Technology in the Translation Class: Introducing CAT Tools to Hungarian Translation Students

Balázs Kis

MorphoLogic, Budapest

Translators' and Interpreters' Training Centre, ELTE University, Budapest

Resumen

Esta ponencia relata las experiencias de los profesores al enseñar a alumnos de traducción húngaros a utilizar y a sacar el máximo partido de las herramientas de traducción asistida por ordenador (TAO). Éste es el segundo año en el que el Centro de Formación de Traductores e Intérpretes de la Universidad ELTE de Budapest enseña el manejo de herramientas de TAO a alumnos licenciados en traducción. Este proyecto requiere prestar particular atención a asuntos relacionados con la UE, ya que este curso se introdujo por primera vez en las clases de traducción especializada de la UE. El programa del curso está adaptado a las necesidades específicas del idioma húngaro y de los estudiantes de traducción húngaros. Los estudiantes que normalmente participan en este curso suelen poseer un Máster en dos lenguas y conocimientos informáticos básicos. Otra característica de este proyecto es que combina de forma única las herramientas de TAO húngaras con las herramientas estándar utilizadas por los servicios de traducción de la UE. Es importante recalcar este hecho, ya que las herramientas de TAO que se desarrollan en Hungría están adaptadas a las características de la morfología y sintaxis húngaras y, en consecuencia, pueden realizar traducciones de textos húngaros de manera más rápida y eficaz. Todo el programa está específicamente diseñado para que los estudiantes sean capaces de utilizar de forma eficaz las herramientas de TAO con las que trabajan. Por esta razón, se imparten muy pocas clases teóricas y la mayoría del tiempo se dedica a que los estudiantes realicen ejercicios prácticos, en grupos de tres o cuatro personas, mientras trabajan en las traducciones que se les han asignado. Tanto nuestra experiencia personal como un estudio al que hace alusión esta ponencia demuestran que los estudiantes que asisten a este curso adquieren la suficiente competencia en materia de herramientas de TAO como para poder enfrentarse con éxito a futuros proyectos de traducción, incluidos proyectos de la UE.

0. Introduction

This paper examines the possibilities of introducing technical training in translation courses, with special regards to circumstances in EU enlargement countries, especially Hungary. Section 1 outlines the translation challenge faced by these countries at the time of writing. Section 2 describes a specialised EU translation training programme at a Hungarian university. Section 3 examines technical skills required by translators to successfully undertake their tasks. Section 4 explains the priorities and methodology of the course the author has devised for the above-mentioned training programme. Section 5 outlines further improvements. Throughout the paper, the circumstances are related as seen by the author. No formal survey has been carried out with regard to this paper. However, independent surveys (Drugan 2004, Fulford & Granell-Zafra 2004) provide solid evidence for the observations.

1. The translation challenge for enlargement countries

In enlargement countries such as Hungary (with only a few days before officially joining the EU at the time of writing), the translation community is still struggling to catch up with the enormous task of translating EU documents as the bulk translation of EU legislation is still in progress. Translation demands are not foreseen to become less pressing as Hungarian will soon be an official language of the EU.

Translation resources are short, so an ever-growing translation infrastructure will be needed over the few years to come, mainly from the Community working languages (English, French, German) into new languages and vice versa. It is by no means surprising that the most used language among the above is English: over 95% of legislation was translated from the English text into Hungarian. The translation challenge consists of a number of major factors. The most important ones are briefly explained below.

Terminology is considered a major problem within EU translations because new terminology was required practically overnight. The task was taken by several organisations and Hungarian

terminology was created mostly in a parallel manner. Although there was an attempt from the Ministry of Justice to co-ordinate the terminology-establishment process, the translation services of various bodies of the EU and different organisations in Hungary are still using inconsistent terminology.

The Ministry of Justice has established a terminological database publicly available over the Internet, and it is gaining momentum as the most authentic source of Hungarian EU terminology. However, its coverage is not always able to meet the requirements of translators, and inconsistent terminology is still created in areas not covered by the database. It is important to note that Hungarian translation organisations recognised this problem and aim at implementing total harmonisation of terminology by non-governmental means, too.

The other major problem is the shortage of foreign language speakers within the Hungarian population. Therefore it is essential to provide translation services for the public, so that Hungarian individuals and enterprises will be able to apply for EU grants in a more efficient manner.

According to the Hungarian Census in 2000, only 17% of Hungarians speak at least one foreign language, and it is suspected that less than 12% speak English. Though this number is rapidly increasing, it will still take years, if not decades, to reach a level where we can say that speaking foreign languages is common to the Hungarian society.

In order to bridge the language gap more quickly than the pace at which the population learns foreign languages, the overall translation capacity must increase. In other words, those that already speak one or more foreign languages must be taught to use them at a professional level. The output of professional translation training must increase, and, at the same time, the translation workflow must become more efficient, mostly through exploiting new technologies. Translation training is now put under increasing pressure by both of these demands.

2. Background: The programme at ELTE University, Budapest

2.1. Translator training in Hungary at present

Translation training is available as postgraduate or supplemental education at several institutions (two of them are to be found in Budapest). There are currently no undergraduate translation training courses.

Translation training focuses mostly on Commission working languages as 'B' languages, namely, English, French, German, although other languages (such as Spanish, Russian or Italian) are also popular.

2.2. The specialised EU translation course

Specialised EU translation training is available as an additional postgraduate degree at ELTE University's Translation Training Centre (http://www.ftk.hu). This two-semester course is a follow-up to 'regular' translation training, i.e. completion of a regular translation course – or proof of equivalent skills – is a prerequisite here.

Within the specialised EU-oriented programme, translators gain practice with specific texts produced by different organisations within the European Union. Translation practice is supplemented with extensive coverage of EU terminology and various introductions to the institutions and the processes of different bodies of the EU. This is also the first such programme in Hungary where technical training, covering computational tools for translators, is compulsory.

2.3. Technology training background

Technical training at the specialised EU translation programme is currently provided by language technology experts aware of the capabilities and limitations of specialised technology, namely, that of computer-assisted human translation. While one may doubt the choice of such individuals as technical instructors, the

availability of translation experts with sufficient technical skills is very limited. In addition, the lack of day-to-day translation experience on the instructor's part might be compensated by sufficient empathy.

Technical training at the specialised translation course is undertaken by experts working mainly for MorphoLogic, a Budapest-based language technology company. MorphoLogic develops translation software itself, with special attention to characteristics of the Hungarian language. The company, however, is not currently present on the market with specialised tools usually required on the translator 's desktop, so that an unbiased coverage of tools from various manufacturers is guaranteed.

3. Technical skills for the translator

When designing a technical course for translators, one must very carefully assess the skills that students should acquire upon completion of the training. However, the resulting curriculum is based on a delicate balance of factors such as existing skills of students, the time frame available for training, the resources available at the training site, both in terms of hardware and software.

A large majority of translators in Hungary are freelancers: this is not expected to change over the next 5–6 years. Freelancers are practically left alone in more than one sense. Firstly, they have no company or team to impose specific requirements on them. As a consequence, they have little awareness of the computational skills they need, and are able to gain this experience only through difficulties. Secondly, they have virtually no technical assistance, or, if they have any, it is by no means formal and there is no guarantee that it will be at their disposal when required.
As a result, they need technical skills for

- productivity: it is vital that they maintain a competitive performance and availability (Austermühl 2001)

- communication: to overcome relative isolation and work in teams, however virtual

- maintenance: to preserve availability of their computer infrastructure even when they have no access to technical assistance

3.1. Core skill components

Production skills, as regards to computer use, include the following components:

- General text manipulation: a literacy with word processing is required in almost all areas of the labour market, and this is even so with translators. In addition, a recurring problem within the translation workflow is the treatment of documents in various formats. Not only must the (freelance) translator be able to extract text from documents of less frequently used formats, but they are often required to preserve the formatting of the original document. Therefore, the ability to use a range of desktop publishing applications is also a requirement.

- Specific translation aids: this means translation memories, terminology management systems and special purpose translation programs in general. Many translators and translation experts include machine translation among the tool set. Since none of these translation tools are by any means perfect – without substantial human assistance, they are not able to produce high-quality translations –, and each of them includes language technology to some extent, a well-prepared translator must have a good understanding of the operation and limitations of such programs. In addition, practical skills must be acquired with more than one program because currently there are substantially different products on the market.

- General research aids: these are various online resources – other than electronic dictionaries – such as Internet

search engines and domain-specific knowledge bases. Translators need a good command of Internet usage and thorough understanding of search engines.

Communication skills include the following:

- E-mail, chat etc.: general interpersonal communication tools, both for receiving and submitting translation jobs, and general consultation.

- Online collaboration tools: there are a number of generalized collaboration tools (such as Blackboard) and groupware (such as Lotus Notes or Microsoft Exchange). In addition, networked translation workflow management systems are gaining ground. As for today, a general awareness of such tools might be sufficient, but further skills will be necessary as such processes become standardized.

As mentioned earlier, a freelance translator must have some maintenance skills in order to avoid downtime when working with tight deadlines. Although a personal computer system does not seem very complicated, several steps must be taken to prevent various problems such as virus attacks. A freelance translator must be aware of tools (such as virus protection programs and firewalls) and actions (such as installing security updates regularly) required for keeping their computers in good shape. To this end, they need some understanding of the operating system and in general the principles of computer operation: technical problems must be identified in time and appropriate actions must be taken to overcome them.

3.2. Multiple levels of technical skills

The previous paragraphs described core skills, i.e. those required in the skill set of all translators. However, there are multiple levels, depending on the task the translator is to undertake. A brief list follows:

- Core skills: described in detail in Section 3.1. Note that at the specialised EU translation training, only these skills are provided for the time being.

- Website localisation or subtitling: specialised tasks requiring more-than-average technical skills. Website localisation requires an understanding of the structure of HTML and XML documents, and the ability to identify translatable parts. Knowledge of web page editing software is also necessary. Subtitling requires knowledge in specialised subtitling programs.

- Advanced software localisation: this requires some programming skills, and good understanding the structure of computer software. The translator involved in software localisation must have knowledge in programs assisting localisation. (Esselink 2000; Kis 2002)

- Research in language/translation technology: for translators who wish to pursue a scientific career, various PhD programmes are available even in Hungary. At the Translators' and Interpreters' Training Centre of ELTE University, a PhD programme in translation studies was started last year.

4 The course

Of the technical course within the specialised training, the following paragraphs briefly describe

1. the starting point: what the students' existing skills are;
2. basic concepts: the priorities of the curriculum;
3. the methodology.

4.1. The starting point

We have looked at the skills translators require to comply with their tasks. It is also important to know what the students' skills

are when they enter technical training. We can start from the following facts:

- A large majority of postgraduate translation students have MA-equivalent degrees in one or more foreign languages.

- Though basic computer literacy is now taught at arts faculties, the typical student has little more than basic word processing, e-mail and Internet skills.

4.2. Priorities

Within the current scheme, we have two semesters to provide as many core skills as possible. Though basic computer literacy is often taught insufficiently at secondary schools and universities, the course, due to the limited time frame, must assume that students have basic skills. Thus general computer literacy is not taught, but emerging problems are addressed, and related skills are reviewed. Most of the course is restricted to skills with specialised translation tools.

If the skill set is identified, it is still a difficult task to provide students with up-to-date knowledge. General production, communication, infrastructure skills do not change much over time, but specialised translation tools are expected to undergo substantial change in the next 5 years. We have mentioned that translation tools (both translation memories and machine translation) are far from perfect, and the developers are presently carrying out fundamental research, which might bring dramatic changes.

The changes include the following:

- More intelligent translation memories: Synergy of example-based translation (EBMT) and language-aware translation memories is being worked on by several research teams, including the one of the author (Grőbler-Hodász-Kis 2003).

- More intelligent termers (Jacquemin 2001) are expected to appear in CAT tools, where automatic term extraction is still more theory than practice.

Both technologies will benefit from (semi)automatic harvesting of online resources. Research in this respect is expected to take a few more years because two significant problems must still be overcome: (1) copyright problems of downloaded resources, (2) automatic extraction of useful data.

In addition to the technical aspects described above, changing the students' attitude toward technology must also be an important goal (Drugan 2004). Several translators with backgrounds in literature can still be considered 'technophobes', unwilling, but not unable, to learn using technology. Thus the course must to some extent develop the ability to learn and adapt to new techniques. This includes the ability to use equipment individually (maintaining infrastructure, overcoming problems), and an insight into technology's capabilities and limitations.

4.3. Methodology

The course is largely built upon student projects: preferably, students are assigned a complex task for an entire semester. The task aims at translating a short piece of text exploiting the most technical means possible, with an emphasis on the integration of technology. The task requires the students to follow the steps below:

- Building terminology
- Collecting and aligning parallel corpora
- Building a translation memory
- Converting and formatting text files of various formats
- Practicing translation: exposed to both user paradigms of translation memories

By 'both user paradigms', we mean that there are basically two types of translation memory user interfaces. A translation memory can either assist the user when they are translating text using their common word processor programs; or, the program can provide its own editor, simplified but more appropriate to the translation task (usually a two-column table). The former requires less learning on the translators' part; however, the latter approach results in far more reliable programs. An example of the former is Trados Translators' Workbench, while Déjà Vu and STAR Transit follow the second scheme.

Students are required to form teams of 3–4 people based on their B language, at the very beginning of the semester. We provide all-practical training: explanations and minimum theoretical backgrounds are provided when an exercise requires them. (Lengyel 2004)

5. Further steps

5.1. Survey

In the introduction, we mentioned that currently there is no hard evidence of the circumstances we are relating through the paper. In order to properly evaluate the technical training, both the clients working with former students and the former students themselves must be surveyed about their work. To date, too few students have left this training to provide sufficient survey data.

5.2. Extension of technical training

At the time of writing, translation courses in Hungary usually do not include technical training. However, it is essential to extend training to lower-level translation courses. Steps are being taken to achieve this.

Training should also be extended to higher levels of technical skills such as web site or software localisation, with special regard to the requirements of translation students.

References

Austermühl, Frank (2001) "Electronic Tools for Translators" Manchester: St. Jerome.

Drugan, Joanna (2004) "Training Tomorrow's Translators" in Proceedings of the IV Conference on Training and Career Development in Translation and Interpreting Madrid: Universidad Europea de Madrid. To be published.

Esselink, Bert (2000) "A Practical Guide to Localization", Amsterdam & Philadelphia: John Benjamins.

Fulford, Heather & Joaquín Granell-Zafra (2004) "The freelance translator's workstation: an empirical investigation", in Proceedings of the Ninth EAMT Workshop, Valletta (Malta): Foundation for International Studies, University of Malta, 53-61.

Hodász, Gábor & Tamás Grőbler & Balázs Kis (2004) "Translation memory as a robust example-based translation system", in Proceedings of the Ninth EAMT Workshop, Valletta (Malta): Foundation for International Studies, University of Malta, 82-29.

Jacquemin, Christian (2001) "Spotting and Discovering Terms through Natural Language Processing" Cambridge (Massachusetts): The MIT Press.

Kis, Balázs (2002) "Training Seminar on Translation and Localisation. Universitat Rovira i Virigli, Tarragona, Spain, 10-11 May, 2002" in Across Languages and Cultures 3 (2) (2002), Budapest: Szent Jeromos Alapítvány.

Lengyel, István (2004) "Group Translation: Exploiting Synergy.", in Proceedings of the IV Conference on Training and Career Development in Translation and Interpreting, Madrid: Universidad Europea de Madrid. To be published.