Linguists Love Art and Management Loves Efficiency -- Can MT be the Solution?

Sachiyo Demizu
Nikon Precision Inc.
1399 Shoreway Rd.
Belmont CA, 94002
sdemizu@nikon.com

Mike Holland
Nikon Precision Inc.
1399 Shoreway Rd.
Belmont CA, 94002
mholland@nikon.com

Abstract

How to achieve the optimal balance of quality and cost when the need for translation is skyrocketing? Can machine translation be the solution? What system to choose? Finding the right MT solution for your organization is not easy. In this paper, we would like to share our experience at Nikon Precision Inc. in quest of the right tool, focusing on rule-based Japanese MT software and the results of a small pilot project, together with our plans for the future and the challenges we are facing.

1 Introduction

We started our MT endeavor about two years ago when we learned that use of free online translation tools in field offices was posing a security threat. Could MT software fill the gap when there is no translator available? How good is the quality of MT? Could we use it to increase our productivity?

2 Business needs and challenges

Before evaluating MT engines, we first examined our current situation and identified the following business needs and challenges.

Increasing workload: Translation volume has been increasing rapidly while the availability of qualified translators and editors is limited. There are piles of documents waiting for translation.

Real-time translation needs: Our field offices and technical support people need rapid translation of e-mails and other information exchanges. Sometimes our turnaround time is 24 hours or less.

Accuracy of translation: Due to our business requirements, it is critical to provide accurate translations.

Confidentiality: Translated content is highly confidential and unauthorized use of online free translation services poses a security threat

Language combination: More than 90% of our translations are from Japanese into English, and the majority of potential MT software users do not understand Japanese.

Source document problems: Most source documents are not well written and are filled with ambiguities. In addition, source language style and terminology are not standardized. Sending feedback all the way to the source document author is always a challenge.

Many-to-many relationship in terminology: The relationships between Japanese and English terms are often many-to-many and human translators need to determine what terms to use based on the context and their knowledge.

Frequent updates: Source documents are updated frequently so we need a method for version control. A good CAT tool would allow us to recycle existing translations, which would reduce the net translation page count.

Large translation memory: We have a large translation memory, all of that content is classified as “technical,” and no sub-grouping is possible at this stage.

3 What MT engine to use for phase 1 implementation?

Based on our findings, we tried to determine what MT engine/service to use. An ideal MT engine would quickly produce quality translations at a reasonable cost. Having the ability to normalize the source document would be a plus. Could we find such a tool?

When we first started our research, the language combination of Japanese to English greatly restricted our choices. Although we were interested in a statistic-based machine translation engine as well, the disadvantages were too high compared to the potential benefits.

After contacting several MT vendors and checking their references, we decided that the best choice for us at the time was rule-based Japanese MT software that could display menus in English.

4 Evaluation of Japanese rule-based translation software

For this evaluation, we selected a desktop-type rule-based MT software that was developed by a Japanese manufacturer.

4.1 Translation test data and evaluation criteria

To evaluate the MT software's translation capability, we used the test we administer to human translators to test their Japanese to English translation skills. The test consists of two parts: part one has content specific to Nikon and part two contains generic computer material. From here on, we will call this document the "original" test.

For the MT software evaluation, we rewrote the original test to create another document we will call the "modified" test. Because MT software in general is weak in handling long convoluted sentences and omissions, we simplified the sentence structure and added back omitted subjects and other parts of speech. Our assumption was that if we could see great improvement in the results of the modified test, then pre-processing either by our translation department or by the source document author might improve the quality of the MT output. This could also help those MT software users who could understand the source language to edit the source language text for more understandable target language output.

The Japanese character count of the source test files and the English translation word counts are given in Table 1.

	Part 1	Part 2	Total
Orig. source	657	616	1273 char.
Mod. source	694	648	1342 char.
Orig. trans.	345	295	650 words
Mod. trans.	345	296	641 words

Table 1. Test character/word counts

The segment information of the test documents is given in Table 2.

	Part 1	Part 2	Total
Original	12	14	26
Modified	22	16	38

Table 2. Test file segment counts

When we evaluate human translators, we count the number of errors and calculate the total error score based on the nature and severity of the errors. Test criteria and error scores are given in Table 3. The first four error types are considered accuracy errors and the remaining six types are considered idiomatic errors.

Error Category	Major	Minor
Incorrect meaning (Accuracy)	6	1.5
Untranslated/Over-translated (Accuracy)	3	0.75
Non-standard terminology (Accuracy)	2	0.5
Inconsistent terminology (Accuracy)	1	0.25
Spelling error (Idiom)	3	0.75
Grammar/syntax error (Idiom)	2	0.5
Inappropriate construction (Idiom)	2	0.5
Punctuation error (Idiom)	1	0.25
Register error (Idiom)	1	0.25
Foreignism (Idiom)	1	0.25

Table 3. Error scores

For human translator evaluation purposes, Nikon-specific terminology is provided at the time of testing.

A passing score is less than 10. The passing rate of the test is about 24%. The people who take the translation test have passed our resume screening process and most of them are professional translators. For those who fail the test, we do not have the total error score data because we stop evaluating the test when the total reaches 10.

4.2 Evaluation of original test

For the MT software evaluation, we first customized the user dictionary with our standard terminology, ran the test documents through the MT software, and evaluated the entire test without stopping at the pass/fail threshold of 10 points. The total processing time of the test file, including the time to open the file, was about 15 seconds. For the original test, we did not use the automatic sentence-splitting feature provided by the MT software.

Two of our translators evaluated the MT results and the averages of the two evaluators were used for the final analysis. Table 4 presents the results of the original test. The first three columns (Part 1, Part 2, and Total) show the results of the MT software, and the 4th column (Pass Ave.) shows the average scores of human translators who passed the test. The accuracy score is the total of the first 4 error types indicated by the asterisk (*) in Table 4.

Error Type	Part 1	Part 2	Total	Pass Ave.
*Incorrect meaning	48	60	108	0.3
*Untranslated/ Over-translated	0	0	0	0.5
*Non-standard terminology	2	2	4	2.4
*Inconsistent terminology	1	0	1	0.1
Spelling error	1	0	1	0.1
Grammar/ syntax error	15	14	29	0.6
Inappropriate construction	12	12	24	0.1
Punctuation error	0	0	0	0.1
Register error	0	0	0	0.1
Foreignism	3	5	8	0.5
*Accuracy Score Total	51	62	113	3.3
Grand Total	80	92	172	5.5

Table 4. MT test results (original)

The total error score of the original test was 172. Since our passing score is less than 10, it is obvious that we cannot use the MT output for our regular translation without massive post-editing.

As expected, the MT software had extremely few punctuation and spelling errors. Only one spelling error was found in the software translation due to an MT software term recognition problem. It also made no register errors. There were no over-translation or missed-translation errors, but there were many inappropriate construction errors. The biggest problems in the MT output were numerous meaning errors and incorrect grammar that drove the total error score very high.

4.3 Evaluation of modified test

Next we evaluated the modified test. The test scores are given in Table 5.

Error Type	Part 1	Part 2	Total
*Incorrect meaning	26	34	60
*Untranslated/Over-translated	2	0	2
*Non-standard terminology	1	1	2
*Inconsistent terminology	0	0	0.1
Spelling error	0	0	0
Grammar/syntax error	12	7	19
Inappropriate construction	4	9	13
Punctuation error	0	0	0
Register error	0	0	0
Foreignism	0	3	3
*Accuracy Score Total	29	35	64
Grand Total	45	53	98

Table 5. MT test results (modified)

Although there was some improvement in test scores, the total score of 98 is way beyond our passing score.

What is interesting is that the total error score of part 1 is lower than that of part 2 in both original and modified tests. Most human candidates consider part 1, which is highly technical with specialized terminology, more difficult than the generic computer-related content in part 2.

4.4 Post-edit vs. translation evaluation

To check the feasibility of a machine translation + human post-edit approach, we divided the tests into segments and evaluated the translation of each segment based on the following criteria:

- The MT output can be used as-is.
- The MT output is editable and can be used af-

ter edit.

- The MT output is useless. It is quicker to translate the segment from scratch.

The evaluation results of the original test are given in Table 6.

	Part 1	Part 2	Total
Number of segments			
Can use	0	0	0
Can edit	2.5	4.5	7
Cannot use	9.5	9.5	19
Percentage			
Can use	0%	0%	0%
Can edit	21%	32%	27%
Cannot use	79%	68%	73%

Table 6. Segment analysis (original)

In this evaluation, none of the segments could be used as-is. The number of segments that could be used after post-edit is small. The majority of the segments were considered not usable at all

We did the same evaluation for the modified test. The results are given in Table 7.

	Part 1	Part 2	Total
Number of segments			
Can use	3.5	0.5	4
Can edit	10	9	19
Cannot use	7.5	6.5	14
Percentage			
Can use	17%	3%	11%
Can edit	48%	56%	51%
Cannot use	36%	41%	38%

Table 7. Segment analysis (modified).

There are some improvements in the number of segments that can be used after editing. A small numbers of segments can be used as-is.

Next, we asked evaluators how long it would take to edit both documents in comparison with translating them from scratch.

Human candidates are given a total of two hours to complete translation of part 1 and part 2. The evaluators are familiar with the subjects and usually need less time to translate these types of documents. The results of the time evaluation are given in Table 8.

	Part 1	Part 2	Total
Original			
To edit	40	60	100
To translate	45	45	90
Modified			
To edit	30	45	75
To translate	45	40	85

Units in minutes.

Table 8. Post-edit vs. translation time

It shows that the total post-edit time of the original test would take longer than translating the entire document.

It may be worth noting that the post-editing time of part 1 is shorter than the translation time for both the original and modified test. As mentioned in Section 4.3, most human translators consider part 1 more difficult than part 2; however, the MT evaluation results are somewhat different. Although the sample size of this evaluation is way too small to see whether such a difference is meaningful, rule-based translation software might be able to handle technical documents better as long as the source documents are well written and use standard terminology.

After this evaluation, we changed the style setting of the MT software to “divide long source sentences automatically” and ran both the original and modified tests again. In part 1 of the original test, 4 segments showed major improvements in accuracy. One of the 4 segments is the longest segment in the document, with 109 characters. This indicates that non-native speakers of the source language can get more understandable sentences out of MT by breaking up long source sentences. No such improvements were shown in part 2 of the original test nor in part 1 or part 2 of the modified test. We did not perform a detailed segment analysis of these outputs because the error count in part 2 of the original test exceeded our pass/fail threshold.

Again, it might be worth noting that part 1 of the original test is considered difficult by human translators but the document itself is relatively well written while part 2, although considered easier in terms of content, is not as clear-cut in terms of grammatical structure compared to part 1. This might be the reason we only saw major improvement in part 1 of the original test using this setting.

4.5 Initial conclusion

Based on the evaluation results, we concluded that 1) the raw translation of the MT software is not suitable for translation of technical documents where accuracy is critical; 2) the use of the sentence-splitting feature of the software and modification of the source document can improve the quality of the translation but not enough to change our conclusion stated above; and 3) machine translation with this software plus human post-editing will not work because the post-edit takes as long as the actual translation, if not longer.

However, after considering the field office needs stated in Section 2, we also concluded that although MT software is not suitable for the regular translation of maintenance/installation procedures, it might have some use for engineers who need to get a rough idea of what the document is about. In addition, implementation of MT software could reduce the risks associated with the unauthorized use of online-based free machine translation services. The reasons behind this conclusion are 1) the software can handle short, simple sentences relatively well, and the sentence-splitting function works to a certain extent; 2) it can provide the right terminology fairly accurately after customization of the MT dictionary; and 3) there is a relatively low risk of information leaks with desktop-type software.

Furthermore, the MT engine could be a useful tool for those who understand both Japanese and English but need some help in writing. These users could perform both pre-edit of the source document and post-edit of the target document by themselves to obtain an adequate level of translation.

5 Phase 1 implementation and user feedback

Based on our analysis, we first implemented the software for a small number of test users. After receiving positive feedback, we implemented the software for other US-based employees for limited applications. They are allowed to use the MT software for informational purposes only and use of MT documents to perform any maintenance/install procedures is strictly prohibited. The use of MT software for translating customer documents is also prohibited.

5.1 User feedback

Contrary to the initial evaluation performed by the translation department, the feedback from field users is positive, and we are currently considering the purchase of additional licenses. Of more than 100 users, more than 95% of them state that the software is useful and they need to keep it.

To find out the reasons for the gap between our negative evaluation and positive user feedback, we interviewed some of the power users. Below is the summary of our questions and their feedback.

Q1) Do you find the MT software really useful?
The majority of the users interviewed say that they like the software and use it almost every day. Those who answer negatively are bilingual speakers who can speak both languages well.

Those who do not understand the source language state that with the MT software, they can understand the contents of technical updates they receive on a daily basis as well as e-mail communication. If they can spot a new procedure or other important information, they can send these documents for official translation. Some people state that the quality of MT translation is usually good enough to decide what action to take next. They also state that MT translations of Power Point presentations are usually not bad because most sentences are short and simple.

Those who have limited English ability state that the software helps them understand the gist of lengthy English documents quickly. They also state that the software is a great tool to check for grammatical and spelling errors.

Q2) How do you work around the limitations of software translation?

Many people state that the key to having the software produce meaningful output is to split long sentences into short ones. Some non-Japanese speakers perform this themselves by using the Japanese comma (、) as a clue and also by trial and error. Some users are not aware of the automatic sentence-splitting feature provided by the software.

They also state that if they do not understand certain sentences, there is always somebody who can help them. Usually they go to bilingual speakers for help. Sometimes, subject matter experts (SME) of the target language can help clarify the contents of the communication as well. Usually SMEs can make sense of fragmented MT output.

One interesting thing we found is the way Japanese speakers use the software to write English. They first write something in English and have the software translate it back to Japanese. If the Japanese does not make sense, they correct the Japanese and translate it back to English again. They repeat this process several times until the English output becomes somewhat decent. They state that they cannot write any meaningful English without the help of the software.

Q3) What are the things you don't like about the software?

Some people complain about the speed of the MT software. They state that the term lookup function is slower than that of similar software and that the translation speed is slow compared to free online translations. Other people state that when the MT software is running, it slows down other applications.

There are several complaints that the contents of the user dictionary are not up-to-date or fully customized for their needs.

Some users who use the software for English to Japanese translation state that the software cannot handle long complicated sentences properly and often mixes up relative pronouns with interrogatives.

There are some complaints about the software menu not being fully compatible with Windows 7. They say that some menu boxes do not pop up when they use the software with MS Outlook while they never had this problem with Windows XP.

We also interviewed a translator who used to translate some of the Japanese users' weekly reports manually. She told us that she used to spend about 14 hours per week translating weekly reports written by four Japanese engineers. After implementation of the MT software, they stopped sending her weekly report translation jobs. She can now spend more of her time on the regular translation jobs for the translation department.

Further, we asked her if she herself used the MT software for her own translation purposes. She states that it is only useful for looking up terms in place of an electronic dictionary.

5.2 Further analysis of the software

Based on the user feedback, we did a further study of the software. First, we checked the software performance with the large files we regularly translate and found it is weak in handling large files. When we tried to open a MS Word file of 8.7 MB, it took 33 minutes. The translation speed itself was not a problem, and it translated 71,181 characters in 254 seconds, which is equal to 16,614 characters per minute. However, if we incorporate the file opening time into this calculation, the translation speed goes down to 1,903 characters per minute.

Then, we studied the shared user dictionary and memory features. Although the software allows us to share dictionaries and memories using either a common folder or a server and dictionaries can be automatically downloaded, it is very difficult to ensure all users have the right dictionary and environment setup for such desktop-type software. Furthermore, we could not find a good way to reflect all the changes made by multiple users to the shared dictionaries. The administrator of the dictionary needs to make changes and upload them to a common folder or a server, while individual users can make changes to their own user dictionary on their computer.

Another problem we found is the incompatibility of the terminology databases and the memories our translation department maintains with those of the MT software. The translation department maintains large terminology databases and translation memories for its regular translation/interpretation purposes. Compared to the memory size of our CAT tool, what the MT software can handle is very small. The recommended size of the user dictionary is up to 10,000 entries and the same applies to the translation memories since translation memories are stored in some of the dictionaries. According to the user guide, the software allows us to have up to 1,000 dictionaries or memories. Therefore, it is possible, in theory, to divide our main memory and term database into hundreds of small dictionaries; however, this approach is not practical. If we divide our memories and term databases into small pieces, it would make the daily management of translation jobs very difficult as well as the management of multiple memory and dictionary files. In addition, the MT software allows us to have only one changeable dictionary per user at a time. If we have to divide our database into multiple pieces, we need to modify the dictionary setting each time we want to make changes to a different dictionary.

Another difficulty we find is the incompatibility between the terminology database and the MT software dictionary. The terminology database is for human translators who translate the concepts behind the language. What is most important in the terminology database is the concept of the terms in a given context, which often appears as a definition. Grammatical information such as part of speech, gender (if applicable), or notes on usage are helpful but not a necessity for those whose language command is near native level. On the other hand, what the rule-based MT software needs for its literal translation of phrases or sentences is the lexicological information of words, most of which our terminology database does not have.

Furthermore, the MT software does not have the capability to interface with the CAT tool the translation department has. One-time conversion of our CAT tool memory into the MT software dictionary may be feasible but that is not enough. What we need is the real-time interface of our CAT tool with an MT engine.

6 Phase 2: Search for an MT engine for translators

As stated in Section 5.1, MT software can be a good tool for those whose language command is limited. However, what we are really looking for is a more powerful MT engine that satisfies the needs of the translation department. Almost two years after the initial MT software implementation, we decide to launch phase 2 of the project to look for a more powerful tool for translators, i.e. something that can produce a high volume of high quality translations quickly.

We are revisiting some of the MT tools we reviewed in the past and have realized that our range of selection has increased a lot. As a first step, we decided to define our basic requirements again. Described below are the requirements, challenges, and questions we have. Some of the details have yet to be refined.

Price within budget + good ROI: This is probably the most basic requirement everybody has; however, due to the various service types and pricing structures offered by MT vendors, comparison of basic prices and ROI calculation is complicated. In addition, unlike desktop-type MT software that usually offers a free trial period, most of the statistic-based MT engine vendors require up-front MT engine development or training fees for a pilot project.

Costs we have to consider for the ROI calculation include pilot costs (engine development /training cost + fee for initial usage, if any) , on-going costs (annual licensing fee or fee per usage, support contract cost, user training costs (if any), server maintenance costs or hosted service fees, additional engine training costs, additional hardware purchases (if any).

The MT returns we expect are 1) MT saving that is calculated as [(Human translation hours * rate + Edit hours * rate) – Post-Edit hours * Rate], and 2) reduced turnaround time calculated as [(human translation hours + edit hours + DTP hours) - (MT processing time (hours) + post-edit hours + DTP hours)] .

Depending on the pricing structure of the MT vendors, it might be necessary to establish a formula to convert source file character count to target file word count and establish a pricing method for fuzzy translations.

Our quick price comparison of potential MT engines indicates that there is a correlation between quality of the MT engines and their prices. It requires a lot of initial investment and/or expensive on-going fees. Due to our extremely high accuracy requirements, we probably need to spend many hours post-editing the MT output. Finding a quality MT engine that can give us a good ROI might be a challenge for us.

Ability to translate quickly: What would be a good benchmark number for the MT engine speed, 2,000 words per minute or 10,000 words per minute? How about file processing time or file conversion speed? The numbers we received from several vendors vary a lot. Perhaps this is something we need to establish by actually testing the MT engines in our own environment and with our own files.

Japanese file handling capability: In addition to having the ability to translate Japanese into English, the MT engine should be able to handle and display Japanese file names and other metadata correctly. No garbled Japanese characters should be allowed. Furthermore, segmentation of the Japanese texts should be adequate.

Large file handling capability: The MT engine should be able to handle large files (up to 15 MB) and the file processing speed should be adequate.

Heavy workload and multiple user access support: The MT engine should have the capability to handle access by multiple users and process files simultaneously.

Easy to edit: Quality of the MT output should be good enough to allow post-editors to work quickly.

This is another basic requirement we have; however, when editors who edit the work of human translators state "the MT output should be easy to edit," what exactly do they mean? Good terminology, good grammar, or free of mistakes? The general consensus is that the MT post-edit usually takes significantly longer than the editing of human translators' work. How much longer is acceptable? Perhaps this is another bench mark we have to determine.

Engine retraining requirements: The retraining of the engine should be performed easily and frequently. Some MT engines allow us to retrain the engine by ourselves for additional fees, others offer engine retraining services either with or without fees. How much effort and resources are required for this process? For some vendors/users, engine retraining simply means adding or updating terms in the user dictionary; for others, this means massive post-editing. This is another area that requires further study.

Accuracy of MT output must be satisfactory: If we employ the machine translation plus human post-editing approach, the ease of editing might be more important than the accuracy of the MT output; however, if we also want field users to use the same MT output for quick informational translation, it is important that the accuracy level of the MT output is decent. However, what level of accuracy would be required for a field user to believe that the translation is decent enough? Many field users perceive that the accuracy level of the MT software we implemented is a little over 30%. Where did this number come from? We do not know the answer yet.

Interface with CAT tool: The MT engine must have the ability to interface with our CAT tool to make the translation workflow seamless.

Compatibility: The MT engine should be able to handle most standard file formats such as TMX, TBX, and various MS Office documents we translate regularly.

Good security and confidentiality: Due to the nature of the information we handle, good system security is another must-have item. However, here is another irony. When we compare the quality of publicly available free translation engines with that of privately developed ones, it seems like the former is superior because they have access to the tremendous amount of linguistic data supplied to them online on a daily basis. Finding a good MT engine without compromising our security requirements is another challenge for us.

Hardware requirements: The MT engine should run on servers currently available in our environment in order to keep the up-front investment to a minimum.

Hosted-service model if possible: A hosted-service solution would be ideal to avoid the extra work required for server maintenance.

System reliability: The MT engine should be able to maintain a high up-time rate, and it should rarely fail or generate errors.

Good customer support: The MT engine should come with good customer support, and the response time for queries and updates should be fast.

Easy-to-use user interface: The MT engine should be easy to operate and should require little time to learn.

Scalability[1]**:** The MT engine should be able to support future expansion of the MT application.

Easy to maintain: Updates and other maintenance of the MT engine should be performed easily or automatically. This is another requirement we have based on our experience with MT software and CAT tools. Implementation of updates and changes to the software, dictionary, and memory should be pushed to the users' computers, if possible.

Transparency: This may not be a real requirement but something we want to have after our experience with the small pilot project described in the next section. Sometimes, MT services are like a black box to us, and we may not have a way to validate the accuracy of the information provided by the vendors. A certain transparency in this area is desirable.

Ability to normalize source document: Because most of our source documents are not standardized, it would be nice to have the ability to normalize the source document prior to the MT. The MT engine should be able to normalize the single-byte vs. double-byte style problems associated with the Japanese language so that we do not have to make

[1] Depraetereet et al. 2010. *Machine Translation Engine Selection in the Enterprise. EAMT May 2010 St Raphael, France,* P3.

additional investment in a third-party's quality-checking software.

7 Small pilot project with an MT vendor

Here are the results of a small pilot project we had with an MT vendor. In this project, we provided our main terminology database and memory in TMX format. The vendor cleaned up the memory and created the MT engine. The post-edit service was also provided by the vendor in this project. Their analysis shows a BLEU score of 56.49; however, we were unable to check the validity of this score and other data. The BLEU score of an engine developed by another vendor using a similar set of data was about half of this score.

The character count and memory match information of the file used for the project are given in Table 9.

Match Type	Characters	Segments
Repetitions	149	14
100%	308	68
95% - 99%	102	3
85% - 94%	129	10
75% - 84%	259	20
50% - 74%	120	7
No Match	4,168	181
Total	5,235	303

Table 9. Source file count and match data

The vendor first processed the source text through a CAT tool, sent the output from the tool to the MT system, and then fed the data back to the CAT tool again for human post-edit. The results are summarized in Table 10:

	Translation (hours)	Edit (hours)	Total (hours)
Human	13.5	4.5	18
MT		12.5	12.5
Reduction	5.5 hours in total (30%)		

Table 10. Potential productivity gain.

The vendor performed an evaluation of 170 segments using their post-editors. They asked the editors to evaluate each segment using a scale of 1 (poor) to 4 (excellent). The average score of all 170 segments was 2.56, somewhere between 2 (having significant errors) and 3 (having minor mistakes).

The types and ratios of the errors they found were wrong word order (29%), grammatical errors (21%), wrong terminology (20%), omissions (12%), superfluous text (10%), and style problems (7%).

The feedback from their editors suggests that MT can handle short sentences relatively well; however, it has difficulty translating long sentences in the right word order. That coincides with the weakness of the MT software we tested in phase1.

Our inhouse editors reviewed the translation post-edited by the vendor and found several meaning errors and style problems. The quality of their post-edited translation is roughly equal to that of our translation vendors except that they have some more style problems. In hindsight, we did not discuss our style requirements prior to this pilot project. Building our style requirements into the MT engine algorithm might be a challenge for us.

We also noticed that the vendor removed a significant amount of translation segments during the cleanup process because the source and translation are identical. Because many alphanumeric expressions are used in our Japanese source documents, removal of such segments may not have been necessary.

In terms of ROI, it is necessary for our inhouse editors to review the translation for the final quality check. Whether we can use the MT plus post-edit business model that this vendor provides depends on how much costs we can save. Another factor we have to consider is the turnaround time. We need to know how many hours (or days) we can save by using this kind of MT plus post-edit service. We also need to see how much improvement the MT engine can make after 1 to 3 months of initial training. With effective MT engine training, quick turnaround time, and proper pricing, this kind of model may have potential. However, this kind of service model may not work for our field users who require translations of technical documents overnight, if not sooner.

We are currently evaluating several other MT engines and software as pilot candidates and are hoping to find one that satisfies our needs soon.

8 Conclusion

Our initial implementation of the desktop-type MT software for limited applications was successful. The software works OK with short simple sentences, and when the users understand both source and target languages enough to modify both input and output accordingly, it works better. Even for those who only speak the target language, the translation quality is good enough to get a rough idea of the content. However, the quality of the raw MT output is not good enough to use as a procedural document for machine maintenance or installation. Also, it is not useful for professional translators.

Another weakness of the software is memory and dictionary management. Centralized MT engine maintenance and update is difficult, and real-time interface with our CAT tool is not possible. We need a more robust machine translation engine that can interface with our CAT tool and increase productivity for translators, if possible.

Our initial search indicates that some products might have potential; however, further study is needed for the final conclusion. Finding a balance between ROI and quality MT output might be a challenge for us.

References

Depraetereet et al. 2010. *Machine Translation engine selection in the Enterprise. EAMT May 2010 St Raphael, France*, P3.