



Russian Translation of Documentation for Siemens' Telecommunication Equipment

O. Y. Vigodsky, G. V. Yaltchick

Oleg Y. Vigodsky

11, Varshavskaya, St.Petersburg, Russia, 196128

LONIIS Translation Centre, Manager

e-mail: "Oleg Vigodsky" olv@elder.loniis.ru

Gleb V. Yaltchick

11, Varshavskaya, St.Petersburg, Russia, 196128

LONIIS Translation Centre, Software Engineer

e-mail: "Gleb Yaltchick" gyv@elder.loniis.ru

Stichworte

Keywords

russian translation, machine translation, filters, dictionaries

Kurzfassung

Summary

The experience of Russian Translation of Documentation for Siemens' Telecommunication Equipment is presented. The translation process is described, namely, telecommunication dictionary development and Machine Translation system customization, filters (for FrameBuilder files).

Russian Translation of Documentation for Siemens' Telecommunication Equipment

Introduction

Today, Machine Translation (MT) tools are very effective while translating voluminous documentation. In the mean time, some features of translation process trouble the use of such systems. The main reason is that the documentation to be translated is already created in a specific format, i.e. contains tables, pictures, cross-references and even hyper-links. In general, MT systems whether do not support specific formats or distort file structure. Here, the process of using MT systems with no distortion of text and control information is disclosed.

Tasks

The ideal documentation translation process is to meet following requirements:

1. Maximization of automatic operations, centralization
2. Excluding an information not to be processed (listings, screens, user interface, menus, commands, messages, etc.) from the translation process
3. Storing and re-using document' service information (document structure, tabs, margins, fonts, text formatting, etc.)
4. Translating text using centralized terminology data base, different dictionaries, providing a consistency (terms and style)

Approach

A document to be translated is split to two parts: text (which is actually translated) and service information (which is used further to create a target document).

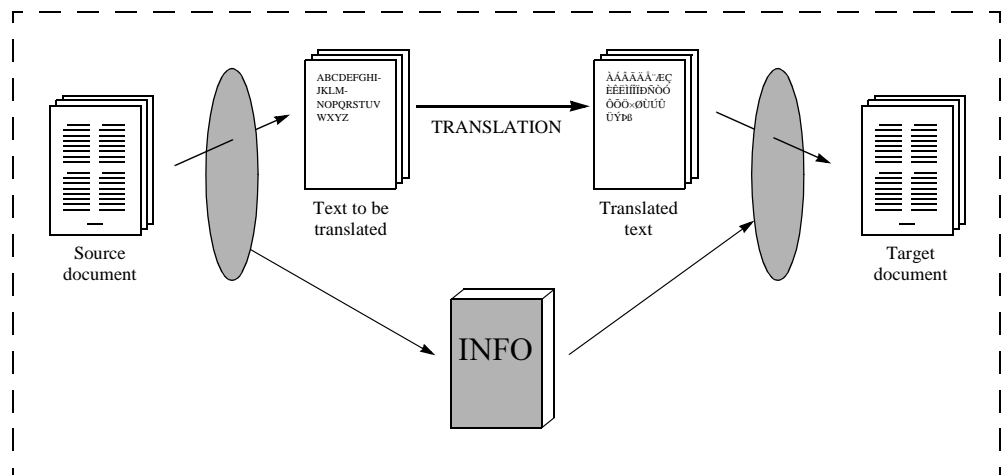


Figure 1: General process conceptio

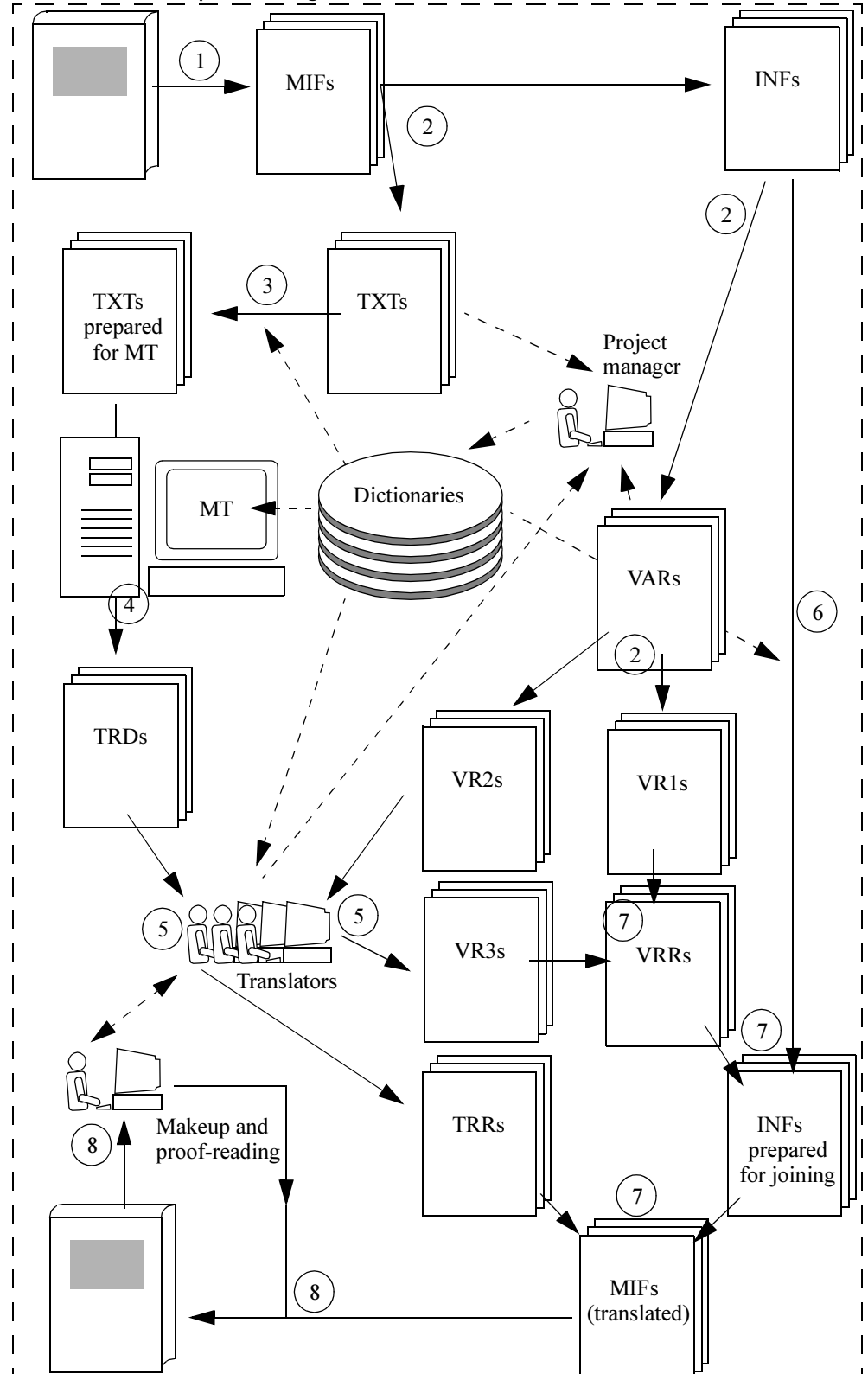
Solutions

Legend:

- Document flows
- - - Information flows and control data
- ① See appropriate process steps

Currently we use a documentation translation process as follows:

Bild 2: Detailed processing



Conversion

A document (or some documents) are converted to one of available text formats, to be further processed. Today we are able to handle RTF, MIF, Ileaf ASCII format, SGML, HTML, TROFF, and some other specific formats.

Below is the example of translating the EWSD document, done in Frame Builder 1.0 for Windows.

At the conversion step a "book", containing several documents, is saved in MIF format, using Frame Builder features.

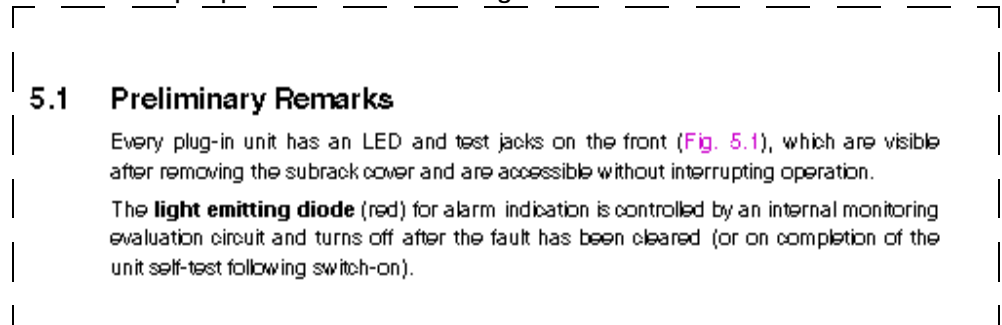
```
doc1. => doc1.mif
doc2. => doc2.mif
doc.idx => doc_idx.mif
doc.abr => doc_abr.mif
```

Filtering

The document is split into service and text parts, i.e. "filtered" (this is done for all MIF files).

```
doc1.mif => doc1.txt + doc1.inf
```

Below a sample part of a document is given



The resulting text (to be translated) is as follows:

Preliminary Remarks

<0>

Every plug-in unit has an LED and test jacks on the front (

<1>

), which are visible after removing the subrack cover and are accessible without interrupting operation.

The \$\$\$

<2>

light emitting diode

<3>

(red) for alarm indication is controlled by an internal monitoring evaluation circuit and turns off after the fault has been cleared (or on completion of the unit self-test following switch-on).

<4>

Some information in source document is stored as variables, i.e. named text fragments, used further via references by the variable name. (e.g., EWSD acronym). To process variables these are extracted.

```
doc1.inf => doc1.var
```

The resulting file contains all the variables, used in the document, even if they are not available in the text, and has a format as follows:

```
SystemName
EWSD
----
ShortDate
10/15/1997
----
Title
Operation Manual
----
Issue
***-7618
----
```

Not all variables are to be translated. For example, SystemName and ShortDate variables are not modified, the Title variable is to be translated, and the Issue variable is to be replaced in all documents by ***-5618. For the whole document the analysis of variables used is performed, and then to select non-modifiable variables we have to design a specific control file.

After processing, two files, namely doc1.vr1 and doc1.vr2, are created.

```
doc1.var => doc1.vr1 + doc1.vr2
```

Preprocessing	Preprocessing depends on the project, document type and its information nature. Here different operations can be performed such as protecting text not to be translated (so called masking), adapting text for machine translation and some other operations, which allow to accelerate or simplify the translation.
Machine Translation	Extracted text is processed by the machine translation system. doc1.txt (source)=> doc1.trd (target)
Post-editing	In general, this is conventional translation process, when the human translator receives a source text file, a target machine translated file, to be edited, and hard copy. In our example, the human translator is to edit the target text file and variables to be translated. <pre>doc1.trd => doc1.trr doc1.vr2 => doc1.vr3</pre>
Note:	More detailed information on the translation process, including terminology development and dictionary customization for machine translation system, is disclosed later in this article.
Post-processing	Similarly to preprocessing, the post-processing depends on the project. On this step some operations are performed such as font replacement, correcting styles, checking a document consistency and so on. Following is the list of operations, performed before unsplitting: font replacement, turning hyphenation off, changing margins and tabs in captions, chang-

ing paragraph autonumbering format, checking consistency for translated files, correcting cross-references, changing templates related to auto indexing.

Unsplitting

The target (post-edited) text file and previously stored service information are combined (unsplit) to form a source format (MIF).

Then we need to combine variables to obtain their complete set.

```
doc1.vr2+doc1.vr3 => doc1.vrr
```

Next, these new variables replace old ones, which are stored in service file.

```
doc1.inf + doc1.vrr => doc1.inv
```

After that, we should combine files and get a resulting MIF file.

```
doc1.trd + doc1.inv => doc1.mfr
```

Finally, we need to restore file names.

```
doc1.mfr => doc1
```

```
doc2.mfr => doc2
```

```
doc_idx.mfr => doc.idx
```

```
doc_abr.mfr => doc.abr
```

Make up and proof-reading

During this step the final make up is performed, as well as changing text, which was not extracted from the source file, for example, in pictures (in PCX, TIFF, JPEG etc). After doing a make up, the target document is printed and forwarded to corrector for proof-reading.

Terminology Development and Dictionary Customization for Machine Translation System

Before starting any translation, it is necessary to develop a terminology. For example, when developing telecommunication terminology we use different documents as follows:

- international recommendations (ITU-T, ETSI),
- available dictionaries,
- articles (from periodical magazines) with term explanation,
- explanatory dictionaries of terms and abbreviations from companies,
- documentation itself.

Note:

For translating a EWSD documentation while developing terminology the telecommunication glossary (for EWSD) created in SIEMENS was used, as well as different manuals to be translated such as Technical Descriptions, Operation Manuals, Maintenance Manuals, Emergency Manuals, Training Courses etc. Meanwhile, the customization of dictionaries for the Machine Translation system (PROMT' STYLUS) is performed. We enter to dictionaries a following information:

- general vocabulary (usually, already available in system general dictionary),
- basic terms for hardware & software, telecommunications, power supply and so on,
- names for subsystems, modules, units, processes, circuits, tests, services, and other specific information related to documentation to be

translated, e.g. frequently used expressions and phrases (microsegments).

Our experience shows that thanks to detailed dictionary customization we are able to:

- provide terminology consistency,
- control a style (it is possible to load standard expressions and phrases to the dictionary)
- simplify (significantly) human work of searching and selecting relevant terminology (today, our specialized telecommunication dictionary contains more than 80% of necessary terms).

Because of using the machine translation system following benefits are achieved:

- consistency management,
- translation capacity acceleration (at least, twice),
- significant minimization of manual work during post-editing (keyboard activities),
- new staff adaptation is simplified.

Using integral system features we generate (and regularly update) electronic text dictionaries, which can be invoked from word processor (these generated files are stored in server and updated by the dictionary manager). For instance, the basic dictionary of telecommunication and computer terms, customized for EWSD-related terminology, contains more than 30,000 entries (words and collocations).

Note that in addition to terminology development and dictionary customization, we create explanatory glossaries on different subjects (for example, today we have such dictionaries for ATM, ISDN, Intelligent Networks), which also can be invoked from the word processor.

Conclusions

Documentation translation process described above allows to:

- minimize significantly a manual work,
- provide a centralized management of terminology and translation style before and during translation (rather than during technical reviewing only),
- simplify and accelerate the translation and publishing process.

