Testing "Prompt": The Development of a Rapid Post-Editing Service at CLS Corporate Language Services AG, Switzerland

Catherine Hyland

CLS Corporate Language Services AG, Elisabethenanlage 11, CH-4051 Basel, Switzerland

CLS Communication Inc., 1500 Harbor Boulevard, Weehawken, NJ 07086-6732, USA
catherine.hyland@cls.ch

Abstract

CLS Corporate Language Services AG recently began offering the rapid post-editing of raw machine translation output to meet the rising demand for this service among clients. What is meant by rapid post-editing is the rough correction of machine translated texts with emphasis on speed and denotative accuracy. In the preliminary phase of the project, CLS conducted a test among four in-house translators. The objective was to gain practical experience, establish workflow requirements and set up efficient post-editing processes. Text samples were selected from several subject categories, and post-edited in English, German and French. The participants were given 10, 15 and 30 minutes per page to complete their tasks. This paper aims to present the results of the post-editing test at CLS Corporate Language Services AG, and to examine the conditions under which a rapid post-editing service is feasible in a commercial environment.

1 Introduction

CLS Corporate Language Services AG caters for a broad spectrum of client needs. While the demand for superior quality human translations has always been high, a new demand for rapid post-editing service has emerged. This is partly the result of increasing time and cost restraints on the client side, as well as thorough considerations in relation to the prospective use of the final text: A large proportion of multi-lingual texts is created for internal communication purposes and the dissemination of information across a global corporate infrastructure. In order to accommodate diverse customer requirements, CLS recently introduced a set of four different service levels. These include:

1) Human Translation with Quality Assurance ("Premium")
2) Human Translation without Quality Assurance ("Professional")
3) Machine Translation with Pre-Editing and Rapid Post-Editing ("Prompt")
4) Customised Machine Translation ("CLS Machine Translation")

This paper describes the in-house testing phase of Service Level 3), Machine Translation with Rapid Post-Editing, referred to as "Prompt".

2 Testing Procedure and Conditions

The "Prompt" testing phase started in December 2002 and was completed in January 2003. It was organised by Nicole Sagan, German Translator at the CLS office in Berne.[1] Four participants from various Swiss CLS offices took part in the test, and covered five language directions in total (En-Ge, Ge-En, Fr-Ge, Ge-Fr, En-Fr). The Machine Translation system used was our in-house CLS Machine Translation System (supplied by Comprendium, Germany). CLS used its own translation memories in conjunction with the CLS Machine Translation System.

The main objectives of the test were to gain hands-on experience with pre-editing and rapid post-editing, co-ordinate workflow processes and apply them in a day-to-day work environment. Special attention was given to identifying the optimum

---

[1] I would like to thank Nicole Sagan for providing me with the detailed results of the post-editing test.

relationship between document quality and time spent on editing, as well as defining a minimum document size, specifying text categories suitable for post-editing, and determining the professional requirements of prospective post-editing staff.

In the first round of the test, the participants were asked to spend no more than 10 minutes per page on the entire revision process. The time budget included pre-editing, running the translation through the MT engine, post-editing and storing the file. Pre-editing involved reading through the text to check for typing errors, adjusting the format where necessary, and providing terms and paragraphs that were not to be translated with a so-called "no proofing" marker. To automate these work steps, CLS developed a combination of pre-editing macros. After running the pre-edited texts through the MT system, the participants began post-editing them. Post-editing activities included replacing unknown words, deleting superfluous translation alternatives, correcting lexical and grammatical errors, and rewriting parts of (or entire) sentences. As the test proceeded, the organiser extended the time scope from 10 to 15, and finally to 30 minutes per page. The document length ranged from one to four pages. The sample texts were selected from three text categories: 1) Investment texts; 2) Mixed finance texts; and 3) Letters, e-mails and memos.

After completing the post-editing job, each post-editor rated the intelligibility of their own texts, putting themselves into the position of a reader with no command of the source language. This was done by calculating mean values for sentences, paragraphs and the entire document. The rating was based on a scale of 1 to 4: 1= I (would) have no problem understanding this sentence/paragraph/text. The quality of the text is comparable to the quality of a human translation. 2= I (would) have no problem understanding this sentence/paragraph/text, however, I consider it to be badly written. 3= I am not sure whether or not I (would) have fully understood this sentence/paragraph/text. 4= I have not (would not have) understood this sentence/paragraph/text at all.[2]

To determine whether or not the post-editors' own ratings were reliable, CLS AG selected single results to be double-checked in a spot test based on the same rating criteria. Of 147 texts post-edited in total, 28 were handed to a jury of (non-participating) in-house volunteers. This 'neutral' test returned similar, if not better results than the previous evaluation: Of 28 second ratings, 6 were approximately the same[3], 6 ratings were lower, and 16 turned out better than the post-editors' own judgment had suggested.

3 Results

As is evident from the post-editors' feedback reports, a time budget of 10 minutes per page for pre-editing, post-editing and administrative steps was not sufficient to produce acceptable text results. Most participants thought that 15 minutes were adequate provided that the post-editor is proficient in the respective subject domain, and that sentence structures are not too complex. An additional 5 minutes for basic formatting jobs was considered reasonable. All participants agreed that a time line of 30 minutes per page is too tight to retranslate the text, and too generous for a rapid post-editing service. In accordance with these results, CLS Corporate Language Services AG set a time budget of 20 minutes per page.

The size of the document does not appear to have any impact on the quality of the result, although some participants mentioned that due to the tight time limits and high work pressure, concentration levels tended to decrease towards the end. However, since the amount of time needed for mere technical processes (opening documents, running the translation through the MT engine etc.) is approximately the same for various document sizes, texts should be at least one page long. An optimum text length for best results is yet to be identified.

As for suitable language combinations, the qualitative ratings have shown that some language directions yielded better MT results than others. On the one hand, this divergence is due to different stages of lexicon development: As machine translation lexicons are not always up-dated simultaneously, one set of lexicons may be more advanced than another. On the other hand, there are system-inherent differences in language

---

[2] The CLS ratings are comparable to the intelligibility ratings used by Wagner (1998: 94) and van Slype (1980:7) for the evaluation of MT systems. See Appendix for a detailed list of the post-editors' ratings.

[3] Second digits after the decimal point were rounded off.

compatibility: Some post-editors pointed out that German-English generally produced better parsing results than English-German.

Internal communication texts such as e-mails, letters and memos received excellent marks and appeared to be best suited for "Prompt". The test participants remarked that texts of this category are usually short and easy to handle; simple sentence structures and general terminology facilitated the revision process. Mixed finance texts received mixed ratings, and, on average, outperformed the quality of post-edited investment texts: Market analyses, bond reports and currency ratings were less suited for MT in that they contained a large number of domain-specific idioms and were often written in an elliptic style. As a result, the post-editing of investment texts turned out to be relatively time consuming. Assessing the MT suitability of the source text is therefore a key prerequisite for setting up a cost-efficient service: "From the commercial point of view, the time and effort required for getting from translated output to a final version of the translation is crucial. The less post-editing is required or the easier it is to correct flaws, the better the cost-efficiency of the whole operation." (Povlsen and Bech, 2002: 284) The ability to work swiftly and methodically proved to be a fundamental skill for achieving good test results, but the test has also shown that rapid post-editing should remain a part-time task for translators to avoid excessive work pressure.

CLS AG noted that the successful implementation of a post-editing service strongly depends on the post-editors' attitude towards MT revision, and shares Vasconcello's view that "long-term commitment, positive attitudes, innovative responses [and] creative problem-solving" are the most essential requirements for post-editing. It was found that participants with a background of many years in translation had difficulty in accepting stylistic incongruities and prioritising comprehensibility. However, the experiment has shown that a basic introduction to Machine Translation processes helped post-editors handle their task more comfortably. Lavorel describes the issue of acceptance as follows: "Most post-editors are translators rather than revisors [...], their work is assessed on the basis of accuracy and quality, perfection of style being one of the main criteria.

[...] they are unwilling to hand in work written in any kind of standardized style [...]. Potential post-editors should therefore be given an objective explanation of what the machine can and cannot (yet) do and why." (1982: 108)

Understanding how the system operates helped to identify typical weaknesses and correct recurring mistakes more efficiently. Errors of this kind were sporadically communicated to the Machine Translation Team. As rapid post-editing leaves little time for research, in-depth expertise in the respective subject domain is regarded an essential requirement for making "Prompt" a successful service.[4] Finally, the participants considered it important to clearly indicate that, in contrast to texts translated by CLS Human Translation Services, the corporation does not assume liability for the contents of post-edited texts. To ensure full acceptance by the client, it was therefore suggested that "Prompt" texts be marked with a Rapid Post-Editing disclaimer.[5]

4 Conclusion and Outlook

The assessment of "Prompt" at CLS Corporate Language Services AG has shown that Rapid Post-Editing is a useful service supplement to Human Translation and pure Machine Translation. Rapid post-editing yields optimum results if texts are checked for MT suitability before they are translated and post-edited. Moreover, the time spent on post-editing and text quality should be optimally balanced. To attain this goal, CLS will continue to automate work steps and rely on highly qualified staff to further enhance the quality of its post-editing service.

Finally, it should be mentioned that the quality of post-edited texts strongly depends on the quality of the Machine Translation output. Therefore, CLS Corporate Language Services AG will continue to focus on lexicon updates and, in addition, establish a systemised feedback loop between post-editors and the machine translation team to facilitate this process. As one of four service levels, "Prompt" has

---

[4] H. Koby (2001: 12 ff.) takes this view further by suggesting that post-editing tasks should only be assigned to experienced translators.

[5] Cf. D. Senez: "Experience has shown that the acceptance of post-edited MT output hinges on the reader being fully aware that the text is essentially the product of a machine." (1998: 293)

become part of a comprehensive solution designed to fully respond to our clients' needs.

5   Bibliographical References

Koby, Geoffrey (Ed.). 2001. Editor's Introduction. Post-Editing of Machine Translation Output: Who, What, Why, and How (Much). In *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*. By Hans Krings. Kent State University Press, Kent, Ohio, pp. 1-24.

Lavorel, Bernhard. 1982. Experience in English – French Post-Editing. In *Practical Experience of Machine Translation.* Ed. By Veronica Lawson. North-Holland Publishing Company, Amsterdam, The Netherlands, pp. 105-109.

Povlsen, Claus and Anneliese Bech. 2002. Ape: Reducing the Monkey Business in Post-Editing by Automating the Task Intelligently. In *Proceedings of the MT Summit VIII. Machine Translation in the Information Age, Santiago de Compostela, Spain,* pp. 283-286.

Senez, Dorothy. 1998. The Machine Translation Help Desk and the Post-Editing Service. In Terminologie *& Traduction,* 1, pp. 289-295.

Wagner, Simone. 1998. Small Scale Evaluation Methods. In *Evaluation of the Linguistic* Performance *of Machine Translation Systems. Proceedings of the Konvens 98 Workshop, Bonn, Germany*, pp. 91-103.

Van Slype, Georges. 1980. Bewertung des Verfahrens SYSTRAN für die maschinelle Sprachübersetzung bei der K.E.G.. In *Lebende Sprachen: Zeitschrift für Fremde Sprachen in Wissenschaft und Praxis*, 25, pp. 6-9.

Appendix: Post-Editors' Test Results

Rating Criteria

1=   I (would) have no problem understanding this sentence/paragraph/text. The quality of the text is comparable to the quality of human translation.

2=   I (would) have no problem understanding this sentence/paragraph/text, however, I consider it to be badly written.

3=   I am not sure whether or not I (would) have fully understood this sentence/paragraph/ text.

4=   I have not (would not have) understood this sentence/paragraph/text at all.

Average ratings per time limit

Average per 10-minute text (48 texts)

Sentences (Mean Value)	2.33
Paragraphs (Mean Value)	2.45
Text	2.56

Average per 15-minute text (53 texts)

Sentences (**Mean** Value)	1.94
Paragraphs (Mean Value)	2.02
Text	2.04

Average per 30-minute text (34 texts)

Sentences (Mean Value)	1.35
Paragraphs (Mean Value)	1.36
Text	1.41

Average ratings per document size

Average per one page (47 texts)

Sentences (Mean Value)	1.80
Paragraphs (Mean Value)	1.86
Text	1.88

Average per two pages (69 texts)

Sentences (Mean Value)	1.98
Paragraphs (Mean Value)	2.07
Text	2.25

Average per three pages (12 texts)
Sentences (Mean Value)   2.08
Paragraphs (Mean Value)   2.18
Text   2.13

Average per four pages (7 texts)
Sentences (Mean Value)   2.04
Paragraphs (Mean Value)   2.17
Text   2.33

Average ratings per text category

Text types

1=   Investment texts (market analyses, bond
       reports, currency ratings etc.)
2=   Mixed finance texts (interviews, editorial
       texts, profiles etc.)
3=   Letters, e-mails and memos

Average per text type 1 (47 texts)
Sentences (Mean Value)   2.07
Paragraphs (Mean Value)   2.16
Text   2.27

Average per text type 2 (34 texts)
Sentences (Mean Value)   1.85
Paragraphs (Mean Value)   1.89
Text   1.95

Average per text type 3 (54 texts)
Sentences (Mean Value)   1.86
Paragraphs (Mean Value)   1.95
Text   1.96

Average ratings per language combination

Average per En-Ge text (57 texts)
Sentences (Mean Value)   2.23
Paragraphs (Mean Value)   2.36
Text   2.43

Average per Ge-En text (36 texts)
Sentences (Mean Value)   1.61
Paragraphs (Mean Value)   1.64
Text   1.63

Average per Ge-Fr text (15 texts)
Sentences (Mean Value)   1.69
Paragraphs (Mean Value)   1.74
Text   1.81

Average per Fr-Ge text (16 texts)
Sentences (Mean Value)   1.77
Paragraphs (Mean Value)   1.93
Text   1.94

Average per En-Fr text (11 texts)
Sentences (Mean Value)   1.99
Paragraphs (Mean Value)   1.88
Text   2.11