MT in real-world practice: Challenges and solutions at Swiss Federal Railways

Nadira Hofmann
STAR Group
Wiesholz 35, 8262 Ramsen
Switzerland
nadira.hofmann@star-group.net

Maryse Lepan
SBB AG
Hilfikerstrasse 1, 3000 Bern 65
Switzerland
maryse.lepan@sbb.ch

Abstract

This user study uses the example of the Swiss Federal Railways (SBB) to show how an MT system is evaluated and introduced in practice. The first part describes the motivation and requirements for the company when it comes to introducing machine translation. Part two explains how the benefits can be determined before the system is launched by reliably analysing real product jobs and how a conclusive pilot phase can be implemented in a company's real-world setting. The third part deals with the findings from the pilot phase and uses specific examples to show how they have been taken into consideration and implemented for the product going live.

1 Introduction

1.1 Multilingualism as a tradition

Switzerland has several official languages and has written the promotion of *"understanding and exchange between the linguistic communities"* into the constitution. As the national rail company, the Swiss Federal Railways (SBB) therefore also has a long tradition of linguistic diversity: Translation has been part of operations at SBB since it was founded over 100 years ago. Multilingual project teams and multilingual communication with employees, customers and suppliers are part of day-to-day business.

1.2 SBB Language Services

The 15 people in the SBB Language Services team are supported by 10 external translation agencies and freelance translators. It is responsible for all translations and for the centralisation and management of the corporate language and terminology, which is developed in collaboration with technical experts and language specialists.
The following language technology is used:

- STAR CLM (Corporate Language Management) for managing the language processes,
- Transit as the translation memory system,
- TermStar and WebTerm as the terminology management systems.

Thanks to this technology, since 2001, SBB Language Services has developed a large, well-structured translation memory and comprehensive dictionaries with validated terminology. The SBB dictionary that resulted from this is available to the entire workforce at SBB.

1.3 Creating added value from existing data

The company was prompted to think about introducing an MT system by an SBB talent programme for first-line managers where the task was to create added value for SBB.
It was an obvious choice to use machine translation to generate added value from the linguistically validated data from the translation memory and terminology. Two approaches were pursued for this:

- Integrating the MT into SBB Language Services' existing translation workflow in order to support professional translators by offering additional MT-generated translation suggestions,
- Integrating the MT into the SBB intranet portal in order to support all employees by offering ad-hoc translations for their communications (*"SBB Translate"*).

The following requirements were present as framework conditions:

- Develop a valid decision-making tool to decide on the sense, practicability and economic efficiency of an MT system,
- Use the company's own terminology and formulations in the railway jargon that is approved by SBB,
- Seamless integration into existing processes and into the IT environment at SBB,
- Scalable and expandable for future requirements.

2 Evaluating the added value

To decide whether the solution actually generates added value, the decision-makers required reliable information regarding the benefits and quality in everyday translation.

For "facts and figures", engine training and an initial analysis with real productive jobs from the SBB Language Services were initially carried out. The prerequisites for integrating the MT solution into its IT environment were also checked.

The pilot phase that followed involved the evaluation of how those involved in the process handle MT in practice.

2.1 Training the evaluation engine

The sample evaluation is carried out with *one* language combination, for which the involved parties can handle and evaluate the source and target language.

Therefore, in this case, German-French was trained with the SBB training material (2,593,609 segments for the translation memory and 42,788 dictionary entries). The engine was trained and hosted by the system provider.

2.2 Initial analysis with real production jobs

846 production jobs from recent months that were human-translated and human-reviewed without MT support were then analysed.[1] The jobs were repeated using MT and automatically compared with the existing results from the human translators and reviewers. [2]

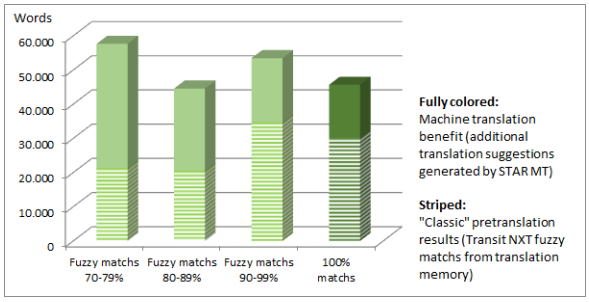

Figure 1. Additional MT-generated translation suggestions from the MT system

Result:

- The "perfect matches" (suggestions that could be applied without any changes) increased by 50%.
- The number of "good" translation suggestions (fuzzy quality 70-99%) more than doubled.

2.3 Pilot phase in translator's everyday work

The figures from the initial analysis proved the theoretical benefits of using MT. The pilot phase had to show whether this would result in real added value in a translator's everyday work. In addition to pure figures, acceptance, usability and "perceived" benefits also play a leading role.

To answer this question under real-life conditions, the MT should and must be integrated into real production jobs that are part of ongoing operations. As for all production jobs, the project management was therefore the responsibility of the customer. However, the engines were hosted by the system provider because, from experience, customers/interested parties in the pilot phase do not want to concern themselves with how the MT works or its infrastructure.

There are several scenarios for integrating the hosted MT solution. In this specific case, the project managers created their projects without machine translation and sent the project packages to the system provider. Here, segments that were not pretranslated from the translation memory were enriched using translation suggestions from the hosted MT system[3] and the packages were sent back to the project manager who was then able forward them to the translator.

[1] The document types were categorised as follows: 72% Word, 20% PowerPoint, 7% Excel, 1% Visio. The total volume was 587,927 words or nearly 4.7 million characters.

[2] In order to get meaningful results, these production jobs were *not* part of the SBB training material and therefore not used to train the engines.

[3] MT suggestions are not only created for "no matches" but also for segments with fuzzy matches from the TM. The procedure is still unusual, but it is logical: High fuzzy quality implies that the segment to be translated is very similar to the TM. Since the TM also acts as the basis for the engine training, the statistical machine translation for these segments provides particularly good results.

With this, nothing changes for internal and external translators: They do not require any additional tools or work steps; the MT suggestions are offered as fuzzy matches in the translation editor window of the translation memory system and can then be used with the familiar functions.

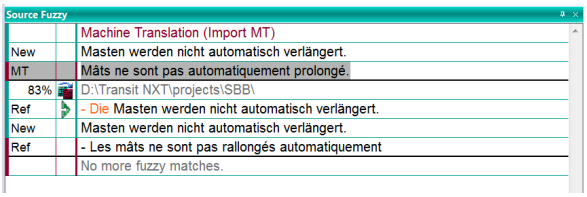

Figure 2. MT suggestion and classic fuzzy match in Transit's translation editor

The translators also do not need to access to the MT system or the hosted engines: They receive the additional MT suggestions automatically with the project packages. This means that external service providers and employees who work from home can be easily integrated into the pilot phases without any technical hurdles.

2.4 Web application for specific translation tests

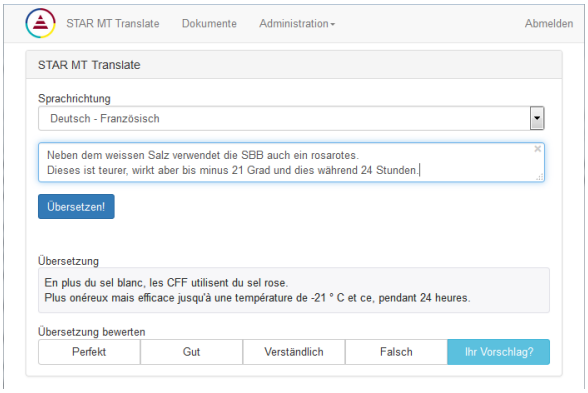

Figure 3. Prototype for the online solution – not yet customised and adjusted to the customer's corporate identity

In parallel to this, a web application was provided in order to request machine translations from the hosted engine via the browser. This means that the experts from SBB Language Services were able to carry out tests to determine how individual sentences or paragraphs are translated by the MT system and how the system performs to different source texts.

At the same time, the application was a prototype for SBB Translate as an online solution for translating individual sentences, but also entire documents[4]. Here, selected SBB employees were able to test whether the solution meets the demands of everyday work.

3 Findings and challenges from the pilot phase

The pilot phase produced the following findings, which could then be taken into consideration when the system goes live.

3.1 All language directions directly and without pivot languages

In accordance with the requirement from SBB, all possible combinations and directions of the four languages (German, French, Italian and English) had to be supported.

Analysis of the existing data (volumes of the text corpora and terminology) showed that, for all translation directions requested by the customer, separate engines can be trained – i.e. a total of 12 engines for four languages.

This means that all translations can be carried out directly. It does *not* result in any of the adverse effects on quality or performance that are expected from machine translations using the "detour" pivot language.[5]

3.2 No differentiation according to subject area

The well-structured translation memory would have allowed a differentiation in order to train separate engines according to subject area.

However, the pilot phase showed that this differentiation is neither necessary nor useful. For this reason, the SBB material was used to train just one engine for each language direction.

[4] All of the file formats that are relevant in practice can be supported. For the specific customer, support for Office documents, PDFs and text files is provided.

[5] It makes sense to use the language that is used most frequently (in this case, German) as the pivot language. Language combinations that do not involve pivot languages would be translated using two successive machine translations.
Example French-Italian translation: The French text would be machine-translated into German and the German text would then be translated into Italian.
This increases potential MT errors and the server load increases because two machine translations would have to be performed for one translation request.

3.3 Generic back-up engines for general-language translation requests

When testing the web applications, the experts from the language department predominantly request translations of railway-specific texts with specialist terminology. In contrast to this, the requests from testers outside of the language department were significantly more general-language.

The engines were not initially trained for this; the translation results did not always meet the expectations of the "translation laymen".

To translate these types of text with better results, additional, generic back-up engines are used for the web application. These are enriched with freely available corpora (e.g. from Europarl) and are automatically taken into consideration if the SBB-specific trained engines cannot generate a suitable translation.

3.4 The human factor

We know that the human factor decides on the acceptability and usability of the MT. The various target groups (professional translators vs "normal" employees) have differing knowledge, expectations and reservations, which all have to be taken into consideration when the MT goes live.

In terms of staff numbers, the target group of translation professionals is a known quantity and these people can therefore be informed directly and individually. A key aspect is the use of MT suggestions in the translation process. This is made easier by the fact that the translation process itself does not change – it is "only" supported by additional translation suggestions.

An individual approach is not possible for the numerous users of the web application (30,000 SBB employees). They are informed about the opportunities and limitations of machine translation via an attractive FAQ area: MT as a tool for understanding foreign-language texts – but not as a replacement for professional translation by SBB Language Services for documents that are to be published.

3.5 Text corpora with a broad range of formats

Many companies call upon the content from the projects from the language department as well as the extensive data stocks from translations that could be used as training corpora. This content is also not usually available in translation-typical exchange formats (e.g. TMX or XLIFF) because the sustainable data usage and managed language processes are not often at the forefront of such translations.

In this specific case, the contents of the SBB website were localised by external agencies, for example, but could not be retrieved from the translation memory for SBB's language department.[6] The content was then only available in the form of more than 20,000 HTML files. [7]

To prepare such contents for engine training, the translation memory system's filter technology is used. It generates format-neutral language files so that content from any source and format can be used.

3.6 Morphologically generated additional information

Terminology plays an important role in engine training and has a significant impact on MT quality: The more validated terminology is used for engine training, the better the translation results provided by MT.

SBB Language Services has carried out extensive terminology work that has, to date, been used in collaboration with the TM. The dictionaries also usually contain the base form of nouns, verbs and adjectives, while the texts that are to be translated usually contain inflected forms. A large proportion of the terminological potential would remain untapped if MT engines were only trained with canonical forms.

Morphology is used to close the gaps between dictionary entries and real texts. The technology for this comes from the translation memory system, which provides morphological support for over 80 languages and language variants. In this case, "morphology" means linguistic expertise mapped out in tried-and-tested rules, and not just simple stemming.

As an example, the values from the French-German engine show what morphology can offer:

- Text corpus: Translation memory with 3,007,240 segments/36,882,122 words

[6] The website contents were particularly valuable for training the English engines: In the SBB Language Services' translation memory, English was heavily under-represented when compared with the other languages.

[7] The HTML files were asynchronous material as the individual languages differed in structure and contents. Therefore the material was used as basis for monolingual data to train the language models.
In scenarios with synchronous bilingual or multilingual documents, bilingual language file pairs can be generated and used for engine training.

- Terminology: 35,815 language entries
- Morphologically generated additional terminology and segments in which they occur: 2,063,522 segments

The larger terminology base and the additional context information means that the BLEU score for this engine increases from 35 to 48. Irrespective of the BLEU score, the translations that were enhanced by extra morphologically generated terminology were clearly preferred by translators during a manual evaluation of sentence BLEU lists.

3.7 Web application with automatic TM pretranslation

The trained SBB engines usually provide good results for texts that were not previously professionally translated. However, the human translations that are validated by SBB Language Services and are available in the translation memory are of a higher quality. By definition, they have a BLEU value of 100.

The web application therefore uses the same process that is common and established for professional translation systems: For segments that are already contained in the translation memory, the translation from the translation memory is used (100% matches); for the rest, the MT engines are used. Thanks to the high performance of the TM indices, the user does not notice any increase in the response times.

In addition to the increased translation quality, the two-stage process has another advantage: Newly translated segments from SBB Language Services' projects immediately flow into the web application from the translation; new formulations, terminology and text types are immediately available. This means that the intervals for retraining the engine can be increased.

3.8 Focus for optimisation strategies

MT processes and MT engines are complex and have many influencing parameters that help to further improve the quality of the translation results. The theoretical opportunities are almost infinite but, in practice, it is useful to focus on the relevant areas.

This decision is supported by analysis functions. Examples:

- The evaluation of the engine-specific access figures shows which language combinations are used particularly intensively.
- The interactive feedback function of the web application reproduces the "perceived" translation quality and allows for targeted improvements, where required.
- With OOV statistics ("out of vocabulary"), SBB Language Services can determine which terms in the terminology work should be prioritised.
- Quantitative evaluations of peak loads and load distribution provide IT with information about the sizing of the system.

4 Long-term perspectives

The introduction of an MT system is no rush job: It may take months before the solution takes hold and is accepted by the employees in their everyday business.

It is even more important that this investment in time and resources offers a long-term perspective and is open to future requirements that cannot currently be foreseen by the parties concerned. Future-proof interfaces that allow the solution to be integrated into changing IT structures are particularly relevant.

In this specific case, the following scenarios are envisaged and, from a technical point of view, could already be implemented:

- Apps for iOS/Android so that the MT can be easily used on mobile devices.
- Integration into Office in order to request MT directly from Word, Excel, PowerPoint or Outlook
- An API that can be used to translate texts and documents using third-party applications.