MANAGING DISTRIBUTED MT PROJECTS TODAY
- A NEW CHALLENGE -

Jennifer A. Brundage, Susan McCormick
SAP AG
P.O. Box 14 61, D-69185 Walldorf
jennifer.brundage@sap-ag.de
susan.mccormick@sap-ag.de

Chris Pyne
S&D
P.O. Box 447
D-24754 Rendsburg
cpyne@iceurope.de

Abstract

The current trend towards globalization means that even the most modern of industries must constantly re-evaluate its strategies and adapt to new technologies. As a long-time supporter of MT technology, SAP has shown that it can make productive use of competitive, commercial MT products along with other CAT products. In making MT work for them, however, SAP has also had to substantially adapt the products that they received from MT companies. The result, after many years, is a full range of peripheral tools and workflow scenarios that support the use of their MT programs.

1.   The Status Quo

While NLP applications generally enjoy a reputation as 'cutting-edge' in the move to a more knowledge-/information-based marketplace, some areas of commercial NLP research and development, notably MT, are actually mature enough to be facing the same pressures to modernize that more traditional industries face. Given the job of translating large volumes of technical documentation, the members of the Multilingual Technology (MLT) group at SAP have proved over the years that they can work more efficiently by taking advantage of what MT has to offer, e.g., consistency in terminology, quick throughput of large volumes of technical documentation. However, this required adapting the MT products substantially, e.g. regarding a special workflow for the translation of customer error notes, a user interface for generating terminology files for mass import for German-to-English.

While the MLT group has been successful in using MT to improve efficiency in translating SAP documentation in-house, it is now introducing a translation memory tool (TM) to provide additional support to the specialized translators, also in particular for language pairs for which commercial MT systems have not proven efficient enough or are not available at all. Both the rapid growth of SAP itself and the changing nature of work in a more global market mean not only that the MT technology and approach must be adapted, but also requires the available tools to be combined and integrated in a

(more) intelligent way. In place of the traditional, monolithic approach to MT translation, there are a number of approaches, each tailored to a specific company need:

- All MT and/or TM translation work is done in-house
- Texts are preprocessed and translated in-house with MT and/or TM, but outsourced for post-editing
- Texts are outsourced for preprocessing, but then translated with MT and/or TM and post-edited in-house
- All MT and/or TM translation work is outsourced

2. Facing the Challenge at SAP

The changing reality of work distribution within a company like SAP would ideally be reflected in new approaches to integrating and using MT. As currently constructed, MT systems are not flexible enough to address a wide (and growing) variety of workflow types. Accommodating them will mean simultaneously supporting scenarios that combine different work environments, e.g. working centrally in a client/server environment, working remote with local tools, working online or offline, developing terminology centrally or locally.

Adapting to a new situation will have to occur on several fronts:

2.1. Terminology Development and Management

While MT provides measurable improvements in terminology consistency, it can be a major investment of time and money to bring an MT lexicon to the level in which it can be used profitably. MT systems often require specialized linguistic knowledge and/or are packaged with interfaces that do not adequately address the needs of high-volume users. If moreover a user requires systems from different vendors, the challenges of MT lexicography are multiplied, since different systems often

> have different platform requirements
> have different lexicon interface approaches
> use different concepts
> use different internal formats
> require different, system-specific knowledge for terminology management

The MLT group manages SAP data in three different systems, the company-internal term database *SAPterm* (which for some language pairs has been imported into the TM-internal term database MultiTerm), the Metal lexicons for German-English, and the Logos lexicons for English-French. In addition to the protocols that have been developed in-house to maintain consistency of data lexicon-to-lexicon, the requirements for terminology maintenance given the new array of workflow types must now be factored in.

Several issues are currently being addressed:

- How to avoid multiple and/or disparate codings of the same entries, given different systems and different sites in which terminology can be generated

- How to adapt MT lexicon interfaces to ensure that professional translators handle terminological data that is relevant to them, and not specific to an MT system only

- How to provide centralized terminology development and administration and at the same time accommodate the needs/input of translators who work in remote locations.

In general, ongoing efforts to standardize terminological/lexical data will benefit the MLT group. Work now being done in the framework of the OTELO project (in which SAP is a full partner) will be the first attempt at standardization to directly benefit SAP's translation work.
The OTELO lexical database will house data from all of the relevant SAP lexical resources; data will be represented in a common exchange format and will be input via a generic interface that supports MT as well as traditional terminological requirements.

2.2. Continued Productive Use of MT

New requirements for MT at SAP will take into account not only the needs of in-house professional translators, but will expand to include translators working remote or any non-translator within the company requiring language support. Currently, the MLT group at SAP is supporting resp. planning the following scenarios:

- Providing quick, roughly post-edited translations involving MT and/or TM for internal use or finalized versions for first customer shipments
- Preprocessing and translation in-house involving MT and/or TM, post-editing done by an external translation bureau, quality control in-house
- Preprocessing, translation involving MT and/or TM, and post-editing done by an external translation bureau, quality control in-house
- Translation done in-house, by a subsidiary, or agency making use of TM (documentation updates)
- Providing translation services for colleagues with an occasional need for translation making use of email and/or the web

The MLT group at SAP has been successfully practicing some of these methods ever since MT was introduced to the company in 1990 for German-to-English and 1995 for English-to-French. In the near future, SAP will introduce further language pairs: English-to-Spanish, English-to-German.

3.   S&D and Machine Translation

S&D involvement began with its first METAL system 1988. The system was incredibly expensive in today's terms - and in hindsight was probably bought for the wrong reasons.
Yet it was used for specific projects for key customers in the early nineties. Here the system operated tolerably well and fulfilled its objectives. During the projects S&D also learnt to lower their sights considerably. The conclusions that were drawn from these early projects were logical and self-evident. In a clear, well written technological and terminological domain it would be possible to be effective. However, S&D would have to rely more on themselves for improvements outside the core MT engine itself.

3.1.   MT - A Reappraisal

How else can you use MT to effect as a serious service provider? Step one was to reanalyze and research out customer base.
Are there customers who fit at least one or preferably several criteria?
 •   Terminology/know-how in-house
 •   Type of translation jobs which fit
 •   Stream-type translation model
 •   Interested in MT as an added-value service to enhance currently offered services

3.2.   Working With SAP

Performing MT outsourcing work with SAP is ideal in relation to the above-mentioned criteria on several levels:

1. SAP has learnt more about MT by doing than any other company S&D knows, thus allowing a quicker kick-off of the project.
2. The SAP terminology, though extensive, could be learnt and cloned from the SAP donor system
3. The job type SAP wanted S&D to do was well-defined and fitted MT workflow
4. The work was and is literally a constant stream. This factor was important for S&D because it meant a medium-term investment in personnel and know-how
5. We could take over the work in stages ...

Stage	Logistics	Personnel/skills
1	Set up digital (European ISDN) line to SAP, Walldorf via router	Recruit for new SAP group out of current staff
2	Set up nailed line to SAP Germany backbone	Initial training in Walldorf
3	Organize X-terminal SW for 586 workstations, install R/3 UI's locally	Further recruiting in England/Ireland
4	Set up dedicated METAL SunSparc server for cloned SAP METAL database	Specific area training in Walldorf

Stage	Logistics	Personnel/skills
5	Organize replication scenarios for term updates Walldorf ↔ Rendsburg	Construction of a "stand-in" personnel pool / each SAP-worker with his/her "shadows"
6	Develop facilities for monitoring statistics and trends in work levels and area concentrations	In-depth / on-job Mt training
...	...	...

3.3. Further goals

As also for SAP, the watchword here is optimization. It must be ensured that turnaround times, pre/post-editing techniques, terminology management are optimized to make the entire system more robust and ready for future challenges.

4. Occasional MT User Project

Providing MT services to "translation professionals" is far less difficult than offering them to non-MT "savvy" corporations or to individuals and groups within a corporate environment who are not language specialists but still need a rapid, scaleable range of language services to meet various needs. They may need translations from or to foreign language texts to support decision-making, to communicate, to understand issues etc.

Experience (of MT) has shown that it is not possible to offer serious MT translation services simply by pushing everything through the MT system and automatically returning the results to the customer.

Both, SAP and S&D have certain assumptions as to the needs of such people. But these assumptions must be validated, and the workflow scenarios geared to the validated findings.

Some of the made assumptions

- Current MT systems are not geared to handling (receiving and returning) masses of documents, sorting them according to urgency or other criteria etc.
- It will be almost a DIY approach. There is no adequate product available on the market
- Users will need to be offered a choice of "human intervention" levels
- An experienced group of in-house translators for peripheral (pre/post-processing and maintenance) work around the MT system is required
- Access to such services must be through the standard software environment of the user.

These assumptions will be verified within the framework of the OTELO project and the results will be used by SAP and S&D to determine an appropriate workflow for occasional users.

5.   Presentation Outlook

In their presentation, SAP together with S&D will present combined translation processes supporting MT and TM. Special focus here will be on the organizational, technical, logistic, and terminological bottlenecks that were encountered during the implementation of the individual and joint projects.