

# One Technology: Many Solutions

**Raymond Flournoy**  
flournoy@adobe.com

**Jeff Rueppel**  
jrueppel@adobe.com

Translation Technology Team  
Adobe Systems Inc.  
345 Park Avenue  
San Jose, CA 95110

## Abstract

Over the last two years, Adobe Systems has incorporated Machine Translation with post-editing into the localization workflow. Currently, the number of products using MT for localization has grown to over a dozen, and the number of languages covered is now five. Adobe is continuing to expand the number of products which use MT for localization, and is also looking beyond localization to other applications of MT technology.

In this paper, we discuss some of our further use cases, and the varying requirements each use case has for quality, customization, cost, and other factors. Based on those varying requirements, we consider a range of MT solutions beyond our current model of licensed, customized commercial engines.

## 1 Introduction

Adobe Systems has spent the previous two years developing infrastructure and workflows to incorporate Machine Translation (MT) technology into the localization process. As Adobe looks to expand the number of products using MT for localization and the number of language pairs licensed, we are also looking at other applications for the technology. These applications include both internal as well as externally-facing uses, with widely differing requirements for quality, customization, cost, and other factors.

The purpose of this paper is to present a selection of Adobe's possible use cases, discussing the varied requirements for each. In response to these varied use cases, different MT solutions are considered, including free online services, vertically-customized engines, and open-source solutions.

## 2 MT and Localization at Adobe

### 2.1 A brief history

When Adobe began investigating using MT technology, the first area of application that was targeted was product documentation localization. Adobe markets products worldwide, localizing the documentation, UI, marketing materials, and support documents into 30 languages in the most extreme case. As a result, localization is a major component of the product development budget and timeline.

After an initial pilot project produced encouraging results, in May 2009 Adobe applied MT to its first large-scale document localization task. Approximately 200,000 words of new documentation for Adobe Flex was translated into French and Russian using licensed engines customized for Adobe terminology. Productivity gains were estimated in the range of 25-40%, and discounted post-editing rates were negotiated based on these productivity gains.

In the time since the first Flex localization project, we have gradually expanded the number of products using MT. As of September 2010, fifteen

products use MT for at least one language's localization, including such major Adobe products as Acrobat, Flash, and Photoshop. The number of languages covered by licensed MT engines has grown to five: French, Russian, Spanish, Brazilian Portuguese, and Simplified Chinese.

## 2.2 Requirements for MT for Localization

The MT solution selected for product document localization was licensed, customized engines translating from English into another language. Each aspect of this solution was determined by requirements of the product documentation localization task itself.

- **Licensed:** All engines were either deployed internally or accessed through secure remote calls. This was necessary because all of the text being translated is unpublished, and thus considered confidential by Adobe. This automatically disqualified any free, online systems, as they are not secure or covered by a non-disclosure agreement.
- **Customized:** The purpose of deploying MT was to be able to realize productivity gains in MT post-editing as compared to human translation from scratch. In order for post-editing to reach a significant level of efficiency, the MT output must be of very high quality. While we did not test uncustomized engines, anecdotal evidence from post-editing agencies supports the need for customization to produce a measurable productivity increase.
- **From English:** All product documentation at Adobe is authored in English and then localized into other languages. Thus we were able to concentrate on language pairs with an English source. This also allowed us to concentrate on methods to improve the quality of the input English text.

Additionally, product documentation localization was chosen as the initial area for applying MT in part because one can show concrete return on investment for the licensing costs. Adobe pays Linguistic Service Providers (LSPs) on a per-word basis for most translation tasks. Based on the productivity gains calculated for MT post-editing, Adobe has negotiated discounts off of this per-word rate. Depending on the cost of the engine, the cost of translating into a specific language, and the productivity gains seen for an MT engine, we

currently calculate that after approximately 2 to 3 million words of new text the cost of the engine and lexicography can be recouped.

## 3 Next Applications for MT

We continue to expand the number of products which use MT for localization and expand the number of language pairs which Adobe licenses. Concurrently, we are investigating other applications for MT technology. Below are brief descriptions of some of the main use cases which are being targeted, concentrating on the varying requirements they place on an MT technology solution.

### 3.1 Customer care

Customer care interactions – in both directions, from a company to consumers and from consumers back to the company – have greatly changed with the advent of the web. While product manuals used to be the primary method for communicating with customers, users now receive more information via online support content which is updated dynamically. Consumers also have more avenues for communicating with Adobe, whether via email enquiries or solicited feedback.

#### 3.1.1 Online support content

Adobe publishes thousands of online support documents which could never be localized for all of the company's target markets. We are considering following a strategy employed by many other major enterprises, to use MT to publish all online documents in other languages.

In this case, the MT output would be raw and unedited, published with the purpose of general understanding of the text ("gisting"). High quality MT is preferred, but the quality requirements are not as stringent as for document localization where post-editing is performed. The MT engines should be customized to Adobe language. Historically, Adobe online support content is derived from many sources, so the quality of the input text is variable, although it is generally of high quality.

The direction of translation is basically English into other languages, and because all of the information is publicly available, security is less of an issue.

Finally, it is difficult to make an ROI argument for this use case. Effective online content can divert expensive customer care calls and email, but in a new market this could also help to grow the consumer base, resulting in more help requests.

### **3.1.2 Incoming email, feedback**

While the online support documents represent communication from Adobe to consumers, customer care also involves a high volume of communication in the opposite direction, from consumers to Adobe.

This communication involves email, product support requests, bug reports from partners and developers, and feedback from beta releases or other customer surveys. The feedback is sometimes product specific, but the subject matter can vary widely, so customization targets can be trickier to identify accurately. Because the communication can involve pre-release information the text must be kept secure.

As with web content, the purpose of applying MT is general understanding, and no post-editing is performed. Here, the input language can potentially be in any language, and the text will be of highly varying quality.

## **3.2 User-generated content**

Adobe is working to develop richer community-derived resources in many markets, including community translations and user community forums. MT has natural integration with both scenarios.

### **3.2.1 Community translation**

In a community translation initiative, MT output can be presented as pre-translations for the community members to edit to create the final translation. High quality MT is preferred, but the quality requirements are not as high as for regular document localization. Customization of the engines for the specific product is best, however it can be difficult to justify the cost of customization. Community-based projects produce many benefits, including increasing customer engagement and product exposure, but many of these benefits are not easily labeled with a dollar amount.

The language pairs of interest are basically English into other languages, and security is generally

not an issue, because the public will be accessing and translating the text.

### **3.2.2 User forums**

With user discussion groups and forums, a considerable amount of information – including Q&A, tutorials, and product reviews – is generated by the user community. However, this rich source of information is limited to speakers of that community's language. MT can provide a way to make these valuable resources available to other language groups.

In these use cases the quality requirements are not as high, because gisting is the goal of the MT. Customization targets are probably going to be difficult to define, unless a community is highly targeted to a specific product.

The text to be translated is publicly available, so security is not an issue.

Especially in the case of forums, the input could potentially be in any language, but the added complication of this use case is that the target language is also open-ended. For example, Adobe has active user communities in both China and Japan, so the Chinese-Japanese bidirectional language pair is potentially of great usefulness to allow communication between the two groups.

As with customer care, it is difficult to calculate the ROI of MT for this use case. Potentially the use of MT could reduce customer enquiries and professional translation costs, but it can also help to grow penetration within a market.

## **3.3 Product integration**

A natural long-range application for MT would be integration with Adobe's products themselves, to allow end users access to MT via the Adobe software. In this case, the MT would be presented as a part of Adobe's native offering, so the quality expectation would be very high. However, in this case there is no customization target, because the input text will come from the end user and will be completely unconstrained. Potentially any language pair direction would be of interest.

In this situation, Adobe would want to guarantee that the users data was kept secure, even through the Machine Translation process.

An ROI argument for this use case would depend on the market effect of offering this incremental feature on Adobe software. That will be

difficult to pin down, and is likely to vary widely from product to product, and from market to market.

The table below summarizes the above use cases, ordered from the highest quality requirements to lowest.

	Quality	Purpose	Customization	ROI?	Security	Language Pairs	Input quality
<b>Document localization</b>	High	Productivity gains	Required	Calculable savings	Required	EN>XX	High, controlled
<b>Adobe product integration</b>	High	Marketable feature for end users	No specific customization target	Varies	Flexible	XX>XX	Varied, uncontrolled
<b>Online support content</b>	Medium	Gisting	Required	Difficult to calculate	Low	EN>XX	Medium, controlled
<b>Community translation, forums</b>	Low-medium	Gisting	Varied subject matter	Difficult to calculate	Low	Primarily EN>XX; also XX>XX	Varied, uncontrolled
<b>Incoming email, bug reports, feedback</b>	Low	Gisting	Varied subject matter	Difficult to calculate	Required	XX>EN	Varied, uncontrolled

## 4 Solutions

Adobe has begun to explore various MT solutions, paying attention to how they can answer the particular requirements of each of the above potential use cases.

### 4.1 Customized licensed engines

Adobe currently licenses engines from commercial providers which have been customized to Adobe terminology. For statistical engines, this customization process involved training the statistical models using all available Adobe TMs as training data. This had the benefit of creating engines which are well-suited for all Adobe products, although the base license for these engines was relatively expensive.

For the rule-based engines, the customization is primarily a manual process, and requires creation of dictionaries based on a specific product's terminology lists and TMs. The rule-based engines have a lower initial license cost, but each incremental product requires additional lexicography charges.

Customized engines are appropriate when quality requirements are high, but they also only make sense when there is a clear customization target.

Because statistical engines can cover a large collection of TMs at once, they are well-suited when a general Adobe engine is required. The rule-based engines can be much cheaper when the customization target is narrow, but when the number of customization targets grows the total cost of ownership can become comparatively more expensive.

Because of the expense of customized engines, they are more easily justified in cases where there is a clear ROI argument, and where the number of language pairs is limited.

Commercial licenses generally allow for internal deployment or secure engine calls, necessary when secure communication is required.

### 4.2 Customization for verticals

A cheaper alternative to a customized licensed engine is a commercial engine customized for an industry category, such as travel or software. This offering has existed for rule-based engines in the form of subject matter-specific dictionaries, but now statistical MT vendors are taking advantage of the greater availability of categorized data to create translation models trained for a narrow category.

This option trades quality for price, and can be appropriate when a clear ROI argument cannot be

made, or when the quantity of language pairs required is large.

### 4.3 Licensed, uncustomized engines

Most MT engine vendors license their baseline, uncustomized engines, allowing a client to avoid the significant costs of dictionary building for client customization. This trades down quality even further for a lower price tag, but can be the best choice when there is no clear customization target or when an ROI argument cannot be made.

### 4.4 Free online services

The ready availability of free, online MT services has proven irresistible for many enterprises because of the price tag. However, their use comes with many drawbacks:

- Generally the use is not governed by any non-disclosure agreement and is not accessed through a secure connection. So online services are not an option when security is a requirement.
- Free services do not allow for any customization, so are not appropriate when the highest quality MT is required.
- Most free services are offered under a very loose usage agreement in which uptime is not guaranteed, volume can be limited, and the service can be terminated at any time.

On the positive side, if there is no customization target, the MT quality may sufficient. If a large number of language pairs are needed, the access to numerous, free engines is a definite advantage.

### 4.5 Open-source MT

Open-source MT solutions are an option that has recently become more realistic even for enterprise clients who are not actively involved in MT research.

The great advantage of open-source packages is the ability to train customized engines flexibly for narrow product lines or broadly across an entire enterprise, in as many language pairs as there is data available. This is a better choice for companies which have sufficient TMs to train engines, or who have access to pooled data (e.g. through the TDA) which is similar enough to be relevant.

The investment to get an open-source initiative up and running is mostly in terms of manpower. In the case of Adobe, the equivalent of one engineer working for three months was able to bring the Moses open-source package to a state where we could build engines of significant size and quality. The initial hardware investment was minimal; the development work was done with an off-the-shelf Mac running OSX, with some amount of scripting changes for our specific environment.

Beyond basic engine development, refining the engine training process and creating useful system integration is the next hurdle of using open-source MT. Because of the lack of supporting tools, it is not a trivial task to train and retrain a significant quantity of engines and to deploy them to be integrable with multiple, varied systems throughout the company. A significant amount of Adobe's development time has been spent creating more extensive corpus building and clean-up tools, a framework for a service deployment of the engines, and software connectors to our Globalization Management System (GMS).

Adobe is currently working on measuring the quality of the Moses-built engines against the commercially-customized engines. Our intuition is that the engine quality will be lower than commercial offerings in general, but for very narrow customization targets the Moses engines might match or exceed the licensed engines.

The table below summarizes key aspects of these various MT solutions.

	Quality	Customization	Expense	Secure	Language Pairs
<b>Licensed, customized</b>	++	Narrow or broad, but costs accrue per customization	Most expensive; costs per language pair direction	+	Costs accrue per language pair
<b>Vertically customized</b>	+	Broad industry customization, if commercially available	Less expensive; costs per language pair direction	+	Costs accrue per language pair

<b>Licensed uncustomized</b>		None	Even less expensive; costs per language pair direction	+	Costs accrue per language pair
<b>Free, online</b>		None	Free	-	Many
<b>Open-source</b>	+	Narrow or broad, if data available	Initial development work, then low cost per engine	+	Flexible, if data available

## 5 Next steps for Adobe

Based on the relative strengths of the various MT approaches discussed above, Adobe is considering the best matches for new applications for MT.

Within the category of licensed, customized engines, Adobe has maintained a best-of-breed approach, and currently licenses from multiple vendors. Similarly, as we move forward with new MT projects, we are open to engaging MT solutions from many of the discussed categories. To accommodate this, Adobe is creating a wrapper MT service unifying calls to our various engines, regardless of their source. This will allow groups within the company to make unified MT calls without knowing the specifics of the underlying technology, and also allows us to swap solutions out when better alternatives are uncovered.

While we have not made any concrete decisions about solutions for the various use cases, in general we are using the cost savings from document localization to justify the expense of licenses and customization. This allows Adobe to invest in the highest quality MT results where possible, making sure that licenses allow for wider use within the company. Localization becomes the economic driver of adopting and expanding MT use within the wider company.

For lower-tier languages, language pairs where the source is not English, and applications where the quality requirements are lower, we are developing solutions from the lower-cost categories. As described, Adobe has begun investigating the Moses open-source package, and has also been working with vendors to mitigate some of the drawbacks of vertically-customized, uncustomized, and free online services.

## 6 Summary

In this paper we presented some of the potential uses for MT technology that Adobe is currently exploring. Each presents various requirements for the MT technology solution employed, depending

on quality, customization, security, cost, and other concerns.

Additionally, we discussed five categories of MT solutions which vary in their relative cost, customizability, flexibility, and security.

Ultimately, Adobe believes that our full portfolio of MT solutions will flexibly pull from many of these categories, and so we are working on an internal infrastructure to support this heterogeneous suite of technology solutions.