Online and Free! Ten Years of Online Machine Translation: Origins, Developments, Current Use and Future Prospects

Federico Gaspari

School of Informatics, University of Manchester
PO Box 88, Manchester M60 1QD
United Kingdom
F.Gaspari@manchester.ac.uk

John Hutchins

89 Christchurch Road
Norwich NR2 3NG
United Kingdom
w_john_hutchins@yahoo.co.uk

Abstract

Marking the ten-year anniversary of the launch of Babel Fish, the first ever free online machine translation (MT) service that went live on the Internet in late 1997, this paper sketches the background that led to its development, giving an account of its origins and of the early stages of its evolution. Several competitors have entered the field of web-based MT over the last decade, and the paper offers a review of the most significant contributions in the literature with a particular focus on two key issues: firstly, the role that these online MT tools have played in meeting the translation needs of the users, and secondly the impact that they have had on the MT-related industry and business. Information coming from a variety of sources, including data on current usage supplied by the online MT providers themselves for the purposes of this study, testifies to the massive increase in the use of the leading multilingual online MT services over the last ten years. On this basis, the conclusion assesses the future prospects of Internet-based MT.

Keywords: online MT, Internet, Web, history

1. Origins and Early Developments

Fulfilling predictions since the late 1980s that MT services would become available on the Internet, the first online MT service was provided from 1988 by the Systran Centre in Paris to subscribers of the French postal service's Minitel network (restricted to France). Users could send texts for translation from their PC or Macintosh and receive results (22 lines a minute, at a charge of about $1.20 per page) – the language pairs offered were French to English, German to English, and English to French (Gachot, 1989). According to Ryan (1987: 100), MT software provided by Systran was potentially accessible to 4.5 million users of Minitel in France.

The next concrete proposal was for an online service that would be available more widely. In September 1992, it was announced that CompuServe was investigating the possibility of offering MT to its subscribers (Harrison, 1992). This project involved a six-month evaluation period to test the output quality and the overall performance of existing MT systems for the language pair English-German (in both directions): "CompuServe's basic goal for MT is to provide draft-quality translation directly to end users. […] We suspect that there is a market for low-cost translations, even if the quality is less than ideal" (Harrison, 1992: 11). Although not yet offering online MT as it is commonly understood today, i.e. in the form of services that users can directly log on to in order to have input texts or webpages of their choice translated, the pioneering experiences at CompuServe laid the foundations for further crucial developments in the following 15 years which are the focus of this paper.

Mary Flanagan, a computational linguist based in the USA, led the Advanced Technologies Group at CompuServe from 1992 until 1998, and regularly reported the groundbreaking developments at CompuServe. For example, Flanagan & Jensen (1994) describe the early implementation of an entirely automated MT process, whereby the messages posted in English on selected CompuServe forums were periodically collected, fed through Intergraph's Transcend system, and the output in French and German displayed in parallel versions that could be read online by people unfamiliar with English. The original forums (in English) and the machine-translated versions (in French and German) presented the same contents and the same structure, i.e. the threading and sequence of the multilingual postings were identical. The usage statistics reported in Flanagan (1995) are quite impressive: MT was used for translation between English and three other languages, namely French, German and Spanish, and during the first month of operation, on just one of the more than 600 specialist interest forums available on CompuServe, more than 900,000 words were translated at a speed of over 3,000 words per minute. Details were also given of further plans and services (e.g. a low-cost post-editing service for email translation) to be launched for CompuServe subscribers, with an assessment of the commercial opportunities offered by the deployment of MT in the online environment, as well as of the associated challenges.

In a later report, Flanagan focuses on the usage patterns and on the reactions of the users who are exposed to MT output for the first time in the online environment: 25% of them abandon the service after receiving the first translations, possibly because they are surprised by the poor quality of the raw output and find it impossible to understand or use effectively. Interestingly, following the launch of their online MT facility, CompuServe received "hundreds of angry e-mail messages, as well as hundreds of resumes from translators" offering their services (Flanagan, 1996a: 193). The report also says that "users were overwhelmingly satisfied with the quality of the translations […] and several large users routinely submit jobs totalling more than 10,000 words per week" (ibid.: 194). Flanagan (1996b: 244) summarises the philosophy of these initial attempts to offer online MT to registered users as follows:

CompuServe has taken a pragmatic approach to MT technology, focusing on finding a market niche for what it can do – generate rapid, very rough draft, information-scanning quality translations in an environment where quick scanning for content is more important than high quality.

However, retaining some quality was still a concern. Consequently, CompuServe's Document Translation Service offered its subscribers the possibility of uploading documents for MT and requesting an optional post-editing service, which was charged at a rate per-word that was ten times higher than the rate for raw MT output (Flanagan, 1996b: 245). Flanagan (1997a: 25) states that 85% of the requests were submitted for unedited translation (i.e. raw MT output), and that additional professional post-editing at the higher rate was requested in only 15% of the cases. Commenting on CompuServe's MT-related services, Bonthrone (1996: 4) reports that post-edited jobs tended to be larger than those for which only raw MT was requested. As a result, in terms of word count there was a more balanced ratio of roughly 60% raw MT output vs. 40% post-edited content. Flanagan perceptively referred to online translations as "MT's new frontier" (Flanagan 1997b) – as the following years were to demonstrate.

Whilst the importance of the role played by CompuServe in introducing MT to a large population of new users (more than two million, according to *MT News International*, 1999: 15) cannot be overestimated, the access to MT provided by CompuServe via the Web was still restricted, in that it was limited only to registered subscribers. The online MT scene changed radically on December 9, 1997 with the launch of Babel Fish, which made MT available free of charge to any Internet user. It was the result of a partnership between Systran Software Inc. and the well-known search engine AltaVista.[1] After its experience with Minitel in France, Systran had started to offer online translations of webpages from its own website since 1996 (Yang & Lange, 1998: 276). Now, it made the service more widely available.

2. Online MT and Its Use

It was Babel Fish that ensured the unprecedented global visibility and accessibility of MT on the Internet. However, it came with a few surprises. Yang & Lange (1998: 282), reporting on users' feedback and usage behaviour in the first few months of its operation, observed significant usage in areas that were either not anticipated by the developers and providers of the service (e.g. as a tool for language learning, cf. McCarthy, 2004 and Somers et al., 2006), or deprecated by them (i.e. as an entertainment tool, getting it to perform 'back-and-forth' translations or to translate idioms). Initially Babel Fish covered ten language combinations involving five major European languages, and Yang & Lange (2003) provide an update on its usage by a growing group of Internet users (more information is available in the Appendix).

They report that between 1998 and 1999 the most popular language combination was for translations from English into Spanish, followed by English-French, German-English, Spanish-English, English-German and French-German. They reported a wide variety of motives by users (assimilation, dissemination, communication, language learning, and of course entertainment), and an amazing variety of texts were fed to the system, from chatroom jargon to X-rated material, adult content, taboo words and risqué terms. They revealed that over half the 'texts' submitted were less than five words, and only a quarter were longer than 20 words (Yang & Lange, 2003).

Within a short time after its launch, the Babel Fish service was no longer unique. Other MT vendors joined in offering free online services. By 2000 there were more than ten companies involved; apart from Babel Fish (AltaVista) and CompuServe, these included FreeTranslation (Transparent Language Inc.), Gist-in-Time (Alis Technologies), iTranslator (Lernout & Hauspie), MT Ave (Toshiba), My Translator (Apex Network), PARS (Lingvistica), ProMT, and Reverso (Softissimo). In many cases, these free services were augmented by charged post-editing and/or human translation services. Currently there are over 30 free online MT services (see current issue of *Compendium of translation software*).[2]

A major factor in the rapid popularity of web-based MT services has undoubtedly been that they are available free to users,[3] and that results are (almost) instantaneous. Limitation of the amount of text has been no impediment for most users. Offering an MT service without charge was clearly seen as a means of promoting sales of full MT systems, both to the general public and to companies – particularly since purchased systems would not have the limitations (in length of texts, functions and facilities) of the online service. To what extent the MT vendors have benefitted in this way from their free online services is unknown. However, it may well be that they have profited from the leasing of their software to other web-based services (e.g. news and current affairs providers).[4]

Information on current use of online MT from a Japanese perspective comes in a report by Yamada et al. (2005) on the MT market in Japan. A questionnaire-based online survey elicited information from 4,000 respondents between February 2003 and February 2005. The data show a slight but steady increase in the use of online MT services in this period. Not surprisingly, the overwhelming majority of MT activity in Japan involves English and Japanese. Those with limited knowledge of English are particularly likely to use online MT to translate content of websites available only in English. However, no specific information is given on MT usage by those with no knowledge of English at all. In this period, on the other hand, there was a 5% rise in the number of Japan-based professional translators using online MT as part of their work.

The availability of free translation services has had some less attractive consequences. It has given the opportunity for users to exploit the known inadequacies of automatic translation. One particularly unfortunate example of the questionable use of Babel Fish – and, for

---

[1] In 2006 the service was acquired by Yahoo!, becoming officially known as Yahoo! Babel Fish (Flournoy, 2006).

[2] Available at http://www.hutchinsweb.me.uk/Compendium.htm.

[3] But not necessarily free to providers; services such as AltaVista and Yahoo! pay a fee to Systran.

[4] Economic aspects are not covered here; for obvious reasons, providers are unwilling to disclose commercially sensitive data.

that matter, of any MT software – is provided by Watters & Patel (1999), who attempted to evaluate Babel Fish by translating proverbs from English into other languages. They argue that:

> using a set of the most commonly known proverbs in English, it should be possible to evaluate how well direct translation systems are able to process semantic information, and whether they correctly select the appropriate sense of a word, where multiple senses exist. (ibid.: 155)

Their evaluation method consisted in translating four proverbs into each of the five target languages initially supported by Babel Fish in combination with English, and then back again into the original source language, and trying to account for the mistranslations that occur in the process. On this fairly impressionistic basis, the authors analysed the data in detail, and, not surprisingly, they found that the results in general were rather disappointing. However, the authors do concede that their conclusions "are limited to the extent that the translation performance of expert human translators was not tested against" the online MT service (ibid.: 159).

In subsequent years, numerous commentators have enjoyed finding fault with online MT – and, by implication, with MT systems in general (e.g. Budiansky, 1998). The principal method is to input sentences which contain one or more ambiguous words or ambiguous syntactic structures. Naturally, the results are garbage and often amusing.[5] A major problem with the use of online MT – one recognised by the MT community – is that most users are not aware of the limitations of MT. Church & Hovy (1993: 246) emphasised that "it should be clear to the users what the system can and cannot do", whatever the type of MT service – but particularly, for a service intended for large-scale use by the general public.

3. Use and Misuse

Before the appearance of free online MT, the acceptability and usefulness of 'less-than-perfect' MT was a matter of discussion. Church & Hovy (1993) argued that the MT community should seek 'niche applications' where poor quality ('crummy') MT would be acceptable, where expectations were not too high, and the service "should be attractive to the intended users" (ibid.). At the time, their suggested applications involved variations on existing practice, such as rapid post-editing and draft versions for human translators. They did refer, however, to previous experience of users with 'raw' unedited output (going back to the 1960s and the Georgetown system – cf. Henisz-Dostert, 1979), who were often happy with unedited MT – particularly if otherwise no translation was available. There were clearly many situations in which 'crummy' MT would be acceptable, and online MT represents just such a 'niche application'.

Such has been the uptake of free online MT that it is now arguable that this form of MT is more widely known to the general public than the corporate use of edited and unedited MT in the production of technical material and

in the circulation of administrative documentation. Even though there are a few exceptions, it is therefore all the more strange that the MT community has largely ignored discussion of these services and their impact on the image of MT in the wider world. (The bibliography of this paper lists a high proportion of the total English-language literature on the topic.)

McLaughlin & Schwall (1998) outline the MT products and services with the greatest potential on the Internet. They present a case study focusing on Lernout & Hauspie, at that time a leading provider of MT solutions on/for the Web, reporting on its customer base, tools and products. Ananiadou (1998), reviewing trends in MT in Europe and Japan, provides details on language coverage, modes of use and conditions of service of four leading commercial providers of online MT (two based in Japan and two based in Europe), and illustrates the dynamism and growing demand in this area.

Prompted by the unanticipated ways and unreasonable expectations of quality which some users have of online MT, Bennett (2000) examines its merits and drawbacks, and explains some of the major challenges involved in processing unpredictable input. A technique often used intuitively by some users to evaluate online MT is the so-called "round-trip translation" (or RTT, also sometimes referred to as "back-and-forth translation"). It is, however, a technique without solid theoretical or empirical foundations. Somers (2005) demonstrates that RTT is not as useful as some lay-users of MT on the Web may think. He conducted two separate experiments, based on a variety of texts and involving five different free online MT services (i.e. Babel Fish, FreeTranslation, Systran, ProMT and Worldlingo). His conclusion is that although non-experts in MT might see some value in it, the RTT technique does neither help to reveal the quality of a particular MT system nor to indicate the "machine translatability" of a specific text.

4. Translators and Online MT

Fulford (2002) reports on an exploratory study conducted between 2001 and 2002 to investigate the uptake of MT among freelance translators based in the United Kingdom. Of the 30 individuals who were interviewed, only two (i.e. less than 7%) "were actively using MT in their work" (Fulford, 2002: 119). Interestingly, however, eight of the professional translators who were interviewed (26%) "stated that they had 'occasionally' made use of web-based MT systems to produce an initial rough draft of a translation, or to 'get ideas' for producing a translation, before polishing the output manually ready for presentation to a client" (ibid.). This suggests that although, in general, professional translators are reluctant to invest in MT software and to integrate it into their working practice, online MT is seen by some as a potentially valuable translation aid.

Fulford & Granell-Zafra (2004a) give a progress report with 390 responses received from freelancers. The authors conclude that "[b]eyond terminology and document consultation / look-up, there was little or no actual use being made of online systems, such as MT" (Fulford & Granell-Zafra, 2004a: 59). In another paper, providing additional data from the same survey, Fulford & Granell-Zafra (2004b) report that only 3% of the 390 freelance translators who took part in the survey made use

---

of online MT systems as part of their work (ibid.: 41), compared with much higher percentages for the use of other Internet-based tools such as e-mail (93%), search engines (85%), online dictionaries and glossaries (78%) and multilingual terminology databases (59%).

5. Online MT and the MT Community

The ready availability of online MT systems on the Internet over the last ten years has had direct and indirect effects on all those involved with MT, namely not only on the users, but also on the wider MT research community and especially on the MT industry and the MT vendors. The earliest indication dates back more than two years before the launch of Babel Fish, recognising the pioneer developments by CompuServe at that time. In a survey of the users and usage of MT in Europe and the Americas, Brace et al. (1995) predict an "upsurge in the use of MT on-line", which they call an "impressive development". The paper also refers to the experience of Internet-based providers who offered added-value services to customers requiring post-edited polished versions of online MT output, demonstrating that in the mid-1990s the MT support services were attracting growing commercial interest.

At the AMTA conference in 1996, two of the speakers on a panel "MT Online: The Future is Now!" (AMTA, 1996: 220 ff.) were David Clements of Globalink and Patricia O'Neill-Brown of the US Department of Commerce. Clements (1996) presents a number of scenarios in which the availability of online MT is the ideal solution to real communication problems. However, the language used in Internet-based exchanges is often stylistically and grammatically sloppy, and thus presents unprecedented and largely unpredictable challenges. On the other hand, he emphasises that translation technology is essential to enable communication and interactions on the Internet, and concludes that the "convergence of the Internet, e-mail and various online services, as well as the increasing popularity of the personal computer over the whole world, presents the greatest opportunity yet to bring MT "to the masses"" (Clements, 1996: 221). O'Neill-Brown (1996) points out that trivial but still important technical aspects can prevent the large-scale or effective deployment of online MT tools, that large volumes of documents are available only in paper form, and that issues of encoding can have adverse effects on MT processing of online texts. In a number of respects, the problems raised by O'Neill-Brown have been resolved in subsequent years.

Westfall (1996) raises a series of interesting questions regarding the legal implications of online MT, with particular reference to the potential risk of litigation and lawsuit against MT companies and providers. They may be liable if web-based MT services are used to translate, disseminate and distribute (on the Internet or otherwise) text protected by copyright, information of a commercially sensitive nature, or content that is illegal for whatever reason. Another potential liability might arise if an incorrect translation provided by an online system leads to safety violations. She concludes that "[t]he legal issues surrounding machine translation on-line will need to be defined within the next year" (Westfall, 1996: 231). Unfortunately, this has not yet happened, as a few years later Yang & Lange (1998: 282-283; 2003: 205-206)

comment on a number of outstanding legal issues with regard to Babel Fish where users are threatening to take legal action for specific incidents resulting from mistranslations, or are demanding that the service should provide a clear disclaimer that it cannot accept any liability in connection with its use (cf. Gaspari, 2004: 69).

In an overview of MT research and development in Canada, Macklovitch (1997: 204) claims that "the Internet is transforming classic MT". In considering the impact of this evolving scenario on the MT industry, he describes the commercial strategy of one major Canadian-based software company. The approach of Alis Technologies was to specialise at this time in the customisation and re-selling of MT products manufactured by other vendors, and to provide a one-stop solution to corporate clients needing Internet-based translation and multilingual online content management. The same author returns to this topic in a later paper, reinforcing these arguments on the basis of the developments and trends detected in the intervening few years (Macklovitch, 2001: 27), describing the attempts of the MT industry to capitalise on the demand for automated solutions to translation needs in the online environment. Similar discussions are also found in Smith (2001a) and Baron (2003: 118).

Allen (2000) focuses on the impact of online MT services on the MT market in the late 1990s, and offers a perspective on their value for the users as well as for MT companies. This contribution sheds light on the connections between the several free online MT services available on the Internet and the commercial strategies of a relatively small number of companies specialising in MT system development and implementation. When Internet users try out free MT services that are based on related commercial products, companies see an opportunity to make them paying customers who for a fee receive higher-quality translation services. They offer, for example, the possibility of translating unlimited amounts of text (whereas length limits and connection timeout constraints apply to non-paying occasional users) and the option of activating domain-specific dictionaries to improve the quality of translations of specialised texts (whilst free web-based MT services typically have only general-language dictionaries that cannot be customised or augmented).

6. Customising Online MT

Limitations on text lengths and on dictionary coverage are obvious impediments to corporate users. However, many of them are reluctant to invest time and money in full-scale in-house MT systems. For this reason, we are seeing increasingly the use of Internet-based MT engines as the basis for enhanced translation services. Smith (2001b) describes a pilot project to test the feasibility of offering an online MT application powered by Systran to the multilingual employees of PricewaterhouseCoopers (PwC) over the firm's intranet. The proposed MT facility covers 20 language pairs and users can select specialised dictionaries to improve the quality of the MT output. The quality of the results is variable, and additional fine-tuning was needed. Further developments are reported in Smith (2003): some customisation at the terminological level to tailor the MT dictionaries to the needs of the company; the number of language pairs covered by Systran nearly doubled, rising to 37; and, the most

popular language combination among PwC employees is for translations from English into Spanish, nearly double the next most popular, from Spanish to English – which is in line with usage for the free online service provided by Babel Fish. The paper also reports that approximately 130,000 translation requests coming from 7,300 individuals around the world have been processed by the MT engine since it was first introduced. User feedback indicates a prevalence of positive reactions, although there is also evidence of negative responses, and some possible future enhancements to the online MT facility are discussed.

Kübler (2002) focuses on similar efforts to enhance the performance of online MT services, but outside the corporate environment. This study addresses the challenges of combining available tools and resources to customise dictionaries used by online MT systems, in an attempt to enable technical translators to become more time-efficient in their work. The experiments make use of Systranet, an online MT service powered by Systran, which gives users access to a dictionary management facility to create and augment personalised dictionaries, to improve the quality of translations of specialised and domain-specific documentation. The paper focuses on the translation of technical material from English into French. After initial investment of time to identify relevant terminology and feed it into Systranet's customisable dictionaries, noticeable improvements were achieved in the quality of the raw output provided by the web-based MT system. The output produced with personalised dictionaries provides a target-language draft good enough for post-editing and polishing into a final text of professional standard, saving time and guaranteeing productivity gains.

7. Current Use of Online MT

The users of online MT are probably the largest group of MT users, and yet we know very little about them. How satisfied are they with the quality of online MT? How often do they use online MT? How well do they know the language(s) they want to translate into or from? What types of translation 'errors' are most irritating? Do they want more specialised services (e.g. for medical texts, for technical documentation)? How much do they use online MT for emails and chatrooms? Do they use MT for business purposes? What are their social, occupational and age groups? etc., etc. Much of the discussion about online MT is based on speculation. As far as we know, there has been no large-scale survey of users and of what they expect from online MT now or in the future. It could be instructive to hear the views of users themselves, particularly as this area of MT provision is now what the majority of the general public consider to be the most (or even 'best') that MT can do.

Very little data regarding the use of online MT services is publicly available, and most companies associated with web-based MT systems seem reluctant to publicise information regarding, for example, the nature of the texts submitted for translation or to reveal feedback received from the users (Zervaki, 2002), although usage patterns are constantly monitored. In addition, the providers of online MT regard information on a wide range of topics, particularly having to do with the inner workings of their web-based MT systems, as proprietary

and confidential, given its sensitive nature in a highly competitive field. To date, Yang & Lange (1998) and Yang & Lange (2003) are the key sources of information in this respect, although the perspective that they offer is rather limited and outdated, being focused exclusively on the usage of Babel Fish on two census days in June 1998 and November 1999.

In an attempt to find more up-to-date and representative information about the overall current usage of online MT services, and with a view to comparing it with the scant reports that are already available, the authors have approached a number of people directly associated with some of the leading online MT services. Eventually, we received usage data regarding three of the major providers of web-based MT systems, namely Yahoo! Babel Fish, FreeTranslation and Systran. The survey was centred on the following four core questions:
1) most popular and frequently used language pairs;
2) number of translation requests per unit of time (day, week or month);
3) ratio of translation types, i.e. for webpages vs. passages of plain text;
4) length of plain-text translations (number of words per translation job).
We also emphasised that we would welcome any other insight or detail that the providers of web-based MT were willing to supply to us for the purposes of this paper. The Appendix at the end of the paper shows the data on current usage that our contacts have kindly shared with us (alongside the information about Babel Fish for 1998 and 1999). As a result, the Appendix allows for some comparison across the leading online MT services, giving at the same time an overall picture of the scale of the traffic and of the volume of translation requests handled at present by the web-based MT systems involved in the survey.

Responses from MT providers about the use of their online services indicate that the most popular language pairs remain English to and from Spanish and English to and from French, and that (as expected) for each particular country, the most popular pair is English to and from the native language. Overall, the volume of online MT continues to grow – Flournoy (personal communication) reports that the advent of Yahoo's Babel Fish has not affected in any way the volume of traffic on AltaVista's MT service, which is still in existence and operating as usual.

What is perhaps surprising is that translation from webpages is much less common than plain text (only 2% of Yahoo! Babel Fish, less than 10% of FreeTranslation and no more than 17% on AltaVista is webpage translation). No less surprising is that most users are using the online services to look up or check translations of single words or very short phrases. This may perhaps mean that most users have some knowledge of the original language and require only occasional assistance, i.e. they use the services as electronic dictionaries. Some of these users may in fact be translators (as described in section 4) or readers seeking only to extract small pieces of information from texts. And a few may be 'testing' MT out of curiosity or for entertainment (as mentioned in section 2). This low use of online MT for translating more substantial texts (whether plain text or webpages) may suggest also that the number of users with poor or no knowledge of the source languages is relatively small.

However, as emphasised above, we really know almost nothing about who the users of online MT services are, what their language knowledge is, what they are looking for when using online MT services, and how much awareness they possess of the limitations and potential of MT software. We hope that this gap will be filled by future research efforts, possibly to be undertaken in close collaboration with the MT providers, who would certainly benefit from a deeper understanding of the reasons and circumstances leading users to access online translation services.

8. Conclusion: Future Prospects of Online MT

The comparison between the current usage of the major online MT services and the volume of translations handled by Babel Fish in the late 1990s, when web-based MT was still a newcomer to the Internet, indicates that over the years the growth in demand on the part of the users has been constant and substantial. To match this demand and offer a reliable and efficient service, the providers invest significant resources to support the growing amount of traffic. In addition, the developers are continuously working to improve the facilities and capabilities offered by the online MT systems (Flournoy, 2006), with a special emphasis on adding new language pairs likely to attract interest and enhancing translation quality, by creating more lexical entries in the dictionaries and including more powerful rules in the systems all the time.

The surprisingly low current use of online MT for translating texts longer than a few words or phrases (as reported by service providers) may mean that these free services do not yet represent as great a revolution in overall MT usage as might have been anticipated when they were first introduced. In other words, online MT is meeting a demand for occasional 'translation' that traditional translation services could not meet – a demand primarily for translation assistance from users already with some knowledge of the 'source' language. Although in volume terms this may currently be the dominant use, there is still a substantial demand for online translation of longer texts and for translation from unknown languages – even if it is only a small fraction of usage, out of the millions of translation requests per day it represents a large demand and one not in any way to be dismissed. This is a demand that the providers have to meet and one which is likely to grow over time.

If the trends observed in the past decade continue, the interest of Internet users for online MT services is set to remain high for the foreseeable future, and it can be safely assumed that the providers will try hard to benefit from this. It is not clear, however, whether over time the users of online MT services, particularly casual or occasional ones, will become more aware of the limitations of these systems, and will have more reasonable expectations regarding the level of quality that they can realistically offer. There is the likelihood that new and extended support services will be offered (e.g. easier options for post-edited or human translation, advice on how to implement MT-based multilingual communication strategies, etc.). A vital challenge consists in the education of general users on how to gain the maximum benefit from using online MT services, as advocated by Somers (2003: 523), with a view to strengthening the confidence of average web surfers in the use of web-based MT (Gaspari, 2006).

As yet, online MT (and, indeed, MT in general) gives poor results for the kind of colloquial and ill-formed language found in electronic mail, chatrooms and blogs. In future, there may be several online MT services devoted to such text types;[6] and it is to be hoped that more research effort will be directed to this neglected area – if only for the sake of the reputation of machine translation itself. Many MT service providers have access to vast caches of emails, webpages and blogs, which could be analysed and utilised by current statistical MT techniques (cf. Yang & Lange, 1998: 280). Perhaps it is being done already – we do not know.

More speculatively, the coming of Internet telephony suggests the ultimate possibility of online MT of spoken language – initially perhaps in constrained domains for some very specific tasks. More immediately, there must be the expectation that free online MT services will provide an ever wider range of language pairs. Translation from English is available into a large number of languages, although there is always room for more – particularly, as always, for languages of Africa and India. Translation into English is less well served – even some European languages are not provided for (Czech, Lithuanian, Polish). Free online MT services are available for some non-English pairs, such as Chinese-Japanese, Korean-Japanese, French-German, Spanish-Portuguese; but there is clearly a need for many more. What can be in no doubt is that online MT will continue to grow, and that it will become the principal focus of MT activity and research in the not too distant future.

References[7]

Allen, J. (2000) "The value of Internet translation portals". *International Journal for Language and Documentation* **6**(August-September): 45-46

AMTA (1996) *Proc. 2nd AMTA Conference*

Ananiadou, S. (1998) "Machine translation trends in Europe and Japan". *Proc. 20th Translating and the Computer*

Baron, N.S. (2003) "Language of the Internet". Farghaly, A. (ed.), *Stanford Handbook for Language Engineers*. Stanford: CSLI Publications. 59-127

Bennett, S. (2000) "Taking the babble out of Babel Fish". *Language International* **12**(3): 20-21

---

[6] The company Translution has launched an online service for translating electronic mail, with limited functionality (see http://www.translution.com).

[7] Papers from proceedings of AMTA, EAMT and MT Summit conferences are available at http://www.mt-archive.info.

Bonthrone, R. (1996) "Translation and the Internet – Part 2". *LISA Newsletter* 5(2). Available at http://www.fb-partners.com/pdfs/02-Translation_and_Net-0396.pdf [accessed 5 July 2007]

Brace, C., M. Vasconcellos & L.C. Miller (1995) "MT users and usage: Europe and the Americas". *Proc. MT Summit V*

Budiansky, S. (1998) "Lost in translation" *Atlantic Monthly* **282**(6): 80-84

Church, K.W. & E. Hovy (1993) "Good applications for crummy machine translation". *Machine Translation* **8**(4): 239-258

Clements, D. (1996) "The value of online MT in the age of the 'cyber society'". *Proc. 2nd AMTA Conference*. 220-221

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

Flanagan, M. (1995) "MT in the online environment: challenges and opportunities". *Proc. MT Summit V*

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

Flanagan, M. (1996b) "System demonstration: online MT at CompuServe". *Proc. 2nd AMTA Conference*. 244-246

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

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

Flournoy, R. (2006) "Say "hello" to my little friend". Posting dated April 27, 2006. Available at http://www.ysearchblog.com/archives/000295.html [accessed 5 July 2007]

Fulford, H. (2002) "Freelance translators and machine translation: an investigation of perceptions, uptake, experience and training needs". *Proc. 6th EAMT Workshop*. 117-122

Fulford, H. & J. Granell-Zafra (2004a) "The freelance translator's workstation: an empirical investigation". *Proc. 9th EAMT Workshop*. 53-61

Fulford, H. & J. Granell-Zafra (2004b) "The uptake of online tools and web-based language resources by freelance translators: implications for translator training, professional development, and research". *Proc. 2nd International Workshop on Language Resources for Translation Work, Research and Training*. 37-44

Gachot, J. (1989) "Machine translation in everyday use". *Proc. MT Summit II*. 72-73

Gaspari, F. (2004) "Integrating on-line MT services into monolingual web-sites for dissemination purposes: an evaluation perspective". *Proc. 9th EAMT Workshop*. 62-72

Gaspari, F. (2006) "The added value of free online MT services: confidence boosters for linguistically-challenged Internet users, a case study for the language pair Italian-English". *Proc. 7th AMTA Conference*. 46-55

Harrison, C. (1992) "CompuServe investigates machine translation". *MT News International*, issue 3, September 1992. 11

Henisz-Dostert, B. (1979) "Users' evaluation of machine translation". Henisz-Dostert, B., R.R. Macdonald & M. Zarechnak (eds.), *Machine translation*. The Hague: Mouton. 147-244

Kübler, N. (2002) "Teaching commercial MT to translators: bridging the gap between human and machine". *Proc. 6th EAMT Workshop*. 155-162

Macklovitch, E. (1997) "MT R&D in Canada". Teller, V. & B. Sundheim (eds.), *Proc. MT Summit IV*. 200-210

Macklovitch, E. (2001) "Recent trends in translation technology". *Proc. 2nd International Conference The Translation Industry Today*. 23-47

McCarthy, B. (2004) "Does online machine translation spell the end of take-home translation assignments?" *CALL-EJ Online* **6**(1). Available at http://www.tell.is.ritsumei.ac.jp/callejonline/journal/6-1/mccarthy.html [accessed 5 July 2007]

McLaughlin, S. & U. Schwall (1998) "Spicing up the information soup: machine translation and the Internet". Farwell, D., L. Gerber & E. Hovy (eds.), *Proc. 3rd AMTA Conference*. Berlin: Springer. 384-397

MT News International. Newsletter of the IAMT (1999) **8**(1), issue no. 22, July 1999

O'Neill-Brown, P. (1996) "Online machine translation". *Proc. 2nd AMTA Conference*. 222-223

Ryan, J.P. (1987) "Systran: a machine translation system to meet user needs". *Proc. MT Summit*. 99-103

Smith, R. (2001a) "Machine translation: potential for progress". *English Today* **17**(4): 38-44

Smith, R. (2001b) "Using information technology to optimise translation processes at PricewaterhouseCoopers Madrid". *Proc. MT Summit VIII*. 341-344

Smith, R. (2003) "Overview of PwC/Systranet on-line MT facility". *Proc. 25th Translating and the Computer*

Somers, H. (2003) "Machine translation: latest developments". Mitkov, R. (ed.), *Oxford Handbook of Computational Linguistics*. Oxford: OUP. 512-528

Somers, H. (2005) "Round-trip translation: what is it good for?" *Proc. Australasian Language Technology Workshop 2005*. 127-133. Available at http://www.informatics.manchester.ac.uk/~harold/RoundTrip.doc [accessed 5 July 2007]

Somers, H., F. Gaspari & A. Niño (2006) "Detecting inappropriate use of free online machine translation by language students – a special case of plagiarism detection". *Proc. 11th EAMT Conference*. 41-48

Watters, A. & M. Patel (1999) "Semantic processing performance of Internet machine translation systems". *Internet Research: Electronic Networking Applications and Policy* **9**(2): 153-160

Westfall, E. (1996) "Legal implications of MT on-line". *Proc. 2nd AMTA Conference*. 231-232

Yamada, S., S. Kodama, T. Matsuoka, H. Araki, Y. Murakami, O. Takano & Y. Sakamoto (2005) "A Report on the Machine Translation Market in Japan". *Proc. MT Summit X*. 55-62

Yang, J. & E.D. Lange (1998) "SYSTRAN on AltaVista: a user study on real-time machine translation on the Internet". Farwell, D., L. Gerber & E. Hovy (eds.), *Proc. 3rd AMTA Conference*. Berlin: Springer. 275-285

Yang, J. & E. Lange (2003) "Going live on the Internet". Somers, H. (ed.), *Computers and translation: a translator's guide*. Amsterdam: John Benjamins. 191-210

Zervaki, T. (2002) "Online free translation services". *Proc. 24th Translating and the Computer*

Appendix

AltaVista Babel Fish (from December 1997 until early 2006)	
Date of launch	9 December 1997
Primary URL	`http://babelfish.altavista.digital.com`
Initial language pairs	5 bidirectional language pairs (i.e. 10 separate combinations) including major European languages, i.e. EN<>FR, EN<>DE, EN<>ES, EN<>IT, EN<>PT
Most popular language pairs	In decreasing order of frequency: EN>ES, EN>FR, DE>EN, ES>EN, EN>DE, FR>DE, EN>IT, EN>PT, IT>EN, PT>EN (Nov. 1999)
Volume of translations	500,000 per day (May 1998); 740,218 (10 Nov. 1999); 1.3 million per day (Oct. 2000)
Types and details of translations	42.3% webpages vs. 57.6% plain text (June 1998); 17.6% webpages vs. 82.4% plain text (Nov. 1999); 50%+ of translations are one- or two-word phrases (May 1998); average length of texts submitted approx. 20 words (Nov. 1999)
User feedback	Between January and May 1998 more than 5,000 emails sent by users with linguistic comments or feedback on translations – estimated 95% positive

Table 1: Information on the use of AltaVista Babel Fish extracted from Yang & Lange (1998) and Yang & Lange (2003)

Yahoo! Babel Fish (since April 2006)	
Date of launch	27 April 2006 (effectively a re-launch under the Yahoo! brand)
Primary URL	`http://babelfish.yahoo.com`
Initial language pairs	38 language pair-directions, including EN<>KO, DU<>FR, GR<>EN, etc.
Most popular language pairs	EN<>ES (main/US sites); EN<>FR (UK site); in every non-English-speaking region, the most popular pair is always the local native language<>EN
Types and details of translations	- 2% URLs vs. 98% text for requests submitted directly via the Yahoo! Babel Fish portal (not including requests from other sources, such as search results or the toolbar) - The launch of Yahoo! Babel Fish in April 2006 has not affected the volume of traffic on AltaVista Babel Fish (which is still in existence and operating as usual) - The language pair traditional Chinese<>simplified Chinese was developed in-house by Yahoo! Babel Fish and is the only one which is not also on AltaVista Babel Fish

Table 2: Information on the use of Yahoo! Babel Fish as of April 2007

FreeTranslation	
Date of launch	21 June 1999 (official announcement in press release), but the service had been running in "stealth mode" for several weeks before this official launch
Primary URL	`http://www.freetranslation.com`
Most popular language pairs	Based on a sample of 1 million translation requests to a single server: EN>ES (34.83% of total); ES>EN (22.17%); EN>FR (11.85%); FR>EN (7.31%); DE>EN (5.09%); EN>DE (4.91%)
Volume of translations	3,500 on first day of stealth operation; 50,000 per day (Dec. 1999); 100,000 per day (Dec. 2000); 1 million per day (late 2002); nearly 3 million per day (Jan. 2006); 3.4 million per day, corresponding to 50 million source words (Sep. 2006)
Types and details of translations	- More than 90% of the requests are for plain text translations; average number of source words per translation request is 15 - Average of 3 translations per minute done in 6-week beta period (mid-1999) - Peak usage is usually between 8:00 and 10:00 PM Eastern Standard Time, with up to 4,000 translation requests per minute received during this time

Table 3: Information on the use of FreeTranslation as of September 2006[8]

Services powered by Systran	
Most popular language pairs	EN<>FR; EN<>ES; EN<>DE; RU>EN; EN<>ZH; EN<>JA; more than 40 language pairs in total
Volume of translations	More than 30 million pages translated per day on all services powered by Systran
Types and details of translations	On average plain-text translations are becoming increasingly shorter, and it is very common for people to use the service to look up and translate single words

Table 4: Information on the use of Systran (all services) as of June 2007

---

[8] Updates about this service are regularly posted by its developers at `http://blog.freetranslation.com` [accessed 5 July 2007].