The contribution of a user group to the evaluation and improvement of an MT system

A. Fourla[1], O. Yannoutsou[1],I. Tsakou[2], S. Stamou[2] and A.Petrits[3]

1   Institute for Language and Speech Processing (ILSP), EUROMAT Office, Artemidos & Epidavrou, 15125, Marousi, Greece
{soula,olga}@ilsp.gr

2   Computer Technology Institute (CTI), Research Unit II, 61 Riga Feraiou str., 262 21, Patras, Greece
{tsakou,stamou}@cti.gr

3   Commission of the European Communities, Translation Service, 200, rue de la Loi, B-1049 Brussels, Belgium
angelique.petrits@cec.eu.int

ABSTRACT

This paper describes the attitude of a Greek public administration user group towards the EC SYSTRAN machine translation (MT) system and its contribution to the system's evaluation and enhancement. EC SYSTRAN was first made available to the Greek public sector in 1994, under an initiative of the Greek Government in collaboration with the EC. Translation services are provided free of charge to the public sector and the system has unproved thanks to the evaluations performed and the terminology provided by the user group. Currently, the group has two sections: the end-users themselves and a team of linguists. The presentation concludes with some statistics on the number of users willing to assist in the enhancement of the MT system together with a description of their profile and an indication of users' reactions (positive/negative comments) to using machine translation services and language technology products.

1. INTRODUCTION

EC SYSTRAN in Greece is targeted at the public services, a user group which has many particularities and is difficult to approach. Since the system's purpose is to meet the real translation needs of officials, it has to be continuously evaluated and unproved. This paper concentrates on the formation of a user group, its contribution to the evaluation and enhancement of EC SYSTRAN, as well as the attitude of Greek public services towards machine translation.

Section 2 describes the historical background of the project and its objectives, Section 3 the campaign to promote MT in the Greek public sector and disseminate the services offered. These efforts resulted in the formation of two evaluation teams which are described in Section 4, consisting of two types of users, each contributing to the evaluation and improvement of the system in a different way. The evaluation

conducted by both teams and their contribution, as well as the contribution of the European Commission's Translation Service, is described in Section 5. Finally, Section 6 presents future goals and conclusions.

2. HISTORICAL BACKGROUND

EC SYSTRAN, that is, the version of Systran developed by the European Commission, is one of the most famous systems in the history of machine translation. However, what is not widely known is the part concerning minority languages, how they were added to the system and their impact on users. In this paper, the Greek case is presented.

In 1988, the Greek government undertook to promote the Greek language and fund initiatives that would incorporate Modem Greek in language technology tools. EC SYSTRAN was chosen on the basis that it was robust and operated with a variety of languages, while one of its major components, English analysis, was ready and well tested. As a result, the English → Greek was launched, and is still under development. This language pair was the result of co-financing between the General Secretariat of Research and Technology and the European Commission.

After a period of 6 years, in 1994, when the quality level had risen to a satisfactory level, the Greek Secretariat of Research and Technology decided to finance a Translation Technology Service Centre for the Greek public sector, which would offer free machine translation services. The language pairs offered at that time were English → Greek and 17 other language pairs. The aims were:

- ► integration of Modern Greek in a robust operational tool;

- ► modernisation of the public sector by making public servants familiar with language technology and assistance in translation tasks.

Furthermore, the European Commission wanted to make EC SYSTRAN available to the public sector of one of its member states on an experimental basis. In exchange, users would have to provide feedback in order to improve the system. This would be of great benefit for both developers and users; the former would have all the necessary feedback in terms of specialised and authorised terminology, which could not be collected otherwise, while the latter would have their requests met.

The whole idea seemed quite risky at the time, as the technological infrastructure of the public sector in Greece was quite poor, while the majority of public servants were in general ignorant of information / language technology. The timing was not bad though, as at that time the use of information technology had just started to grow in many public directorates.

There are two main points that should be made clear regarding the way the translation technology service functions:

- ► Translation services were provided for *free* and *strictly* to the Greek public sector owing to the question of intellectual property rights.

- ► The promotion strategy was low-key, since only a limited number of requests could be satisfied at the beginning. Furthermore, the main idea was to acquire

a steady number of users, who would contribute and evaluate the system on a regular basis.

As EC SYSTRAN ran on a mainframe at the time, it was decided at the pilot stage not to install the system in Greece, as this would be quite expensive for something experimental. The solution adopted was to be connected via X25 to the server in Luxembourg. The team that manned the Service Centre consisted of two linguists, one of whom specialised in public relations, plus an engineer. In the next couple of years, when the pilot phase was over and considered successful, another person joined the team for handling texts in hard copy (scanning, OCR).

Many problems were encountered, especially in the first few years, the hardest one being breaking down the misconception of what machine translation is, revealing its limits and its potentials. What made things even more complicated was ignorance of information technology in general.

Another major problem was bureaucracy in the public sector. Although the procedure was quite simple (submitting a text to the translation Service Centre), employees often faced difficulties because they would have to obtain official permission from their superior, which meant re-presenting the system to the next level of the hierarchy.

Despite the difficulties encountered in the beginning, raw machine translation services have been offered successfully to the public sector in Greece for the last six years and the system has been enhanced accordingly with feedback from experts. Due to this initial experiment, three more language pairs have been added in the course of tune: Greek → English, Greek → French and French → Greek.

The financing of the new language pairs as well as English → Greek, the provision of translation services, the promotion of the system and other related peripheral activities are covered by the EC's Multilingual Information Society (MLIS) framework and are co-financed by the General Secretariat for Research and Technology in Greece.

By the end of 2001 the system will be made available to the public sector directly, as the Commission plans to change platform from Assembler to C, meaning that EC SYSTRAN can be run on a PC.

A more detailed presentation of the interaction with users follows in the following Sections.

3. PROMOTION TACTICS

As mentioned in Section 2, one of the project's goals and anticipated results is to relieve the Greek public sector of a large part of the translation burden by developing a machine translation system adapted to its needs, the services being provided free of charge to the public authorities.

Consequently, a user group drawn from the Greek public administrations had to be formed, to which translation services would be provided. In return they would contribute to the enhancement of the MT system by providing evaluation comments and feedback to the development team. The main difficulties in the formation of such a group were: a) the lack of awareness of public officials employees regarding

language technology services and b) overcoming their preconceptions about the performance of a machine translation system.

Trying to surmount the aforementioned difficulties, the project team had to employ various promotion techniques in order to inform employees of the services provided by EC SYSTRAN and to encourage them to use it for their daily translation needs. It should be noted that informing public services and enlarging the user group is an on-going process. The most important promotion tactics are summarised below.

Some of the promotion activities take place on a daily basis, contributing to a continuous interaction with end-users, whereas others take place periodically, aiming at spreading information about the system to a wider audience and attracting employees from different public administrative fields.

Daily promotion activities involve either direct mailing to public services or contacting key persons from public administrations in order to spread information of EC SYSTRAN'S capabilities and the services provided by it. The letters sent to them include general information – such as the language pairs covered by the system, the project's historical background, and other public services using it – a brief description of the system's characteristics and performance, as well as a detailed description of the procedure that needs to be followed, and the people that should be contacted by the administration's employees in order for them to obtain translations of their documents.

In addition to the aforementioned activities, a web site is maintained, providing potential users with all available information regarding the provision of MT services to the Greek public sector.

Periodic activities are either of a general nature, i.e. aiming at a broader audience, or specific - aiming only at public employees. In the former case, presentations of the system take place during more general events, for example information days or conferences related to language, translation or terminology, where the audience is varied and only a small percentage of the target group is present. This proved to be particularly fruitful in the first few years, as translators who attend this kind of event are keener on asking for further information and promoting the use of MT.

Specific periodic activities involve the organisation of awareness meetings between members of the project team and end-users. For these events, the project team invites public officials who are already using the system as well as potential users with probable translation needs. During these meetings, the main objectives and capabilities of EC SYSTRAN are demonstrated. Furthermore, a simplified presentation of how the system works and an overall description of its potential are provided, with a view mostly to increasing the public sector's awareness of MT's potential. Users who are present at the meetings are encouraged to actively participate by expressing their opinion on how useful they find the MT system, asking questions about its performance and commenting on the quality of the translations produced.

Users' comments and other useful information are also obtained in the form of questionnaires designed by the project team, which concentrate on the improvements to MT desired by users so as to cover their translation needs.

At this point it should be noted that user response towards the above promotion activities is quite positive (more details regarding their participation in meetings are given in Section 5), and their willingness to actively participate in the evaluation of the MT system is beneficial both for the developers and the end-users themselves.

Among the promotion activities, information days are included, which take place either at ILSP and CTI or at the ministries (following an invitation). In the first case, employees from different services and ministries are invited, while in the second the presentation is focused more on the needs of the particular ministry.

The results of the awareness campaign are quite encouraging, as almost all Greek ministries have used the system at least once. The biggest difficulty faced during these activities is to maintain the interest of the target group, as many of them lose enthusiasm if they see a few poor translations.

As a result of this campaign, 700 - 1 000 pages per month are translated for various public administrations.

Encouragingly, the Greek administration is in the process of being modernised, and officials are starting to use more electronic aids to facilitate their work.

4. FORMATION OF THE USER GROUP

EC SYSTRAN is an operational system that meets actual needs, and is used by both European Union institutions and the Greek public sector. One of the project's aims was an evaluation of the English → Greek language pair from the users' point of view. In order to carry out this evaluation, a group of users from the Greek public sector was formed.

The EC SYSTRAN user group is divided into two sections acting in complement to each other. The first section, or linguistic team, consists of four people, linguists and translators. All participants come from the project teams of CTI (Computer Technology Institute) and ILSP (Institute for Language and Speech Processing). Two of them work on the development of the system by coding lemmas and rules in its dictionaries. The team's knowledge of the system varies according to experience. There are members who have been working on the system for several years, while others have only recently started working on it.

The second section consists of end-users of the translation output, public officials who have only a general knowledge of how the system works. As before, some members of this team have been using the system for the last 3 or more years, while others are new users.

The main reasoning behind the decision of forming a group consisting of both end-users and developers had to do with the evaluation process. The first group would evaluate the system knowing its capabilities and limitations, while the latter would be almost 100% concentrated on the final result, without always having in mind "why" the system makes a mistake and "whether" trying to fix it is a realistic task. Thus, they would have a more objective and uninfluenced attitude towards the outcome.

The second reason had to do with system improvement. The development team needed direct feedback from end-users for the system's dictionaries and programs, in order to make enhancements and produce high-quality results.

Different criteria were used for the recruitment of members in each section. For the linguistic team, the level of experience in developing the system, as well as the linguistic ability and studies of the candidates played an important role, while for the end-users the criteria were the frequency and volume of their translations, their linguistic ability and – most of all – their willingness to accept such an invitation.

The formation of the second section was made after an information day targeted at users of EC SYSTRAN from the public sector, during which the whole concept was explained to the participants. After this, they were asked to fill in a form which included a question on whether they wanted to join such an evaluation group and if so, how often they could meet. The rest of the questions concerned their field of speciality, volume of documents for translation, use of MT, text type and domain, as well as the type of feedback they could provide to the development team.

The group members were given a detailed presentation of the system in order to acquaint them with its features and capabilities.

Regardless of the common goal the two sections share, i.e. to evaluate EC SYSTRAN and build a powerful system based on the needs of the Greek public sector, there are two points in which they differ greatly.

The first difference has to do with the nature of evaluation that each group provides. The linguist team provides *quantitative feedback,* evaluating the system on a regular basis and indicating a percentage of improvement or deterioration. The end-users perform *qualitative evaluations,* they do not meet regularly and give feedback regarding the system quality by contacting the development team on a personal basis to discuss the problems they encountered.

The second difference refers to the fact that the first section consists of language specialists (linguists, translators, computational linguists), whereas the second is composed of a variety of public employees who use the system for their own needs. This is considered an asset for the system's evaluation, since representatives of different professions provide texts for translation from different domains and consequently, the quality of EC SYSTRAN'S output for each domain can be evaluated. Sometimes, this variety of domains and text types results in a wide range of opinions about the system on the part of the end-users. Requesters who provide administrative documents in domains which are covered by the system (and which are therefore better translated) have a totally different view of the system's performance compared to those who have, for instance, very technical documents, or manuals.

At present, the public officials' section has 15 members. It is an open group, meaning that it can be enlarged if a new user wants to join, or reduced if a user wants out. 9 out of the 15 are translators, while the rest are administrative staff from various fields such as mechanical engineering, policing, chemistry, and economics.

The number of pages that each user translates varies between 20 and more than 100 pages per month. Their domains of interest include sports, administration, culture,

geology, mineralogy, law, economy, chemistry, pharmaceuticals, medicine, agriculture, fisheries, biology, forestry, and environment. Amongst the text types they need to translate are: memos, circulars, letters, law articles, contracts, EU guidelines, manuals, studies, and reports. As a result, users provide feedback for the aforementioned domains.

The corpus gathered by EC SYSTRAN users in the Greek public sector consists of about 2 000 texts of different types and domains. The 10 main text domains, which represent almost half of the corpus, are (in order of frequency): administration, police matters, aviation, agriculture, sports, law, crime, education, EC policy issues and medicine.
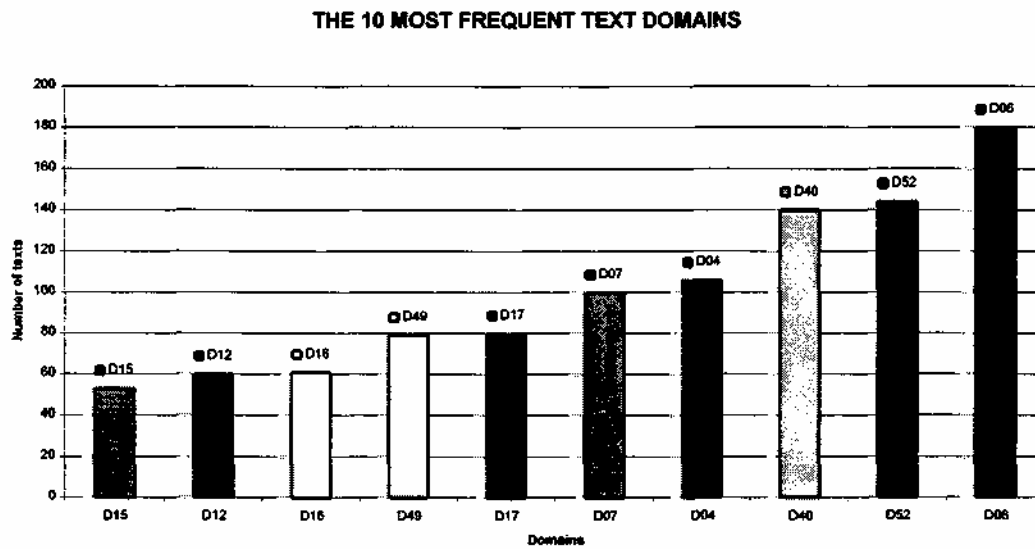

Figure 1: The 10 most frequent text domains - number of requests

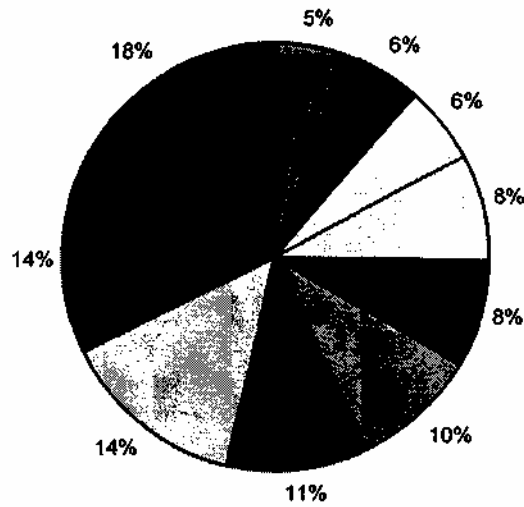**PERCENTAGES OF THE 10 MOST FREQUENT TEXT DOMAINS**

Figure 2: The 10 most frequent text domains - by percentage

In terms of text types submitted for translation, the 10 most frequent text types, which also represent almost 50% of the corpus, are (in order of frequency): reports, manuals, press articles, project guidelines, books, studies, law articles, databases, papers and leaflets.
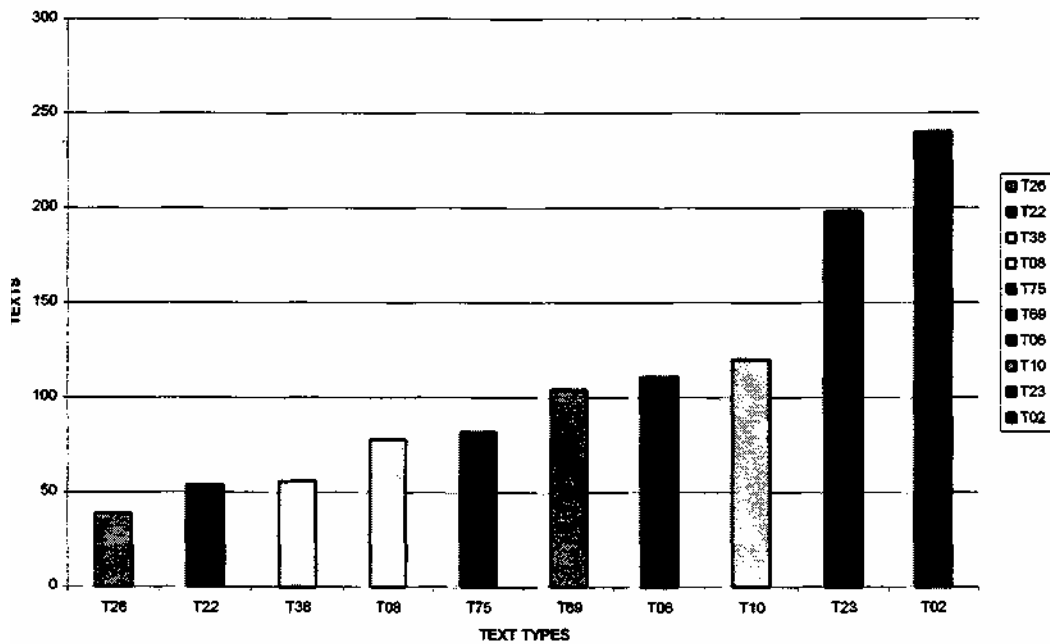
THE 10 MOST FREQUENT TEXT TYPES

Figure 3: The 10 most frequent text types - number of requests

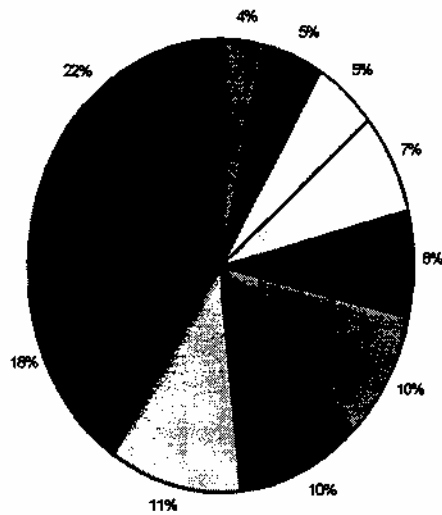**PERCENTAGES OF THE 10 MOST FREQUENT TEXT TYPES**

Figure 4: The 10 most frequent text types - by percentage

Regarding the use of the machine translation output, it is used for browsing, assistance in terminology problems, getting the gist of a document, or as a base for an official translation (with post-editing).

5. THE EVALUATION AND CONTRIBUTION OF THE USER GROUP TO THE ENHANCEMENT OF THE SYSTEM

As previously mentioned, two sections participate in the system's evaluation and contribute to its improvement: a small group of linguists / translators and a larger team of end-users from the Greek public sector. The contribution of both sections is considered vital to the enhancement of the system.

For reasons of convenience, the first section, namely the group of linguists and translators, is part of the project team and performs regular evaluations of EC SYSTRAN. This group's role involves – among other things – testing and evaluating different versions of EC SYSTRAN. In order to do so, they use a specific software (NLPWin) developed for the evaluation of EC SYSTRAN by the European Commission.

The evaluation tool is a comparison browser and interface, which reads files generated by a comparison procedure run by the development team and displays them in order to allow evaluators to comfortably read, sentence by sentence, translations produced by both old and new system versions. NLPWin also provides the evaluators with commenting and marking facilities, which allow for easy sentence extraction for development feedback. As said before, it is used for carrying out regular evaluations of all language pairs in the EC SYSTRAN system.

Using the NLPWin evaluation software, the group of linguists and translators carries out regular quarterly evaluations of the system. To be specific, their contribution consists in using the evaluation software so as to compare translations of both the most recent and previous versions of EC SYSTRAN, marking differences as better, worse or equal, according to their impact on the translation's comprehensibility, and producing statistics indicative of the system's general improvement or deterioration.

However, since this kind of evaluation is not considered adequate, they additionally comment in writing on the general translation quality and the comprehensiveness of the system's output and submit quarterly reports to the development team. In these reports, they also point out characteristic difficulties in the translations (e.g. recurrent syntactic or grammatical errors) that require further improvement and suggest terms and expressions that should be covered in the system's dictionaries. The results of such evaluations allow the developers of EC SYSTRAN to concentrate on problematic aspects of the system's output and have proven to be valuable for enhancement of the system.

As mentioned in Section 4 of this paper, the second section, which is much larger than the linguist team, consists of end-users from Greek public services. Their main contribution to the enhancement of the system lies in evaluation, providing direct feedback and helping developers (by means of domains and text types) to determine the direction in which the system should be developed, so as to meet the particular needs of the Greek authorities, EC SYSTRAN'S target public.

In particular, the group of end-users provides the development team with specialised terminology from their field of expertise, which is often the result of many years of work and would not have otherwise been accessible to developers. As shown in Figure 5, thanks to this group's feedback, the system's lexica have been greatly enhanced with terminology from many different domains such as Fire Brigade, technical manuals, databases, computer science, agriculture, and finance. As developers encode this feedback in the system's dictionaries, updating the system with users' terminology not only improves the performance of the MT system, but also gives the end-users an incentive to provide even more help, as translation quality improves.
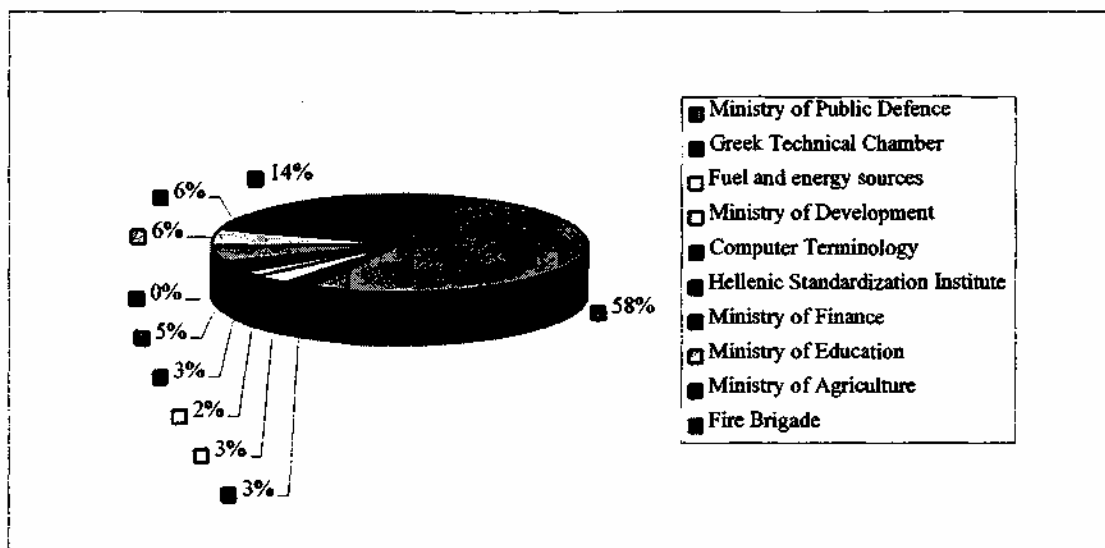
Figure 5: Domain terminology already encoded in EC SYSTRAN'S dictionaries

In addition to providing domain terminology, end-users also contribute to the evaluation process of EC SYSTRAN. Most end-users (~93%) who participated in the information day were willing to attend at least two meetings per year (~28,5%), up to four quarterly meetings (~71,4%). During the meetings (and by completing the relevant questionnaires), end-users can provide the development team with their general opinion of the MT system, comments on the overall quality of output, its usage within the public service (e.g. drafting, browsing purposes) and the improvements that they would like to see carried out. Moreover, all end-users, even those who are not willing to take part in regular meetings, are encouraged to communicate their comments to the development team on a more personal basis, whenever it is convenient for them.

In this way, an evaluation of the translations is achieved, and although not very detailed or specific, it nonetheless gives developers an idea of the system's usefulness from the users' point of view.

Finally, texts sent by public administrations to EC SYSTRAN for translation form part of a corpus used for testing, evaluating and improving the MT system. This corpus comprises those text types and domains that Greek public services use most, namely reports, manuals, articles, books, EC recommendations / guidelines, studies, legal articles, papers and leaflets from the domains of administration / management, policing, aviation, agriculture, sports, law, crime, education and medicine - to mention only the most frequent text types and domains. Such a corpus is invaluable for the development team since, under any other circumstances, it would be very difficult to have access to such texts, the public services in Greece being reluctant to provide texts concerning their internal matters - memos, circulars, letters, contracts, reports, etc. - unless they can benefit from it. Texts sent by end-users for translation contain standardised expressions, terminology and other valuable information used by the development team for the enhancement of the EC SYSTRAN.

The aforementioned evaluation procedure, i.e. active involvement of users, is an adaptation of the one used in the European Commission, where the situation is not very different. Development and evaluation of the EC SYSTRAN system is carried

out by a group of 18 translators, called 'MT correspondents', who use MT in their daily work and dedicate half of their time to the improvement of the system.

All MT correspondents work into their mother tongue and cover all 18 language pairs provided by EC SYSTRAN. They are responsible for providing regular feedback to the Commission's development team (an outside contractor, SYSTRAN Luxembourg S.A.) and for controlling the MT output.

Their involvement in MT varies considerably from one language to another. In most cases, they choose a domain in which EC SYSTRAN can provide satisfactory results for a specific language pair (antidumping, answers to parliamentary questions, etc) when the relevant terminology is introduced. This is the case for English → Greek and French → Italian. For other language pairs, such as English and French into Spanish, a high percentage of all documents are sent to MT by translators and feedback is provided in several domains.

MT correspondents have direct contact with the linguist responsible (at SYSTRAN Luxembourg) for the development of the language pair they work on. In cooperation with the development team, they set priorities and draw up a working plan.

Among their responsibilities is a twice-yearly evaluation of the system using NLPWin, thus enabling them to determine to what extent their priorities have been met and what remains to be done in order to improve the system further.

As far as English → Greek is concerned, developers introduce feedback from both the MT correspondents and end-users in the Greek public sector, keeping a watchful eye on the coherence of the system.

6. CONCLUSIONS

Evaluating an MT system is by no means an easy task as there are no hard-and-fast rules of procedure to be followed. Evaluations are extremely difficult to perform and even then, they are hardly ever made public.

Forming a user group to evaluate an MT system is even more difficult owing to the following:

a) Participation in the user-group is on a voluntary basis and for free.

b) The users do not benefit *directly* or *immediately* from the whole evaluation procedure, i.e. it takes tune to incorporate changes in the system, as its development takes place on two sites, for different types of users, public servants and EC administrative staff. Therefore, lexical update files have to be monitored and double-checked before changes, modifications or additions are permanent.

c) There is generally a tight time limit, prominent in all activities that take place, as users do not want to spend a great deal of extra working hours on evaluation, presentations etc.

The points above cannot be overlooked in terms of their impact on effectiveness. However, they do not seem to have had a negative influence on the evaluation

procedure as a whole, due to the fact that the users who participate in the group find the whole idea interesting and understand the importance of assisting in the modernisation of the public sector, and above all of sharing their knowledge and expertise as this will eventually reflect on their daily work.

On the other hand, as the domains and text types translated are highly specialised, the quickest and most cost-effective way to collect relevant feedback could only come from the users themselves, as EC SYSTRAN is a corpus-based system. Since it can well handle administrative texts from various disciplines or fields, this proved to be the *only way* to truly enhance it.

All in all, it is not the quantity of remarks that counts but the quality, and the ability to improve one area without causing deteriorations elsewhere. It is by far better to have a small team of committed users than large groups that will waste time with more general remarks, which in the long run might also prove useless. That said, it is part of the project's objectives to form new groups of specialists, e.g. chemical engineers, who will be more deeply involved in the evaluation.

In addition, the user group's two sections are at present engaged in lengthy debate over how methodology can be improved to ensure a more objective evaluation.

Last but not least, the remarks from linguists and end-users have served to complement each other as the former focus more on general language matters, the latter on terminology issues. Blending their remarks has proved a successful formula.

References

Hutchins W.J & Somers H. (1992) *An Introduction to Machine Translation,* Academic Press. ISBN 012 362830 x

Piggot, I. M. (1992) *"Systran development at the EC Commission: 1976 to 1992",* Luxembourg.

Senez, D.(1994), 'Developments in Systran' in *Translating and the computer* 16, pp:65-77

Senez, D. (1998), 'Post-editing service for Machine Translation users at the European Commission, in *Translating and the computer* 20.

Linguistic description of Systran, Luxembourg, April 1993

List of Abbreviations

EC SYSTRAN: European Community SYstem TRANslation

CTI: Computer Technology Institute

ILSP: Institute for Language and Speech Processing

MT: machine translation