Using free on-line services in MT teaching

Federico Gaspari

Centre for Computational Linguistics
UMIST
PO Box 88
Manchester M60 1QD
United Kingdom
F.Gaspari@postgrad.umist.ac.uk

Abstract

This paper discusses the role that can be played in teaching machine translation (MT) by free on-line MT services, emphasising the advantages that they offer for didactic purposes as well as their drawbacks in comparison with traditional PC-based MT software. Some of the key reasons behind the huge potential for the successful integration of on-line systems into MT teaching activities are explored. The paper argues that exposing students to Internet-based MT can raise their awareness on crucial issues in today's multilingual communication processes, which are relevant *inter alia* to web localisation and multilingual on-line content management.

1   Introduction

The present author has taught modules focused on MT within graduate courses aimed at trainee translators, which had never included MT components before (cf. Gaspari, 2001). These modules comprised a theoretical part which served as a general introduction to the subject of MT, and covered the following main areas: historical development, design methods, linguistic and computational issues in MT systems, their practical use with reference to post-editing techniques, principles of evaluation, a review of the main approaches (e.g. rule-based vs. example-based and statistics-based MT) and a few case studies of some of the systems (e.g. Meteo, Systran, Eurotra), etc.

At the end of the modules there were hands-on lab sessions in which the students were invited to experiment with some on-line MT services available free of charge on the Internet, testing and comparing their performance. This strategy had to be adopted because of the lack of financial resources, owing to which it was not possible to purchase PC-based commercial MT systems for the students to experiment with. As a result, for demonstrations or practical experiments with MT, the instructor had no choice but to deploy free on-line systems available through the Internet.

The students were asked to choose the input texts and the language combinations they wanted to work with among those available, and to examine the output given by the free on-line MT services in the target language under the supervision of the instructor, following some simple pre-defined guidelines (described in detail in Gaspari, 2001:37-38).

This final practical session proved extremely popular with the students, who at the conclusion of the theoretical part of the course could appreciate the difficulties faced by MT systems in simulated productive scenarios. On the whole, the tests carried out with the web-based MT services turned out to be very stimulating for all the students, who were eager to share with the rest of the class and the

instructor their comments and remarks about the performance of the systems they were testing.

Drawing on the experience briefly described above, this paper will argue that integrating the use of free on-line MT services into MT courses can be highly beneficial, irrespective of the background and syllabus of the students, at both undergraduate and postgraduate level.

2   Background: the role of on-line MT

For the vast majority of Internet users the access of on-line texts written in an unknown language calls for swift on-the-fly translations in real time for gisting or skimming purposes. In these cases raw MT output into a more familiar language offered by free on-line services can prove a viable assimilation tool for many users.

At the same time, many designers of monolingual web-sites rely on free web-based MT for dissemination purposes, i.e. they are willing to use it to spread the informative contents of their virtual pages in more than one language. An amazing number of monolingual Internet sites today suggest to their visitors to take advantage of free web-based MT services to translate the contents of their web-pages into other languages.

From the perspective of the MT-related industry and vendors, free on-line services provide an ideal showcase for MT technology, and may represent a privileged channel through which casual users are exposed to it for the first time. Free web-based MT systems are maintained and powered by vendors of commercial MT products, which may win new market shares of users who get to know about MT through the World Wide Web, and appreciate the valuable help it can provide.

Finally, it is worth noting that web-based MT services have recently received attention in a number of non-specialist publications that are very widely circulated.   Yang   &   Lange

(forthcoming:190), for instance, present a list of articles that appeared in the early part of 1998 in assorted US newspapers and journals. Over the last few years magazines, newspapers, newsletters, computing journals, etc. published in a variety of languages have included articles and special reports dealing with on-line MT, which apparently could be of interest to a wide readership.

3   On-line MT services in teaching MT

The attention devoted by the media to Internet-based MT reinforces the point that it has a strategic importance to promote the knowledge of translation software altogether among non-specialists, such as casual and "one-off" users. All the more reason, then, to present the strengths and weaknesses of free on-line systems within MT courses, making sure that students get familiar with them by being themselves exposed to such popular web-based applications.

Even though the scenario depicted in the previous section suggests that free on-line MT attracts the interest of a diverse and large population of users as well as of important players in the MT community, it is surprising that on-line MT services do not seem to have so far gained credibility and popularity within MT instruction as a useful and convenient resource.

To the best of my knowledge, in the literature devoted to teaching MT there are very few superficial references to free web-based systems, possibly because of their recent introduction and relatively new proliferation on the Internet.

In the survey presented in Balkan et al. (1997), for instance, the possibility to use on-line MT systems is referred to twice[1], but only in passing. Similarly, Balkan

---

[1] Cf. sections "Course Resources" and "Availability of systems" at the following URLs [accessed 7 October 2002] http://clwww.essex.ac.uk/group/projects/MTfor Teaching/index_8.html   and http://clwww.essex.ac.uk/group/projects/MTfor Teaching/index_12.html.

(2001:9), Clavier & Poudat (2001), Kenny & Way (2001:16), Yuste Rodrigo (2001) and Gaspari (2001) also mention various ways in which on-line MT services can be exploited for teaching purposes, but without providing a specific focus or discussing the implications at length. This paper, then, intends to draw attention to the potential of their use in teaching MT, discussing both the advantages and the drawbacks of integrating and using free on-line services in MT courses.

4   Advantages of on-line MT

The strongest point in favour of some of the most popular on-line MT services available on the Internet is of course that they are accessible for free (cf. the appendix, which presents some more detailed information about six of them). This is essentially the reason why the present author used them for the MT courses he gave, as has been mentioned in Section 1. As a matter of fact, in order to have access to on-line MT services for demonstrations in the classroom or practical hands-on sessions there is only a need for a computer lab whose machines have Internet access.

This crucial competitive advantage over PC-based software allows instructors to circumvent a serious problem that is often mentioned in the literature about teaching MT. That is to say, courses usually take place under the constraint of very limited budgets, which makes it sometimes hard to have software for the students to carry out experiments, exercises, projects, etc.

In this respect, a number of MT instructors have in fact reported on various occasions that most of the time it is difficult to buy commercial MT software and acquire the necessary licenses at affordable prices, i.e. negotiating convenient agreements with vendors, who seem to be little inclined to grant discounts or favourable purchasing conditions to

academic partners for teaching purposes. As a result, free on-line MT systems can represent successful resources to be employed in MT courses for demonstrations or hands-on experimental sessions, overcoming financial difficulties altogether.

Another key advantage that free web-based services have, if compared to traditional PC-based software for teaching and learning purposes, is the wide coverage of language pairs and the choice of possible combinations, certainly much more extensive than what is offered by most commercial MT systems available for purchase. A number of free on-line MT services cover a wide variety of language pairs, and for instance the six web-based MT systems described in more detail in the appendix cover between 10 and 20 language pairs. Hutchins & Hartmann (2002) provide a comprehensive and up-to-date listing of MT systems that translate Internet and Web content, some of which are accessible for free over the Internet.

On the whole, the spectrum of language combinations on offer is quite wide, and on-line MT services can be helpful to cover a range of languages that students may be familiar with, also giving them the possibility to experiment with others, if they want to. Along these lines, for instance, Somers (2001:27) mentions the possibility of testing MT software "for assimilation" purposes, i.e. when students use as input source texts written in languages that are unknown to them.

As might be expected, English is the most widely available language in web-based MT, especially as a source language, but it is very often found as the target in several other combinations, too. By and large, the languages covered by free MT services mirror those offered by commercial PC-based software. This situation is not surprising, since vendors and manufacturers of commercial MT systems offer on-line services to promote the sales of their own PC-based products, as has been briefly mentioned in Section 2.

One crucial difference, however, is that the light versions available on-line of the mainstream MT products do not require users' licences or any installation process, even though they are usually based on or derived from the core engines of recognised commercial general-purpose MT systems. As a result, they provide access to a realistic working environment that can give a flavour of the challenges involved in using MT for a variety of language pairs in real productive scenarios.

On top of that, on-line MT systems are fairly flexible tools since they usually enable the user to translate either passages of plain text or entire web pages. In the former case, students involved in a lab session could for example type in challenging sentences (cf. Somers, 2001:26-27), or alternatively copy and paste into the appropriate fields paragraphs or excerpts of their own essays, and request the output in a target language of their own choice, in order to test the system; in the latter instance, on the other hand, students can provide URLs and have the textual contents of the corresponding web-page machine-translated on demand (see Section 6 for a more in-depth discussion of this point).

5 Disadvantages of on-line MT

One crucial factor that differentiates on-line MT services from traditional PC-based translation software and deserves special attention here concerns the possibility to augment or customise the internal components of the system at various levels, e.g. in terms of adding or modifying the information contained in the internal lexicons. This seems to be a noticeable drawback of web-based MT systems for teaching purposes, since with commercial PC-based MT software it is usually possible to provide extra linguistic information or change internal rules, with the purpose of increasing the quality of the output or establishing user-specific

preferred translations for some terms and phrases.

When working or experimenting with this kind of commercial MT software, students can for instance be asked by the instructor to change some rules of the system's internal architecture, extend the coverage of the topical glossaries and update the entries of the internal dictionaries, etc. (cf. Somers, 2001:27 and Yuste Rodrigo, 2001:46-47).

Web-based MT systems, on the other hand, are offered as so-called "black-box" applications, i.e. the internal components of the engine are not flexible or adaptable in any way, so that fine-tuning or customisation based on user-specific requirements and preferences are not possible. This point is summarised by the convincing observations provided by O'Connell (2001:ev) as follows:

> COTS [commercial off-the-shelf (MT systems)] users can add to the system's store of linguistic knowledge by teaching the system new words or phrases. […] Free Web MT users must depend on the provider's vocabulary selections.

This applies to Internet users and as a consequence to students when on-line MT systems are integrated into courses and teaching activities for practicing or testing purposes. It should be noted, however, that sophisticated adjustments to the internal architecture of MT systems (e.g. adding an entry to the lexicon or changing an existing one to account for subtle differences in meaning) may be quite time-consuming, and would require some familiarity with the internal formalisms and rules of the MT engine, as well as in-depth knowledge of the languages involved in the translation process.

These activities may thus be irrelevant to some teaching environments or to certain courses in which the students are not expected to possess the necessary technical expertise and advanced linguistic

skills to supply the system with extra information of this kind. Augmenting or customising on-line MT services is not an available option, thus leaving no room for manoeuvre for end-users in this respect, but under certain circumstances this drawback may not affect the effectiveness of their use for teaching activities.

6   On-line MT of entire web-pages

Just like on-line MT for passages of plain text, the translation of entire web-pages takes place in real time on an on-demand basis; in this case, however, the output is a web document (identified through its URL which is supplied by the user) whose textual contents are displayed in the target language selected by the on-line MT user, but the graphic appearance of the web-page is the same as the original (e.g. in terms of hyperlinks, frames, banners, etc.).

In other words, when translating entire URLs, on-line free MT services provide the user with a "clone" of the original web-page which looks exactly like the source document, in that it presents unchanged font and colours, has unaltered active hyperlinks, banners, etc., but at the same time displays textual contents in the target language chosen by the user. Web-surfers can therefore access the machine-translated version of an *ad hoc* created document, which would otherwise be unintelligible in the original language.

This function will be specifically focused on in the rest of the paper, since its success extends beyond the linguistic performance of the on-line MT service, and covers a wide range of sophisticated operations and a number of layers that interact during the web-based translation process.

Furthermore, this mode of use is particularly popular as an aid to facilitate Internet navigation among surfers with limited multilingual abilities, thus representing a very helpful resource and a powerful tool for the assimilation of multilingual information on the World Wide Web. It is quite apparent that employing on-line MT for this purpose has a promising future intertwined with the potential for further development of the World Wide Web.

7   Web localisation and multilingual on-line content management

Providing multilingual versions of the contents of a web-site for which originally only a monolingual validated source exists is the task commonly referred to as "*web localisation*". The term "localisation" indicates in general the complex process of adapting products to meet the linguistic, cultural and legal requirements of specific environments or target regional markets scattered around the world, which are called *locales* (cf. Esselink, 2000 and Esselink, forthcoming).

Localisation processes do not only affect web-sites, but a variety of other products as well, such as software packages and multimedia products (e.g. CD-ROMs, video-games, etc.). However, due to the main topic of this paper, that is to say the role of on-line MT services in teaching MT, the emphasis will be laid on the localisation of web content. As a matter of fact, in order to be accessed and operated by people belonging to different cultures and located in several areas across the world, web-sites require extensive and careful adaptations which extend well beyond switching to a new target language.

Web localisation, then, encompasses translation as one of its core components but is not restricted to it, since it adds to the linguistic aspects a technical and a content dimension. Localisation projects are usually very complex, and can often require teams of people with diverse expertise working simultaneously on each of the languages into which a web-site needs to be localised. Web-sites that rely on extensive budget and human resources to localise their contents can invest

substantial amounts of money and time in *multilingual on-line content management.*

The need to localise web material arises in fact in connection with multilingual on-line content management, which refers more specifically to the work, the techniques and the tools that are necessary to create and maintain up-to-date information that may be accessed on-line in a variety of languages. However, problems crop up in keeping this information accurate and simultaneously aligned in all the multilingual versions of the web-site, especially when frequent updates of the contents take place.

One point to be taken into account in this respect is that the versions in various languages need not be parallel: on the one hand, some general information may be necessary in all versions of the web-site (translated into, say, four new languages added to the original one, which may supposedly be English); on the other hand, however, some sections may be language- or even country-specific, so as to request particular treatment according to local needs, and should therefore not be included for the other languages:

> In the ideal world, you would translate your entire site into each of the languages important to your audience as well as add local content of interest to your customers in those countries (Nielsen, 2000:322).

The problem becomes very serious and absorbs substantial resources for highly dynamic web-sites, such as those with frequent extensive up-dates that often re-style or re-arrange their contents, or tend to add new sections and remove old ones at short intervals of time. In this respect, one may think for instance of the web-sites of on-line press agencies, governmental bodies, electronic journals with editorials and regular weekly or daily columns, etc.

For such complex tasks which require very tight deadlines for up-to-date information to be available without delay in more than one language to virtual visitors, hiring localisation service providers all the time would simply be financially not feasible, and also very difficult to manage effectively time-wise. Seen from this perspective, then, Internet-based MT offers real advantages from the standpoint of multilingual on-line content management of web-sites.

8 Web localisation, multilingual on-line content management and Internet-based MT

Checking and synchronising the available textual information in multiple languages is a very demanding and laborious task, especially for large corporate, e-commerce and institutional Internet sites with a complex and dynamic nature (see e.g. Cheng, 2000 for an interesting discussion on globalising and making multilingual the contents of an e-commerce web-site).

However, provided that web-masters keep on-line an up-to-date version of the original or official version (e.g. in English) of a web-site, on-demand MT requested by Internet users who provide the appropriate URL of a web document is always based on its latest update available on-line. As a result, synchronisation will always be maximised through the employment of on-line MT services, and web-surfers can navigate through exactly parallel machine-translated "cloned" multilingual versions of the original web-site in as many target languages as the on-line MT service supports.

The use of free Internet-based MT technology integrated into web-sites is becoming increasingly popular, even though it cannot effectively deal with the difficult management of region- or country-specific content. Moreover, one could argue that exposing Internet users to MT output in order to disseminate information in multiple languages entails controversial issues.

However, adopting on-line MT services to translate the contents of web-sites provides a time-saving and effective strategy to circumvent the need for more traditional approaches to multilingual on-line content management, which would be unavoidable if the multilingual versions of the site were to be provided by web localisation. Being a language-oriented task, on-line MT performs a very basic *pseudo*-localisation of Internet documents at the linguistic level, without however adapting their cultural features or graphic style in any way.

In this perspective, Internet-based MT contributes to bridge an existing communication gap when the contents of web-sites cannot be (or more simply are not already) professionally localised into multiple languages. In summary, the focus on some of the main issues brought about by web localisation and multilingual on-line content management should have helped to establish some links with the usefulness of free on-line MT services, showing that there is an interesting common area.

9   On-line MT and related areas

The present author is convinced that showing some functionalities of on-line MT services (like for instance those illustrated in the previous section) is useful to raise a number of issues that are not exclusively relevant to MT. This can be done by exposing the students to real-life applications in which web-based MT can be employed, irrespective of what course they are taking. As a matter of fact, providing some insight into the role played by free on-line MT technology within the wider framework of multilingual communication processes on the Internet can provide food for thought for students of different disciplines.

The broader approach that is referred to here is of course to be provided or encouraged by the MT instructor, for instance by means of lectures on the topics of web localisation and multilingual on-line content management, presented to the students according to their background, previous experience and main interests. Otherwise, if the group of students can be more receptive and thus likely to come up with clever insights and intuitions, MT instructors can adopt more indirect strategies, e.g. prompting discussions and spontaneous debates in the classroom, or giving essays and assignments, aimed at providing the students with opportunities to comment and express their views on the usefulness of on-line MT services in a wider perspective. In this way, a reflection on the *role* (rather than only on the *performance*) of web-based MT in Internet-based multilingual communication processes can be elicited from the students, by means of an inductive process.

In any case, if this wider approach is pursued, the responsibility of making clear to the students what are (if any) the points of contact between the role of web-based MT services and, for instance, on-line multilingual content management and web localisation, lies with the instructor. Insofar as it brings together a variety of perspectives, this approach goes beyond the consideration of on-line MT *per se*, or simply as a particular form in which MT systems can be accessed and used. Rather, it hinges on the potential for the actual exploitation of MT technology in multilingual communication processes.

Along these lines, it is the contention of this paper that looking at free on-line MT services within this broader framework can be of interest to students with a number of different backgrounds taking various courses at both graduate and undergraduate level: translation studies, applied and computational linguistics, natural language processing, information technology, computer science, information retrieval, localisation, etc.

For instance, while teaching trainee translators about MT, the present author has made an attempt to illustrate more

general points by means of the on-line MT services that the students were using for their hands-on lab sessions. Exposing undergraduate students majoring in translation studies to the use of web-based MT systems has proved successful in raising their awareness on a number of issues that are crucial in today's multilingual communication processes, and which they would come across in their profession.

It was felt for instance that the role played by web-based MT services translating entire web-pages could shed some light (either by contrast or by analogy, depending on what particular aspect was taken into consideration) on the priorities in the management of web localisation projects, also providing insights into the typical translation workflow of companies, institutions and government bodies that adopt multilingual policies and have multilingual web-sites.

10 Conclusion

This paper has discussed in what ways free on-line MT services could be used in teaching MT courses. Drawing on the experience of the author in designing assignments and hands-on practical sessions aimed at trainee translators that relied on the use of free web-based MT systems, their main advantages and drawbacks have been described.

In conclusion, the paper argues that integrating the use of on-line MT services into MT courses can be highly beneficial irrespective of the background and syllabus of the students, at both undergraduate and postgraduate level. Their integration into teaching and practicing activities is a very feasible and convenient option that can contribute significantly to enhance the overall quality of MT instruction.

At the same time, the paper has presented some of the reasons why the horizons within which MT is taught can be widened by exposing the students to on-line MT services. As a matter of fact, web-based MT technology provides the opportunity to illustrate a number of issues that have a crucial importance in today's multilingual communication processes that take place over the Internet (e.g. in connection with web localisation and multilingual on-line content management), and as such can be of interest to many types of students.

Acknowledgements

References

Balkan, L., D. Arnold & L. Sadler (1997). *Tools and Techniques for Machine Translation Teaching: A Survey.* [Available on-line at `http://clwww.essex.ac.uk/group/projects/MTforTeaching/index.html` - Accessed 7 October 2002].

Balkan, L. (2001). "Exploiting the WWW for MT teaching". In M.L. Forcada et al. (2001). 7–11.

Cheng, S. (2000). "Globalising an e-Commerce Web Site". In R.C. Sprung (2000). 29-42.

Clavier, V. & C. Poudat (2001). "Teaching Machine Translation in non Computer Science Subjects: Report of An Educational Experience within the University of Orleans". In M.L. Forcada et al. (2001). 19–23.

Esselink B. (2000). *A Practical Guide to Localization.* Amsterdam: John Benjamins.

Esselink B. (forthcoming). "Localisation and Translation". To appear in H. Somers (forthcoming). 233-251 (provisional page numbers).

Forcada, M.L., J.A. Pérez-Ortiz & D.R. Lewis, eds. (2001). *MT Summit VIII Workshop on Teaching Machine*

Translation. Santiago de Compostela, Spain. [Available on-line at `http://www.dlsi.ua.es/tmt/proceedings.html` - Accessed 7 October 2002].

Gaspari, F. (2001). "Teaching Machine Translation to Trainee Translators: a Survey of Their Knowledge and Opinions". In M.L. Forcada et al. (2001). 35–44.

Hutchins J. & W. Hartmann (2002). *Compendium of Translation Software. Commercial machine translation systems and computer-aided translation support tools. Version 1.4 (Revision June 2002).* Genève: European Association for Machine Translation. [Can be downloaded from `http://www.eamt.org/compendium.html` - Accessed 7 October 2002].

Kenny, D. & A. Way (2001). "Teaching Machine Translation & Translation Technology: A Contrastive Study". In M.L. Forcada et al. (2001). 13–17.

Nielsen, J. (2000). *Designing Web Usability: The Practice of Simplicity.* Indianapolis: New Riders Publishing.

O'Connell, T. (2001). *Preparing your Web site for machine translation. How to avoid losing (or gaining) something in the translation.* [Available on-line at `http://www-106.ibm.com/developerworks/library/us-mt/?dwzone=usability` - Accessed 7 October 2002].

Somers, H. (2001). "Three Perspectives on MT in the Classroom". In M.L. Forcada et al. (2001). 25–29.

Somers, H., ed. (forthcoming). *Computers and Translation A Handbook for Translators (provisional title).* To be published by John Benjamins.

Sprung, R.C., ed. (2000). *Translating Into Success: Cutting-edge strategies for going multilingual in a global age.* Amsterdam: John Benjamins.

Yang, J. & E. Lange (forthcoming). "Going Live on the Internet". To appear in H. Somers (forthcoming). 187-205 (provisional page numbers).

Yuste Rodrigo, E. (2001). "Making MT Commonplace in Translation Training Curricula – Too Many Misconceptions, So Much Potential!". In M.L. Forcada et al. (2001). 45–49.

Appendix

The following table provides some information about six free on-line machine translation services. For each web-based MT system the following details are given[2]:

- *Name of the free on-line MT service*
- *URL of the default home page*
- *Number of language pairs offered (as stated on the web-site)*

Free on-line MT service	BabelFish	
URL	`http://babelfish.altavista.com`	
Number of language pairs	19	

Free on-line MT service	FreeTranslation	
URL	`http://www.freetranslation.com/web.htm`	
Number of language pairs	11	

Free on-line MT service	Google Translate BETA	
URL	`http://translate.google.com/translate_t`	
Number of language pairs	12	

Free on-line MT service	Gist-In-Time	
URL	`http://www.teletranslator.com`	
Number of language pairs	18	

Free on-line MT service	Lycos/Systran	
URL	`http://translation.lycos.com`	
Number of language pairs	18	

Free on-line MT service	Voila	
URL	`http://tr.voila.fr`	
Number of language pairs	13	

---

[2] Disclaimer: the information presented in this table has been checked on-line and is correct as of 7 October 2002.