

The MT developer/provider and the global corporation

Terence Lewis¹ & Rudolf M.Meier²

¹ Hook & Hatton Ltd, ² Siemens Nederland N.V.

Terence.Lewis@language-engineer.co.uk, Rudolf.M.Meier@siemens.com

Abstract. This paper describes the collaboration between MT developer/service provider, Terence Lewis (Hook & Hatton Ltd) and the Translation Office of Siemens Nederland N.V. (Siemens Netherlands). It involves the use of the developer's Dutch-English MT software to translate technical documentation for divisions of the Siemens Group and for external customers. The use of this language technology has resulted in significant cost savings for Siemens Netherlands. The authors note the evolution of an 'MT culture' among their regular users.

1. Introduction

In 2002 Siemens Netherlands and Hook & Hatton Ltd signed an exclusive agreement on the marketing of Dutch-English MT services in the Netherlands. This agreement set the seal on an informal partnership that had evolved over the preceding six-seven years between Siemens and the Developer (Terence Lewis, trading through the private company Hook & Hatton Ltd). Under the contract, Siemens Netherlands undertakes to market the MT Services provided by the Developer to divisions within the Siemens Group and to prospective external customers.

This arrangement has proved to be a win-win situation for both Siemens Netherlands and the Developer. It has given Siemens Netherlands access to translation technology at a time when it has been involved in major cross-border infrastructure projects with consortium partners that need rapid access to usable English translations of Dutch technical and legal documentation (technical requirements, relevant legislation and regulations etc.). Siemens is responsible for the complete administration of the MT customers, and this set-up has provided the Developer with a stable stream of revenue enabling him to devote more time to further development of the software. This paper provides a mostly practical account of this relationship, first through the eyes of the Developer, and then from the perspective of the global corporation.

2. The Developer's perspective

Under the current arrangement, the Developer (working from home in the United Kingdom) currently acts as the sole operator of an Automatic Translation Environment (Trasy), processing documents forwarded by e-mail by the Translation Office of Siemens Netherlands in The Hague. Ideally, these documents are in MS Word or rtf format, but the Translation Office also receives pdf files or OCR output files for automatic translation. After conversion to a machine-processable format, these files need to be cleaned up, otherwise the MT output will be markedly inferior to the agreed baseline. This clean-up is done by either party by arrangement.

The documents are processed and returned on the basis of an agreed delivery date. Short documents are often returned within the hour. They are not post-edited or revised line by line, but the Developer does quickly scan each page to spot any gross errors or omissions. Sentences containing gross errors are usually sent back to the MT system after further preparation work. Broadly speaking, users receive automatically generated documents. These are either edited by the end users themselves or used "as is" by engineers familiar with the technical background to the documents.

In addition to the run-of-the-mill jobs that pass through a corporate translation office, the automatic translation environment has been used to handle huge volumes of documentation

from such major projects as the HSL-Zuid (the high-speed rail link between Amsterdam and Paris), the RandstadRail light-rail project in the Hague Region, and a huge construction project for an external customer, Shell International.

Trasy is built around a small LAN comprising 4 PCs:

- a Linux gateway to the Internet
- two Linux translation servers
- a translator workstation/client running under Windows 98

This configuration has been set up to facilitate the handling of a flow of jobs in an office environment. Of course, the number of translation servers and workstation/clients can be increased to meet any required number of translator/operators. All the dictionary building work is done on the translator workstation and the up-to-date lexical resources are then uploaded together with the files to be translated to the Linux servers. Typically, the Developer/Provider will start the terminology work for the next job while the current job is being processed. This client/server arrangement has been found to be the most practical one, but it is possible (although somewhat inconvenient) to put the whole set-up on a powerful notebook running either Linux (or Solaris) or Windows XP (or both) that functions as a mobile translation workstation.

The use of two translation servers by a single operator makes it easy to jump the queue to handle shorter jobs when a long job is already running. It also enables the Developer to run the same job simultaneously on two successive builds of the software, thus providing a simple mechanism for observing the success or failure of intended software improvements. This is vital since the Dutch-English MT application, written entirely in Java, is seen as a dynamic process – most large jobs will generate at least one improvement in the software. And the process of improvement has been customer-driven: mistakes that were tolerated six years ago are now pointed out with alarm.

3. Software tools

All jobs are processed using the already described approach combining Machine Translation and Translation Memory. This involves using a translation memory application as the fi-

nal delivery tool and only sending to the MT engine sentences that are not already present in the translation memory database. The MT software produces a file that can be imported directly into the translation memory. This selection, extraction and integration is performed semi-automatically. Every job the Developer does is ultimately stored in the translation memory database. Figure 1 shows an example of translation unit in an aligned file produced by the MT software. It should be noted that the MT software can also be used on a stand-alone basis, producing an html output file.

```
<TrU>
<Quality>98
<CrU>LEWIS!
<CrD>15042004,08:38:04
<Seg L=NL_NL>De hoofddraagconstructie is van
staal.
<Seg L=EN-GB>The main supporting structure is of
steel.
</TrU>
```

Figure 1: Translation unit produced by MT software

The following software tools are used in this translation environment:

- Resource Building Tools
- Dutch-English machine translation software
- Trados Translator's Workbench

The Resource Building Tools, written by the Developer, include conventional single-word dictionary-building tools plus tools for building resources consisting of multiword expressions up to idiomatic translations of complete sentences. The utility called "ResourceBuilder" can produce useful sub-sentence segments from translation memory databases, aligned source/target files and even different language versions of web pages. It is the **key factor in securing user acceptance of the automatically generated translations** since many of the constituent parts of sentences are taken from "human translations". The single-word dictionary and the Resources Repository are basically simple resource files (enhanced text files) which are only compiled at run-time. The Resources Repository contains all lexical items that are not single-word entries.

Examples of entries in this repository are:

“Tot het werk behoort”, “<MF>The work includes</MF>,,
(literal translation =,to the work belongs”)
 “van den Berg”, “<N1 semType= „proper_noun”>van den Berg</N1>,,
(Dutch surname, not “of the Mountain”!)
 “in dit document wordt beschreven”, “<MF>this document describes</MF>,,
(literal translation: “in this document is described”)
 “Centrale Velsen”, “<N1>Velsen Power Plant</N1>,,
 “certificerende instantie van de opdrachtgever,
 “<N1>client’s certification agency</N1>,,
(literal translation: “certifying instance of the client”)
 “Dit systeem is gelaagd opgebouwd”, “<MF>this system has been built up in layers</MF>,,
(literal translation: “This system is layered built up”)

Figure 2: Examples of repository entries

The Dutch-English translation software is a modular application incorporating a hybrid approach to machine translation. Since the Developer believes that this approach is critical in achieving an acceptable quality level for Siemens Netherlands users, a few paragraphs will be devoted to the internal flow of the application.

The program first of all applies known “data-driven” techniques to match the strings in the input file with segments and sub-segments stored in the Resources Repository. The software attempts to provide grammatical tagging for the results of these matching operations, and applies some rules for extracting usable translations from the Resources Repository by consulting the single-word dictionary. This module is a self-contained application in its own right,

„tot het werk behoort”=„<MF>the work includes</MF>,,
 „het programmeren en laden van software”= „<N1>the programming and loading of software</N1>,,
 „die gebruikt wordt voor de sturing van”= „<MF>which is used to control</MF>,,
 „fabrieken”=„<N2>factories</N2>,, <single word look-up>

Figure 3: Example of a sentence constructed using ‘data-driven techniques’

and, with sufficient data, can produce complete translations independently within limited domains. Figure 3 shows the translation of a complete sentence produced by mining the Resources Repository with one word being looked up in the single-word dictionary.

After this module has run, the sentence is then passed to a Rule-Based Machine Translation module which applies various known MT techniques to the sentence. These include the parsing of “clusters” or chunks of text, recognition of known patterns and techniques for achieving semantic disambiguation based on the identification of the subject-matter of the document. At this stage, for instance, the string “niet zal kunnen worden gekozen” would already be identified and tagged as a verbal phrase and the provisional translation “will not be able to be chosen” would be allocated for it (the word-for-word translation being “not will can be chosen”). The term “provisional” is used since other modules might “decide” to change this translation at a later stage. These various techniques are applied “sentence down”, i.e. the last series of rules (contained in an IssueResolver module) deals with single words. In the final part of the ‘run-time sequence’, the sentence is passed to the EnglishStyleEditor module which works on the target sentence and performs some of the tasks of a human post-editor. For example: “the decision is taken by the board” becomes “the board takes the decision”.

Of all these modules, the semantic disambiguation module, particularly the business of getting the technical terms right, is a very important aspect for the non-Dutch Siemens engineers who are the main users of the MT output. If they know they are dealing with transformers, substations, catenaries and the like, they are not too bothered with a clumsily constructed English sentence, nor are they in the least interested in the software tools used to generate their document. **Much of the success of the use of MT at Siemens Netherlands is due to the Developer’s commitment to getting the terminology right.** In the important field of railway engineering, for instance, every major piece of railway infrastructure, including substation and transformer names, has been entered in the repository, which even includes the names of the key personnel involved in the HSL-Zuid pro-

ject. So high is the confidence in this terminology that Siemens staff translators working on “human translations” sometimes send the Developer their terminology queries. But this is a two-way process, and the Developer is also able to forward terminology queries to the Siemens Translation Office before processing a job. To conclude this first part of the paper, the main advantages for the Developer in this close relationship with Siemens Netherlands are:

- a guaranteed flow of revenues
- an opportunity to test new developments on “real texts”
- the possibility of building in knowledge of Siemens’s core areas of expertise
- a constant challenge to meet the growing quality demands made by Siemens customers over the years

4. The Siemens perspective

4.1. General situation of in-company translations departments

The increasing globalization, liberalization of markets, and ever growing global (tele)communication speeds and possibilities as well as global competition have fundamentally changed the way in which enterprises conduct their business. Lifetime employment in a company or even in a department is a phenomenon of the past.

Instead, companies are in a constant move of internal reorganizations, assessing outsourcing paths, redefining core business, selling activities to other companies.

Internal departments such as a translation department can no longer trust that the once made decision to set it up is a future-proof decision. Internal translation departments are subject to shareholder-value-driven profitability considerations. They must prove to be able to produce an added value and increase profitability in order to safeguard their survival.

4.2. Situation of the in-company translation office at Siemens Netherlands

The translation department of Siemens in the Netherlands was founded in 1988. Its mission is to ‘satisfy the translation requirements of Sie-

mens Nederland at market prices while covering its costs in full.’

Although the translation department has succeeded in fulfilling its mission over the years, this task has become more and more difficult on account of the developments described above and the specific situation of the translation market which is characterized by a constant downward movement of prices, while at the same time the internal cost allocations have constantly risen.

4.3. Impact of introducing the MT service of Hook&Hatton (‘Trasy’)

4.3.1. General description of the Trasy service

In the early 90’s, the translation department came into contact with Hook&Hatton which offered MT services from Dutch into English. The incidental cooperation between the translation department and Hook&Hatton has developed into a firm relationship on the basis of an exclusive contract for the delivery of MT services for the Netherlands which was concluded in 2002.

The Trasy service of Hook&Hatton which the translation department of Siemens is offering to internal as well as to external customers is a unique selling point. Constant reviews of existing competing MT system show that the Trasy service is the best available choice for machine translations from Dutch into English.

There is no organization in the world which can offer machine translations from Dutch into English of higher quality than the translations department of Siemens!

Through its preediting approach the system is constantly learning and thus preserving its competitive edge over the other systems in the market.

The recognition of the system under internal customers has grown over the years. Situations in which customers by themselves ask for a MT translation are everyday practice. The price/performance of Trasy is second to none. In the most favorable case Trasy can deliver 150.000 words a day, seven days a week!

4.3.2. Financial impact for the translation department

The MT business has in the meantime grown to a volume of €220.000 (in fiscal 2004), that is 24% of the total turnover of about €900.000.

Roughly 30% of this turnover is realized with external customers.

In other words, 24% of our sales come from a sector in which the translation department is and will stay the top-quality supplier.

4.3.3. Financial impact for Siemens Netherlands

For Trasy machine translations the translation department charges 40% of the price charged for human translations. On the basis of the MT turnover with internal Siemens Netherlands customers of about € 154.000 in 2004 this means cost savings of about € 231.000 for this period for Siemens Netherlands.

4.3.4. Organizational impact for the translation department

These figures contributed to the fact that the translation department of Siemens Netherlands has been able to prove its added value for the organization.

4.3.5. Side-effects of offering MT services

The translation department of Siemens Netherlands has experienced a very positive side-effect of offering MT services to big customers. These customers have decided to also order human translation from us. The advantages of one-stop-shopping may have been the reason for this. It may be claimed without any exaggeration that the introduction of the Trasy MT service has safeguarded the existence of the translation department of Siemens and caused a material cost reduction not only for Siemens Netherlands but also for external customers.

4.3.6. Translators' professional considerations

Does Trasy threaten the existence of human translators? This is a very difficult question which cannot be answered satisfactorily here. Technical and scientific developments have always had an effect on existing professions. The rise of the motor-driven vehicles has caused the disappearance of coachmen, for example.

On the other hand, technical developments are continuously creating new professions.

As far as machine translations systems at the current stage of development are concerned: they often create additional translation needs. Documents for which no human translation would be ordered, become translation jobs. Cost savings due to machine translation free funds for – in much cases more interesting – human translations.

5. Conclusion

The authors are aware of the fact that the subject of this paper is limited to the effect of working with the Trasy MT system in the translation department of Siemens Netherlands. This paper does not claim to be a general study about the implications of MT systems for translation agencies and for the whole translation market. It does, however, describe the practical benefits of collaboration between a developer working with a 'lesser used language' and a major corporation with a requirement for rapid, low-cost translations from that language.

The following paragraph was delivered to Siemens Netherlands by Trasy on 28/01/2005:

“The Contractor should be fully aware that the activities to be carried out by him at the locations take place during the normal operational management of the production facilities. The production and the related supply logistics may never be impeded by activities relating to the access control. Therefore installation operations should be planned at all times in proper consultation with the Principal, and where necessary temporary provisions should be made by the Contractor to limit obstacles for the production to an acceptable level for the Principal.”

6. References

LEWIS, Terence (2001). 'Combining Tools to Improve Automatic Translation'. Paper given at the 'MT Summit VIII' conference, 18-22 September 2001, Santiago de Compostela, Galicia, Spain.