[From: *Translating and the Computer*, B.M. Snell (ed.), North-Holland Publishing Company, 1979]

IMPLEMENTING MACHINE AIDS TO TRANSLATION

Eberhard H. Tanke

Language Services Department, Siemens AG, Munich

The machine aids to translation developed by the Language Services Department of Siemens AG are intended solely as tools for the translation of special-subject material. The TEAM system with its associated programs provides the translator with terminology and phraseology, with target-language equivalents and examples of their usage in contexts, and with definitions.

Since text processing and editing systems are also valuable aids in translating, it is a logical step to combine them with machine aids to translation. Translators working on their own or in small translation bureaus can derive more or less equal benefit from the TEAM system by using such automatic text processors, or by cooperating with larger translation services.

1. General considerations.

The computer applications discussed here, whose function is to aid the human process of translation, reflect the multilingual realities of a large transnational concern. Translation is spoken of in this context in a broader sense than is generally understood by the term. It encompasses all the processes entailed in conveying the meaning of texts from one natural language to another, including purely cerebral processes that find no expression in writing. A description is given of all computer application that serve to improve special-language communication within a multinational, multilingual concern, and between the concern and its customers and partners.

The core of this concept is the TEAM* program system developed by the Language Services Department of Siemens AG.

Within this system, the problems associated with preparing special-language entities for use in the translation process - on the one hand by standardizing special language, and on the other by doing the opposite, viz., by adapting special language to the particular characteristics of the addressee's language usage and language competence - play as Important a role as questions concerning the concept or meaning of terms and the resolution of ambiguities.

Although special vocabulary often accounts for more than 50%

***TEAM is an acronym for Terminology Evaluation and Acquisition Method.**

of a specialized text, it is obvious that special-subject languages differ from general language not only in their vocabularies, but also in their morphology, syntagmatic structures, phraseology, and, to some extent, their idiom. These features must therefore be considered in designing, developing, and implementing machine aids to translation. They are as essential to programming as to the acquisition of special-language data.

Machine aids to translation - in the broader sense defined at the outset – must take into account two extremes of language and subject-matter competence on the part of users, as well as all gradations in between. One extreme is personified by the translator who exhibits a high level of competence in the source and target languages, but possesses little or no knowledge of the special subject area. At the opposite pole is the scientist, engineer, or programmer – or even the economist or jurist – who is highly knowledgeable and competent in his own field but possesses only a passive knowledge of a given foreign language.

In the first case, the computer that is to function as a translation tool will have to provide not only the necessary linguistic elements – such as semantic units, terms, syntagms, or phrases – of the special-subject language, but also as much information as possible on the subject itself. In the second case, the linguistic information is of primary interest to the user. This information embraces not only orthography, including diacritics, but also parts of speech, gender, prepositional usage, and the like. Only when these heterogeneous requirements have been met can a system of machine-aided translation be employed universally with success.

What happens in the extreme situation in which the user of the terminology data bank has no extensive knowledge of the field being dealt with? In analyzing the source-language text, he will have difficulty recognizing syntagmatic units. He will likewise be ill-equipped to distinguish between endocentric and exocentric syntagms or idioms. A semantic entity is endocentric if its meaning is identical with the sum total of the meanings of its constituents, and exocentric if it is not. In the first case, the translator can often transfer the constituents of an entity individually from source to target language and link them together without impairing the entity's meaning. In the case of exocentric entities, which are often idiomatic and occur frequently in special-language contexts, he cannot.

If the translator does not recognize a semantic entity as such in the source language, but tries to understand the text by linking the meanings of the entity's constituents, the result will be a misunderstanding of the concept. The target-language version of the text will then contain misinterpretations of concepts, often rendering it incoherent or even incomprehensible.

But even if we leave idiomatic expressions aside and limit our

observations to syntagmatic structures in general, comparative analysis of special-subject language, whatever the field, reveals considerable variance in "semantic construction principles" and concepts between natural languages. Not surprisingly, the prerequisites and rules for transferring the syntagmatic units of special-subject language from one natural language to another are just as variable. However, it is generally true that specialist terms which have evolved in a given language in connection with scientific or technological development in innovation-intensive sectors are introduced directly into other languages by a process of "transliteration." A case in point is data processing, a sector in which the U.S. acquired a huge headstart over Europe and the U.S.S.R.

The result is illustrated below by some compounds or multiword terms from this field in English and German:

binary coded character
binär codiertes Zeichen

defining word mark
begrenzende Wortmarke

library element
Bibliothekselement.

This phenomenon is observed mainly in phases of primary research and development. Once an innovation has become common knowledge in a region, original terms will be coined in the language of that region.

A few examples from the field of power engineering will show how special-language terms that have been coined in different languages from the most disparate points of view still convey the same semantic content to specialists, but not to pure linguists:

dynamometer test
input-output test
Belastungsverfahren

off-peak load
Belastung außerhalb der Spitzenlastzeit

flame drilling
Sauerstoffbohren

power-factor angle
Phasenverschiebungswinkel

lifting nut
Korbmutter.

As these simple examples confirm, it is essential for a system of machine-aided translation to provide the user with complete information of the type which helps him to recognize that a

queried word may in fact, in a given context, be only a constituent of a syntagmatic element, only a part of a larger language entity in the source-language text. A word-for-word translation of multiword terms or constituents of compounds is a dangerous practice and seldom results in the correct rendering of special-language terms and syntagms in the target language.

This consideration is crucial for the user in the opposite camp as well, i.e., the specialist who is versed in the grammar and rules of a foreign language, but is unfamiliar with its idiom. The provision of complete information on the term or syntagm – its concept, its usage, and its environment – helps him to penetrate the idiom of specialist jargon and discover the correlation between semantic entities in the source language and their equivalents in the target language.

One of the most notorious traits of specialist language is its propensity to proliferate synonyms. Synonyms evolve and are used for many reasons: because parallel developments are pursued simultaneously by different researchers who describe their innovation in their own terms; because there is a need to find differentiating labels for competing products of the same kind; because style in writing demands variety of expression; because usage within the same language varies from region to region (e.g., English in America, Great Britain, South Africa, Australia; French in France and Belgium; Spanish in Spain and the sundry countries of Latin America), etc.

Another difficulty of special language is its numerous homographs. Since the text producers (in the case of multilingual communication, the author-translator team) quite naturally have a different mental set and orientation from those to whom the text is addressed, homographs are ambiguous in the source language. They must therefore be resolved in the translation process. For it is seldom that the ambiguities of a homograph in the source language are congruent with the ambiguities of the equivalent in the target language.

In designing the data base and programming for a system of machine-aided translation that is intended to serve not only translators, but anyone able to act as a link in the process of multilingual communication, it is useful to know the statistical distribution of language elements (terms, syntagms, phrases) which are revealed by the linguistic analysis of special-language texts.

These were a few of the key factors considered in laying the groundwork for the development of the TEAM program system. The voluminous quantity of special-language terminology and its rate of change give rise to other aspects of no less importance.

Since it was obvious from the start that not every future need, application, and technical development could be foreseen, the system was kept open-ended, and was developed pragmatically, step-by-step, with every stage of development being put in practical application immediately.

The same is true of the linguistic data stored. As already mentioned, all possible sources of ambiguity were eliminated. To this end, the linguistic data of the various languages - the terms, syntagms, phrases, etc. - are linked by their concepts or meanings, i.e., by "semantic connectors." Ambiguities are further minimized by supplying supplementary information, such as subject fields, definitions, contexts, explanatory notes, and the like.

2. TEAM program system.

2.1. Entry format.

The entry format used, that is, the structure of the records, which are variable in length, is shown in Pig. 1 and 2. These sample entries have been written on a typewriter and read by an OCR reader for input into the computer. This form of input is preferred because it allows the flexible use of freelance typists without incurring high office overhead. Codes are used for all diacritics to prevent reading errors. The entry is divided into categories from 00 to 99. These numbers enable the computer to distinguish between the various types of information.

These records show the general principle governing the information categories. Single-word terms, compounds, syntagms, or complete phrases can be stored and made accessible via their significant constituents.

The first ten categories form a header which contains the date of the entry, the person responsible for input, the subject field or fields to which the terms in the entry refer and any other information relevant to the entire record. A six-digit alphanumeric code in category 00 is used to identify the entry. This is required whenever entries have to be addressed individually for modification, completion or deletion.

Categories 10 through 19 are set aside for the German term and the items of information referring to it, such as part-of-speech labels, sources, and index or search words in the case of phraseological entries and other explanatory matter. The remaining groups of categories are allotted to English, French, Spanish, Russian, Italian, Portuguese and Dutch. The nineties group can be used at option for passive languages on which only target information can be offered.

Typewritten records are read by an optical character recognition device, and the data is input into the computer and stored on magnetic tape for subsequent checking for formal errors. The terms, syntagms, and phrases can then be retrieved from the data bank.

2.2. System configuration.

Fig. 3 is a schematic diagram of the TEAM program system and its hardware. Paper tapes and punched cards are used only as interrogation media in batch processing. Video terminals and magnetic tape can be used for input in addition to the OCR records described. Video terminals also offer interactive inter-

```
00  AB1235
03  d
04  0179
05 DRUS
06 E4250
10 mischen
11   v.
14   Das Zusammenführen zweier oder mehrerer sortierter
     Dateien von äußeren Speichern einer Datenverarbeitungs-
     anlage zu einer ebenfalls sortierten Datei wird als
     Mischen bezeichnet und von besonderen Mischprogrammen
     ausgeführt. Ein Mischprogramm übernimmt für auf Speichern
     untergebrachte Dateien ähnliche Aufgaben wie ein Loch-
     kartenmischer in der Lochkartentechnik für auf Lochkarten
     enthaltene Daten.
20   merge
21   v.
24   An operation performed on two, or more, ordered sets of
     records to create a single set or file. The two original
     sets must first be arranged into the same sequence by
     sorting on a common key. This operation may be performed
     on punched cards by a collator, or on any ordered file of
     records held in memory or on a backing store medium.
26  collate;v.
50   ob£8edinjat£1
51   v.
56  sme£7sivat£1;v. ; slivat£1;v.
99£a
```

Fig. 1 TEAM entry with terms in German, English and Russian, shown in categories 10, 20, and 50 respectively. Categories 26 and 56 contain synonyms in the corresponding languages. The Russian terms are transliterated according to ISO Recommendations and their diacritics are coded to prevent reading errors.German and English definitions are shown in categories 14 and 24, the serial number of the entry in category 00, the quality code of the entry in category 03 and the subject field in category 06.

```
00  AB1111
3    A
4    1078
5    BIND
6    E4310, E4207
7    DIN 19226 (Ausg. 5.68) (D/E)
8    Dat Switch
9    SIMATIC 99
10   AnaLog/Digital-*Wandler *(ADW)
11   m.
12   FREE.21
14  Setzt analoge E.-Signale in digitale A.-Signale um.
16  Analog/Digital-*Umsetzer;m. ; A-D-Wandler;m .
20   analog/digital *converter *(ADC)
21   n.
22   DICT.43
24   Changes analog signals into digital output signals.
26   analog-to-digital *converter
99£a
```

Fig. 2 TEAM entry containing multi-word terms. The * is a control character to provide for automatic permutation of these multi-word terms or of abbreviations and their long forms. As shown in this example these two possibilities can also be combined. In this case the system generates entries automatically and places them in their proper alphabetical order. For example, in a German list the term "ADW ..." would be shown preceding the abbreviation "Analog ... " and "Wandler ..." will be listed under "W ...".

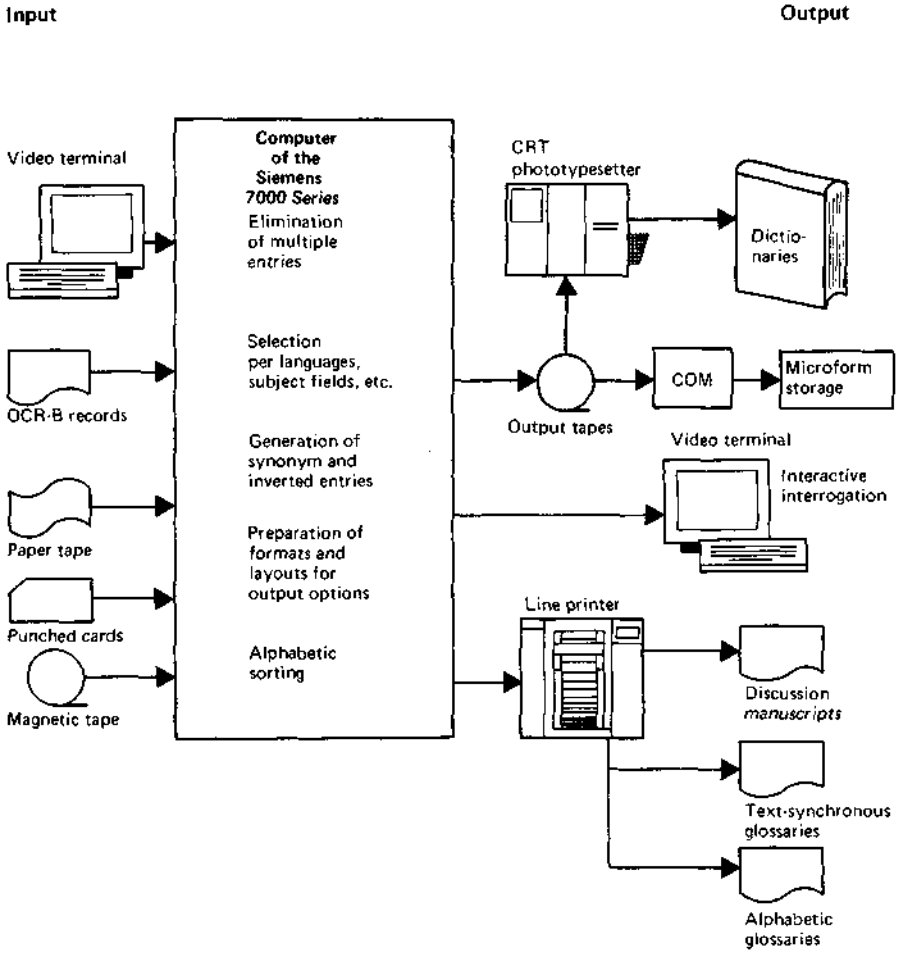

Fig. 3 The TEAM Terminology Data Bank System
A characteristic feature of the system is the variety of its input and output facilities.

rogation facilities. The output facilities of the system are many and varied. Lists can be produced by line printers.

A CRT* phototypesetter makes it possible to typeset entire dictionaries. Microfiches can be produced with the aid of a COM** device.

2.3- Output options.
All output devices, line printers, the COM system and the CRT phototypesetter are directly controlled by the data bank system. This means that the continually updated contents of the data bank can be printed immediately on paper or film in any form required. To illustrate the speed with which this is done, a two-language dictionary with 200,000 entries, whose contents were revised up to the last minute before editorial deadline, was typeset in just a few hours in a format that features running titles, pagination and clusters, etc. as Fig. 4 and 5 show.

In this process all typesetting commands determining layout, type of font, type size and face, and so on, are generated by the computer according to parameters which are input prior to the typesetting run.

Pig. 6 shows and explains part of the classification labels used in the data systems dictionary, a page of which is shown in Fig. 5. It should be noted that these labels are generated from the alphanumeric codes in category 06 of the TEAM entry classifying the subject field or fields to which this entry, i.e. the terms and their concept, belong.

Dictionaries which are produced at such high speed are up to date and must therefore be considered machine aids to translation. Microfiche and lists can be produced in an even shorter time, regardless of the selection, merging, and sorting parameters applied. Thanks to the flexibility of the program, parameters are freely selectable and the output format is in no way dependent on the structure of the input record. The same is true for the data contained in the output.

The TEAM system consists of some 60 individual programs which are modular in character and can be linked to produce any desired result, such as text-synchronous lists, lists in alphabetical order, etc. Examples of these are shown in Fig. 7, 8 and 9.

All stored terminological and phraseological data is likewise accessible via video terminals. The translator can freely determine the languages and language directions in which he wishes to see terms or other semantic elements displayed.After keying in the sought term, he can request to see all target-language entries contained in the data bank in which the term

---

* Cathode ray tube
**Computer output microfilm

P

работа в истинном масштабе времени *DigDV, Sw* / Echtzeitbetrieb *m*, Realzeitbetrieb *m*, schritthaltender Ablauf / real-time operation, real-time working; - **в комплексе с машиной** *DigDV* / direkter Betrieb, rechnerabhängiger Betrieb, On-Line-Betrieb *m* / on-line mode, on-line operation; - **в реальном масштабе времени** *DigDV, Sw* / Echtzeitbetrieb *m*, Realzeitbetrieb *m*, schritthaltender Ablauf / real-time operation, real-time working

работоспособный *adj* / funktionsfähig *adj* / operative *adj*

рабочая лента *Mb* / Arbeitsband *n* / work tape, user tape, scratch tape, utility tape

рабочее состояние *Hw* / Arbeitszustand *m* / operating state

рабочий адрес *Sw* / Bezugsadresse *f* / reference address; - **режим** *SystSw* / Betriebsart *f* / mode of operation, computer control mode; - **файл** *Sw* / Arbeitsdatei *f* / work file; - **цикл** *StW* / Arbeitszyklus *m* / operation cycle

равнозначный оператор *FORTRAN* / Äquivalenzanweisung *f* / equivalence statement

равноценность *f Boole* / Äquivalenz *f* / equivalence *n*, equivalency *n*

разбивать блоки на записи *Sw* / entblocken *v* / deblock *v*

разбивка по слогам / Silbentrennung *f* / hyphenation *n*

разблокировать *v Sw* / entblocken *v* / deblock *v*

развёртывание *n ZchErk* / Abtastung *f* / scanning *n*, sampling *n*

разветвление *n DigDV, Sw* / Sprung *m*, Verzweigung *f* / jump *n*, branch *n*, control transfer

разговорный фортран *PrSpr* / Dialog-FORTRAN *n* / interactive FORTRAN

разгружать *v PeSp* / entladen *v* / unload *v*

раздел банка данных *Info* / Informationspool *m* / information pool; - **данных** *Sw* / Datenteil *m* / data division; - **оборудования** *COBOL* / Maschinenteil *n*, Vorlaufteil *m*, Geräteteil *m* / environment division; - **определений** *COBOL* / Erkennungsteil *m* / identification division; - **процедур** *COBOL* / Algorithmenteil *m*, Prozedurteil *m* / procedure division

разделение по главным группам *Lk* / Hauptgruppentrennung *f*, Hauptgruppenwechsel *m* / intermediate control change, intermediate control break; - **программ на секции** *Sw* / Programmunterteilung *f* / program sectioning

разделительный знак *Sw* / Trennzeichen *n* / data delimiter, information separator; - **маркер** / Bandmarke *f*, Abschnittsmarke (AM) *f* / tape mark, magnetic tape mark

размер данных *Sw* / Datenformat *n*, Datenstruktur *f* / data format; - **страницы** *COBOL* / Seitenbegrenzung *f* / page limit

размерность массива *FORTRAN* / Anzahl der Dimensionen / dimensionality *n*

размерять (напр. библиотечную область) *Sw* / einrichten *v* (z.B. Bibliotheksbereich) / initialize *v*, prepare *v*, preformat *v*

разметка ленты *SystSw* / Vorbereiten eines Bandes (TPINIT) / tape initializer

размечать *v Sw* / etikettieren *v* / label *v*, to write labels on, initialize *v*

разрад *m Math, Theor* / Ziffer *f*, Stelle *f*, Ziffernstelle *f* / digit *n*; - **десятичной дроби** *Math, Theor* / Dezimalstelle *f*, Dezimalbruchstelle *f* / decimal place; **двоичный** - *Theor* / Binärziffer *f*, Binärstelle *f*, Bit *n*, Dualziffer *f* / binary digit, bit *n*; **знаковый** - *Theor, Sw* / Vorzeichenbit *n* / sign bit; **младший** - *Math, Theor* / niederwertiges Bit / low-order bit, low-order digit; **самый младший** - *Math, Theor* / niedrigstwertiges Bit / lowest-order bit, least significant bit, lowest-order digit; **старший** - *Sw* / höchste Bitstelle / high-order bit position, most significant bit position;

Fig. 4 A page from a Russian-German-English dictionary, showing the automatically made-up page with pagination, running title, and clusters.

193 **perforation**

pass *n* [PuCa] Durchlauf *m*, Kartendurchlauf *m*, Arbeitsgang *m*‖ ~ [ProLan] Durchlauf *m* (Übersetzer)
passband, channel ~ [Transm] Kanaldurchlaßbereich *m*
passive station [Transm] Wartestation *f*‖ ~ **station** [RePro] Wartestation *f*‖ ~ **transducer** [ProcDP] passiver Wandler
password *n* [Sw] Kennwort *n*, Schlüsselwort *n*‖ ~ **protection** [Prot, Sw] Kennwort-Dateischutz *m*‖ **computer** ~ [Sw] Rechnerkennwort *n*
paste-up *n* [COM] Klebemontage *f*
patch *v* [Sw] korrigieren *v*‖ ~ *n* [Sw] Korrektur *f*, Phasenänderung *f*‖ ~ **bay** [PuCa] Schaltplatte *f*, Schaltfeld *n*, Stecktafel *f*‖ ~ **card** [Sw] Korrekturkarte *f*, Änderungskarte‖ ~ **card** [Sw] Korrekturkarte *f*‖ ~ **loader** [Sw] Korrekturlader *m*‖ ~ **panel** [PuCa] Schaltplatte *f*, Schaltfeld *n*, Stecktafel *f*‖ **alternative** ~ [Sw] Wechselkorrekturbefehl *m*
patchboard *n* [PuCa] Schaltplatte *f*, Schaltfeld *n*, Stecktafel *f*
patchcord *n* [PuCa] Steckschnur *f*
patching jackfield [Hw] Stöpselfeld *n*
path data [NC] Weginformationen *f,pl*‖ ~ **selection** [Switch] Wegeauswahl *f*‖ **short** ~ [Switch] Kurzweg *m*
patient monitor [IO] Patienten-Überwachungsanlage *f*
pattern of reflection [Cyb] Reflexionsmuster *n*‖ ~ **recognition** [OCR] Flächenmustererkennung *f*, Strukturerkennung *f*
pattern-sensitive fault [Sw] datenabhängiger Fehler, datenbedingter Fehler
pause exit [Switch] Pausenausgang *m*‖ ~ **message** [Switch] Pausenmeldung *f*‖ ~ **statement** [FORTRAN] Pausenanweisung *f*‖ ~ **supervision** [Switch] Pausenüberwachung *f*
PBX (s. private branch exchange)
pc board [Circ] Leiterplatte *f*‖ ~ **board with floating output** [RePro] Schaltungsplatte mit potentialfreier Ausgangsschaltung
PCI (s. program-controlled interrupt)
PCM (s. punched card machine)‖ ≙ (s. pulse code modulation)
P-counter *n* [Arithm] Befehlszählregister *n*, Befehlszähler *m*
PDM pulses [ProcDP] längenmodulierte Impulse
peak amplitude [Switch] Spitzenamplitude *f*‖ ~ **hours** [RePro, Switch] Hauptverkehrsstunden *f,pl*‖ ~ **load** [DigDP] Spitzenbelastung *f*‖ ~ **load periods** [RePro] Hauptverkehrsstunden *f,pl*‖ ~ **load periods** [RePro] Stoßzeiten (Stoßbetrieb)‖ ~ **traffic hours** [RePro] Hauptverkehrsstunden *f,pl*
peck feed [NC] Pick-Vorschub *m*
pecker *n* [ProcDP] Abfühlstift *m*
peek-a-boo *n* [PuCa] Sichtprüfung *f*, Sichtkontrolle *f*, Blickkontrolle *f*
Peirce function [Boole] NOR-Funktion *f*, NOR-Verknüpfung *f*, Peirce-Funktion *f*
pel *n* [OCR] Bildelement *n*
pen, electrographic ~ [PuCa, OCR] stromleitender Grafitstift, Elektrografitstift *m*
pencil mark [OCR] Strichmarkierung *f*, Bleistiftmarkierung *f*‖ **conductive** ~ [PuCa, OCR] stromleitender Grafitstift, Elektrografitstift *m*
pending interrupt [Sw] anstehende Unterbrechung, schwebende Unterbrechung‖ ~ **interruption** [Sw] anstehende Unterbrechung, schwebende Unterbrechung
penetration *n* [OCR] Anschlagstärke *f*‖ **fault** ~ **range** [Switch] Fehlerwirkungsbreite *f*
pentade *n* [Theor] Pentade *f*
PER (s. program event recording)
percent sign [Sw] Prozentzeichen *n*
perception *n* [Cyb] Wahrnehmung *f*
perforate *v* [PuTa, PuCa] lochen *v*, stanzen *v*
perforated paper tape [PuTa] Lochstreifen *m*‖ ~ **tape** [PuTa] Lochstreifen *m*‖ ~ **tape equipment** [PuTa] Lochstreifengeräte *n,pl*‖ ~ **tape reader** [PuTa] Lochstreifenleser *m*, Lochstreifenabfühler *m*, Lochstreifenabtaster *m*‖ ~ **template** [NC] Lochschablone *f*
perforating *n* [IO] Ablochen *n*
perforation *n* [PuTa, PuCa] Loch *n*, Perforation *f*, Lochung *f*‖ ~ **rate** [PuCa]

Fig. 5 A page from the Data Systems Dictionary, an example for data bank-controlled CRT phototypesetting.

Classification Labels Sachgebietsschlüssel

ALGOL	ALGOL (Programming language)	ALGOL (Programmiersprache)
AnDig	Analog-to-digital converters	Analog-Digital-Umsetzer
AnDP	Analog data processing, analog computers	Analogverarbeitung, -rechner
AnSt	Analog storage devices	Analogspeicher
Arithm	Arithmetic unit	Rechenwerk
ASSEMBLER	ASSEMBLER (Programming language)	ASSEMBLER (Programmiersprache)
Autom	Automata theory	Automatentheorie
Boole	Boolean algebra	Schaltungsalgebra
CAT	Computer-aided instruction	Computerunterstützte Unterweisung
Circ	Circuit technology	Schaltkreistechnik
COBOL	COBOL (Programming language)	COBOL (Programmiersprache)
COM	Computer output to microfilm	Computerausgabe auf Mikrofilm
ConsTy	Console typewriters	Blattschreiber
ComSw	Commercial and administrative programming	Kommerzielle/administrative Programme
Contr	Control unit	Steuerwerk
CU	Central (processing) unit	Zentraleinheit
Cyb	Cybernetics	Kybernetik
DigDP	Digital data processing, digital computers	Digitalverarbeitung, -rechner
DP	Data processing	Datenverarbeitung
FORTRAN	FORTRAN (Programming language)	FORTRAN (Programmiersprache)
Hist	History of information processing	Geschichte der Datentechnik
Hw	Hardware	Hardware
HyCom	Hybrid computers	Hybridrechner
Info	Information systems, documentation	Informationssysteme, maschinelle Dokumentation
IO	Input/output devices	Ein-/Ausgabegeräte
KeyTT	Key-to-tape stations	Magnetbandbeschriftungsgeräte
MaTransl	Machine-aided and machine translation	Maschinenunterstützte/maschinelle Sprachübersetzung
MCa	Magnetic card storage devices	Magnetkartenspeicher
MDi	Magnetic disk equipment	Magnetplattengeräte
MDr	Magnetic drum storage devices	Magnettrommelspeicher
MMem	Main memory	Arbeitsspeicher
MTa	Magnetic tape equipment	Magnetbandgeräte
NC	Numerical control	Numerische Steuerung
OCR	Optical character recognition	Optische Zeichenerkennung
OR	Operations research	Unternehmensforschung
OS	Operating systems	Betriebssysteme
Periph	Peripheral storage devices	Periphere Speicher

Fig. 6 Classification labels used in the Data Systems Dictionary, with explanations in English and German.

P31 Bedienungsanleitung /f. / /E4220, E4250 /AI9846

1 instruction manual
FACH E5380, E9320

2 operating guide
DEF: A description containing all functions required to run a system. /RUSC/
FACH E4010, ED240

3 ..operating manual
SYN: ..operator's handbook
FACH E4380

4 operator's instruction booklet
FACH E4370, E4385

5 Operators' Guide
FACH E4250

232 Bedienungsanweisung /f. / /E4256 /AI8761

1 operating instructions
FACH E5380, E9320

233 Bedienungsblattschreiber /m. / /E4100 /AI8762

1 console *typewriter
DEF: A combination of console printer and keyboard for operator/system communication.
FACH E4000, E4130, E4101, E4222, E4490

2 console typewriter
SYN: operator's *console
FACH E4101

3 console typewriter
FACH E4380

4 operators' teleprinter
FACH E4380

Fig. 7 Result of batch interrogation of the data bank (queries in alphabetical order; text-synchronous arrangement of questions and answers is also possible).

```
Drehmelder   m.                                  AE9010
      QUE.      VDI   3257                      d 0968
      KAT8  PrDV., AnDV.                          0183
   E: synchro n.
      QUE.  VDI 3257
      SYN.  selsyn;n.; magslip;n.
   F: synchro m.
      QUE.:  VDI 3257
      SYN.:  transformateur tournant;
          transformateur rotatif
   I: trasmettitore sincrono
   .QUE.:  HAND
      SYN.  magslip;m.
   KATG.02  12
   E4221, E4260


Drehmomentverstärker m.                          AD2168
      QUE.: IFIP; A902-35                       d 0173
      KAT8: MB., AnDV.                            DU10
   E: torque *amplifier
      QUE.: IFIP; A902-35
   I: amplificatore di coppia
      QUE.: HAND
   KATG.02: 12
   E4260, E4114


*Drehmomentverstärker , Andruckrollen-~ m.    AQ3505
      QUE.: A 902-34                            d 1176
      KAT9: PrDV.                                 UE10
   E: capstan type *torque *amplifier
      QUE.: A 902-34
   E4221


Drehrichteranlage f.                             AW6406
      QUE.: OBERT                               A 0775
      KAT8: DSw.                                  9999
   E: .,3-phase power *inverter
      QUE.: OBERT
   KATG.09: EDS (1975)
   E4391
```

Fig. 8 Data bank printout used as discussion manuscript in terminology work. The right-hand column contains information such as the identification code of the lexical unit.

console n. DEF: That part of a computer used for communication between the operator and the computer.	**Bedienungsplatz (Konsole) m.**
console *operator	**Operator m.** SYN: Operateur; m.
console *printer DEF: A printer used for system-to-operator communication.	**Konsoldrucker m.**
console *processor DEF: An autonomous picoprocessor with its own arithmetic and logic unit (ALU) and memory. It controls the operation of display, keyboard, control panel, and control memory loading device. /Unidata product/ SYSTEM: 7.000	**Bedienungsprozessor m.**
console typewriter DEF: A combination of console printer and keyboard for operator/ system communication.	**Bedienungsblattschreiber m.**
***console, central operator**	**zentraler *Bedienungsplatz**
***console, main** DEF : +	**Hauptbedienungsplatz m.**
contention mode DEF: A communication procedure allowing two or more stations to compete for use of the line.	**Konkurrenzbetrieb (DIN 44302E Nov. 1973) m.**

Fig. 9 Glossary covering a special subject field and serving as a basis for terminology standardization in the translation, documentation, and sales departments etc .

appears. Compound expressions can be interrogated under any significant element of the term or phrase. He will then be shown the sum total of all relevant entries and can page through the entire inventory.

2.4. Cooperation in terminology acquisition and evaluation. Special-subject terminology is voluminous in every sector of knowledge. By the beginning of this century, the Association of German Engineers (VDI) had already compiled some 3-5 million technical terms. Vocabularies of special-subject languages grow and change rapidly in direct proportion to the rate of innovation- in science and technology, and the rate of change in other fields of learning.

The practical value of any collection of terminology data, which is the basis of machine-aided translation in special-subject languages, depends on its comprehensiveness and currency. This in turn depends on how successfully the living language of special subject areas is recognized, evaluated, recorded and stored. This is not a task that can be handled by individual language services acting alone. It requires the cooperation of many translators and experts who are versed in numerous languages and disciplines and who possess widely differing orientations. As a consequence, the Language Services Department of Siemens AG has for years worked successfully in cooperation with highly competent language service departments of other companies and government agencies, with independent translation bureaus, with technical associations - indeed, with all types of linguistically qualified groups specialized in various subject fields or working in heterogeneous areas. A joint venture operating on the principle of mutual give and take has resulted from this effort. In it, terminology is evaluated in accordance with established quality criteria, recorded, and stored in a common TEAM System data bank.

In this way, every member of the venture adds his contribution of special-language data to a common intellectual fund which is at the disposal of all members.

In addition to the contents of this common terminology data bank the TEAM program facilities are made available to all participants. Input records are generally written on OCR typewriters by the members and the Language Services Department of Siemens AG provides for their input and processing. All partners are supplied with lists, microfiches or tapes as desired. Self-employed translators and those working in private translation bureaus play an important part in overcoming language barriers.

Special-subject translators who fall into this category could make a valuable contribution to the expansion and enhancement of the common data bank, while at the same time profiting from its shared resources, which at present comprise well over a million entries of various types in the languages dealt with. Membership in the joint venture portrayed here has the advantage of providing effective machine aids to translation and

making special-language expertise available in great breadth and depth, without requiring a large investment of funds.

5. Small computers and machine aids to translation. Thanks to new technology, computers have become not only more efficient and faster, but also less expensive and much smaller. Word processors are a result of this development. They are very useful tools for the translator and can even provide machine aids to translation.

Word processors are essentially small text processing and editing machines equipped with video screens and printers. Their textual data is stored on floppy disks, cards or tape. The Siemens Text System 580, for example, uses two floppy disks capable of storing 240,000 characters. Part of one floppy disk is occupied by programs. The remaining space on this disk is available for text which can be edited, modified, updated, etc. These features are useful in all stages of translation and revision. Terminology and phraseology can be stored on the second disk and retrieved during translation. Text-related lists or terms in alphabetical order can likewise be retrieved from this floppy disk to aid the translation process.

Text modules can, of course, be stored in the same way and retrieved during translation.

The Siemens Text System 580 can be connected to the Siemens Program System TEAM as shown in Fig. 10. Any terminology required by the translator using a Text System 580 can be selected from the TEAM data bank and transmitted via telephone line to the word processor where it is stored on floppy disks. Using word processors in this way represents an inexpensive machine aid to translation which is well suited for small translator groups and translators working on their own. They could, for example, set up disk libraries containing special-language terminology in their own fields and access them whenever required.

For these types of translation support a central data bank is used where all terminology is continuously updated, processed and output according to the needs of the individual partners. Solutions of this type are desirable not only because of the enormous volume and rapid rate of change characterizing special-subject languages, but also for reasons of cost effectiveness .

It is of course also possible to produce machine aids to translation directly, in the form of floppy disk libraries, with the help of a word processor. By this method the user can himself collect, edit and update his term-and-phrase data and store it on floppy disks. It can then be retrieved as required. This solution naturally imposes restrictions of computing power and storage capacity as compared with large computers.

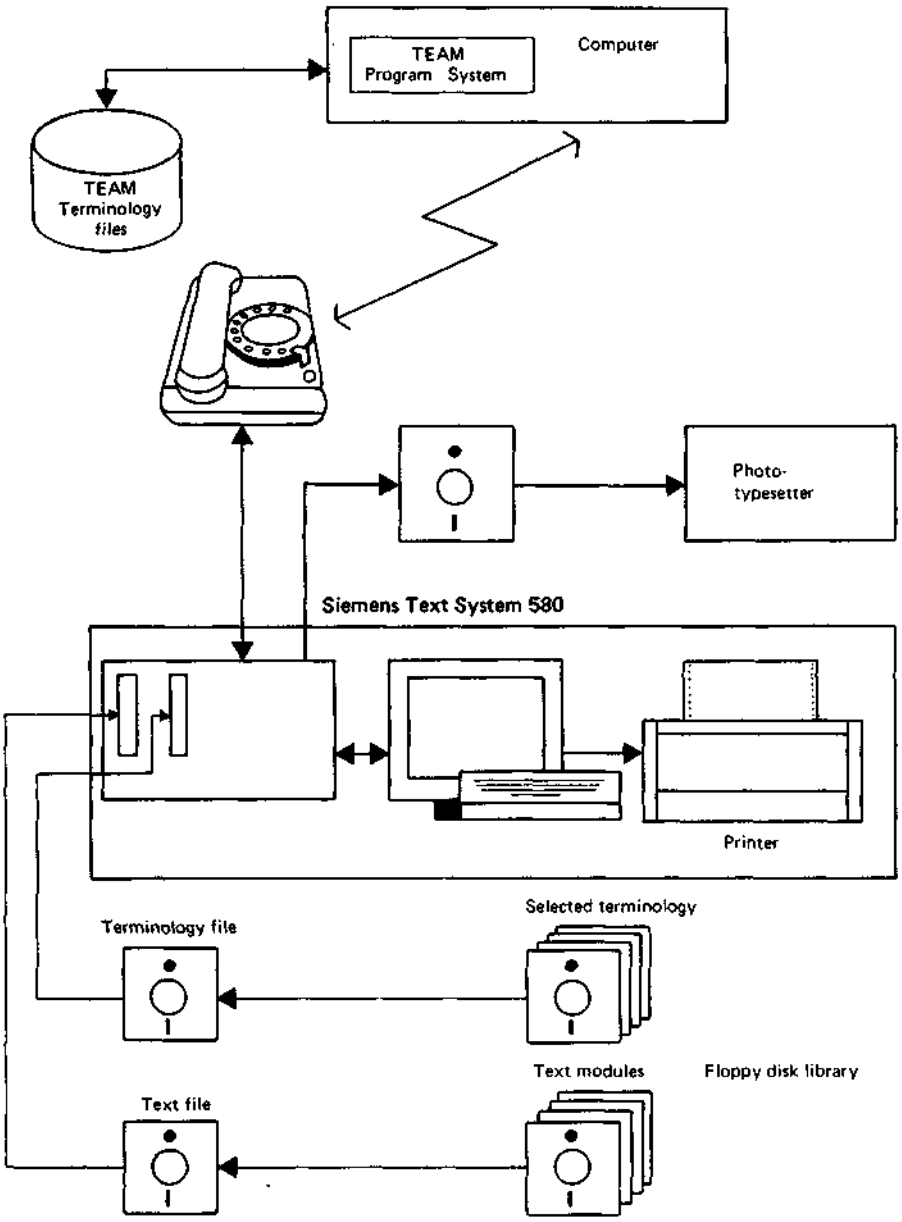

Fig. 10 Interconnection of TEAM data bank system with Siemens Text System 580

4. Combination of TEAM program system with text editor and processor.

If one breaks down the translating process - including revision, intermediate, and final copy typing - into its individual phases, it becomes apparent that some of them are intellectual in nature, while others are mechanical and routine tasks.

By transferring the latter to the computer the overall efficiency of the translation process can be increased and the translator's task will become more rewarding. With this in mind, the TEAM terminology data bank system was combined with a text editor and processor as shown in Fig. 11. On the right is a simplified representation of the TEAM data bank system. It is linked with a text editing and text processing system. Texts can be input using magnetic tape, OCR, or video terminals. The latter can likewise be used for text editing. This, then, is the branch used for translation and revision. Wordprocessors can also be connected to the text system to serve as "intelligent" input or output devices and of course for the editing of text. Standard output is via line printers, CRT phototypesetters, COM equipment, or on magnetic tape.

This combined system makes it possible to select terras and phrases from the TEAM data bank and transfer them to the text editing system. Here, specific pools are built up - for example, on a particular system or subject area - and stored on magnetic disks for direct access. In addition, the terminology data bank of the TEAM system can be directly interrogated via the terminal of the text system. The formation of subject or system-specific pools serves to eliminate terms and syntagms not wanted, such as synonyms and wrong terminology. This is done before actual translation begins. Every text is examined in the source language for uniformity of terminology. It is compared automatically with the contents of the system-specific pools and all deviations from standard terminology are shown on lists together with their line numbers in the text. The general language portions of texts are excluded from this comparison. Text analysis and concordance programs of the TEAM program system are employed for determining what is required in the system-oriented pools. The source language text is transferred from the text editing system to the TEAM program system for analysis. If terms not contained in the system- or subject-oriented pools are discovered in the text, they are entered in both the TEAM data bank and the terminology pools of the text system. They are then available for the translating process. Wrong terms in the source language text are replaced by the proper ones. As a result of text analysis, the work entailed in proofreading is also reduced. This applies for every level of the translating process, from source language to target language text.

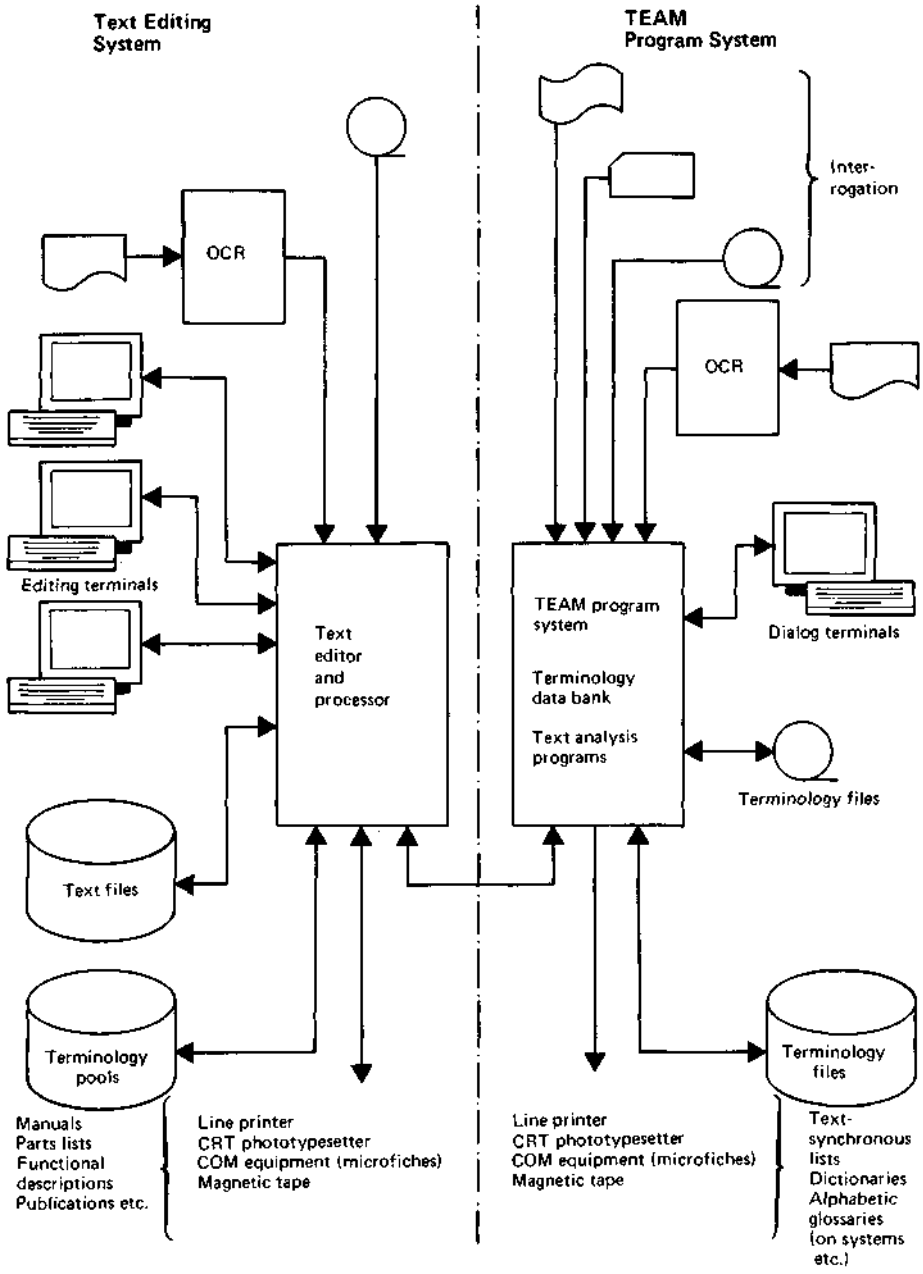

Fig. 11 Interconnection of TEAM program system and text editing system

In any stage of the translation process, including revision and final typing, the text can be changed in various ways. The translated text can also be changed if modifications of the technical equipment, the system, or the like require it. All these functions are supported by the text editor. The text processor, together with the editor, is used when terminology contained in a text is to be changed, or when translations are to be prepared for automatic typesetting.

In this text editor and processor - used in connection with the TEAM terminology data base - all editing functions are controlled on screen. These include typing error correction - where the original character(s) are simply overwritten - and insertion or deletion of text of almost any length - words, lines, or paragraphs.

For the computer to carry out these functions, the translator or revisor enters commands, which can affect entire files, or delimited vertical and horizontal ranges. The vertical range is delimited by screen columns and the horizontal by means of line numbers, the so-called index.

This index can be used to address particular portions of the text, such as lines or paragraphs, for the purpose of restructuring text files, e.g. to change the sequence of text modules. Lines and paragraphs may be copied as often as desired in order to re-use them either in the same text file or in any other file, for example, if the same object or the same facts are described in different places.

A similar method is used when text modules are copied into a text that is being translated or revised. In all stages of the translation process the translator or the revisor may either work directly on screen or on paper. In the latter case a hard copy of the text is output either by way of line printer or a hard copy device connected to the terminal. Any changes entered into this hard copy are transferred into the computer-stored file with the index serving as line address.

If specific strings are to be checked, substituted, or located within a text, a concordance or string search is carried out. This function is initiated by keying in the string, following a short command. The computer then searches the entire file for this string. If so desired, the strings can be automatically replaced by different strings of any length, either one-by-one throughout the entire file, or within delimited ranges. This string search and substitute facility is also used to prepare a source language text when, for example, the text is to be parcelled out to a number of translators. Before the actual translation process begins, special language terms, phrases, etc. - i.e. strings - are inserted into the source language text.

In this case the term or syntagm in the source language text, i.e. the string sought, is not overwritten but the wanted target language equivalent is inserted following the source lan-

guage concordance. The target strings are obtained from the special subject terminology pool which is frequently updated. This search/substitute process may again be carried out one by one. If so desired, the string sought can also be overwritten. However, this method is disadvantageous, since the translator is then no longer able to resolve ambiguities such as homographs .

Even if a source language text comprises thousands of lines with frequent occurrences of the sought string, this operation, if carried out in one go, only takes a few seconds. This process is illustrated by the following text which is first shown in its original form and then a second time after pre-editing, that is, after the target language strings have been inserted.

The text editor and processor greatly facilitates text editing functions. For example, long character strings, such as modules, which are stored in the computer can be identified and retrieved by a specific character or character combination. Each time the translator or revisor enters this character or character combination in a text, the computer will automatically insert this long string.

All these facilities do, of course, necessitate file protection, so that text cannot be destroyed accidentally or altered intentionally without authorization of the responsible translator or revisor.

As already mentioned, output can be directed to various output media as desired, such as hard copy printers - printing one screen after the other - line printers, magnetic tapes, discs, and phototypesetters. If, following translation and revision, typeset output is requested, the text editor and processor offers additional features. A copy of the target language file is analyzed by the computer and according to input format criteria, typesetting commands are automatically inserted. A specific output format including font is then produced. This output format may, however, be changed by simply altering a few parameters that interpret the text format analysis programs .

The Text Editor and Processor **Textbearbeitungs- und -Verarbeitungssystem** greatly facilitates text editing **Textbearbeitung** functions. For example, long character strings **Zeichenfolge**, such as modules, which are stored in the computer can be identified and retrieved by a specific character or character combination. Each time the translator or revisor enters this character or character combination in a text, the computer will automatically insert this long string.

All these facilities do, of course, necessitate file protection, so that text cannot be destroyed accidentally or altered intentionally without authorization of the translator or revisor.

As already mentioned, output can be directed to various output media as described, such as hard copy printers **Bildschirmdrucker**, - printing one screen after the other - Line printers **Schnelldrucker**, magnetic tapes **Magnetband**, discs **Magnetplatte** and phototype-setters **Lichtsatzanlage**. If, following translation and revision, typeset output is requested, the Text Editor and Processor **Textbearbeitungs- und -Verarbeitungssystem** offers additional features. A copy of the target language file is analysed by the computer and according to input format criteria, typesetting commands **Satzbefehl** are automatically inserted. A specific output format including font **Schriftart** is then produced. This output format may, however, be changed by simply altering a few parameters that interpret the format analysis programs.

After these typesetting commands have been inserted in the text, a magnetic tape is produced which controls the CRT phototype-setter DIGISET. As an end product, properly laid out text, including running titles, automatic pagination, syllabification, etc., is received, either on photo paper or on film.

In the next stage of program development, this procedure will be automated. All special-language terms and all syntagms contained in the text will be found by the computer. The terminology data pools of the text system will be interrogated, and will supply the target language equivalents, automatically. These in turn will be automatically inserted in the source language text in the manner described. This text will be the one used for translation.

5. Concluding remarks.

The multilingual communication of transnational organizations whose operations are worldwide cannot be handled solely by the limited facilities of their translation offices and language services. Specialists in foreign sales and marketing, and of course all employees working abroad, play a major role in the process of multilingual communication. It is obvious therefore that a common unifying language basis is required to assure that the transparency of special-language communication is preserved. This common basis is the TEAM program system providing all participants in the communication process - i.e., both specialists and translators - with mechanical support for their translating activities.

The TEAM program system is likewise used to prepare crash courses aimed at developing the foreign-language reading ability of employees. The assumption underlying these courses is that the specialist will understand foreign-language texts dealing with his own field, if he possesses a minimal vocabulary of combined special and general language terms and is acquainted with the key morphological and syntactical charac-

teristics of the special language. To determine this minimum, the TEAM system's statistical analysis programs are employed. For example, the most frequent concordances in a large volume of textual material are identified and then made available in two languages for study by interrogating the data bank.

The machine aids to translation described here are employed by Siemens AG. They are being further developed with the goal of relegating to the computer all routine and mechanical tasks in the language sector, thus allowing translators and all others engaged in multilingual communication more time to devote to its intellectual and creative aspects.

Note:
Because of the difficulty of reproducing actual computer printouts, these have been simulated with an OCR typewriter.

References:

1. Brinkmann, K.H., Schulz, J., Tanke, E.: Das Wörterbuch aus der Maschine. data report 4 (1969) No. 4

2. Brinkmann, K.H.: Überlegungen zum Aufbau und Betrieb von Terminlogie-Datenbanken als Voraussetzung der maschinenunterstützten Übersetzung. Nachrichten für Dokumentation 25 (1974) No. 3

3. Brinkmann, K.H.: The TEAM Program System. Philips Terminology Bulletin, Vol. 4 (1975) No. 2/3

4. Brinkmann, K.H.: Perspektiven der maschinellen Übersetzung. Nachrichten für Dokumentation 29 (1978) No. 3

5. Göricke, H., Schulz, J.: The Dictionary in the Computer - Possibilities of Directly Interrogating a Multilingual Terminology Data Bank via Video Display Units. Babel - Revue international de la traduction, No. 7 (1977)

6. Schmidt, R. , Vollnhals, 0.: Der Einsatz des lexikographischen Zweigs eines Datenbanksystems zur Herstellung eines phraseologischen Fachglossars. Lebende Sprachen XIX (1974) No. 3-4

7. Schulz, J.: Le Systeme TEAM, une aide à la traduction. Actes du colloque international de linguistique et de traduction, Montréal 30.9. - 3-10.1970. META Journal des traducteurs, Vol. 15 (1971) No. 1-2

8. Schulz, J.: Reversible Transliteration kyrillischer Buchstaben. Nachrichten für Dokumentation 24 (1973) No. 6

9. Schulz, J.: Lexikographie *mit* TEAM - Automatischer Satz von Wörterbüchern. data report 9 (1974) No. 1

10. Schulz, J.: Der Computer hilft dem Übersetzer - Ein Verfahren zur Abfrage eines auf Magnetband gespeicherten mehrsprachigen Fachwörterbuches. Lebende Sprachen XX (1975) No. 4

11. Schulz, J.: Automatische Abfrage einer Terminologie-Datenbank. Nachrichten für Dokumentation 27 (1976) No. 1

12. Tanke, E.: Das aktuelle Wörterbuch aus der Datenbank - Ein Beitrag zur Lösung sprachlicher Probleme im internationalen Informationsaustausch. Deutscher Drucker No. 38 (1971)

13. Tanke, E.: Electronic Data Processing in the Service of Translators, Terminologists, and Lexicographers. Philips Terminology Bulletin, Vol. 4 (1975) No. 2/3

14. Tanke, E.: Siemens Linguistic Data Bank. Proceedings - Aslib Technical Translation Group, London (1976)

15. Vollnhals, Otto: Die datenbankgesteuerte Herstellung einer Wörterbuchreihe, Babel - Revue international de la traduction, No. 2 (1978)