Introduction to Post-Editing: Who, What, How and Where to Next?
Dr. Sharon O’Brien,
Centre for Next Generation Localisation
Dublin City University

Definitions of Post-Editing

- The “term used for the correction of machine translation output by human linguists/editors” (Veale and Way 1997)

- “…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)
Different from “Pre-editing”

- Pre-editing: modifying the input text *before* automatic translation to facilitate machine processing

- Pre-editing techniques include:
  - Use of style guides
  - Use of controlled terminology
  - Use of controlled language rules

Different from “Revision”?  

- Overlaps, but differences too:

  - Differences:
    - Types of errors
    - Time available
    - Level of final quality
Different from “Revision”? 

• Overlaps?  
  – Revisers check for (Mossop 2001):  
    • Accuracy  
    • Completeness  
    • Logic  
    • Facts  
    • Smoothness (cohesion)  
    • Tailoring (target audience)  
    • Style  
    • Idiom  
    • Mechanics (grammar etc.)  
    • Layout  
    • Typography  
    • Organisation

Degrees of Post-Editing

• “Fast Post-Editing”:  
  – Quick turn-around  
  – Essential corrections only

• Also called:  
  – Gist Post-Editing  
  – Rapid Post-Editing  
  – Light Post-Editing
Degrees of Post-Editing

• “Conventional Post-Editing”:
  – Slower turn-around
  – More corrections leading to higher quality

• Also called:
  – Full Post-Editing

Degrees of Post-Editing

• Decided by:
  – User Requirements
  – Volume
  – Quality Expectations
  – Turn-Around Time
  – Perishability
  – Text Function

(Allen 2002)
Light vs. Full?

• Is the distinction useful?
  – Evidence that most MT users engage in full post-editing (TAUS Report 2010)
  – Scenarios for light post-editing are few?
  – Raw MT or Full post-edit?

<table>
<thead>
<tr>
<th>Source Text</th>
<th>Raw MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Un vaste réseau qui piratait les codes de déverrouillage des téléphones portables a été démantelé, ont annoncé, dimanche 26 septembre, les enquêteurs.</td>
<td>A vast network hacked unlock codes for mobile phones has been dismantled, announced Sunday, Sept. 26, investigators.</td>
</tr>
<tr>
<td><strong>Example of Light Post-Edit</strong></td>
<td>A vast network which hacked unlock codes for mobile phones has been dismantled, <em>it was</em> announced Sunday, Sept. 26, <em>by</em> investigators.</td>
</tr>
<tr>
<td><strong>Example of Full Post-Edit</strong></td>
<td>A vast network which hacked security codes for mobile phones has been dismantled, <strong>according to an announcement by</strong> investigators <strong>on</strong> Sunday, Sept. 26.</td>
</tr>
</tbody>
</table>
Examples of post-edited text

• ST: If an error occurred, the error code is displayed.
• MT: Si une erreur se produit, le code d’erreur est affichée.
• MT: Si une erreur se produit, le code d’erreur est *affichée*.
• PE: Si une erreur se produit, le code d’erreur est affiché.

Examples of post-edited text

• ST: Click this to decompress, or expand, compressed files as they are backed up.
• MT: Cliquez sur cette option pour decompress ou développer, les fichiers compressés ils sont sauvegardés.
• MT: Cliquez sur cette option pour decompress ou développer, *(φ)* les fichiers compressés ils sont sauvegardés.
• PE: Cliquez sur cette option pour décompresser ou développer les fichiers compressés, tandis qu’ils sont sauvegardés.
Examples of post-edited text

• English-German: Example of variability across PE solutions:
  – ST Select the C drive.
  – MT Wählen Sie das C- Laufwerk aus.
  – P1 Wählen Sie -Laufwerk C aus.
  – P2 Wählen Sie das Laufwerk C aus.
  – P3 Wählen Sie das C- Laufwerk aus.
  – P4 Wählen Sie das C- Laufwerk aus.
  – P5 Wählen Sie das C- Laufwerk aus.
  – P6 Wählen Sie das Laufwerk "C:" aus
  – P7 Wählen Sie das Laufwerk C aus.
  – P8 Wählen Sie das Laufwerk C: aus.
  – P9 Wechseln Sie zu -Laufwerk C-

Examples of post-edited text

• English-Japanese (from Midori Tatsumi’s PhD work) – Example of pronoun being replaced by noun

  • ST: You must have the Folder Full Control role in the folder to give other users access to it.
  • MT: それへの他のユーザーアクセスを与えるフォルダのフォルダのフルコントロールのロールを持たなければなりません。
    [Gloss: it]
  • PE: フォルダへの他のユーザーアクセスを与えるにはそのフォルダのフルコントロールのロールを持たなければなりません。
    [Gloss: folder]
Examples of post-edited text

• Example of a phrase being shifted from one location to another to increase naturalness of text

  • ST: ... show data ingestion progress, and the status of the automatic categorization.
  • MT: ...自動類別のデータ取り込みの進行状況とステータスを現します。[...show data ingestion progress of the automatic categorization and the status]
  • PE: ...データ取り込みの進行状況と自動類別のステータスを現します。[...show data ingestion progress and the status of the automatic categorization]

Quality Expectations

• The received wisdom:
  — MT + PE will generally not produce the same high level quality as HT + revision
  — But, things are changing...?

• Raw MT quality & PE effort will vary depending on:
  — System
  — Language Pair
  — Domain
  — Text Type
  — Degree of control of input text
    • Degree of suitability?
Quality Expectations: System-Type Dependencies

RBMT Systems:
- Level of dictionary coding
- Level of linguistic coding via rules
- Customisability
- Quality of source input

Data-Driven Systems
- Quality of training data
- Domain of training data
- Volume of training data
- Linguistic rules

Terminology, Terminology, Terminology

Quality Expectations: System-Type Errors

RBMT Systems:
- Incorrect word/term selected
- Incorrect attachment (e.g. of preposition phrases)
- Meaning is not disambiguated

Data-Driven Systems
- Words added
- Words omitted
- Loss of capitalisation
- Loss/incorrect punctuation
- Some phrases very fluent, others not at all
Quality – Different User Perspectives

<table>
<thead>
<tr>
<th>Role</th>
<th>Method</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developer</td>
<td>Automatic Metrics</td>
<td>BLEU, NIST, TER, GTM…</td>
</tr>
<tr>
<td>User</td>
<td>Utility, Acceptability</td>
<td>User surveys, crowd consensus</td>
</tr>
<tr>
<td>Buyer</td>
<td>Financial, practical</td>
<td>ROI, throughput, standard quality measurements</td>
</tr>
<tr>
<td>Linguist/LSP</td>
<td>Financial, Human evaluation</td>
<td>Word rate, productivity, standard quality measurements</td>
</tr>
</tbody>
</table>

TAUS Report 2010

Ways of Measuring Quality for PE

• 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
Ways of Measuring Quality for PE

• Which method is best?
  – Types of errors:
    • Good for system development
  – Changes made:
    • Good for system development
    • Good for post-task assessment of effort
  – Estimated effort:
    • Good for estimating PE productivity prior to task commencement

Ways of Measuring Quality for PE

• A note on automatic metrics:

  – Different “currency” from “Fuzzy Match” method
  – Further research on correlations between metrics and PE effort required
Quality – Classifying Errors for PE

- Minor, Major, Grey (Green 1982)
- Single word errors; errors of relation; structural or informational errors (Loffler-Laurian 1983)
- Incorrect verb forms, mistranslation of prepositions, literal rendition of common idioms, consistent translation of a word in one manner when context demands another (Lavorel 1982)

Quality – Types of Changes Made

- De Almeida & O’Brien 2010: Pilot Study - Preliminary Findings:

<table>
<thead>
<tr>
<th>Essential changes</th>
<th>French</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>17%</td>
<td>21%</td>
</tr>
<tr>
<td>Consistency</td>
<td>6%</td>
<td>2%</td>
</tr>
<tr>
<td>Format</td>
<td>13%</td>
<td>13%</td>
</tr>
<tr>
<td>Language</td>
<td>49%</td>
<td>47%</td>
</tr>
<tr>
<td>Mistranslation</td>
<td>13%</td>
<td>12%</td>
</tr>
<tr>
<td>Terminology</td>
<td>2%</td>
<td>3%</td>
</tr>
</tbody>
</table>

- Based on LISA QA model
Quality – Estimated Post-Editing Effort

• E.g. Symantec’s Human evaluation metrics
  
  – Four categories:

• Excellent
• Good
• Medium
• Poor

Managing Expectations – Quality vs. Productivity?
Managing Expectations – Quality vs. Productivity?

• Krings (2001):
  – Some evidence to suggest that medium quality MT output was more demanding than poor quality.
  – The relationship between number of errors and post-editing difficulty is not linear, but exponential.

Managing Expectations - Productivity

• How do you measure post-editing effort?
  – Temporal measurement only?
  – +Technical
  – +Cognitive

• Recurring questions:
  – Is post-editing throughput faster than translation?
  – Is post-editing more or less keyboard intensive than translation?
  – Is post-editing more or less cognitively demanding than translation?
Managing Expectations - Productivity

• Is post-editing throughput faster than translation?

  – Resounding evidence: Yes

  – Throughput rates vary from:

    • 3,000 to 9,000 words per day

Managing Expectations - Productivity

• Is post-editing throughput faster than translation?

  – But:

    • Comparisons are often of first pass translation vs. post-editing, i.e. no revision
    • You will see individual variation
    • It will vary across systems and languages
    • And, one important question remains:

    – Can these throughput rates be sustained over one day, the entire week, or several months?
Managing Expectations - Productivity

• Is post-editing more or less keyboard intensive than translation?

  – Experiments using keyboard logging
    • (e.g. Autodesk, De Almeida & O’Brien 2010, O’Brien 2006)
    – Post-editing clearly involves less typing than translation
    – But, note that translators are usually very fast typists anyway

Managing Expectations - Productivity

• Is post-editing more or less cognitively demanding than translation?
  – Rarely considered (cf. research agenda)
  – Translators report being “more tired” after post-editing –three texts vs. two
  – PE is “more tedious”?
Managing Expectations – Pricing Methods

• Two most popular approaches (TAUS 2010):
  – Paying as fuzzy segment matches
  – Paying a fee based on time spent

• Variations on the 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

Managing Expectations – Pricing Methods

• Important Questions on Pricing Methods:
  – Is the level of effort required for post-editing comparable with Fuzzy Match editing?
  – At what level of Fuzzy Match (50%, 70%, 80%..)?
**Linking Quality, & Productivity to Levels of PE**

<table>
<thead>
<tr>
<th>Light Post-Editing</th>
<th>Full Post-Editing</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Low to medium quality</td>
<td>• Medium to high quality</td>
</tr>
<tr>
<td>• Throughput could be at least double normal translation rate?</td>
<td>• Throughput could be faster than translation, but rate would probably be lower than rate for “light” edits</td>
</tr>
</tbody>
</table>

**Post-Editing Guidelines – Current Challenges**

- No standard guidelines
- Guidelines tend to be too vague or too detailed
- The “2-second” rule is unhelpful
Post-Editing Guidelines – Current Challenges

• 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 Guidelines (General)

• Retain as much raw translation as possible
• Don’t hesitate too long over a problem
• Don’t worry about style (?)
• Don’t embark on time-consuming research
• Make changes only where absolutely necessary,
  – i.e. correct words or phrases that are (a) nonsensical, (b) wrong, (c) omitted or added unnecessarily, and if there’s enough time, (d) ambiguous.
Post-Editing Guidelines (Light)

• The message transferred should be accurate
• Grammatical problems are not a big concern, unless they interfere with accuracy
• Ignore stylistic problems
• Do not spend time researching terms
• Edit any offensive, inappropriate or culturally unacceptable information
• All basic rules regarding spelling still apply
• Textual standards (cohesion, coherence, standard word order etc.) are not so important
• Throughput expectations: very high
• Quality expectations: low

Post-Editing Guidelines (Full)

• The message transferred should be accurate
• Grammar should be accurate
• Ignore stylistic and textuality problems
• Ensure that key terminology is correctly translated
• Edit any offensive, inappropriate or culturally unacceptable information
• All basic rules regarding spelling, punctuation and hyphenation still apply
• For tagged formats, ensure all tags are present and in the correct positions
• Throughput expectations: high
• Quality expectations: medium
Training – Current Challenges

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

Training – Current Challenges

• Who is the best post-editor?

• My intuition:
  – Good post-editor = good translator, but...
Training – Current Challenges

• Who is the best post-editor?
  – Evidence suggests that less-experienced translators may benefit more from MT than long-term professional translators
  – More experience = faster, but...
  – 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 – Skill set

• Excellent knowledge of SL (≡ translator)
• Excellent command of TL (≡ translator)
• Specialised domain knowledge (≡ translator)
• Excellent key-boarding skills (≡ translator)
• Good revision skills
• Ability to make quick quality assessment and to adhere to guidelines
• Tolerance
• Positive attitude to MT
Training – Where should it be done?

• We are in transition...

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

Disconnects between translation and post-editing

• 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
PE Tools & User Interface

• Is there really a need for a “Post-Editing Tool”?

• Translators like familiarity, so
  – Post-editing in familiar editing environments is a plus
  – Also, current workflow usually involves integration with TM environment

PE Tools & User Interface

• Benefits of post-editing in TM environment:

  – Familiarity
  – Mixing HT and MT
  – Access to approved glossary
  – Edits recorded in TM
    • subsequent use for training MT
  – Context
Alternatives

• E.g. PAHO’s use of MS Word, customised toolbar for PE
  – Statistics for post-editor
  – Customised Search and Replace
  – Browse related dictionaries
  – Switch right and left
  – Lower/upper case change
  – Delete next the
  – Change its to their etc.
  – Send problem report to system developers

Alternatives

• Re- or de-capitalise
• Change inflection (plural vs. singular)
• Change gender
• Add/delete punctuation symbol
• Change word order
• Change formatting
• Remove/add words
Research Agenda –
A Selection of Questions

- What UI support would post-editors benefit from?
- Does controlling source input reduce PE effort?
- How does cognitive effort for post-editing compare with fuzzy match effort?
- Are there correlations between automatic MT metrics and post-editing effort?
- Can reviewers differentiate between human translation and MT+PE?
- Can MT automatic confidence scores accurately predict PE effort?
- How do we best deliver training for PE?
- Is there a particular psychological profile most suited for PE?
- How do you get translators to buy into MT/PE?
- How do you (fairly?) price PE?
- Can Statistical Post-Editing (SPE) really help reduce PE effort?

Research Agenda

- What UI support would post-editors benefit from?
  - Not necessarily keyboarding support (Karamanis et al 2010)
  - Is predictive matching really useful to post-editors (e.g. Koehn and Haddow 2009, Caitra experiment)?
  - Support similar to PAHO’s Word macros?
  - Confidence scores from MT system which are calibrated with PE effort?
  - Highlighting of typical errors?
  - Automatic feedback to system developers?
Research Agenda

• Does controlling source input reduce PE effort?
  – Yes (O’Brien 2006)
  – But, controlling source is not an easy task
  – Some controls are more effective than others
  – It does not eliminate PE
  – New question: relation between controlled source and SMT?

Research Agenda

• How does PE cognitive effort compare with editing Fuzzy Matches?
  – Similar to 80-90% fuzzy match for high quality raw output (O’Brien 2006)?
  – If so, what are the pricing implications?
Research Agenda

• Are there correlations between automatic metrics and post-editing effort?
  
  – Preliminary tests suggest there might be correlations between low and high GTM scores, but medium level GTM scores were questionable (O’Brien, forthcoming?)

  • Is medium-quality MT harder to process than low/high quality?
  • If so, what are the implications for pricing?

Research Agenda

• Can reviewers differentiate between HT and MT+PE
  
  – No (Autodesk Experiment)
  – No (Fiederer and O’Brien, 2009)
    • But they have a distinct preference for HT when style is taken into consideration
Research Agenda

• Can Statistical Post-Editing (SPE) really help reduce PE effort?

  – Current research shows significant improvements in automatic metrics (Dugast et al. 2007, Roturier and Senellart 2008)

  – Little research on correlations with human PE effort

Research Agenda

• Can MT automatic confidence scores accurately predict PE effort?

  – Very little research to date
  – Where is the best place to put an MT confidence score?
  – Preliminary study (O’Brien, forthcoming?) suggests that translators want to see scores in a familiar format, i.e. Fuzzy Match %, not 0.5391
Research Agenda

• How do you get translators to buy into MT/PE?
  
  – Learn from the success stories, e.g. PAHO, Symantec
  
  – Commonalities:
    • Long-term project, hard work
    • Buy-in from technical writers
    • Ongoing research
    • Attempts to unify processes (n.b. terminology)
    • Evolving guidelines
    • Incorporation of feedback from post-editors
  
  – Give post-editors a stake in the process

Research Agenda

• How do you (fairly?) price PE?
  
  – Empirical research into post-editing effort (not just throughput based measurements)
  
  – Question assumptions about linearity of quality/productivity