



## Manual MT Post-editing: "if it's not broken, don't fix it!"

by Rafael Guzmán

This article was originally published in *Translation Journal*, Volume 11, No. 4, October, 2007. (ISSN: 1536-7207). URL: <http://translationjournal.net/journal/42mt.htm>  
(Last retrieved: 1 October 2007)

### Introduction

It is widely accepted that MT has its full potential in translating domain-specific text in controlled environments (Hutchins, 1995). This is particularly true when dealing with vast amounts of text of an instructional or descriptive nature, such as technical manuals and user guides. On the other hand, MT is not meant for translating poetry, marketing texts, or press-releases.

MT can provide greater benefits and less frustration as long as reasonable expectations are set and complementary strategies are put into place.

However, Machine Translation (MT) keeps raising hopes and disillusionment. If readability and accuracy are to be guaranteed, it is necessary to invest heavily in post-editing what MT has produced. Post-editing has become by now one of the most disliked tasks by translators. As pointed out by an experienced researcher, translators still find it easier retranslating than post-editing MT (O'Brien, 2006). As a result, the benefits of using MT remain being questioned all the time.

In this context, it should be explored how manual MT post-editing can be addressed in a way that makes it an easier task and enables companies to draw real benefits. There are a number of parallel strategies that can be implemented in order to achieve this. One of them is specifying the scope of manual MT post-editing and sticking to it stoically.

### Narrowing down the scope of manual post-editing

If post-editing MT aims at delivering the same high quality as human translation, does it not defeat the purpose of using MT?

On the other hand, is language quality no longer important as long as MT makes good business sense? (Schaler, 2004).

Whether we like it or not, the scope and limitations of MT need to be understood and accepted. We need to "stop dreaming (...) and take our expectations to a reasonable level" (Champollion, 2001). But where is the right balance?

A close look at the types of corrections made during manual MT post-editing reveals the vast amounts of time, effort, and money "unnecessarily" spent in making merely stylistic corrections. While high stylistic quality is to be expected from human

translations, it does not seem reasonable to expect the same from MT... or from manual MT post-editing. It needs to be examined, therefore, to what extent the scope of post-editing could be limited to what is strictly necessary in order to transmit the information with accuracy and correction.

For instance, it could be agreed that readers of manuals and user guides can tolerate a certain level of "artificial" language as long as it is intelligible, accurate, and grammatically correct. Based on this assumption, linguistic corrections could be reduced to just fixing the following types of errors:

- Grammatical and syntax errors: e.g. wrong concordance (number, gender) and word order causing grammatical problems.
- Misspellings and punctuation errors: e.g. missing accents, wrong capitalization.
- Mistranslations: e.g. wrong use of key terminology; correct sentences with a different meaning from the source.

### Examples of unnecessary manual post-editing

The following examples (bold text) will hopefully illustrate this point. They are post-edited segments after MT. Although their relevance applies mostly to Spanish, the principle behind them remains valid for other languages. Of course, the reader of this article is entitled to completely agree or disagree with these suggestions.

#### *Infinitives vs. nouns*

In Spanish, nouns are often preferred to infinitives in titles. However, both forms are grammatically acceptable.

English	Acceptable MT output	Unnecessary post-editing
Setting host name in remote items	<b>Configurar</b> nombre de host en elementos remotos	<b>Configuración</b> del nombre de host en elementos remotos

The same applies if an infinitive occurs instead of a noun in the middle of a sentence. If the sentence is already grammatically correct, and the change will not add any significant clarity to the meaning of the sentence, the stylistic change could be dispensed with.

English	Acceptable MT output	Unnecessary post-editing
You can run the tool from the command line to facilitate the process of distributing the update to multiple host files.	Usted puede ejecutar la herramienta desde la línea de comandos para facilitar el proceso <b>de distribuir</b> la actualización a los archivos varios del host.	Se puede ejecutar la herramienta desde la línea de comandos para facilitar el proceso <b>de distribución</b> de la actualización a varios archivos de host.

#### *Personal vs impersonal style (use of usted)*

It is true that too many occurrences of "usted" (formal use of "you") in a large piece of text can sound too obsequious (Butt & Benjamin, 2004) and become a bit irritating for the reader. However, since the reader will normally understand it and it is not grammatically incorrect, post-editing could be regarded as unnecessary.

English	Acceptable MT output	Unnecessary post-editing
If you installed the Access Server without importing a license file, you can import the license file through the Access Server Manager.	Si <b>usted instaló</b> el servidor de acceso sin importar un archivo de licencia, <b>usted puede</b> importar el archivo de licencia a través del Administrador del servidor de acceso.	Si <b>se instaló</b> el servidor de acceso sin importar un archivo de licencia, <b>se puede</b> importar el archivo de licencia a través del Administrador del servidor de acceso.
You can undock a host from the Access Server.	<b>Usted puede</b> desacoplar un host del servidor de acceso.	<b>Es posible</b> desacoplar un host del servidor de acceso.
You must install the compatibility update to configure the connection to the Access Server.	<b>Usted debe</b> instalar la actualización de la compatibilidad para configurar la conexión al servidor de acceso.	<b>Es necesario</b> instalar la actualización de compatibilidad para configurar la conexión al servidor de acceso.

There can be exceptions, of course. Sometimes, the occurrence of "usted" + verb may sound too direct or a bit awkward for the Spanish user. If this is the case, a stylistic correction might be advisable.

English	Unacceptable MT output	Required post-editing
You use pcAnywhere to configure the hosts.	<b>Usted utiliza</b> pcAnywhere para configurar los hosts.	<b>Se utiliza</b> pcAnywhere para configurar los hosts.
You need to supply the Access Server name and user password to connect to the Access Server.	<b>Usted necesita</b> proporcionar el nombre del servidor de acceso y la contraseña del usuario para conectarse al servidor de acceso.	<b>Es necesario</b> proporcionar el nombre del servidor de acceso y la contraseña del usuario para conectarse al servidor de acceso.

### *Active vs passive voice*

The active voice is usually more suitable than the passive voice in technical manuals. However, there are many instances in which the difference is so small that post-editing the MT translation would not add any significant amount of clarity to the sentence.

English	Acceptable MT output	Unnecessary post-editing
If you are running Packager on Windows Vista, .pmi files from previous versions prior to 12.1 cannot be imported; however, you can import them if you are running Packager on XP.	Si está ejecutando Packager en Windows Vista, los archivos .pmi de las versiones anteriores 12.1 anteriores <b>no pueden ser importados</b> ; sin embargo, <b>es posible importarlos</b> si usted está ejecutando Packager en XP.	Si está ejecutando Packager en Windows Vista, los archivos .pmi de versiones anteriores a 12.1 <b>no pueden importarse</b> ; sin embargo, <b>sí pueden importarse</b> si usted está ejecutando Packager en XP.

### *Generic words and phrases*

Generic words and phrases do not need to be paraphrased with nicer words, as long as they are clear enough and grammatically correct.

English	Acceptable MT output	Unnecessary post-editing
To re-establish a connection with the Access Server, the host user must manually dock to the Access Server again.	Para reestablecer una conexión con el servidor de acceso, el usuario del host debe acoplarse manualmente al servidor de acceso <b>otra vez</b> .	Para reestablecer una conexión con el servidor de acceso, el usuario del host debe <b>volver a</b> acoplarse manualmente al servidor de acceso.

### *Prepositions and articles*

As a general rule, if prepositions and articles used by the MT system are grammatically correct, and changing them would not add any substantial clarity to the sentence, there is no need to make any changes. Unfortunately, this may cause some minor terminology inconsistencies in some segments.

English	Acceptable MT output	Unnecessary post-editing
The compatibility update installs a configuration tool in the pcAnywhere program folder.	La actualización <b>de la</b> compatibilidad instala una herramienta de configuración en la carpeta del programa de pcAnywhere.	La actualización <b>de</b> compatibilidad instala una herramienta de configuración en la carpeta de programa de pcAnywhere.

On the other hand, if a term is very widely used and there is one accepted way of using it ("opciones **de** Inicio", instead of "opciones **del** Inicio"), it would be appropriate to post-edit this segment accordingly.

In other cases, the use of the preposition in the MT output maybe grammatically correct, but may give the sentence a different meaning. In these cases, post-editing is required. This is one of the trickiest areas. In case of a doubt, it is always better to post-edit the segment to make it clear enough.

English	Unacceptable MT output	Necessary post-editing
Service-response times can fluctuate with the volume of sessions and session activity.	El tiempo de respuesta puede fluctuar con el volumen de sesiones y <b>de</b> actividad de la sesión.	El tiempo de respuesta puede fluctuar con el volumen de sesiones y <b>la</b> actividad de la sesión.

### Word order

If a sentence is grammatically correct and perfectly understandable by the user, there is no need to change the word order to make it more stylistically correct. For instance:

English	Acceptable MT output	Unnecessary post-editing
On the computer on which the host is installed,...	En el equipo en el cual <b>el host está instalado</b> ,...	En el equipo en el cual <b>está instalado el host</b> ,...
Documentation version 1.0	<b>Versión 1.0</b> de la documentación	<b>Versión</b> de la documentación <b>1.0</b>

As it can be imagined, avoiding these type of post-editing in large amounts of text can save a lot of time and reduce costs quite significantly.

## Training MT post-editors

Training MT post-editors is crucial. They need clear guidance in order to identify the scope of manual MT post-editing that will be expected from them. Getting the balance right is a challenge in itself, because MT post-editors cannot be expected to deliver the same linguistic quality as human translators, nor is bad quality acceptable.

The following are some practical recommendations that can help in this regard.

1. Create clear guidelines with detailed examples of what needs and does not need to be post-edited. As said earlier, post-editors need to be very clear on what they are to look for during post-editing.
2. Anticipate potential issues and appropriate solutions. One of the most likely questions MT post-editors will always ask is how to reconcile the style guidelines meant for human translation and MT guidelines when a project uses both. For instance, suppose that a project leverages 70% of the translations from a TM and the remaining 30% is machine translated. In this scenario stylistic inconsistencies will occur in the product. Post-editors need to be reassured that this is all right. This is part of the price that has to be paid if MT benefits are to be harvested.

Another potential issue is key terminology inconsistencies. This goes beyond merely stylistic considerations and demands immediate attention. Also, whenever a global change of a term is made in the MT output, this change must be replicated in any TM that is used in the same project.

3. Before post-editing starts, give your post-editors a test with some tricky MT segments to post-edit. Then give them feedback. This will provide them with a clearer understanding and more confidence to apply the MT post-editing guidelines. As Jeff Allen points out, "experienced translators find it more difficult to accept translations with a level of quality that is lower than what they have done for years." (Allen, 2003)
4. If there are several post-editors working on the same language, make sure that they all share the same knowledge and understanding of the post-editing requirements.
5. After post-editing, it is advisable to let a post-editing coordinator proofread the final output, at least perfunctorily. If unnecessary corrections are noticed, update your MT post-editing guidelines and inform your post-editors for future reference.
6. Ask your post-editors to report linguistic patterns that required severe post-editing. This feedback will help the writers to make the necessary adjustments in the source for next time.

## Final considerations

All the suggestions discussed in this article will have to be fine-tuned and agreed depending on whether MT is used in-house or outsourced to translation vendors.

Limiting the scope of manual MT post-editing does not mean that other complementary strategies cannot be implemented to improve the linguistic quality of the MT output.

MT preparation starts with the writers. If the source is MT-friendly, the MT output quality will be better. In addition, there is plenty of opportunities to automate the correction of many typical grammatical and stylistic errors (i.e. *linguistic patterns*). This can be done just before starting manual post-editing. For instance, a script could be developed to search and replace linguistic patterns (like the ones described in this article) occurring in the MT output. If post-editors learn to create *regular expressions* to automatically fix linguistic patterns, they will soon find that less manual post-editing is required. This is definitely something that translation programs should teach in college.

## Conclusion

MT can provide greater benefits and less frustration as long as reasonable expectations are set and complementary strategies are put into place. One of these strategies consists in avoiding unnecessary stylistic corrections during manual MT post-editing. This requires training and clear guidelines that must be learned and put into practice. Finally, automatic post-editing can be implemented before starting manual post-editing. This will improve the final linguistic quality and reduce manual post-editing.

## Acknowledgments

I thank Dr. Johann Roturier, Javier Mallo, Nora Aranberri, and María González for theoretical discussions of this topic and for their comments.

## References

1. Allen, J., 2003. Retrieved: June 16, 2007, from: <http://www.mail-archive.com/mt-list@eamt.org/msg00578.html>
2. Butt, J. & Benjamin, C., 2004. *A New Reference Grammar of Modern Spanish*. Hodder Arnold, London.
3. Champollion, Y., 2000. 'Machine translation (MT), and the future of the translation industry', *Translation Journal*, vol. 5, No. 1. Retrieved: May 10, 2007, from: <http://accurapid.com/journal/15mt.htm>
4. Hutchings, J., 1995. 'Reflections on the History and Present State of Machine Translation'. Retrieved: June 16, 2007, from: <http://www.hutchinsweb.me.uk/MTS-1995.pdf>
5. O'Brien, S., 2006. 'Controlled Language and Post-Editing', "*Multilingual - Getting Started Guide*", vol. 17, Issue 7, No. 83, p.19.
6. Schaler, R., 2004. 'A New Dawn for MT?', *Localisation Focus*, vol.3, Issue 1, p. 17.

## About the author

*Rafael Guzmán works for Symantec Corporation (Ireland) on terminology management, Spanish MT and linguistic quality. He also provides e-learning courses on localization tools organized by Austin Community College.*

*His personal Website is at [www.invenis.net](http://www.invenis.net)*