[*Translating and the computer 11*. Proceedings of a conference … 16-17 November 1989, ed. Catriona Picken (London: Aslib, 1990)]

METAL – machine translation in practice

Patrick Little

Philips Kommunikations Industrie

In 1986, Philips Kommunikations Industrie AG, one of the major suppliers of the German PTT, decided to increase its export activities considerably in the area of public telecommunications systems; up until then exports had accounted for less than 2 per cent, of turnover. This decision was made in view of the completion of the initial development phase for ISDN and the consequent long-term fall in orders from the Deutsche Bundespost, our main customer. The deregulation of the European PTTs and the setting up of the Single European Market in 1992 were also decisive factors.

In the area of export, as with all business transactions, the customer always comes first. The customer nowadays is not only interested in buying equipment, but also in actually starting up production in his or her own country through projects such as technology transfers or joint ventures undertaken in cooperation with the suppliers.

However, as almost all the product and manufacturing documentation at Philips Kommunikations Industrie was available only in German, this had to be translated into English as quickly as possible. The first project encompassed approximately 6,000 pages, which had to be translated using standardised terminology.

After a careful analysis of the translation systems available on the market, the METAL system (Machine Evaluation and Translation of Natural Languages), developed by Siemens, was chosen. Siemens worked with the University of Austin, Texas, for over 10 years on developing the system, and just as we entered into negotiations, were looking for a partner in industry in order to conduct a pilot test. PKI had already been negotiating with other machine translation firms. These negotiations, however, had repeatedly broken down due to unacceptably high costs or a lack of user autonomy.

Unlike online systems, METAL is a system which can be installed and expanded within the firm itself. The hardware configuration is preallocated by Siemens AG (see Figure 1) and consists of a Symbolics 3620 central processing unit with a 9 Mbit/s memory and an MX 300/20 central processing unit with an 8 Mbit/s memory, to which 12 work stations can be connected. We have one console connected to the Symbolics central processing unit and 10 Sinix work stations in operation. A variety of input and output facilities, i.e. magnetic tape, disks, laser printers, etc. are used. In addition, we are in the process of linking

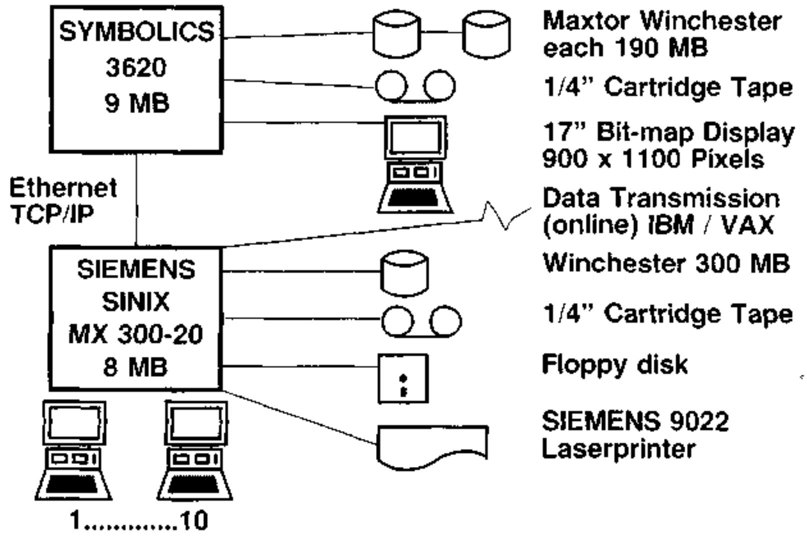
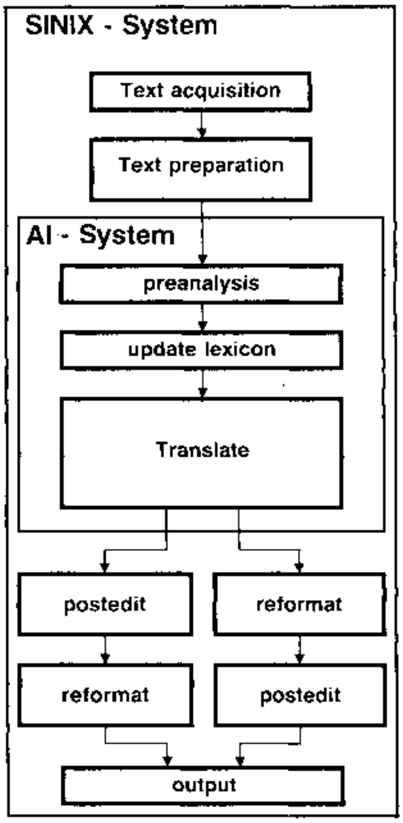


Figure 1. METAL configuration used at Philips

our METAL LAN to the IBM and VAX networks within the firm, in order to gain direct access to various databases, containing the data to be translated.

Figure 2 illustrates which tasks are carried out by the respective central processing units, and how a translation is run through METAL. After the text is input, it is split up into two categories: layout specific font information and text, including word specific font information. The former is filed on a data template. The latter is sorted into translation units and undergoes a consistency check. The shortest translation unit would be a single word headline, the longest a complete sentence. These pre-editing processes are carried out on the MX 300. On completion, the text is sent to the Symbolics central processing unit, where a lexicon search is carried out. This produces a file of unknown terms listed in alphabetical order. This function is also useful for texts which are not going to be translated using the system (such as press statements, etc.), as these lists can be researched quickly and methodically for terminology. After the terminology has been agreed upon, it is then input into the respective lexica and becomes standard.

The technical data sheet, shown on Figure 3, is a typical example of the documentation which is put through METAL. This type of document will be the main input in the METAL system in the coming year. Since technical data sheets are a mixture of both text and graphics, METAL is a particularly suitable system, as it has the advantage of maintaining the format of the text, or in other words, the original layout. Before METAL begins to process the document, the graphics must be removed by a special filter which is being developed at the moment. I shall return to this point later. If the technical data sheets, which have been produced in Interleaf, were to be translated manually, each translator would have to be equipped with top-grade Interleaf software and the appropriate



Text Input
- data transmission (online)
- input via storage medium (floppy disk, magnet tape, font reader)
- (keyboard input)

Text Format
- extracting sentences
- treating special formats
  - diagrams
  - tables, flow-charts etc.

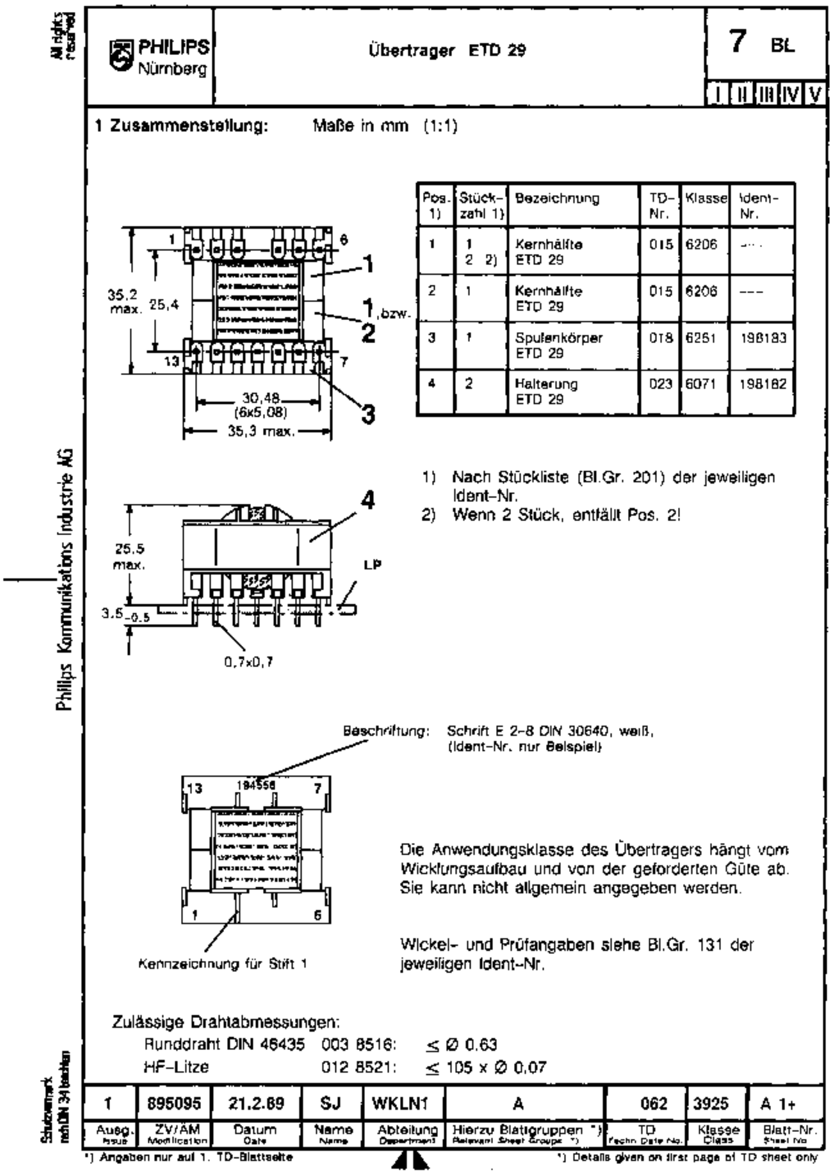
Lexicon
- check system lexicon for completeness
- Intercoder

Translation
- translate text

Postediting
- restore original format
- postedit format and machine translated text
- printed output
- transfer to other text processing systems, publishing systems

Figure 2. Translation procedure using METAL

hardware just to insert the translated text. Consistency in the area of terminology could not be guaranteed either; a problem which does not arise with METAL. The input document (here our technical data sheet) is sorted according to graphics and text in an Interleaf-METAL filter and the graphics are stored on a template. The text is then processed further in the METAL system. As mentioned above, the input or source text is deformatted by a program which extracts both tables and graphics, as well as the words to be translated. Two types of data are produced by the deformatting process, one of which is the template file with format information. The other is the MTI file (see Figure 4), the input file for the translation. As far as pre-editing the text is concerned, it has



All rights reserved

PHILIPS Nürnberg

Übertrager ETD 29

7 BL

I II III IV V

1 Zusammenstellung: Maße in mm (1:1)

Pos. 1)	Stück-zahl 1)	Bezeichnung	TD-Nr.	Klasse	Ident-Nr.
1	1 2 2)	Kernhälfte ETD 29	015	6206	---
2	1	Kernhälfte ETD 29	015	6206	---
3	1	Spulenkörper ETD 29	018	6251	198183
4	2	Halterung ETD 29	023	6071	198182

1) Nach Stückliste (Bl.Gr. 201) der jeweiligen Ident-Nr.
2) Wenn 2 Stück, entfällt Pos. 2!

Beschriftung: Schrift E 2-8 DIN 30640, weiß, (Ident-Nr. nur Beispiel)

Kennzeichnung für Stift 1

Die Anwendungsklasse des Übertragers hängt vom Wicklungsaufbau und von der geforderten Güte ab. Sie kann nicht allgemein angegeben werden.

Wickel- und Prüfangaben siehe Bl.Gr. 131 der jeweiligen Ident-Nr.

Zulässige Drahtabmessungen:

Runddraht DIN 46435	003 8516:	≤ Ø 0.63
HF-Litze	012 8521:	≤ 105 x Ø 0,07

1	895095	21.2.89	SJ	WKLN1	A	062	3925	A 1+
Ausg. Issue	ZV/ÄM Modification	Datum Date	Name Name	Abteilung Department	Hierzu Blattgruppen *) Relevant Sheet Groups *)	TD Techn Data No.	Klasse Class	Blatt-Nr. Sheet No

*) Angaben nur auf 1. TD-Blattseite

*) Details given on first page of TD sheet only

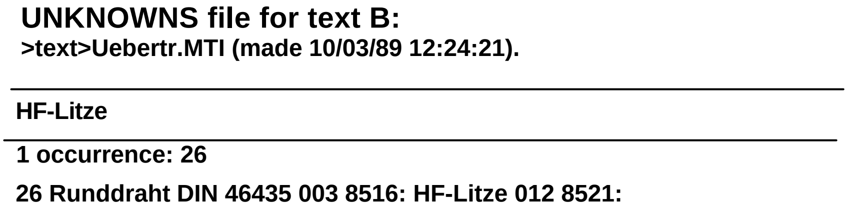
Philips Kommunikations Industrie AG

Schutzvermerk nach DIN 34 beachten

Figure 3. Typical example of the documentation put through METAL

been decided that all the preparation required up to the MTI file stage can be done by a typist and not necessarily by a technical translator.

Our technical data sheet, in the form of the MTI file, is finally sent to the Symbolics central processing unit so that it can be checked for new terminology.



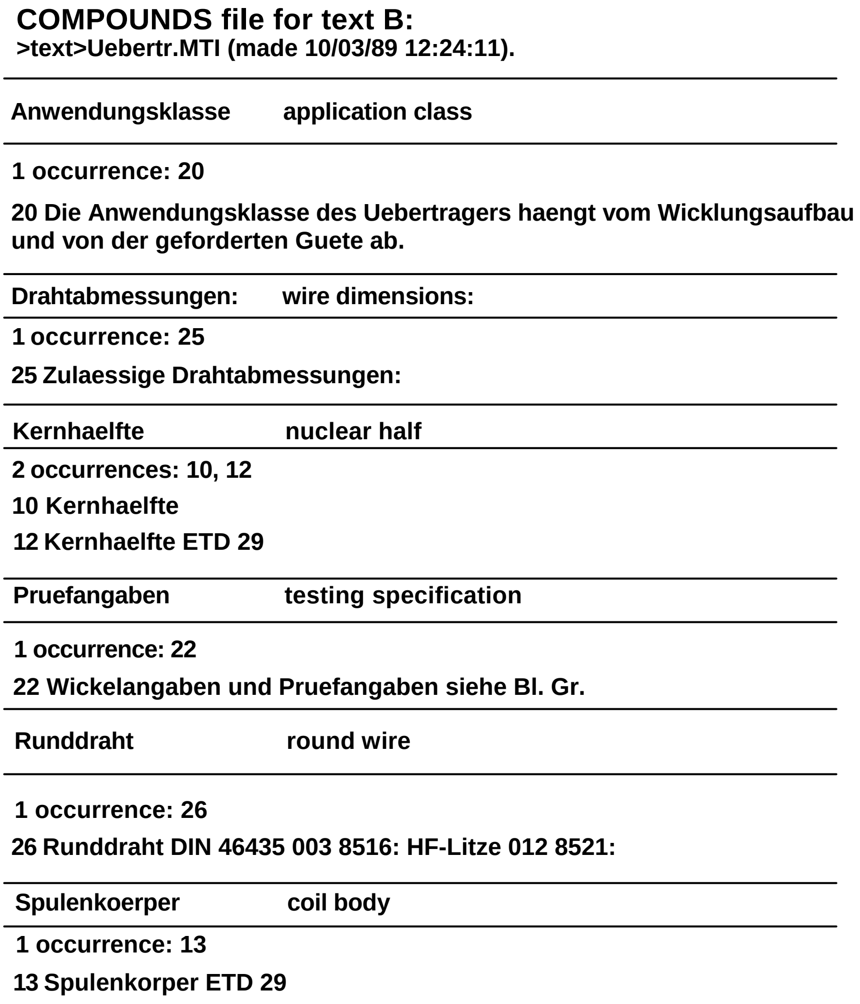
UNKNOWNS file for text B:
>text>Uebertr.MTI (made 10/03/89 12:24:21).

HF-Litze

1 occurrence: 26

26 Runddraht DIN 46435 003 8516: HF-Litze 012 8521:

Figure 4. Unknowns file



COMPOUNDS file for text B:
>text>Uebertr.MTI (made 10/03/89 12:24:11).

Anwendungsklasse application class

1 occurrence: 20

20 Die Anwendungsklasse des Uebertragers haengt vom Wicklungsaufbau und von der geforderten Guete ab.

Drahtabmessungen: wire dimensions:

1 occurrence: 25

25 Zulaessige Drahtabmessungen:

Kernhaelfte nuclear half

2 occurrences: 10, 12

10 Kernhaelfte

12 Kernhaelfte ETD 29

Pruefangaben testing specification

1 occurrence: 22

22 Wickelangaben und Pruefangaben siehe Bl. Gr.

Runddraht round wire

1 occurrence: 26

26 Runddraht DIN 46435 003 8516: HF-Litze 012 8521:

Spulenkoerper coil body

1 occurrence: 13

13 Spulenkorper ETD 29

Figure 5. Compounds file

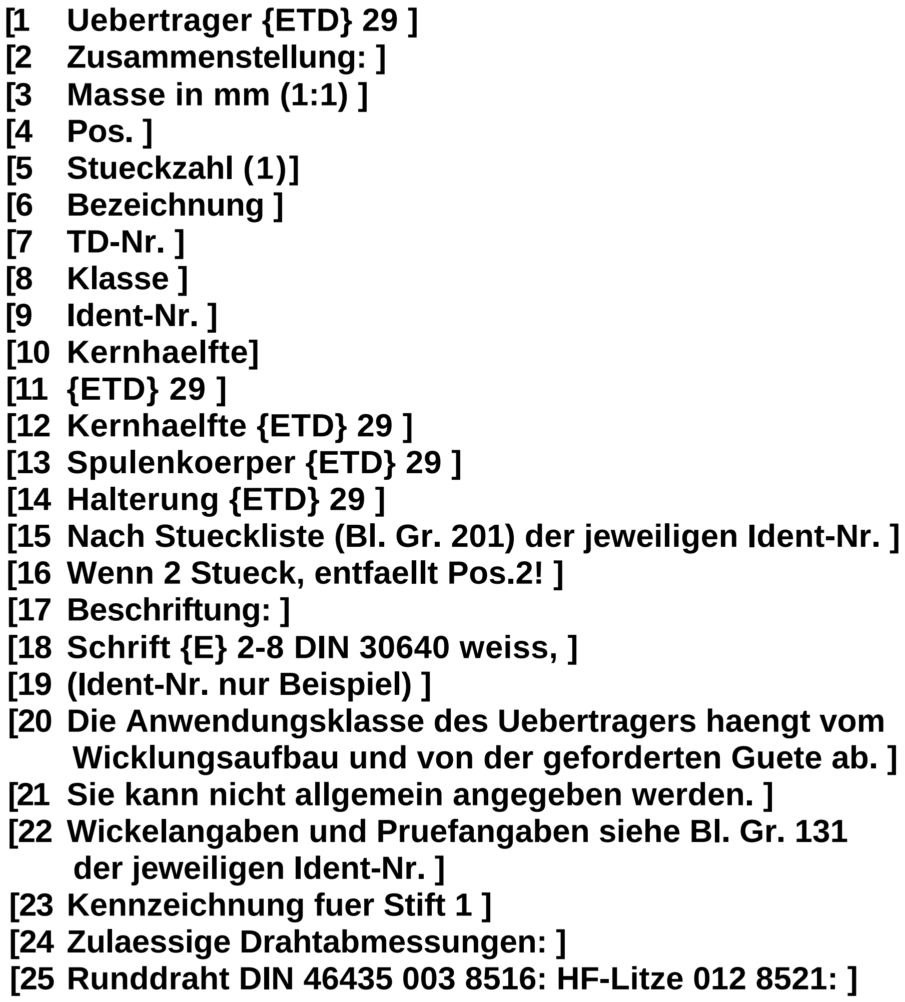
```
 "TEXT GLOSSARY file for text B:
 >text>Uebertr.MTI (made 10/03/89 12:23:52)."

[Context: 20]
Guete:
                    → quality factor (TEL)
[Context: 14]
Halterung:
                    → mount (CTV)
[Contexts 15 19 23]
Ident-Nummer:
                    → I.D. number (CTV)
[Context: 24]
Kennzeichnung:
                    → identifier (DP)
[Context: 31]
Masse:
                    → ground (CTV) / dimension (CTV) /
                      mass (CTV)
[Context: 24]
Stift:
                    → pin (CTV)
[Context: 15]
Stueckliste:
                    → parts list (CTV)
[Context: 1]
Uebertrager:
                    → transformer (CTV)
```

Figure 6. Glossary of known technical words

We get three different sets of data from this: (a) the list of unknown words (see Figure 4), (b) the list of compounds (see Figure 5), for which METAL suggests a translation, using as a basis the information already coded, in conjunction with (c) the glossary or list of known technical words, which are translated and listed in the appropriate technical category (see Figure 6). The compound list has often been a source of amusement, for instance when words such as ‘Darmstadt’ and ‘Tastverhältnis’ are translated as ‘bowel town; and ‘touching relationship’ respectively.

After the terminology has been researched and standardised with the help of these three lists (and at a later stage coded), the MTI file (Figure 7), is sent to the Symbolics central processing unit once more to be translated. Hence we get the

MTO file (Figure 8), which like the MTI file, comprises the text and the word specific font information. In order to get the text back to its original form and layout, the MTO and template file of the output text must be combined; a step which gives us the Ref file (Figure 9). There are several ways of post-editing the text, such as in the MTI file, in one window, or using two windows and making a comparison between the MTO and MTI files. Another possibility is to post-edit the text in the Mix file with alternating translation units, target and source language, or in the Ref data which contains the original layout.



```
[1   Uebertrager {ETD} 29 ]
[2   Zusammenstellung: ]
[3   Masse in mm (1:1) ]
[4   Pos. ]
[5   Stueckzahl (1)]
[6   Bezeichnung ]
[7   TD-Nr. ]
[8   Klasse ]
[9   Ident-Nr. ]
[10  Kernhaelfte]
[11  {ETD} 29 ]
[12  Kernhaelfte {ETD} 29 ]
[13  Spulenkoerper {ETD} 29 ]
[14  Halterung {ETD} 29 ]
[15  Nach Stueckliste (Bl. Gr. 201) der jeweiligen Ident-Nr. ]
[16  Wenn 2 Stueck, entfaellt Pos.2! ]
[17  Beschriftung: ]
[18  Schrift {E} 2-8 DIN 30640 weiss, ]
[19  (Ident-Nr. nur Beispiel) ]
[20  Die Anwendungsklasse des Uebertragers haengt vom
      Wicklungsaufbau und von der geforderten Guete ab. ]
[21  Sie kann nicht allgemein angegeben werden. ]
[22  Wickelangaben und Pruefangaben siehe Bl. Gr. 131
      der jeweiligen Ident-Nr. ]
[23  Kennzeichnung fuer Stift 1 ]
[24  Zulaessige Drahtabmessungen: ]
[25  Runddraht DIN 46435 003 8516: HF-Litze 012 8521: ]
```

Figure 7. MTI file

We would like to turn our attention next to the lexicon and coding. First of all, however, we shall take a look at the tag hierarchy (Figure 10). The three lower levels (function words, general vocabulary and common technical vocabulary) form the three basic lexicon modules, which are delivered as standard by Siemens to the user. It is possible to add further technological categories or in other words, define new modules, as they are needed. For each translation run through METAL, the specialised areas are selected in the order required and put onto the docket in this order. The specialised lexicon areas are checked for specific words right down to the most general.

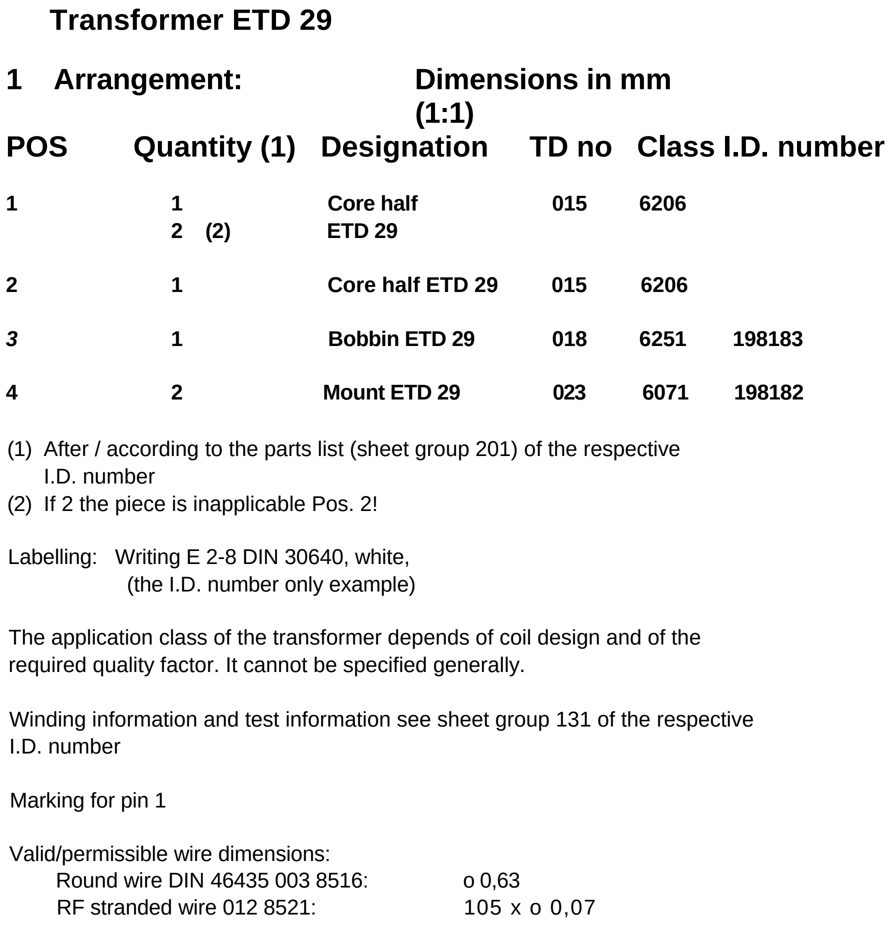
The METAL lexicon is in modular form and consists of two monolingual dictionaries for the source and target language respectively, as well as one bilingual dictionary for the transfer compounds. The monolingual lexica contain morphological, syntactic and semantic details and to a certain extent are comparable with an Oxford or Duden dictionary. This information is necessary during the actual translation process, for analysing the language of the source

```
[1   Transformer ETD 29 ]
[2   Arrangement: ]
[3   Dimensions in mm (1:1) ]
[4   POS ]
[5   Quantity (1)]
[6   Designation ]
[7   TD no ]
[8   Class ]
[9   I.D. number ]
[10  Core half]
[11  ETD 29 ]
[12  Core half ETD 29 ]
[13  Bobbin ETD 29]
[14  Mount ETD 29]
[15  After / according to the parts list (sheet group 201)
      of the respective I. D. number ]
[16  If 2 the piece is inapplicable Pos. 2! ]
[17  Labelling:]
[18  Writing E 2-8 DIN 30640, white, ]
[19  (the I.D. number only example) ]
[20  The application class of the transformer depends of coil
      design and of the required quality factor. ]
[21  It cannot be specified generally. ]
[22  Winding information and test information see sheet group
      of the respective I.D. number ]
[23  Marking for pin 1 ]
[24  Valid/permissible wire dimensions: ]
[25  Round wire DIN 46435 003 8516:
      RF stranded wire 012 8521: ]
```

Figure 8. MTO file

text and finding the appropriate translation for the target text. The bilingual lexicon includes, in addition to the respective translations, examples of context, which have been established by so-called ‘tests’. When required, additional details relating specifically to the product or customer in question, or a choice of dialect (UK/US), are available. The bilingual dictionary can be compared to the Langenscheidt or Collins dictionary. The three dictionaries delivered by Siemens contain approximately 30,000 entries each.

Figure 10 is followed by a further three figures which show the screen of the Intercoder, which is the interactive coding system. In Figure 11, the Intercoder is shown without any entries on the screen, as it appears at the start of a coding session. In Figure 12, we can see that on the lower part of the screen in the Coding History, some entries have already been retrieved from the database.



Transformer ETD 29

1 Arrangement: Dimensions in mm (1:1)

POS	Quantity (1)	Designation	TD no	Class	I.D. number
1	**1** **2 (2)**	**Core half** **ETD 29**	**015**	**6206**	
2	**1**	**Core half ETD 29**	**015**	**6206**	
3	**1**	**Bobbin ETD 29**	**018**	**6251**	**198183**
4	**2**	**Mount ETD 29**	**023**	**6071**	**198182**

(1) After / according to the parts list (sheet group 201) of the respective I.D. number
(2) If 2 the piece is inapplicable Pos. 2!

Labelling: Writing E 2-8 DIN 30640, white,
(the I.D. number only example)

The application class of the transformer depends of coil design and of the required quality factor. It cannot be specified generally.

Winding information and test information see sheet group 131 of the respective I.D. number

Marking for pin 1

Valid/permissible wire dimensions:
Round wire DIN 46435 003 8516: o 0,63
RF stranded wire 012 8521: 105 x o 0,07

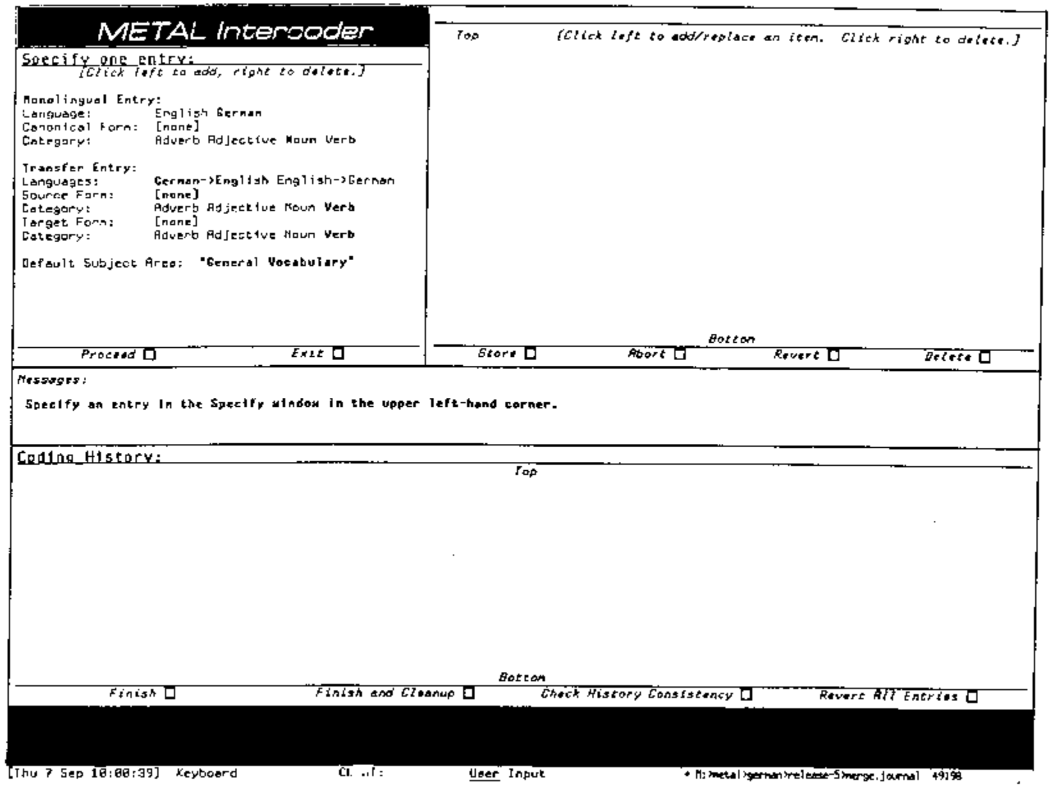
Figure 9. Ref file

Here the English verb 'test' is being modified in the edit mode. Figure 13 illustrates a transfer entry where the context, the so-called 'feature test', is specified as well as the subject area. As can be seen in the dark area towards the bottom of Figure 13, the Intercoder has an interactive coding assistance program which provides explanatory information. A further merit of the Intercoder is that all important classes of words can be coded and a default automatically appears, which makes the coding easier. In addition, it is a menu-based system and therefore user-friendly. Thorough and consistent coding is a fundamental part of achieving standardised terminology. With time, the amount of coding done will diminish, as documentation such as technical data sheets, for example, only contain a limited vocabulary. The founding of a METAL users' club in the near future will facilitate the exchange of lexica among the users. This automated procedure, the 'merge dictionary' function, is supported by the system and will prove important at a time when terminology is becoming increasingly expensive.

```
Command: Show Tags

  FW Function Words
      GV General Vocabulary
           CTV Common Technical Vocabulary
                COMM commercial documentation
                DP Data Processing
                     DP-HW Computer Hardware
                     DP-SW Computer Software
                     DP-TR   Data Processing - Data
                            Transmission
                       TEL Telecommunications
                            Tel-PW Pruefwesen
```

Figure 10. Tag hierarchy

The major drawback we have suffered over the past year is the lack of compatibility with other documentation systems (word processing and desktop publishing systems). We have decided to have direct interfaces developed which will overcome the problem in the short term and have been assured that Siemens will introduce an interface in accordance with the ODIF standard by 1991 which will mean that all the systems supporting this standard can be used in conjunction with METAL without loss of data. The sooner METAL becomes fully compatible with all widely used documentation systems, the better.



```
Specify one entry:
        [Click left to add, right to delete.]

Monolingual Entry:
Language:          English German
Canonical Form:    [none]
Category:          Adverb Adjective Noun Verb

Transfer Entry:
Languages:         German->English English->German
Source Form:       [none]
Category:          Adverb Adjective Noun Verb
Target Form:       [none]
Category:          Adverb Adjective Noun Verb

Default Subject Area:  "General Vocabulary"
```

[Click left to add/replace an item. Click right to delete.]

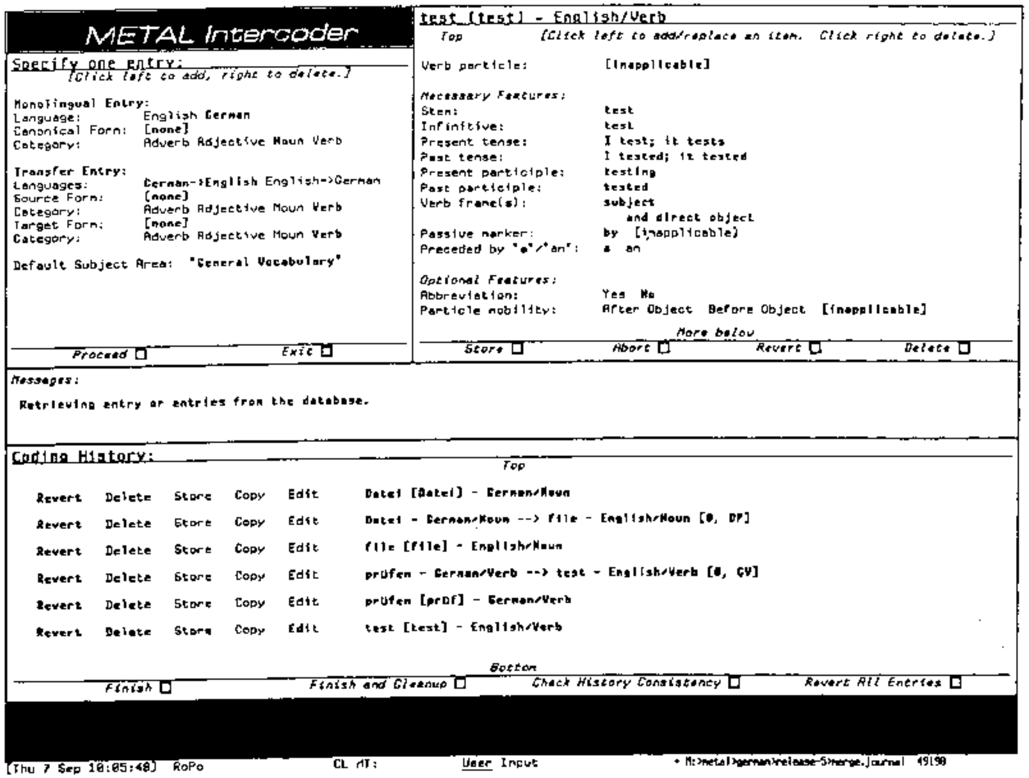
Messages:

Specify an entry in the Specify window in the upper left-hand corner.

Coding History:

```
[Thu 7 Sep 10:00:39]  Keyboard        CL ..I:        User Input        • M:>metal>german>release-5>merge.journal  49198
```

Figure 11. Intercoder screen at the start of a coding session



```
Specify one entry:
        [Click left to add, right to delete.]

Monolingual Entry:
Language:            English German
Canonical Form:      [none]
Category:            Adverb Adjective Noun Verb

Transfer Entry:
Languages:           German->English English->German
Source Form:         [none]
Category:            Adverb Adjective Noun Verb
Target Form:         [none]
Category:            Adverb Adjective Noun Verb

Default Subject Area:  "General Vocabulary"
```

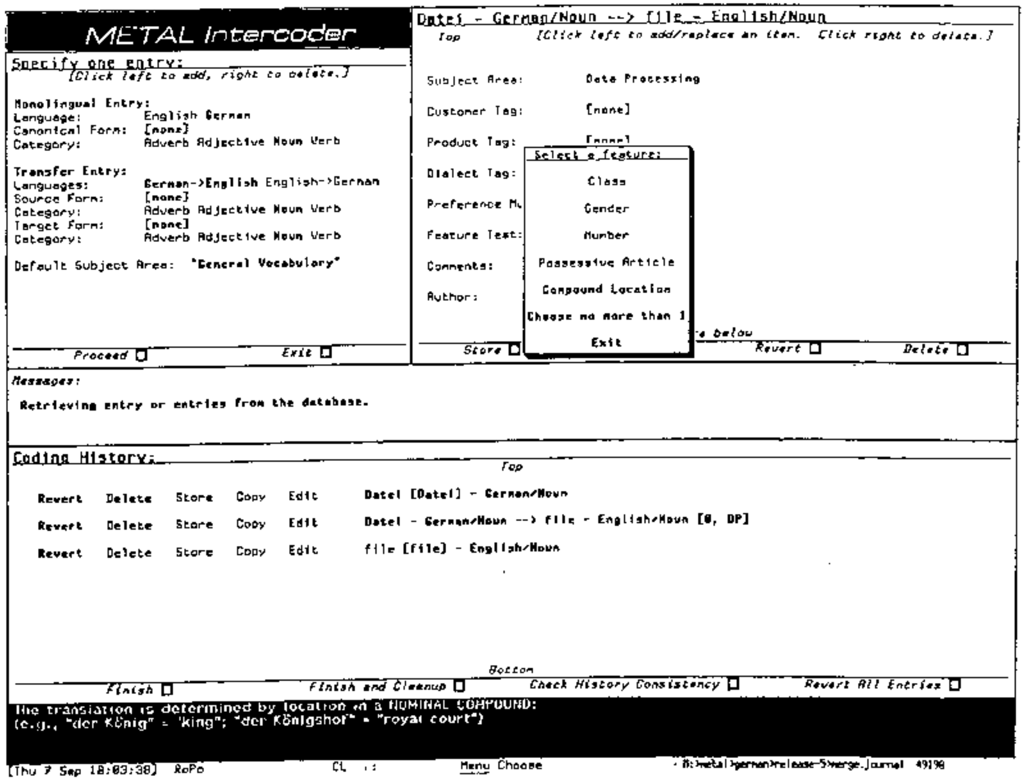
```
test [test] - English/Verb
  Top          [Click left to add/replace an item.  Click right to delete.]

Verb particle:             [inapplicable]

Necessary Features:
Stem:                      test
Infinitive:                test
Present tense:             I test; it tests
Past tense:                I tested; it tested
Present participle:        testing
Past participle:           tested
Verb frame(s):             subject
                              and direct object
Passive marker:            by  [inapplicable]
Preceded by "a"/"an":      a  an

Optional Features:
Abbreviation:              Yes  No
Particle mobility:         After Object  Before Object  [inapplicable]
                                  More below
```

```
Messages:
 Retrieving entry or entries from the database.
```

```
Coding History:
                                   Top

 Revert  Delete  Store  Copy  Edit    Datei [Datei] - German/Noun
 Revert  Delete  Store  Copy  Edit    Datei - German/Noun --> file - English/Noun [0, DP]
 Revert  Delete  Store  Copy  Edit    file [file] - English/Noun
 Revert  Delete  Store  Copy  Edit    prüfen - German/Verb --> test - English/Verb [0, GV]
 Revert  Delete  Store  Copy  Edit    prüfen [prüf] - German/Verb
 Revert  Delete  Store  Copy  Edit    test [test] - English/Verb

                                   Bottom
```

```
[Thu 7 Sep 10:05:48]  RoPo          CL AT:          User Input          + M:>metal>german>release-5>merge.journal  49190
```

Figure 12. Intercoder screen after some entries have been retrieved from the database



Specify one entry:
[Click left to add, right to delete.]

Monolingual Entry:
Language: English German
Canonical Form: [none]
Category: Adverb Adjective Noun Verb

Transfer Entry:
Languages: German->English English->German
Source Form: [none]
Category: Adverb Adjective Noun Verb
Target Form: [none]
Category: Adverb Adjective Noun Verb

Default Subject Area: "General Vocabulary"

Datei - German/Noun --> file - English/Noun

[Click left to add/replace an item. Click right to delete.]

Subject Area: Data Processing
Customer Tag: [none]
Product Tag: [none]
Dialect Tag:
Preference M.
Feature Test:
Comments:
Author:

Messages:

Retrieving entry or entries from the database.

Coding History:

```
Revert  Delete  Store  Copy  Edit   Datei [Datei] - German/Noun
Revert  Delete  Store  Copy  Edit   Datei - German/Noun --> file - English/Noun [0, DP]
Revert  Delete  Store  Copy  Edit   file [file] - English/Noun
```

The translation is determined by location in a NOMINAL COMPOUND:
(e.g., "der König" = "king"; "der Königshof" = "royal court")

[Thu 7 Sep 18:03:38] RoPo CL : Menu Choose · B:>metal>german>release-5>merge.Journal 49198

Figure 13. Transfer entry – 'feature test' specified as well as subject area

When this problem is overcome and when large volumes of complex technical documents are to be translated, as is the case at Philips Kommunikations Industrie, the METAL system is clearly a rational and effective solution to the problem.

AUTHOR

Patrick Little, Project Manager, Export Documentation, Philips Kommunikation Systems Industrie AG, Public Communication Systems, Thurn-und-Taxis-Straße 10, D-8500, Nürnberg 10, Federal Republic of Germany.