

# **Making a sow's ear out of a silk purse: (mis)using online MT services as bilingual dictionaries**

**Federico Gaspari**

Discipline della Mediazione Linguistica  
Facoltà di Lettere e Filosofia  
Università degli Studi di Macerata  
62100 Macerata  
Italy

F.Gaspari@unimc.it

**Harold Somers**

School of Languages, Linguistics and Cultures  
University of Manchester  
Oxford Road  
Manchester M13 9PL  
United Kingdom

Harold.Somers@manchester.ac.uk

## **Abstract**

This paper concerns the misuse of online machine translation (MT) systems for lexical look-up, as if they were bilingual dictionaries. Following a review of the literature on online dictionaries, the paper reports (part of) a survey carried out among 104 university students in the United Kingdom investigating their usage of free online MT services. This paper focuses in particular on the widespread use of these MT tools for a purpose that they were not designed for, i.e. the translation of single lexical items. The 104 respondents (from an original survey of 280) had used web-based MT services in the past, and 65 of these (62.5%) reported using them for single-word lookup. This finding suggests that designers and developers of online MT services should seriously consider taking a proactive approach by treating single-word translation requests as dictionary look-up rather than translation, and/or raising the awareness of users with regard to the most (in)appropriate ways of using web-based MT software. The paper argues that it would be in the interests of those who have a stake in offering and promoting MT in the online environment (e.g. system designers, developers, and ultimately the MT vendors themselves) to manage the expectations of naïve users.

## **1. Introduction and background**

### ***1.1 Machine translation and lexicography***

The two areas of machine translation (MT) and lexicography have historically been closely linked (Steffens, 1995; Wanner, 1996), and lexicography has traditionally played a key role in the development of successful operational MT systems with the provision of machine-readable lexical components (Meekhof & Clements, 2000; Gdaniec & Manandise, 2002; Koeling et al., 2003; Zajac et al., 2003). Producing lexicographic resources and rules for MT systems is a laborious and expensive process (Kilgarriff & Tugwell, 2001: 187), and a number of approaches have been proposed in MT-related research and development to overcome the challenges posed by what Farwell et al. (1992: 532) among others have called the "lexical acquisition bottleneck". These consist, for example, in reusing

already existing lexical resources to design new MT systems (Bond et al., 2001) or in fine-tuning customisable software to specific domains by adding the relevant terminology to augment or refine coverage (Ayan et al., 2003). In particular, cost-effective approaches that rely on repurposing available lexical components have been successfully applied to the development of MT systems covering low-density languages, as reported in Diaz de Ilarraza et al. (2000), Weerasinghe (2002) and Karagol-Ayan et al. (2003).

In recent years the Internet has spurred and enabled the appearance of a host of tools and resources that rely on intensive lexicographic work, including mono- and multilingual online dictionaries and lexical look-up facilities, browsable term banks as well as web-based MT systems. At the same time, over the last decade or so, the advent of the Internet has had a considerable impact on the ways in which translation software on the one hand and lexicographic resources on the other are offered to, and deployed by, the public. In addition, the growth of the World Wide Web has also reshaped to some extent the relationship between MT and lexicography (Dillinger, 2001: 85): the migration of already existing services, resources and tools onto the new Internet platform has taken place with varying degrees of adaptation and consideration of the requirements that are germane to online interaction, giving rise to unprecedented and multi-faceted usage patterns, some of which may not have been fully anticipated and accounted for by the designers of the online applications concerned.

Despite the availability of online dictionaries, it is apparent that a substantial number of people employ online MT systems as dictionary-like look-up facilities, in an attempt to find out or check the semantic bilingual equivalence of individual isolated lexical units out of context. This trend represents one of the latest, and arguably one of the most surprising, developments in the long-standing collaboration between the two traditionally related areas of MT and lexicography. Against this background, with the help of data gathered from a survey among 104 university students based in the UK who are MT users, the rest of this paper addresses the issues raised by this distorted use of web-based MT services.

## ***1.2 Research into online bilingual dictionaries***

The presence of dictionaries of various kinds on the Internet has attracted significant scholarly attention over the last few years, and a number of studies have approached this topic from a variety of angles and with different levels of detail, mainly looking at the potential of these online resources and the advantages that they offer compared with traditional paper-based or offline lexicographic works (e.g. Docherty, 2000; Harley, 2000; Geisler, 2002; Menagarishvili & Coggin, 2003; Li, 2005). In an early study, Storrer & Freese (1996) looked at dictionaries available on the Internet combining qualitative and typological considerations, to present a wide-ranging overview of this emerging field which links it with the long-established tradition of printed dictionaries. They identified and analysed in detail 216 online dictionaries offered on the Internet in mid 1996. Their survey showed that on the whole bi- and multilingual dictionaries were much more widely available than monolingual ones, and that overall there was a higher number of general dictionaries than specialised ones. In particular, they emphasised that

the vast majority of online bilingual dictionaries involving German were for the language pair English-German.

Carr (1997) is another early example of the interest in the new possibilities offered by the Internet as a medium to provide and consult dictionaries with a clear emphasis on lexicographic issues. This paper is a forward-looking attempt to raise lexicographers' awareness of the opportunities and challenges presented by the Internet, and also provides an overview of online sources of mono- and bilingual lexicographic information. Nesi (1998) was particularly concerned with mono- and multilingual electronic dictionaries for learners of English, and argued that the role played by such works becoming available on the Internet could be expected to grow, in spite of the pressures exerted by market forces to slow down this process. Interestingly for the study presented in the rest of this paper, Nesi's conclusion (1998: ev) argued that "[a]s we receive more and more of our daily supply of information on-line, it will become an increasingly obvious alternative to access word knowledge on-line too, using whatever search routes suit our needs". This prediction seems to be substantiated by the strong inclination of the students involved in the survey presented here to use online MT services when looking up the meaning of individual words.

De Schryver (2003a) discusses the benefits offered to dictionary making by present-day technological advancements, emphasising the advantages of mono-, bi- and multilingual electronic dictionaries over traditional paper-based ones, and devotes one section specifically to the reasons for the appeal of online dictionaries (*ibid.*: 157-160). De Schryver (2003b) is an overview of online dictionaries available for nearly 120 languages grouped under the umbrella definition of "African languages", revealing that 182 of them are bilingual or multilingual, and that only one is a monolingual dictionary (*ibid.*: 9). De Schryver & Joffe (2004) discuss the issues involved in the design of a freely-available online bilingual dictionary for the language pair English and Sesotho sa Leboa; they emphasise the benefits of providing feedback forms and of implementing an unobtrusive log file analysis system to reveal usage patterns and difficulties experienced by users, so that lexicographers can improve the compilation of the dictionary, as needed. In a similar vein, Kiatisevi et al. (2003) present an ongoing project geared towards the creation of a multilingual online dictionary service for Thai in combination with a range of other languages, discussing issues related to the implementation of the system on the Internet and how users' queries can be fed back into the development process to improve the overall performance of the service.

In spite of the interest of these studies, however, it is not clear according to what criteria online lexical resources are classified and considered as dictionaries by the authors. The literature is surprisingly silent on the factors that contribute to drawing the line between "proper" online dictionaries on the one hand and "amateur" websites which provide some sort of lexical information but cannot be considered dictionaries as such on the other (e.g. word-lists, glossaries, etc.). This distinction is far from obvious, and is particularly crucial for less-resourced languages, especially in the case of online lexical resources that are not associated with well-established publishers or providers of printed dictionaries.

As noted in Somers (1997), for example, there are several webpages and websites claiming to give access to “online dictionaries” for a number of minority languages, which however turn out to consist of a few hundred entries, and appear to be prepared by keen amateurs without professional training in lexicography. Many of these online resources are of questionable usefulness, and can hardly be considered dictionaries in their own right.

Finally, Wooldridge (2004) deals with lexicography in the era of the World Wide Web within a general overview of computer-assisted lexicography, including the latest developments in online dictionaries. Summing up the discussion, the author argues that the “most radical effect that the computer has had on lexicography [...] has been to supplement the limited number of paths for information retrieval determined in advance by author and publisher with the infinite number of paths chosen by the dictionary-user” (ibid.: ev). This conclusion draws attention to the key issues flagged up by our survey, regarding the questionable use of web-based MT services as dictionary-like providers of bilingual lexicographic information and semantic equivalence for single lexical units taken out of context.

### ***1.3 Relevant usage reports for Babel Fish***

Part of the motivation to conduct the survey reported here and for the strong focus of this study in particular on the translation of single words by means of web-based MT services came from the information in Yang & Lange (2003).<sup>1</sup> They reveal some data on the usage of Babel Fish for two randomly chosen census days in June 1998 and November 1999, which is of special interest here because Babel Fish is arguably the most well-known online MT system. Yang & Lange (2003) provide information focusing on the users’ general demographic profiles, their reactions to the quality of the service based on feedback sent to the developers, the most frequent language combinations that are requested, and the kind of input that gets translated, i.e. plain text vs. entire webpages, which is constantly in favour of the former option.

In particular, they explain that “more than 50% of translations are of one- or two-word phrases” (Yang & Lange, 2003: 199), which leads the developers of Babel Fish to consider “adding a button to the web page [of the service] to distinguish dictionary look-up from translation. One-word translation requests would be treated as dictionary look-up, and a list of alternatives, perhaps with glosses, will be returned” (ibid.). At the time of writing of this paper, i.e. roughly four years after the information about Babel Fish contained in Yang & Lange (2003) was published, no look-up facility separate from the translation option for blocks of running text or entire webpages is as yet offered by Babel Fish; and, if a one-word “text” is offered for translation, only one (supposedly) equivalent lexical item in the target language is provided, as indeed is the case with the other major web-based MT systems (e.g. FreeTranslation and Google Language Tools).

---

<sup>1</sup> This is an extended and updated version of a paper that was originally published five years earlier (Yang & Lange, 1998).

It seems reasonable to imagine that the unorthodox use of online MT technology as a dictionary-like look-up tool to check the meaning and translation of individual lexical items reported in Yang & Lange (2003) is also found among Babel Fish's competitors. Some information on this specific aspect is presented in a survey carried out by Gaspari & Hutchins (2007) to investigate current use of online MT, eliciting usage data from the providers of free web-based translation services. A tendency for users to translate very short phrases or look up individual words is in fact evident from recent data regarding other Internet-based MT tools as well (ibid.: 203). The research undertaken for this paper looked at this issue directly in more detail involving end users of a range of Internet-based MT systems, in order to understand more accurately how widespread this dictionary-like use of online MT tools actually is. After presenting the results of this study, the paper discusses the challenges that the Internet poses in terms of unanticipated usage of web-based MT technology with a bearing on online dictionary use and lexicography.

## **2. The survey**

### ***2.1 Design and distribution of the questionnaires***

This survey was conducted with students at three British Universities, namely Manchester, Salford and Liverpool Hope. The questionnaires were circulated among undergraduate and postgraduate students in the spring of 2005 at different stages and under slightly different settings. The overall survey covered several areas, as it was designed to test and evaluate users' reactions to the perceived quality of a range of online MT services, asking the participants to perform a number of different tasks (the results of these separate experiments are reported in Gaspari, 2005 and Gaspari, 2006). The parts relevant to the information presented here were all identical, and this section provides more detail on the design and distribution of the questionnaires that were employed to collect the data included in this study, as well as on the profile of the respondents.

The questionnaires were administered on paper and contained 50 questions. Apart from collecting demographic data and assessing the respondents' skills and experience in Internet browsing, the focus was also on whether they had ever used any web-based MT system before. If this turned out to be the case, the participants were asked to provide specific information about their usage (i.e. language combinations required, type and length of the texts they translated, such as technical documentation or personal letters, etc.). The data presented here concerns in particular the sections of the questionnaires with a direct impact on dictionary use and lexicography, referring specifically to the scenario in which online MT services are employed as surrogate bilingual dictionaries, with a view to discussing the relevant implications.

The majority of the questionnaires were circulated during lectures and tutorials to undergraduate and taught postgraduate students, whilst supervisors helped with the distribution to their postgraduate research students, and a number of additional copies were also made available through pigeon holes or departmental common rooms. Whilst this approach to the distribution of the questionnaires entailed

a self-selection effect that might have affected the reliability and integrity of the data (e.g. the students who had already used online MT services and had possibly found them useful might have been more keen to participate in the survey than others), it was felt that this was the best way to ensure a wide coverage and to maximise the number of completed questionnaires that would be returned.

## ***2.2 Profile of the respondents***

The student population was deliberately targeted for the survey because, regardless of their discipline and level of study, they have to use the Internet on a regular basis as part of their courses, to carry out research, access official university information and courseware prepared by tutors, etc. In addition, most students also use the Internet for personal or social purposes (reading newspapers, online purchases, e-banking, etc.), and easy access to state-of-the-art IT facilities is provided on campus by their universities, which by and large meant that all students had at least reasonable Internet browsing skills, and could be assumed to represent fairly accurately a segment of regular and relatively experienced Internet users. This was essential to make sure that the students would be able to carry out some of the tasks on the Internet focusing on web-based MT systems that formed part of the tests and evaluation experiments included in the questionnaires.

In total 280 questionnaires were filled in and returned by students. The design of the questionnaire enabled the researchers to ascertain that out of the total sample of 280 respondents initially involved in the survey, 104 of them (37.1%) had used online MT services at some point in the past before taking part in the survey. As a result, the rest of the paper discusses exclusively the usage data provided by the 104 questionnaires containing validated answers that were definitely about the previous use of online MT software on the part of the respondents, and this section gives demographic information only on these 104 students, who provided the data on which this study is based.

The majority of the respondents (65 of them)<sup>2</sup> considered here were attending undergraduate courses at the time of the survey, whereas the remaining 39 were doing taught or research-based postgraduate programmes leading to masters and doctoral qualifications. The same ratio was found in terms of gender, with a prevalence of female individuals (65). The mean age of the participants was 23.81 years (median 23.00), the youngest being 18 years of age and the oldest, a postgraduate student, 66; 99 of the respondents were under 30 years of age.

The 104 participants involved in this study displayed a very diverse range of language backgrounds, because the sample included a proportion of international and visiting students (e.g. Erasmus). Not surprisingly, however, English was the single most common native language with 42 native speakers, and all the participants were fluent in English by virtue of being students at British universities. In addition, 80 of them were students of foreign languages at the university (given the links of the authors

---

<sup>2</sup> Since the group of respondents considered here consists of 104 individuals, for the sake of brevity we refrain from giving both raw totals and percentages which would always be similar figures.

with language departments at the three institutions that were involved, the distribution of the questionnaires took place during their own classes and those of their colleagues).

The questionnaires also elicited information regarding the level of online experience and the typical Internet use of the respondents. As expected, the frequency with which the sample population accessed the Internet in a typical week turned out to be very high on average: 70 respondents reported using the Internet on a daily basis, 27 individuals used it regularly five or six days a week, while the remaining 7 accessed the Internet three or four days in a typical week.

### 3. Results

#### 3.1 Online MT services used in the past

The online MT system most often used by the respondents prior to taking part in the research was Babel Fish, which was explicitly mentioned alone by 41 of the 104 respondents. The next most popular online MT service was Google Language Tools, which was indicated alone by 21 users, while the third most popular translation tool was FreeTranslation, which had been used on its own by 12 people. In addition, all these three online MT services had also been used in different combinations alongside other web-based translation tools (including less popular ones) by some of the respondents. Table 1 shows how many participants had used one online MT service or a variety of them (in combinations of up to four, which was the case for only one participant).

<b><i>Combinations of online MT services used in the past</i></b>	<b><i>Number</i></b>
<i>Only one</i>	84
<i>Two</i>	15
<i>Three</i>	4
<i>Four</i>	1
<b>Total</b>	<b>104</b>

*Table 1. Combinations of online MT services*

Table 2, on the other hand, shows all the online MT services that the participants reported using before taking part in the survey.<sup>3</sup>

---

<sup>3</sup> Each of the online MT services mentioned by the participants has been counted separately, so the total count exceeds the number of respondents because 20 of them had used more than one system, as shown by Table 1.

<b>Online MT services used in the past</b>	<b>Number</b>
<i>Babel Fish</i>	55
<i>Google Language Tools</i>	33
<i>FreeTranslation</i>	20
<i>Systran Online</i>	6
<i>Yahoo!</i>	4
<i>Worldlingo</i>	3
<i>Reverso</i>	2
<i>Voila</i>	2
<i>www.freetranslator.com</i>	2
<i>ProMT</i>	1
<i>www.1-800-translate.com</i>	1
<i>www.appliedlanguage.com</i>	1

Table 2. Online MT services used by the respondents

### 3.2 Usage patterns revealed by the survey

The results of this survey confirm and reinforce the findings presented in Yang & Lange (2003) and also Gaspari & Hutchins (2007), according to which users routinely interrogate Internet-based MT software with single-word translation requests. The data that we gathered shows in fact that 65 of the respondents in the sample of 104 people had used online translation tools in the past to translate single words. Table 3 shows in more detail the Internet-based MT tools which the 65 respondents had used to translate single words.

<b>Online MT services used to translate single words</b>	<b>Number</b>
<i>Babel Fish</i>	34
<i>FreeTranslation</i>	17
<i>Google Language Tools</i>	17
<i>Yahoo!</i>	4
<i>Reverso</i>	2
<i>Systran Online</i>	2
<i>Voila</i>	2
<i>Worldlingo</i>	2
<i>www.freetranslator.com</i>	2
<i>www.1-800-translate.com</i>	1

Table 3. Online MT services used to translate single words

It should be pointed out that in the majority of cases this mode of use was also mentioned alongside others involving longer passages of text. In fact, only four respondents indicated that they had used online MT services exclusively to translate “single words” or “individual terms”, whilst in all the other cases this specific usage was mentioned explicitly alongside others as well (the translation of “sentences”, “paragraphs”, “articles”, “letters”, etc.).



## 4. Discussion and conclusion

### 4.1 (Mis)using online MT services as bilingual dictionaries

These findings seem to reveal a lack of understanding on the part of the users regarding the kind of service that web-based MT software is designed to provide, and also, perhaps, about language and translation in general: non-linguists often assume that there is a one-to-one equivalence of vocabulary items between languages. The large number of online bilingual dictionaries that are available on the Internet covering a wide range of language combinations, some of which can be accessed free of charge, would seem the most suitable tools to find out the meaning or translation of single words and isolated lexical units. As a matter of fact, most online dictionaries provide users with detailed lexicographic information and semantic insights that are meant to guide them towards a thorough understanding of the meaning and correct use of the lexical items by offering definitions, glosses, explanations, part-of-speech information, possible target-language equivalents based on sense disambiguation, examples of typical in-context occurrences, unmarked collocational patterns, frequent metaphorical uses, idiomatic expressions, etc. All these features derive from long experience in dictionary making and development, and high-quality online bilingual dictionaries disseminate over the new Internet medium essentially the same kind of lexicographic information that traditional high-quality paper-based dictionaries have been supplying to their users (language learners, translators, technical writers, etc.) for a long time.

On the other hand, in spite of relying on a significant amount of lexicographic work and lexical processing to operate, typical online MT services (like all those listed in Table 3) are designed to hide lexicographic processing and information from users, and simply to provide target-language output for any input, so much so that for any single word in the source language they will provide a supposed equivalent in the target language strictly on a one-to-one basis. In order to interpret the data provided by the survey and discuss the potential problems that it reveals, one can differentiate scenarios in which this modus operandi might not affect users badly after all, as would for instance be the case when they are simply checking or reminding themselves of the meaning and translation of a word between two languages with which they are already familiar to some extent. If the users have at least a basic knowledge of both the source and target languages concerned, then a straightforward one-to-one correspondence for a particular lexical item between the two languages might serve the purpose of backing up the user's intuition or assumption about its possible meaning(s) and translation(s), because they should be able to vet the system's suggestion.<sup>4</sup>

However, the straightforward equivalence provided by online MT services when individual lexical items without context are given as input in effect hides the range of context-dependent senses (and therefore different possible translations into the target language). At the same time, web-based MT systems cannot provide accurate or reliable information for such cases as homography or polysemy, for which a

---

<sup>4</sup> Nevertheless, in this respect it should be noted that recent studies investigating from a variety of perspectives the habits of advanced learners, language students and trainee translators using both paper and online dictionaries have lamented serious shortcomings in their lookup strategies, occasionally resulting in severe consultation problems (e.g. Nesi & Hail, 2002; Sánchez Ramos, 2005; Frankenberg-García, 2005).

proper online bilingual dictionary would be able to account much more rigorously, for example thanks to cross-references to related entries. Pérez-Ortiz & Forcada (2001) present a laboratory assignment expanding on these issues and designed to raise students' awareness of how commercial MT software actually works. Taking as an example the language pair English-Spanish, they use a simple technique to demonstrate that real MT programs are not designed to translate on a word-for-word basis, but that context-dependent target-language lexical choices producing good-quality translations are selected depending on the analysis of larger units such as phrases found in source-language sentences.

In summary, it seems reasonable to suggest that in some cases for users familiar with both languages involved in the translation process using online MT services for simple lexical look-up may be a valid technique, because their bilingual knowledge would enable them to be in control of the process and vet the acceptability and correctness of the output provided by the system for an individual lexical item. On the other hand, however, naïve monolingual users or those with a weak understanding of either or both of the languages involved in the translation process should be very wary of taking at face value the straightforward translation on a one-to-one basis offered by online MT systems for individual words. This is due, in particular, to the lack of extra lexicographic details that prevents users from appreciating crucial subtleties, which might give rise to misinterpretations and translation errors. Unfortunately, the design of our questionnaires did not make it possible to investigate thoroughly under what circumstances the respondents had used online MT services to translate single words (i.e. what was the level of their bilingual knowledge, if the translation direction involved translating individual words into a target language with which they were very familiar or did not know at all, etc.). As a result, it is not possible to expand on these issues in more detail here on the basis of the data gathered in the survey.

#### ***4.2 Need for further research***

Our findings reinforce and confirm the information about the usage of Babel Fish presented in Yang & Lange (2003) and the data given in Gaspari & Hutchins (2007) for other web-based MT systems, showing that in fact Internet users frequently make a questionable use of a range of online MT services. Using MT services available on the Internet as if they were bilingual dictionaries or lexical look-up facilities is a misguided strategy that is liable to provide users with partial and misleading information, in that MT software is designed to hide lexicographic information from the users and to provide one-to-one target-language equivalents. In particular, the common behaviour observed in the respondents raises the question of why Internet users would rather rely on web-based MT services to translate individual words, although bilingual dictionaries exist on the Internet and some of them, including those associated with high-quality prestigious printed dictionaries, are free of charge. Even though our survey did not investigate the reasons behind the usage patterns reported by the respondents, one possibility might be that online MT systems represent convenient "one-stop shops", in that they cover a wide variety of language pairs, particularly in combination with English as source or target language, whilst on the whole online bilingual dictionaries tend to be restricted to specific language pairs.

However, users may not realise the pitfalls associated with using web-based MT software for a purpose that it is not designed to serve, and it is difficult to assess the practical consequences of this inappropriate usage (e.g. in terms of misunderstandings, incorrect interpretations, wrong translations), given that the survey did not go into this level of detail. At any rate, the data gathered with the questionnaires shows that this is an interesting topic worthy of further research and additional experimental observation, to look deeper into the motivations for the choices made by Internet users in this area at the intersection between Internet user behaviour, MT, dictionary use and lexicography. In this respect, addressing some of the issues raised in Pérez-Ortiz & Forcada (2001) and Forcada (2002), it would be interesting to carry out some additional empirical research into the practical consequences (problems in terms of potential mistranslations, misunderstandings, misinterpretations, etc.) that occur when people with different linguistic abilities (e.g. language students at beginners', intermediate or advanced levels) translate or read a text in a foreign language using online MT simply as a look-up dictionary-like lexical resource to translate single words during the process.

For example, a set of experiments could be designed to evaluate the performance of a range of language learners working under these conditions against that of control groups that instead are allowed to use proper online and/or printed bilingual dictionaries to perform the same reading or translation tasks (an empirical study involving online MT along similar lines, although with a specific focus on post-editing activities to produce polished translations, is described in Somers et al., 2006). Niño Alonso (2006) discusses the implications and potential of using MT post-editing in foreign language teaching, with a special emphasis on the pedagogic challenges posed by the possibility of utilising raw MT output to develop the students' skills in foreign language written production. Recognising that currently it is not uncommon for language students to rely on the output provided by free online MT services for the preparation of written assignments, the overall aim of the research is to suggest how the output produced by MT software could be put to good use within the context of language teaching and learning.

It is hoped that closer investigations will address the issues raised by this survey in the future, following up on some of the suggestions put forward here. The ultimate goal of this further research work would be to increase the users' awareness as to what Internet-based tools can best serve their linguistic needs. This would ensure that their choice of the lexicographic resources, reference works and MT services that are available on the Internet is properly motivated and carefully selected, on the basis of the most appropriate tools that can be used for the tasks in which the users need linguistic support.

### ***4.3 Implications for the design of online MT services***

Finally, the issues raised here are also of direct relevance to the developers, designers and providers of online MT services. Due to the large number of requests for the translation of input consisting of one- or two-word phrases revealed by their monitoring of usage, Yang & Lange (2003) had already suggested that the default webpage through which users access Babel Fish should provide a clear distinction between the proper translation service (for passages of running text or entire webpages), and a separate dictionary look-up facility offering more detailed lexicographic information (for individual words). However, no developments in this direction have as yet been seen in Babel Fish, or indeed in any of the other popular online MT services, whilst users seem to be consolidating their habit of using them as bilingual dictionaries.

This survey has revealed that a significant number of users tend to use a range of online MT services to translate single words, and it seems likely that this pattern of usage would not stop unless the users are informed and educated about the potential and capabilities of web-based MT software, so that they are aware of its limitations and understand how best to use it. This in turn should enable Internet users to take advantage of the most appropriate linguistic resources that can meet their needs, particularly online dictionaries, lexical look-up facilities, multilingual glossaries and term banks, when they require translations of single words and isolated terms or lexical items without context. MT services could even take the more radical step of automatically diverting single-word translation requests to a suitable bilingual dictionary, or at least asking users if this is what they wanted (though that might become tiresome for a regular user). Translations of shorter phrases (two or three words) are more delicate, and the software would perform best if it had some means of detecting whether the input was effectively a single multiword lexical item or a short phrase that it would be legitimate to attempt to translate as such.

The findings of this survey suggest very strongly that the designers and developers of online MT services should seriously consider taking a proactive approach towards raising the awareness of the users with regard to the most appropriate (or, otherwise, deprecated) ways of using web-based MT software. After all, it would seem to be in the interests of those who have a stake in offering and promoting web-based MT technology (e.g. system designers, developers, and ultimately the MT vendors themselves) to manage the expectations of naïve users, informing them of the types of usage with which MT software is not capable of dealing well and for which it can provide, at best, only a service of questionable quality.

### **Acknowledgements**

The authors would like to thank the members of staff at the Universities of Manchester, Salford and at Liverpool Hope University for their assistance in distributing the questionnaires to their students. Special thanks also to all the students who took the time to fill in the questionnaires on which this study was based.

## References

- Ayan, N.F., B.J. Dorr & O. Kolak (2003) "Evaluation Techniques Applied to Domain Tuning of MT Lexicons". In *Proceedings of the Workshop "Towards Systematizing MT Evaluation"*, New Orleans, LA. 3-11
- Bond, F., R.B. Sulong, T. Yamazaki & K. Ogura (2001) "Design and Construction of a Machine-translatable Japanese-Malay Dictionary". In *Proceedings of MT Summit VIII*, Santiago de Compostela, Spain. 53-58
- Carr, M. (1997) "Internet Dictionaries and Lexicography". *International Journal of Lexicography* **10**(3): 209-230
- De Schryver, G.-M. (2003a) "Lexicographers' Dreams in the Electronic-dictionary Age". *International Journal of Lexicography* **16**(2): 143-199
- De Schryver, G.-M. (2003b) "Online Dictionaries on the Internet: an Overview for the African Languages". *Lexicos* 13: 1-20
- De Schryver, G.-M. & D. Joffe (2004) "On How Electronic Dictionaries Are Really Used". In *Proceedings of the 11th Euralex International Congress*, Lorient, France. 187-196
- Diaz de Ilarraza, A., A. Mayor & K. Sarasola (2000) "Building a Lexicon for an English-Basque Machine Translation System from Heterogeneous Wide-coverage Dictionaries". In *Proceedings of the International Conference on MT and Multilingual NLP: MT 2000*, Exeter, UK. 2-1 – 2-9
- Dillinger, M. (2001) "Dictionary Development Workflow for MT: Design and Management". In *Proceedings of MT Summit VIII*, Santiago de Compostela, Spain. 83-87
- Docherty, V.J. (2000) "Dictionaries on the Internet: an Overview". In *Proceedings of the 9th Euralex International Congress*, Stuttgart, Germany. 67-74
- Farwell, D., L. Guthrie & Y. Wilks (1992) "The Automatic Creation of Lexical Entries for a Multilingual MT System". In *Proceedings of the fifteenth [sic] International Conference on Computational Linguistics COLING-92*, Nantes, France. 532-538
- Forcada, M.L. (2002) "Explaining Real MT to Translators: between Compositional Semantics and Word-for-word". In *Proceedings of the 6th EAMT Workshop "Teaching Machine Translation"*, Manchester, UK. 137-143
- Frankenberg-Garcia, A. (2005) "A Peek into What Today's Language Learners as Researchers Actually Do". *International Journal of Lexicography* **18**(3): 335-355
- Gaspari, F. (2005) "Embedding Free Online Machine Translation into Monolingual Websites for Multilingual Dissemination: a Case Study of Implementation". In *Proceedings of Translating and the Computer 27*, London, UK. [pages not numbered]
- 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". In

*Proceedings of AMTA 2006, 7th Conference of the Association for Machine Translation in the Americas*, Cambridge, Mass. 46-55

- Gaspari, F. & J. Hutchins (2007) "Online and Free! Ten Years of Online Machine Translation: Origins, Developments, Current Use and Future Prospects". In *Proceedings of MT Summit XI*, Copenhagen, Denmark. 199-206
- Gdaniec, C. & E. Manandise (2002) "Using Word Formation Rules to Extend MT Lexicons". In *Proceedings of the 5th Conference of the Association for Machine Translation in the Americas, AMTA 2002*, Tiburon, CA. 64-73
- Geisler, C. (2002) "Reversing a Swedish-English Dictionary for the Internet". In Borin, L. (ed.), *Parallel Corpora, Parallel Worlds: Selected Papers from a Symposium on Parallel and Comparable Corpora at Uppsala University*. Amsterdam: Rodopi. 123-133
- Harley, A. (2000) "Cambridge Dictionaries Online". In *Proceedings of the 9th Euralex International Congress*, Stuttgart, Germany. 85-88
- Karagol-Ayan, B., D. Doermann & B.J. Dorr (2003) "Acquisition of Bilingual MT Lexicons from OCR'd Dictionaries". In *Proceedings of MT Summit IX*, New Orleans, LA. 208-215
- Kiatisevi, P., V. Ampornaramveth, A.I. Kovács & H. Ueno (2003) "Searching in the Longdo Thai Online Dictionary Service". In *Proceedings of the 4th International Conference on Information Technologies (InTech 2003)*, Chiang Mai, Thailand. 1-6
- Kilgarriff, A. & D. Tugwell (2001) "WASP-Bench: an MT Lexicographers' Workstation Supporting State-of-the-art Lexical Disambiguation". In *Proceedings of MT Summit VIII*, Santiago de Compostela, Spain. 187-190
- Koeling, R., A. Kilgarriff, D. Tugwell & R. Evans (2003) "An Evaluation of a Lexicographer's Workbench: Building Lexicons for Machine Translation". In *Proceedings of the 7th EAMT Workshop "Improving Machine Translation through Other Language Technology Tools"*, Budapest, Hungary. 9-16
- Li, L. (2005) "The Growing Prosperity of On-line Dictionaries". *English Today* 21(3): 16-21
- Meekhof, T. & D. Clements (2000) "L&H Lexicography Toolkit for Machine Translation". In *Proceedings of the 4th Conference of the Association for Machine Translation in the Americas, AMTA 2000*, Cuernavaca, Mexico. 213-218
- Menagarishvili, O. & W.O. Coggin (2003) "Electronic Resources and Aids for Writing, Editing, and Translation". In *Proceedings of the 50th Annual Conference of the Society for Technical Communication*, Dallas, TX. 438-442
- Nesi, H. (1998) "Dictionaries on Computer: How Different Markets Have Created Different Products". In Wilson, A. & J. Schmieid (eds.), *Proceedings of the Chemnitz Symposium "Language Learning and Computers"*, Chemnitz, Germany. Available at <http://www.tu->

chemnitz.de/phil/english/chairs/linguist/real/independent/llc/Conference1998/Papers/Nesi.htm  
[accessed 15 October 2007]

- Nesi, H. & R. Haill (2002) "A Study of Dictionary Use by International Studies at a British University". *International Journal of Lexicography* **15**(4): 277-305
- Niño Alonso, A.M. (2006) *Evaluating the Suitability of Using Raw Machine Translation Output as Input for Foreign Language Written Production*. PhD Thesis, School of Education, University of Manchester, Manchester, United Kingdom
- Pérez-Ortiz, J.A. & M.L. Forcada (2001) "Discovering Machine Translation Strategies Beyond Word-for-word Translation: a Laboratory Assignment". In *Proceedings of MT Summit VIII Workshop on Teaching Machine Translation*, Santiago de Compostela. 57-60
- Sánchez Ramos, M.d.M. (2005) "Research on Dictionary Use by Trainee Translators". *Translation Journal* **9**(2), April 2005. Available at <http://accurapid.com/journal/32dictuse.htm> [accessed 15 October 2007]
- Somers, H.L. (1997) "Machine Translation and Minority Languages". In *Translating and the Computer 19: Papers from the Aslib Conference*, London, UK. 1-13
- Somers, H., F. Gaspari & A. Niño (2006) "Detecting Inappropriate Use of Free Online Machine Translation by Language Students – a Special Case of Plagiarism Detection". In *Proceedings of the 11th Annual Conference of the European Association for Machine Translation*, Oslo, Norway. 41-48
- Steffens, P. (1995) (ed.) *Machine Translation and the Lexicon*. Berlin: Springer
- Storrer, A. & K. Freese (1996) "Wörterbücher im Internet". *Deutsche Sprache* **24**(2): 97-153
- Wanner, L. (1996) "Lexical Choice in Text Generation and Machine Translation". *Machine Translation* **11**(1-3): 3-35
- Weerasinghe, R. (2002) "Bootstrapping the Lexicon Building Process for Machine Translation between 'New' Languages". In *Proceedings of the 5th Conference of the Association for Machine Translation in the Americas, AMTA 2002*, Tiburon, CA. 177-186
- Wooldridge, R. (2004) "Lexicography". In Schreibman, S., R. Siemens & J. Unsworth (eds.), *A Companion to Digital Humanities*. Oxford: Blackwell. 69-78. Available at [http://www.chass.utoronto.ca/~wulftric/articles/lao/lexico\\_eng.htm](http://www.chass.utoronto.ca/~wulftric/articles/lao/lexico_eng.htm) [accessed 15 October 2007]
- Yang, J. & E. Lange (1998) "SYSTRAN on AltaVista: a User Study on Real-time Machine Translation on the Internet". In *Proceedings of the 3rd Conference of the Association for Machine Translation in the Americas, AMTA-98*, Langhorne, PA. 275-285
- Yang, J. & E. Lange (2003) "Going Live on the Internet". In Somers, H. (ed.), *Computers and Translation: A Translator's Guide*. Amsterdam: John Benjamins. 191-210
- Zajac, R., E. Lange & J. Yang (2003) "Customizing Complex Lexical Entries for High-quality MT". In *Proceedings of MT Summit IX*, Santiago de Compostela. 433-438