LANGUAGE TECHNOLOGY AND THE NEW TRANSLATOR

IS ON-THE-JOB-TRAINING THE BEST APPROACH?

Robert Clark, **Praetorius**

It is no secret that the overwhelming majority of graduates completing modern languages degrees or postgraduate programmes in translation studies must wait until they enter the field to acquire the necessary skills to cope with the advances in language engineering that are now taking place. The down-sizing of Machine Translation systems to run on platforms available to the individual translator, along with the new generation of translation memory-based productivity tools, has major implications for the way commercial translations are produced. Would it not be better for new translators and the translation industry to provide these skills before the student graduates? The attitudes of translators and the teachers of translation to language technology have an important influence on student expectations and student orientation. This paper will attempt to examine the steps that are being taken to remedy the situation in Europe and will discuss whether or not these measures will adequately equip translation students to meet the challenges of the 21st century.

Language technology and the new translator - is on-the-j ob-training the best approach? Experience shows that on-the-j ob-training has been, and always will be, the best possible method of teaching new translators current methods and procedures. There is no argument that, in theory, on-the-job-training is the best approach, but is it the most realistic in practice? Do companies have the time and resources to actually put this type of training into practice? Just about everyone I know seems to be working at least twice as hard as they were five years ago - probably for less money! Nowadays, any activity that does not produce an immediate return slips right down to the bottom of the priority list. Unfortunately, training is quite often one of the areas to suffer.

Today we are specifically discussing the training of new translators in the use of language technology, which, broadly speaking, includes just about any computer-based tool that assists in the translation process. But, because of the ever-changing nature of language technology, we must also consider the training of existing translators.

Before any product training can begin, the product in question must be identified and acquired. This may sound painfully obvious, but the point must be made. After all, one of the reasons most of us are here at this conference today is to discover what is already available on the market and what will be available in the future. We all appreciate the fact that keeping abreast of new developments in this area is not at all straightforward. We read journals, we attend conferences, we eavesdrop on the Internet, all of which takes time and costs money. Granted, information is more readily available now than it used to be, but, like wild mushrooms, you still have to hunt for it. It will not magically appear on your plate! Once the product is acquired, there must follow a thorough evaluation in order to establish its suitability for a given environment. Again, this process requires time and resources. Up to this point, the whole procedure has been one big gamble - in the end, there is no guarantee that the product will be appropriate. If it transpires that it is not suitable, one often hears choruses of 'I told you so, if it ain't broke, don't fix it!'

By examining the various types of translation organisations, we can see that the smaller the organisation, the more acute the problem is. We can safely group translation activities into three categories: in-house translation services of large organisations, translation companies/agencies and freelance translators.

The first group seems to be shrinking lately, resulting in the growth of the other two groups. Large companies are feeling the pinch along with everyone else and, as a result, in-house translation departments are either getting trimmed down or disappearing altogether. More translation work is

being out-sourced to translation companies or directly to freelance translators. In some respects, the criteria for language technology in this group is slightly less complex than that of the other two. Large organisations usually maintain a house style for document production and they are not faced with a plethora of document types to deal with. In general, larger companies tend to devote a greater portion of their budgets to training and the acquisition of technology. Those of you that work for such organisations may dispute the fact, but, relatively speaking, this tends to be the case.

Where translation companies and bureaux are concerned, the picture is totally different. They must cater for any document format that is thrown at them. This requires expertise in every major word processing or desk top publishing system, along with translation tools that integrate with those systems. Again, training budgets vary with the size of the company concerned.

When we look at the final group, the freelance translators, we see that, not only do they have the same requirements as the translation companies as regards expertise, but their resources are virtually non-existent.

So what is this all leading up to? I would suggest that, because of the lack of resources and, in some cases, the historical lack of enthusiasm to embrace new technology, the translation industry is far behind the stage of development that could be possible. How could this be remedied? One obvious solution would be to have more time and money - a nice thought, but not likely. Another possible solution would be for new translators entering the industry to bring the expertise with them. Instead of having to acquire this knowledge on the job, they could learn it while they are still at university. In a small way, this is already happening. We do see recent graduates with a working knowledge of most mainstream word processors, some computer-assisted translation tools. Unfortunately, we also find recent graduates lacking even the most basic computer skills. Not long ago, a language graduate from a well respected university came to our company looking for translation work. I asked her what experience she had with computers. She replied that she had used a word processor, but she could not remember which one it was.

Let us examine programmes offered by two institutions that do seem to have it right. Dorothy Kenny of Dublin City University was kind enough to furnish me with the details of courses covering translation technology offered during the fourth year of their BA in Applied Languages:

Week Activities

1   **OVERVIEW**

2   **INTRODUCTION TO COMPUTER SKILLS hardware: the PC and peripherals inside the PC - CPU and main storage the keyboard - key sequences secondary storage: hard and floppy disks software: operating systems vs. applications MS-DOS and Windows**

4   **INTRODUCTION TO APPLICATIONS Word Processing: Word for Windows**

6   **MACHINE CODE AND NATURAL LANGUAGE machine code, ASCII, extended character set, other character sets**

7   **OFFICE AUTOMATION AND COMMUNICATIONS modems, communications software, the Internet**

8,9   **ELECTRONIC SOURCING AND STORAGE OF TERMINOLOGY electronic corpora and tools (ALPS source text analysis, Microconcord), terminology management systems for the PC, termbanks**

10-12 **INTRODUCTION TO MACHINE (AIDED) TRANSLATION brief history of M(A)T, basic design issues, computational issues**

13-15 **LINGUISTICS FOR MT the lexicon, monolingual problems in analysis, contrastive problems**

16,17 **MT DEMOS Globalink, ALPS: import text, analyse text, prepare dictionary, translate text, post-edit, export, print**

18,19 **RECENT DEVELOPMENTS IN M(A)T translation memory, tools for localisation, other MT systems**

20   **ANALYSING TRANSLATION NEEDS cost, volume, nature of texts to be translated, etc. Week 22 PREPARING TEXT FOR M(A)T, POST-EDITING**

I think we can all agree that we would be very pleased, and pleasantly surprised, if a graduate seeking employment as a translator walked into the office armed with this information. An even more ambitious approach is that of the University of Saarland at Saarbrücken, Germany. Their translation technology component runs the full four-year term of the degree programme for future translators and interpreters that lead to the academic degree of 'Diplom-Übersetzer' and 'Diplom-Dolmetscher'. Karl-Heinz Freigang of the University of Saarland supplied the following details:

FOUNDATION COURSE (GRUNDSTUDIUM)

1. Lecture and practical classes: Introduction to Electronic Data Processing for Translators and Interpreters (2 hours per week, 1st semester)

The course consists of a one-hour lecture and a one-hour practice session. It is designed for first-semester students. The main aim is to impart basic knowledge about computer hardware, the methods of data representation and data encoding as well as operating systems and the basic principles of computer programming. Moreover, together with a variety of standard word-processing, DTP, and data bank applications, translation-specific programs are presented. The practice sessions are intended to consolidate the material of the lectures. Specific applications are demonstrated and basic programming principles are exemplified in simple programs.

2. Lecture and computer sessions: Introduction to Word Processing (3 hours per week, 2nd semester)

The course consists of a one-hour lecture and two hours of computer sessions. It is addressed to students who have attended the 'Introduction to Electronic Data Processing for Translators and Interpreters'. Besides more general topics such as the concept and history of word-processing, the lecture primarily provides an overview of various WP-systems and discusses the performance of advanced programs.

3. Lecture and practical classes: Introduction to NLP (2 hours per week, 3rd semester)

The course comprises a one-hour lecture and a one-hour exercise. It is addressed to students in the Foundation Course (approx. 3rd semester) who have attended the 'Introduction to Electronic Data Processing for Translators and Interpreters' and the 'Introduction to Word Processing'. The lecture provides an overview of topics such as different types of man-machine-interaction in NLP (e.g. speech understanding, optical character recognition, etc.), methodological problems in analysing and representing linguistic data, models for human parsing, introduction to machine parsing and language-processing systems in AI.

The practical classes first discuss approaches to 'simple' problems of NLP, such as spell-, grammar- and style-checking. Then, students develop simple algorithms 'on paper' to improve 'algorithmic thinking'. These algorithms are finally translated into BASIC-programs. Recently the course has been completed by an introduction to PROLOG. The intention is to demonstrate some approaches to problem-solving in AI. However, it has to be stressed that this is by no means a traditional course in programming. It is rather meant to give a first insight into different programming techniques and to provide the opportunity of practical PC-training.

4. Lecture and computer sessions: Software Applications for Translators and Interpreters (3 hours per week, 4th Semester)

The course is made up of a one-hour lecture and two hours of computer sessions. It is addressed to students in the Foundation Course (approx. 4th semester) who have attended the 'Introduction to Electronic Data Processing for Translators and Interpreters', the 'Introduction to Word Processing', and the 'Introduction to NLP'. The lecture provides a survey on software applications other

than word processors and terminology management systems that might be relevant for the translator's workbench. These applications include data base systems, spreadsheet programs, graphic applications, grammar- and style-checkers, DTP-programs, and data communication software.

5. Seminar in NLP (2 hours per week, 3rd or 4th semester)

The course consists of a one-hour lecture and two hours of computer sessions. It addresses students in the Advanced Course (approx. 5th semester). The lecture deals with the following aspects of computer-assisted terminography: terminology management at the translator's workbench, introduction to practical work with terminology management systems (TMS), market survey on TMS, entry structures of different TMS, access to existing terminology such as terminological data banks and machine-readable glossaries on disk and CD-ROM.

In the computer sessions students learn how to use two or three different TMS. The major aim is to enable students to independently manage terminology. First experiences have shown that a number of students make use of TMS to store the terminology of their LSP-translation classes in order to compile glossaries for the final exams.

ADVANCED COURSE (HAUPTSTUDIUM)

6. Lecture and computer sessions: Introduction to Computer-assisted Terminography (3 hours per week, 5th semester)

The course consists of a one-hour lecture and two hours of computer sessions. It addresses students in the Advanced Course (approx. 5th semester). The lecture deals with the following aspects of computer-assisted terminography: terminology management at the translator's workbench, introduction to practical work with terminology management systems (TMS), market survey on TMS, entry structures of different TMS, access to existing terminology such as terminological data banks and machine-readable glossaries on disk and CD-ROM.

In the computer sessions students learn how to use two or three different TMS. The major aim is to enable students to independently manage terminology. First experiences have shown that a number of students make use of TMS to store the terminology of their LSP-translation classes in order to compile glossaries for the final exams.

7. Lecture and practical classes: The Theory of Computer-assisted Terminography (2 hours per week, 6th semester)

The course comprises a one-hour lecture and a one-hour practice session. It addresses students of the Advanced Course (approx. 6th semester) intending to specialise in the field of terminography. The lecture treats basic concepts in computer-assisted terminography, examines different concepts of terminology management systems and terminology data banks, discusses entry models and data categories as well as macro-structures such as classification systems, thesauri and subject fields. Terminographic work methods and the problems of exchanging terminographic data are further topics of this lecture.

In the practical classes students develop and analyse terminological entries, analyse and evaluate terminological diploma-theses and work out concept systems.

8. Lecture and practical classes: Machine-aided and Machine Translation (3 hours per week, 7th Semester)

The course, which is organised and held in close cooperation with the institute's chair of Machine Translation, consists of a one-hour lecture and two hours of practical classes. It addresses students in the Advanced Course. After an overview of the various concepts of MAT- and MT-systems the lecture presents the major systems either currently available on the market or under development.

Many of the systems, such as SUSY, SYSTRAN, LOGOS, EUROTRA, GLOBALINK, or TRANS-LATOR'S WORKBENCH II, are available at the institute and can be directly demonstrated on the computer. Other systems are discussed with the help of extensive documentation material.

In this course the main emphasis is put on the practical application and usability of the systems. More theoretical questions that are rather related to computer-linguistics are discussed in other courses offered by the chair of Machine Translation.

9. Seminar on NLP (2 hours per week, 6th, 7th or 8th semester)

Besides the introductory courses in the three major subjects of the component, i.e. EDP/NLP, computer-assisted terminography and machine-aided/machine translation, seminars providing a more detailed analysis of selected aspects are offered in both the Foundation Course and the Advanced Course. Whereas seminars in the Foundation Course focus on basic principles of academic work in general, seminars in the Advanced Course rather intend to allow students a deeper insight into various problems. Besides a paper on the chosen topic, the oral presentation of this topic by using adequate presentation techniques plays an important part.

The following list exemplifies the range of topics dealt with in the seminars (FC = Foundation Course, AC = Advanced Course):

Machine translation essentials and problems (FC) New technologies in the translator's working environment (FC) Introduction to terminography (FC) Linguistic models in NLP (AC) MT-systems (AC) Automation at the translator's workplace (AC) Terminology data banks (AC) Computer-assisted terminology work in theory and practice (AC) Project seminar: Writing a computer manual (AC)

10. Colloquium for advanced students and exam candidates (2 hours per week)

The colloquium, which has been offered since October 1989, serves as a forum for lecturers and advanced students to discuss the contents, structure and preliminary results of current diploma theses in the field of NLP.

Among students and lecturers there is an increasing interest in the colloquium, so that by now it has become an established course at the department.

The following list contains a few examples of diploma-theses supervised by members of the project team:

- Terminological entry structures in Multiterm
- Terminological entry structures in Termex/MTX
- Analysis of entry structures in terminological diploma-theses and their possible realisation in the terminology system CATS
- Computer-assisted terminology work in translation practice
- The exchange of terminographic data
- Analysis and evaluation of standard PC-based terminology management systems in translation practice
- The use of scanner- and OCR-technology at the translator's workbench
- Analysis of an automatic translation system English-French translation results and post-editing
- Analysis of an automatic translation system English-German with special emphasis on system dictionaries
- Linguistic evaluation of a commercial MT-system German-English

The University of Saarland have been offering their programme for over six years. During that time, Karl-Heinz Freigang and his colleagues at Saarbrücken have shared their experience with other European universities with an almost missionary zeal. As a result of their efforts, some universities are beginning to recognise the importance of integrating technology into their language programmes. For example, the University of Germersheim have introduced courses dealing with computer-assisted terminography in to their degree programme. If a survey of universities offering similar programmes were done, we might discover that there are more institutions than we realise following this approach. If this is the case, we would almost certainly find that the language degrees concerned are not traditional university modern languages degrees, but applied programmes. Several universities in the United Kingdom are now forming Masters programmes in Translation Studies to supplement their traditional modern languages degrees.. The University of Leeds will offer such a degree in the 1996 academic year. The programme will follow the current trend of 'modularisation' and will have as one of its core modules a Translation Technology module. The 'good news' is that the situation is changing for the better. The 'bad news' is that is taking a long time to happen. What could be done to speed up the process?

Universities have the same problem as the translation industry - before any training can take place, the trainers need to be trained. How can we assist the universities to acquire the expertise required to offer students translation technology courses? One solution would be closer cooperation between the translation industry and academia. Professional translators could contribute to the cause by giving seminars at universities outlining the changes in the demands of the translation market, particularly regarding skills in using modern technologies. If universities have a better understanding of what .is expected of their graduates, they might be more prepared to tailor their degree programmes to reflect those requirements. If some universities do not have staff with the necessary expertise to teach such courses, perhaps they could draw from the professional translation community to fill this gap. This is the approach that the University of Leeds is planning to adopt. One other factor that could be considered is the software required for such training. At the moment, universities must purchase software that would be used for teaching purposes. This is just one more burden on an already stretched budget that could be alleviated by a different approach on the part of the software vendors. Companies, such as IBM, Systran or Trados, could supply academic licences to universities at a greatly reduced rate, or, better yet, free of charge. As long as the academic institutions agreed to observe the licence agreement and enter into an ongoing technical support contract, this arrangement would provide the vendors with invaluable feedback. I think most developers of translation technology would like to expand their user base. This could be the answer.