From Testbench to Workflow: Relocating MT in Education and Training

Tony Hartley	Klaus Schubert
School of Languages	Studiengang Technikiibersetzen
University of Brighton	Fachhochschule Flensburg
Falmer	KanzleistraBe 91-93
UK - Brighton BN1 9PH	D-24943 Flensburg
Tony. Hartley@itri. brighton. ac. uk	schubert@fh-flensburg.de

Introduction

In this paper we describe current initiatives in the teaching of machine translation (MT) at Flensburg and Brighton. We believe that our teaching programmes are innovative and realistic, presenting to students both an academic challenge and skills that are in demand in the world of multilingual document production.

The key to our approach is that we do not consider MT systems "in vacuo", as black boxes to be tested on a laboratory workbench. Rather, we foreground the process of translation, creating a variety of workflow scenarios which mirror current industrial practice as well as developments in the foreseeable future. These give our students experience of using commercial MT systems as tools in a document production cycle that begins with authoring and ends with maintenance.

Adopting this approach enables us to pursue our education and training goal of producing reflective practitioners familiar with (cf. Schubert 1993: 308):

• computer-assisted tools for document processing

• computer-assisted tools for translation;

• techniques of system evaluation from a user's perspective;

• good practice in making a user's contribution to system development.

The workflow approach increases the number of roles as stakeholders in MT that students can play. Instead of casting them solely as developers, we enable them to play the parts of translators, revisers and managers also.

Furthermore, we recognise the importance of Sager's (1989: 91) stance on translation quality in general, that "there are no absolute standards of translation but only more or less appropriate translations for the purpose for which they are intended". This comment is echoed by King with respect to defining quality requirements in MT, when she argues that the needs that count are "the needs not of the users of the translation software, but of the translations produced." (King 1997: 261).

Accordingly, the key features of our programmes are:

- an emphasis on all types of evaluation of MT systems, from feasibility to usability in operational settings;

- the existence of end users for the translations.

Background

The educational environments for the teaching of translation and MT are different in our two countries.

In the German-speaking countries, computer aids to professional translation work, and MT in particular, have long had a recognised place in the curriculum of translator training establishments. In Germany, the *Bundesverband der Dolmetscher und Übersetzer e. V.* (BDU) was instrumental in urging the introduction of current working techniques in the curricula. In their 1986 "Memorandum" (Koordinierungsausschuß 1986) the BDU mentioned this need, to further elaborate on it in much more detail e.g. in the *Wiesbadener Profil* of 1993 (Bayer et al. 1993). The University of the Saarland at Saarbrücken played an important role with its 1988-1993 pilot study of language technology in translator and interpreter training (cf. Fischer et al. 1993; Haller 1995). While the Saarbrücken test was still going on, other establishments of academic translator and interpreter training simply introduced the relevant components to their curricula. At present, terminology databases, text analysis software, translation memories and translator's workbenches as well as MT systems both on PCs and on UNIX workstations are used in most translator training programmes (cf. e.g. Neubert 1996; Schmitt 1996; Ahrens 1997; Bohm 1997).

In the UK, however, there is no such common framework for university education and professional translators themselves remain relatively ignorant of the state of the art of MT development and deployment. While the application of low-cost MT systems for language learning purposes is quite widespread (cf Mitkov 1996; Lewis 1997), the use of "industrial grade" systems in translator training remains restricted to a few universities, including Leeds, UMIST and, in the Republic of Ireland, Dublin City University. Indeed, a survey by Essex University in 1996 to gauge support for a national initiative in MT teaching roused little enthusiasm. In what is, in any case, a highly decentralised tertiary system, individual establishments make their own judgments as to the best agenda to pursue. Concertation between universities and professional bodies, such as the *Institute of Translation and Interpreting* (ITI), is also relatively low-key.

Guiding aims and principles

In neither Flensburg nor Brighton are translation and machine translation a means towards an end of language learning. Although coming from different starting points, we have converged on common aims:

- to give students the competence to carry out professional translation work – with or without computer aids;
- to give students practical experience of using particular MT systems;

- to enable students to take a strategic view of the use of MT within a particular organisational setting.

The principles to which we subscribe are:

- to engage our students with authentic tasks which mirror those performed by professionals in the field;
- to encourage our students to reflect on the wider implications of the particular task with which they are engaged at any one time.
- to equip our students with the competence to judge the desirability and the functioning of automation in language-related work.

Creating workflow scenarios allows us to respect all of these principles, insofar as they provide a chain of interdependent activities for students to perform and integrate tasks executed by both humans and machines. To be a reflective practitioner means not only being able to decide which workflow scenario best meets the needs of the moment. It also means being able to judge and explain the relative merits of the possible alternatives, since workflow is essentially about a relationship between clients and service providers, and between service providers and their language professionals, whether in-house, subcontracted or freelance. Reflection, then, is both a necessary feature of academic education and a prerequisite for a successful career.

The principal activity for promoting reflection is evaluation. We now turn to the different types of evaluation that can be applied to machine translation.

Types of MT evaluation

We review briefly the different perspectives from which MT systems can be evaluated and the correspondingly different capacities in which students can be called upon to act.[1] The account below is based on White (1997).

Feasibility evaluation

This is a task for researchers. The goal is to assess whether a particular approach (i.e. a linguistic theory or a computational-linguistic formalisation) is capable of achieving a translation objective. The answer will lie in the coverage of the theory and its extensibility, and the method will rely on test suites or corpus tests.

Internal evaluation

This is a task which engages both researchers and developers. The goal is to measure whether modifications to a system improve or degrade its performance. The evaluation may reveal necessary changes to the functionality of the system to be made by the developers. Equally, it may focus on modifications carried out during the running and maintenance of the system by users, such as updates to the dictionaries and changes to the

---

[1] Accessible introductions to evaluation issues are to be found in Sparck Jones & Galliers (1996), Machine Translation 8 [1-2] 1993 special issue on evaluation of MT systems, and at http://issco-www.unige.ch/ewg95/ewg95.html.

settings of user-definable system parameters. The answer lies again in coverage, and the method will rely on error analysis, using both "black box" and "glass box" approaches.

Declarative evaluation

This is a question for developers, end users (translators, revisers and MT lexicographers) and translation managers. The goal is to assess how well a system can translate text that is representative of its intended use. The answer lies in the fidelity and intelligibility of the output. The method will rely on error analysis and judgments of "goodness" by consumers of the translated texts.

Usability evaluation

This is an issue for end users and managers. The question addressed is: How usable is the system? It is answered by measuring the ease with which users can learn, perform and remember to use the system for specified tasks. The methods employed include observation, protocol analysis, questionnaires and interviews - techniques developed by Human-Computer Interaction (HCI) specialists.[2]

Operational evaluation

This is yet again a question for end users and managers. They aim to assess how cost-effective the system is. For this, they must consider factors such as end-to-end time, environmental compatibility (with software, hardware and workflow), training requirements and cost.

Comparative evaluation

This is a consideration for buyers, who must determine which system is best for their current or future needs. It calls upon internal, declarative, usability and operational methods, using common benchmarks and test data.

Possible workflow scenarios

In the previous section we presented the full range of evaluation types. Feasibility and internal evaluations can be conducted on the laboratory testbench. The remaining forms of evaluation (and thus education) are only possible in the context of some real or simulated workflow. Workflow involves clients. Ideally, clients are real; if not, their attributes and requirements must be clearly specified in order to provide a reference point for the scope and findings of any evaluation.

At a closer look, the typical translation workflow is not just a bidirectional relationship between a client and a service provider. There are more than these two players. The roles include the client, the service provider (or translation agent), the translator and the reviser. In professional, high-quality translation work, all four are normally present. This is true whether they are employed in departments or profit centres of a single company, belong to

---

[2] A good introduction for students is Hix & Hartson (1993).

separate companies, or are freelance professionals. Often they are each located at a different site and communicate using computer aids.

When computer support is also applied to the translation work itself, new roles come into play. Translation work on a translator's workbench requires a lexicographer or terminographer and a translation memory maintainer, while an MT system calls for an MT operator, a dictionary coder and possibly a translation memory maintainer. If raw MT output is not acceptable (i.e. in most cases), a reviser, or post-editor, is needed. Where the source documents are authored using a controlled language, there may also be a pre-editor whose job it is to ensure that the text does indeed conform to the prescribed syntactic and lexical rules. As with most of the activities in these workflows, the pre-editing may be carried out either manually or with computer aids. (The issue of simplified writing or controlled language for MT applications is further discussed by Pym 1990; Siemens Nixdorf 1993; Adriaens & Macken 1995; Lehrndorfer 1996; Bernth 1997; Knops & Depoortere 1998.)

Such considerations show that complex translation workflows can include up to nine or ten professionals, each with a different role and each with purpose-made computer tools at their disposal to partially automate their specific share of the overall work. An intricate pattern of workflow scenarios arises. (We have shown some ten variants in Schubert 1997.[3])

These workflow scenarios depict more than just a piece of work being sent from one of the nine or ten players to the next, undergoing another operation at each desk, with the final translation popping out at the end. Translating engages not only the piece of work and the professional but also a variety of knowledge sources essential for such an intellectually demanding activity. They are important at every step in the overall workflow, whether these are automated or not. Relevant knowledge sources include dictionaries and terminology databases (for translators, translator's workbenches and MT systems), translation memories (for translator's workbenches and MT systems) and all kinds of document repositories for reference purposes. These in turn entail information retrieval systems connected to electronic archives and libraries, as well as systems for finding, filtering, routing and summarising information. Most of these knowledge sources need to be built up and maintained prior to the translation work proper and queried during that work. Some sources, e.g. translation memories, are updated in a fully or partially automated way during translation.

The knowledge sources may be physically located at the place where the relevant activity is carried out. However, the workflow picture is much complicated by the fact that this is far from always the case. A knowledge source such as a client's proprietary terminology database may be sent on diskettes to the translator who is obliged to use it for a certain assignment. It may equally be uploaded via the Internet. Or the translator could access and view the database on-line or queries could be submitted and answered by e-mail.

---

[3] An R&D project with the objective of designing and prototyping a user interface for a remote-access translation service of the kind described by Schubert (1997) started on 1 November 1998 at Fachhochschule Flensburg in cooperation with the International Communication Europe GmbH in Rendsburg (Germany). The *Virtueller Übersetzungsdienst* project is carried out with funding from the German Ministry of Education, Science, Research and Technology (BMBF).

Similarly, updates to the database may be carried out on the spot, by real-time on-line work, or by upload and download of patches. It is in these possibilities of made-to-measure resource sharing that Internet and intranet technologies have enhanced the quality (rather than simply the speed) of today's technology-mediated translation work.

We believe that at least the more frequent of these scenarios should be analysed, taught and practically experienced in translator training.

Enacted workflow scenarios

Now we look at the ways in which, in our two institutions, we have created scenarios to exercise our students in these different types of evaluation.

Flensburg

The forms of evaluation practiced in Flensburg include both system evaluation and the evaluation of partially automated workflows. Especially in the case of students' six-month diploma projects, there are good opportunities to perform evaluations in cooperation with companies engaged in technical translation work or other forms of multilingual business communication.

Here are a few projects typical of those undertaken in recent times:[4]

- Declarative evaluations of MT systems using test texts taken from medium-sized companies. The purpose was to establish whether and under which conditions the company in question could benefit from using the system under review. (Systems evaluated since 1992: PC-Translator, Globalink, Globalink Power Translator, Personal Translator plus, T1, T1 professional, Metal, Logos).
- Evaluation of an Internet MT service with virtually no user functionality. (System evaluated: Systran.)
- Internal evaluation of an MT system, tuning it to a new text type, with access to the user functionality only. (System evaluated: Metal).
- Partially automated pre-editing of texts with a view to reducing the manual post-editing effort. This project was conducted in the workflow of an industrial translation service. (System evaluated: Metal.)
- Operational evaluation of the trade-off between manual translation and MT plus post-editing. (System evaluated: LMT.)

---

[4] With thanks to the companies and authorities who hosted the Flensburg students: AHP/ Gesellschaft für Informationsverarbeitung mbH Munkbrarup, Boehringer Ingelheim GmbH, Drägerwerk AG Lübeck, GMD/Gesellschaft für Mathematik und Datenverarbeitung mbH Sankt Augustin, GMS/Gesellschaft für multilinguale System mbH Berlin, IBM Deutschland Informationssysteme GmbH Stuttgart, International Communication Europe GmbH Rendsburg, Kraftfahrtbundesamt Flensburg (German Federal Bureau of Road Traffic), Mercedes-Benz AG Stuttgart, New Zealand Translation Centre Wellington, Siemens AG Automatisierungstechnik Erlangen, Software and Systems Engineering Ltd. Dublin, Volkswagen AG Wolfsburg.

- Design of a controlled language grammar for German as a source language in software localisation.[5] (System evaluated: T1 professional.)
- Design and implementation of a tailor-made terminology database to be used as an on-line reference tool for translators and as an upload facility for MT systems. (MT systems involved: Logos, Metal.)

Brighton

The BA in Language Studies programme at Brighton is one which offers students the opportunity to specialise in translation and MT, but it is not oriented solely to the training of translators. So some students opting for the course in MT wish simply to extend the knowledge acquired in earlier modules in Linguistics and Language Engineering; their interests may focus more on feasibility, internal and usability evaluation. Others, however, are following a parallel course in translation and are more apt to conduct declarative, operational and comparative evaluations. From next year, nearly all students in MT will have already completed a two-semester course in technical authoring and web document design which includes controlled languages and writing for an international audience.

As undergraduates, the students do not have the possibility of a work placement.[6] How, then, can they participate in workflow scenarios? We have been able to create two situations that engage students in authentic activity.

- Support for International Relations

  Like any university, Brighton is an enterprise with many actual and potential research partners in non English-speaking countries. One such context of collaboration is the INTERREG programme, which funds a variety of initiatives between partners located in Sussex and Normandie/Picardie. Incoming and outgoing partner-search enquiries, in the form of outline research proposals, were handled by the International Relations Office. Since unreasonable demands for translation, particularly from French, were being made on their limited resources, we undertook a study to see if MT could assist.

  This investigation involved a number of small-scale internal and declarative evaluations, working with International Relations staff to enhance the Systran dictionaries with an INTERREG-specific lexicon. Similarly, students worked with individual research groups to create subject-specific lexicons and assess the quality of the output using measures that the students themselves had designed. We envisage building on these modest beginnings with usability and operational evaluations. Another project is to design and test authoring guidelines to improve the translatability of the English and – if we are able to forge closer relations with partners in France – French source texts.

  A project using T1 professional built lexical resources in the domain of the European Credit Transfer System and undertook a declarative evaluation of the output.

---

[5] Some of the special features of translation work in software localisation are discussed by Freigang (1996).

[6] This possibility will exist from 1999/2000.

- Support for software localisation:

    A joint venture associates translators in Brighton with software engineers at the IUT of Bayonne in software internationalisation and localisation. A pilot project with the goal of establishing the collaborative framework resulted in the localisation of an educational shareware package and an HCI evaluation of the interface. The continuation will involve the re-design and re-implementation of the interface (mostly in Bayonne), and the production and translation of new documentation (mostly in Brighton). This second stage will be supported by Systran and Corel Catalyst.

Reflection on the experience

We have shown that it is possible to provide students with an authentic experience of using tools for document processing and MT to enhance multilingual communication. The fact that these tools are embedded in a workflow extends considerably the range of activities in which students can engage. The participation of end user in particular opens the possibility of extending evaluation into new realms. Students can elicit judgments of output quality, assess usability and measure cost-effectiveness. This authenticity is clearly easier to achieve if the students are placed in a working environment. But, with imagination, this can be done without their setting foot outside the university campus. The products generated by students may not always be of professional quality, especially in the case of undergraduates, but the benefits they gain by engaging in the process are great. We believe also that our emphasis on evaluation equips our students with generic skills that are applicable to the workflow scenarios of today and tomorrow.

Looking to the future, we are considering a scenario that we believe to be novel in the world of education, if commonplace in the commercial world: Flensburg and Brighton would undertake different parts of a translation job, sharing resources, and taking turns to manage projects.

The success of our experiences to date relies in no small part on the good will of companies and departments who have "hosted" students, and on the material support provided by some MT vendors. We hope that we can continue to enjoy and repay their confidence.

References

Adriaens, Geert & Lieve Macken (1995) "Technological Evaluation of a Controlled Language Application: Precision, Recall, and Convergence Tests for SECC" *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation* Leuven: Katholieke Universiteit Leuven, Vol. 1: 123-141

Ahrens, Helga (1997) "Der Computereinsatz in der Ausbildung von Übersetzern und Dolmetschern unter Berücksichtigung translationsrelevanter Aspekte" Eberhard Fleischmann, Wladimir Kutz, Peter A. Schmitt (eds), *Translationsdidaktik* Tübingen: Narr, 344-351

Bayer, Ruth K., Gerhard Frechen, Ingo Hohnhold, Marian Hutny, Angelika Jaeger, Vinca Meskemper, Gabriele Notzold (1993) "Rechnerunterstützte Dokumentation und

Nutzung fachtextlicher Daten am Übersetzungsarbeitsplatz (Wiesbadener Profil 1993)" *Mitteilungsblatt für Dolmetscher* und *Übersetzer* 39 [2]: 1-12

Bernth, Arendse (1997) "EasyEnglish: A Tool for Improving Document Quality" *Fifth Conference on Applied Natural Language Processing* Association for Computational Linguistics, 159-165

Bohm, Edgar (1997) "A Translation Memory System in the University Context - Practical Applications and Didactic Implications" Eberhard Fleischmann, Wladimir Kutz, Peter A. Schmitt (eds.), *Translationsdidaktik* Tübingen: Narr, 361-367

Fischer, Inge, Karl-Heinz Freigang, Felix Mayer, Uwe Reinke (1993) *Abschlußbericht (1.4.1988-31.3.1993). Modellversuch Studienkomponente Sprachdatenverarbeitung in der Übersetzer- und Dolmetscherausbildung* Saarbrücken: Universität des Saarlandes

Freigang, Karl-Heinz (1996): "Software-Lokalisierung. Ein Gegenstand übersetzungs-wissenschaftlicher Reflexion?" Angelika Lauer, Heidrun Gerzymisch-Arbogast, Johann Haller, Erich Steiner (eds.), *Übersetzungswissenschaft im Umbruch* Tübingen: Narr, 135-146

Haller, Johann (1995): "Computerlinguistik und maschinelle Übersetzung in einem Studiengang für Übersetzer und Dolmetscher" *LDV-Forum* 12 [1]: 29-34

Hix, Deborah & H. Rex Hartson (1993) *Developing User Interfaces: Ensuring Usability through Product and Process* New York/Chichester: John Wiley

King, Margaret (1997) "Evaluating Translation" Christa Hauenschild & Susanne Heizmann (eds.), *Machine Translation and Translation Theory* Berlin: Mouton de Gruyter, 251-263

Knops, Uus & Bart Depoortere (1998) "Controlled Language and Machine Translation" *CLAW'98: 2nd International Workshop on Controlled Language Applications* Carnegie Mellon University: Pittsburgh

Koordinierungsausschuß "Praxis und Lehre" des Bundesverbandes der Dolmetscher und Übersetzer e.V. (BDU) (1986) "Memorandum" *Mitteilungsblatt für Dolmetscher und Übersetzer* 32 [5]: 1-8

Lehrndorfer, Anne (1996a) *Kontrolliertes Deutsch. Linguistische und sprachpsycho-logische Leitlinien für eine (maschinell) kontrollierte Sprache in der Technischen Dokumentation* Tübingen: Narr

Lewis, Derek (1997) "MT Evaluation - Science or Art?" *Translating and the Computer 19* London: Aslib

Mitkov, R., J. Higgins-Cezza & O. Fukutomi "Towards a More Efficient Use of PC-based Machine Translation in Education" *Translating and the Computer 18* London: Aslib

Neubert, Albrecht (1996) "Computers, Dictionary Makers and Translators" Angelika Lauer, Heidrun Gerzymisch-Arbogast, Johann Haller, Erich Steiner (eds.), *Übersetzungswissenschaft im Umbruch* Tübingen: Narr, 155-168

Pym, P. J. (1990) "Pre-Editing and the Use of Simplified Writing for MT: An Engineer's Experience of Operating an MT System" Pamela Majorcas (ed.), *Translating and the Computer 10* London: Aslib, 80-96

Sager, Juan-Carlos (1989) "Quality and Standards - the Evaluation of Translations" Catriona Picken (ed.) *The Translator's Handbook* 2nd ed. London: Aslib, 91-102

Schmitt, Peter A. (1996) "Computereinsatz in der Translation" Angelika Lauer, Heidrun Gerzymisch-Arbogast, Johann Haller, Erich Steiner (eds.), *Übersetzungswissenschaft im Umbruch* Tübingen: Narr, 187-195

Schubert, Klaus (1993) "Zwischen Benutzerschulung und Wissenschaft: Sprachtechnologie in der Übersetzerausbildung" Horst P. Pütz, Johann Haller (eds.), *Sprachtechnologie: Methoden und Werkzeuge* Hildesheim/Zürich/New York: Olms, 304-311

Schubert, Klaus (1997) "Remote-Access Translation Services: Software Design with the User in Focus" *Translating and the Computer 19* London: Aslib

Siemens Nixdorf (1993) *Verständlich und übersetzungsfreundlich schreiben* München: Siemens Nixdorf Informationssysteme

Sparck Jones, Karen & Judith Galliers (1996) *Evaluating Natural Language Processing Systems: an Analysis and Review* Berlin: Springer

White, John (1997) "MT Evaluation" Tutorial given at MT *Summit VI* San Diego: Association for Machine Translation in the Americas