

Can Language Technology Co-exist with Traditional Translation Studies?

Integrating Language Technology into a Postgraduate Translation Program

by Andrew Rothwell and Mark Shuttleworth

In early 1995 the School of Modern Languages and Cultures centres at the University of Leeds, among the largest centres for the study of Modern Languages and Linguistics in the UK, decided to launch a postgraduate translator training Masters. The idea was to combine the School's unrivalled range of languages, which would allow students to study say Japanese and Portuguese, or Chinese and Bulgarian, with its cross-departmental expertise in Linguistics and Translation Theory, and the experience of the many practicing translators on the staff, into a program that would have unique strengths and a distinctive practical identity. The two present authors, Mark Shuttleworth (then of the Leeds Department of Russian) and Andrew Rothwell (French) were invited to design and co-ordinate the program.

Quickly, however, we realized that there was something missing from our plans: the kind of specialist computerized translation tool that had recently been demonstrated to a research group in the School. The software was the then new IBM TranslationManager, and the speaker



was Robert Clark, Leeds languages graduate in Russian and Arabic and (fortunately for us) Leeds-based professional translator and writer on Language Technology (LT). Without Bob's intensive professional input from outside the academic sphere, the MA in Applied Translation Studies (MAATS)¹ at Leeds could not have developed the software-intensive slant which proved to be its most distinctive selling-point and allowed it to exceed the recommendations of the European Union's LETRAC (Language Engineering for Translators' Curricula) project well before its 1999 report was published². In the first year of operation (1996-7) Bob taught the whole of the compulsory Translation Tools module on his own, and only gradually did a few academic staff begin to acquire the expertise to contribute to that component of the course. The learning curve was steep and is still ongoing, since every year brings new products, new versions of existing tools, new operating systems and/or new IT training facilities, all of which introduce unknowns and fresh problems and require teaching materials to be rewritten.

Now that MAATS has spawned offspring of its own³, it seems appropriate to stand back and review the lessons learnt from five years of incorporating LT into post-graduate translator training programs, starting with three general observations:

1. LT skills are definitely in demand from the translation professions and graduates who have them have much greater employability;
2. LT can and should be taught in an academic environment, but with strong support from the translation profession and software developers, as part of a rounded and integrated program;
3. Expert technical backup is crucial.

Observation 1 has now been conclusively demonstrated by experience. Observations 2 and 3 go together: although it cannot fully mimic the "real world" of commercial translating, a university course provides an ideal environment in which fledgling translators can experience a whole range of LT tools, rather than just those adopted in a particular company, and explore their comparative functionality and design principles in depth, rather than just learning the features required to "get a job done". Installing and maintaining LT software in an academic networked environment for which in many cases it was not designed does, however, place extremely heavy demands on technical support staff. Once it is working, the benefits (collaborative working by students, access to programs and data from anywhere on campus) can be invaluable, but the difficulties involved in getting it right should not be underestimated⁴.

Objectives

Realistic goals need to be set when planning the syllabus for a postgraduate training course of this type. After all, there is typically only one calendar year to play with, and participants are likely to come from a variety of backgrounds as far as nationality, age, previous translation experience and IT expertise are concerned. Clearly, attempting to cover every aspect of translator training in equal detail is an unrealistic goal given these very real limitations. However, over the past four to five years we have generally found it perfectly feasible to produce graduates who have at least a basic amount of experience translating a range of text-types and are equipped with a range of skills which make them highly desirable to potential future employers. In addition, we try to

ensure that their new-found knowledge and skills are underpinned with a significant grasp of theoretical issues.

In practice, though, the main focus of our courses has been on providing training in LT. Somewhat surprisingly, given the overtly technology-intensive nature of our programs, our participants have ranged from the highly computer-literate to the out-and-out technophobe. One of our greatest challenges has therefore been to ensure that all students commence the course with at least a basic proficiency in Windows skills and some familiarity with a number of standard Windows-based applications. Before students can work with TRADOS Translator's Workbench, they quite clearly need to be able to resize a window; before attempting to develop a multilingual terminological database, they must know how to enter accented characters.

**...over the past four to five years
we have generally found it perfectly
feasible to produce graduates who
have at least a basic amount of
experience translating a range of
text-types and are equipped with a
range of skills which make them
highly desirable to potential future
employers.**

Within this context our chief aim is to provide our students with a thorough understanding of and wide-ranging experience in the use of industry-standard translation software. Hand in hand with this, we seek to make participants aware of a range of modern working methodologies and get them used to the idea of teamwork and project management. It also goes without saying that familiarity with the Internet occupies a place high up on our agenda.

On a more theoretical level, we aim to familiarize students with all aspects of the translator's decision-making and to introduce them to a range of theoretical and historical thought on the subject of translation. In addition, we seek to make explicit the interface between LT, transla-

tion theory and linguistics, as we hope will be made clear below.

Delivery

Depending on the main subject-focus of the program—which will in turn be determined to a large extent by the constraints operating at an institutional level—it is possible to cover a wide range of subject areas and text-types. Obvious candidates for coverage include product documentation, institutional documentation, scientific, legal and medical topics, localization projects (involving both software and Web pages), marcom materials and so on. Covering a range of texts—and including plenty of less repetitive material to highlight to students this fundamental limitation on translation memory technology—will enable students to reflect on the limited usefulness of the software when dealing with certain text-types. Students can be encouraged to put into immediate practice the skills and techniques which they have been acquiring in the computer laboratory when working on their practical translation tasks; there is also plenty of scope for coordinating tool with text-type so as to play to the strengths of each particular package.

Because of the control over language settings which it gives to individual users, Windows 2000 is rapidly becoming the operating system of choice. We have also found the use of a networked facility with dedicated server space and qualified technical support to be indispensable. This not only enables classes to be conducted in an environment in which the software can be delivered to users in a reliable and controlled manner but also affords unlimited opportunities for students to practice and complete assignments out of class time.

Training sessions are usually fully hands-on, although sometimes feature a presentation or demonstration from the front. The teaching team includes not only lecturers but also demonstrators, who are typically former students of the course; in this way we aim to offer language-specific expertise for most or all the languages which are being actively used by students studying on the programs. This approach is backed up by detailed courseware with step-by-step explanations of procedures and functionalities; however, knowledge is taken to be cumulative, with skills and conceptual understanding acquired while working on one package being assumed to be available for all applications subsequently covered.

LT Syllabus

For reasons which have been outlined above, our programs start with a few sessions on Windows basics, fonts and codepages, and both basic and more advanced word processing skills. This is swiftly followed by a section on how to exploit the Web for translation-related purposes. This includes such topics as term databases, search engines and creative terminology mining. Once students complete this component the stage is set for the introduction of specialized LT tools.

There is such a bewildering variety of tools available now that it would be infeasible to cover all specific application types, let alone each individual product within every single category. Our three programs all concentrate above all on translation memory applications. The Leeds program, which with five years under its belt is by far the oldest of the three, has gradually extended its coverage from a single tool to four. Thus students now cut their teeth on IBM TranslationManager, the application on which they learn all about how to produce a translation within the translation environment, how to work with dictio-

naries and translation memories, how to import and export information, and so on. The advantage of introducing them to IBM TranslationManager first is that they

The teaching team includes not only lecturers but also demonstrators, who are typically former students of the course; in this way we aim to offer language-specific expertise for most or all the languages which are being actively used by students studying on the programs.

can see everything functioning within a single integrated application without having to worry about the sometimes significant challenges of making as many as three separate applications function in a coordinated manner. IBM TranslationManager is followed by STAR TermStar

and TRANSIT, which in turn are followed by TRADOS Translator's Workbench. Passolo is also introduced as the software localization-specific tool. Swansea uses the same basic arrangement with minor alterations. Imperial will on the other hand be taking a slightly different approach, aiming to provide students with an overview of a wider range of tools.

While translation memory systems form the main focus for all three programs, in each case the syllabus includes significant coverage of terminology management tools. The aim is for students to become proficient in using the (often stand-alone) terminology component of the range of workbench tools covered to create, consult, import and export bi- and sometimes multilingual databases. Some time is devoted to the principles of database structure, while high flyers have the chance to master how to manipulate data between formats using the global search and replace and mailmerge facilities of Microsoft Word. Finally, the explosion of Web-based machine translation systems has made incorporation of this important area of translation technology on a hands-on basis something which is fairly easy to



implement, at least at a basic level. However, in order to give students a feel of the true potential of machine translation it is probably necessary to have a locally-installed system available as well.

Significant staff effort is required to design and update student workbooks for each LT product on the syllabus. It is important for the exercises chosen to be not just procedural (which buttons to press to

In an integrated translator training program it is important for LT training not to be seen as a discrete component unrelated to the other skills students are learning.

accomplish a particular task) but conceptual (how the product has been engineered to accomplish a particular task, and the advantages and disadvantages of doing it that way). With students of a dozen or more languages working in the same lab, it is important to write the workbooks and practice texts (often software Help files) in the common language (in our case English), to avoid having to produce multiple versions. This does mean that for practice purposes mother-tongue English students are translating the "wrong way"; others, however, find themselves translating into their mother tongue, which is some compensation to them for the fact that the rest of the program is in English. Tasks can be made cumulative, i.e. the practice text for a second workbench-style package can be a new version of the one translated with the first package and, after suitable data conversion routines have been applied, re-use the same dictionary and translation memory (this really brings home the twofold lesson that translation data is reusable, but also that bad translations come back to haunt the translator!). Workbooks facilitate self-directed study and allow students to work at their own pace, but it is important that they be substantially bug-free (otherwise large amounts of demonstrator time are spent talking students through the errors) and properly "localized" (in terms of drive letters etc.) for the networked system on which the product is being taught.

LT "Interface" Issues

In an integrated translator training program it is important for LT training not to be seen as a discrete component unrelated

to the other skills students are learning. In particular it should be designed to interface both with practical translation exercises and with the study of linguistics and translation theory, so that students can apply both their practical experience as translators and their theoretical understanding of the translation process to the complex problem of how to optimize LT use in the real world.

A number of exercises have been devised to link LT with practical translation classes, including building in MT assignments in which students select texts for automatic translation, critique the output in linguistic terms, post-edit it to different quality levels and report on their experience, and Terminology Acquisition Projects in which students use Web and paper-based resources to research a technical domain about which they have absolutely no prior knowledge, then produce for assessment a bilingual term-base in one of the terminology management packages they have been taught to use, together with a short methodological commentary (both linguistic accuracy and use of the software are assessed). In this latter case, knowledge of structural semantics and basic lexicography gained from the Translation Theory module also comes into play. Finally, the two summer Extended Translations which complete the MA syllabus are (in the Swansea program at least) performed using two different workbench systems, and the

It is important for the exercises chosen to be not just procedural (which buttons to press to accomplish a particular task) but conceptual (how the product has been engineered to accomplish a particular task, and the advantages and disadvantages of doing it that way).

assessment gives equal weight to the quality of the data files and of the translations themselves.

LT interfaces with translation theory in a number of other ways, beginning with the broad notion of cultural equivalence as it relates to the complex issues underlying localization. The theory module at

Swansea in particular also places heavy emphasis on the linguistic principles and problems of MT design (POS tagging, parsing, transformational syntax, structural semantics and knowledge representation), to give students an insight into why successful MT is so profoundly difficult to achieve, to deepen their awareness of what they as human translators are doing every time they perform a successful translation, and to reinforce their awareness of the differences between MT and the workbench tools that they are using in LT classes. Recent research on problems such as automatic lexicon-generation from non-aligned and aligned corpora (conceptually identical of course to translation memories) is discussed to stimulate thinking about the likely future convergence of MT and CAT technologies, and students are invited to consider how useful it would be if the translation memory tools they have learnt to use were enhanced by different types of MT-like linguistic analysis, rather than being, as they are today, simply dumb string-matching routines.

(This article is based on a presentation given at the third Encontros de Tradução, held on 25-26 May 2001. The conference's title was "Training the Language Services Provider for the New Millennium", and it was jointly organized by the Portuguese Translators' Association and the Faculdade de Letras at the University of Porto.)

Mark Shuttleworth is Senior Lecturer in Scientific, Technical and Medical Translation at Imperial College, London. Dr Andrew Rothwell is Professor of French at the University of Wales Swansea.

1. See <http://smc01.leeds.ac.uk>
2. See <http://iaisun.iai.uni-sb.de/letrac>
3. The MA in Translation with Language Technology at University of Wales Swansea, co-ordinated by Andrew Rothwell (www.swan.ac.uk/sel) and the MSc in Scientific, Technical and Medical Translation with Translation Technology at Imperial College London, co-ordinated by Mark Shuttleworth (www.hu.ic.ac.uk/translation/intro.html). Bob Clark remains a Co-Director of MAATS.
4. We gratefully acknowledge here the vital technical contribution made to MAATS in its first three years by Michael Beddow, then Professor of German at Leeds.