

LTC Communicator

Philippe NOZAY and Adriane RINSCHÉ

Language Technology Centre Ltd
Kingston upon Thames, UK
pnozay@langtech.co.uk
arinsche@langtech.co.uk

Abstract

Combining machine translation (MT), translation memory (TM), XML, and an automation server, the LTC Communicator enables help desk systems to handle multilingual data by providing automatic translation on the fly. The system has been designed to deliver machine-translated questions/answers (trouble tickets/solutions) at an intelligible level. The modular architecture combining automation servers and workflow management gives flexibility and reliability to the overall system. The web server architecture allows remote access and easy integration with existing help desk systems. A trial was funded within the framework of the EU project IMPACT.

1 Multilingual applications

1.1 Necessity

With the awesome growth of the Internet, more and more applications and websites are becoming truly global and have to handle input in several languages. They have to provide some sort of translation to enable people to communicate whatever their mother tongue.

1.2 Issues

Machine translation is the answer to instant translation; however, the output quality is often low and very often difficult to understand, if not misleading. Human translation provides excellent output but it very slow and expensive. Other tools in the translation industry, such as translation memory, provide very good output but need human interaction.

Most websites/ applications would like to provide instant, understandable and cheap translation, but this is hard to achieve unless a machine translation system has been tuned to a well defined domain and integrated into a set of tools automating the interaction between (in our case) users of a software product with support issues entered in one language and solution engineers providing solutions in another language.

1.3 LTC Communicator

We have developed the LTC Communicator to provide a practical solution to these issues. It offers multilingualism to any application or website that requires automatic translation at an intelligible level, using a combination of customised tools in a well defined, automated sequence. This system is designed as a web service accepting XML requests and returning translations in XML messages.

LTC Communicator is a product that includes the automation server and the LTC translation memory. Systran or any other MT system can be plugged in, depending on customer preference or the language combinations required. LTC Communicator can be purchased outright or run by LTC as a hosted service.

1.4 Fields of application

In order to provide “intelligible translation” the system has been tailored to technical questions/answers (trouble tickets/solutions). It is at its best in technical support applications such as a web help desk where a user posts his question in his own language (e.g. English); the LTC Communicator then translates the question for the engineer to read it in his language (e.g. German). When the engineer posts the response in German, LTC Communicator translates it into English for the user to read.

The system can also benefit technical search engines, where the user enters search criteria in

his own language; the LTC Communicator then translates the criteria into multiple languages, and the search engine queries its database for any of the translated criteria. If matches are found, the summary of the document produced by the search engine can be translated back into the user's language for him to judge the relevance of the result.

LTC Communicator could be used in any application needing to provide multilingual communication between users. To ensure reasonable output, the vocabulary should be restricted to specific domains if possible.

1.5 System background

The LTC Communicator emerged from the EU project IMPACT, which was very successful in terms of system design and output quality in the well-defined software domain selected (Helpdesk application for a CRM product).

The full name of the project was "Impact of a web based multilingual portal for helpdesk applications", and it was a trial project within the European "User-friendly Information Society" programme (project number IST-2000-30110); it was funded by the European Commission as part of the 5th Framework Programme (under "Human Language Technologies"). CAS Software AG in Germany cooperated with the Language Technology Centre (London), Infoworld Srl (Milan), a CAS sales partner for Italy, and the systems house New Emphasis (Greece).

IMPACT links a helpdesk based on the CAS workflow product genesisWorld via the Internet to a multilingual web service developed by LTC using Systran MT. A customer can report a problem via the web, entering the information in English. The trouble ticket is automatically translated into German before it is viewed by the helpdesk operative. After processing, the helpdesk agent's solution, which is in German, is automatically translated back into English in the same way and is then available for the customer in a protected area of the web pages. The customer can also find information there on the current status of the problem they reported. The customer additionally has access to a multilingual knowledge database (FAQs etc.) here.

Based on its genesisWorld system, CAS Software operates a helpdesk for processing problems reported by the international partners for genesisWorld. Via the web, the partners can initially search for existing solutions and report problems if required. These are processed in the CAS in-house genesisWorld system, following a structured workflow.

In IMPACT, genesisWorld exchanges the content of problem reports with the LTC system for automatic translation, using an interface based on Internet protocols (HTTP, XML). The LTC system processes the language entry and compares it against existing translations.

The tools used guarantee a linguistic quality that makes everything easily understandable to the user and ensures efficient communication. They are also considerably different in terms of architecture and complexity from the consumer web based translation services, the results of which are often unsuitable for professional users. For users who require a translation of publishable quality rather than just an intelligible result, the system can be temporarily or permanently set to include revision by professional, human translators.

If completed solutions are to be stored permanently in the solution database, they are translated from German into English by the LTC software and manually revised to yield publication quality, and they are then available in both languages.

The trial was carried out in English and German, and the LTC modules were developed into the product now called LTC Communicator, which can cover many language combinations.

2 System Architecture

2.1 Overview

To meet the contradictory requirements of speed and quality, the system makes use of a variety of technologies:

- Machine translation (MT) is used to provide instant translation.
- Translation memory (TM) is used to improve the quality of repetitive text.
- A spell checker can be used to improve the input quality, and improve the chance of TM and MT matches.

- Custom dictionaries for the MT are built to cover the specific domains of application.
- Databases populated with bilingual documents on the domain, software, existing FAQs etc. are used by the TM.
- Input guidance and validation is used to limit the amount of free text entered by the user and also to reduce the grammatical complexity of the input. The web interface is structured so that the user needs to provide basic information about the version of the software he/she is using, the operating system used etc, in drop down lists. In addition, style guidelines were developed for solution engineers and system integrators on the user side to avoid complex syntactic and grammatical structures that are difficult for a machine translation system to handle.
- Optional human revision where publishable quality is required. The raw translation is automatically forwarded to a reviser for post-editing; recommended when the system is first introduced to tune the TM memories and the MT dictionaries, but also if the application/website needs to publish the information in many languages to a wide audience.
- The post-edited MT output is automatically fed back into the TM.

The machine translation system used in the trial was SYSTRAN; the TM system used is an in-house development by LTC, used for batch processing only, and replacing identical matches with target text. Interactive TM processing is not desirable within this workflow as the system automatically processes new text with no TM matches via the MT system. All human post-editing results are fed back into the TM.

A Systran user dictionary was customized with the all the key software terms for the CRM system, and the translation memory was populated with bilingual online help files and software documentation, and the content of a bilingual FAQ database.

The output quality was monitored initially by human post-editors, who added entries to the Systran custom dictionary and populated the TM database with post-edited material.

2.2 Workflow

Taking as an example 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 software;
- 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.

In addition, the user can set the system to route the translated request and/or solution to an optional human post-editing service, either by default or based on some other criteria such as every 5th message or messages with low translation quality.

Depending on the request parameters (priority and revision level), a fully automated or a semi-automated workflow is instantiated. The workflow server decides on the next step between each process. The system has been optimized so that each message took only a few seconds to translate in the trial (without post-editing).

Our translation service does not use SOAP as a protocol but instead uses XML over the standard HTTP Post protocol. It was necessary to convert the XML messages to RTF format as the version of Systran on which the trial was carried out did not accept XML input.

2.3 Interface

2.3.1 HTTP Protocol

To access the translation system, the HTTP protocol is used. The application server sends (posts) the text to be translated to the LTC Communicator server through a web server. HTTP allows for:

- Distributed processes: the LTC web server can be hosted anywhere in the world.
- Concurrency: the LTC web server can receive many messages simultaneously.
- Workload balancing: multiple LTC web servers can be grouped in clusters to handle more concurrent messages.
- Implementation simplicity: HTTP has existed for years and it is very easy to

implement web servers and to generate HTTP invocation (POST) from clients.

After translation, the translated text needs to be returned from the LTC automation server to the application server. For the same reasons, we decided to use HTTP invocation to send back the translation.

2.3.2 XML message

The message containing the text to be translated and posted to the web server is in XML format. XML allows for easy exchange of structured information between applications. With XML, it is possible to add tag information describing the data contained in the message avoiding confusion as with text files. New tags can be added in future versions without making older applications incompatible. XML parsers are now widely available and XML is therefore relatively easy to implement.

2.4 Technical Advantages

In addition to the translation requirement, the system has the following advantages:

- Scalability: possible to add/remove servers according to the load.
- Load balancing: automatically balance the load between several servers.
- Easy support: quick to pinpoint any errors.

- Modifiable workflow: possible to add new steps or replace existing ones, e.g. MT.

3 Conclusion

LTC Communicator provides an easy solution to enable any website/application to handle multilingual communication between users by offering:

- An automated support environment for many web-based multilingual applications.
- Ability to plug in any MT system, depending on customer preference and language combination.
- High quality output due to customisation in well defined domain.
- Automatic handling of otherwise time-consuming management of typically short and informal communication.
- Intelligible output quality; no need for publishable quality, therefore post-editing not needed for most question/answer.
- Style guidelines for authoring to further improve quality of target messages.

In the help desk industry, it allows software publishers to run a support centre from headquarters rather than training and paying many support engineers in the different countries or hiring (expensive) multilingual support engineers centrally.