

[From: Makoto Nagao (ed.) *Machine translation summit*. Tokyo: Ohmsha, 1989]

5 • 7 Translation of Engineering Documentation with METAL

Raymond Shah

Compulex, Computer Integrated Translations, Switzerland

Switzerland's mechanical and electrical industries export up to 90% of the products developed in the country. For these products all technical documentation has to be translated into English. Hence Siemens selected a Swiss company, Compulex in Zurich, to operate the first test facility outside Siemens to evaluate METAL.

Compulex is specialized in multilingual technical documentation. One major activity is the translation of complete customer documentation for complex control systems for leading Swiss companies. Another is that of setting up terminology for specialized fields. An example is the "Dictionary of Microelectronics & Micro Computers" produced for the Society of German Engineers.

The METAL pilot facility at Compulex has been in operation since February 1987. After a start-up period of roughly 3 months, the installation has successfully completed several translation jobs for Swiss industry, e.g. manuals and test reports as well as licence manufacturing documents. In the latter case, the first 300 pages of the 22,000-page job have shown that the quality of machine translation can reach a high standard, providing the system — in particular the system lexicon — is adequately tuned to the subject field and phraseology of a specific application.

Initial experience shows that machine translation of technical documentation has the following 3 main advantages over conventional methods:

- Lead time, i.e. the total time required for producing the documentation is drastically reduced. In practice, lead times of as short as 1 week for a 1,000-page document can be achieved.
- Consistency of terminology and wording of text is significantly improved. In the case of complex systems employing advanced electronics and computer technology, this improved consistency results in improved functionality of the documentation.
- Overall costs are significantly lower, particularly when one considers that no additional word processing is needed to attain the page format of the original document. Because about 30% of the total costs of a translated document are due to word processing, and a further 10 to 20% are due to terminology checks, overall costs are at least 40% lower.

As shown in Fig. 5-11 we have implemented METAL in a computer integrated environment.

Basically, the integration concept consists of 3 main modules: The METAL system with the modules text translation and formatting software and the third module comprising transformation software for pre- and post-editing.

Fig. 5-12 summarizes the major advantages of METAL in a computer integrated environment and Fig. 5-13 shows the hardware configuration at Compulex in Zurich.

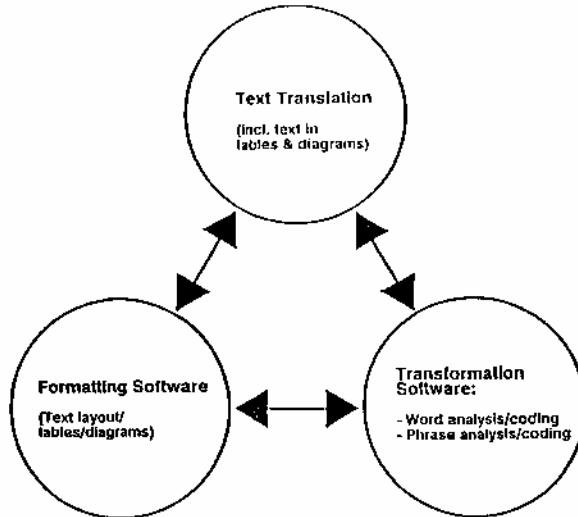
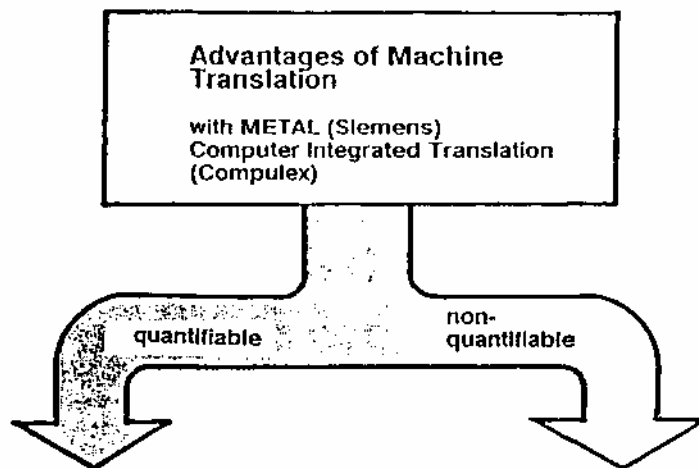


Fig. 5-11 Integration of numerous software modules in the METAL Computer Integrated Translation system



- | | |
|---|---|
| <ul style="list-style-type: none"> ● Lead time: very short (approx. 1 week/1000 pages) ● Lower overall costs due to automatic reformatting (approx. 30-40% savings) ● Lower translation costs due to controlled vocabulary (approx. 20%) | <ul style="list-style-type: none"> ● Unsurpassed terminology consistence ● Improved functionality due to consistent phraseology |
|---|---|

Fig. 5-12 Major advantages of machine translation with METAL-CIT

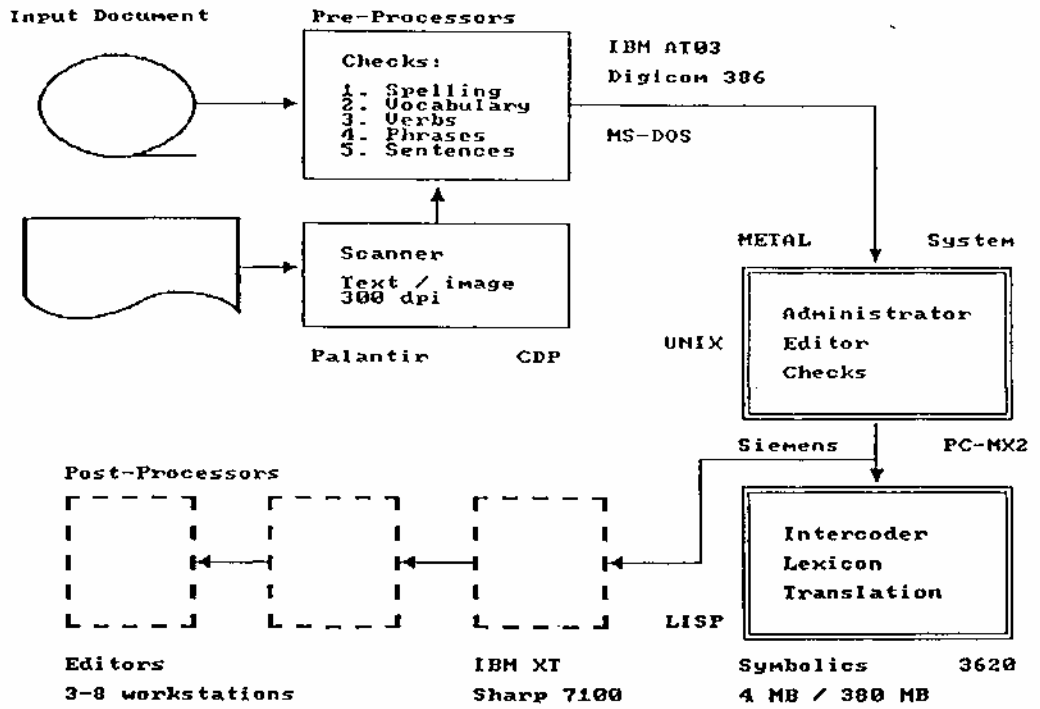


Fig. 5-13 Hardware configuration of METAL-CIT

On the basis of our initial experience we have set up a check list for our clients. As shown in Fig. 5-14, we recommend machine translation for all translation jobs with a high volume (upwards of 300 pages) and a high percentage of subject- specific terminology (upwards of 3,000 terms).

It is obvious that the actual cost of machine translation depends primarily on the ability of the system to produce an acceptable translation without extensive post-editing. Here is an example of a machine produced text without any editing. Because of the relatively simple phraseology of the source text, the raw translation is acceptable as it stands and needs no post-editing. In our experience less than 20% of the raw translation produced by METAL requires extensive post-editing.

CHECK LIST

Machine translation is economically viable providing the following conditions apply:

1. Extensive documentation employing consistent terminology and phraseology
2. Multiple variants of basic documents (e.g. manuals for a series of product models, test reports, tender documentation)
3. Texts with a high percentage of subject-specific terms
4. Highly structured documents, i.e. with a high proportion of tables and diagrams

Analysis Steps:

1. ABC analysis of required translations (incl. variants of same document)
 - A - High-volume documents (usually 60-70% of total)
 - B - Medium-volume documents (usually 20-30% of total)
 - C - Low-volume documents (usually 10-20% of total)

Volume is defined as size of basic document times number of versions needed per year

2. In the case of A and B documents, the suitability of machine translation employing METAL/CIT (Computer Integrated Translation) can be estimated with the help of the following diagram:

Terminology (No. of terms)	<1000	ca.3000	>5000
Volume -low (<100p.)	MAN	MAN/CIT	CIT
- medium (ca. 300 p.)	MAN/CIT	CIT	CIT
-high (>1000 p.)	CIT	CIT	CIT

MAN = manual translation
CIT = Computer Integrated Translation

Fig. 5-14 Check list for selecting documents suitable for machine translation

METAL Translation (Example 1)

Source Text:

Prüfanweisung für Leiterplatten
Diese Vorschrift gilt für das Anbringen von Bezeichnungen auf Baugruppen und Leiterplatten. Vorwiegend handelt es sich um Schutz Erde-Zeichen.

Anforderungen:

Die Loetoesen sind mit Schmirgeltuch an geeigneter Stelle zu markieren. Bei Klebstellen ist es notwendig, den Isolierkörper zu entfernen. Mangelhaft ausgetrocknete Klebstellen können einen Ausfall der Bauelemente verursachen.

Vorrichtungen und Werkzeuge:

Für das Abtrennen von Glasdurchführungen wird ein Schneidgerät benötigt. Die Kurzverbindungen sind mittels Lot zu sichern.

Target Text (unedited METAL translation)

Test procedure for PCBs

This procedure applies to the marking of designations on assemblies and on PCBs. Mostly it concerns protective ground symbols.

Requirements:

The solder tags are to be processed with abrasive cloth. At bonding points, it is necessary to remove the insulating part. Imperfectly dried bonding points can cause a failure of the components.

Fixtures and tools:

A cutting device is required for cutting off glass lead-in bushings. The jumpers are to be secured by means of solder.

Fig. 5-15 Example of a raw translation which requires no post-editing

Job: Crypto Acceptance Test UC30-50-78		BBCUCAT3.WRD	Customer: BBC		
		In: 87-07.24 Out: 87-07.28	Pages: 24 Diskette: 21 kB		
Format: Word Diskette					
- Text: 21 kB					
- Tables: 3					
- Diagrams: 3 = 380 kB (300 dpi)					
Step No.	Task	MS-DOS	Time (min)		
			SINIX	LISP	
1.	File format check	8	-	-	
2.	Spelling check	28	-	-	
3.	Vocabulary check	11	-	-	
4.	New words: translate, pre-code word list	43	-	-	
5.	Enter word list in Lexicon	-	3	5	
6.	Code key words individually	-	-	26	
7.	Check/modify sentence structure, phrases, etc.	15	-	-	
8.	Transfer to SINIX; check structure, tables, extract format data	-	7	-	
9.	Translate text	-	3	122	
10.	Generate mix file (Germ/Eng)	-	4	-	
11.	Reformat file	-	5	-	
12.	Transfer file to MS-DOS	-	3	-	
13.	Post-edit file	210	-	-	
Sub totals.....		315	25	153	
Total time.....		493 min			
Report on Editing Operation:					
1. Spelling check: 25 spelling errors + 7 grammatical errors					
2. Format: 9 hard hyphens eliminated					
3. Vocabulary: 185 words					
4. Sentence construction:					
- Brackets used for complete sentences: 3					
- Bracketted phrases interrupting sentences: 8					
- Long sentences (> 20 words): 16					
5. Pre-editing:					
- Bracketted sentences/phrases: 11					
- Long sentences: 16					
6. Post-editing:					
- Sentences completely re-written: 21					

Fig. 5-16 Job sheet for a 24-page document requiring extensive post-editing

The job sheet shown in Fig. 5-16 details the steps carried out in the computer integrated translation system for a 24-page document which required considerable post-editing work (the job sheet shows 210 minutes were needed). Yet the total job time was only

slightly over 8 hours. Since an average technical translator produces roughly 8 pages/day, even in this extreme case machine translation with METAL including post-editing is 200% more productive.

Case Study: METAL-CIT vs. Conventional Methods		
Producing English version of Licence Manufacturing Documents		
Package A: 16,000 p. Process Sheets (8 lines/page, i.e. total 128,000 lines)		
- Draft translation		
manual (320 lines/day/translator)	400	
CIT (200 pages/day)	80 days	
- Editing		
manual (720 lines/day/editor).....	178	
CIT (720 lines/day/editor)	178 days	
- Word processing		
manual (all 16,000 p.)	320	
CIT	0 days	
- Total time		
manual method	898	= 100%
CIT	258 days	= 29%
- Costs		
manual method	320	= 100%
CIT	190 U.	= 60%
Savings: 71% time savings / 40% cost savings		
Package B: 6,000 p. Test Specifications & Procedures (20 lines/page, i.e. total 120,000 lines)		
- Draft translation		
manual (240 lines/day/translator)	500	
CIT (200 pages/day)	30 days	
- Editing		
manual (720 lines/day/editor)	167	
CIT (720 lines/day/editor)	334 days	
- Word processing		
manual (all 6,000 p.)	188	
CIT	20 days	
- Total time		
manual method	855	= 100%
CIT	384 days	= 45%
- Costs.		
manual method	450	= 100%
CIT.....	265 U.	= 59%
Savings: 55% time savings / 41% cost savings		

Fig. 5-17 Case study METAL-CIT vs. conventional methods for producing the English version of licence manufacturing documents

Finally, Fig. 5-17 demonstrates the economics of machine translation of complete technical documentation on the basis of an actual case study. This case study concerns all documents required for licence manufacturing a complex control system. The documents involved can be split up into two packages: Package A with 16,000 pages of process sheets, and

Package B with 6,000 pages of test specifications and procedures. Our estimates, based on 300 pages of sample texts actually processed with METAL-CIT, indicate that the system can achieve time savings of over 70% (Package A) and roughly 55% (Package B) as well as cost savings of roughly 40% in both cases.