

"... to boldly go ..."

Daniel Grasmick, SAP AGⁱ and Chris Pyne, S&D/ICE GmbHⁱⁱ

Description of paper

This brief paper focuses on one area of SAP's ongoing implementation and enhancement of language processing tools. It examines the concept of outsourcing / distributing technology and services to a partner company. Finally it looks beyond the current long-term project to ambitious new avenues of language services to be explored based on experience gained to date.

Initial idea

SAP is the 4th largest software vendor in the world. It has some 7,500 customers in over 85 countries and employs over 11,000 people world-wide. Significantly, its head office is not in the United States but in Germany, and of its group sales total of \$ 2.4 billion for 1996, 75% was generated outside of Germany. The company therefore has a logical policy of producing information in both German and English. SAP has a comparatively large pool of salaried translators, especially for German to English translations.

As early as 1982, SAP began to investigate whether machine translation could work successfully for its documentation types. Although the first attempt was not entirely successful, a second project, this time using the Siemens METAL system in 1990 did succeed, and production began in 1991.

To date some 70,000 terminology "transfer" entries have been coded for German->English and about 20 million German source words have been translated for document types ranging from Help files and release notes to implementation guides.

"Hinweise" aka "Notes"

A fourth document type was somewhat different to those which continue to be handled by SAP's Multilingual Technology Group (MLT Group). These are the so-called "Hinweise" or "Notes" - comparatively small documents, ranging from a few lines to a few pages. They can be written by any of SAP's army of R/3 software developers and thus usually first appear in German. They are written as a response to a customer problem, as a bug fix, a workaround, update information etc. Since SAP is not only expanding its customer base at a phenomenal rate, but is also dramatically expanding the range of its software into new business segments, the sheer volume of information is increasing concurrently.

These "pieces" of information are only useful if they are available to SAP customers very quickly once they have been written and are updated to reflect the latest information on a subject. The majority of SAP customers will be depending on the English translation of the original German "Hinweis".

The project goals were:

1. Translate all incoming German "Hinweise" into English within 24 hours
2. Maintain as near to 100% terminological consistency within the specific area of the R/3 system (Treasury, Oil, Hospital, Logistics, Human Resources ...)
3. Be able to handle high and fluctuating demand levels

High volume, low, fast turnaround, a case for language tools? And if so, which? In 1994, the language tools market was very "binary" - either one used a "translation memory system" or a "machine translation system" - it was as if the one negated the use of the other.

SAP began by researching the repetitivity in the "notes" corpora that it already had. This revealed a 1% level of repetitivity/redundancy - reflecting the bandwidth of input which these "notes" contained. This suggested that just translation memory would be unsuccessful.

The idea would be to use machine translation on the first translation run, leveraging the well defined R/3 terminology base already coded into METAL by the MultiLingual Technology Group.

Once translated by METAL, translators with a grounding in R/3 and an awareness of text and terminology issues would perform postediting, archiving a first bilingual version of the respective "note".

Many "notes" have a history. Once they have been created, machine translated, post-edited and archived, new information is entered by the author, and a memory cycle is initiated. Here, of course, the level of "legacy" repetition is very high and the lead tool is now translation memory. The translation memory tool excludes 100% text string matches and hands off the new text (the "delta" text) to the MT system.

The MT system translates the delta text, this is reintegrated in the bilingual "notes" file and the newly translated passages are flagged for the translator to review and edit.

METAL system "client-server"

The METAL system runs on a Sun workstation. True clients did not really exist for the METAL system; all the server required to perform the translation was the file to be processed and a parameter file telling the server how to process it. Working with Sietec, the division of Siemens Nixdorf responsible for METAL at that time, SAP specified changes to the software periphery so that a group of translators using X windows clients could download "Hinweise" from the R/3 system, process them and upload them for release.

The result was a hybrid MT/TM system with multi-user document submission, and master control of translation jobs in process.

Outsourcing/distribution?

The "Hinweis" in-house solution was proving itself to be successful. However, the range of new linguistic challenges which was confronting the MLT Group and the sheer volume of the "Hinweis" project suggested that now that model had been defined and tested the project could be outsourced.

It proved to be an advantage that the MLT Group had its own translation specialists. These people were focussed on the issues relating to Hinweise and could oversee the project transition from in-house to outsourced solution.

SAP's outsourcing partner is the firm S&D/ICE, based in Rendsburg in North Germany. This translation and localisation company became involved in machine translation in 1988 with the purchase of the Siemens METAL system running on proprietary hardware. The company had already performed machine translation projects for various industrial customers from the defence, aeronautics, chemicals and automobile industries. Unlike many translation companies, this company has a high level of staff translators for its main translation languages, German and English all of which operate from one central base. Since SAP was looking for a long-term commitment from its partner to offset the necessary training and logistics time and outlay during project handover, S&D/ICE looked to be a sensible choice.

The logistics

The model foresaw a gradual transition of know-how and technology from the SAP headquarters in Walldorf in Southern Germany to Rendsburg in the North.

The first stage was to establish a double ISDN (128 kB/s) nailed line to enable data transfer and allow S&D/ICE access to R/3 servers located in Walldorf. This was vital since sometimes process and transactions have to be researched during post-editing by the translator and some "Hinweise" that had not previously been translated by the MLT Group needed to be updated "online" in the R/3 system itself.

Initial training followed in Walldorf, both theoretical and "learning by doing", enabling the Rendsburg translators to get the hang of post-editing MT output.

Several months into the project came the next big step; the replication of the METAL database generated by SAP since 1991 and its installation onto a METAL server at S&D/ICE. Parallel with this, the "Hinweis" archive comprising some 600,000 files was also installed in Rendsburg.

The programs covering master job control and job downloading and uploading were rewritten to enable tasks to be performed flexibly either from Walldorf or Rendsburg, making it possible to selectively release certain "Hinweis" areas for translation in Rendsburg, increasing this as confidence and know-how grew.

METAL has a terminology database structure which enables new terminology out of a terminology coding session to be transported as a file "patch" for updating/replication purposes on the clone machine's database. At present this terminology updating traffic is one way; Walldorf to Rendsburg, in future it will be bi-directional as confidence and competence increases.

The learning curve of the Rendsburg translators involved has had to be moderately steep. Thanks to the excellent work being done at the Fachhochschule in Flensburg by Dr. Klaus Schubert and his colleagues on introducing translation tools and systems to undergraduates, the most recent addition to the S&D/ICE translation team already have a grounding in METAL. Following on from the basic SAP R/3 training received in Walldorf, the next level of training for SAP was more specific, involving a focus on a certain area of R/3 technology (Logistics, Treasury etc.). This reflects the complexity of the R/3 system; whereas some terms can be seen as being "generic" to SAP across the board, other terms have an entirely different meaning depending on the R/3 area.

Translators are becoming specialists for separate areas whilst being able to cover for colleagues specialising in other areas due to peaks in certain areas, illness, vacations etc.

Results

Where are we now?

S&D/ICE started on this project in the Autumn of 1996, allocating two of its translators to it, whilst the MLT Group in Walldorf still handled the majority of the work. During the year a shift in the workload has taken place. A year later 6-10 translators are working full time on the project in Rendsburg, supported by "half" a person from the MLT Group.

Production is now running at some 300,000 words processed per month, a figure which is certain to increase again in the coming months. In terms of the key goals set for the project it is very successful. Quality, cost and time targets have been met and this model is probably scalable to handle any foreseeable volume increases in "Hinweise".

One very important element which must be included in the success paradigm is the "people" factor. Can translators obtain long-term job satisfaction in working within this type of framework? The question is vital to the long-term success of this and any future machine translation venture. Feedback to date has been positive: the subject matter is technically

complex and demanding and S&D/ICE will encourage translators to analyse, modify and improve the work model by "teaching" the machine new terms, improving the input text by devising new and better pattern matching routines and liaising with "Hinweis" authors.

Beyond outsourcing - distribution

The idea in the future is that SAP not only receives translated work from S&D/ICE but can also retain its investment in the hard work it has put into the project so far. Correctly coded terminology, new automatic prediting routines for source texts, and a massive archive of bilingual "Hinweise" will also be coming from Rendsburg. Even in cases where the entire system in Rendsburg were to fail, a backup service based on exactly the same data stock could be implemented within minutes.

And where to after that...?

The current work at the MLT Group and at S&D/ICE is providing insights into new services and larger-scale projects. For the multilingual technology group at SAP this means new machine translation directions, support for memory tools across the global corporation, some R&D of their own (SAP is one of the lead partners in the European OTELO project) and possibly new corporate clients for machine translation services.

Both MLT Group and S&D/ICE are interested in focusing on what is being described in the current OTELO project as "the occasional user" - people whose own foreign language skills may be low or non-existent for certain languages, who may require a range of language services running from "gisting" through to email translation to high-quality translation with a high-speed turnaround.

To more fully examine the requirements which such a user poses, SAP has, within the framework of the OTELO project, created a questionnaire which can be accessed and filled out via the Internet (<http://www.otelo.lu/quest/>). The authors of this paper would like to encourage as many people as possible to access this page, so that the questionnaire results can give the widest insight into this important field.

And with what system...?

Even assuming that we are successful in defining requirements which help us to understand how language tools might be able to be used on a wider scale, the question which still arises is whether there are tools and systems capable of meeting the linguistic and technical problems that this poses.

To be able to move from beyond the current, rather limited use of machine translation we feel that several things must happen:

- it has clearly been easier to "port" and downsize the mainframe and UNIX-based systems developed over the last 30 years to fit the size (and price) of today's personal computers than it has been to "upsized" the quality of the lingware. Therefore,
- we would make a plea for the parallel development of both lingware and software we would encourage large corporations to involve themselves with these developments, rather than waiting for someone else to do it,
- we would ask language tools manufacturers to enable their tools to be more open. This would allow for the increasing use of shared resources (such as terminology) across hybrid systems (TM and MT).

Conclusion

The future for both the language services vendor and the language tools manufacturer looks to be interesting and challenging. The challenge for the services vendor will be to find these new, broader service markets; the challenge for the tools manufacturer will be to meet him there.

¹ Daniel Grasmick can be contacted via email at daniel.grasmick@sap-ag.de

" Chris Pyne can be contacted via email at cpyne@iceurope.de