Computer-assisted multilingual e-communication in a variety of application areas

Adriane Rinsche
The Language Technology Centre Ltd (LTC)
5-7 Kingston Hill
Kingston upon Thames
Surrey KT2 7PW
Tel. +44 20 8549 2359
Email: ARinsche@langtech.co.uk

Abstract

The paper describes the architecture and functionality of LTC Communicator, a software product from the Language Technology Centre Ltd, which offers an innovative and cost-effective response to the growing need for multilingual web based communication in various user contexts. LTC Communicator was originally developed to support software vendors operating in international markets facing the need to offer web based multilingual support to diverse customers in a variety of countries, where end users may not speak the same language as the helpdesk. This is followed by a short description of several additional application areas of this software for which LTC has received EU funding: The AMBIENT project carries out a market validation for multilingual and multimodal eLearning for business and innovation management, the EUCAM project tests multilingual eLearning in the automotive industry, including a major car manufacturer and the German and European Metal Workers Associations, and the ALADDIN project provides a mobile multilingual environment for tour guides, interacting between tour operators and tourists, with the objective of optimising their travel experience. Finally, a case study of multilingual email exchange in conjunction with web based product sales is described.

1 Multilingual help desk facilities

Until recently, there were basically three possible strategies for multilingual support:

- Maintaining a separate support centre in each target country; this is very expensive, and only really feasible for a large organisation with an established international presence;
- Running a centralised help desk employing multilingual support staff; also potentially expensive (people with the right technical and linguistic qualifications are not easy to find, and user numbers may not justify experts in every language to be covered); a variant of this is the distributed 'virtual help desk' with remote support staff connected to a central database;
- Using telephone interpreting services to offer multilingual support from a central location; expensive on a case-by-case basis, but perhaps justified where the volume of multilingual support is small and the demand sporadic.

LTC-Communicator now offers a fourth possibility:

- Using translation software to incorporate multilingual capability into the help desk software itself, and integrating with web-based workflow and knowledge base facilities – reducing the dependence on human experts to manage communication.

▪ 1.1 Business case:

LTC-Communicator allows a software vendor (or any service organisation) to run a centralised support centre without the expense of training and equipping local support desks in different countries, or employing multilingual staff; the end-user and the support engineer do not need to share a common language.

For example, for an English-speaking help desk supporting German users:

- The user enters a trouble ticket in German, via the software company's support portal;
- The trouble ticket is routed through the LTC-Communicator translation components;
- The request is displayed to the support engineer in English;
- The engineer prepares the solution, also in English, to be automatically routed back through the translation environment;
- The user can then view the solution (or status information) in German.

An attraction of this solution is that it can integrate with any existing web-based *e*-communication solution that is capable of communicating via XML and can be configured to provide flexible workflow; incoming messages in the company's 'native' language are passed directly into the normal communication workflow, while 'foreign' messages are routed through the LTC-Communicator automated translation environment. Similarly, response messages may be directed straight to the requesting user or translated first, as applicable.

LTC-Communicator can also be integrated with established knowledge base functionality, supporting user 'self-help' with multilingual query facilities against existing documentation and FAQs, and continuously enhancing the content of the knowledge base with new queries and solutions.

The overall workflow is shown below:

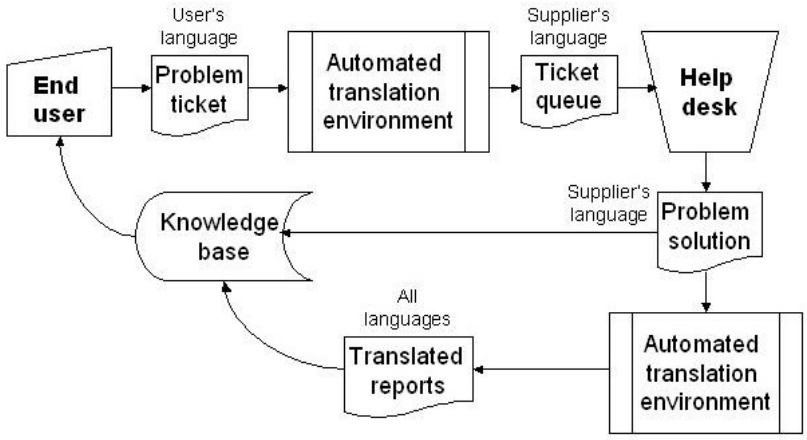

Figure 1: LTC Communicator helpdesk workflow

1.2 Technical architecture:

At the technical level, LTC-Communicator itself comprises several components:

- A *translation memory* populated with relevant bilingual material;
- An interface to a *machine translation* program (e.g. SYSTRAN) including a custom dictionary holding key terms relevant to the tourist industry;
- A *workflow server* responsible for workflow management consisting of the following components:

  - an automation server which interfaces external translation services with the translation bus[1],
  - a web server which receives the translation request wrapped in XML messages,
  - an optional post-editing service which routes machine translated text to available human revisers, if high quality and publishable output is required as opposed to understandable output.

[1] A transmission channel through which data is carried from one automation server to another.

Together, these function as a 'black box', with input and output in XML format; the only real changes that have to be made to an existing application to integrate with LTC-Communicator are:

- To generate (and read) XML messages in the specified format;
- To tag the messages with meta information such as source and target languages, quality required, speed required, to route them correctly through the automated translation environment.

The LTC Communicator architecture is shown schematically below:

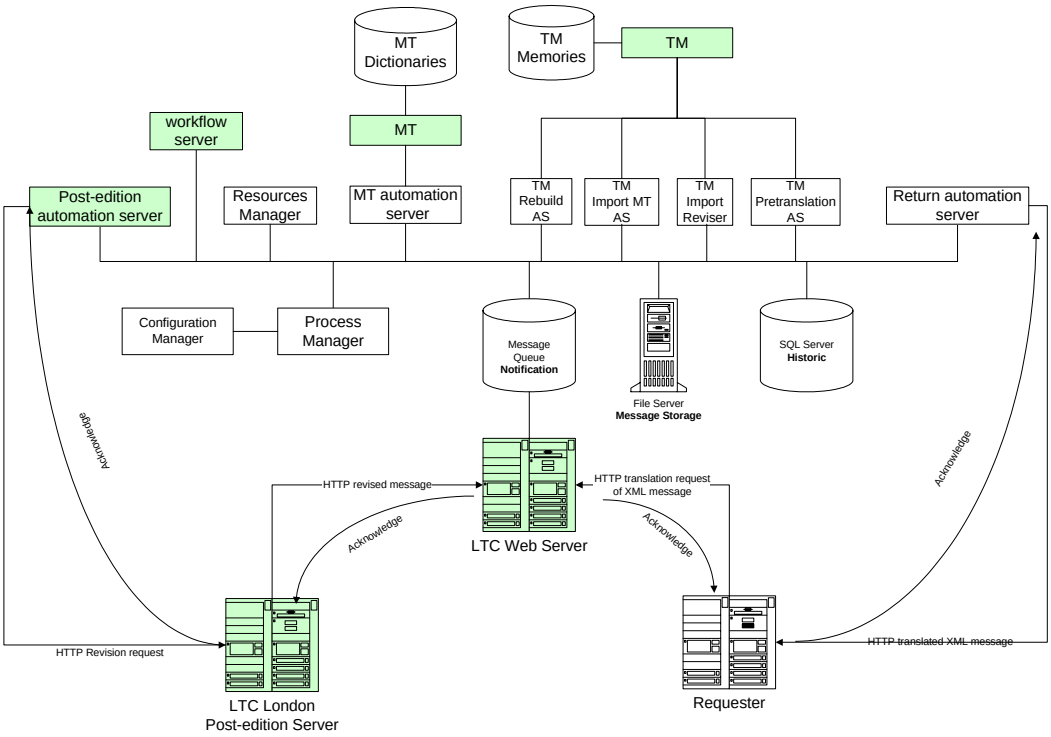

Figure 2: LTC Communicator Architecture

In principle, the system can handle any language combination, limited only by the capabilities of the machine translation software available; between them, the existing MT systems cover all the major European languages, and increasing numbers of non-European languages also.

Once the system is set up, it is relatively easy to add further languages, with the workflow server ensuring that each incoming message is directed to the appropriate MT module within the translation environment. A single help desk can then support users in many different countries at little additional cost.

1.3 Use of machine translation:

Machine translation is particularly well-suited to help desk and other e-communciation requirements, where:

- Output needs to be intelligible, and not generally of 'publishable' quality;
- The subject area is usually quite restricted (e.g. a range of software products), and custom dictionaries delivered with the MT software can be tailored to the specific 'domain';
- Existing material can be re-used; databases populated with bilingual documents, localised software, existing FAQs etc. are used by the translation memory;

- Similarly, each 'new' solution is potentially re-usable; it can be added to existing FAQs, or to a multilingual knowledge base, enabling users to resolve an increasing number of queries without calling the help desk at all.

Moreover, the LTC-Communicator architecture allows results to be optimised in a number of ways:

- Input can be structured by providing pick lists for key items such as application/version, type of fault, severity etc., producing more matches against the translation memory, and minimising the amount of free text to be translated;
- Style guidelines can be applied to produce more consistent input, enforcing standard use of terminology and avoiding complex or ambiguous constructions that the MT software may not resolve correctly;
- Spell checkers can also be used to improve input quality, further increasing potential 'hit rate' against the TM;
- Where output of 'publishable' quality is required (e.g. to populate a multilingual knowledge base, or to be incorporated in formal documentation), a post-editing option is available; post-editing is also recommended when the system is first installed, to help 'tune' the TM and MT dictionaries;
- Post-edited output is automatically fed back into the TM.

2 Multilingual eLearning

2.1 The AMBIENT Project (eTen 510749) is a market validation study, funded at 50% by the European Union. The consortium consists of technology partners and private and public training institutions from various European countries. The project started in August 2004.

The objective of the AMBIENT LEARNING project is to provide a pragmatic, easy-to-use eLearning web service, which allows access any time, any where and any how to personalised, high quality learning content. The AMBIENT learning service is based on stable and mature technology and offers ambient, multimodal, multilingual, personalised and context-sensitive access to learning at work, at home, at a training institution or on the move. The main purpose of the project is the market validation of the AMBIENT learning service by demonstrating the various services in several European regions and therefore preparing the ground for successful market development.

LTC proposes to adapt the LTC Communicator to include multilinguality in an eLearning environment as proposed in the AMBIENT project as follows: Users can request eLearning material via the personalised user interface provided by one of the technology partners and will be notified which material is available in which languages, or will be translated as part of the current request.

The following aspects are relevant:

1. Multilingual keyword search**:** In order to be able to return information about materials available in other languages, the current structure of the context manager and the semantic tagging applied needs to be adapted in order to provide an appropriate multilingual key word search.

2. Translation workflow and output quality: It is to be decided in a user trial during the course of the project which of the following possible workflows will be most suitable:

A. User receives raw translated document within a couple of minutes

B. User orders documents that need translating, these will be machine translated, routed to the human post-editing services and returned in good quality. The revised language strings will be saved in the translation memory. The next time the same document is requested in the same target language, the translated version will be immediately available, either directly from the knowledge database because multilingual content is stored there after high quality translation, or via the automated translation process where the translation memory contains all translated strings at high quality and replaces the source strings by the correct target strings available in the translation

memory rather than routing the document through the machine translation system.

The disadvantage of this high quality approach lies in the fact that there will be a cost for the post-editing and (depending on the size of the document) a certain delay for post-editing the material that has never been translated before. The advantage is obviously a high quality output and increasing speed and quality over time, for instance when returning previously translated documents, that have been updated since the last translation request, as most of the strings are already in the translation memory or knowledge database, and only the updated parts would have to be machine translated and revised.

Another interesting aspect is also the question whether raw translated material would be acceptable, if the user requests speech output, for instance to listen to learning content while driving to a remote destination. Whilst the user interface including access to LTC Communicator has been implemented, these trials are about to start.

2.2 The EUCAM project:

The EUCAM project (eContent EDC 22238) examines multilinguality in eLearning for the automotive industry, with international partners from several European countries and Daimler Chrysler as co-ordinator. DC developed a production learning system which now needs to be localised and used in various countries with customised content depending on the products manufactured. The German IG Metall and European Metal Workers Associations have been involved from the very beginning, as learning by doing and access to learning content via terminals at the workplace needs to be acceptable to the workers as opposed to more traditional learning methods off site via training institutions. The project aims at building a technical communication and multilingual leaning infrastructure by

a) optimising horizontal technical communication between manufacturers, suppliers, and engineers – especially by integrating small suppliers into the learning environment and

b) optimising vertical technical communication between engineering, production planning, quality management and the shop floor level.

LTC will provide a localisation strategy for the user interface of the production learning system. Besides, interactive features within the system, allowing users to add content and managers to interact with their (possibly international and multilingual) team members may require the use of machine translation via LTC-Communicator. The use of highly restricted language (although not via controlled language tools) may suggest that MT could be an option, including human post-editing for quality control where necessary. A decision regarding the use of tools will be made after a detailed evaluation of the required content.

3 Multilingual destination management in the tourist industry

The ALADDIN project (CRAFT 017566) is a co-operative research effort of 5 SMEs, of which three are technology partners and the other two users providing a test bed for the completed prototype. Several research partner institutions will provide technology and tourism market knowledge. The academic partners will provide valuable input in terms of system requirements and evaluation.
From the consortium's perspective, an optimal mobile destination management is a combination of four main components:

1. The mobile workspace for incoming agencies and tour operators
2. The mobile leisure environment for the tourist
3. The seamless integration between both workspace and leisure environment for the advantage of the tourist (providing personal characterization data)
4. A multilingual access facility in real time or near real time.

All components are faced with different technical challenges which the consortium defines as the objectives which shall be solved through the partnership of the universities, the SME technology providers and the users.

After the usual lengthy contract preparation phase, the ALADDIN project is now about to start. The project is of considerable relevance, as European SMEs in the tourism and travel industry in Europe are facing serious challenges from large international companies. Incoming tour operators, tourist offices, restaurants, hotels, museums and comparable establishments are typically small or medium enterprises which need to provide different services for the same customer – the tourist – in a cost efficient but attractive way.

The multilingual requirements in this industry are quite obvious. First of all, within the scope of the ALADDIN project, tourists and incoming tour operators can be nationals of any European country, interacting with nationals of any other country. Therefore, the user interface of the planned mobile workspace needs to be localized into a number of relevant languages. Secondly, content relevant to tourists needs to be made available either on the fly or with minor delay in their own language.

By including necessary transnational cooperation aspects in the tourism area, by perfecting tourist relationship management by making CRM features mobile and furthermore by including localised content through the use of LTC Communicator's web based translation service, ALADDIN will lead to a sustainable enhancement of the competitiveness of European SMEs in the tourism area.

4 Multilingual email exchange

One of LTC's long standing clients sells golf equipment via the internet. LTC localised their web site into several languages, and this led to email enquiries in all the languages in which the website was made available. The initial workflow was such that emails arrived at the customer's service in-box and needed to be forwarded to LTC's service team for translation. The translation was then returned to the customer who prepared the answer for the client, which was then returned to LTC for translation into the user's language. The translation was returned to the customer who forwarded the text to the original requester. This is obviously a very awkward and time consuming administrative exercise, particularly because such emails tend to be quite short.

By implementing LTC Communicator's automatic email translation facility with a custom dictionary of golf terminology, which is integrated with Microsoft Outlook, most of the manual email handling and human translation can be eliminated. If the quality of the machine translation is not sufficient for whatever reason, it is forwarded to a reviser by a simple mouse click. The human revision is done within a few seconds, and the email is then automatically forwarded back to the requester (which from LTC's point of view can either be the seller or the potential buyer of the products concerned).

5 User Feedback

Users of the original helpdesk application reported overall understandable content and acceptable delays when requesting human post-editing. The other applications described above are being implemented and user feedback is not yet available.

6 Conclusion

The awesome growth of the Internet and increasing use of mobile technology means that more and more applications are becoming truly global in reach, and companies have to operate in an increasingly international - and multilingual - environment. The demand for instant translation (on demand) threatens to outstrip the capacity of human translators, who are an expensive resource in any case.

A highly efficient workflow implementation automating as many steps as possible already saves time and resources, combined with machine translation, is part of the answer. LTC-Communicator offers an innovative combination of technologies to overcome these linguistic limitations whilst at the same time minimising the reliance on human translators.

Although the focus so far has been on help desk systems, the LTC-Communicator would fit into any workflow application with a need to support users in more than one

language; it could also be used to add multilingual capability to technical search engines and knowledge base products, both of which play an increasingly central role in the corporate IT environment.

7 About LTC:

The Language Technology Centre is a limited company based in the UK, specialising in building multilingual websites, software localisation, consultancy in language technology, technical translation and software development. Clients include telecommunications companies, software developers, mechanical engineering businesses, the medical and automotive industries as well as European and international institutions.