-Editing: MT Quality Measurement

22

## □ Human evaluation metrics

### ▣ Four categories:

- Excellent
- Good
- Medium
- Poor

# Human Evaluation MEtric

23

## □ Excellent

- ▣ Your understanding is not improved by the reading of the ST because the MT output is syntactically correct; it uses proper terminology; the translation conveys information accurately; minimum style requirements for Doc & Help or software content comply with the MT post-editing guidelines
- ▣ Effect: no post-editing required

# Human evaluation metric

24

## □ Good:

- Your understanding is not improved by the reading of the ST even though the MT segment contains minor errors affecting any of these: grammatical (article, preposition), syntax (word order), punctuation, word formation (verb endings, number agreement), unacceptable style
  - ▣ An end-user who does not have access to the source text could possibly understand the MT segment
- Effect: Only minor post-editing required in terms of actual changes or time spent post-editing



# Human Evaluation Metric

25

## □ Medium

- Your understanding is improved by the reading of the ST, due to significant errors in the MT segment (textual coherence/ textual pragmatics/ word formation/ morphology)
  - You would have to re-read the ST a few times to correct these errors in the MT segment
  - An end-user who does not have access to the source text could only get the gist of the MT segment
- Effect: Severe post-editing is required or maybe just minor post-editing after spending too much time trying to understand the intended meaning and where the errors are

# Human Evaluation Metric

26

## □ Poor

- Your understanding only derives from the reading of the ST, as you could not understand the MT segment. It contained serious errors in any of the categories listed above, including wrong POS
- You could only produce a translation by dismissing most of the MT segment and/or re-translating from scratch
- An end-user who does not have access to the source text would not be able to understand the MT segment at all
- Effect: It would be better to manually retranslate from scratch (post-editing is not worthwhile)

# Post-Editing: Types Of Changes (De Almeida & O'Brien 2010)

28

Essential changes	French	Spanish
Accuracy	17%	21%
Consistency	6%	2%
Format	13%	13%
Language	49%	47%
Mistranslation	13%	12%
Terminology	2%	3%

# Post-Editing: Quality vs. Productivity (Kring 2001)

29

- Some evidence to suggest that medium quality MT output was more demanding than poor quality
- The relationship between number of errors and post-editing is not linear, but exponential

# Post-Editing: Productivity

30

- How is post-editing effort measured?
  - ▣ Temporal
  - ▣ Technical?
  - ▣ Cognitive?
  
- Recurring questions
  - ▣ Is post-editing faster than translation?
  - ▣ Is post-editing more or less keyboard intensive than translation?
  - ▣ Is post-editing more or less cognitively demanding than translation?

# Post-Editing: Productivity

31

- Is PE faster than translation?
  - ▣ Yes, rates vary between 3000-9000 words/day
  - ▣ But, comparisons are often of first pass translation vs. PE, i.e. no revision
  - ▣ There is individual variation
  - ▣ Varies across systems and languages
  - ▣ Can these rates be sustained over one day, the entire week or several months?
- Is PE more or less keyboard intensive than translation?
  - ▣ PE involves less typing than translation
  - ▣ But, translators are usually fast typists
- Is PE more or less cognitively demanding than translation?
  - ▣ Rarely considered
  - ▣ Translators report on being more tired after PE three texts vs. two
  - ▣ PE is more tedious?

# Post-Editing: Pricing

32

- Two approaches (TAUS 2010)
  - ▣ Based on fuzzy segment matches
  - ▣ Based on time spent
  
- Variations in payment per word/segment rate
  - ▣ Between 15% and 25% of fuzzy match rate
  - ▣ Per word discount on price
  - ▣ Percentage of no-match word rate
  - ▣ 50% of human translation rate
  - ▣ Rate based on productivity

# Post-Editing: Quality & Productivity For Different PE Types

33

## Light Post-Editing

- ❑ Low to medium quality
- ❑ Output/result could be least double normal translation rate

## Full Post-Editing

- ❑ Medium to high quality
- ❑ Output/result could be faster than translation, but rate lower than rate for light PE



# Post-Editing: Guidelines for Post-Editing

34

- There are no standard guidelines for post-editors
- Guidelines are either too vague or too detailed
- Guidelines may need to be system- and language-specific
- How to differentiate between essential and preferential changes?
- How to differentiate guidelines for different degrees of post-editing?

# Post-Editing: Training Post-Editors

35

- Who is the best post-editor?
- Where should training be done?
- What training is required?
- Disconnects between translation professionalism and post-editing demands

# Post-Editing: Finding Post-Editors

36

- Post-editor's professional profile not very developed yet
  - ▣ Traditional translators often perform PE task
- Frustration element
  - ▣ Caused by recurring MT output errors
    - If unreported, require repetitive editing
  - ▣ Quality plateau with rules-based engine reached for some languages
- Closer collaboration with post-editors required to resolve issues quickly

# How to be a Good Post-Editor

37

- Training?
- Good post-editor = good translator?
- Training on PE & good translation skills

# Training Post-Editors - Skills

38

- Excellent knowledge of SL ( $\equiv$  translator)
- Excellent command of TL ( $\equiv$  translator)
- Specialised domain knowledge ( $\equiv$  translator)
- Excellent key-boarding skills ( $\equiv$  translator)
- Good revision skills
- Ability to make quick quality assessment and to adhere to guidelines
- Tolerance
- Positive attitude to MT

# Proposal For Training (O'Brien 2002)

39

- Additional skills
  - ▣ Knowledge of MT
  - ▣ Term management skills including
    - MT dictionary coding (RBMT)
    - Corpus quality assessment (SMT)
    - Term management (exchange formats, tools)
  - ▣ Pre-editing / Controlled Language skills
  - ▣ Basic programming skills
    - E.g. macros for automated text correction

# Training Post-Editors

40

- Who is the best post-editor?
- Intuition:
  - ▣ Good post-editor = good translator, but...

# Training Post-Editors

- Evidence suggests that less-experienced translators may benefit more from MT than long-term professional translators
- More experience = more preferential (i.e. stylistic) changes
- More experience sometimes = negative opinion of MT & PE
- Are bilinguals to be preferred over translators?
  - ▣ Some may be good post-editors, others will not be good (i.e. same as translation community)
  - ▣ If PE is mixed with HT in a TM environment, translators are still preferred



# Training Post-Editors - Where

42

- In transition
  - ▣ Currently: mostly in-house, on-the-job
  - ▣ Post-editing is creeping into university curricula

# Training Post-Editors - Difficulties

43

- Essentially, translators are asked to unlearn much of what they are taught regarding quality and professionalism:
  - ▣ Ignore style, fluency, cohesion, coherence, text function, context, end user...
  - ▣ Do more, of lower quality, for much less pay
- Post-editors are “self-selecting”
- Post-editing is best mixed with “regular” translation
- Success: post-editors are “part of” the dialogue and process

# Tools for Post-Editing

44

- Why a tool?
  - ▣ Familiarity
  - ▣ Mixing human translation with machine translation
  - ▣ Measuring cognitive and time effort
  
- Tools (any CAT tool)
  - ▣ MateCAT
  - ▣ SMARTMate
  - ▣ PET
  - ▣ TransLog

# Feedback from the Post-Editor

45

- Feedback gathering is a separate task
- Unstructured feedback can be difficult to process
- Unstructured feedback can be difficult to implement
- Possible feedback questions:
  - ▣ Revisions procedures
  - ▣ Post-editing learning curve
  - ▣ Post-editing proficiency
  - ▣ Post-editing effort
  - ▣ Price satisfaction
  - ▣ Job satisfaction

# Literature

46

- Jeffrey Allen. Post-editing. In Computers and Translation - A Translator's Guide. John Benjamins Publishing, Amsterdam, Netherland, 2003. ISBN 978-9-027-21640-3
- Wilker Aziz, Sheila Castilho Monteiro de Sousa, and Lucia Specia. PET: a tool for post-editing and assessing machine translation. In Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12), Istanbul, Turkey, may 2012. European Language Resources Association (ELRA). ISBN 978-2-9517408-7-7.
- Giselle de Almeida and Sharon O'Brien. Analysing post-editing performance: Correlations with years of translation experience. In Proceedings of the 14th Annual Conference of the European Association for Machine Translation (EAMT 2010), Saint-Raphaël, France, May 2010
- GALE. Post editing guidelines for gale machine translation evaluation. Technical report, GALE, 2007. URL [http://projects.ldc.upenn.edu/gale/Translation/Editors/GALEpostedit\\_guidelines-3.0.2.pdf](http://projects.ldc.upenn.edu/gale/Translation/Editors/GALEpostedit_guidelines-3.0.2.pdf)

# Literature

47

- Roy Green. The MT errors which cause most trouble to post-editors. In Proceedings of the ASLIB conference 1982, London, UK, November 1982.
- Hans Krings. Repairing Texts - Empirical Investigations of Machine Translation Post-Editing Processes. Kent State University Press, Ohio and London, 2001. ISBN 978-0-873-38671-5.
- Anne-Marie Laurian: Machine translation: what type of post-editing on what type of documents for what type of users. In Proceedings of Coling84: 10th International Conference on Computational Linguistics & 22nd Annual Meeting of the Association for Computational Linguistics, 1984, Stanford University, California.
- Uwe Muegge website: <http://www.muegge.cc/controlled-language.htm>
- Sergio Penkale and Andy Way. SmartMATE: An Online End-To-End MT Post-Editing Framework. In AMTA 2012 Workshop on Post-Editing Technology and Practice (WPTP 2012), San Diego, USA, October 2012. Association for Machine Translation in the Americas (AMTA).

# Literature

48

- Sharon O'Brien. Teaching post-editing: A proposal for course content. In Proceedings of the 6th EAMT Workshop Teaching Machine Translation, Manchester, UK, 2002
- Sharon O'Brien. Introduction to post-editing: Who, what, how and where to next? In AMTA 2010 Workshop on Post-Editing Technology and Practice (WPTP 2010), Denver, USA, October 2010
- Sharon O'Brien, Johann Roturier, and Giselle de Almeida. Post-editing machine translation output. In SMT 2009 Tutorial on Post-editing Machine Translation Output (SMT 2009), Ottawa, Ontario, Canada, August 2009.
- TAUS. Postediting in practice. a taus report. Technical report, TAUS, 2010a. URL <http://www.translationautomation.com/articles/>
- TAUS. Results post-editing survey. Technical report, TAUS, 2010b. URL <http://www.translationautomation.com/articles/>

# Questions?



# Assignment – Get acquainted with MateCAT

50

- Go to MateCAT (use Google Chrome)
- Create a translation project in MateCAT
  - ▣ Use the settings of your choice
  - ▣ Use a text text of your choice
  - ▣ Use a language combination of your choice
- What is the difference between LookUp and Update?
- What is the difference between enabling MT and not enabling MT?