Dr. Sharon O'Brien, CNGL/Dublin City University

Dr. Johann Roturier, Symantec

Ms. Giselle de Almeida, Dublin City University

Post-Editing MT Output
Views from the researcher, trainer, publisher and practitioner

Overview

- Researching and Teaching Post-Editing (40 mins)

- Publisher's View on Implementing PE (Dr. J. Roturier – 40 mins.)

- Break (10 mins)

- The practitioner's viewpoint (G. de Almeida – 40 mins)

- Hands-on post-editing task (30 mins)

- Summary and Wrap-Up (20 mins)

Why talk about post-editing?

- Recent growth in demand for PE due to globalisation (Allen 2003)

- "…we contend that many companies and government agencies will consider automated translation as a way to maximize the amount of information available to customers and constituencies who speak other languages." (Common Sense Advisory Report: "Automated Translation Technology", 2006)

Why talk about post-editing?

- CSA Survey 2006:

  - Half of consumers use MT to understand Anglophone websites

    - Source: Common Sense Advisory, Inc. "Can't Read, Won't Buy" (Sep06) (Online survey of 2,430 consumers in eight non-Anglophone countries)

Definitions of Post-Editing

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

- "checking, proof-reading and revising translations carried out by any kind of translating automaton". (Gouadec 2007)

- A process of *modification* rather than *revision*. (Loffler-Laurian 1985)

- "an unfortunate human intrusion into a fully-automatic process"

Different from "Revision"?

- Differs from traditional translation revision on several fronts, e.g.

    - The types of errors to be corrected are different

    - Misconstruction of meaning will occur at different levels

    - Post-editing seeks the minimum steps required for an acceptable text

    - The final function of the two texts often differs

                                            (McElhaney and Vasconcellos 1988)

Degrees of Post-Editing

● Classified according to :

Degrees of Post-Editing

- "Fast Post-Editing":

  - Quick Turn-around
  - Essential corrections only (Loffler-Laurian)

- 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 (Loffler-Laurian)

- Also called:

  - Full Post-Editing

Degrees of Post-Editing

- Decided by:

    - User Requirements
    - Volume
    - Quality Expectations
    - Turn-Around Time
    - Perishability
    - Text Function

(Allen 2002)

Post-Editing "Rules" (General)

- Retain as much raw translation as possible
- Don't hesitate too long over a problem
- Don't worry if style is repetitive
- 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, and if there's enough time left, (c) ambiguous.

(Wagner 1985)

Light Post-Editing Guidelines

- 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 on researching terms
- Textual standards (cohesion, coherence etc) are not so important
- Throughput expectations: very high
- Quality expectations: low
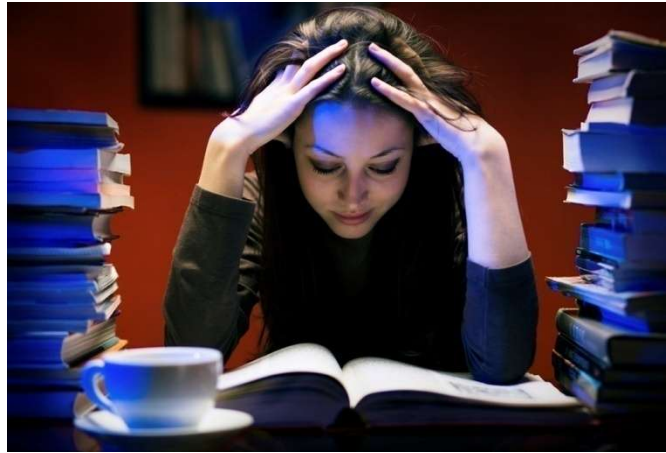
Full Post-Editing Guidelines

- The message transferred should be accurate
- Grammar should generally be accurate
- Do not worry too much about style, standards of textuality
- Spend little time on researching terms etc.
- Throughput expectations: high
- Quality expectations: medium

Diving Deeper

- Post-Editing and Quality (Krings)
- Error Categories (Loffler-Laurian)
- CL and related PE Effort (O'Brien)
- TM vs MT effort (Guerberhof)

Diving Deeper - Quality

- Krings:

  - For MT quality rated as "poor":
    - Ratio of number of errors to words: 1:2
  - For MT quality rated as "medium":
    - Ratio of number of errors to words: 1:3
  - For MT quality rated as "good":
    - Ratio of number of errors to words: 1:4

Diving Deeper - Quality

- Krings:

  - Although there is not much difference in ratios, Krings found that the difference exerts a considerable effect on the post-editing process.

  - The relationship between number of errors and post-editing difficulty is not <u>linear</u>, but <u>exponential</u>.

Diving Deeper - Quality

- Krings:

  - POS errors were found in all sentences rated as "poor" => POS errors particularly critical for post-editing difficulty?

  - When measured by the number of "source text processes" required of the post-editor, medium quality MT output was more demanding than poor quality.

Diving Deeper – Error Categories

- Errors are categorised in order to:

    1. Gauge how "good" the MT output is
    2. Model the post-editing process so we can understand what needs to be fixed, how, and how demanding each type of fix is.

- However, error categorisation is notoriously difficult.

Diving Deeper – Error Categories

Examples:

- 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)

Loffler-Laurian: most detailed, 12 categories, e.g.

3. Vocab/Terminology
4. Abbreviations/Proper Nouns
5. Prepositions
6. Determiners; verb modifiers
7. Verb forms
8. Tense etc..

Diving Deeper – Controlled Language

- CL improves MT output

- But, does it reduce post-editing effort?

(O'Brien 2006)

Diving Deeper – Controlled Language

- Post-editing effort measured by:

  - Temporal (time on task - Translog)
  - Technical (keyboarding effort - Translog)
  - Cognitive (brain power! – the two above plus "Choice Network Analysis")

Diving Deeper – Controlled Language

- ## Findings: overview
  - In general, yes, CL reduces post-editing effort
  - But, individuals differ in post-editing speeds
  - Also, some ST features appear to lead to more post-editing effort than others, e.g.
    - Gerund
    - Ungrammatical constructs
    - Use of "(s)" to mark plural
    - Non-Finite verbs
    - Long noun phrases
    - Short segments (≤4 words)

Diving Deeper – TM vs MT

- Combining MT post-editing with TM editing and unassisted translation (Guerberhof)

- Preliminary findings:
  - Productivity for post-editing MT is greater than for editing fuzzy matches!
  - Translators leave more errors in TM segments than in MT segments!
  - Translators' experience has an impact on processing speed
  - Errors: *slightly* higher in MT output for experienced translators

How to be a good post-editor

- A question of training?

  - To be a good post-editor, you must first be a good translator?
  - Post-editors do not need to be trained translators
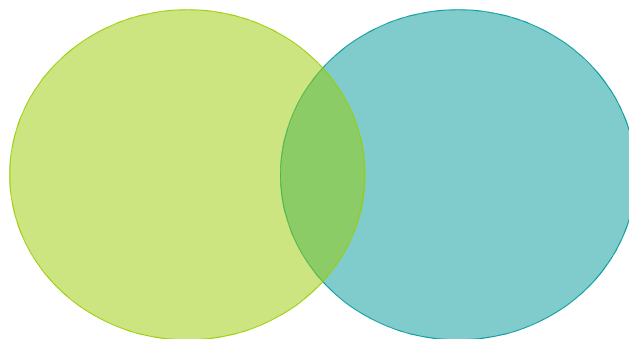 - It doesn't matter, it'll all be done via SPE soon anyway

Allow One Indulgence: Intuition

- My intuition says:

  - Good post-editor = good translator
  - But…
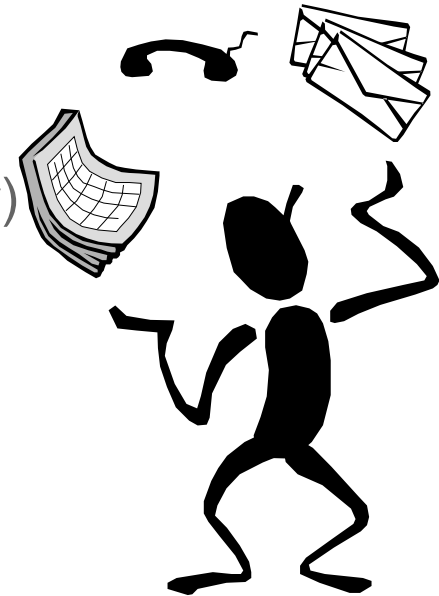
Proposal for training (O'Brien 2002)

● Skill set (Wagner 1987; Vasconcellos & Léon)?

- Excellent knowledge of SL (=translator)
- Perfect command of TL (=translator)
- Specialised domain knowledge (=translator)
- Word-processing skills (=translator)
- Tolerance! (?)
- Positive attitude to MT (?)

Proposal for training

- Additional Skills required (O'Brien 2002):

  - Knowledge of MT
  - Term management skills including
    - MT dictionary coding (RBMT)
    - Corpus quality assessment (SMT)
    - Term management (exchange formats, tools etc)
  - Pre-editing/CL skills
  - Basic Programming skills (e.g. macros for automated text correction)

To Finish: 1 Important Question

- Why do translators dislike post-editing?
- Mismatch between the nature of post-editing and the profession of translating?

Overview

- Impact of PE on the translation profession

- Implications for quality

- Implications for productivity

- Alternatives to post-editing

- Post-editing guidelines

- Example of post-editing guidelines

- Post-editing training

- Future perspectives

Impact of PE on the Translation Profession

Negative aspects:

- New activity, still relatively unknown to many translators

- Not much information or training available

- New methodologies and tools to learn

Impact of PE on the Translation Profession

Negative aspects (cont.):

- Possibly lower rates

- Expected higher productivity

- Less room for creativity, due to the nature and limitations of PE

Impact of PE on the Translation Profession

Positive aspects:

- Acquisition of knowledge and experience about MT in general

- PE as a productivity tool

- Possibility to avoid repetitive tasks and translations

- A new set of skills that can be developed and offered by translators

Implications for quality

Negative aspects:

- Expected level of language quality determined by each project (i.e. types of errors to be corrected or not)

- Potentially difficult adaptation (having to overlook some of the errors instead of correcting them)

- Danger of getting used to the output produced by the computer

- Tools to check if guidelines have been followed do not exist yet

Implications for quality

Positive aspects:

- Standardisation of terminology, leading to fewer inconsistencies

- Possibility to automatically correct repetitive or predictable error patterns

- Fine-tuning of MT engines over time, leading to better quality of the output

Implications for productivity

Negative aspects:

- More limited scope of changes: expected daily productivity for PE higher than for translation or review

- Expectation of higher productivity (5000 words post-edited per day) => more pressure on translators

- Very high daily productivity levels may not be sustainable

Implications for productivity

Positive aspects:

- PE could be used by translators in their favour, as a productivity tool

- Possibility to combine PE with other activities to avoid a negative impact on individual skills and to increase the overall productivity and job satisfaction

Alternatives to post-editing

Translators willing to explore new areas can get involved in different aspects of MT:

- Maintenance of MT dictionaries

- Creation of regular expressions to fix common error patterns

- Assessment of PE quality performed by third parties

Post-editing guidelines

Negative aspects:

- No internationally adopted standard guidelines

- Each company tends to have their own PE guidelines

- Examples and explanations not always clear

- Need to find a balance between vague guidelines vs. too many details

Post-editing guidelines

Positive aspects:

- As PE becomes more widespread, it is likely that a consensus will be reached for the development of internationally adopted guidelines

- Translators can/should provide feedback for the improvement of existing PE guidelines

Example of post-editing guidelines

- Degree of subjectivity: changes considered essential by a reviewer might be considered preferential by another one
- Preferential changes (such as stylistic changes) should be avoided
- Information not contained in the source text should not be added
- Information contained in the source text should not be omited

Post-editing training

- Courses not available for those who already have a translation degree

- Companies provide only guidelines, not training

- New trend: a few universities beggining to add PE training to translation courses

Future perspectives

- More training (universities, courses) should be available to prepare translators for PE

- Better PE guidelines should be developed, leading to international standards

- Demythification of post-editing – incorporation to other activities performed by translators

Post-Editing: The Content Owner's Perspective

Post-Editing: the content owner's perspective

- Why MT and PE?

- How were MT and PE implemented?

- Challenges

- Next steps

Why use MT?

- Increase productivity (higher throughputs) by pre-translating (MT) all source content

- Reduce time to market (TTM) through increased productivity

- Lower localization cost

- Increase terminology consistency by enforcing terms through MT dictionary
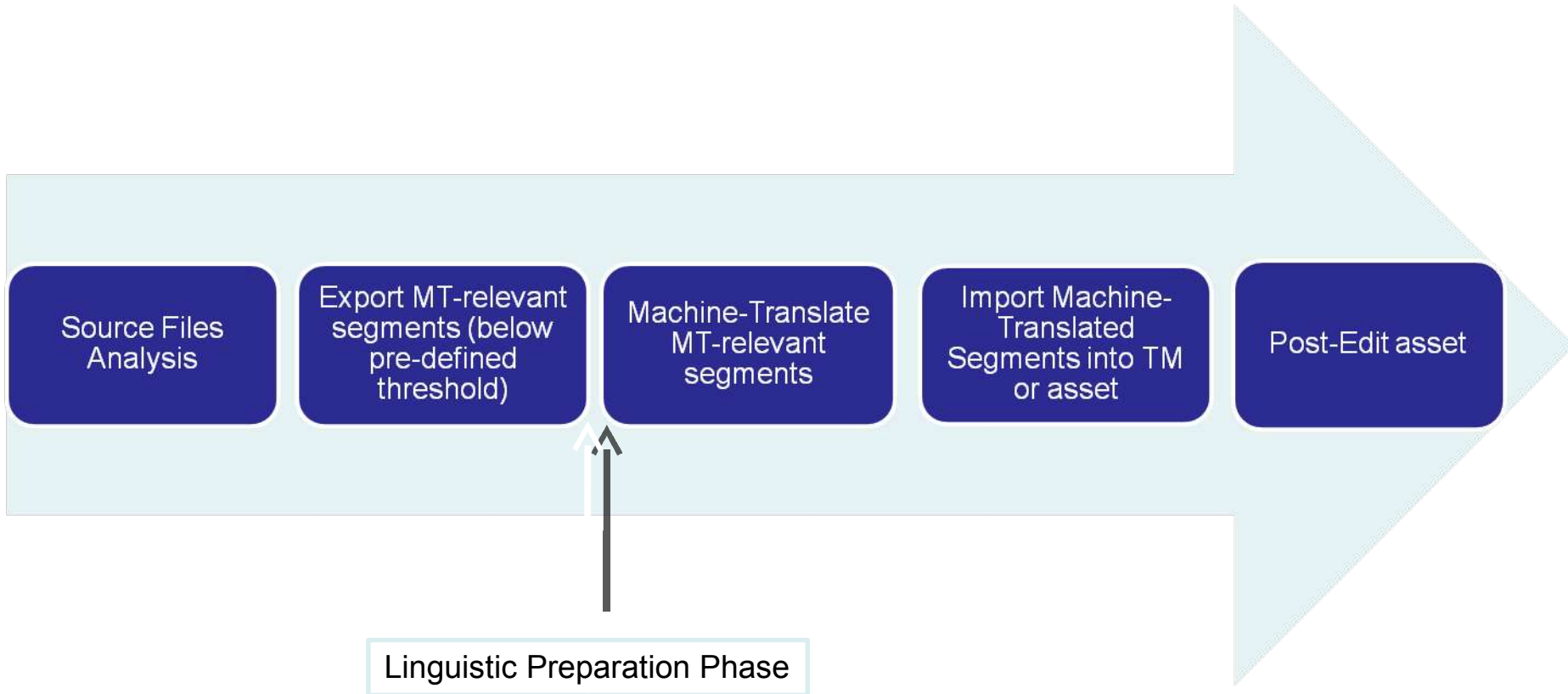
Initial findings

- Increased translation productivity proven

- One system needed to cater for 7 languages
  - SYSTRAN: Rule-based MT system (RBMT) chosen

- RBMT system's scope limited to internal rules and lexicon:
  - Internal dictionaries must be fed with domain-specific terminology

- Post-Editing (PE) required to bring MT output to commercial quality

How is MT technology implemented?

Source Files Analysis	Export MT-relevant segments (below pre-defined threshold)	Machine-Translate MT-relevant segments	Import Machine-Translated Segments into TM or asset	Post-Edit asset

Linguistic Preparation Phase

The Linguistic MT Preparation Phase

1. Source Text Analysis

Source Control Check

Terminology Harvesting

Exclude unsuitable content

2. Terminology Preparation

Terminology Implementation and Testing (including correct SW references)

3. Testing and Refining

Research missing terms

4. Ready for PE

Hand off along with relevant instructions

Is content suitable?

Essential Tuning

Further Fine-Tuning

Release

Maintaining quality

- The following points must be addressed:
  - Design and use specific MT guidelines describing the acceptable level of quality (including examples)

  - Use MT-friendly content
    - Yes = Software, Documentation, Help, Support Information
    - No = Marketing, Advertising, content where regional adaptation (re-writing) is needed and lexical variation required

  - Ensure a close collaboration with Post-Editors: sample checks, post-partum evaluation feedback, constant alignment

  - Use reviews and integrate post-editors' feedback

Specific MT guidelines

- Subset of guidelines given to post-editors:
  - Style is not a primary consideration - even when repetitive or pedestrian, but information accuracy is.
  - All the words are probably present in the MT output, but possibly in the wrong order.

Specific MT guidelines

- Make changes according to these guidelines:
  - Make sure that all information is transferred accurately.
  - Modify what is grammatically deviant from an output of commercial quality.
  - Modify what is lexically essential for the understanding of the target text (wrong or nonsensical words and phrases).
  - Make sure that all links are localized.
  - Remember that the terminology has been imported into the MT user dictionaries and should be correct. There is no need to use synonyms for the sake of originality.
  - Finally: do not spend too much time over a style problem. If you cannot think straightaway of a way to improve the output, leave it unchanged (there is no point in trying a few alternatives and reverting eventually to the initial suggestion).

Finding appropriate source content

- Requirements for Product documentation:
  - Consistent Key Terminology
  - Focus on Features and Product Names (enforced through Terminology Preparation/MT)
  - Correct Software References
  - Translated content correctly reflects source

  content
  - No negative impact on comprehension

- Our quality criteria accept less variance in style and a certain degree of repetitiveness

(PE) Quality Measures

- Human evaluation metrics
  - Four categories: Excellent, Good, Medium, Poor

- Automated evaluation metrics
  - General Text Matcher (GTM)
  - Precision, recall, and their average, the F-measure (taking word order into account)
  - The unigram-based F-measure has good correlation with human judgments

Human evaluation metrics

- Excellent:
  - Your understanding is not improved by the reading of the ST because it 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 metrics

- 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 metrics

- 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 metrics

- 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 Parts Of Speech. 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).

Measuring (PE) Quality

source segment	refined segment	post-edited segment	GTM score	SymEval (EGMP)	Vendor Editing Quality	Comment
Thereafter, the reports display the description instead of the numeric code.	**Danach zeigen die Berichte die Beschreibung anstelle** vom numerischen Code an.	Diese wird in den Berichten dann anstelle des Fehlercodes angezeigt.	0.0952381	G	over-edited	**Danach zeigen die Berichte die Beschreibung anstelle** des Fehlercodes an.
The products that you have installed determine which agents and plug-ins appear in the Software Catalog.	**Die Produkte, die Sie installiert haben, legen fest,** die Agenten und Plugins im Softwarekatalog erscheinen.	Die von Ihnen installierten Produkte legen fest, welche Agents und **Plug-ns** im Softwarekatalog angezeigt werden.	0.3054093	M	over-edited	**Die Produkte, die Sie installiert haben, legen fest,** welche Agents und Plugins im Softwarekatalog angezeigt werden.
The service is not currently installed on the site server.	Der Dienst wird derzeit nicht auf dem Site-Server installiert.	Der Dienst **ist** derzeit nicht auf dem Site-Server installiert.	0.795808	G	correctly edited	
It may take several hours for the changed permissions to become effective.	**Er** kann einige Stunden in Anspruch nehmen, damit die geänderten Berechtigungen wirkungsvoll werden.	**Er** kann einige Stunden dauern, ehe die geänderten Berechtigungen wirksam werden.	0.5457925	M	under-edited	s ...

62

Measuring (PE) Quality

Editing Practice Q3
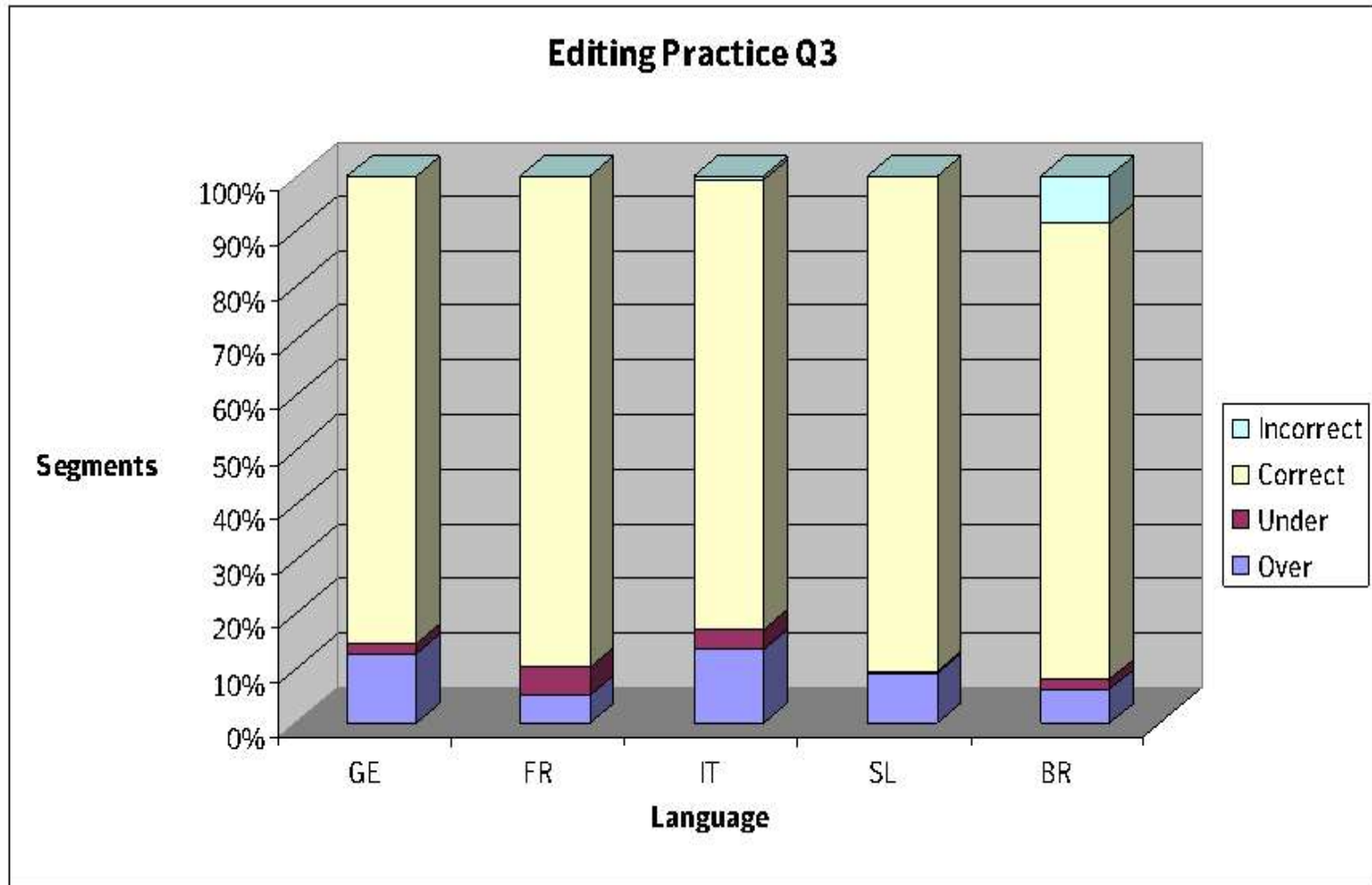
Measuring (PE) Quality

German GTM - Symeval
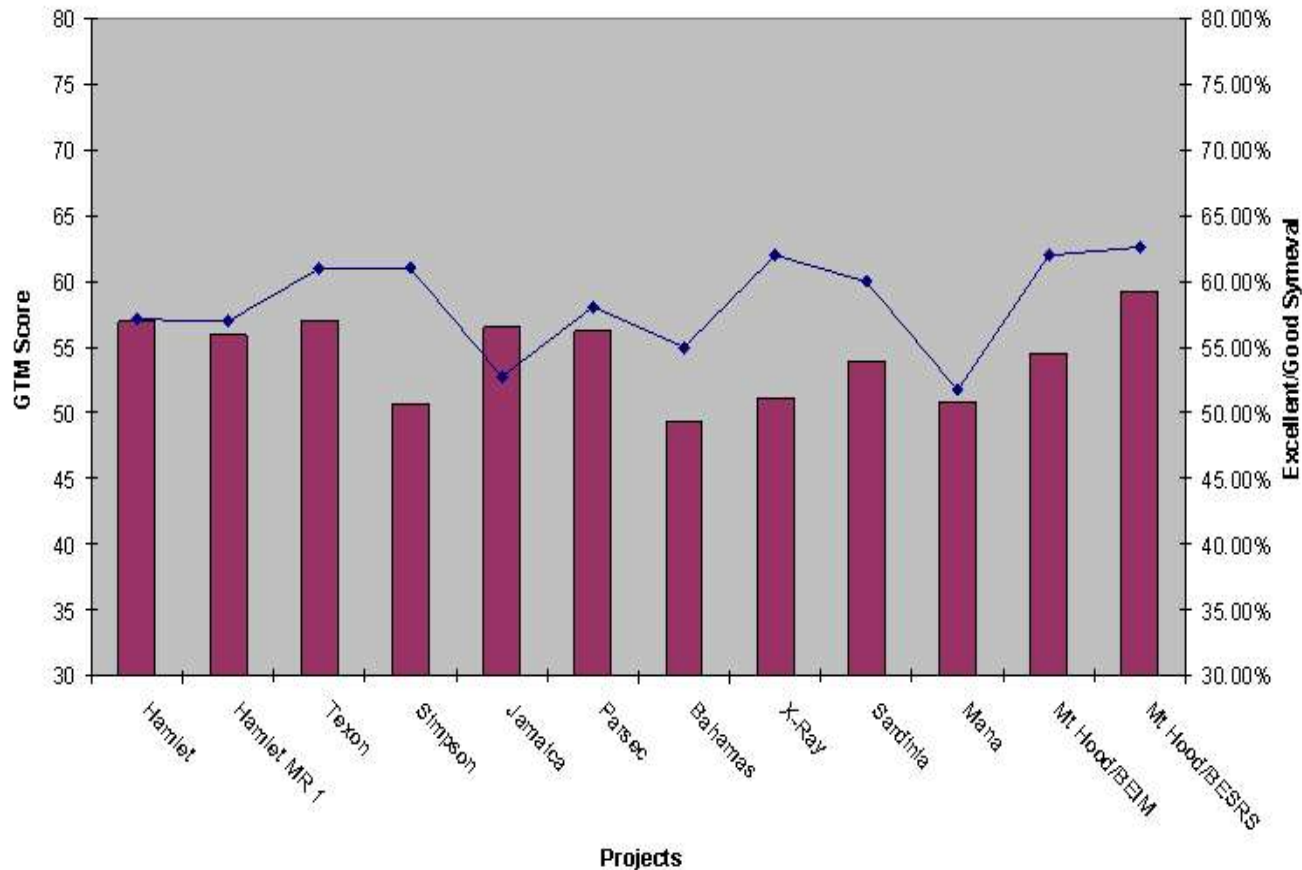
Impact of source compliance on MT quality

Source words	MT quality	Evaluation type	acrocheck™ project score
1083	Excellent	Human	28
3677	Good	Human	79
2546	Medium	Human	118
2129	Poor	Human	150
10972	Greater than 0.6 GTM scores	Automatic	64
9926	Less than 0.6 GTM scores	Automatic	147

Integrating post-editors' feedback

- Feedback gathering is a separate task

- Unstructured feedback can be difficult to process

- (Unstructured) feedback can be difficult to implement
  - No access to internal SYSTRAN rules

- Difference between specific issues (low coverage) and patterns (high coverage)

Dealing with frequent and frustrating errors

- Frequent MT errors require repetitive and tedious post-editing effort
  - Word level or phrase level
  - Specific errors vs. repetitive errors

- RBMT system is based on rules so output is predictable (to some extent)

- Post-editors are willing to correct errors as long as the MT system used is perceived to be learning from their work

Automated Post-Editing

- Global find and replace can be used to address frequent issues automatically:
  - Word reordering example (file names):
  - Standard SYSTRAN output: Die Local.cfg Datei unterstützt nicht Sprachen, die DoppeltByte Zeichensätze verwenden.
  - Search pattern: (?<!\boder)(\b\w+?\.\w{1,5}) (Datei\b)
  - Replace pattern: $2 $1
  - Standard SYSTRAN output + APE: Die Datei Local.cfg unterstützt nicht Sprachen, die DoppeltByte Zeichensätze verwenden.

Statistical Post-Editing

- Limitations of current APE step:
  - Find and replace rules are hand-written; complex rules are error-prone
  - High precision but limited coverage

- Possible solution
  - Learn from PE activity and train statistical post-editing engine on aligned MT/PE content

Statistical Post-Editing

Source Segment	SYSTRAN 6.06	SYSTRAN SPE	REFERENCE (from TM)	GTM (v6)	GTM (SPE)	GTM Diff	Source Words
The AOFO files are installed on the remote computer in the following directory:	Les fichiers d'AOFO sont installés sur l'ordinateur distant dans le répertoire suivant :	Les fichiers AOFO sont installés sur l'ordinateur distant dans le répertoire suivant :	Les fichiers AOFO sont installés sur l'ordinateur distant, dans le répertoire suivant :	0.713	0.824	0.111	13
To delete a media server group	Pour supprimer un groupe de serveur de supports	Pour supprimer un groupe de serveurs de supports	Pour supprimer un groupe de serveurs de supports :	0.747	0.941	0.194	6
Select the Redirect Microsoft SharePoint sets check box.	Sélectionnez la case à cocher de Rediriger les jeux Microsoft SharePoint.	Cochez la case de Rediriger les jeux Microsoft SharePoint.	Cochez la case Rediriger les jeux Microsoft SharePoint.	0.668	0.843	0.175	8

Statistical Post-Editing challenges

- Use clean data
  - What does "clean" mean?
  - How do we clean TMs?

- Define domain boundaries
  - When is a new model required?

- Limit degradations
  - Terminology
  - Meaning
  - Tags

Finding Post-Editors

- 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

Next steps

- Adding statistical post-editing (SPE) to raise output quality further

- Address remaining structural rule issues with MT provider

- Establish more structured feedback mechanism between post-editors
  - Maintain a strong relationship with PE vendors

- Campaign for development of Post-Editor's and MT Linguist's professional profile

Post-Editing MT Output
Hands-On Practical Task

Summary & Conclusion

- Translation and Post-Editing are different tasks
- Translators' views of post-editing are largely negative because…

Summary & Conclusion

Image of Translation:

- Highly skilled

- Takes time

- Quality is important

- Paid reasonably well

- Is a skilled task

- Is rewarding

Image of Post-editing:

- Almost anyone can do it, even a machine

- Do it as fast as you can

- Quality is not so important or we want the same quality for less money and time

- Constant downward pressure on pay rates

- Not as skilled

- Is tiring, tedious, unrewarding

Summary & Conclusion

- The negativity can be addressed by:

    - More dialogue between publishers and translators
    - Making translators stakeholders in the MT process
    - Standardising guidelines
    - Having realistic throughput and quality expectations
    - Mixing post-editing tasks with translation
    - Better training for post-editors