**Exploring Machine Translation on the Web**

Ana Guerberof Arenas
PhD programme in Translation and Intercultural Studies
Universitat Rovira i Virgili, Tarragona, Spain
ana.guerberof@gmail.com

**Abstract**
This article briefly explores machine translation on the web, its history and current research. It briefly examines as well four free on-line machine translation engines in terms of language combinations offered, text length accepted and document formats supported as well as the quality of their raw MT output.

**Keywords**
machine translation, free on-line service, web, internet, MT

## 1. Introduction

Machine Translation (MT) and the web have had a long and endurable relationship that continues to grow and consolidate. Users frequently use free machine translation on the web to understand texts coming from another language (often referred to as assimilation) or to publish a text into another language (dissemination). Moreover, they also use MT on the web to communicate instantly with other users in different languages through instant messaging. MT providers use the readily available bilingual data on the web to boost the performance of their engines in different language combinations. And finally, many language and translation teachers and students are using MT on the web to help their language learning as well as translation and post-editing strategies. In this article we will look at free MT services offered on the web and explore their performance through a fast comparison of four different engines. The purpose of this article is not to setup a scientific experiment but to give general information on this topic.

## 2. Brief history

In the 1980s the MT provider Systran implemented a service in France to be used on the postal service Minitel that allowed basic paid translation of short strings on a few language combinations. Systran was also being used in different intranets in US Government agencies and in the European Commission. From 1992 to 1998, CompuServe offered online MT translation services to their customers into major language pairs. CompuServe used MT in different forums and setup a document translation service where customers could upload online documents for quick translations at competitive prices. Flanagan reports on different reactions from users, some quite adverse and others very positive. However, the experience was in general very successful considering the volume that was translated using this service (Flanagan 1994-1996). 1997 saw the creation of the first free on-line machine translation service when Altavista and Systran teamed up to create Babelfish (now Yahoo! Babelfish) to offer free MT technology to the general public in ten major European languages. Yang and Lange reported on the increased use of this system and its success story (2003). Since then, many other MT vendors have added their free machine translation services to the web. Google launched Google Translate in 2007 followed by Google

Translator Toolkit in 2009 making free MT available into even more language combinations, and offering a free online machine translation and translation memory system for users on the web. Hutchins' Compendium of translation software (2009) offers a thorough list of all MT online services available, product characteristics, language combinations and prices. He cites more than fifty MT services online, some offering only one language combination, while others offer many, such as Yahoo! Babelfish, Click2Translate, Prompt-Online, Microsoft Bing Translator, Transledium, or Wordlingo amongst many others.

Gaspari and Hutchins (2007) offer an interesting historical analysis of free online MT as well as a description of its use today. They confirmed that there is an increasing demand of this service and that free online MT providers constantly improve the service offered (for example by adding new language pairs) guaranteeing that the general interest in on-line MT services remains high.

## 3. Research

Gaspari (2006) has studied the additional value offered by free online machine translation services when users consult web pages in another language that they do not understand. His conclusions were that MT did help users in finding basic information on web pages where they did not understand the language and MT increased their confidence level on the information found.

Also Gaspari (2007) has also carried out an extensive and interesting research on the role of MT in webpage translation. His research offered valuable information on available MT services, a survey on the use of free online MT, the use of MT for assimilation and dissemination purposes as well as guidelines for the design of a website powered by MT. Gaspari reports on comparative analysis of Online MT services carried out by: Blekman et al (2001), Zervaki (2002), Guyon (2003), Craciunescu et al (2004), Wießner (2004), Uneson (2005) and White (2005). These authors compared different MT engines in different language combinations from a linguistic perspective offering diverse results. Gaspari himself (2004) presented a comparative evaluation from a different perspective mainly based on user interaction and ease of use and not necessarily on the linguistic quality of the raw output. Gaspari also found that Internet users often made "questionable use of a range of free online MT services" (2007:117) as they tended to use it as bilingual dictionaries or lexical lookups, and he questioned the appropriateness of this use, which was also questioned by Yang and Lange (2003). Surprisingly, users had a positive perception of online MT services when consulting information in another language, even though MT had sometimes "and adverse effect on their correctness" (2007:241). The survey carried out also underlines the fact that users had an overall positive impression on the service rendered by online MT engines.

More recently Ethan Shen did a comparison of three online Machine Translation engines in the Gabble On Research Project (http://www.gabble-on.com). In the first phase of this project he reported that for longer texts Google was the engine generally preferred by users, but that for certain language pairs the results for Google did not differ greatly from other engines like Microsoft Bing or Yahoo! Babelfish. However, if the text length was shorter, Bing Translator or Yahoo! Babelfish "gain ground and engines showed different strengths according to the language pairs" (Shen 2010). Interestingly he looked as well into brand bias to discover that it did have an effect on how the users graded the engines, and this positively favoured Google Translate. After improving the methodology, the second phase of this project is still on-going and anyone can participate in the research by accessing their Internet site.

## 4. Comparison of four engines

The engines chosen for the English-Spanish language combination are Yahoo! Babelfish, Google Translate, Apertium and Microsoft Translator, two rule-based engines and two statistical-based engines with the intention of evaluating if the type of engine did have an impact on the results.

For the comparison, a general text found on the web from a magazine specialized in computers was used as we found this type of text could be representative of the type of webpages translated using free online MT services. The article chosen was selected from PC World: http://www.pcworld.com/article/200866/why_microsoft_should_thank_apple_for_windows_7s_success.html. The language style used in the article was informal and technical at the same time. Only four sentences were selected from the text and run through each engine to be able to analyze them in depth. Since accuracy is an important factor when deciding which engine to use to obtain better results, we chose the English to Spanish language combination as we had skills to validate how accurate the MT output was from each particular engine.

According to the Gabble project, Google was the preferred engine for the language combination English-Spanish in all categories (under 2000 characters, under 500 characters, under 150 characters and under 50 characters) by a 50% margin in three categories and by 28% for texts under 50 characters when compared to Bing and Babelfish.

### 4.1. Yahoo! Babelfish

Yahoo! Babelfish is powered by Systran, a rule-based translation engine. This portal http://babelfish.yahoo.com/ offers a very friendly and simple main screen divided into two: one box can be used to insert text (up to 150 words) while the other one can be used to translate web pages by inserting the URLs. Once the user clicks Translate, Babelfish shows the webpage translated.

Babelfish offers 38 language combinations. Twelve of those have English as the Source language and 7 have French, the other 19 include Simplified Chinese, Traditional Chinese, Dutch, German, Greek, Italian, Japanese, Korean, Portuguese, Russian and Spanish. The target languages include all of these.

### 4.2. Google Translate

Google translate uses a statistical-based machine translation engine. Google's main screen (http://translate.google.com/) is also user friendly. It presents a single text box where you can introduce text, a webpage or URL or upload a document (in PDF, TXT, DOC, PPT, XLS or RTF), although it warns the user that the original format might be lost.

The engine offers 52 languages in both directions. Some of them are labelled as "ALPHA languages" which means that the engine has not yet been perfected and it does not work as well as with the other languages.

Once the user clicks Translate, Google Translate shows the full webpage translated. If the user runs the mouse over the translated text, the original text for the highlighted segment is displayed in an information bubble just above the translated text. The user can also contribute and edit the text by clicking Contribute a better translation.

### 4.3. Microsoft Bing Translator

Microsoft Bing Translator is a linguistically informed statistical MT system. Its main screen (http://www.microsofttranslator.com/) offers two boxes, one for the source text and the other one for the raw output. The user can enter text and a webpage or URL. The application will automatically detect the Source language. Bing Translator offers 32 languages in both directions.

When a URL is introduced, Bing Translator opens a new window showing the text in English and the progress of the translation. Once it is finished it shows the translated webpage. Four different views can be selected "Side by Side", "Top, Bottom", "Original with hover translation" and "Translation with hover original". When the user runs the mouse over one target sentence, it highlights this sentence and the corresponding source sentence.

## 4.4. Apertium

Apertium is an open source, ruled-based, shallow transfer machine translation system. It offers a free and open source machine translation platform aimed initially at related language pairs but that has recently expanded to include other language combinations (up to 35).

Apertium lets the user translate text by inserting it in a box online and choosing the language combination. It does not support URL or webpage translation directly. The user needs to cut and paste the text and it provides a translation at the bottom of the input box.

## 5. Examining the output

These are the results for the chosen sentences provided by the four engines. Errors are marked in red.

| # | Source Text | Yahoo! Babelfish | Google Translate | Apertium | Microsoft Bing Translator |
|---|---|---|---|---|---|
| 1 | More than half of software developers are already building applications for Windows 7and nearly 80% will do so within the next year, a new survey has found. | Más que la mitad de los analistas de programas informáticos están construyendo ya los usos paraWindows 7 y el casi 80% harán así que dentro del próximo año, un nuevo examen ha encontrado. | Más de la mitad de los desarrolladores de software ya la creación de aplicaciones para Windows 7 y casi el 80% lo hará en el próximo año, una nueva encuesta ha encontrado. | Más que medio de software está construyendo aplicaciones para Ventanas 7 y nearly 80% hará tan dentro del año próximo, una encuesta nueva ha encontrado. | Más de la mitad de los desarrolladores de software ya están construyendo aplicaciones para Windows 7 y casi 80% será hacerlo dentro del año siguiente, se ha encontrado una nueva encuesta. |
| 2 | And somewhat ironically, Microsoft might have Apple to thank for developer interest in its new operating system. | Y algo irónico, Microsoft pudo tener Apple a agradecer por revelador interesa en su nuevo sistema operativo | Y con cierta ironía, Microsoft podría tener Apple dar las gracias por el interés de desarrolladores en su nuevo sistema operativo. | Y un poco irónicamente, Microsoft podría tener Manzana a thank para developer interés en su sistema operativo nuevo. | Y un poco irónicamente, Microsoft podría tener Apple para agradecer al desarrollador de interés en su nuevo sistema operativo. |
| 3 | Overall, 54% of developers said they are already developing applications for Windows 7, and another 25% said they will do so within the next year. | Total, el 54% de reveladores dijeron él está desarrollando ya los usos para Windows 7, y otro 25% dijeron que él hará tan dentro del próximo año. | En general, el 54% de los desarrolladores dijeron que ya se están desarrollando aplicaciones para Windows 7, y otro 25% dijeron que lo harán en el próximo año. | Overall, 54% de developers dicho ya están desarrollando aplicaciones para Ventanas 7, y otro 25% dijo harán tan dentro del año próximo. | En general, 54% de los desarrolladores dijo ya están desarrollando aplicaciones para Windows 7, y otro 25% dijo que lo harán dentro del año siguiente. |
| 4 | Ten percent of respondents have no plans at all to develop Windows 7 applications. | El diez por ciento de respondedores no tiene ninguÌn plan en absoluto para desarrollar | El diez por ciento de los encuestados no tienen planes para desarrollar en | Diez per cent de respondents tiene ningún plan nada para desarrollar Ventanas 7 | Diez por ciento de los encuestados no tiene planes para desarrollar aplicaciones de |

| | | Windows 7 usos". | todas las aplicaciones de Windows 7. | aplicaciones | Windows 7 |
|---|---|---|---|---|---|

In general, none of the engines solved well two problematic structures in sentence 2: "And somewhat ironically" and "might have Apple to thank for". The first is a set expression that could be translated as "*Resulta irónico* que" and the second expression is quite a complicated verbal structure with a proper noun inserted in the middle of them that could be translated as "*podría tener que dar las gracias a Apple por*". Both type of expressions are quite complex and they represent known problems for MT engines. The two statistical-based MT engines showed less number of errors than the rule-based MT engines and they were much easier to read. In fact, both texts (provided by Google Translate and Bing Translator) could be very well understood without reading the English. Both engines showed almost perfect translations of sentence 3 and Bing Translator presented no errors on sentence 4. The errors were not only less but they were also minor and therefore did not affect in equal measure the reading of the text as the rule-based engines did.

If we looked at the errors in Babelfish, we could see that there were terminological errors (the translation for developer, applications, survey and respondents were wrongly translated in this context) and it also presented some problems with word order (developer interest, Windows 7 applications). Google Translate, apart from the common errors initially mentioned, showed some issues with verbal structures and with the expression "at all". Apertium left some words directly in English (developers, nearly, overall, thank, per cent) and translated some proper nouns (Windows and Apple) which made the text quite difficult to understand. Microsoft Bing Translator showed some problems with more unusual verbal structures.

## 6. Conclusions

In this very brief exploration of Machine Translation on the web, we could see that there is a vast array of free online MT services available in many language combinations, and this number is constantly increasing together with the quality of the target text. This offers real help to Internet users. There is important research and studies that have been carried out to analyze the use and quality of online services although there is indeed a need for more. After briefly comparing four engines in the English to Spanish language combination using only four sentences, both statistical engines, Google Translate and Microsoft Bing Translator, provided more understandable translations than the rule-based Babelfish and Apertium. Bing Translator performed slightly better overall and it even offered a perfect translation with no errors for one of the proposed sentences. On the other hand, Babelfish performed better than Apertium showing less untranslated words. Finally, both Google Translate and Bing Translator showed a more complete and resourceful interface as well as more language combinations.

## References

Blekhman, M.S., O. Bezhanova & M. Byezhanova (2001) "Comparative Analysis of the Translation Quality Produced by Three MT Systems". http://www.lec.com/downloads/transanalysis.doc [Date of last visit: 15 June 2007]. No longer available online.

Craciunescu, O., C. Gerding-Salas & S. Stringer-O'Keeffe (2004) "MachineTranslation and Computer-assisted Translation: a New Way of Translating?".Translation Journal 8(3). Available at http://www.accurapid.com/journal/29computers.htm [Date of last visit: July 25 2010].

Flanagan, M. & P. Jensen (1994) "Machine translation on CompuServe Forums". *Proceedings 1st AMTA Conference*. 228

Flanagan, M. (1996) "Two years online: experiences, challenges and trends". *Proceedings 2nd AMTA Conference*. 192-197

Flanagan, M. (1997a) "MT today: emerging roles, new successes". *Machine Translation* 12(1-2): Kluwer Academic Publishers. The Netherlands. 25-27

Flanagan, M. (1997b) "Online translation: MT's new frontier". *Proceedings 19th Translating and the Computer*. London. UK.

Gaspari, F; Hutchins, J. (2007). "Online and Free! Ten Years of Online Machine Translation: Origins, Developments, Current Use and Future Prospects". MT Summit XI. Copenhagen,Denmark. Proceedings.199-206.

Gaspari, F. (2007). *The role of Online MT in Webpage Translation*. A thesis submitted to The University of Manchester for the degree of Doctor of Philosophy in the Faculty of Humanities. http://eamt.org/archives.php [Date of last visit: July 10 2010]

Gaspari, F. (2004) "Online MT Services and Real Users' Needs: an Empirical Usability Evaluation". In Frederking, R.E. & K.B. Taylor (eds.). *Proceedings 6th AMTA Conference*. Berlin: Springer. 74-85.

Guyon, A. (2003) "Machine Translation and the Virtual Museum of Canada (VMC)". http://www.chin.gc.ca/English/Pdf/Digital_Content/Machine_Translation/Machine_Translation.pdf [Date of last visit: 15 June 2007]. No longer available online.

Hutchins, J. (2009). "Compendium of Translation Software" http://www.hutchinsweb.me.uk/Compendium.htm [Date of last visit: July 25, 2010]

Shen, E. (2010) "Real World Comparison of Online Machine Translation" http://www.gabble-on.com/files/phase1_full_research_report.pdf [Date of last visit: July 17 2010]

Uneson, M. (2005) "InterTran vs SysTran: Evaluation of Two Online MachineTranslation Services". http://person.sol.lu.se/MarcusUneson/hlt/courses/gslt-nlp/ass4/uneson_gsltnlp_ass4_report.pdf [Date of last visit: 15 June 2007]. No longer available online.

Wießner, E. (2004) "Automatic WWW-translations – an Empirical Comparison".Available at http://www.tu-chemnitz.de/phil/english/chairs/linguist/independent/kursmaterialien/translation/Automatic_wwwtranslations/Automatic www-Translations - An EmpiricalComparison.doc [Date of last visit: July 25 2010].

White, S. (2005) "An Evaluation of Online Machine Translation". Available at http://www.dcs.shef.ac.uk/intranet/teaching/projects/archive/ug2005/pdf/u1sw.pdf [Date of last visit: July 25 2010].

Yang, J; Lange, E. (2003) "Going live on the internet". *In Computers and Translation: A Translator's Guide*. Harold Somers, ed. Amsterdam & Philadelphia: Benjamins. 191-210.

Zervaki, T. (2002) "Online Free Translation Services". *Proceedings of the 24th International Conference on Translating and the Computer*. London: Aslib/IMI.