Reliably Assessing the Quality of Post-edited Translation Based on Formalized Structured Translation Specifications

Alan K. Melby
Brigham Young University
Department of Linguistics and English Language
akmtrg@byu.edu

Jason Housley
Brigham Young University
Translation Research Group
housleyjk@gmail.com

Paul J. Fields
Brigham Young University
Continuing Education
pjfphd@byu.net

Emily Tuioti
Brigham Young University
Department of Linguistics and English Language
emily.tuioti@gmail.com

Abstract

Post-editing of machine translation has become more common in recent years. This has created the need for a formal method of assessing the performance of post-editors in terms of whether they are able to produce post-edited target texts that follow project specifications. This paper proposes the use of formalized structured translation specifications (FSTS) as a basis for post-editor assessment. To determine if potential evaluators are able to reliably assess the quality of post-edited translations, an experiment used texts representing the work of five fictional post-editors. Two software applications were developed to facilitate the assessment: the Ruqual Specifications Writer, which aids in establishing post-editing project specifications; and Ruqual Rubric Viewer, which provides a graphical user interface for constructing a rubric in a machine-readable format. Seventeen non-experts rated the translation quality of each simulated post-edited text. Intraclass correlation analysis showed evidence that the evaluators were highly reliable in evaluating the performance of the post-editors. Thus, we assert that using FSTS specifications applied through the Ruqual software tools provides a useful basis for evaluating the quality of post-edited texts.

1 Research Question

The progression of globalization has produced an ever increasing demand for materials to be translated. Moreover, there are now an insufficient number of highly skilled translators to handle the total demand for translation services. In other words, the world has moved beyond the point when there were more than enough translators to meet the demand (Hutchins, 2007), into an era where the need for translation has made the use of machine translation (MT) widespread.

One practical application of MT that allows it to be applied to a variety of situations is post-editing, that is, the addition of a human editor to correct raw MT output to meet a set of requirements that the MT system would not be able to fully meet on its own. Post-editing presents a set of problems and challenges, not the least of which is assessing how well human editors can perform as post-editors. If the target-text production performance of human post-editors cannot be assessed reliably, then other measures, such as post-editing speed, are meaningless. Therefore, this study was focused on reliability.

As a step toward a general answer, we asked a specific research question: How reliably can non-expert human evaluators assess the quality of post-edited machine translations given three conditions:

1. The initial English target text was generated by a free and publically available machine translation system;
2. The source text was a medium difficulty (Interagency Language Roundtable (ILR) Level 2) document in Japanese; and
3. The evaluators assessing the performance of the post-editors were given a rubric based on a set of structured translation specifications?

Here performance is defined as "the ability to produce a target text that meets agreed-upon project specifications." Therefore, a quality target

text is one that meets the specifications. Note that this is a functional view of quality, not an absolute view of quality, which would require a target text to be completely accurate and perfectly fluent, regardless of audience and purpose.

To serve as the basis for assessing translation quality, we used the following specification-based definition developed within our research group:

> A quality translation achieves sufficient accuracy and fluency for the audience and purpose, while, in addition, meeting all other negotiated specifications that are appropriate to meet end-user needs.

This novel definition, which goes beyond a strictly industry-neutral, ISO 9000 approach to quality, makes the implicit claim that translation quality cannot be assessed without pre-determined specifications about the *process* of translation and the resultant *product*. Building on this definition, even a target text that is somewhat awkward yet usable could be a quality translation, if it fully meets the agreed-upon specifications. To this end, we developed and tested a methodology for formalizing structured translation specifications to support post-editing assessment. This methodology involves two software applications which we have developed: the Ruqual Specifications Writer, which aids in the authoring of post-editing project specifications, and the Ruqual Rubric Viewer which provides a graphical user interface for filling out a machine readable rubric file. Since this project uses a rubric to assess quality, the name of the software is Ruqual, which is a blend of "rubric" and "quality."

2   Previous Work

Up until the last ten years, very little research had been done on the subject of post-editing (Allen, 2003). However, advances in MT have prompted an increase of interest in the subject (Alves, 2003; Guerberof, 2009; O'Brien, 2002, 2005, 2011; Ramos, 2010; Rios et al., 2011; Specia et al., 2009, 2011).

Most previous studies have focused on post-editing effort and the quality of the raw MT target text. This post-editing effort may be defined in a number of ways; most notably the work of Krings (2001) divides post-editing effort into three categories: temporal, technical, and cognitive. Temporal effort measures how long it takes the post-editor to finish editing the target text, whereas technical effort measures the changes made to the MT-generated text. The cognitive load is difficult to measure because techniques designed to measure the thought processes of translators/post-editors often make the task of translating more difficult (O'Brien, 2005). However, O'Brien has found that a measure of cognitive effort can be obtained from other measures, such as comparing the differences between the changes of multiple post-editors and accepting pauses in the timed record of changes as an indication of increased cognitive activity (2005). Specia and Farzindar (2010) have also developed a system of measuring expected post-editing effort so that companies can estimate whether a particular machine translated text is worth sending to post-editors.

Measuring the effort—or the time it takes—to post-edit a text assumes that the post-edited target text has sufficient and similar quality in all cases compared and that the post-editor followed all of the procedures necessary for the project. Measuring strictly "the time it takes to post-edit a text" must be based on a definition of translation quality and a method of measuring it. The "transcendent" view of quality assumes that every translation exhibits the same high levels of accuracy and fluency. Once it is recognized that translation quality is not transcendent but relative, measuring post-editing effort is only useful when the specifications are the same. One machine translated text may be useless for a particular set of specifications while being suitable for another set of specifications. The amount of effort necessary to successfully post-edit a text in accordance with a set of specifications will probably change when the specifications change, even if the source, raw MT text, and post-editor are the same. Hence any measure of post-editing effort must be based on a foundation of defining and measuring quality applicable to raw and post-edited translation.

The approach to measuring the quality of post-editing espoused in this project rejects the transcendent view of quality. This project provides a way to organize the information necessary to clarify which quality factors are relevant to a particular post-editing project. One study may investigate how much time it takes to post-edit raw MT output into documentation strictly for internal use in a software company. Another study may involve producing translations for general public consumption. If the specifications are not explicitly stated, then the results of one study may be misinterpreted to be directly relevant to the subsequent project.

Moreover, if the specifications are stated but not organized, a comparison of two studies would be difficult. If one study concluded that post-editing should take less than 10 minutes to be cost-effective, then such a measure might discriminate against good post-editors who take 20 minutes to post-edit a text in a different study. The reason for the difference in time may have less to do with individual post-editors than it does with the project specifications. It takes more effort to post-edit a text for a general audience than it does for a small audience that has more background knowledge and is more tolerant of errors. Explicit, structured project specifications and quality measures based on them are needed to complement on-going research in post-editing effort.

Translation specifications and quality measures must not only be explicit, they must be reliable. If evaluators cannot agree on the quality of a translation, human, raw machine, or post-edited, then the notion of quality is useless.

Colina has proposed a rubric for assessing the quality of human translation in a healthcare environment. Colina's approach is compatible with the definition of quality used in this project (Colina, 2008). The TAUS Labs have recently developed a Dynamic Quality Evaluation Framework (TAUS Labs, 2012) that may be compatible with the approach in this paper, but it is not available to the public. The EU-funded QT Launchpad project (2012) is also working on translation quality assessment. Collaboration among these related efforts would be beneficial to the translation industry.

3 Structured Specifications

This project proposes a format for formalizing structured translation specifications in order to support post-editing assessment. The basic components of the formalized structured translation specifications (FSTS) format are derived directly from the recently published ISO document ISO/TS 11669 (General Guidance -- Translation Projects) and the status descriptors in the Linport STS format (Linport, 2012; Melby et al., 2011), based on the earlier Container Project.

As shown in Table 1, the top-level categories in the FSTS format are Linguistic (divided into Source Content Information and Target Content Requirements), Production tasks to be performed during the project, Environment requirements, and Relationships between the requester (sometimes called the client, although "client" is ambiguous) and the translation service provider.

A. Linguistic [1–13]
    Source content information [1–5]
        [1] textual characteristics
            a) source language
            b) text type
            c) audience
            d) purpose
        [2] specialized language
            a) subject field
            b) terminology [in source]
        [3] volume (e.g. word count)
        [4] complexity (obstacles)
        [5] origin [of the source content]
    Target content requirements [6–13]
        [6] target language information
            a) target language
            b) target terminology
        [7] audience
        [8] purpose
        [9] content correspondence
        [10] register
        [11] file format
        [12] style
            a) style guide
            b) style relevance
        [13] layout
B. Production tasks [14–15]
        [14] typical production tasks
            a) preparation
            b) initial translation
            c) in-process quality assurance
        [15] additional tasks
C. Environment [16–18]
        [16] technology
        [17] reference materials
        [18] workplace requirements
D. Relationships [19–21]
        [19] permissions
            a) copyright
            b) recognition
            c) restrictions
        [20] submissions
            a) qualifications
            b) deliverables
            c) delivery
            d) deadline
        [21] expectations
            a) compensation
            b) communication

Table 1. List of 21 formalized structured translation specifications (FSTS).

The Source category describes the source content. The Target category is concerned with the

language into which the material is translated and various other requirements for how the translation is to be carried out. The Production category lists the tasks to be performed during the translation project. The Environment category includes any technology that must be used, all reference materials that must be consulted by either software or human, and any security requirements, such as the need to conduct the work in a particular location. The Relationships category refers to the project expectations and work requirements for all team members, including the post-editor.

The five FSTS categories (Source, Target, Production, Environment, and Relationships) arrange the 21 translation parameters into logical groups, as is shown in Table 1. All parameters have two attributes that assist in determining its importance for a particular project: Status and Priority. The value of the status attribute can be one of four options: Incomplete, Not Specified, Proposed, and Approved.

One of the key components of the development of our methodology was the use of "Directives," or prose descriptions of specific instructions that could be assessed by an evaluator during the translation workflow process. Our methodology makes a distinction between *process-oriented* directives, or instructions to the post-editor concerning the steps he or she should follow while modifying the translation, and *product-oriented* directives, which relate to the final state of the target text.

The Ruqual Specifications Writer allows for the development of post-editing project specifications which are both process- and product-oriented. In its design, several parameters and attributes in the FSTS take a list of directives as their value. A directive has two attributes: Request and Priority. The request consists of natural language content describing the post-editor's task. The priority indicates how important it is that the request be fulfilled. Each directive can be modified based on project specifications.

The FSTS naturally support the generation of a rubric for evaluating post-editing that can handle a high degree of variability in the specifications of various projects. The rubric developed in our methodology, the Ruqual Rubric Viewer, is composed of a list of directives pertaining to the top-level FSTS categories previously mentioned. When using the rubric for assessment, an evaluator simply specifies whether a particular directive was fulfilled or not. If it was fulfilled, the value of the priority is awarded; otherwise, no points are awarded. The final score for a given category is the number of points received divided by the number of points possible.

With these tools it is possible to write translation project specifications and consequent rubrics that will allow non-experts to quickly and straight-forwardly assess the translation quality of post-edited texts.

The software developed for this research is hosted as an open source Google Code project at: http://code.google.com/p/ruqual/. Collaboration with other projects and extensions of the software are welcome.

4 Study Design

In the structured assessment of our methodology and accompanying Ruqual software, a Japanese source text was translated by Google Translate (Google, 2012) to produce a raw machine translated text. With attention to real post-editing data, five different potential post-edited texts were developed from the machine translation to simulate the work of five fictional post-editors, whom we named Editors A-E. Five different scenarios describing the translation process experiences of the five fictional post-editors were also developed. Errors were purposefully introduced into the post-editors' scenarios such that some violated process-oriented directives while others violated product-oriented directives. The source text, raw machine translation, and five post-edited texts are shown in the appendix.

Space limitations for this paper do not allow inclusion of the full FSTS used in this experiment. Some of the key elements of the specifications were that the translation was for a general audience, that it should be fluent and not obviously a translation, that a particular bilingual glossary must be used, and that the translation product must be delivered by a certain date and in a particular format. The authors are quite aware that in many post-editing environments the translation can be less than fluent and can be an obvious translation.

The definition of a non-expert assessor in this study was an individual who was:

1. A non-native speaker of the source language, but who had studied the source language for at least two years;
2. A native speaker of the target language;
3. A high school graduate or higher; and
4. A novice in the professional translation industry.

The focus of the study was reliability, that is, how consistent the non-expert assessors were with each other. In order to assure that the assessment was also reasonably valid, the assessments were compared with those of an expert assessor.

In total, 17 non-experts provided complete assessments of the five work products of the simulated post-editors A-E. The data were gathered via a questionnaire that was accessible from the Ruqual website (ruqual.gevterm.net).

The first portion of the questionnaire asked for some basic demographic information, and then participants were directed to an instructions page that included four items:

1. A video demonstrating the Ruqual Rubric Viewer;
2. A text walk-through with the same content as the video in case the participant lacked the software necessary to display the video;
3. A location from which to download the source materials and terminology files; and
4. A link to a zipped version of the Ruqual Rubric Viewer.

Participants were instructed to familiarize themselves with the software and source materials before proceeding with the questionnaire.

The second portion of the questionnaire presented the work of each of the five fictional post-editors in random order. The evaluators were instructed to assess the performance of each post-editor independently.

5   Analysis

The analysis of the data was twofold. First, we examined reliability among the non-expert evaluators as well as the concordance of the non-expert evaluators with an expert evaluator. Second, we examined the similarities and differences among the five fictional post-edited target texts and a human translation provided by an expert human translator.

Figure 1 shows a comparative box plot of scores given on a scale ranging from 0.0 to 1.0 for the five post-editors ordered by median score.
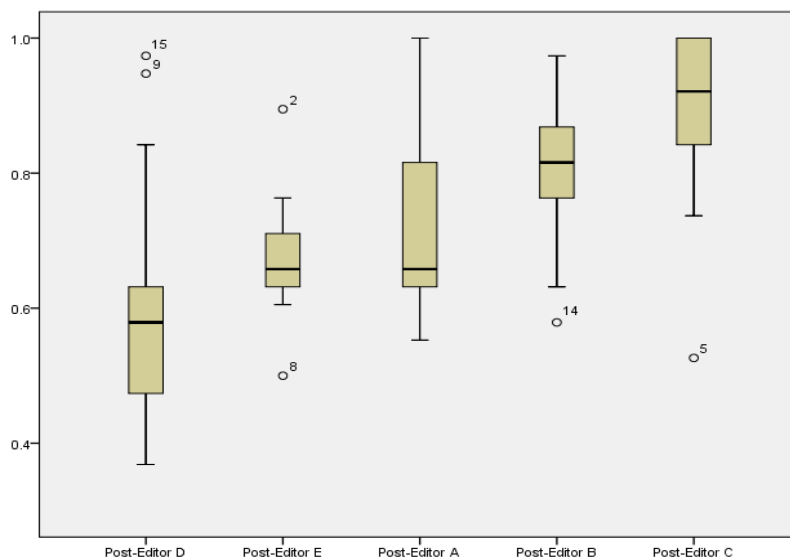

Figure 1. Scores Given by the Non-Expert Evaluators for Each Post-Editor Ordered by Median Score.

Overall, non-expert evaluators gave the highest scores to post-editor C and the lowest score to post-editor D, which is as would be expected based on the number of specifications these post-editors were designed to violate. Assessments for post-editors E spanned a much smaller range around the median than the scores for post-editors A, B, C, and D. It should be noted that no post-editor received a score lower than 0.35 from any grader, which could be due to the fact that all of the target texts were, in our opinion, reasonably grammatically correct.

In order to test reliability, which was the focus of the research question, we calculated the two-way random Intraclass Correlation Coefficients (ICC) as described by Shrout and Fleiss (1979).

ICC values can range from 0.0 to 1.0 analogous to percentages. The ICC is a measurement of the agreement between evaluators, or in other words, the percentage of variability in the scores that represents the quality of the post-editing. Using ICC values provides a measure of the agreement and consistency of the evaluations. The question of reliability in this case is not simply whether evaluators assigned the same relative scores to the post-editors, but to what degree they assigned the same scores. Since all 17 non-expert evaluators assessed all five fictional post-editors and these evaluators can be considered to be a sample of potential non-expert translation evaluators, the ICC values calculated utilized a two-way random effects model with evaluator effects and measurement effects. Table 2 shows the single and average ICC scores for the non-experts evaluators subdivided by rubric categories.

Target	0.167	0.773
Production	0.148	0.747
Environment	0.529	0.95
Relationships	0.607	0.963
Total	**0.426**	**0.927**

Table 2. Single and Average Intraclass Correlation Coefficients (ICC) for Non-Expert Evaluators.

The single ICC is a measure of the reliability of a single evaluator from this set of evaluators, if we were to accept his or her score alone. The average ICC indicates at the percentage of agreement among the evaluators with each other as a group.

The key statistic in Table 2 is the average ICC for the total score, which is ICC (2, 17) = 0.927. This is a strong indicator that the non-expert evaluators were reliable as a group. However, the single ICC for the same category was only ICC(2, 1) = 0.426 suggesting that if one evaluator was to be selected from this set, he or she would be expected to be reliable only about 43% of the time.

Looking at the rubric categories, there appears to be a split between the Target/Production specifications and Environment/Relationships specifications. This is worth noting because the specifications as constructed for this research generally include product-oriented directives in Target/Production and process-oriented directives in Environment/Relationships. The evaluators might have had an easier time agreeing on whether a post-editor followed the specified processes than they did deciding whether a particular text sufficiently corresponded with another text.

In addition to reliability there is also the question of whether non-expert evaluators were assessing the post-editors in a manner similar to how an expert would do so. The expert evaluator's assessments are provided in Table 3 along with 95% confidence intervals for the non-experts.

	A	B	C	D	E
Expert Scores	0.447	0.684	0.763	0.315	0.605
Non-Expert Upper Confidence Limit	0.784	0.863	0.957	0.693	0.716
Non-Expert Lower Confidence Limit	0.648	0.755	0.829	0.514	0.627

Table 3. Expert Evaluator's Scores and 95% Confidence Intervals for the Non-Expert Evaluators' Scores.

The expert evaluator assessed post-editor C as the best and D as the worst, with post-editor B assessed as next best after post-editor D. The non-expert evaluators matched these rankings. Although the expert and non-experts reversed the rank order of post-editors A and E, it should be noted that there was no statistical difference between post-editors A and E for the non-experts. These data suggest that the goal of providing the evaluators with simulated post-editors showing distinct differences and a progression from higher quality to lower quality was achieved.

In addition, we calculated a coefficient of concordance for each non-expert evaluator and the expert evaluator. Two evaluators showed a strong concordance with the expert and ten evaluators showed a moderate concordance. One evaluator showed a weak concordance and four evaluators

showed little or no concordance. The four evaluators who showed no concordance with the expert, were also the evaluators who gave the most extreme evaluation scores (evaluators 2, 5, 9, and 15 in Figure 1). This suggests that these evaluators were perhaps less skilled or less trained (i.e., they did not go through the prescribed training) than the other evaluators.

However, none of the expert evaluator's scores fell within the confidence intervals of the non-experts' scores. In fact, the expert provided a lower score than the non-experts in all cases, indicating that the expert may have allowed the post-editors less leeway in evaluating their work products. Since the expert would have had a better understanding of the importance of following proper procedures and fully meeting the translation specifications, perhaps the expert was either more inclined to find fault with the performance of the fictional post-editors or was more aware of the failures present in the text and scenario.

As an addition to our study, a second expert was sought out to provide a human translation of the source text used in the study without reference to the raw machine translated text. In fact, the second expert was only given the source text and specifications. The purpose of requesting an expert human translation was to obtain a reference translation for the source text. Since none of the post-edited texts were intended to meet all of the specifications, it was worthwhile to identify how closely these fictional post-edited texts were to an actual human translation. If the human translation did in fact meet the specifications, and if post-editing is worthwhile, then the post-edited text should have been generally similar to the HT reference text.

In our comparison it appeared that the human translator took advantage of the flexibility provided by the specifications that allowed for some awkward sentences as long as the target text fulfilled its purpose for the intended audience. (The expert human translation is also shown in the appendix.) The human translator also rendered some sentences in a way that typically would be described as run-on sentences, but these sentences closely matched the flow of the source text. In fact, such sentences may facilitate automatic alignment and processing better than the sentence breaks provided by the machine translation. Overall, it appeared that post-editor C and the human translator were generally similar, but the requirement to not change sufficiently translated phrases in the initial machine translation could have limited the latitude of a post-editor.

6 Results

Overall, the research results support the hypothesis that non-expert evaluators can reliably assess the quality of fictional post-edited translations when taken as a group. This is a promising outcome since it shows that it is possible to obtain agreement about the quality of post-editing when using formalized structured translation specifications and multiple evaluators. Moreover, the fact that a majority (12 out of 17) of the non-expert evaluators showed at least moderate concordance with an expert evaluator suggests that there was evidence of the validity of the non-experts' evaluations.

Consequently, we assert that using FSTS specifications provides both a practical and realistic basis for evaluating the quality of post-edited texts. Therefore, if appropriate specifications are provided and structured via the Ruqual tools developed in this research, then evaluators can be expected to reach generally reliable and valid conclusions.

Finally, the fact that the text judged to be the work of the best post-editor was similar to the text produced by a human expert translator supports the assertion that post-edited MT text can be of sufficient quality to compete with HT alone.

7 Conclusions and future work

Overall, the research results provide evidence that non-experts can reliably assess the quality of post-edited machine translation relative to structured specifications. Further studies need to be conducted using the same approach to determine the effect of more training for the assessors and the effect of more specific rubrics.

More important than the particular experiment described in this paper are the methodology and tools used in the experiment. We anticipate working with other teams to conduct a series of experiments using various source texts, alternative machine translation systems, and widely varying project specifications, but applying the same methodology as in this study, and including the same standard set of translation parameters from

ISO/TS 11669 as well as the same definition of translation quality for human and machine translation. This will allow meaningful comparison of results. If there are problems with the translation parameters, suggestions should be made to the ISO project 11669 team as they prepare the next version. Extensions to ISO 11669, such as much more detailed and narrow assessment specifications, can be developed. Other work in post-editing, such as measures of effort, needs a widely used, reliable approach to translation quality assessment.

References

Allen, Jeffrey. 2003. Post-editing. In H. Somers (Ed.), *Computers and translation: A translator's guide* (297-317). Philadelphia: John Benjamins.

Alves, F. 2003. *Triangulating Translation*. Amsterdam: John Benjamins.

Colina, Sonia. 2008. Translation Quality Evaluation: Empirical Evidence for a Functionalist Approach. *The Translator*, 14: 97-134.

Google. 2012. Google Translate. http://translate.google.com/. (Accessed 2012).

Guerberof, Ana. 2009. Productivity and quality in MT post-editing. Paper presented at the MT Summit XII Workshop: *Beyond Translation Memories: New Tools for Translators*. Ottawa, Ontario, Canada.

Hutchins, John. 2007. Machine Translation: A concise history. In C. Wai (Ed.), *Computer aided translation: Theory and practice*. Hong Kong: Chinese University of Hong Kong.

ISO/TS 11669: Translation projects. General guidance. 2012. ISO. Geneva, Switzerland.

Krings, Hans P. (Ed.) 2001. *Repairing texts: Empirical investigations of machine translation post-editing processes*. Kent, OH: Kent State University Press.

Linport. 2012. Linport: The Language Interoperability Portfolio Project. www.linport.org. (Accessed 2012).

Melby, Alan K., Arle Lommel, Nathan Rasmussen & Jason Housley. 2011. The Container Project. Paper presented at the First International Conference on Terminology, Languages, and Content Resources. Seoul, South Korea.

O'Brien, Sharon. 2002. Teaching Post-Editing: A Proposal for Course Content. Paper presented at the 6th EAMT Workshop: *Teaching Machine Translation*. Manchester.

—. 2005. Methodologies for Measuring the Correlations between Post-Editing Effort and Machine Translatability. *Machine Translation,* 19: 37-58.

—. 2011. Towards predicting post-editing productivity. *Machine Translation,* 25:197-215.

QT Launchpad. 2012. Motivation webpage. http://www.qt21.eu/launchpad/. (Accessed 2012).

Ramos, Luciana. 2010. Post-Editing Free Machine Translation. Proceedings from the Ninth Conference of the Association for Machine Translation in the Americas (AMTA): *From a Language Vendor's Perspective.* Colorado.

Rios, Miguel, Wilker Aziz & Lucia Specia. 2011. TINE: A Metric to Assess MT Adequacy. Paper presented at the 6th Workshop on Statistical Machine Translation. Edinburgh.

Shrout, Patrick E. & Joseph L. Fleiss. 1979. Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin,* 86: 420-28.

Specia, Lucia & Lucas Nuna Vieira. 2011. A Review of Machine Translation Tools from a Post-Editing Perspective. Paper presented at the 3rd Joint EM+/CNGL Workshop: *Bringing MT to the USER: Research Meets Translators.* Luxembourg.

Specia, Lucia & Atefeh Farzindar. 2010. Estimating Machine Translation Post-Editing Effort with HTER. Paper presented at the AMTA 2010 Workshop: *Bringing MT to the USER: MT Research and the Translation Industry*. Denver, Colorado.

Specia, Lucia, Dhwaj Raj & Marco Turchi. 2009. Machine translation evaluation versus quality estimation. *Machine Translation,* 24: 39-50.

TAUS Labs. 2012. About DQF webpage. http://www.tauslabs.com/dynamic-quality/about-dqf. (Accessed 2012).

Appendix

Source Text (from a 2009 source accessed June 2012 http://www.asahi.com/topics/%E3%82%A2%E3%83%83%E3%83%97%E3%83%AB.php):

解説

アップルとは

パソコンから携帯電話まで

マッキントッシュ（Macintosh）などのパソコンや携帯音楽プレーヤー・アイポッド（iPod）、携帯電話アイフォーン（iPhone）などを販売する、米コンピューターメーカー。世界中で直営店（アップルストア）やネット店を展開し、アイチューンズ（iTunes）ストアでは音楽や映画などの販売、アップストア（AppStore）ではiPhone向けのソフト販売も行う。世界市場ではマイクロソフトのウィンドウズで動くパソコンが9割以上を占め、同社のパソコンは数％のシェアしか得ていないが、斬新なデザインや使い勝手の良さなどでファンを獲得し独自路線を歩んできた。カリスマ的な経営者スティーブ・ジョブズ氏の言動が常に世界中で注目されることでも有名。

しかし彼は２０１１年10月に死去した。

Raw Machine Translated Target Text

Commentary
Apple and
From PC to mobile phone
Macintosh personal computer or portable music player such as iPod (Macintosh) (iPod), mobile phones to sell, such as iPhone (iPhone), the U.S. computer maker. Expand the net and shop (Apple Store), at (iTunes) store selling music and movies, also performs in software sales for the iPhone (App Store) Apple retail store in the world up. In the world market a personal computer running on Windows Microsoft accounted for more than 90%, PC company is not only get share of a few percent, has come a maverick won the fan, such as the difference between ease of use and innovative design. Well known for their words and deeds of Mr. Steve Jobs, the charismatic owner is always attention throughout the world. He died in October 2011, however.

Post-Edited Texts and Scenarios

Post-Editor A

Scenario:
When the post-editor received the source materials, he/she sent them to an acquaintance who converted them to MS Word 2003 format because the post-editor did not own MS Word 2007 or greater. After finished post-editing, the post-editor sent the post-edited target text to the acquaintance to have it converted to .docx format, but this delayed the project meaning that the finished product was returned on March, 28 2012.

Target Text:
Company Description
Apple
From PC to mobile phone
Apple is the American computer maker that markets the Macintosh computer series, the iPod MP3 player, and the iPhone smart phone. Around the world, Apple has opened an internet store and company stores (Apple Stores). Music and movies are sold at the iTunes store whereas at the App Store software for the iPhone is sold. In the world market personal computers running on Microsoft Windows account for more than 90%. However, Apple only controls a few percentage points of market share, but Apple has taken its own approach to consumer electronics and won fans for its ease of use and innovative design. The company is well known for the words and deeds of Mr. Steve Jobs, the charismatic CEO, who was always earning the attention of the world. However, he passed away in October of 2011.

Post-Editor B

Scenario:
After getting the source materials, the post-editor verified via email that he/she had received all of the source materials and would follow all of the specifications. The post-editor owns and uses MS Word 2007 and Acrobat Reader. The post-editor completed and returned the post-edited text on March, 24 2012. He/she also included the source text and machine translation. Finally, the post-editor deleted all of the project files after hearing back that that project was complete.

Target Text:
Commentary
What's up with Apple?
From computers to mobile devices
Apple Inc. is a consumer electronics manufacturer that sells personal computers (Macs), iPod portable music players, and mobile phones such as the iPhone at pure bead prices. They also have retail stores (Apple Stores) and an internet store. In the iTunes store they charge extra for music and movies, and at the App Store they sell sub standard software for the iPhone. In the world market, machines running on Windows make up more than 90% of all machines. Apple Inc., on the other hand, only controls a small percentage of the market. Apple Inc likes to do its own thing and has won some fans for its perceived ease of use and innovative design. Ultimately, the company gets a lot of attention for the words and deeds of Mr. Steve Jobs, the ostentatious owner, but he died in October of 2011.

Post-Editor C

Scenario:
Upon receipt of the source materials, the post-editor confirmed that he/she had received all of the source materials and would follow all of the specifications. The post-editor used Acrobat Reader to view the appleTerms.pdf file and MS Word 2010 to edit the target text. When finished post-editing, the post-editor returned the post-edited text alongside the source and machine translation on March, 23 2012. After completing the project, the post-editor deleted all of the related texts.

Target Text:
Company Overview
Apple
From PC to mobile phone
Apple is a U.S. computer maker that sells products such as the Macintosh personal computer, iPod portable music player, and mobile phones such as the iPhone. Apple has opened an internet store and company stores

(Apple Stores) across the world. At the iTunes store they sell music and movies, while at the App Store they sell programs for the iPhone. In the world market personal computers running on Microsoft Windows account for more than 90% of the market, whereas Apple has only a small percentage of market share, but Apple has marched to the beat of its own drummer and won fans via ease of use and innovative design. The company is well known for the words and deeds of Mr. Steve Jobs, the charismatic owner, who is always getting attention throughout the world. However, he died in October of 2011.

Post-Editor D

Scenario:
The post-editor never agreed to the specifications. Instead, the post-editor returned the post-edited target text (by itself) on March, 26 2012 before a formal agreement was complete. After the project was complete, it was discovered that the post-editor had posted a copy of the source materials and his/her translation on his/her blog for the public to write comments about.

Target Text:
Commentary
And Apple?
From personal computers all the way to mobile devices: Apple Inc. is in the business of consumer electronics. They make Macs, iPod portable music players, and the iPhone intelligent cell phone. They expanded their net shop and retail stores (Apple Stores); in the iTunes store they offer tunes and flicks, and at the App Store they have a place for applications that run on the iPhone. In the world market a personal computer generally is running on Microsoft Windows, which accounts for more than 90%. Apple Inc. is a maverick to its fans, who love its innovative design and dang good usability. The company is well known for its attention getting owner Mr. Steve Jobs, but he died in October of 2011.

Post-Editor E

Scenario:
Because the post-editor did not own MS Word (and did not realize that he/she could download Acrobat Reader for free), he/she sent the source materials to a friend

asking for help. The friend was late returning the post-edited translation and did not include the original source text and raw machine translation. The post-editor hurriedly returned the finished post-edited translation on March 26, 2012, but he/she forgot to delete any of the project documents after the project was finished.

Target Text:
Company Overview
Apple
From PC to mobile phone
Apple is the U.S. computer maker that sells the Macintosh personal computer, iPod portable music player, and mobile phones such as the iPhone. They have company stores (Apple Stores) the world over and an internet store; at their iTunes store they sell music and movies, and at the App Store they sell software for the iPhone. In the world market, computers running Microsoft Windows account for more than 90%, while Apple has only a small percentage of the market, but Apple has followed its own path and won fans for its ease of use and innovative design. The company is well known for the words and deeds of Mr. Steve Jobs, the charismatic CEO, who is always getting attention throughout the world. He died in October of 2011, however.

Expert Human Translated Reference Text:

Commentary
Apple
From Personal Computers to Mobile Phones
Apple is a US computer manufacturer that sells Macintosh and other personal computers, the portable music player iPod, the mobile phone iPhone, and other products. Apple operates Apple Store outlets worldwide as well as an Internet shop, and sells music , movies, and other media at the iTunes store, and software for the iPhone in the Apps Store. Personal computers that operate Microsoft Windows account for over 90% of the world market, and Apple computers have only a small percentage of the market share, but due to their novel designs, ease of use, and other features, Apple computers have acquired fans and the company has walked an independent path. The words and actions of the charismatic CEO Steve Jobs were famous, being heard and seen worldwide. However, Steve Jobs passed away in October, 2011.