Machine Translation and Document Localization at Adobe: From Pilot to Production

Raymond Flournoy   **Christine Duran**

Adobe Systems, Inc.
345 Park Avenue
San Jose, CA 95110-2704 USA

flournoy@adobe.com   cduran@adobe.com

Abstract

Adobe recently began exploring the integration of Machine Translation (MT) technology into its localization workflow. The primary question that we sought to answer was whether post-editing Machine Translation output was faster than translating the text from scratch. The exploration occurred in two stages: a small pilot followed by a larger project localizing Adobe product documentation.

The project used two MT engines (one statistical and one traditional) both of which were trained with Adobe data and lexicons. Initially, a small test set of 800-2000 words of documentation was machine translated and post-edited, and based on the positive results we proceeded to the second stage, localizing about 200,000 words of new text.

The second stage completed successfully, but with some complications, including:

- The post-editing rate and the MT quality varied significantly between files. Additionally, differences in the first and second pilot test data created differences in MT quality.

- The technical integration of MT with the localization management system has been much more difficult than expected.

- The translation vendor reported that post-editing required more-highly-skilled translators to perform the task, which was contrary to our expectations.

1 Introduction

Adobe Systems localizes over 70 products into upwards of 32 languages. As a result, localization is a significant portion of the product development budget and the product release timeline.

Recently, Adobe has begun exploring the introduction of Machine Translation into the localization process, beginning with a two-stage pilot project designed to measure the benefits of MT. The goal of the project was to show what efficiencies were gained from using MT plus post-editing as compared to Human Translation (HT) from scratch. If such efficiencies were realized then Adobe would ultimately see three benefits:

- Faster turn-around for the localization process;
- Lower per-word cost for translating new text;
- And based on the faster, cheaper localization, easier expansion of the number of products and number of languages localized by Adobe.

The pilot was broken into two stages. In the first stage, we tested the concept of using MT plus post-editing on a small test set. Based on the results of this stage we moved on to a larger set of actual

product documentation to test using MT in the production localization workflow.

The results of each stage were positive, and the second stage reinforced the findings of the initial, smaller test. However we noted some interesting differences in moving from the small-scale to large-scale tests.

2 First-Stage Pilot

2.1 Methodology

Adobe commissioned two MT engine providers to customize their translation engines according to Adobe data. For Russian we enlisted PROMT, a fundamentally traditional MT system, which was customized for Adobe through a combination of manual rule writing and automated lexicon gathering. For Spanish and French, Language Weaver's statistical MT engines were trained with large quantities of Adobe translation memories.[1]

For the test set, text was sampled from documentation for Adobe's Flash product. Each test set was translated with the MT engine, and the time required to post-edit the text was recorded. Additionally, for Language Weaver a head-to-head test was performed to compare the post-editing rate with the HT rate.

2.2 Results

For both the PROMT and Language Weaver engines, the first-stage pilot produced positive results showing significant speed-up from using MT. The following table summarizes the post-editing rates for both engines.

	PROMT	Lang. Weaver[2]
Test set size	867 words	~2000 words
Time for 2500 words	70 min.	~115 min.
Words/hour rate	2142 w/h	~1000-1700 w/h

The unit of 2500 words is significant because that is the approximate daily output of a translator.

Thus these numbers seem to imply a translator's daily output can be produced by a post-editor in less than two hours. That translates to a 7-fold speed-up for PROMT and a 4-fold speed-up for Language Weaver.

These numbers seemed overly optimistic, because actual translation requires the overhead of project set-up, research, and proofreading. So a more instructive comparison was performed with the Language Weaver's engine. Here, the rates for MT post-editing and for HT were compared directly on comparable texts. In this head-to-head comparison the speed-up was judged to be between 22% and 51%.

3 Second-Stage Document Localization

Based on the encouraging results from the first-stage pilot, we proceeded to a significantly larger localization task. In the second-stage, we translated product documentation for Adobe's Flex developer framework as part of an actual product release, concentrating on documentation for the ActionScript programming language.

This localization included almost 1.5 million words total, but after the translation memories were leveraged the amount of new text was less than 200K words. The Flex documentation was translated into Russian and French.

Final results are not available at publication time, but preliminary results indicate that the MT post-editing was performed approximately 40% to 45% faster than human translation for comparable text or for the same text into comparable languages. Thus, indications are that the second-stage pilot also successfully showed significant efficiencies from inserting MT into the localization workflow.

4 Observations and Differences

While the first-stage of the pilot included a very small data set in an artificial environment, the second-stage experiment was a large-scale, real-world localization task. The results of the second stage matched and reinforced many of our results from the first stage, but there were some interesting dif-

---

[1] The training was performed by Cross Language, a reseller of Language Weaver technology.
[2] Language Weaver's results are averages for multiple editors for the two languages, French and Spanish.

ferences that we noted between the two experiments.

- Our suspicions were correct about the overly-optimistic nature of the statistics which only looked at the throughput of editors. Those numbers do not take into account the overhead that is included in the 2500-word daily output of a translator. The head-to-head comparison of post-editing speed and human translation was more accurate, and closely reflected the results we saw in the second stage.

- When we began the second-stage localization, early feedback from the post-editors indicated that the MT quality was significantly lower than expected and the editing speed was significantly slower than the initial pilot showed. However, as the work progressed the editors reported that this impression was erroneously based on a small number of more problematic files. Apparently the MT quality and the editing speed varied significantly between files, with some requiring very little editing and others requiring re-translation from scratch. Thus, it is important to gauge productivity over as large a data set as possible, and to be aware of the high variance in quality and speed between texts. This variance would be an issue for smaller jobs, where the editors might end up spending significantly more time than the average would indicate.

- The first-stage pilot test set was drawn from documentation for the Flash platform, but the second-stage pilot text consisted of documentation for a different, but related, product, Flex. Some drop in the MT quality was attributed to some Flex-related vocabulary being absent from the engines' lexicons. Additionally, the Flex documentation was originally chosen as a test case for MT localization because it has certain characteristic which are felicitous for machine translation: The sentences are mostly simple, declarative constructions which are understood in isolation, without much pronominal

reference. However, the Flex text is also not run through authoring tools, as is the case with other product documentation. As a result, the text contained misspelling and other errors, and was not authored with a concern for sentence length, clarity, and grammatical complexity. This showed that MT output quality can be sensitive to the specific characteristics of the input text.

- The second-stage project localization was performed using Adobe's standard Globalization Management System (GMS). The integration of the GMS with the MT engines has proved to be much more complicated than expected, and there are still outstanding technical issues which have not been resolved by this paper's publication date. The exact issues are not of general interest, but any company which is considering a similar integration should keep in mind how the editors will use the system and judge the integration accordingly. An example of how the editors' needs can be overlooked is that our GMS' default integration simply sent all new strings to the MT engines and stored the outputs returned. In doing so, it overwrote all repetition information so editors were forced to re-edit every occurrence of an MT-translated string.

- During the post-editing phase, we asked the editors to gather examples of bad translations and missing terminology, in anticipation of retraining and tuning the engines before the next round of translations. The amount of feedback and the quality of the feedback received varied between the editors, raising an issue which we had not originally anticipated. By gathering the feedback, the editors were providing a service to Adobe, but in doing so they were slowing down their editing pace. The net result was that they were not properly incentivized to provide the feedback properly. Moving forward, pricing will need to take into account the effort to generate this feedback.

- The Linguistic Service Provider which performed the post-editing expressed the opinion that post-editing required more senior, and more skilled translators. One reason given was the feeling that inexperienced translators would be too likely to trust the MT output. This observation was counter to our expectation that MT post-editing would require less experience. We are continuing to explore this issue.

5 Conclusion

Adobe has begun testing the introduction of MT into the localization workflow. Both an initial small pilot and a larger test with real-world documentation have shown significant productivity increases through the use of translation technology.

Acknowledgments