ACCESSING INTERNET-BASED BUSINESS COMMUNICATION SERVICES IN YOUR OWN LANGUAGE: OUTLINE OF AN R&D EFFORT FOR SMALL ENTERPRISES

Klaus Schubert

Abstract. The ALSBALT Project aims to develop a system for providing translation services, both raw and post-edited, as a value-added extension of available business communication services in the Internet and Internet-like networks. Small and medium-sized enterprises are the primary target group. The system will be composed of available modules and missing links will be developed. For the first prototype, the languages Finnish, Swedish and German are envisaged. Links to existing MT facilities for English, French, Spanish etc. will be provided as well as interfaces to modules now being developed for languages of the Baltic Sea Region.

Language as a Business Resource

The pronounced political aims of the European Union include support to

- o   small and medium-sized enterprises,
- o   peripheral regions across national boundaries,
- o   lesser-used languages,
- o   multilingual access to information,
- o   the integration of the new member-states.

A consortium has been formed to contribute to these objectives with the instruments of language technology. The ALSBALT Project[1], which is in the fund-raising phase, aims to set up an integrated translation service which will be tailored to the needs of the less frequent user. This type of user is typically found among small and medium-sized enterprises.

The Less Frequent User: A Scenario

The ALSBALT Project encompasses the scenario of small companies for whom the markets have become international.   The companies use Internet-based network services to

- o skim through calls for tenders they might be interested in,
- o look for patent information on a specific development,
- o browse through requests for business cooperation,
- o search large product vendors' point-of-sale and catalogue displays for technical information on spare parts they want to order,
- o scan point-of-information displays and user news groups for maintenance advice and technical hints to get quick help with a machine tool or a software system which they need to get up and running again within half an hour,
- o exchange informal business correspondence with partner companies,
- o send quick preliminary reactions to calls for tenders, calls for business cooperation and requests for technical help, etc. etc.

Most of these activities lie not so much at the level of the company management but are carried out by the individual employee, a salesperson, a technician, a secretary.

This is the first feature of the problem addressed by the ALSBALT Project: in international markets, business communication needs to overcome language barriers. Large companies can give their employees language training, they can afford language departments and use the services of translation agencies. In small companies, all skills need to be concentrated in a single person. Therefore, small companies in particular need a ready-to-use translation service:

Feature 1 A translation service.

The second feature is closely related: international business communication via the networks makes it necessary to bridge the language barrier instantly. Ordering a translation from a traditional translation agency is often too costly and much too slow for somebody who is looking for immediate help in an on-line service. Therefore, the translation service should be integrated into the business communication systems used in small companies:

Feature 2 A translation service integrated in on-line business communication systems.

The developments witnessed in multimedia business communication technology today suggest a scenario where the employees of small companies increasingly make use of the Internet or some closed network service which is based on an Internet-like technology without being open to the public. The latter type of network service is gaining ground as the information and services offered become more and more professional in nature. It can be expected that networks for business subscribers will be established outside the Internet as the Internet becomes more and more overcrowded and slow. A point in case is the communication networks being set up by the Chambers of Commerce for their member companies in various parts of Europe and worldwide. These are networks with tens and

hundreds of thousands of participants, each of whom may use the services only every now and then.

The system should therefore be equipped for servicing a large number of users:

Feature 3 Very high number of users.

As mentioned the ALSBALT service should be tailored to the needs and capabilities of less frequent users who cannot be especially trained for using it:

Feature 4 A translation service for the less frequent, untrained user.

The next feature is implicit in the above list of communication tasks: in the trade-off between speed and quality, the users of the ALSBALT service will often make do with a less-than-perfect translation quality. The service should therefore include the option of receiving a fast, fully automatic raw translation:

Feature 5 Fully automatic raw translation as an option.

Once small and medium-sized companies have got a translation service at their fingertips built into their Internet business communication, they are also likely to use it for purposes where a raw translation is not good enough. There is a gradual transition from the tasks of the above list to tasks like the following.

- o   Drafting and finalizing technical notes, inspection reports, release notes,
    progress reports etc. in joint projects.
- o   Posting product marketing information, technical updates, product notes etc.
        for customers, subcontractors and business contacts.
- o   Enabling remote access to computer-based product presentations.
- o   Providing and receiving computer-based training.

This gives rise to another requirement. As a sixth feature the ALSBALT Application should provide the option of ordering a high-quality translation which entails the need for the carrying out of human tasks, post-editing in particular, along with the machine translation function:

Feature 6 High-quality translation as an option.

An Integrated Translation Service Centre

A quality translation facility at the service of a large community of less frequent users is in itself an enterprise which needs to be maintained by professionals. The ALSBALT Project therefore envisages the setting up of an Integrated Translation Service Centre to provide the services offered and maintain the systems needed.

Having derived features for the envisaged application directly from user requirements thus far, one should take a closer look at the type of service that could meet the requirements, the type of system that could underlie the service and the corollaries for the underlying application which derive indirectly from the identified requirements and features.

A quick raw translation functionality obviously calls for machine translation. The integration into business communication networks requires facilities for making the machine translation system(s) accessible on an Internet-type server with full job management facilities:

Feature 7 An Internet server with full job management facilities.

The ALSBALT Project does not aim at an Internet server open to the public for free (cf. the Globalink or the CompuServe experience: Berberich [1], Flanagan [2]) but at an economically viable service to be provided to paying customers. The application will therefore need an advanced accounting and billing facility along with a functionality to ensure access rights to the service as such an to individual resources such as company dictionaries or customized translation memories (see Features 9 and 10 below):

Feature 8 Advanced accounting and billing facilities.

While all this may be automated to a very high degree, feature 6 (and features 9 and 10, below) shows that it would be insufficient simply to link an unattended translation server to the Internet and let the users cope with the output it generates. The step from raw to post-edited output should always be possible.

However, human interference is needed not only for post-editing (plus the inevitable tasks of technical system maintenance, troubleshooting etc.). The experience of professional use of machine translation systems shows that there are at least two major tasks to be carried out by skilled professionals in the Translation Centre rather than by the users.

The first of these tasks is dictionary coding and management. Machine translation experience shows that virtually every text contains new words, i.e. words not previously entered into the dictionaries of the system or composed of elements unknown to the system. Every single machine translation task therefore comprises the necessity of coding the unknown words with their grammatical features into the dictionaries of the system. In the more advanced machine translation systems for the professional user, the grammatical

information which must be provided with a dictionary entry can reach quite far not only into the structural transfer mechanism but also into the decision criteria and mechanisms of lexical transfer (Schubert [3]). The task is so decisive for the quality of the output that the Logos Corporation for instance have reserved the task of coding the grammatically central entries (the verbs) to their central development team, so denying both their customers and even their own European staff the possibility to enter the grammatical information for a new verb entry (Logos [4]). Other manufacturers do provide their customers with this possibility, supplying extensive documentation for the purpose (e.g. Metal [5]). In the ALSBALT scenario, dictionary coding is therefore a task to be reserved for very well trained machine translation professionals.

Lexical transfer in machine translation relies very much on a structured, often hierarchic, portioning of the vocabulary. The systems normally allow for a structuring into domains or subject areas, used with some success for circumventing decisions which would otherwise be of a semantic nature. Along with this structuring there is often a possibility for the user to create special dictionary compartments for the terminology used by individual companies, manufacturers, customers or the like. A network service provider like the ALSBALT Translation Centre will have to make use of this facility to a large extent in order to keep their customers' vocabularies apart. This is needed to ensure a high output quality, especially in lexical transfer. It may well become an especially delicate task when the still modestly budding ideas of claiming copyright on one's terminology files gain ground:

> **Feature 9** Dictionary coding and management.

Machine translation quality can be greatly improved by incorporating the implicit knowledge applied by the post-editing translator into the system. The manufacturers have therefore been working for quite some time at solutions which integrate a translation memory into a machine translation system. The ALSBALT Centre should make use of this technology. In much the same way as for dictionaries, the Centre will have to set up efficient management procedures, both automated in the job management and manual in the Centre's workflow organization, to make sure that the right translation memory is used with the right translation job. As for dictionaries, this facility is needed both for ensuring a high output quality and for copyright reasons[3]:

> **Feature 10** Translation memory management.

The Baltic Sea Region ... and Beyond

The ALSBALT Project aims in the first place to develop a prototype with the facilities needed for an Integrated Translation Service Centre of the type outlined above. The idea for the project derives from two organizations engaged in purveying technological facilities to

small and medium-sized enterprises at the regional level[2]. For the prototype phase, the Baltic Sea Region has been chosen as the first field in which to test the application. To this end, the project will address the languages Finnish, Swedish and German and open up connections to further languages both within the Region, viz. Danish, Russian, Polish, Estonian, Latvian, Lithuanian etc., and beyond the Baltic Sea Region, such as English, French, Spanish etc.

The prototype will be composed of available resources as far as possible. The basic technology will be taken from two existing machine translation systems, T1 and KIELIKONE. T1 was recently developed by the German Gesellschaft fiir multilinguale Systeme GmbH (GMS) out of the better known METAL system of the Siemens AG and some of its subsidiaries. KIELIKONE is a development of the Finnish Kielikone Oy. In addition to these machine translation systems and their Finnish and German modules, ALSBALT will make use of existing modules for Swedish, viz. the University of Gothenburg's GULD database of lexical data and the Finnish Lingsoft Oy's Constraint Grammar modules for Swedish morphological and word-formational analysis (Karlsson et al. [6]) to be further developed within the Project by the University of Helsinki.

The language pairs chosen for ALSBALT's first prototype development break new ground in connecting the languages of the two new member states of the European Union to one of the major languages of the Union.

In a second phase, ALSBALT will make use of the language pairs and modules available from each of the two machine translation systems involved, i.e. KIELIKONE and T1, possibly along with the other derivative of the former METAL system, the LANTMARK system being developed by the Belgian LANT N.V.[4]

When efforts to develop vendor-independent user interfaces to various machine translation systems and translation tools make progress, the ALSBALT Project will in a third phase be in a position to link up to language pairs and modules provided by other manufacturers as well. This will open up a broader range of facilities to the users of the ALSBALT on-line service.

Notes

1   ALSBALT stands for "Advanced Language Services in Business Tele-communication around the Baltic Sea".

2   Technologie-Transfer-Zentrale Schleswig-Holstein GmbH, Kiel, Germany, and Nelikon Oy, Vaasa, Finland. The project consortium consists of these two, the developers mentioned and a group of test users and validation sites. It is coordinated by the Technologie-Transfer-Zentrale and the Fachhochschule Flensburg.

3   Obviously the list of features in this contribution is far from complete.

4   Kielikone have an available MT system for English-Finnish. METAL modules exist for German-English, English-German, German-Spanish, English-Spanish, German-Danish, French-English, French-Dutch, Dutch-French and developments are underway for German-French, Russian-German, German-Russian, Spanish-English, English-Catalan, English-French. There will be no

further development of METAL as such, but GMS have developed T1 Standard modules for German-English and English-German and will market modules for English-Spanish, Spanish-English and Russian-German and other language pairs, while LANT N.V. aim to develop their LANTMARK system for German-English, English-German, French-English, English-French, German-French, French-Dutch, Dutch-French and English-Spanish.

References

1   Berberich, F., 1996, Die digitalen Dolmetscher bieten ihre Dienste mittlerweile in Echtzeit an. *Computer-Zeitung* 18 January 1996, 8

2   Flanagan, M., 1995, MT in the Online Environment. MT Summit V. [Luxembourg: Commission of the European Community, unpublished conference proceedings]

3   Schubert, K., forthc., Lexikon und grammatischer Regelmechanismus in der maschinellen Übersetzung. R. Walczak (ed.), *Forschungsbericht 1995/96.* Flensburg: Fachhochschule Flensburg

4   Logos, 1995, Logos Version 7.7. Alex. User's Guide. Mt. Arlington: Logos Corporation, 119

5   Metal, 1993, METAL/X-Codierung English-Deutsch V2.0. Benutzerhandbuch. München: Sietec, 139-232

6   Karlsson, F., Voutilainen, A., Heikkilä, J. and Anttila, A. (eds.), 1995, *Constraint Grammar.* Berlin/New York: Mouton de Gruyter

Author

Klaus Schubert is professor of computational linguistics and technical translation at Fachhochschule Flensburg.

Address: Studiengang Technikübersetzen, Fachhochschule Flensburg, Am Bundesbahnhof 1, D-24937 Flensburg, Germany.

E-Mail: schubert@fh-flensburg.de